UNIVERSITY OF ZAGREB
FACULTY OF ORGANIZATION AND INFORMATICS
VARAŽDIN

UDC 004:005:3
DOI: https://doi.org/10.31341/jios

ISSN 1846-3312
e-ISSN 1846-9418

Journal of Information and Organizational Sciences

JIOS VOL. 48, NO. 1, PP. 1-252, JUNE 2024

JIOS
Journal of Information and Organizational Sciences

UDC 004:005:3
ISSN 1846-3312
e-ISSN 1846-9418
DOI: https://doi.org/10.31341/jios

JIOS
Journal of Information and Organizational Sciences

CONTENTS

VOL. 48, NO. 1, PP. 1-252, JUNE 2024 ISSN 1846-3312

JIOS
Journal of Information and Organizational Sciences

From the Editor

Dear readers,

I had the pleasure of preparing the articles published in the Journal of Information and Organizational Sciences in the last four years. This is my issue, and I want to thank all who helped, every one in its way, the Journal to prosper.

Firstly, I want to thank the authors who recognized JIOS as a decent journal adequate to publish their paper. Thanks to all the reviewers who helped me to classify received papers and raise the quality of the journal. Many of them were members of the editorial board which I want to address my thanks, too. The advisory board helped me to widen the pull of reviewers. At the end, I want to thank members of the publishing board who invested their time and work to improve the design of the journal and web page, and made it visible through the different journal bases.

In the next issue, the function of the editor will be taken by Prof. Igor Balaban. I am sure that the Journal will become much better under his leadership. I wish good luck and pleasant work on our Journal to him and his team.

Alen Lovrenčić

Editor

VOL. 48, NO. 1, PP. 1-252, JUNE 2024 ISSN 1846-3312

10.31341/jios.48.1.1 UDC 004.775
Open Access Preliminary Communication

Relevancy between Anchor Text and Wikipedia: A Web Search Framework

Falah Al-akashi *falahh.alakaishi@uokufa.edu.iq*
Faculty of Engineering
University of Kufa, Najaf, Iraq

Diana Inkpen *Diana.Inkpen@uottawa.ca*
Faculty of Engineering
University of Ottawa, Ottawa, Canada

Abstract

The overall volume of data available on the Internet is growing rapidly while finding relevant documents is becoming increasingly difficult. Moreover, queries entered by users are unique, unstructured and often ambiguous while the process has changed dramatically from standard query languages that governed by strict syntax rules to unstructured strings. In Web information retrieval, search paradigms used term occurrences to weight document content prior to any boosting stage. PageRank algorithm, for instance, was used integrated techniques to enhance post retrieval document relevancy to adequately compromise the overall process in two stages. Nevertheless, hypertexts in Web have been used for improving the quality of search results for the most common type of queries. Our main premise is that hypertexts play an important role for ranking documents in IR such as margining between user queries and consensus hypertext. We propose a new algorithm that uses term impact technology for compromising hypertext weighting in Web along with Wikipedia for efficiently find most relevant documents among large set of results. Our experimental results showed that Wikipedia could efficiently improve document relevancy rank when combined with hypertexts for exhibit robust and very good short-term process capability.
Keywords: World Wide Web, Entity Weighting Schema, Data Fusion, Anchor Text

1. Introduction

Over the past two decades, the benefits of technology advances on the Internet have grown exponentially. Although casual Internet surfing is typically enough to determine the relevancy of documents, there are huge volume of information and innumerable links available online can complicate the privileges of users for processing access to information requests. In addition, search engines faced high volume of difficulty to determine the relevance rank of documents on the Internet, whilst information has become useless unless involved to a validation technique of weighing and sorting. From the first indexing approach in 1990 (Archie) to the modern search algorithms used today, the problem for deciding the relevance of information

remains a critical issue because Web is fuelled by vast troves of data and becomes a surveillance platform and gatekeepers of innovation. An effective way to compute the relevancy of documents is to determine the empirical evidences in the content of the documents using an automated algorithm for predicting web topics. Researchers had developed an efficient method for computing the relevance of hypertext along with a link structure technique to weight a hypertext graph (Yadah et al. 2009). Often, weighting documents is assigned for every on-bound and outbound links on a directed graph, and the inter-links between webpages have been improved to govern the information that target the links for observing the results. We believe such algorithms might be very useful when incorporated with other existing algorithms as web usage mining has been increased exponentially in the last decade. This tremendous collection of web pages along with the degree of topics between websites could pose challenges to the information retrieval systems.

Often, indexing starts by terms have supposedly been appeared more significant. Because search engines are software programs, they work according to the rules established by their designers to determine which terms are usually important in a broad range of documents. First, the titles of pages usually convey specific and helpful information to a specific user/audience. Second, terms positioned at the beginning of documents give more weight as they repeat several times on the document. Third, hypertexts could significantly enhance Web page retrieval by improving the semantic retrieval of Web data (Halpin and Lavrenko, 2011). Fourth, optimizing hypertexts could improve Web ranking and reduce search engine impact traffic. Sometimes hypertext used typically to indicate the subject matter of pages to which they were topically linked (Yadav et al, 2009); for instance, the keywords in hypertexts could improve the relevancy of target pages. This pattern of Web data has been exploited by search engine algorithms to enhance the relevancy of target pages due to some relevancy of keywords appeared on some pages, this was boosted our previous framework (Al-akashi and Inkpen, 2014) using the linked keywords on the target URLs. Traditional search engines, e.g. Google, Yahoo, etc., pretended to exploit the content of hypertexts for a valid indexed page, which configured to index hypertexts in a separate table to make it evident. Recent studies showed that weights given to hypertext have been optimized among traditional search algorithms. The basic technique to materialize a portion of a web page is to navigate from known hyperlinks using regular expressions while hypertexts played a vital boosting to improve the retrieval performance (Craswell et al., 20sf88u11). The experiments have been conducted that ranking based on hypertexts was twice as effective ranking based on document contents in finding the main entry point of a specific topic. Hypertexts are considered relevant and consequently the Page Rank algorithm was applied in hypertext contents to improve document relevancy because hypertexts typically match or share similar properties of titles and that's why documents targeted by different hypertexts. This is a significant indicator that hypertexts acquire high relevant text for targeting the relevancy of documents. Retrieval approaches that leverage extra evidences include a well-known PageRank algorithm (page et al., 1999; Kleinberg, 1999) and HITS (Boytsov and Belova, 2012) to estimate the credibility of pages based on the number of ongoing and outgoing links. However, TREC Web track provided a

subset collection that involved more hyper textual content than the whole collection (Margin and Jaap, 2011). Other studies showed that hypertext annotators somewhat useful for the homepage finding task (McBryan, 1994). Other external evidences such as outgoing links pointed to the retrieval documents were also useful for topic distillation tasks (Chakrabrti et al., 1998). As a result, links and hypertext are important for increasing the relevancy of ranking algorithms but queries entered by users showed more challenges to determine user intent since they focus on multiple viewpoints at different times. Moreover, keywords entered by users might involve problem for distinguishing between some terms spelled similarly when involved different meanings, synonyms, or sometimes not entered by a query string. For instance, a query about 'heart' in the class of diseases might return results related to a term 'cardiac'.

The rest of this paper is organized as follows: Related work will be outlined in Sections 2. Hypertexts and Hyperlinks evidences are elaborated in Section 3. Section 4 will discuss our overall approach. Section 5 will discuss our experimental results. Hypertexts diversification will be demonstrated in Section 6, and finally, conclusion and future insights will be outlined in Section 7.

2. Related Work

Hypertexts are used for various Web information retrieval tasks. When the Internet became established, McBryan (1994) proposed that hypertexts were important to a Web searching. Most current Web search engines use hypertexts as evidences to improve search relevancy ranks. Contextual text in a specific vicinity of the hypertexts would automatically compile lists of authoritative Web resources for a range of topics (Atsushi, 2008). Nadav and Kevin (2003) conducted experiments to investigate several aspects of hypertexts, including their relationship to titles, the frequency of the queries that could be satisfied by hypertexts alone, and the homogeneity of the results acquired by using hypertexts only. They found that hypertexts were typically less ambiguous than other types of texts and enabled more coherent and focused results for queries solved by hypertexts more than for those based on other features in the corpus. In addition to providing a better match than titles or content, hypertexts also have a potential delivery on authoritative search results. Atsushi (2008) proposed modeling hypertexts and classifying queries to identify synonyms of query terms in the hypertexts. He also assumed using synonyms for smoothing purposes could enhance the document relevancy ranks. As hypertext window considered important for other tasks (Davison, 2000; Attardi et al, 1999; Shuming et al, 2009; Ganeshiya and Sharma, 2014); Shuming et al., (2009) proposed that hypertext were not necessarily needed in window to improve relevant/non-relevant documents. Although they approached windows implicitly, they adopted Pseudo URLs and machine learning approaches to exploit the citation relationship and extract pseudo hypertexts for academic articles. They also proposed that the extracted pseudo anchors were useful for improving search performance. Clarke and Cormack (1995) explored the important of hyperlinks and various resources or parameters could be used to effect on-page ranking factors, e.g. in-links and out-links, anchor text, anchor-related text,

etc. The hypertext in a cloud computing paradigm is important since they induced topic selection for addressing the abundance problem (Albi and Silvester, 2017).

3. Hypertext vs. Hyperlink

Hyperlink and hypertext are a fundamental core of the internet and also a foundation of SEO. Using hypertext evidences to rank web documents rather than a document full-text search seems significant and constant for increasing effectiveness of homepage finding task and topic distillation task (Hiemstra and Hauff, 2010). Removing aggregated hypertexts length normalization altogether or normalizing according to a full-text document length was also found to improve the retrieval effectiveness. The most effective usage of hyperlink scores was to reduce the corpus size without decreasing homepage finding tasks. Document evidences should include full-text evidences and other useful document level evidences while Web based evidences might use incoming hypertext and other helpful external document features. For a home page finding task, URL depth features are measured literally for re-ranking documents within URLs' lengths under n% characters, or by adding a normalized URL length score to a query dependent score. The results of such experiments showed that (i) the importance of both hypertext and URL length for home page finding tasks, (ii) both PageRank and in-degree depth performed similarly and are highly correlated, (iii) using hypertext evidences to rank documents rather than document full-text provide significant effectiveness improvements in homepage finding and topic distillation tasks, and (iv) hyperlink recommendation evidences are far less effective than URL based measures.

However, search engines have been used multiple evidences for facilitating the retrieval performance while popular models focused on full-text document content for evidences. Regardless of document content is relevant, other content, such as videos, audios, graphics, etc., were also indexed (Anh and Moffat, 2010). While some approaches focused on hypertext, the visible keywords in hyperlinks and text appeared within bounds of tags when linked to other documents. They could provide search engines and users with relevant contextual information regarding the content of target websites. It was observed that hypertext in web documents were very useful to improve web search quality for most types of queries. Search engines use external hypertext for reflecting the issue how people view pages and what topics of pages they focused on? While websites typically could not control links, hypertext used in websites is useful, descriptive, and relevant. If many websites assumed that a particular page is relevant for a given set of terms, the page might be ranked highly even if terms were not appeared in the text itself. Search engines algorithms have had been developed extremely and they can now identify more metrics to determine relevant web pages. One important metric is a link relevancy, means how a topic in page 'A' is related to page 'B' if two pages are linked together and took users and search engine robots/crawlers from page 'A' to page 'B'. A highly relevant link could improve the rank of both pages 'A' and 'B' for queries related topics. Links pointed to a content related to the topic of source pages often send stronger relevance signals than links pointed to unrelated content. For example, a page about the best lattes in

Seattle might pass a better relevance signal to Google when it links to the coffee shop's websites rather than web sites with pictures of baby animals. Traditional search engines considered different hypertext variations that linked back to the original articles, and used them for additional weights for what topics of articles are and which search queries are relevant. If too many inbound links contain similar hypertexts, it could appear suspicious signs for links were not acquired truly. In general, it is still best practice to obtain and use keywords and text and topics of specific anchors. However, search engine optimization might achieve better results when search for a variety in hypertext phrases rather than similar keywords each time. Current studies proposed that indexing documents based on their term frequency is not adequate to determine their relevancy. Some search engine algorithms use different strategies to increase the impact of documents on some websites rather than others, for example, Google PageRank algorithm used inbound and outbound hyperlink in their experiments (Kamps et al., 2010) whereas impact of terms on document's header was used to weight the associated document's content (Serdyukov and Vries, 2009). However, developing algorithms to improve retrieval effectiveness and answering the prompt queries from large collections is critical especially when algorithms preserved to improve precision and recall of algorithms. Several systems are already used different relevance ranking models than conventional information retrieval models.

4. Our Overall Approach

The TREC Web track coordinators at the University of Twente provided researchers with a dataset from the ClueWeb09 collection named subset (A) which was 500 million documents, approximately involved 2.5 billion hyperlinks and hypertexts in a file of 10 Gigabytes (Cormack, 2010). This was a set of links used experimentally in the articles of 440,678,986 documents, 87% of English documents had hypertexts while the collection of subset (B) was 45,077,244 documents (approximately 90% of collection B). To process such corpus, hyperlinks referred to the same destinations were normalized by discarding the duplicated links from the whole collection. Then, hypertexts and hyperlinks had been compromised to focus on the user's intention. However, researchers showed that indexing only hypertexts outperformed indexing the whole collection.

4.1. Weighting Hypertexts in Web

Hyperlinks in the shared hypertexts could be used to represent the unique URLs for the target documents, but it is important to compute the frequency of each shared link to calculate the number of outbound hyperlinks. To do this, all hyperlinks have been aggregated and combined in a single hash-table using a unique identifier while markup characters (e.g., https://, http://, www) stripped out initially. According to the Page Rank algorithm, the procedure was repeated recursively to compute the inbound and outbound value of each page or URL. From our perspective, computing the impact of each hyperlink is not limited by inbound and outbound values, which means, the impact is not specified by the Web designer solely but also by the user behavior. In

some cases, documents with only few links were visited more often according to the Alexa.com global rank. We approached this fact by weighting the total ranking value of each URL by creating a three featured hash table: hypertext, hyperlink, and its frequency. Recursively, the hyperlinks would aggregate each key shared similar URLs. Table 1 shows an example of frequency and Alexa networking weights of websites in a subset of aggregated outbound links.

Hypertext	Hyperlink	Local Weight	Global Weight
Bbcarabic	bbcarabic.com	0.19	0.61
BBC Arabic	bbcarabic.com	0.13	0.61
BBC News From bbcarabic.com	bbcarabic.com	0.04	0.61
prestige news mission	bbcarabic.com	0.08	0.61
Daily News BBC	bbcarabic.com	0.06	0.61
CNN Arabic	cnnarabic.com	0.12	0.48
CNN News	cnnarabic.com	0.05	0.48
CNN	cnnarabic.com	0.01	0.48
Hawaw	hawaw.com	0.01	0.00
Kids hats	flickr.com	0.06	0.67
handy pouches	flickr.com	0.02	0.67
Shopping bag	flickr.com	0.03	0.67
Arabic news	aljazeera.net	0.70	0.57
Arabic news	alarabiya.net	0.56	0.52

Table 1. Hypertext, local weights and global weights

Anchors that shared similar hyperlinks could combine the corresponding text to improve their weight efficiently. First, we stripped out all inappropriate phrases (e.g., click here) or implicitly contain that phrases since they decrease the impact of target documents. Second, the phrases that shared ongoing links are considered important to determine the keywords and the descriptions of the target documents. They referred to good resources for weighting document keywords rather than weighting document content itself. However, all phrases in hypertexts shared similar ongoing links that combined together to annotate the overall description of hyperlinks in which texts were changed to lowercase. For example, the hypertext "T" in hyperlink "L" was combined together to treat it as a bag of words. However, weight assigned to the similarity between the hypertext "A" and a query string "B" might be different than those between the content of the document "A" and the query string "B". In other words, the importance of hypertexts in the same collection was treated and weighted differently. The relevancy of hypertext "T" that pointed to hyperlink "L" was computed as follows:

$$Score(D) = \beta \sum_{i=1}^{n} W(T_i, L) \quad (1)$$

where β denotes the weighted factor used to normalize the weight of phrase 'T' in anchor 'L', and n is the number of phrases pointed to the target document 'D'.

The following algorithm shows how to combine and seize the similar phrases.

SELECT *d.URL, e.URL, a.LABEL*
FROM Document *d* SUCH THAT
'www.mysite.com' →* *d*,
Document *e* SUCH THAT *d* → *e*,
Anchor *a* SUCH THAT *a.base* = *d.URL*
WHERE *a.href* = *e.URL* **AND** *a.label* = 'label';

Vector Space Model (SVM) was used to map each document on a unique vector to be counted once a time, as shown in Table 2.

Hypertext	Hyperlink	Local Weight	Global Weight
Bbcarabic, bbc, arabic, news, from, bbcarabic.com, prestige, mission, daily, bbcarabic.com	bbcarabic.com	0.50	0.61
Cnn, arabic, news	cnnarabic.com	0.18	0.48
kids, hats	flickr.com	0.06	0.67
Handy, pouches	flickr.com	0.02	0.67
Shopping, bag	flickr.com	0.03	0.67
arabic, news	aljazeera.net	0.70	0.57
arabic, news	alarabiya.net	0.56	0.52

Table 2. Hypertexts combination, local weights and global weights

Figure 1 shows the execution time for indexing 2.5 billion of anchors in 7 days.

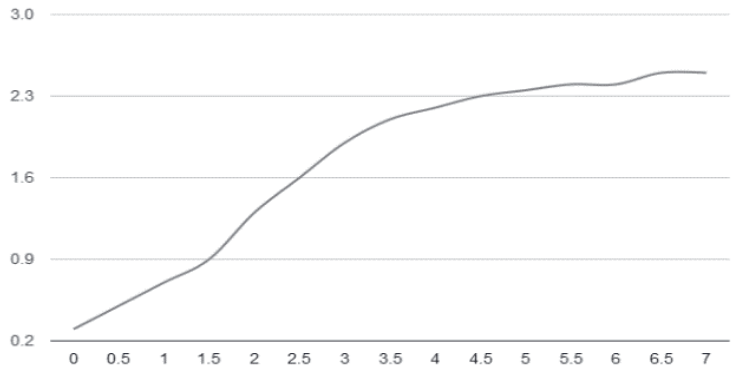


Figure 1. Hypertext indexing period within 7 days

4.2. Weighting Links in Wikipedia

Hypertext in a conventional or classical Web is different from that available in Wikipedia. Wikipedia repository is a significant source for information retrieval since

contains information very worthy. Researchers have had used external links in Wikipedia to successfully retrieve high relevant pages potentially works better than searching hyper information for the same task (Kamps et al., 2010)), and also, its content used for entity and home page finding tasks (Serdyukov and Vries, 2009). Most registered websites are referenced or cited by Wikipedia writers on the External Links and References sections to use URLs to link directly to any web pages. No page should be linked from a Wikipedia article unless its inclusion is justifiable according to the Wikipedia's guideline and copyright role. Longevity of links is a one of the most important feature that considers the link is likely to remain relevant and acceptable to the article in the foreseeable future. However, as Wikipedia is a major part of ClueWeb09 corpus, we approached our index to involve all external web pages. Wikipedia uses a standard appendices and footers template of terms, such as 'Official', 'Website', or 'Home Page', to represent an external website of relevant resource pages (example below). A regular expression was used to parse the literals in the standard terms and extract the corresponding text. Often, the hypertexts refer to the titles of the remote pages which extracted to create a class tag of index nodes. Hence, the indexed node would enclose all target pages referenced by articles alongside with home pages.

```
== External links ==
* [https://example.com Official website]
* [https://example.net/link Interview with Subject] in ‘’Example Magazine’’
* [https://example.org/repository/Subject Subject’s papers] at [[Example Museum]]
```

4.3. Improving Quality and Relevancy

Based on our research, the algorithm in the previous section is not applicable for all purposes since hypertext involves other features particularly when similar hypertext targets different hyperlinks, unlike different hypertexts target similar hyperlinks. As shown in Table 1, 'Arabic news' points to different hypertexts. Alexa Network Global Rank[1] is a rich information repository that weighted resources based on website's visitors. As a result, our approach exploits such worth information for improving the relevancy ranks of all collected hypertexts. We proposed a weighted threshold for each hypertext, if it exceeded a frequency "45", it was discarded. Equation 2 used to strip out the irrelevant hypertexts and to compute the score of relevant documents:

$$P(D)=\frac{|A|^{\pi}}{\sum_{A\in D}|A|^{\pi}} \quad (2)$$

where π denotes ALEXA normalization impact factor on hypertext A.

The final retrieval weight of a document D pointed by a hypertext A is computed as follows:

$$Weight(D, A)=Score\ (D)\ .\ P(D) \quad (3)$$

[1] www.alexa.net

Some irrelevant hypertexts were weighted too high; to address this issue all duplicated terms on the hypertexts alongside with all the following references were stripped out in the earlier stage:

1. **Harmful hyperlinks:** When there an excessive short or long hypertext repeated with some promoting text that targeted some particular sources.
2. **One way hypertext backlink:** When some hypertexts have backlinks to external web pages.
3. **Keyword stuffing**: When hypertexts on a page associate with many resources and target a particular page but with different hypertexts. This causes an algorithm to pay more attention to avoid spam links or black hats (e.g., '*pay day loans*', '*buy now', 'viagra*').
4. **Poor frequency:** Some hyperlinks encountered poor frequency (e.g., 2 or 3), such links typically contain generic text, such as: '*click here*', '*read more*', '*check this out*', '*more info here*', '*login here*', '*email me*', '*admin*', etc.

Hence, the index was minimized to 77,048 tables, in which, each contained an approximately 1,000 vectors (80 Kb in size). Experimentally, we noticed that all hyperlinks pointed to a similar website should be combined in a similar index table and might be shared with other tables. This influenced our approach to use a hash table which is a block index orienting schema for improving the indexing performance and accelerating the query processing time. Each table was represented in a bag-of-words model (BoW) and defined by a particular identifier. The Apache Lucene[2] indexing framework, as a high performance and full-featured indexing library, then used to index all tables. As a result, 77,048 index tables were stored to be used later for a query processing. Figure 2 shows the tradeoff between hypertexts frequencies and types of relevancy. The fallacies of relevancy, spams for example, clearly fail to provide adequate result for true information. Although they often used to attempt to persuade users by irrelevant means, the predisposed users are apt to be fooled.

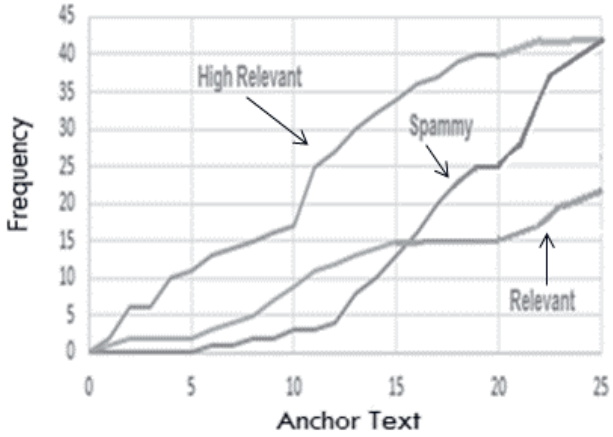


Figure 2. Hypertext relevancy

[2] www.luecne.com

4.4. Query Processing

Our experiments indicated that Apache Lucene is not an ideal tool for hypertext indexing because it deals with text like a bag of words and ignores other related features, such as bigrams and synonyms. Furthermore, the weight given to the similarity between hypertext and query might be different than the weight between documents and query because the relevancy of hypertext and query on document *D* might be different. Apatche Lucene typically uses vector space model for dealing with data as set of vectors and then invokes cosine similarity function for ranking, which is a suitable tool for indexing documents rather than hypertext. We deal with each index as a block content that composited each vector for each document. However, Lucene retrieved block of documents that matched every user's query despite the table implied several vectors (e.g., hypertext, hyperlink, etc.). Thus, we needed to examine vectors sequentially and track it more precisely to determine the relevant ones whose match a query string because the query string might be showed sparsely and distributed over the content of index table. A common way to measure the similarity of two vectors is to compute the similarity distribution of terms; by which, we used a cosine similarity function for vector similarity rather than for block or table similarity. Given two probability distributions, document D and query Q with a set of terms *t*, the cosine distance of two non-zero vectors was derived for two vectors of attributes as follows:

$$\text{Similarity (Q, D)} = \text{Q.D} = ||Q||\ ||D|| = \frac{\sum_{i=1}^{n} Q_i D_i}{\sqrt{\sum_{i=1}^{n} Q_i^2}\ \sqrt{\sum_{i=1}^{n} D_i^2}} \quad (4)$$

where Q_i and D_i are components of vector Q and D, respectively.

Thus, the cosine similarity between a query Q and documents was ranged between 0 and 1.

4.5. Query Refinement

We emphasized how Wikipedia articles are connected to form an online resource in which Wikipedia writers join similar/ related articles using hyperlink. This means Wikipedia articles can expand current topic by transferring the current articles to other articles using interconnected hyperlink (inbound links); similarly, the target articles often point back to reliable articles using outbound links. For instance, if article *A* point to article *B* while article *B* point back to article *A*, we assume articles *A* and *B* are related topically. However, all ingoing and outgoing links in each article are indexed using a custom hash table index platform. The article titles would utilize the keys in the table whilst the outbound links would store the content of the corresponding keys and hence the index would map the query for processing. Let's consider an article *Global Warming* points to the articles *Carbon Dioxide*, *Air Pollution*, *Greenhouse Gas,* and *Alternative Fuel*; these articles would point back to the source article *Global Warming*. Similarly, the article *Automobile* points to three articles while the destination articles points back to the source article. This strategy was used to expand the original query and refine other results from the related topics.

Figure 3 shows an example of expanding a query *Global Warming* by *Carbon Dioxide*, *Air Pollution*, *Greenhouse Gas* and *Alternative Fuel*; whereas, the query *Automobile* was expanded by *Carbon Dioxide*, *Air Pollution* and *Greenhouse Gas*. However, query expansion was not used for all queries but only when the resulting list was short.

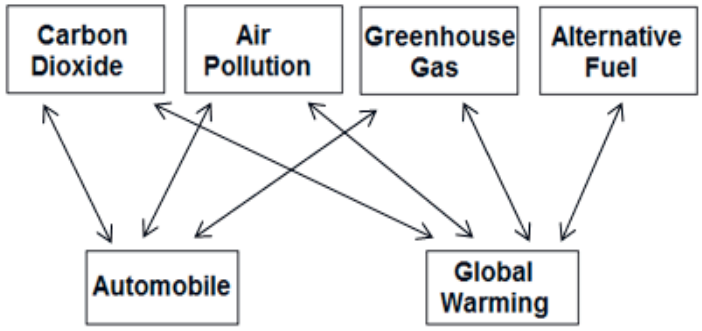


Figure 3. Query Expansion

5. Experimental Results

Often, traditional search engines are judged by some people and candidates. Human relevance judgment is popular for evaluating search algorithms in the market. The TREC Web Track queries were used to compare the results of our algorithm with the results of other algorithms in a task that annotated the best algorithm for each test set using relevant judgement. Our algorithm was filtered out scams and spams using an impact factors identified by ALEXA networking, in which, a document with a ranking score greater than the threshold 0.4 was considered a spam, therefore, discarded from the final results (Cormack, 2010). Table 3 and Table 4 show our best results compared with the results from the related algorithms. We used a precision score at different recall levels P@k to denote a proportion of relevant items amongst other items located on the first k retrieved documents, while Main Average Precision (MAP) denotes the effectiveness performance of adhoc retrieval tasks. The effectiveness performance of diversity tasks[3] for the same queries is shown in Table 5 and Table 6. Our run achieved a significant and substantial improvement in the effectiveness and relevance of retrieval performance in regard to 50 test queries (Table 7).

Run	Description	MAP	P@5	P@10	P@20
Muadanchor	Anchor only	0.0256	0.296	0.250	0.179
Our system	Anchor only	0.0293	0.321	0.284	0.307

Table 3. Effectiveness performance of Web track Adhoc tasks (query list 2011)

[3] The diversity task asked to return a ranked list of pages that together provide complete coverage for a query, while avoiding excessive redundancy in the result list.

.

Run	Description	MAP	P@5	P@10	P@20
Uma10BASF	Anchor + content	0.088	0.383	0.356	0.321
Uma10IASF	Anchor + content	0.087	0.394	0.358	0.319
Our system	Anchor only	0.113	0.471	0.477	0.421

Table 4. Effectiveness performance of Web track Adhoc tasks (query list 2012)

Run	nDCG@5	nDCG@10	nDCG@20	P-IA@5	P-IA@10	P-IA@20
mudvimp	0.220	0.241	0.268	0.091	0.073	0.061
Our system	0.227	0.411	0.392	0.131	0.112	0.195

Table 5. Effectiveness performance of Web track diversity tasks (query list 2011)

Run	nDCG@5	nDCG@10	nDCG@20	P-IA@5	P-IA@10	P-IA@20
Uma10BASF	0.275	0.336	0.379	0.162	0.152	0.131
Uma10IASF	0.281	0.335	0.380	0.165	0.144	0.130
Our system	0.311	0.370	0.472	0.201	0.179	0.218

Table 6. Effectiveness performance of Web track diversity tasks (query list 2012)

Test Queries	P@1	P@5	P@10	P@20
403b	1.00	0.80	0.60	0.60
angular cheilitis	1.00	0.60	0.70	0.50
Pocono	0.00	0.40	0.20	0.40
Figs	0.00	0.60	0.30	0.20
last supper painting	1.00	0.40	0.30	0.60
university of phoenix	1.00	1.00	1.00	1.00
the Beatles rock band	1.00	0.60	0.30	0.10
Grilling	1.00	0.60	0.50	0.60
furniture for small spaces	0.50	0.20	0.60	0.70
Dnr	0.00	0.60	0.50	0.80
Arkansas	1.00	1.00	1.00	1.00
hobby stores	1.00	0.80	0.70	0.40
blue throated hummingbird	0.00	0.60	0.30	0.15
computer programming	1.00	0.40	0.60	0.60
Barbados	1.00	1.00	1.00	1.00
Lipoma	1.00	0.40	0.80	0.80
battles in the civil war	1.00	0.40	0.70	0.80
Scooters	0.00	0.30	0.30	0.30
ron howard	1.00	1.00	1.00	1.00
becoming a paralegal	1.00	0.40	0.70	0.70
hip fractures	1.00	0.90	0.90	0.90
septic system design	1.00	1.00	0.60	0.50
rock art	1.00	0.60	0.50	0.40
sings of a heart attack	1.00	0.80	0.80	0.50
weather strip	1.00	0.60	0.40	0.30
best long term care insurance	1.00	0.90	0.80	0.70
pork tenderloin	1.00	1.00	0.70	0.70

black history	0.00	0.70	0.80	0.60
new york hotels	1.00	1.00	1.00	1.00
old coins	1.00	1.00	0.80	0.70
quit smoking	1.00	1.00	0.80	0.80
kansas city mo	1.00	1.00	0.80	0.70
civil rights movement	1.00	0.20	0.60	0.60
credit report	1.00	0.80	0.60	0.40
Unc	1.00	0.80	0.90	0.80
Vanuatu	1.00	1.00	1.00	0.80
internet phone service	1.00	0.70	0.60	0.40
gs pay rate	1.00	0.50	0.60	0.50
brooks brothers clearance	0.00	0.20	0.50	0.40
churchill downs	1.00	0.60	0.60	0.60
condos in florida	0.00	0.40	0.50	0.60
Porterville	0.00	0.70	0.50	0.20
dog clean up bags	0.00	0.30	0.50	0.70
designer dog breeds	1.00	1.00	0.60	0.50
pressure washers	1.00	1.00	0.70	0.80
sore throat	1.00	1.00	0.50	0.60
idaho state flower	1.00	0.70	0.50	0.50
indiana state fairgrounds	1.00	1.00	0.80	0.80
Fibromyalgia	1.00	0.30	0.50	0.50
ontario california airport	1.00	0.60	0.80	0.40

Table 7. The precision of 50 test queries

6. Hypertext Diversification

Some applications have become unpopular due to their focus on hypertext lack of diversity. Google, for example, devalued the backlinks of websites that excessively used promotional keywords as hyperlinks and placed links on spammer websites. This is why diversification is essential (i.e. the mixing of hypertext and backlinks quality). Research has been conducted to determine whether websites that experienced good rankings had optimized hypertext over instead of using natural mixed hyperlinks.

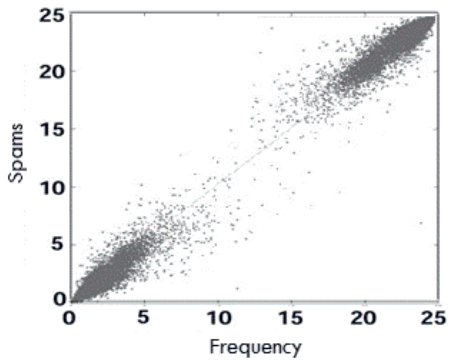


Figure 4. Anchor frequency vs. spam distribution

Figure 4 shows the trade-off between the frequency of hyperlinks in websites and spam validation. The percentage of some keywords (e.g., money) should not exceed the exact threshold. This does not mean a keyword could not be used more than once but it should not rotate that keyword one to five in 100% of the links acquired. To determine what types of hypertexts work from long term perspective and in what proportion, the competitive keywords could be selected, e.g. 'quick weight loss', 'real estate for sale', 'cheap hotels', or 'web design company'.

Figure 5 shows six major types of hypertexts involved different distributions in the selected dataset classified as follows:

- **Branded:** Brand hyperlinks use the actual name of a brand or business. For example, "Outreachmama.com" would use the brand hyperlink "OutreachMama". As the Internet grew, many companies adopted names that would capitalise on industry keywords, such as www.Londonseocompany.com, the branded hypertext might legitimately read as "Find out more about London SEO Company".
- **Exact match:** In some cases, it was difficult even impossible to determine if the hyperlink features exactly match text or simply a brand name. Other examples are domain names, such as "smoking-cessation.org" that ranked high for "smoking cessation".
- **Partial match:** Most websites successfully implemented the diversification strategy by mixing some targeted keywords with synonyms. The number of partially matched hyperlinks, and anchors with synonyms, and long tail keywords were available up to 50% across all websites.
- **Naked links:** As the name implies a naked URL without an http://, it is a URL that only uses "www" and excludes "http://" (e.g. www.youthnoise.com). The quantity of naked URLs in hyperlinks ranged from 2% to 15%. Many URLs (15-35%) as hyperlinks were used by websites with famous brands; such as, namecheap.com (40% of URLs in hyperlinks), hotels.com (53%) and smoking-cessation.com (90%).
- **Non-descriptive hyperlinks are rare:** The quantity of non-descriptive hyperlinks was low across all websites ranged between 2% to 5%.
- **Other languages:** The quantity of keywords in other languages was few ranged from 0.2% to 0.8% per website.

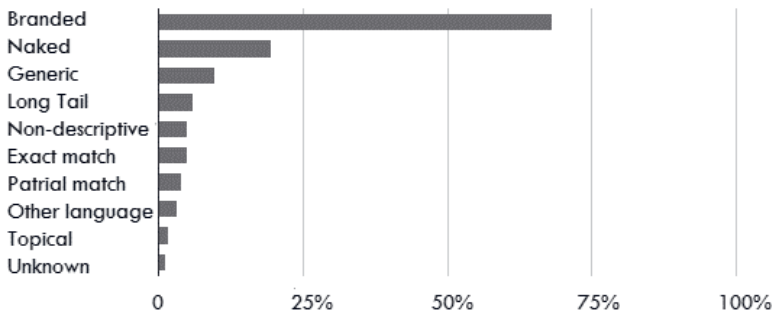


Figure 5. Types of hypertexts

7. Conclusion

We demonstrated how important of hypertext in Web documents for improving the quality of Web search for many types of queries and emphasized the importance of indexing in hypertext systems. We highlighted this by examining the properties of hypertext in a large subset of ClueWeb09 collection. Our main premise is that hypertext has an impact very similar to real user queries and consensus titles, and also, unlike linear text as closely resembles the networked and associational organization of information on the Web. Thus, understanding the relationships of hypertext in the documents leads to better understanding how to find high quality search result to a query string. We tested our approach experimentally in several types of queries from TREC Web track including the significant stream of queries versus the Web content. We conducted our experiments to investigate several aspects of hypertext, including the relationship to its domains, the frequency of queries that could be satisfied by hypertext alone, and the homogeneity of results acquired by hypertext index. We think that a reason makes hypertext more efficient for Web search is that most users used short queries and tend to choose minimal number of terms that annotated by meta-tags or concisely summarized pages as the same way that Web designers choose hypertext.

References

[1] Yadav, D., Sharma, A., and Gupta, J. (2009). "Topical web crawling using weighted anchor text and web page change detection techniques". *WSEAS Transactions on Information Science and Applications*, 6(2).

[2] Halpin, H. and Lavrenko, V. (2011). "Relevance feedback between hypertext and Semantic Web search: Frameworks and evaluation". *Journal of Web Semantics*, 9(4):474-489. DOI:10.1016/j.websem.2011.10.001.

[3] Al-akashi, F. and Inkpen, D. (2014). "Term Impact-Based Web Page Ranking". *In Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS)*. DOI; 10.1145/2611040.2611060.

[4] Craswell, N., Hawking, D., and Robertson, S. (2001). "Effective Site Finding using Link Anchor Information". In *Proceedings of the SIGIR'01*.

[5] Page, L., Brin, S, Motwani, R., Winograd, T. (1999). "The PageRank Citation Ranking: Bringing Order to the Web". *Stanford InfoLab*, DIO; 1.1.31.1768.

[6] Kleinberg, J. (1999). "Authoritative sources in a hyperlinked environment". *Journal of the ACM*, Vol. 46, No. 5, Pp 604-632.

[7] Boytsov, L. and Belova, A. (2012). "Does Category (A) Anchor Text Improve Category (B) Results". In *Proceedings of the Twenty- first Text Retrieval Conference (TREC)*.

[8] Marijn, K. and Jaap, K. (2011). "Reducing Redundancy with Anchor Text and Spam Priors". *In Proceedings of the Twentieth Text REtrieval Conference (TREC).*

[9] Upstill, T., Craswell, N., and David, H. (2003). "Predicting Fame and Fortune: PageRank or In-degree". *ACM Transactions on Information Systems (TOIS).* Vol. 21, No. 3, Pp. 286-313).

[10] McBryan, O. (1994). "Genvl and wwww: Tools for taming the web*".* In *Proceedings of the First International World Wide Web Conference*, Pp. 79-90.

[11] Chakrabarti, S., Dom, B., Gibsonm D., J., Kleinberg, P., and Rajagopalan, S., Gibson, D. Kleinberg, J. (1998). "Automatic resource list compilation by analyzing hyperlink structure and associated text". *Computer Networks and ISDN Systems*. Vol 30, No. 1-7, Pp. 65-74.

[12] Nadav, E. and Kevin, S. (2003). "Analysis of Anchor Text for Web Search". In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (Pp. 459-460).

[13] Atsushi, F. (2008). "Modeling Anchor Text and Classifying Queries to Enhance Web Document Retrieval". In *Proceedings of the 17th International World Wide Web Conference.* DOI: 10.1145/1367497.1367544.

[14] Amitay, E. (1998). "Using common hypertext links to identify the best phrasal description of target web Documents". In *Proceedings of the SIGIR'98 Post Conference Workshop on Hypertext Information Retrieval for the Web*. Melbourne, Australia.

[15] Haveliwala, T., Gionis, A., Klein, D., and Indyk, P. (2002). "Evaluating strategies for similarity search on the web". In *Proceedings of the 11th International Conference on World Wide Web.* Pp. 432–442.

[16] Davison, B. (2000). "Topical locality in the web". *In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 272- 279), New York, ACM.

[17] Attardi, G., Gulli, A., and Sebastiani, F. (1999). "Theseus: categorization by context". In *Proceedings of the 8th International World Wide Web Conference*.

[18] Shuming, S., Fei, X., Mingjie, Z., Zaiqing, N., and Ji-Rong, W. (2009). "Anchor Text Extraction for Academic Search". In *Proceedings of the Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, ACL-IJCNLP (Pp. 10–18).

[19] Ganeshiya, D. and Sharma, D. (2014). "A survey: hyperlink analysis in webpage ranking algorithms," *International Conference of Soft Computing*

Techniques for Engineering and Technology (ICSCTET), Pp. 1-8, DOI: 10.1109/ICSCTET.2015.7371192.

[20] Costa, J., Rodrigues, J., Simões, T. and Lloret, J. (2016). "Exploring Social Networks and Improving Hypertext Results for Cloud Solutions," *Mobile Network Application,* 21, 215–221. DOI: 10.1007/s11036-014-0513-z

[21] Clarke, C., Cormack, G., (1995). "Dynamic Inverted Indexes for a Distributed Full-Text Retrieval System". *TechRep MT-95-01*, University of Waterloo.

[22] Albi, D., and Silvester, H. (2017). "PageRank Algorithm". *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. 19, No. 1, Ver. III (Pp. 01-07).

[23] Hiemstra, D. and Hauff, C. (2010). "MIREX: MapReduce Information Retrieval Experiments". Technical Report TR-CTIT-10-15. ISSN 1381-3625. http://eprints.eemcs.utwente.nl/17797

[24] Anh, V. and Moffat, A. (2010). "The Role of Anchor Text in ClueWeb09 Retrieval". In *Proceedings of the Nineteenth Text Retrieval Conference Proceedings (TREC).*

[25] Kamps, J., Kaptein, R., and Koolen, M. (2010). "Using Anchor Text, Spam Filtering and Wikipedia for Web Search and Entity Ranking," in *Proceedings of the 19th Text Retrieval Conference (TREC)*, USA.

[26] Serdyukov, P. and Vries, A. (2009). "Delft University at the TREC 2009 Entity Track:
Ranking Wikipedia Entities," in Proceedings of the 18th Text Retrieval Conference (TREC), USA.

[27] Cormack, V., Smucker, D., and Clarke, L. (2010). "Efficient and effective spam filtering and re-ranking for large web datasets".
http://arxiv.org/abs/1004.5168, 2010.

10.31341/jios.48.1.2 UDC 004.4:005.2:658
Open Access Survey Paper

Entrepreneur's strategy leveraged and monitored by the interrelation of ERP and BSC

Cidália Oliveira *cidalia.o@hotmail.com*
REMIT – University Portucalense, Porto, Portugal
Instituto Europeu de Estudos Superiores Medelo/ Fafe Portugal

Márcio Pereira *mjpereira@ipca.pt*
Higher School of Management Polytechnic of Cávado and Ave (IPCA)

Margarida Maria Mendes Rodrigues *margarida.rodrigues@iees.pt*
CEFAGE-UBI Research Center, Covilhã, Portugal
Instituto Europeu de Estudos Superiores Medelo/ Fafe Portugal

Rui Rodrigues *ruisilva@utad.pt*
CETRAD Research Center, NECE-UBI,
Univeristy of Trás-os-Montes e Alto Douro, Vila Real, Portugal

Abstract

Nowadays business operates in a turbulent environment, demanding quality products at low cost in short delivery times. Bearing this in mind, entrepreneurs devote special attention to define strategies, monitor entrepreneurial indicators and recruiting qualified employees. In this sense, organizations need to rely on an efficient measurement system with vast and complex information. Considering this need, the Balanced Scorecard (BSC) embraces a short-, medium- and long-term time horizon and guides entrepreneurs in the measurement of their Key Performance Indicators (KPI), as acknowledged by the well-known company like Apple, which is using BSC for planning and performance measurement. Furthermore, as the KPIs need to reflect organisational performance, the interconnection of BSC with SAP Enterprise Resource Planning (ERP) is considered of benefit. By means of this interlinkage, performance is monitored in an integrated way, as the flow of material and information of the whole chain is integrated and interlinked via ERP. Stressing out that several BSC implementations have failed, this research, intends to detail how ERP implementation can leverage the BSC's features, namely to monitor entrepreneur's strategy. As several studies have focused their attention in an isolated way and several BSC implementations failed, this research sheds light on entrepreneur's strategy leveraged and monitored by the interrelation of ERP and BSC.

Keywords: Balanced Scorecard, Entrepreneurial strategy, Enterprise Resource Planning, SAP

1. Introduction

Considering the current challenging macroeconomic context for competitors, entrepreneurs face the need to focus on management tools that identify product cost on the one hand and measure organizational performance on the other [1] [2] stated the need for faster and more competitive adaptation in the face of competition. This adaptation requires human resources willing to train and adapt, to absorb, adapt and reconfigure competences [3]as well as investment in information technology Following the need to adapt, intangible resources sustain growth in tangible returns [4].

Several developments of the Enterprise Resource Planning (ERP) have taken place over the past few decades, aiming to support companies of different countries and different markets. Sustained on ERP, organisations are able to integrate emerging technologies Big Data, social networking and mobile telecommunications [5]. Through the ERP, several records are performed in an integrated way, enabling collection in a faster and easier way [6][7].

The perfect decision-making condition is to have in integrated way, to interlink with the BSC, in order to monitor performance.

The added value of BSC in areas of management has been confirmed by several studies and been disseminated in high profile publications such as the Harvard Business Review [8] presented as a very promising management tool [9]. Organizational knowledge of resources and competencies is a valuable key factor, so, BSC is widely recognized and used to monitor performance [10][11]. [12] report that worldwide 62% of large organizations apply BSC. Similarly, [13] emphasize that 60% of Fortune 1000 organizations have implemented BSC. The adoption of BSC in Fortune 500 organizations corresponds to a 40% rate. Despite the noted advantages, there are several organisations that failed the BSC implementation [14]–[16].

Considering the divergent current of the literature, we intend to interlink theory and practice of this research field, following the path of [17] to understand how an ERP can leverage BSC in its features, namely in monitoring organizational strategy and performance in operational management. For this purpose, the following specific objectives were defined: Identify auxiliary tools for BSC implementation;

Identify the role of ERP in leveraging BSC – in monitoring of organizational strategy and performance

Given the specific objectives, the following Research Question was defined: (RQ): Can ERP leverage BSC in operational management?

Considering the research question, we selected a qualitative approach, of interpretative and explorative nature, supported by the literature review and interviews of organisational entrepreneurs.

This is followed by a literature review, methodology, findings and discussion, and conclusions. Furthermore, the limitations and expectations for further research are pointed out.

2. Entrepreneurial Management and Performance Systems

Enterprise and Corporate Performance and Management (EPM/CPM) systems are now basic tools for efficiently running a company, but digitalization has the potential to completely transform procurement, production, sales, logistics or even financial accounting or corporate management (planning, reporting). EPM/CPM is a tool to integrate operational and financial information into a single decision-support and planning framework [18].

Many organizations face the difficulty to realize the value of a data and waste the valuable and usable data, lack of the integration of multiple single improvement methods. EPM/CPM enhance to fill up the gap between the leader's vision, mission, the investors' expectations and the employees' work. In addition, these methods can abolish some problems that receive great attention recently. EPM/CPM refers systems like activity-based costing (ABC) and lean accounting that enhance the correctness of costs and bear out the causes of the hidden costs. Traditional internal management accounting systems often cannot lay out cost transparency and visibility. Regarding budgeting, it is often a task done by the accountants, isolated from the executive team's strategy [18].

Business intelligence & analytics technologies as a reshaping factor in business gain a wider space for themselves nowadays, as a tool for supporting reporting and decision-making processes. As managerial accounting is supportive and one key pillar for decision-making and control in an organization, it can take clear advantages of adopting BI&A technologies. In other words, BI&A is defined as a technology and process that make it easier to collect data, to analyze and to deliver actionable information, and so better data analysis and decision support create value for companies [19].

2.1. Implementing the Balanced Scorecard for Strategic Management

BSC is a management tool composed of four perspectives, namely the financial, customer, internal and learning and development perspective [20]. Its strategic objectives are grouped into four perspectives, which allows entrepreneurs to aggregate focus on objectives according to the organizational strategy.

BSC is an integrated management system, so many successful cases are known that can improve organizational performance [21]. More and more organizations worldwide are choosing to implement BSC [22]. [16] report that the adoption rate in Germany, Austria and Switzerland is close to 25%. Equally, [23], in their study of the 250 largest companies, reported that the level of knowledge about BSC is almost 50% and the level of implementation is approximately 20%. In turn [24] in an analysis of the 250 most exporting organizations in Portugal concluded that over 40% of them had BSC implemented.

[25] point out that it is essential that BSC evolves to meet new organizational needs and respond to them more effectively and efficiently.

Organizations have a need to value intangible assets, so this justifies increasing the BSC implementation rate. There are a number of organizations, such as Walt

Disney, HJ Heinz, Johnson, Johnson, and Merck, which use the measurement and analysis of the organization's success based on BSC with up-to-date and relevant information [25]. For the implementation of the Balanced Scorecard one of the primary steps is the selection and structuring of the work team as well as the financing of the implementation project. Implementation work should be divided into strategic plans, financial plans, customer segmentation plans, human resources plans, and quality plans. The implementation of this tool enables the aggregation of various levels, namely control, information and communication, as well as learning, so BSC can be considered to measure, control, inform, communicate and stimulate structural changes [26] . BSC focuses on communication and liaison, business planning, vision translation, and customer satisfaction as the central focus of organizational strategy to achieve goals [27]. Regarding the BSC diffusion and implementation, data confirm that the BSC continues to be implemented in Norwegian municipalities [28].

BSC has been selected as a tool of choice by various entrepreneurs not only for measuring performance, integrating budgeting and planning activities, but mainly because of the possibility of interconnecting ambitious capital and investment goals [29]. On the other hand, certain organizations have selected BSC as a strategic management system to monitor current performance, compare past performance and project future performance [30]. This tool is distinguished by its ability to balance organizational information through indicators from various perspectives, conveying results concisely [31], Entrepreneurs are given the opportunity to tailor the indicators to their strategic needs [32].

Besides being a strategic and communicational performance measurement tool, [33], identified perceived benefits of the BSC, as shown in the table below:

Managerial focus	Helps entrepreneurs focus on what is important in the long run Helps managers prioritize and make decisions
Sense of 'balance'	Balanced and holistic view of the organization's performance
Communication and visualization	Common language Common frame of reference Facilitates discussions
Alignment of goals	Helps improve goal congruence Increased awareness of the organization's long-term goals
Cultural and motivational tool	Changes how the organization 'thinks' Captures the attentions of organizational members Motivational effects as a result of more explicit targets and incentives
Organizational change catalyst	Rhetorical tool that can be used to justify organizational changes Well-known concept

Table 1. Perceived benefits of the BSC (Source: 31, p. 85)

Organizations that need to implement corporate strategies should define a timeline in detail and have an opinion on the effectiveness of the new strategy and its strategic indicators [35]. It is recognized that entrepreneurs who follow the definition and implementation of the indicators and the whole process from roll-out enable the BSC to continue its causal links. The rolling out of the implementation process is time

consuming as it takes on average terms, after clear identification of the strategic indicators, from eight to twelve weeks.

There are several types of software that can be selected for BSC implementation, as [36] refer, applying the tool through interface integrations developed in Excel or Access, so the data upload may occur automatically or manually. The degree of compliance of each indicator against the target is flagged with a certain colour. They often opt for red, yellow and green [37]. Red marks indicators whose objectives have not been achieved or are very far from the objective. Yellow characterizes indicators that are at risk of being missed due to slight deviations. Green represents indicators whose targets have been met [36]. Some organizations add blue to symbolize indicators whose objectives have been exceeded. This whole process proceeds simply and demonstrates the fulfilment of each indicator in more or less detail as intended by each manager [36]. Research shows difficulty in confirming the effect of BSC on financial performance, as it leads to different interpretations and practical ways of being used [38].

Despite this, after implementation, it is confirmed that organizations have more transparent information, which facilitates interaction with their partners (suppliers, customers and other entities) and more complex analysis and data. Through monitoring, rather than tracking performance, you can highlight inappropriate strategic decisions and feedback from entrepreneurs [39]. In this sense, Entrepreneurs consider BSC as a crucial tool [40], provided that the conditions for implementation are met in order to vitalize the expected benefits from the tool itself [39].

Translating the Vision - corresponds to a consent between vision and strategy in order to operationalize the strategy, usually supported by strategic maps [26:155] Feedback and Learning - consists of the strategic implementation of a continuous and progressive process sustained in the ability to adjust to cultural changes [26:155].

Communication and Linking - characterizes communication throughout the hierarchy by integrating the various levels individually and in aggregate, respecting the key processes and goals to be achieved [26:155].

Business Planning - consists of allocating resources and identifying internal priorities to meet strategic goals [26:155].

It is possible to elucidate business practice, vision and organizational strategy defined by entrepreneurs [41]. However, mere strategic clarification alone is not sufficient, given that many entrepreneurs effectively develop their strategic plans, yet their operational implementation is not yet effectively fulfilled. Recognizing these assets, BSC may hold the key to success, as it demonstrates how to link organizational strategy through the implementation of indicators [42]. BSC is composed of organizational principles and corporate objectives, individual to each organization, so it is not standard, but adaptable to each organization [43], and adjusted by users [44].

BSC implementation has to be adapted to the technical, cultural and political characteristics of the organization [45]. Regarding the individual indicators of each business unit, the alignment of the vision of each productive unit and the definition of its own objectives have to be defined and subsequently measured through indicators. To be tailored to the needs of each process, everyone involved must be motivated and contribute to these indicators. Expectations are difficult to manage, so the customer

should be involved in setting their expectations, as this will enable the business [46]. Once individual indicators have been defined, the units' BSCs serve as a monitor. By measuring the indicators, deviations from the targets are identified so that managers make their respective decisions [46].

BSC authors [47] report that there are five principles that transform BSC from performance measurement to driving strategic performance, namely: translating operational strategy; strategic alignment; understanding of strategy and incorporation by all employees; continuous and mobilizing process of change sustained in a dynamic management of new cultures and strategies [46][47].

Given the importance of BSC implementation, [49] states that there are eight key factors for BSC to be successfully implemented: (1) involvement of Top Management; (2) involvement of Department entrepreneurs and participation of their employees; (3) culture of excellence that enables the desired performance; (4) learning and development; (5) keeping the tool relatively simple and easy to use and understand; (6) clarity of strategic vision and expected returns; (7) linking the incentive system to the BSC indicators and (8) the existence of resources needed for implementation. In addition to these factors [50] list the following success factors for BSC implementation: (a) adaptation of BSC to the organizational and strategic context; (2) constant learning and training, as also referred to by [49] and (3) review of BSC indicators after implementation.

Strategic execution goes beyond an organizational phenomenon, as it has to be clear and objective for it to be executed accordingly [51]. In this context of strategic clarity, [51] states that whenever the strategy is clear it is possible that consequently the performance achieved will be fostered. Controllers play an active role in organizations by focusing on a strategy driven by the objectives of [47]. BSC is not only a reporting tool, but also a tool that combines executive management strategy with operationalization, enabling communication with other members of the organization [37].

Defining indicators and communicating objectives is crucial, so that all employees know which strategy to communicate, so indicators should be reviewed regularly and updated [48]. In Controlling departments, entrepreneurs raise questions in order to identify possible improvements, notably identifying what is being done and how it could be improved. Controllers ascertain the existence of resources, namely the needs of physical and human resources, to enable an efficient allocation respecting the defined organizational strategy. [52] found in the Mobil Case Study that the CEO said that through Scorecard parameterization he can define the organization's strategic priorities of individual and global performance, as well as the anticipation of possible turbulence [52].

Monitoring of both short- and long-term indicators [29] balances both timeframes [37]. As detailed in this research most of the cited Case Studies, are related to private organisations, but public organisations are also implementing the BSC, even so, in public organisations strategic directions are missing, which are crucial for the BSC to sustain decision-making and performance improvement [50].

Furthermore, the lack of information and knowledge on how to implement the tool triggers disinterest in its implementation, so attention must be paid to the

characteristics and requirements needed for it to be implemented in order to contribute to organizational performance [53].

2.2. Balanced Scorecard Limitations and Implementation Failure

In the literature review, several authors, such as [54], [55] highlight multiple success stories of BSC implementation. However, there are publications that mention that there are organizations that do not implement or intend to have this management tool implemented. Considering this divergence of opinions, [56] conducted an investigation to identify some of the reasons that lead large companies to dispense with the implementation of BSC. The authors conclude that these motives may be related to contingency theory, as well as to the fact that they consider the tool to be unsuitable for measuring performance in their specific context [57]. However, [56] report that the lack of experience makes it impossible for entrepreneurs to implement BSC effectively. This gap causes [58] to emphasize the relevance of investigating the critical factors inherent in BSC implementation. In order to identify the reasons [56] conducted some research and found that there are cases in which entrepreneurs and administrative employees are unfamiliar with the indicators and how they are measured, leading to difficulties in relating the indicators to the strategy. Consequently, they present difficulties in establishing cause-effect relationships between them, as they cannot obviously identify their interconnections [56] In addition, one reason is that some administrative staff are resisting the implementation of BSC by avoiding the change and pressure required for BSC monitoring. Given the diverse reasons for possibly not implementing the tool, it is relevant to distinguish reasons related to inexperience from reasons related to resistance to change and consequent lack of implementation initiative [56].

Furthermore, [59] have identified in a summarized way a portfolio of problems related with the BSC implementation, which can be seen in table 2 below.

Issue type	**Problem**	**Explanation**
Conceptual issues	Contextualization	The BSC is a “general model” which may be difficult to customize to fit the organization
	Causal relationships	Organizations struggle with understanding and testing causal relationships
	Strategy maps	Organizations have difficulties understanding how to implement strategy maps
Technical issues	Technical aspects	Organizations have problems with data gathering and automation, or become too focused on the technical aspects of the concept

Social issues	Organizational culture	The BSC may be incompatible with the organizational culture, e.g. lack of acceptance of measurement
	Participation	Organizational members remain passive and delay or block the implementation process
	Commitment	Lack of commitment from central actors in the organization, such as the top management group or the BSC project group
Political issues	Time and resource Resistance	The implementation of the BSC consumes a lot of time and resources
	Concept champion	The organization lacks a person that is spearheading the BSC project
	Continuity	The continuity of the BSC project is threatened by turnover
	Resistance	Organizational members resist the implementation of the BSC

Table 2. Perceived problems associated to the BSC (Source: 57, p. 123)

Organizations that really want the implementation need employees with willingness and learning capacity, in order to fulfil the defined goals related to their organizational functions.

However, while entrepreneurs recognize the potentialities of BSC, they are also aware of the requirements and complexity inherent in implementing such tools. Because of these difficulties, it is advisable for organizations that choose to implement this tool to hire experienced consultants to guide them through implementation. It is important that a senior project manager is appointed and that the number of employees involved in the project is not insufficient or otherwise excessive. Consultants must rely on a robust experience to meet their stated goals and can translate and convey the strategy to all employees so that everyone can work motivated and committed [47].

There are difficulties of familiarization by the employees, when organizational change happens, or is required. Given that learning is a lengthy process, it is important for employees to move out of their comfort zone and adapt to their needs, which will impact organizational performance as they influence the working climate and employee motivation [60]. To this end, the working climate must be favourable for knowledge sharing and a commitment to teaching, motivating and educating employees [47] which is made more difficult in older organizations.

Shortcomings in the conceptual configuration of BSC related to BSC design are often identified, such as the excessive number of strategic objectives defined for each perspective [56]. There may also be failures related to inconsistent strategic alignment

along organization's processes. Considering the dynamism of markets and customer's requests BSC will have to be constantly adapted and improved, to fit to organizational reality [56]. Outdated financial indicators may induce organizations to make misleading decisions, given that indicators, once set, remain static until they are updated again [61]. Given this need, outdated indicators lead to inefficient management and consequent mis-decision making of reality [61]. On the other hand, [13] state that the indicators refer to specific scenarios, so it would be convenient to be uniform in different situations.

The literature also points to other limitations regarding the imbalance between perspectives, because according to [62], BSC underestimates the employees' perspective, considering that it overvaluates control and trade. Contradicting this statement, [63] highlight BSC implementation studies also in universities, which clearly shows the broad applicability of BSC. Marketing and business objectives pay insufficient attention to human resources despite playing a key role in BSC implementation [27]. Learning capacity generally remains underestimated [11]. In this sense, [64] highlight the impact of the learning and development perspective on organizational performance. There is still a gap regarding the link between performance and employee compensation, as well as the uncertainty of the indicators to be selected and the internal resistance to change [50]. On the other hand, an exaggerated number of indicators, as already mentioned, could make it difficult to monitor them in the Scorecard (Nørreklit, 2003). Through practice, namely testing and analysis, theoretical and practical concepts can be developed considering that indicators differ according to perspectives [65].

Beyond conceptual and empirical limitations, the lack of financial resources leads to the implementation of inadequate information systems or organizational resistance to change. Because it is closely intertwined with the organizational philosophy of each organization, the identification of indicators and their metrics cannot be automated because it must be parameterized in accordance with the mission, vision and specific objectives of each organization. Given that it is not possible to aggregate all the knowledge and strategic information of a given organization in the way a liquid can be stored in a bottle, BSC allows knowledge melting, objective measurement and deviation analysis [66]. BSC's organizational perspectives must also be viewed in aggregate so that weaknesses are addressed in accordance with the [47] strategic plans, recognizing that value creation depends on the supply chain which has to be addressed in depth [67].

Some criticisms of BSC are reported in the literature, as they reveal that BSC may favour or drawback organizational control, as it is incomplete in certain indicators [68], such as: (1) undefined how employees and suppliers collaborate to achieve the objectives; (2) undefined contribution of the community in the environment surrounding the organization; (3) lack of definition of performance indicators that demonstrate shareholder contribution.

[69] states that the tool is not implemented in small companies in the Czech Republic for the following reasons: (1) implementation of alternative systems; (2) small size of the organization, (3) insufficient resources for information interconnection and (4) entrepreneurs' lack of time for effective implementation.

Clearly, limitations are more significant for small and medium-sized organizations because of the time required for implementation, as well as the financial investment required for implementation [69]. Successful implementation can be achieved by minimizing additional costs and efforts through the interconnection of human capital with structural capital, thereby achieving the intellectual capital needed for market sustainability [70]. Knowledge management is critical to future decisions and organizational progress, so it is extremely important to identify how knowledge is absorbed [70]. Short term analysis has been criticized as it does not allow the efficient identification of value creation of organizational activities [71]. [67] state that the BSC should be composed not of four, but of five nuclear perspectives. (see Figure 1).

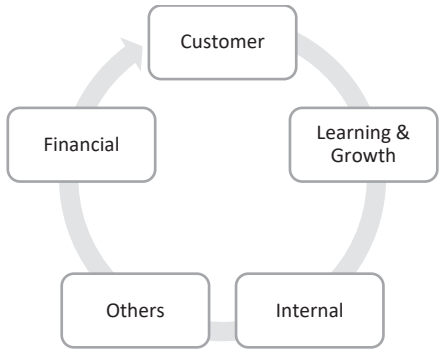


Figure 1. The Five Innovative Perspectives of BSC (Source: Mingming et al., 2010: 66)

For organizations, it is crucial that employees understand and agree with the indicators, so that they do not hinder their implementation and monitoring. In this sense, it is important for employees to be involved, so that they feel part of the project and consequently collaborate efficiently [36].

Communication must be motivating so that in an organization all members work towards the same goal. Although BSC is considered a strategy communication tool [72], this relevant feature is sometimes overlooked when faced with a quantitative measurement system [65].

In this context of quantitative measurement, entrepreneurs often focus almost exclusively on the definition, monitoring and achievement of the defined quantitative targets and disregard the surroundings and all the innovations that the competition exhibits in the meantime. Due to this lack of attention to the external environment, when the organization faces competition threats it finds that the response time is very long and the business opportunity may no longer be profitable [73]. Today the BSC fails to help organizations speed up, meet market challenges, and meet customer expectations. The most successful organizations are those with highly qualified employees who work closely with stakeholders and pay greater attention to the variables that ensure success [74].

[75] consider that the lack of consideration for external factors is because BSC focuses primarily on the correct definition and control of operational tasks, which may consequently make it impossible to adapt to market demands. Given these limitations [76] report that the BSC follows the efficiency-seeking logic, noting certain

limitations such as: (1) BSC is an inflexible tool and tends to fit indicators into one of four perspectives; (2) BSC statically shapes an organization and disables competitive opportunities from the surrounding environment; (3) difficult connection with the external environment as it focuses mostly on an internal analysis; (4) BSC proves gaps in value creation, learning and growth, as it follows a very traditionalist logic of innovation and (v) BSC indicates linear thinking, with no room for the intellection of complex organizational processes.

Given some of the limitations mentioned, [11] report that the BSC is not sufficient in performance management, so they recommend that other resources be used with this tool, while maintaining the BSC as the foundation.

As a complement to existing management tools on the market, a more comprehensive management tool has been launched to enable performance measurement and relationships with local communities and government and their respective stakeholders. This tool, called the Triple Bottom Line, enables the joining of social indicators (such as the impact of the company and its suppliers on society), as well as environmental indicators (such as energy, area and geographical conditions). However, although comprehensive, it was not well received due to its complexity [77] and is therefore not considered an alternative to the BSC.

The literature lacks empirical studies that detail the implementation stages of the tool, so the use of the tool needs attention and practical analysis, and the empirical validation must be reinforced [65]. Additionally, the cost-benefit analysis remains to be determined [31], despite the recognized advantages inherent in the management tool.

Current investigations point to the BSC as a virus, on the one hand and as a tool of fashion, on the other. Although they seem antagonistic, the characteristics should not be seen as mutually exclusive, as they complement each other. Macro-oriented fashion theory describes the contagious nature of the BSC; in turn, micro-oriented virus theory best characterizes post-adoption evolution of the BSC as a practice in organizations [38].

2.2.1. ERP Implementation

In the early 90s of the twentieth century, with the emergence of new technologies that companies were able to enjoy systems that enable the interconnection of their entire business. Initially, these systems were called Enterprise Wide Systems (EWS), currently known as ERP systems [78].

These systems can be used to manage and coordinate all of an organization's information and resources from a shared data set. Data is organized into different modules and integrates all corporate information into a single central database. Thus it is possible to cover in a single system very distinct areas such as: Financial, Logistics, Production, Sales, Human Resources, Quality, Marketing and Maintenance [79].

ERPs are software packages that enable the complete integration of information generated and processed within organizations. The operation of ERP systems is relatively simple and can be summarized, as mentioned above, by the existence of a

database, which collects and feeds this data in applications organized by modules, supporting most of the company's activities. As new information is entered into the system, all related information is automatically updated [80].

According to Rom and Rohde (2007), the ERP implementation lead several manual processes to become automatized, reducing the manual repetitive tasks of accountants. [81] stated that based on an ERP implementation a new way of practicing accounting was reached.

This new version supports the execution of business functions inside and outside companies, not only for employees, but also for all entities that interact with the organization (Suppliers and Customers). The primary focus of ERP is to support organizations throughout the supply and distribution chain, including the integration of social networks, with the goal of creating direct channels for the promotion and sale of goods and services to current and potential customers [82].

This was also very relevant for the positioning of Financial Entrepreneurs and Directors, which in the past have been several times disregarded. Therefore, these entrepreneurs used the implementation of the ERP to have the needed information in an integrated and aggregated way and to get rid of the minimizing title of "beancounters" [83].

However, the decision to implement an ERP system must be taken into account, as it is a very complex decision, which cannot be taken without considering all the pros and cons that these systems can bring to the organization, in terms of both business processes, the tasks and behaviours of employees, as well as the financial structure [84].

In fact, these systems bring many benefits to various levels, including the introduction of organizational best practices, the implementation of standardized procedures throughout the organization, and the use of current and dynamic tools. All this enables companies to be able to integrate, analyse and report information from all business areas, enhancing organizational excellence through full integration [85].

In addition to the benefits, implementing ERP systems has several risks and issues that need to be considered. Therefore, the installation processes must be well planned, managed and constantly monitored by qualified personnel. The implementation of an ERP system can become an endless process if entities take on an unfinished project logic, where information generates the need for more information [79].

It is also crucial that companies do not forget that if all information is integrated, as data enters a part of the system, there may be multiple impacts throughout the system and thus produce harmful effects in various sectors of the organization. Therefore, the implementation of these systems will necessarily have to be accompanied by a constant reengineering of all business processes [86].

ERP systems are seen as levers to exert greater management control and impose more uniform processes and are used to inject more discipline into organizations. These systems are often also used for the opposite purpose, in an attempt to break down hierarchical structures, freeing up their employees, making information available to lower hierarchical levels of management, making them more innovative and flexible. For multinational and geographically dispersed companies, ERP is also

an advantage, as these companies often use these systems to introduce and ensure process uniformity across multiple dispersed units [87].

Currently, information systems in general are implemented in all organizations in a more or less developed way, being part of the daily routine of any element in the organization and assisting in the most diverse tasks [88].

The success of these systems was exponential and even during the 1990s, transforming the famous German company SAP (Systems, Applications and Products in Data Processing) into the leading global company in this field with its flagship SAP ERP product and becoming fastest growing software company in the world [87].

The modules offered by SAP are many and fully integrated, allowing its users to access real-time information about any part of the business. The main modules are divided into accounting and finance, human resources, production and logistics, sales and distribution. According to several studies that have questioned the most implemented modules, they concluded that the financial accounting and materials management modules are the most chosen [84].

2.2.2. Connecting the Balanced Scorecard to the ERP

Research has confirmed that there are some changes in the practice and implementation of advanced management accounting techniques. In this study, regarding the practices used after the ERP installation, the organizations responded that they introduced cost centers, profit centers, profitability analysis by product and by business activity / segment and financial ratio analysis. Regarding the introduction of new, more advanced practices, the results of this study indicate the introduction of financial and non-financial performance appraisal indicators and customer satisfaction surveys. The adoption of BSC and ABC were other techniques chosen after the implementation of the SAP system in the organizations studied [84].

However, another study consisting of ten exploratory case studies found that most advanced management accounting techniques, or even the most traditional ones, were performed in systems independent of the ERP. The implementation of techniques such as the BSC and the ABC method are not being included with the implementation of ERPs. This is not because these techniques are not installed, but because there is a preference to associate them with specialized and standalone software, which are more flexible, pleasant to use and reduce the complexity of operationalization brought by the ERP. These authors consider that despite the potential of ERP systems, changes have only been noted in the reliability of information, faster access to operational data, and automation in data collection and processing. Although not changing the logic of management accounting, ERP has allowed a deeper analysis of results, improving their understanding [89].

[90] dedicated themselves to interconnecting the BSC with ERP systems to monitor performance and enable process automation through the SAP software program. The main objective was to identify how the benefits of an ERP implementation can be measured, so [90] developed an ERP Scorecard.

It is the focus of organizations to maintain competitive advantage, as pointed out by [91], recognizing that the market requires orders to be executed quickly and at the

lowest possible cost. This requirement underlies the main organizational strategies based on the recruitment of qualified employees [74]. Consequently, performance measurement has been the subject of much concern [92]. For it is essential that the measurement system be efficient and provide vast and complex information [69]. Entrepreneurs need to adapt to globalization and to update to efficient management tools [93].

[94] suggest the implementation of a visual map to measure organizational performance. Strategic maps are aligned with the BSC, as they help define their vision and strategy [95]. The ultimate goal is to provide real-time performance information, as supported by the Strategy Board Map [94].

Given any strategic change, it is relevant to take into account organizational culture in organizational performance, given that it has a strong impact on performance [96].

Given the noted relevance in the literature, BSC has effectively demonstrated some measurement-related limitations [97], so ERP may assist in the strategic implementation of the BSC today, researchers focus on finding efficient measurement tools.

3. Methodology

Considering the Research Question that triggered this research, namely: Can ERP leverage BSC in operational management? The Case Study was selected as an exploratory step, encompassing method selection, question formulation, data collection and subsequent analysis so that the respective conclusions can be recorded [82][83]. Following the path [17]qualitative methodology can affect the quality of a research paper, so all methodological steps need to be followed. Temporal phase is one of the five methodological phases identified by [17] that describe aspects of the research phenomenon, fitting with the aim of this research.

Based on the initial question, we opted for a qualitative approach, namely an interpretative and explorative one, based on the literature review and the experience of entrepreneurs in order to identify the specific objectives, namely:

Identify auxiliary tools for BSC implementation;

Identify the role of ERP in leveraging BSC – in monitoring of organizational strategy and performance.

BSC needs information to measure performance from different perspectives through indicators. Successful implementation of this tool requires organizations to produce information in a reliable and timely manner. ERP, through its modules, enables users to obtain potentially more reliable and real-time integrated information. Thus, ERP can serve as an important source for obtaining and feeding indicators and enhancing BSC functionality.

The data collection phase is considered a primordial and very exhaustive phase, because it needs a lot of rigor and preparation time. In this specific case, we then carefully selected the interviewees, namely, the Entrepreneurs of the various Processes.

[100] confirm that the methodology consists in exploring and studying methods to develop academic research. Case Studies enable an in-depth understanding to explore and interpret the reality [99].

As the methodology of this research is of qualitative nature, the first part of this research devotes special attention to a more conceptual methodology in order to reconstruct theory, concepts, and ideas to develop theoretical foundations [101]. Afterwards, the empirical part of the research, namely the interviews were performed. Through the interview, based on semi-structured questions, as Appendix A, the interviewees have greater freedom to comment on the topics covered, as well as an easier way to observe and capture certain details, which may have enormous importance in relation to the questionnaires, which in essence are much more objective, not allowing extraordinary information. [102] state that there is no single method for data collection, despite the large range of methodological procedures for research in social sciences, often case studies follow their own methodological perspective. For [103] the case study has become a commonly used method in accounting and management control research, where interviews are one of the most important primary sources of information [99].

Semi-structured interviews are one of the most common methods in qualitative research, aiming to provide deeper knowledge of interviewees' experiences [104]. Following this path, in order to achieve empirical evidence in this interpretative and exploratory study, five semi-structured interviews were performed with the Process Entrepreneurs and Key-Users of the ERP implementation.

Although authorization, as Table 3, was given for the recording of the interviews, permission was not granted for the identification of the interviewees so they were coded as I1, I2, I3, I4 and I5, as below mentioned table, Table 4.

Company	Activity Sector
Case 1	Textil

Table 3. Organization characteristics

Concerning the interviewee's profile, all of the interviewees work since years at the company and know about the processes and procedures related to organisational control.

Code	Sex	Years of experience	Academic degree	Interview Duration
I1	Female	43	Bachelor	45 minutes
I2	Male	42	Secondary	50 minutes
I3	Female	44	Master	35 minutes
I4	Male	41	Master	45 minutes
I5	Male	44	Bachelor	30 minutes

Table 4. Interviewee profile

Additionally, in order to compare the data collected during the Interviews, direct observation was made of the productive sector, in four distinct periods, during the two shifts, on two consecutive working days. Direct Observation is a method that allows the observer to verify, on the spot, the way employees behave, ascertaining the information flow, ascertaining possible divergences from what was expected or referred to during the interviews.

3.1.1. Analysis of results

In order to simplify the link between the Research Objectives and Interview Questions we decided to accomplish the below mentioned table.

Interview-Questions	Research Objectives
Beyond the importance of defining strategy, which is the importance of monitoring strategy?	Identify auxiliary tools for BSC implementation;
This institution is dynamic and entrepreneurial which led to the need to implement a management control tool	
The implementation of management control tools is crucial, but implies a detailed process	
In order to reach a successful implementation of management control tools other auxiliary tools are needed	
Auxiliary tools need to be integrated to provide accurate data	
Through an integrated database organizational objectives and goals are monitored	Identify the role of ERP in leveraging BSC – in monitoring of organizational strategy and performance
ERPs are updated immediately on-time which provides consequently an on-line monitoring of performance indicators	
ERPs provide integrated information which provides consequently an immediately following of the defined strategy	
Any operational deviation is immediately detected when ERP is linked to BSC	
Any financial deviation is immediately detected when ERP is linked to BSC	

Table 5. Interview Questions linked to Research Objectives

Knowing that the implementation of these tools require huge financial effort, because it is a long process and requires highly qualified human resources, it is also essential that employees understand the goals set. Furthermore, all collaborators should behave in a motivated way to achieve a positive working climate, sustained by effective involvement of top management. This need to motivate and to ensure employees are dedicated and believe in the benefits of the implementation prove to be a difficult but important goal said I02 and I03. For I02, the most difficult was "to convince the employees that the effort that they were doing, would benefit them in the future". In

the same vein, also I03 confirmed "at the beginning it was a huge challenge to motivate our teams, making them believe that this effort and initial trouble would result in better work conditions in the future. The implementation of all the changes related to the new system relies on qualified collaborators; as otherwise, the implementation can be a never-ending process [79].

Literature recognizes that very often the annual Budget are performed by Financial departments, having no link to the executive strategy [18]. Along this research also I04 referred this limitation, namely that due to the pressure of time and the missing link to further information, the Budgets and Forecast are established without an internal informational link. I01 stated that nowadays, organisations have indeed a huge range of possibilities to integrate information, so there is no need to rely on estimation when accurate information exists.

Knowing that organisations and their managers seek for accurate and on-time information to take decisions. On hand, of the BSC implementation managers are able to measure, control, inform, communicate and stimulate structural changes [26].

Along the interviews this specific need to have an implementation of a management control tools was highlighted, specially beside the interviewees that were more linked to provide information for top management's decision, as I04 and I05. Interviewees told that every time that forecasts need to be performed a huge difficulty arises, as the information is not gathered.

Considering that each implementation consists in a special project and singular attention is devoted to each implementation phase, managers usually see advantages in the implementation, but on the other side see the need to have members of their teams affected to this implementation process. I02 referred "Every time we plan to have a new project, it is not just to think about and quickly start with the implementation. We need to meet with the whole project group, define a project timespan and define key-users for each phase". Indeed, it is crucial to define a detailed timeline to follow the defined strategy divided into its strategic indicators [35].

ERP implementation enabled that several manual processes were gathered and showed on-time information, avoiding non-value-added activities [105].

Manager I05 confirmed "We spent several hours designing and analysing the flow of this integrated tool. Even for us this whole process was new". Entrepreneurs knew that after all these definitions, in future once a new data is inserted in one area, it would be automatically updated [87]. [81] stated that based on an ERP implementation a new way of practicing accounting was reached.

As shown in the literature review, both the BSC implementation process and the SAP ERP implementation process are of a high organisational relevance.

According to the Entrepreneurs' perception, I01 confirmed "the first milestone was to identify the team to work with, for the ERP implementation. The first criterion was knowledge of the specific key area and the second criterion was ability to communicate". Also, entrepreneurs I03 and I05 were of the same opinion, as I03 referred "It was challenging to be selected as a key-user, as we were challenged to detail the future process of the organisations and also to transmit and communicate all the afforded changes in the system and employees' mindset".

Regarding the need to have an integrated database to gather objectives and monitor goals, indeed all of the interviewees agreed, that this is the source of success. Especially interviewees from financial departments highlighted the need of integrated database.

Managers benefit from integrated management tools, as on hand of this integration they are able to measure and anticipate the impact of all sectors that compose the organization. In order to keep the management tools updated, constant upgrades of all processes are afforded [86].

Related to the benefit of having updates constantly, in order to assure an on-line monitoring of performance indicators, interviewee I1 referred that "nowadays, constant updates are crucial as otherwise we'll be removed from the supply chain, we need to be very dynamic. Therefore, for sure, we need to rely on trustful information". I02 following the same path, stated, "all Processes need to be aligned, but furthermore, need to understand the needs and demands of the Processes that operate before and after. If we look in an isolated way, we'll never guide the whole organization". Performance Measurement tools, need to integrate operational and financial information, so that this information is available to support in decision making processes [18].

Based on the common alignment between processes, ERP provides integrated information to follow the defined strategy.

I03 referred the concern about updating and following the strategy, saying "every day, or sometimes seve4ral times along a specific day, we face changes in de requirements of customers, so that naturally, as a consequence, we need to adjust very fast in a dynamic way. Considering the ability to adapt, to my understanding, this is not enough, as we need also to follow the strategy and confirm, if the path still keeps the same". Following the same idea I05, highlighted, that based on an integrated system, the fears of missing information will be automatically reduced. Based on the BSC organisational information, shared into four perspectives will be performed in order to reach accurate results [31]. Considering the features of the BSC, the ERP might be an excellent source to promote this alignment. In order to assure strategic needs, indicators should reflect organisational interest and entrepreneur's need [32]. On hand of the BSC, individual indicators ought to be defined and consequently monitored, in order to detect deviations of targets and to define actions [106]. As previously highlighted, interviewees stated that one of the mayor advantages of having a management tool interlinked to an integrated enterprise resource planning is the possibility to react very fast to any possible deviation, this was specially highlighted by I01-I03, as these colleagues referred that without this interlinkage when a deviation is noted it is already to late to react, it just occurred. To provide the BSC with information, several integrated enterprise resource planning softwares might be selected as [36]. The relevance relies in having the interlinkage updated. Based on this link, BSC may be considered the success path for managers to link strategic indicators to organisational's success [42]. Regarding the financial benefits, these were highlighted mainly beside interviewees I04 and I05, referring that I05 stated "having the ERP linked to a management tool any deviation is immediately identified". Regarding the role of accounting, the introduction of ERP has brought

some changes in the most routine tasks, eliminating them with the automation promoted by the system, giving them more time to perform other more relevant tasks that can contribute to the improvement of the information provided. Another important change was in the day-to-day accounting role of the company. In the ERP implementation process, the opportunity was given to accounting professionals to be more involved in designing all implementation processes to ensure that information introduced and handled in the system would produce more reliable and timelier financial and management reporting. Therefore, considering that this implementation enables the analysis and integration report, this could serve as a basis for the BSC indicators, as it levers management control, guides the implementation of new procedures and assures effective internal control. Regarding the ERP implementation process, the main factors for choosing ERP were the company's strategy, the business's rapid growth and the fact that the previous system could not cope with the increased workload and growth. As many studies demonstrate, and interviewees confirm, ERP introduces improvements in organizational practices, standard procedures across all units (national and international), enabling companies to integrate.

Throughout this work we found that the specific objective of characterizing the necessary steps for BSC implementation was fulfilled, as well as identifying auxiliary tools for BSC implementation, considering that we evidenced that ERP could be a lever for BSC in operational management. Therefore, the third specific objective is also fulfilled, namely to identify the role of ERP in BSC.

Based on the specific objectives, we can answer the research question that guided all this research, namely that the ERP could leverage BSC in operational management, as both tools complement each other.

4. Conclusion

This research shed some light on this relevant managerial topic, emphasizing the research question of this work - can ERP leverage BSC in operational management and its specific objectives: Identify auxiliary tools for BSC implementation; Identify the role of ERP in leveraging BSC – in monitoring of organizational strategy and performance Considering the literature review and the information extracted from the interviews, the perception of the interviewees confirm that ERP has a facilitating role when implementing BSC. The aggregation of information in an ERP enables the consequent implementation of the BSC, as it may serve as a ground for BSC's feature, namely to provide communication and monitoring of organizations' strategy.

BSC and ERP are both integrated management systems that complement each other, as ERP has the ability to manage and interconnect data of the entire business and BSC, on the other hand, needs this vast information provided by ERP in order to measure performance in its perspectives.

Based on this research we noted that both ERP and BSC foster organizational practices and common procedures, leveraging organizational excellence. Considering that a meaningfully positive correlation among economic, legal, ethical and discretionary social responsibility and development of SMEs are recognized, the

implementation of ERP could also foster SMEs performance [107]. Before implementing ERP, SME managers have to solidify knowledge and their processes [108]. Furthermore, it is noted, that even in start-ups the implementation of ERP is relevant, but investigations dedicated to startups are rare [109]. However, highly qualified and motivated human resources are needed. As a long implementation process needs to be avoided, concrete and realistic objectives have to be defined. Furthermore, top management needs to support and provide a good environment, giving confidence to their employees and ensuring that the goals to be achieved are understood by all.

To sum up Big Data might be an opportunity to adapt fundamentals and business strategy of the company as it demands for information management and consequently professional growth and positioning [110].

Along this research, interviewees have confirmed that the ERP might be a powerful tool to foster the BSC implementation with success, so as the future monitoring and control of performance. Considering that based on the ERP, all organizational information is available and gathered together, BSC will be able to have on-time and online information to guide managers to take decisions ad to follow organizational performance. The aggregation is crucial and might, for some tasks, reduce the work of the accountants [110]. Therefore, accountants would have the possibility to concentrate on analysis and other more analytical tasks [19]. Bearing this adaptation in mind also academia should adjust their lecturing programmers to focus on more analytical skills [110].

Currently, BSC implementation in Portugal is still in a growth phase, so several studies should be carried out in the future to provide a broader knowledge of these management tools [24], [97]. Despite the recognition of BSC and the fact that several organizations have successfully implemented BSC, there are other organizations that have failed to implement it [16], [111], [112] so we suggest that on hand of the interlinkage with ERP these barriers night be overcome.

Limitations and future research

This empirical study was performed based on a single Case Study, even if it enriches the research with detailed information, it does have a clear limitation related to the single focus on this organisation. Indeed, we selected to research focusing on a single case study, in order to have in depth analysis, as to our understanding, in this field, is still missing detailed information.

Further Case Studies of organisations that already have the ERP and BSC implemented should be performed to confirm the perception of the entrepreneurs of this Case Study.

Acknowledgement

The author Cidália Oliveira gratefully acknowledge REMIT (Research on Economics Management and Information Technologies), Porto, Portugal financial support from

National Funds of the FCT – Portuguese Foundation for Science and Technology within the project UIDB/05105/2020.

The author Rui Silva acknowledged CETRAD (Centre for Transdisciplinary Development Studies) and under the project UIDB/04011/2020 (https://doi.org/10.54499/UIDB/04011/2020), NECE-UBI, Research Centre for Business Sciences, Research Centre under the project UIDB/04630/2022 and by FCT under project CEECINST/00127/2018/CP1501/CT0010.

References

[1] R. S. Kaplan and S. R. Anderson, "Time-Driven Activity- Based Costing," *Harv. Bus. Rev.*, vol. 82, no. 11, pp. 131–138, 2004, [Online]. Available: https://hbr.org/2004/11/time-driven-activity-based-costing.

[2] P. M. Senge, "The Leader's New Work: Building Learning Organization," *Sloan Manage. Rev.*, vol. 23, pp. 1–17, 1990.

[3] D. J. Teece, G. Pisano, and A. Shuen, "Dynamic capabilities and strategic management," *Strateg. Manag. J.*, vol. 18, no. 7, pp. 509–533, 1997, doi: Doi 10.1002/(Sici)1097-0266(199708)18:7<509::Aid-Smj882>3.0.Co;2-Z.

[4] R. S. Kaplan and D. P. Norton, *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Boston: Harvard Business Press., 2004.

[5] D. Romero and F. Vernadat, "Enterprise information systems state of the art: Past, present and future trends," *Comput. Ind.*, vol. 79, pp. 3–13, 2016, doi: 10.1016/j.compind.2016.03.001.

[6] E. L. Wagner, J. Moll, and S. Newell, "Accounting logics, reconfiguration of ERP systems and the emergence of new accounting practices: A sociomaterial perspective," *Manag. Account. Res.*, vol. 22, no. 3, pp. 181–197, 2011, doi: 10.1016/j.mar.2011.03.001.

[7] A. Kanellou and C. Spathis, "Accounting benefits and satisfaction in an ERP environment," *Int. J. Account. Inf. Syst.*, vol. 14, no. 3, pp. 209–234, 2013, doi: 10.1016/j.accinf.2012.12.002.

[8] A. Neely, K. Platts, A. Neely, M. Gregory, and K. Platts, "Performance measurement system design: A literature review and research agenda," *Int. J. Oper. Prod. Manag.*, vol. 25, no. 12, pp. 1228–1263, 2005, doi: 10.1108/01443570510633639.

[9] S. Tangen, "Performance measurement: from philosophy to practice," *Int. J. Product. Perform. Manag.*, vol. 53, no. 8, pp. 726–737, 2004.

[10] B. Marr and G. Schiuma, "Business performance measurement - past , present and future," *Manag. Decis.*, vol. 41, no. 8, pp. 680–687, 2003, doi: 10.1108/00251740310496198.

[11] M. R. Cabrita, V. C. Machado, and A. Grilo, "Leveraging knowledge management with the balanced scorecard," in *IEEM2010 - IEEE International Conference on Industrial Engineering and Engineering Management*, 2010, pp. 1066–1071, doi: 10.1109/IEEM.2010.5674245.

[12] R. Lawson, D. Desroches, and T. Hatch, *Scorecard Best Practices: Design, Implementation, and Evaluation*. New Jersey: John Wiley & Sons., 2008.

[13] M. Lipe and S. Salterio, "The Balanced Scorecard: Judgemental Effects of Common and Unique Performance Measures," *Account. Rev.*, vol. 75, no. 3, pp. 283–298, 2000.

[14] C. Ittner and D. Larcker, "Customer safisfaction and future financial performance discussion of are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction," *J. Account. Res.*, vol. 36, p. 37, 1998, [Online]. Available: http://proquest.umi.com/pqdweb?did=40003711&Fmt=7&clientId=23852&RQT=309&VName=PQD.

[15] C. D. Ittner, D. F. Larcker, and M. W. Meyer, "Subjectivity and the Weighting of Performance Measures," *Account. Rev.*, vol. 78, no. No. 3, p. 34, 2003, doi: 10.2308/accr.2003.78.3.725.

[16] T. Speckbacher, G., Bischof, J. and Pfeiffer, "A descriptive analysis on the implementation of Balanced Scorecards in German-speaking countries," *Manag. Account. Res.*, vol. 14, pp. 361–387, 2003.

[17] A. Salamzadeh, "Five approaches toward presenting qualitative findings," *J. Int. Acad. Case Stud.*, vol. 26, no. 3, pp. 1–2, 2020.

[18] G. Cokins, "Enterprise Performance Management (EPM) and the Digital Revolution," *Perform. Improv.*, vol. 56, no. 4, pp. 14–19, 2017, doi: 10.1002/pfi.21698.

[19] P. Rikhardsson and O. Yigitbasioglu, "Business intelligence & analytics in management accounting research: Status and future focus," *Int. J. Account. Inf. Syst.*, vol. 29, no. June 2016, pp. 37–58, 2018, doi: 10.1016/j.accinf.2018.03.001.

[20] R. S. Kaplan and D. P. Norton, "The balanced scorecard - measures that drive performance," *Harvard Buiness Rev.*, no. January-February, pp. 71–79, 1992.

[21] R. S. Kaplan and D. P. Norton, "Leading change with the balanced scorecard," *Financ. Exec.*, vol. 17, no. 6, pp. 64–66, 2001, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=11861223&site=ehost-live.

[22] D. Rigby, "Management Tools and Techniques: A survey," *Calif. Manage. Rev.*, vol. 43, no. 2, pp. 139–160, 2001.

[23] P. R. Quesado and L. L. Rodrigues, "Factores Determinantes na Implementação do Balanced Scorecard em Portugal," *Rev. Universo Contábil*, vol. 5, no. 4, pp. 135–146, 2009, doi: 10.4270/ruc.2009433.

[24] C. P. de Oliveira, "Balanced Scorecard, Cultura Organizacional e Desempenho: O Caso das Maiores Exportadoras de Portugal," 2018.

[25] S. Bose and K. Thomas, "Article information :," *J. Intellect. Cap.*, vol. 6, no. 2, pp. 267–284, 2007.

[26] A. Lima, A. Cavalcanti, and V. Ponte, "Da onda da Gestão da Qualidade a uma Filosofia da Qualidade da Gestão : Balanced Scorecard Promovendo Mudanças," *Rev. Contab. Finanças*, no. Edição Especial, pp. 79–94, 2004, doi: 10.1590/S1519-70772004000400006.

[27] R. S. Kaplan and D. P. Norton, *The Balanced Scorecard: Translating Strategy into Action.* Boston: Harvard Business School Press, 1996.

[28] P. DiMaggio, "An Ecological Perspective on Nonprofit Research," in *The Study of the Nonprofit Enterprise: Theories and Approaches*, H. K. Anheier and A. Ben-Ner, Eds. Boston, MA: Springer US, 2003, pp. 311–320.

[29] R. S. Kaplan and D. P. Norton, "Using the balanced scorecard as a strategic management system," *Harv. Bus. Rev.*, vol. 85, no. February 1996, pp. 150–161, 1996, doi: 10.1108/10878570410699825.

[30] R. Poll, "Performance , Processes and Costs : Managing Service Quality with the Balanced Scorecard," *Libr. Trends*, vol. 49, no. 4, pp. 709–717, 2001, doi: Article.

[31] S. Mooraj, D. Oyon, and D. Hostettler, "The balanced scorecard: a necessary good or an unnecessary evil?," *Eur. Manag. J.*, vol. 17, no. 5, pp. 481–491, 1999, doi: 10.1016/S0263-2373(99)00034-1.

[32] D. J. Cooper, M. Ezzamel, and S. Qu, "Creating and poularizing a management accounting idea: The case of the Balanced Scorecard," *Unpubl. Pap. Alberta Sch. Bus.*, 2011.

[33] D. Ø. Madsen and T. Stenheim, "Perceived problems associated with the implementation of the balanced scorecard: Evidence from Scandinavia," *Probl. Perspect. Manag.*, vol. 12, no. 1, pp. 121–131, 2014.

[34] D. Ø. Madsen and T. Stenheim, "The Balanced Scorecard: A Review of Five Research Areas," *Am. J. Manag.*, vol. 15, no. 2, pp. 24–41, 2015.

[35] E. N. Johnson, P. M. J. Reckers, and G. D. Bartlett, "Influences of Timeline and Perceived Strategy Effectiveness on Balanced Scorecard Performance Evaluation Judgments.," *J. Manag. Account. Res.*, vol. 26, no. 1, pp. 165–184, 2014, doi: 10.2308/jmar-50639.

[36] M. Donangelo, G. Macedo, and D. Moreto, "Implementação de indicadores estratégicos : proposta de uma ferramenta simples e eficiente," in *Encontro Nacional de Engenharia de Produção*, 2004, pp. 3375–3380.

[37] M. Green, J. Garrity, A. Gumbus, and B. Lyons, "Pitney Bowes calls for new metrics," *Strateg. Financ.*, vol. May, pp. 30–35, 2002.

[38] D. Madsen and T. Stenheim, "The Balanced Scorecard: A Review of Five Research Areas," *Am. J. Manag.*, vol. 15, no. 2, pp. 24–41, 2015.

[39] Ľ. Lesáková and K. Dubcová, "Knowledge and Use of the Balanced Scorecard Method in the Businesses in the Slovak Republic," *Procedia - Soc. Behav. Sci.*, vol. 230, no. May, pp. 39–48, 2016, doi: 10.1016/j.sbspro.2016.09.006.

[40] M. L. Frigo and K. R. Krumwiede, "The balanced scorecard: A winning performance measurement system," *Strateg. Financ.*, vol. 81, no. January, pp. 50–54, 2000.

[41] R. S. Kaplan and D. P. Norton, "Knowing the score.," *Financ. Exec.*, vol. 6, no. December, pp. 30–33, 1996.

[42] A. Gumbus and R. N. Lussier, "Entrepreneurs Use a Balanced Scorecard to Translate Strategy into Performance Measures," *J. Small Bus. Manag.*, vol. 44, no. 3, pp. 407–425, 2006, doi: 10.1111/j.1540-627X.2006.00179.x.

[43] K. F. Cross and R. L. Lynch, "For good measure," *Eur. Manag. J.*, vol. 66, no. 3 OP-CMA-the Management Accounting Magazine. April 1992, Vol. 66 Issue 3, p20, 4 p. chart, p. 20, 1992, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsgao&AN=edsgcl.12630255.

[44] W. B. Tayler, "The balanced scorecard as a strategy-evaluation tool: The effects of implementation involvement and a causal-chain focus," *Account. Rev.*, vol. 85, no. 3, pp. 1095–1117, 2010, doi: 10.2308/accr.2010.85.3.1095.

[45] O. J. Evangelista de Barros and C. de A. Wanderley, "Adaptation of the balanced scorecard: Case study in a fuel distribution company," *Rev. Contab. Finanças*, vol. 27, no. 72, pp. 320–333, 2016, doi: 10.1590/1808-057x201602200.

[46] M. Chavan, "The balanced Scorecard: a new challenge," *J. Manag. Dev.*, vol. 28, no. 5, pp. 393–406, 2009.

[47] R. S. Kaplan and D. P. Norton, "Having Trouble With Your Strategy," *Harv. Bus. Rev.*, vol. 5, pp. 167–176, 2000.

[48] R. S. Kaplan and D. P. Norton, "The Strategy- Focused Organization," *Strateg. Leadersh.*, vol. 29, no. 3, pp. 41–42, 2001.

[49] Y. L. Chan, "Performance Measurement and adoption of balanced scorecards: A survey of municipal governments in the USA and Canada," *Int. J. Public Sect. Manag.*, vol. 17, no. 3, pp. 204–221, 2004.

[50] D. Northcott and T. Ma'amora Taulapapa, "Using the balanced scorecard to manage performance in public sector organizations: Issues and challenges," *Int. J. Public Sect. Manag.*, vol. 25, no. 3, pp. 166–191, 2012, doi: 10.1108/09513551211224234.

[51] J. Strohhecker, "Factors influencing strategy implementation decisions: an evaluation of a balanced scorecard cockpit, intelligence, and knowledge," *J. Manag. Control*, vol. 27, no. 1, pp. 89–119, 2016, doi: 10.1007/s00187-015-0225-y.

[52] R. S. Kaplan, "Leading Change with the Strategy Execution System," *Harvard Bus. Publ.*, vol. 12, no. 6, pp. 1–16, 2010.

[53] Ľ. Lesáková and K. Dubcová, "Knowledge and Use of the Balanced Scorecard Method in the Businesses in the Slovak Republic," *Procedia - Soc. Behav. Sci.*, vol. 230, no. May, pp. 39–48, 2016, doi: http://dx.doi.org/10.1016/j.sbspro.2016.09.006.

[54] R. S. Kaplan and D. P. Norton, "Linking the balanced scorecard to strategy," *Calif. Manage. Rev.*, vol. 39, no. I, pp. 53–79, 1996, doi: 10.1016/S0024-6301(96)00116-1.

[55] B. Hu, U. Leopold-Wildburger, and J. Strohhecker, "Strategy Map Concepts in a Balanced Scorecard Cockpit Improve Performance," *Eur. J. Oper. Res.*, vol. 0, pp. 664–676, 2017, doi: http://dx.doi.org/10.1016/j.ejor.2016.09.026.

[56] R. Lueg and L. Vu, "Success Factors in Balanced Scorecard Implementations - A Literature Review **," *Manag. Rev.*, vol. 26, no. 4, pp. 306–327, 2015, doi: 10.1688/mrev-2015-04-Lueg.

[57] R. H. Chenhall, "Management control systems design within its organizational context: findings from contingency-based research and directions for the future.," *Account. , Organ. Soc. Organ. Soc.*, vol. 28, pp. 127–168, 2003.

[58] R. Lueg and L. Vu, "Success Factors in Balanced Scorecard Implementations – a Literature Review," *Manag. Rev.*, vol. 26, no. 4, pp. 306–327, 2015, doi: 10.1688/mrev-2015-04-Lueg.

[59] D. Ø. Madsen and T. Stenheim, "Perceived benefits of balanced scorecard implementation: Some preliminary evidence," *Probl. Perspect. Manag.*, vol. 12, no. 3, pp. 81–90, 2014.

[60] G. Speckbacher, J. Bischof, and T. Pfeiffer, "A descriptive analysis on the implementation of Balanced Scorecards in German-speaking countries,"

Manag. Account. Res., vol. 14, no. 4, pp. 361–387, 2003, doi: 10.1016/j.mar.2003.10.001.

[61] R. H. Chenhall and K. Langfield-Smith, "The relationship between strategic priorities, management techniques and management accounting: an empirical investigation using a systems approach," *Accounting, Organ. Soc.*, vol. 23, no. 3, pp. 243–264, 1998, doi: 10.1016/S0361-3682(97)00024-X.

[62] A. Wicks, L. S. St. Clair, and C. Kinney, "Competing values in healthcare: balancing the (Un) balanced scorecard/practitioner application," *J. Healthc. Manag.*, vol. 52, no. 5, pp. 309–325, 2007.

[63] J. Fijałkowska and C. Oliveira, "Balanced Scorecard in Universities," *J. Intercult. Manag.*, vol. 10, no. 4, pp. 57–83, 2018, doi: 10.2478/joim-2018-0025.

[64] A. Oliveira, C., Pinho, J., Silva, "The relevance of learning and growth in organizations that adopt and do not adopt the bsc- characterization of the cultural profile," *Revi sta Eletrônica Gestão Soc.*, vol. 12, no. 33, pp. 2584–2602, 2018.

[65] A. C. Maltz, A. J. Shenhar, and R. R. Reilly, "Beyond the balanced scorecard: Refining the search for organizational success measures," *Long Range Plann.*, vol. 36, no. 2, pp. 187–204, 2003, doi: 10.1016/S0024-6301(02)00165-6.

[66] R. S. Kaplan and D. P. Norton, "Using the Balanced Scorecard as a Strategic Management System," *Harv. Bus. Rev.*, vol. 74, no. January-February, pp. 75–85, 1996, doi: 10.1016/S0840-4704(10)60668-0.

[67] H. M. H. Mingming, W. X. W. Xiong, L. S. L. Shun, D. Y. D. Yingliu, Y. Q. Y. Qingquan, and T. Q. T. Quanfu, "Research on Performance Evaluation of Logistics Enterprises Based on the Balanced Scorecard," *Intell. Comput. Technol. Autom. (ICICTA), 2010 Int. Conf.*, vol. 3, no. 1, 2010, doi: 10.1109/ICICTA.2010.241.

[68] A. A. Atkinson, H. Waterhouse, John, and R. B. Wells, "A Stakeholder Approach to Strategic Performance Measurement," *Sloan Management Review*, MIT Press, 1997.

[69] O. Zizlavsky, "The Balanced Scorecard : Innovative Performance Measurement and Management Control System," *J. Technol. Manag. Innov.*, vol. 9, no. 3, pp. 210–222, 2014.

[70] L. Edvinsson, "Developing intellectual capital at Skandia," *Long Range Plann.*, vol. 30, no. 3, pp. 366–373, 1997, doi: 10.1016/S0024-6301(97)90248-X.

[71] R. S. Kaplan and D. P. Norton, "The Balanced Scorecard - Masures That Drive Performance," *Harvard Buiness Rev.*, vol. 70, no. January-February, pp. 71–79, 1992.

[72] R. S. Kaplan and D. Norton, "Linking the balanced scorecard to strategy (Reprinted from the Balanced Scorecard)," *Calif. Manage. Rev.*, vol. 39, no. I, pp. 53-, 1996.

[73] M. Kennerley and A. Neely, "Measuring performance in a changing business environment," *Int. J. Oper. Prod. Manag.*, vol. 23, no. 2, pp. 213–229, 2003, doi: 10.1108/01443570310458465.

[74] J. Gomes and M. Romão, "How benefits management helps balanced scorecard to deal with business dynamic environments," *Tour. Manag. Stud.*, vol. 9, no. 1, pp. 129–138, 2013.

[75] N. Bontis and N. C. Dragonetti, "The Knowledge Toolbox : A Review of the Tools Available to Measure and Manage Intangible Resources," *Eur. Manag. J.*, vol. 17, no. 4, pp. 391–402, 1999, doi: 10.1016/S0263-2373(99)00019-5.

[76] S. C. Voelpel, M. Leibold, and R. a. Eckhoff, "The tyranny of the Balanced Scorecard in the innovation economy," *J. Intellect. Cap.*, vol. 7, no. 1, pp. 43–60, 2006, doi: 10.1108/14691930610639769.

[77] G. Hubbard, "Measuring organizational performance: Beyond the triple bottom line," *Bus. Strateg. Environ.*, vol. 18, no. December 2006, pp. 177–191, 2009, doi: Doi 10.1002/Bse.564.

[78] R. S. Kaplan and R. Cooper, *Cost & effect: using integrated cost systems to drive profitability and performance*. Harvard Business Press, 1998.

[79] N. Dechow and J. Mouritsen, "Enterprise resource planning systems, management control and the quest for integration," *Accounting, Organ. Soc.*, vol. 30, no. 7–8, pp. 691–733, 2005.

[80] S. C. Voelpel, M. Leibold, R. A. Eckhoff, T. H. Davenport, D. O 'donnell, and L. B. Henriksen, "The Tyranny of the Balanced Scorecard in the Innovation Economy – Intellectual Capital Stream CMS4 The Tyranny of the Balanced Scorecard in the Innovation Economy Intellectual Capital Stream," pp. 1–19, 2005.

[81] M. Liguori, S. Rota, and I. Steccolini, "Public sector accounting reforms: a discourse analysis perspective. Bringing the Gattopardo back to life?," in *Public sector accounting reforms: a discourse analysis perspective. Bringing the Gattopardo back to life?*, IRSPM (International Research Society for Public Management), 2011.

[82] J. Vasilev, "The change from ERP II to ERP III systems," 2013.

[83] J. Gammelgaard, F. Mcdonald, A. Stephan, and H. Tu, "The impact of increases in subsidiary autonomy and network relationships on performance," *Int. Bus. Rev.*, vol. 21, pp. 1158–1172, 2012, doi: 10.1016/j.ibusrev.2012.01.001.

[84] M. Alves and S. Matos, "An Investigation into the Use of ERP Systems in the Public Sector," *J. Enterp. Resour. Plan. Stud.*, vol. 2011, pp. 1–6, 2011, doi: 10.5171/2011.950191.

[85] R. Parlakkaya, H. Cetin, and H. Akmese, "Enterprise resource planning systems' impact on accounting processes in Turkey: A research on the largest 500 industrial firms," *Bus. Rev.*, vol. 17, no. 2, pp. 167–174, 2011.

[86] C. Spathis and J. Ananiadis, "Assessing the benefits of using an enterprise system in accounting information and management," *J. Enterp. Inf. Manag.*, vol. 18, no. 2, pp. 195–210, 2005.

[87] H. Davenport Thomas, "Putting the Enterprise into the Enterprise System," *Harv. Bus. Rev.*, pp. 121–132, 1998, [Online]. Available: http://facweb.cti.depaul.edu/jnowotarski/is425/hbr enterprise systems davenport 1998 jul-aug.pdf.

[88] M. Granlund, "Extending AIS research to management accounting and control issues: A research note," *Int. J. Account. Inf. Syst.*, vol. 12, no. 1, pp. 3–19, 2011.

[89] J. Frei, D. Greiling, and J. Schmidthuber, "Reconciling field-level logics and management control practices in research management at Austrian public universities," *Qual. Res. Account. Manag.*, 2022, doi: 10.1108/QRAM-10-2020-0167.

[90] D. Chand, G. Hachey, J. Hunton, V. Owhoso, and S. Vasudevan, "A balanced scorecard based framework for assessing the strategic impacts of ERP systems," *Comput. Ind.*, vol. 56, pp. 558–572, 2005, doi: 10.1016/j.compind.2005.02.011.

[91] A. Czuchry, M. Yasin, and D. L. Khuzhakhmetov, "Enhancing Organizational Effectiveness Through the Implementation of Supplier Parks: the Case of the Automotive Industry," *J. Int. Bus. Res.*, vol. 8, no. 1, pp. 45–61, 2009, [Online]. Available: http://search.proquest.com/docview/215465478?accountid=12253.

[92] I. C. K. Drongelen and J. Bilderbeek, "R&D performance measurement: More than choosing a set of metrics," *R&D Manag.*, vol. 29, no. 1, p. 35, 1999, doi: 10.1111/1467-9310.00115.

[93] G. Samkin, G. Baldvinsdottir, J. Burns, H. N∅rreklit, and R. Scapens, "Professional accounting media: accountants handing over control to the system," *Qual. Res. Account. Manag.*, vol. 7, no. 3, pp. 395–414, 2010, doi: 10.1108/11766091011072819.

[94] A. Neely and M. Al Najjar, "Management learning not management control: The true role of performance measurement?," *Calif. Manage. Rev.*, vol. 48, no. 3, 2006, doi: 10.2307/41166352.

[95] G. Creamer and Y. Freund, "Learning a board Balanced Scorecard to improve corporate performance," *Decis. Support Syst.*, vol. 49, no. 4, pp. 365–385, 2010, doi: 10.1016/j.dss.2010.04.004.

[96] J. P. Kottler and J. L. Heskett, *Corporate Culture and Performance*. Free Pass, 1992.

[97] P. Quesado, B. Guzmán, and L. Rodrigues, "Fatores Determinantes da Implementação do Balanced Scorecard em Portugal: evidência empírica em organizações públicas e privadas," *Rev. Bus. Manag.*, vol. 16, no. 51, pp. 199–222, 2014, doi: 10.7819/rbgn.v16i51.1335.

[98] G. Dube, L. e Parê, "RIGOR IN INFORMATION SYSTEMS POSITIVIST CASE R ESEARCH : CURRENT PRACTICES, TRENDS AND RECOMENDATIONS," *MIS Q.*, vol. 27, no. 4, pp. 597–635, 2003.

[99] R. K. Yin, *Qualitative research from start to finish*. Guilford Publications, 2015.

[100] C. C. Prodanov and E. C. De Freitas, *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. Editora Feevale, 2013.

[101] L. Saletti-cuesta *et al.*, "No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title," *Sustain.*, vol. 4, no. 1, pp. 1–9, 2020, [Online]. Available: https://pesquisa.bvsalud.org/portal/resource/en/mdl-20203177951%0Ahttp://dx.doi.org/10.1038/s41562-020-0887-9%0Ahttp://dx.doi.org/10.1038/s41562-020-0884-z%0Ahttps://doi.org/10.1080/13669877.2020.1758193%0Ahttp://sersc.org/journals/index.php/IJAST/article.

[102] B. Ryan, R. W. Scapens, and M. Theobold, "Research Methof and Methodology in Finance and Accounting," *Second Ed.*, 2002, doi: 10.1186/1687-6180-2011-78.

[103] R. W. Scapens, *Doing Case Study Research*. 2004.

[104] M. Q. Patton, *Qualitative evaluation and research methods*. SAGE Publications, inc, 1990.

[105] A. Rom and C. Rohde, "Management accounting and integrated information systems: A literature review," *Int. J. Account. Inf. Syst.*, vol. 8, no. 1, pp. 40–68, 2007, doi: 10.1016/j.accinf.2006.12.003.

[106] P. Quesado and L. Rodrigues, "Fatores Determinantes Na Implementação Do Balanced Scorecard Em Portugal Determining

Factors of the Bsc Implementation in Portugal," *Rev. Universo Contábil*, no. 54, pp. 94–115, 2009, doi: 10.4270/ruc.2009433.

[107] M. Doshmanli, Y. Salamzadeh, and A. Salamzadeh, "Development of SMEs in an emerging economy: does corporate social responsibility matter?," *Int. J. Manag. Enterp. Dev.*, vol. 17, no. 2, pp. 168–191, 2018.

[108] S. T. Raharjo and M. S. Perdhana, "SME's Enterprise Resource Planning Implementation, Competitive Advantage, and Marketing Performance," *J. Entrep. Bus. Econ.*, vol. 4, no. 1, pp. 22–44, 2016.

[109] A. Salamzadeh, Z. Arasti, and G. M. Elyasi, "Creation of ICT-based social start-ups in Iran: A multiple case study," *J. enterprising Cult.*, vol. 25, no. 01, pp. 97–122, 2017.

[110] G. Richins, A. Stapleton, T. C. Stratopoulos, and C. Wong, "Big data analytics: Opportunity or threat for the accounting profession?," *J. Inf. Syst.*, vol. 31, no. 3, pp. 63–79, 2017, doi: 10.2308/isys-51805.

[111] C. D. Ittner and D. F. Larcker, "Innovations in Performance Measurement: Trends and Research Implications," *J. Manag. Account. Res.*, vol. 10, no. 2, pp. 205–238, 1998, doi: 10.2139/ssrn.137278.

[112] C. D. Ittner, D. F. Larcker, and T. Randall, "Performance implications of strategic performance measurement in financial services firms," *Accounting, Organ. Soc.*, vol. 28, no. 7–8, pp. 715–741, 2003, doi: 10.1016/S0361-3682(03)00033-3.

10.31341/jios.48.1.3
Open Access
UDC 311.13: 004.912=93C++
Original Scientific Paper

Exploring the Access to the Static Array Elements via Indices and via Pointers — the Introductory C++ Case Expanded

Robert Logožar *robert.logozar@unin.hr*
Dpt. of Electrical Engineering
University North, Varaždin, Croatia

Matija Mikac *matija.mikac@unin.hr*
Dpt. of Electrical Engineering
University North, Varaždin, Croatia

Danijel Radošević *darados@foi.unizg.hr*
Faculty of Organization and Informatics
University of Zagreb, Varaždin, Croatia

Abstract

We revisit the old but formally still unresolved debate on the time efficiency of accessing the elements of 1D arrays via indices versus accessing them via pointers. To analyze that, we have programmed benchmarks of minimal complexity in the C++ language and inspected the machine code of their compilation in the x86 assembly language. Before exploring the performance, we briefly compared a few methods used for the execution time measurements. The results on the Wintel platform show no significant advantage in using pointers over indices except for some benchmarks and array (data) types. In other cases, the exact opposite may be true. The cause of this inconsistency lies in the compilation of the source code into the rather nonorthogonal x86 instruction set. Furthermore, the execution speed does not clearly relate to the instruction length. The parallel aim of this work is to provide a ground for further analysis and measurements of this kind using different compilers, languages, and computer platforms.

Keywords: static arrays, pointers, C/C++, accessing the array elements, benchmarks, execution time measurements.

1. Introduction

Algorithms and data structures are the very foundation of software. The simplest and most ubiquitous data structure, which resembles the organization of the computer's main memory itself, is the *array*. It is unavoidable in computer programming, and all general-purpose, high-level languages implement it one way or the other. The simplest way to access the array elements—which follows the mathematical notation of vectors and matrices—is by "subscripting" the arrays with their indices. In Pascal and the C language, this became possible also via pointers, which are the addresses of

the defined data types. The pointers allowed programmers to directly manipulate the memory addresses and their contents in a similar way possible in assembly languages but having left the organizational details to the compiler.

The fathers of the C language, B. Kernighan and D. Ritchie, in their C language "bible," known as K&R, devoted the whole of chapter 5 to the pointers and arrays and their relation [1]. In §5.3, they say: "*Any operation that can be achieved by array subscripting can also be done with pointers. The pointer version will in general be faster but, at least to the uninitiated, somewhat harder to understand.*"

This claim must be echoing in the minds of many computer scientists and practical programmers who have read valuable classical literature and cared to write the most efficient code. Namely, if we follow its implicit suggestion, should we insist on accessing the array elements by using pointers? If so, what are the time efficiency improvements? Kernighan and Ritchie did not support their statement with any explanation or proof. They have even relativized it in the further elaboration on the array element access in the rest of that chapter and the textbook. Besides that, when trying to find out if someone else attempted to corroborate or dispute this thesis, we could not find any other systematic work or firm results covering this topic.

Having said that, this rather elementary dilemma may still intrigue the minds of many pedantic programmers who want to ensure that they write programming code that is as fast as possible. This may also intrigue the minds of the teachers who want to motivate their students to grasp the concept of pointers. In either case, to answer that properly, one should get a deeper insight into the topic and reach the final verdict only after the concrete time measurements. That is, for the appropriately tailored benchmarks of the two approaches, we must measure their *benchmark execution times*, which we will abbreviate as BETs.

We did some simple, preliminary time measurements roughly a decade ago. They showed that the access to the elements of one-dimensional arrays of some integer types (`short int`, `int`) via pointers was approximately 15% to 20% faster than the access via indices. We did not pursue the measurements more systematically nor did we investigate the causes of this behavior at that time. We left that — seemingly trivial research — for some "future work," which finally continues with this paper. Thus, our first aim is to provide an introduction for a detailed analysis of the provided benchmarks and their execution time measurements. At the same time, we will expose the first results obtained by using a modern integrated developing environment (IDE) on a relatively contemporary computer and today's common operating system (OS).

In this text, we expand our preliminary and abbreviated description of the topic given in the conference paper [2] and provide the full version of this exploration. Besides correcting several (minor) errata and noticed omissions, it includes a more detailed description of the results of compilation and the analysis of the obtained machine instructions, as well as a comprehensive insight into the measurement results.

Concerning the outline of this text, in section 2, we shortly expose the basics of the array data structure and the pointer data type and their implementation in the C/C++ languages. Section 3 presents and explains our benchmarks and exposes their assembly language code, often paying attention to several minute details to better understand the following performance measurements. Section 4 discusses the

methods, program setup, and prerequisites for the time measurements. There, we present and discuss the results of our BET measurements for assigning and retrieving array elements of several data types. Section 5 concludes this paper and opens several new directions and topics for future work.

2. Arrays and Pointers

Before elaborating on our benchmark details, we will briefly outline the basics of the arrays and pointers, emphasizing their implementation in C/C++.

2.1. Arrays

One-dimensional (1D) arrays serve for storing the series of n equal a_i elements, where the nonnegative integer index i spans through n consecutive values. For example, with $i = 1, 2, \dots, n$, the elements a_i could be interpreted as the components of vector $\boldsymbol{a}$ in an n-dimensional space. Following this mathematical notation, the standard syntax for accessing the array elements in high-level programming languages requires stating the array name and the element index. Based on that, the compilers ensure the run-time calculation of i-th array element memory address as:

$$A(i) = A_0 + l_T \times (i - i_0), \quad i = i_0,\ i_0 + 1, \dots, i_0 + n - 1\ . \tag{1}$$

Here, A_0 is the address of the starting element, the one with index i_0, and l_T is the size of the element data type (T) in the number of bytes (B)—here, of course, in their original meaning of (the size of) a basic memory location.

The value $x(i)$ of the i-th element is the memory $M_T\big(A(i)\big)$ content of the address $A(i)$ interpreted as the data type T,

$$x(i) = M_T\big(A(i)\big)\ . \tag{2}$$

2.2. C/C++ Arrays and Pointers

The C and C++ languages fix the starting index to $i_0 = 0$, so its value need not be subtracted from i. This leads to the simplest possible formula and the fastest possible address calculation:

$$A(i) = A_0 + l_T \times \mathrm{i}, \quad i = 0, 1, \dots, n - 1. \tag{3}$$

As hinted in the introduction (§1), the elements' addresses and contents can also be operated by the variable of the special, *pointer data type*, usually simply called *pointers*. They hold the addresses with assigned data types and can also be described as *typed addresses*.

Altogether, this leads to the three possible ways of accessing the array elements. We immediately write them in the syntax of the C/C++ languages [1] (1988), [3], [4], in the same way as they appear in the benchmarks' source code in Listing 1:

1) by using the element index: `iX[i]`;
2) by using the pointer arithmetic and then by dereferencing the obtained pointer: `*(pI0 + i)`;
3) by incrementing the pointer and dereferencing it: `*(++pI)`.

The first two ways are of the *random access* kind because they can access array elements directly via their indices, regardless of the previously accessed element. The third access applies only to the passage through successive array elements. To enable the comparison of the time efficiency between all three access types, here we restrict the consideration to only the successive passages through the array elements for all the access types. Of course, this kind of passage is also the most common in practice.

The above general deliberation is applicable to both *static* and *dynamic* arrays, but in the further text, we restrict our consideration to the former.

3. Our Benchmarks

In this section, we describe and analyze our benchmark routines. They are the core of our three C++ test programs created in MS Visual Studio, Community Version 2019 (further on, VS-CV 2019). The slight variations between the programs served to investigate the many peculiarities on which the execution times depend (§4.3). Because of the repetitive code for each data type, the programs built up to a bit more than 1000 lines each.

3.1. Benchmark Source Code

Listing 1 shows *our standard benchmarks* for the int type of 1D arrays. To avoid the user stack overflow, very large arrays—like ours—must be declared static or global. In each version of our test program, there are benchmark loops similar to those in Listing 1, but for the arrays of all the six standard data types — four integer and two floating-point types:

- `char` (1B), `short int` = `short` (2B), `int` (4B) (presented in Listing 1), and `long long int` = abbr. `llint` (8B);
- `float` (4B), and `double` (8B).

For each of the three described and implemented access methods, there are two benchmarks:

A. for storing a value into the `uI`-th array element, and
B. for retrieving (fetching) a value from the `uI`-th array element into a variable.

To investigate primarily the influence of different element-accessing mechanisms, we have kept the A and B types of operations as simple as possible. In A, instead of storing the quasi-random numbers into the array elements that serve as *l*-values, we assigned them the value of the unchanging `iRVal` variable. In action B, one and only one *l*-value, the variable `iLVal`, is assigned the values of the array elements.

```
 // Compiler Optimizations OFF!
// Definitions:
#define intMid     1111
#define intLrg 22222222
typedef unsigned int uint;
// Declaring the static int array:
const uint cuiN = 20000000;   // cuiN = 20 × 10^6.
uint uiN1 = cuiN;
static int iX[cuiN] = {0, }; // Array with 20 × 10^6 int elements.
iLVal = intMid, iRVal = intLrg;
// Pointers to the 0-th and last element:
int *pI0 = iX,  *pI1 = iX + uiN1;
```

```
// A. Storing into array elements
// 1) Via index, iX[uI]:
for (uint uI = 0; uI < uiN1; ++uI)
     iX[uI] = iRVal;
// 2) Via pointer arithmetic:
for (uint uI = 0; uI < uiN1; ++uI)
     *(pI0 + uI) = iRVal;
// 3) Via incrementing the pointer:
for (int* pI = iX; pI < pI1; ++pI)
     *pI = iRVal;
```

```
// B. Retrieving from array el.
// 1) Via index, iX[uI]:
for (uint uI = 0; uI < uiN1; ++uI)
    iLVal = iX[uI];
// 2) Via pointer arithmetic:
for (uint uI = 0; uI < uiN1; ++uI)
    iLVal = *(pI0 + uI);
// 3) Via incrementing the pointer:
for (int* pI = iX; pI < pI1; ++pI)
    iLVal = *pI;
```

Listing 1. Our standard benchmarks for accessing the elements of a (large) C/C++ `int` array by: 1) indices, 2) pointer arithmetic, 3) incrementing the pointer.

In the release version of the program, the compiler would optimize both actions if set so, especially B, because its outcome is the single assignment: `iLVal = iX[uiN1 - 1]`. That is why these benchmarks must run without any optimization.

The objection holds that such circumstances can be considered artificial and that the arrays could have been filled up in other ways. However, if our *r*- or *l*-values were more complex—to prevent the extreme optimization efficiency—the net effect of compilation without the optimization would have been the same. On the other hand, if the optimization were on, the investigation of the assembly language code (see the next subsection) would not be possible directly in VS-CV 2019, as it is now, for the non-optimized, debug version. Of course, some other versions of the benchmarks may be proposed, but not without complicating them at least a bit. For instance, in the assignment of the values to the array elements (action A), the *r*-value could be an element of another array filled with randomly generated numbers. However, in that case, both the *l*- and *r*-value would require accessing the array elements. Such a benchmark would perform double access to the array elements and thus equalize the actions A and B. While this is worth investigating, we leave it for future work and stick here to the proposed benchmarks with the described elementary actions.

Finally, a word about the use of the increment operator (++) that acts on an isolated *l*-value. We have been systematically using it in the prefix form, mostly for the sake

of uniformity and pedantry.[1] We will discuss other details of the concrete program implementation as needed.

3.2. Benchmark Machine Code Disassembled

When running the programs in the debug mode, the VS IDE enables the in-place presentation of the Intel assembly language instructions in symbolic form after disassembling the machine code of the debug version of the program [5]. The x86 (32-bit) and x64 (64-bit) versions of assembly languages are available, but in this paper, we focus on the still-standard 32-bit version. A concise review of the x86 assembly language can be found in [6], whereas the exhaustive reference is in [7]. For our six benchmark routines (A/B.1, 2, 3) with the array of `int` data type, this is shown in Listing 2.

3.2.1. Compilation of the `for`-loops (`int` arrays)

Of the six `for`-loops from our benchmarks in Listing 1, the first two of the A and B types (A/B.1 and A/B.2) have the standard `for`-loop *heads*, with (rising) indices. Because of the same source form, their compilation results in a series of the same eight (8) machine instructions, as presented for the A.1 `for`-loop head. There are only the expected slight differences in the instruction operands. The general-purpose registers used in A.1 are cyclically permuted in B.1 as follows: `EAX` → `ECX`, `ECX` → `EDX`, and `EDX` → `EAX`. Additionally, the loop compilations differ in the relative displacements of the jump addresses in their unconditional (`jmp`) and conditional (`jae`) jumps to the: i) loop *conditions* (`cmp` instruction), ii) *exits* (behind the last, `jmp` instruction in the loop body), and iii) *loop-expressions*, i.e., the index increments, starting at the second `mov` instruction.

To summarize, all these `for`-loops consist of eight (8) machine instructions in the loop head and one (1) additional at the end of the loop body, totaling nine (9) instructions, which are stored in the total o 43B (cf. type `int` benchmark A.1, the `for`-loop control part).

Concretely, in the A.1 `for`-loop head, the first `mov` places the initial, `0` value, stored in the instruction itself (immediate addressing mode) to the memory location of the loop index. The index address is expressed relative to the address contained in the base (frame) pointer register, `EBP`, which is constant during the execution of the main function. In the case shown here, `EBP = 212FFA38h`, so the address of the local `uI` index is: `A(uI) = EBP – 7E8h = 212FF250h`.

The next, `jmp` instruction, performs the jump to the condition-checking part of the loop head, which starts with the `mov` instruction at the address `main + 20E3h =`

[1] The discussion about the performance difference between the pre- and post-increment operators on the isolated *l*-values can be found e.g., in [8]. With the prefix `++`, our benchmarks mostly run from 0% to 2.5% faster. However, on rare occasions (for the longer data types!), they are sometimes slower, but within the standard deviation of the measurements (see sec. 4).

```
  // Listing 2
 // A. Storing into array elements
// A.1 Via index: iX[uI]
for (uint uI = 0; uI < uiN1; ++uI)
00C738B8  mov dword ptr [ebp-7E8h],0   ; uI = ML(EBP - 7E8h) ← 0.
00C738C2  jmp main+20E3h (0C738D3h)    ; GOTO 0C738D3h (loop condition).
00C738C4  mov eax,dword ptr [ebp-7E8h] ; EAX ← ML(EBP – 7E8h) = uI.
00C738CA  add eax,1                    ; EAX ← EAX + 1.
00C738CD  mov dword ptr [ebp-7E8h],eax ; uI ← EAX (uI ← uI + 1).
00C738D3  mov ecx,dword ptr [ebp-7E8h] ; ECX ← uI.
00C738D9  cmp ecx,dword ptr [ebp-77Ch] ; tmp ← ECX – ML(EBP – 77Ch) = uI – uiN1.
00C738DF  jae main+2106h (0C738F6h); If tmp ≥ 0 then GOTO 0C738F6h (exit loop).
        iX[uI] = iRVal;
00C738E1  mov edx,dword ptr [ebp-7E8h] ; EDX ← ML(EBP - 7E8h) = uI.
00C738E7  mov eax,dword ptr [ebp-738h] ; EAX ← ML(EBP - 738h) = iRVal (= intLrg).
00C738ED  mov dword ptr iX (0C826A8h)[edx*4],eax; iX[uI] = ML(iX + 4*uI) ← EAX
00C738F4  jmp main+20D4h (0C738C4h)      ; GOTO 0C738C4h (loop expression).
// A.2 Via pointer arithmetic: *(pI0 + uI)
for (uint uI = 0; uI < uiN1; ++uI) // As in A.1.
       *(pI0 + uI) = iRVal;
00C73B5B  mov eax,dword ptr [ebp-7F8h] ; EAX ← ML(EBP – 7F8h) = uI.
00C73B61  mov ecx,dword ptr [ebp-76Ch] ; ECX ← ML(EBP – 76Ch) = pI0 [= A(iX)].
00C73B67  mov edx,dword ptr [ebp-738h] ; EDX ← ML(EBP - 738h) = iRVal (= intLrg).
00C73B6D  mov dword ptr [ecx+eax*4],edx; *(pI0 + uI) = ML(ECX + 4*EAX) ← EDX
00C73B70  jmp main+234Eh (0C73B3Eh)    ; GOTO 0C73B3Eh (loop expression).
// A.3 Via incrementing the pointer: *(++pI)
for (int* pI = iX; pI < pI1; ++pI)
00C73DD1  mov dword ptr [ebp-808h], offset iX (0C826A8h); pI = ML(EBP - 808h) ← A(iX).
00C73DDB  jmp main+25FCh (0C73DECh)    ; GOTO 0C73DECh (loop condition).
00C73DDD  mov edx,dword ptr [ebp-808h] ; EDX ← ML(EBP – 808h) = pI.
00C73DE3  add edx,4                    ; EDX ← EDX + 4.
00C73DE6  mov dword ptr [ebp-808h],edx ; pI ← EDX (pI ← pI + 4).
00C73DEC  mov eax,dword ptr [ebp-808h] ; EAX ← pI.
00C73DF2  cmp eax,dword ptr [ebp-770h] ; tmp ← EAX – ML(EBP – 770h) = pI – pI1.
00C73DF8  jae main+261Ah (0C73E0Ah); If tmp ≥ 0 then GOTO 0C73E0Ah (exit loop).
       *pI = iRVal;
00C73DFA  mov ecx,dword ptr [ebp-808h] ; ECX ← ML(EBP - 808h) = pI
00C73E00  mov edx,dword ptr [ebp-738h] ; EDX ← ML(EBP - 738h) = iRVal (= intLrg).
00C73E06  mov dword ptr [ecx],edx      ; *pI = ML(ECX) ← EDX.
00C73E08  jmp main+25EDh (0C73DDDh)    ; GOTO 0C73DDDh (loop expression).

 // B. Retrieving from array elements
// B.1 Via index: iX[uI]
for (uint uI = 0; uI < uiN1; ++uI) // As in A.1.
        iLVal = iX[uI];
00C73A09  mov eax,dword ptr [ebp-7F0h] ; EAX ← ML(EBP – 7F0h) = uI.
00C73A0F  mov ecx,dword ptr iX (0C826A8h)[eax*4]; ECX ← ML(iX + 4*uI) = iX[uI].
00C73A16  mov dword ptr [ebp-784h],ecx ; iLVal = ML(EBP – 784h) ← ECX.
00C73A1C  jmp main+21FCh (0C739ECh)    ; GOTO 0C739ECh (loop expression).
// B.2 Via pointer arithmetic: *(pI0 + uI)
for (uint uI = 0; uI < uiN1; ++uI) // As in A.1.
        iLval = *(pI0 + uI);
00C73C9B  mov edx,dword ptr [ebp-800h] ; EDX ← ML(EBP – 800h) = uI.
00C73CA1  mov eax,dword ptr [ebp-76Ch] ; EAX ← ML(EBP – 76Ch) = pI0 [= A(iX)].
00C73CA7  mov ecx,dword ptr [eax+edx*4]; ECX ← ML(EAX+4*EDX) = *(pI0 + uI).
00C73CAA  mov dword ptr [ebp-784h],ecx ; iLVal = ML(EBP – 784h) ← ECX.
00C73CB0  jmp main+248Eh (0C73C7Eh)    ; GOTO 0C73B3Eh (loop expression).
```

```
 // Listing 2 continued.
// B.3 Via incrementing the pointer: *(++pI)
for (int* pI = iX; pI < pI1; ++pI) // As in A.3.
          iLval = *pI;
00C73F3D  mov ecx,dword ptr [ebp-810h] ; ECX ← ML(EBP – 7F0h) = pI.
00C73F43  mov edx,dword ptr [ecx]      ; EDX ← ML(ECX) = *pI.
00C73F45  mov dword ptr [ebp-784h],edx ; iLVal = ML(EBP – 784h) ← EDX.
00C73F4B  jmp main+2730h (0C73F20h)    ; GOTO 0C739ECh (loop expression).
```

Listing 2. Intel x86 assembly language code of the crucial parts of our benchmarks from Listing 1. The order of the code snippets is rearranged for the sake of comparison (see §4.2.1). In the code comments, $A(x)$ denotes the address of (variable) x and $M_T(A)$ denotes the contents on the address A interpreted in the data type T.

`00C738D3h`, where `main = 00C717F0h` represents the starting address of the C++ program main function.[2]

This `mov` prepares the operands for the `cmp` instruction behind it by moving the current `uI` index value to the `ECX` register.

Now `cmp` compares `uI` with `uiN1` variable value (placed at the address `EBP – 77Ch`,[3] by executing the operation `tmp ← uI – uiN1`, and setting the flags accordingly (cf. the x86 instruction reference in [6]). The next instruction, `jae`, performs the jump to the loop exit if the previous result is *above* or *equal to* zero. In other words, for `uI < uiN1`, there is no jump, and the loop enters its body. The body contains the statement `iX[uI] = iRVal;`, compiled into three instructions to be discussed soon.

The last instruction in A.1 `for`-loop is the `jmp` to the address `00C738F4h`, which is the unconditional jump to the loop-expression part of the loop head at the address `0C738C4h`. Here, the index `uI` will be incremented. Because in the x86 instruction set there are no explicit arithmetic operations with the operands in memory, the index value is first moved to the `EAX` register, and after that, it is incremented by the instruction `add eax, 1`.

The next (`mov`) instruction places the obtained result as the new value of index `uI`. After that, the execution of the `for`-loop comes again to its condition-checking part, starting with the preparatory `mov` instruction described at the top of this paragraph.

After observing this control part of the `for`-loop more closely, a reader might wonder why the second `mov` in the pair of the move instructions (the one at the address `00C738D3h`) is not omitted, i.e., why the `cmp` instructions were not of the form:

```
cmp dword ptr [ebp-7E8h], dword ptr [ebp-77Ch].
```

[2] From that address on, in our case at the addresses `00C717F0h – 00C71806h`, there are the first seven instructions of the `main` function, which implement the standard calling convention (see e.g., [6]). As the result of some compilations, the symbolic address of the next (`push`) instruction, `__$EncStackInitStart = main + 00C71807h`, can also appear in this place.

[3] In the VS 2019, the address of a local variable can be found by entering the C/C++ expression for getting its address into the address field of one of the four available disassembly memory windows. Here, `uiN1 = cuiN, A(uiN1) = &uiN1 = 212FF2BCh`.

However, that is invalid because the `cmp` does not allow the second operand to be accessed via a memory address! On the other hand, if it were a constant, which would happen if the `cuiN` (declared as `const`) were the upper limit of the index, as in the following condition,

```
for (uint uI = 0; uI < cuiN; ++uI),
```

then the compiler would indeed omit the "extra" `mov` instruction and prepare the `cmp` instruction as:

```
cmp dword ptr [ebp-7E8h], 1312D00h .
```

Here, the constant value of `cuiN = 1312D00h = 20 000 000d` is stored in the instruction itself. This case exemplifies how the x86 set of instructions is *not orthogonal*, i.e., how many operations are not allowed with an arbitrary addressing mode.

In our old benchmarks, we wrote the `for`-loops in just the above way, and their loop heads were translated into the following seven (7) instructions, with their lengths in bytes in parenthesis:

- `mov`(10), `jmp`(2), `mov`(6), `add`(3), `mov`(6), `cmp`(10), `jae`(2), plus
- `jmp`(2) at the end of the loop body,

adding up to eight (8) instructions in 41B.

This total is one (1) instruction and 2B less than in our standard benchmarks (cf. Table 1). However, whether such loops will run faster or not remains to be determined (§4.3.3). Nevertheless, they do not resemble the general case in which the upper limit can and often will be a non-constant value, so we have abandoned them.

A reader can also detect that the index value was incremented in `EAX`, moved back to the `uI` variable, to be moved into another (`ECX`) register. Why? Would it not be better to operate `cmp` with `EAX` as its first operand if the index value is already there? Furthermore, if we upgrade this by changing the first instruction to `mov EAX, 0`, then the "second" `mov` (second after `add`) is not needed at all. This simplification leads to code optimization, which is beyond the scope of this paper.

To get back to the present state of our benchmark loops, in Listing 2 we observe that the `for`-loop in A.3 benchmark is very similar to the previous ones and that the one in B.3 is equivalent. They both have the same number of instructions (8 + 1), with the same lengths as the loops in the first two benchmarks (cf. Table 1). Furthermore, all instructions in the A.3 `for`-loop are of the same type as those in A.1, with just a few minor differences. In the first instruction, the second operand must be addressed differently because it assigns the address of the `iX` array to be the initial value of the `pI` pointer: `pI ← A(iX) = iX = 00C826A8h`. The second instruction executes the jump to the condition-checking part of the loop head (at the address `00C73DECh`), which — as in A.1 — starts with the preparatory `mov` instruction. It moves the value of `pI` into `EAX`. Then `cmp` executes the subtraction `tmp ← pI – pI1`. `jae` checks the flags and if `tmp ≥ 0`, it exits the loop, else (`CF = 1`), the execution proceeds to the loop body. `jmp` at the end of the body directs the execution to the loop-expression, which first moves the value of `pI` to `EDX`, then increases `EDX` for 4, because `sizeof(int) = 4`, and finally stores the enlarged value into `pI`.

Overall, the differences between manipulating the pointer and the indices are only subtle. The compilation of all `for`-loops in our standard benchmarks gives the same number of instructions, with the same operations and the same or very similar addressing modes, resulting in the instructions of the same lengths. From this, one could expect them to execute within the nearly same time, but whether this is so — we still have to see.

3.2.2. Compilation of the array-element-accessing statements (`int` arrays)

In the `for`-loop bodies of all the presented benchmarks, there is just one C/C++ statement. First, we focus on the compilation of this statement in the three benchmarks of type A, in which an *r*-value provided by a variable `iRVal` is stored into the `uI`-th array element of the integer array `iX`.

In the benchmarks A.1 and A.3 for the `int` array, this is accomplished by three (3) and in the benchmark A.2 by four (4) `mov` instructions (cf. Listing 2 and Table 1). Therefore, we compare A.1 and A.3 cases first. In A.1, the three `mov` instructions together are 19B long, and in A.3 (only) 14B. The first two instructions in both benchmarks are not only of the same length but also of the exact same type. In A.1 (A.3) benchmark:

- the first `mov` instruction places the value of the index `uI` (pointer `pI`) into `EDX (ECX)`;
- the second `mov` instruction, in both A.1 and A.3, places the value of the local variable `iRVal = intLrg = 22222222d = 0153158Eh`, stored at the address `A(iRval) = EBP – 738h`, into `EAX (EDX)`.

The difference is only in the third `mov` instruction. In A.1, it moves the value of `EAX (= iRVal)` to the address of the `uI`-th array element, which is formed by the index addressing (eq. 3). In this case:

```
A(iX[uI]) = iX + 4*uI = 00C826A8h + 4*EDX.          (4)
```

It requires a 7B-`mov` instruction, which is 1B longer than the previous two. Namely, besides the information of the source operand (here `EAX`), it must store the information of the index register (`EDX`), the length of the (integer) array element (4), and—as the longest data—the 32-bit address of the `iX` array.

In A.3, the third `mov` instruction in the loop body is much simpler, and because of that, it is 5B shorter. Namely, the address of the `uI`-th array element is already prepared in the `pI` pointer (which was moved into `ECX` in the first `mov` instruction). So, the content of `EDX` moves to the address shown by `pI`.

To summarize this briefly, in the third `mov` instruction, we note an apparent simplification in benchmark A.3, comparing it to benchmark A.1. However, keeping in mind that this 32-bit code will be executed on the 64-bit platform (and particularly on the 64-bit processor!), it remains to see if the shorter and simpler last `mov` instruction will bring some speed benefits.

As for the loop body of the benchmark A.2, the four instructions are 6B, 6B, 6B, and 3B long, in total 21B, i.e., one instruction and 2B (7B) more than in A1 (A3). The first `mov` instruction places the value of `uI` into `EAX`. A careful reader will note that

Instruc. – op. mnem. (bytes)		*A:* `Arr. el. = const r-value`			*B:* `l-value = Afrr. el.`		
Array type	*for-loop*	A.1 `Arr[i]`	A.2 `*(p0+i)`	A.3 `*(++p)`	B.1 `Arr[i]`	B.2 `*(p0+i)`	B.3 `*(++p)`
char	ctrl:	Equiv. to int A.2.	Equiv. to int A.2.	Analog. to int A.3.	Equiv. to int B.1.	Equiv. to int B.2.	Sim./Analg. to int A.2.
	body:	mov(6), mov(6), mov(6): 18B. (Use of CL reg.)	mov(6), add(6), mov(6), mov(2): 20B.	mov(6), mov(6), mov(2): 14B. (Use of CL reg.)	mov(6), mov(6), mov(6): 18B. (Use of CL reg.)	mov(6), add(6), mov(2), mov(6): 20B.	mov(6), mov(2), mov(6): 14B. (Use of AL reg.)
short	ctrl:	Equiv. to int A.1.	Equiv. to int A.1.	Analog. to int A.3.	Equiv. to int B.1.	Equiv. to int B.2.	Analog. to int B.3.
	body:	mov(6), mov(7), mov(8): 21B. (Use of AX reg.)	mov(6), mov(6), mov(7), mov(4): 23B. (Use of CX.)	mov(6), mov(7), mov(3): 16B. (Use of DX reg.)	mov(6), mov(8), mov(7): 21B. (Use of CX reg.)	mov(6), mov(6), mov(4), mov(7): 23B. (Use of CX.)	mov(6), mov(3), mov(7): 16B. (Use of DX reg.)
int	ctrl:	mov(10), jmp(2), mov(6), add(3), mov(6), mov(6), cmp(6), jae(2); +jmp(2): 43B.	Equiv. to A.1, but with permuted reg.: EAX→ECX→EDX (↑ ← ← ↓)	Simil. to A.1, but with permuted reg.: ECX→EAX→EDX (↑ ← ← ↓)	Equiv. to A.2.	Equiv. to A.1.	Equiv. to A.3.
	body:	mov(6), mov(6), mov(7): 19B.	mov(6), mov(6), mov(6), mov(3): 21B.	mov(6), mov(6), mov(2): 14B.	mov(6), mov(7), mov(6): 19B.	mov(6), add(6), mov(3), mov(6): 21B.	mov(6), mov(2), mov(6); 14B.
ll-int	ctrl:	Equiv. to int A.1.	Equiv. to int A.2.	Analog. to int A.3.	Equiv. to int A.1.	Equiv. to int A.2.	Equiv. to int B.3.
	body:	mov(6), mov(6), mov(6), mov(7), mov(7): 32B.	mov(6), mov(6), mov(6), mov(6), mov(3), mov(4): 31B.	mov(6), mov(6), mov(6), mov(2), mov(3): 23B.	mov(6), mov(7), mov(7), mov(6), mov(6): 32B.	mov(6), mov(6), mov(3), mov(4), mov(6), mov(6): 31B.	mov(6), mov(2), mov(3), mov(6), mov(6): 23B.
float	ctrl:	Equiv. to int A.1.	Equiv. to int A.1.	Equiv. to int A.3.	Equiv. to int A.1.	Equiv. to int A.1.	Equiv. to int B.3.
	body:	mov(6), movss(8), movss(9): 23B.	mov(6), mov(6), movss(8), movss(5): 25B.	mov(6), movss(8), movss(4): 18B.	mov(6), movss(9), movss(8): 23B.	mov(6), mov(6), movss(5), movss(8): 25B.	mov(6), movss(4), movss(8): 18B.
double	ctrl:	Equiv. to int A.1.	Equiv. to int A.2.	Sim./Analg. to int A.1.	Equiv. to int A.1.	Equiv. to int B.1.	Analog. to int A.3.
	body:	mov(6), movsd(8), movsd(9): 23B.	mov(6), mov(6), movsd(8), movsd(5): 25B.	mov(6), movsd(8), movsd(4): 18B.	mov(6), movsd(9), movsd(8): 23B.	mov(6), mov(6), movsd(5), movsd(8): 25B.	mov(6), movsd(4), movsd(8): 18B.

Table 1. The x86 machine instructions of our six standard benchmarks and their lengths for the six standard array types. The benchmarks' `for`-loops are divided into the control part (the loop head and the unconditional jump at the end of the loop body) and the loop body contents. For the two parts, the instructions are represented as: `opMnem`(*l*), where `opMnem` = the operation mnemonic, *l* = the instruction length in bytes. Equiv. = equivalent up to the address specs., Analog./Sim. = as equiv. but adapted to the data type/with a different addressing mode.

this `uI` index is different from the one in A.1, because all indices are declared locally to their `for`-loops. Here, `A(uI) = EBP – 7F8h` [as stated in §3.2.1, we omitted the A.2 `for`-loop head compilation because it is equivalent to that in A.1, except for the circular replacement of the registers (cf. Table 1)].

The next `mov` places the starting `pI0 = A(iX)` pointer value into `ECX` [`A(pI0) = EBP – 76Ch`]. The third `mov` stores `iRVal` into `EDX`. Now everything is prepared for the action of the fourth `mov`, which will form the address of the `uI`-th element by using the index addressing in a very similar way as in eq. 4, but with all the operands in registers:

```
A(iX[uI]) = pI0 + 4*uI = 00C826A8h + 4*EAX.          (5)
```

To recapitulate this shortly: despite one more instruction in this case, the last instruction—which performs the final assignment of the value to the `uI`-th array element—is relatively short (3B) and probably very effective.

The three benchmarks of type B remain, in which the `uI`-th array element value is assigned to the *l*-value provided by the variable `iLVal`. Again, the assignment statements in B.1 and B.3 bodies consist of three (3), and in B.2 of four (4) `mov` instructions. Thus, we start with the former two benchmarks.

In the B.1 loop body, the three `mov` instructions are the same type as those in A.1 and have the same lengths, but the last two are in the reversed order and have different source and destination operands. The first `mov` places the value of `uI` (now with `A(uI) = EBP – 7F0h`) into `EAX`; the second stores the value of `iX[uI]` into `ECX`; the last one moves the value from `ECX` into the `iLVAl` variable; `A(iLVal) = EBP – 784h`.

The analogous situation is in B.3. Again, the instructions there are of the same type as those in A.3 but permuted. First, the pointer `pI` value is moved to `ECX` [`A(pI) = EBP – 810h`]. The second `mov` retrieves the contents from the address in `ECX`, i.e., the value of `*pI`, and stores it into `EDX`, making this crucial fetch of the array element very effective. Finally, the third `mov` places the `EDX` value into `iLVal`.

Similarly, in the B.2 loop body, the first two instructions do the same thing as those in A.2 but with `EDX` and `EAX` as destinations. The third, the shortest one (3B), stores the `*(pI0 + uI)` value into `ECX`. The fourth instruction moves this value from `ECX` to `iLVal`. Judging solely by the number of instructions and their lengths, this solution is longer, but—as suggested previously—its time efficiency may also depend on other factors.

3.2.3. Compilation of the benchmarks with arrays of other types

A detailed inspection of the machine instructions for all other array (data) types would be very voluminous. For them, we will only briefly comment on the data summarized in Table 1, in which the above-elaborated `int`-type array benchmarks serve as the referent ones.

Regarding the implementation of the control part of the `for`-loops, the reader can note that it is the same for all 36 benchmark variations. Because they have indices (counters) and incrementing pointers within the range of the `int` type, they are all compiled to the same instructions, with the same lengths, possibly differing from each other only as follows:

Equiv. — stands for the cases where the implementation is *equivalent* in every way except in the concrete choice of the general-purpose registers (GPRS) and the address offsets.

Analog. — stands for the *analogous* cases, which differ from the equivalent only in that some instruction operands are adapted to the concrete data type, e.g., by changing the address increment according to the data type length: `incr = sizeof(type)`.

Sim. (similar) — stands for the case with the same types of instructions but with different addressing modes.

The inspection of Table 1 confirms that all `for`-loop implementations are — if not equivalent — either analogous or similar in the above sense. The biggest difference, attributed as "similar," concretely means that only the first instruction in the `for`-loop head is different (see the comparison of the loop heads for A.1 and A.3 benchmarks in §3.2.1).

In the implementation of the single C/C++ statement within the loop body, there is much greater diversity, as we have already shown for the `int` arrays. Similarly, the observation that the type-2 benchmarks (A.2 and B.2) have one instruction more and the type-3 benchmarks (A.3 and B.3) have the shortest total length of their instructions also holds for the other array data types. Besides that general remark, we note that with the same number of instructions for the `short`- and `int`-type arrays, the former has a net length that is systematically 2B longer. These instructions use the 16-bit `X`-type registers (`AX`, `CX`, `DX`). The situation levels up for the `char` type. These benchmarks have instructions of the same or faintly shorter length than those for the `int` type, and they use the lower byte of the `X` registers (`AL`, `CL`).

The benchmarks with the `llint` arrays systematically have two (2) instructions more than the corresponding ones of `int` type and are, because of that, considerably longer. In the x86 mode, the VS-CV 2019 compiler produces the x86, 32-bit machine code without using the available 64-bit (`R`) registers, so moving this type of data requires the engagement of two 32-bit registers instead of one, and the use of the standard `dword mov` instructions (`double word` = 32-bit word). In our concrete example, the value of the `llint` variable `llRVal` is stored into the two registers: `ECX:EDX`, of which the left (right) holds the more (less) significant half of the 64-bit value.

The remaining two types of arrays are of the floating-point type. For accessing the array elements of the `float` (`double`) type, the compiler uses `movss` (`movsd`) instructions to move the 32(64)-bit contents to the lowest (lower) portion of the 128-bit `xxm` registers (in our case, it was the `xmm0` register (Intel 2022-2). From their short description in Table 1, we can see that the numbers of these instructions are equal to the numbers of instructions in the corresponding benchmarks for the `int` arrays and that their length is equal for both the 32-bit `float` and 64-bit `double float` type.

3.2.4. A Glimpse to the x64 Compilation

In VS-CV 2019, the default compilation option is for the x86 set of machine instructions, as a still de-facto standard for many sorts of applications. In addition, there is also the x64 compilation, which translates the C++ source code into the Intel's x64 set of 64-bit instructions. To explore this option thoroughly, we would have to inspect the additional instructions in the i64 set, which have access to the quadword (64-bit), R-registers. It would also require the analysis of the compiled machine code similar to the above one. We will leave this out of this work, and here just briefly describe the crucial differences between the 64-bit version and the 32-bit version of the assembly code.

The structure of the `for`-loops in our benchmarks remained the same: there are eight (8) instructions in the loop heads and one (1) unconditional jump at the end of the loop bodies. For our referent, `int` type, and the A.1 benchmark, the 3B-instruction `add eax,1` is now replaced with the 2B-instruction `inc eax,1` (making us wonder why it was not used in the x86 code, too). Furthermore, the loop indices are stored locally with the displacement counted now from the `RSP` register (the 64-bit stack pointer), instead of from the `EBP` register. Overall, the x64 version of our `for`-loops is realized similarly to before. They have the same number of instructions (9), and their total length is 42B, 1B less than in our standard, x86 version loop. This structure holds for all `for`-loops with indices. The `for`-loops with the incrementing pointers are organized somewhat differently. They have, in total, ten (10) instructions and a much longer length of 55B.

As for the assignment statement in the loop body, it is compiled into four (4) instructions: `mov`(7), `lea`(7), `mov`(4), `mov`(3), totaling 23B, and having one (1) instruction and 4B more than the x86 version. Here, all instructions but the first one work with the `R`-type of registers for preparing the operand addresses. For the benchmarks with the `int`-type arrays, their *r*-values and array elements are manipulated according to their type, i.e., as the 32-bit memory and register values. The same instructions, but ordered differently, are in the loop bodies of the B-type benchmarks.

The situation is very similar for the shorter (integer) types, for which the loop bodies have one instruction more than in the x86 version. Likewise, for the `float` type, the loop body of A.1 benchmark is realized by four (4) instructions: `mov`(7), `lea`(7), `movss`(6), `movss`(5), totaling 25B, and having one (1) instruction and 2B more than the x86 version. Generally, it is clear that for the types equal to or shorter than 32-bits, the use of x64 architecture does not improve the structure of the compiled machine code.

The benefits of the 64-bit architecture should—if anywhere—become obvious for the 64-bit data types. Indeed, the loop body of the `llint` A.1 benchmark is realized by (only) four (4) instructions: `mov`(7), `lea`(7), `mov`(8), `mov`(4), totaling 28B, which is one (1) instruction and 4B less than in the x86 machine code. However, for the type `double` of A.1 benchmark, the instructions in the loop body are: `mov`(7), `lea`(7), `movsd`(9), `movsd`(5), totaling 28B, which is one (1) instruction and 5B more than in the x86 version. The rest of this analysis might be presented elsewhere.

3.2.5. Possible Benchmark Pitfalls

The benchmarks written for this research were not intended to perform some standard operations[4] but to be as simple as possible and thus eliminate all unnecessary consumption of the processor time unrelated to the purpose of this testing. In such a reduction, one must pay attention to the generality of the written program code. Otherwise, it may easily happen that the obtained benchmarks favor some of the testing options before others. We will illustrate this in the following example.

In the early versions of our A-type benchmarks, instead of the `iRVal` variable, we have used a numerical constant as a source of the value to be stored in the array element. In a slightly simplified version, shown here in Listing 3, the constant is an integer defined by `intLrg`, as can be seen in the C/C++ assignment statement above the corresponding assembly language code.[5] Unlike the code in Listing 2, the `jmp` instruction on the bottom of the `for`-loop body is not shown because the `for`-loop context is omitted here. In all three cases, there is one `mov` instruction less than in our standard benchmarks (in Listing 2). The numbers and the lengths of the instructions (in parenthesis) are as follows:

A.1 – 2 instructions: `mov`(6), `mov`(11), in total 17B;

A.2 – 3 instructions: `mov`(6), `mov`(6), `mov`(7), in total 19B;

A.3 – 2 instructions: `mov`(6), `mov`(6), in total 12B.

```
 // WARNING! Not a general case because of assigning intLrg, which is de-
// fined as a numerical constant (see Listing 1) and is addressed literally.
// A.1 Storing into:
iX[uI] = intLrg;
00426078 mov eax,dword ptr [ebp-0A20h] ; EAX ← ML(EBP – 0A20h) = uI.
0042607E mov word ptr iX (03D6E850h)[eax*4], 153158Eh
                      ; iX[uI] = ML(iX + 4*uI) ← 153158Eh = 22 222 222d.
 //    ...           ...
// A.2 Storing into:
*(pI0 + uI) = intLrg;
004262E2 mov eax,dword ptr [ebp-0A7Ch]
004262E8 mov ecx,dword ptr [ebp-0A64h]
004262EE mov dword ptr [ecx+eax*4],153158Eh
 //    ...           ...
// A.3 Storing into:
*pI = intLrg;
00426571 mov eax,dword ptr [ebp-0AD8h] ; EAX ← ML(EBP – 0AD8h) = pI.
00426577 mov dword ptr [eax],153158Eh  ; *pI = ML(EAX) ← 153158Eh = 22 222 222d.
//     ...           ...
```

Listing 3. Intel x86 assembly language code of the B-type array element assignments with the less general *r*-value, in the form of numerical constant.

[4] Though the A-type benchmarks do perform the initialization of the array elements to the same value.

[5] In the test programs, we use two such constants: one for the `for`-loops pre-runs (§4.2.2), and another for their time-measuring runs.

The first `mov` instruction in A.1 and A.2 places the value of index `uI` into `EAX`. In A.3, the first `mov` places the value of the pointer `pI` into that register. The second `mov` in A.1 (A.3) finishes the job, by storing the value of the numerical constant (immediately addressed) `intLrg = 153158Eh = 22 222 222d` into the place of the `uI`-th (next) array element. In A.1, it is done by a more complex, 11B-`mov` instruction because of the longer form of the index addressing (4B are needed for the address of `iX`, and 4B for the stored literal). In A.3, the same action is achieved by a 6B-`mov` instruction, thanks to the fact that the fully formed array-element address was already in `EAX`. In A.2, the second of the three `mov` instructions must store the value of `pI0` into `ECX`, and after that, the third `mov` instruction stores the literal at the array element address formed by the index addressing. One might expect that, in this case, the A-type benchmarks would perform faster. We will comment on this in the next section (§4.3).

Another example of a pitfall in the benchmark code was already commented in the last part of §3.2.1. In it, a variable declared as `const` (`cuiN`) has enabled immediate addressing mode and thus — comparing it to the use of a general variable — decreased the number of instructions for one. Besides that, the variables should always be (re)declared in local blocks to assure similar addressing modes in all benchmarks.

In conclusion of this section, when writing benchmarks, one should pay great attention to their generality and follow all the rules of good programming practices. Additionally, before applying the newly created benchmarks, it is important to check their assembly language form.

4. Execution Time Measurements

This section will first explain the methods used for our BET (benchmark execution time) measurements and the prerequisites needed for consistent and precise results. Then, it presents those results and comments on them.

4.1. Time-Measurement Methods

There are several ways to measure the elapsed time of certain program parts in C++. In our preliminary and motivational measurements (mentioned in sec. 1), we have used the MS Windows `SYSTEMTIME struct`, available in VS IDE after including the `windows.h` header file (details in [9]). The way to apply it for the measurements of the benchmark execution times is shown in Listing 4. Because its smallest time division is a millisecond (ms), this method is useful when the measured times approach the order of one second (s). In that ad-hoc measurement, we explored the `short` and `int` arrays with `700 0000h = 117 440 512d` elements, i.e., five times larger than now. Thus, the execution times on the older and slower computers did approach `1s`, so such accuracy was acceptable. Recently, we repeated similar measurements on the arrays with roughly five times fewer elements because, in the same program, we were investigating six different types of arrays, most of them with much larger element types. The execution times of those measurements became too short for this

method. Sometimes, especially when the operation system (OS) was loaded more heavily, the function `GetSystemTime` failed to return a proper value and the calculated difference became negative.

More precise time measurements should get more accurate "time stamps" from the hardware timers, which rely directly on the processor's time cycles [10]. Such are the methods (ii) and (iii) in Listing 4. In (ii), the `HRTimer` class uses the member function `QueryPerformanceCounter` to do the job. We used this method for the time measurements in [11], [12]. By assuming that its accuracy should approach the execution time of the order of (several) instruction cycle(s), we appraised its precision to be around 10ns. However, the comparisons with the other methods used here put a question mark on this optimistic "guesstimate" (see below).

```
// Declarations and definitions as in Listing 1.
//    ...   ...
// Variables for the elapsed times in ms.
float fDltTmSYS, fDltTmHRT, fDltTmHRC;

// i) Time measurement by SYSTEMTIME (SYS)
#include "windows.h"

// _SYSTEMTIME structures and ptrs. to struc.:
_SYSTEMTIME sT1, sT2, *pST1 = &sT1, *pST2 = &sT2;
GetSystemTime(pST1);      // Stopwatch on.
for (uint uI = 0; uI < cN; uI++)
     iX[uI] = intLrg;
GetSystemTime(pST2);    // Stopwatch off.
fDltTmSYS = (float)(pST2->wSecond - pST1->wSecond)*1000 +
            (float)(pST2->wMilliseconds - pST1>wMilliseconds);

// ii) Time measurement by CHRTimer class (HRT)
#include "HRTimer.h" //High-Resolution Time.
CHRTimer hrTimer;   // CHRTimer object.
hrTimer.StartTimer(); // Stopwatch on.
for (uint uI = 0; uI < cN; uI++)
     iX[uI] = intLrg;
fDltTmHRT = (float)hrTimer.StopTimer()*1.e3f; // Stopwatch off.

// iii) Time measurement by high_resolution_clock class (HRC)
#include <chrono>  // XYZ
using namespace std;
using namespace chrono;
duration<float, milli> durTDlt;
auto t1 = high_resolution_clock::now();
auto t2 = high_resolution_clock::now();
t1 = high_resolution_clock::now();// Stopwatch on.
for (uint uI = 0; uI < cN; uI++)
     iX[uI] = intLrg;
t2 = high_resolution_clock::now();// Stopwatch off.
fDltTmHRC = (durTDlt = t2 - t1).count();
```

Listing 4. Examples of the three time-measurement methods in C++, applied to the A.1 benchmark for `int` array: i) `_SYSTEMTIME` structures, ii) `CHRTimer` class, and iii) the `high_resolution_clock` class.

The third time-measurement method tested in this work used the `high_resolution_clock` class from the C++ `std::chrono` library [13]. To get the elapsed time, the programmer uses the template class member `duration`, with which she defines a variable to accept the time difference (`t2 – t1`) of the recorded time points in desired time units. Since our results are of the order of tens of milliseconds, we have chosen that unit. There is a warning about the possibly inconsistent and inferior implementation of this class in different libraries and compilers, as mentioned in [13]. However, we have kept it anyway without investigating how it is realized in our case.

In a separate, short test C++ program, we have compared the results of the three time-measuring methods on the A-type benchmarks 1, 2, and 3 for the `short` and `int` types. On the "outmost side" of the benchmarks, we placed the "measuring points" for the `SYSTEMTIME`, in the middle, the points for the `HRTimer` class, and closest to the benchmarks, we put the `high_resolution_clock` class.

The first method, using `SYSTEMTIME`, gives only roughly correct individual results and nearly correct averages, which is not bad regarding its above-stated deficiencies. The second and third methods give very consistent results of satisfying accuracy for our measurements. However, it is quite surprising that they produce results with differences of the order of $10\mu s$, which is 10^3 times more than our estimate above and probably much more than it should be. The possible culprit is the dubious implementation of the method (iii). Furthermore, the absolute time values can especially have much greater errors than the above estimate ($10ns$) because of the overall imprecision of the clock crystal's frequency [10]. In these measurements, the relative time accuracy is the important one. It requires only the steadiness of the processor clock frequency and not its absolute precision. Anyhow, because the relative accuracy of method (iii) is more than satisfactory for our needs, we were using it as the method of choice for the time measurements in this investigation.

4.2. Time-Measurement Setup

To achieve consistent and accurate time measurements, a programmer should comply with some program- and system-wise conditions. Our approach tends to measure the "average best results," i.e., the optimal benchmark execution times for the given computer.

4.2.1. Benchmark Execution Order

In the test programs, we have placed the benchmarks for each of the six array types in an order different from the one shown in Listing 1. In that order, there are also the *pre-runs* (explained in the next section), as follows:

Repeat for the array element access type $k = 1, 2, 3$:
A. Pre-run A.k, then **A.k with BET** measurements;
B. Pre-run B.k, then **B.k with BET** measurements;
Repeat end.

In this way, we have simulated a somewhat more realistic situation in which the A- and B-type benchmarks exchange first, and then, there is a change in the type of array element access, according to the enumeration given in §2.2.

4.2.2. Pre-runs

The measured BETs can significantly depend on the momentary state of the memory system of today's computers with multitasking OSs. If the multilevel caches are already optimally filled with the data used in the benchmarks, the execution times will also be optimal, i.e., close to the shortest possible. If this condition is not fulfilled, the measured times can increase severely, making the results prone to erratic changes. The first way to deal with this problem is to exclude the "statistical outliers" from the measurements. In practice, it usually boils down to excluding one or a few highest and lowest values from the measurement set. Obviously, in our case, only the worst results are to be excluded.

Here, we have used another method — the benchmark *pre-runs*. They execute the benchmark fully or partly before measuring its execution time in the same or very similar way (cf. [11], [12]). Here, the pre-runs executed the benchmarks by assigning the array elements different number values. Obviously, because the longest BETs will mostly happen at the execution of the first benchmarks, these two approaches produce similar results.

In our C++ programs, the user can control the pre-runs by switching them on and off for the A- and B-types of benchmarks separately, which come one after the other for each array (data) type. The user can also change the number of iterations, i.e., she can specify the number of elements the pre-runs will access.

Without the pre-runs, we have observed that the A.1 benchmark — being the first of the six benchmarks in the row (§4.2.1) — in the first of the (10) measurements for each array type executes much longer, compared to the standard deviation of the rest of the measurements in the set. We have investigated this influence by running the release version of our Program 1 (directly from VS-CV 2019), first with both pre-runs on and then with both or just one of the pre-runs off. With both pre-runs off (F, F) and when only the B pre-run was on (F, T), the BETs of the first A.1 measurement in a series prolonged for approx.:

(F, F): `char`: 27%, `short`: 39%, `int`: 56%, `llint`: 179%, `float`: 62%, `double`: 116%;

(F, T): `char`: 13%, `short`: 29%, `int`: 50%, `llint`: 168%, `float`: 58%, `double`: 120%.

When A pre-run was on and B off, there were no delays. On the contrary, all the first measurements in the series were shortened by about 1.5%!?

If the pre-run executes only a part of the iterations, which the user can control by decreasing the number of iterations in them — and thus access only a part of our huge arrays — the slow-down effect is still present, more so the lesser the chosen number of iterations.

This brief analysis shows that the BETs without the pre-runs delay more for the arrays with the longer data types because they occupy larger memory space.

Moreover, the B-only pre-run (F, T) cannot cover for the delays of the upcoming A benchmarks of longer types. In addition, it turns out that only the pre-run before the first A.1 benchmark is necessary, while the rest of them are superfluous. However, in the present versions of our test programs, we have simply left all of them in the order specified in §4.2.1.

As a final remark on the pre-runs, although introducing them might seem like an artificial intervention into the programming code, without them, we would obtain unstable and less accurate results.

4.2.3. Computer and OS specifications

To complete the BET measurements, one must record the computer hardware and OS specifications. In this paper, we mainly present the measurements that are made on one computer with the following specifications:

Processor:	Intel® Core™ i7-8700K CPU @ 3.70 GHz, with 6 cores.
Installed RAM:	32.0GB
System type:	64-bit OS, x64-based processor.
Op. System:	Windows 10 Pro Education (build 19044.1766).

4.2.4. Measurement Preconditions

Our aim is to measure the execution times of our benchmarks only, as "pure" as possible. To achieve that, we want to eliminate all side effects that could interfere with the computer hardware and OS performance. For that matter, there are a few things to consider. The first is which program version to use — debug or release, and how to run it — from the IDE or by starting the executable code file.

The measurements showed that the influence of both of these options in our VS-CV 2019 is rather minor. By directly running the exe-files of the release and debug versions (reminding the reader that the compiling optimization was turned off) we have noticed no significant changes in the measured BETs. Typically, the time differences for most of the measurements were less than the standard deviation of the measurements. The debug versions for some of the six benchmarks for the six array data types were sometimes slower and sometimes even faster, resulting in the average delta-percentages of the measured corresponding mean values around 0.5% to 1.0%. However, the debug version is prone to delays for the arrays of certain data types, noticeably for the arrays of `char`, `short`, and not so often of the `double` types. These outbursts of delays have ranged chiefly from 10% to 13%, but sometimes escalating to even 20%, then raising the average delay (of all 36 mean values) for up to 1.5%.

The comparison of running the programs directly from the exe-files vs. starting them from the VS-VC19 IDE and leaving the IDE on during the test program executions gives similar results. The former approach does provide a bit shorter BETs

(0.3% to 1.0%), but its true advantage is better stability. Namely, the latter is prone to sporadic upsurges of BETs up to 10%.[6]

The conclusion is that these differences would be irrelevant were it not for rare but possible erratic delays in the inferior options, that is, in running the debug version and running either version from the VSCV-2019 IDE. Thus, as a precaution to achieve better precision, our choice for the final measurements was as follows:

- to run the release versions (without compiler optimizations);
- to start the exe-files with all other applications turned off.

Furthermore, for today's standard multitasking OSs, with the otherwise "standard" execution environment, we take the following precautionary procedure:

- turn off all user applications;
- disable Ethernet and WLAN;
- check the task manager for the possibly demanding background processes and try to turn them off;
- run the test program with benchmarks.

4.3. Results of the BET Measurements

With all the preconditions being fulfilled, we run our benchmark test programs. Once the user enters the required input parameters, the remaining free run of one program on our computer lasts approximately 25s. It performs a benchmark pre-run and a BET measurement, normally both through the arrays with 20×10^6 elements, it does that for six (6) different benchmarks, repeats the whole action ten (10) times to get the averages, and repeats the same thing for six (6) different data types of array. This totals to the $6 \times 10 \times 6 = 360$ pre-runs and the same number of measured iterations, resulting in 14.4×10^9 assignment operations on the array elements. This number is halved if both the A- and B-type benchmark pre-runs are turned off. Each program lists 6×10 BETs for the six array (data) types, with their averages and standard deviations.

4.3.1. BETs of Our Standard Benchmarks

The typical results of the BET measurements for our six standard benchmarks (from Listing 1), extracted from a single run of (test) Program 1, are presented in Table 2.

For the A- and B-type assignments and the three different methods of accessing the array elements, six average $\overline{\Delta t}_{BET}$ times and (corrected) standard deviations are calculated for the set of 10 non-successive BET measurements for each array type.[7] In the extra columns for the A/B.2 and A/B.3 benchmarks, which access the array

[6] Although we did not practice it, the VS-VC19 IDE can be turned off after launching the desired (debug or release) version of the test program, and before proceeding with the program, i.e., before entering the few required input values.

[7] With 20×10^6 iterations performed in ≈ 30ms, one loop iteration takes 1.5ns, or 5.55 clock cycles (CC) of our processor, executing $12 - 13$ machine instructions. This means that the working-core's pipeline has an average throughput of 2.25 instructions per CC.

elements by using pointers, we can track the relative difference of their BETs comparing to the A/B.1 BETs.

The measured times and relations between them are similar for the A- and B-type benchmarks marked with the same numbers. A/B.1 is only slightly slower than A/B.3, meaning that the successive access to the array elements by incrementing the pointers did not considerably shorten the execution, although the lengths of their loop bodies are shorter by roughly a quarter (cf. Table 1). The BETs of A/B.3 benchmarks are shorter than those of A/B.1, but only for the amounts comparable to the measurement standard deviations.

Quite a surprise is that the execution is fastest for the A/B.2 benchmarks (except for the `llint`), although the number of the instructions in their loop body is greater by one-third than those in A/B.1 and 3. In addition, their total length is roughly 10% (30%) longer than in A.3 (B.3). This example clearly shows how the number and the lengths of instructions do not directly dictate their execution times. The decrease is quite significant for the `short` and `int` integers and for both floating-point types, ranging from roughly 30% to more than 40%. The big exception for the `llint` type (+57%) is caused by the simultaneous lack of comparable improvement for its A/B.2 benchmarks and the sudden 38% (34%) decrease of its BETs of A.1 & 3 (B.1 &3), resulting in the significant raise of the relative values for A/B.2.

For the `char` type, the BETs of A/B.2 benchmarks are just a little faster than for A/B.1 & 3. Furthermore, they are $\approx 45\%$ longer than for all other, longer types, except the `llint`, which is quite surprising. Namely, one would expect the BETs for this data type to be similar to the (other) integers but not slower.

From the above deliberation and by looking at the absolute values of those times, we come to the following conclusion.

For our standard benchmarks (Listing 1), the fastest way to access the array elements of type:

- `char`, `short`, `int`, `float`, and `double` is by using **pointer arithmetic** [`*(p0 + i)`], as in A/B.2, though for `char` it is only slightly better;
- `llint` type is by using
  - **pointer incrementing** [`*(pI)`], as in A/B.3, or just a little bit slower by using
  - **indices** (`Arr[i]`), as in A/B.1.

4.3.2. BETs of the Modified Benchmarks

In our Program 2, we have modified the benchmarks from Listing 1 by replacing the assignment (=) operator with the compound addition-assignment (+=) operator. For those benchmarks, Table 3 contains the average values of their typical BET measurement set.

The times of the A- and B-type benchmarks are still close to each other, but not as much and as consistently as for the set of our standard benchmarks. In addition, there are a few greater discrepancies, as in the following cases: for `int` — cf. A.1 vs. B.1 and A.3 vs. B.3; for `char` — cf. A.2 vs. B.2. A big difference is also that now A.2

20×10^6 *elements*	*A: Arr. el. = r-Value,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms					*B: l-Value = Arr. el.,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms				
Array Type	A.1 `Arr[i]`	A.2 `*(p0+i)`	Δrel to A.1	A.3 `*(++p)`	Δrel *to* A.1	B.1 `Arr[i]`	B.2 `*(p0+i)`	Δrel *to* B.1	B.3 `*(++p)`	Δrel *to* B.1
`char`	35.50 ±0.45	33.68 ±0.43	−5.1%	34.90 ±0.47	−1.7%	36.96 ±0.32	35.36 ±0.37	−4.3%	35.95 ±0.29	−2.7%
`short`	35.45 ±0.45	23.89 ±0.43	−32.6%	35.05 ±0.48	−1.1%	36.78 ±0.36	22.63 ±0.26	−38.5%	35.90 ±0.26	−2.4%
`int`	38.45 ±0.50	22.65 ±0.34	−41.1%	37.81 ±0.34	−1.7%	36.78 ±0.53	23.28 ±0.23	−36.7%	35.96 ±0.32	−2.2%
`llint`	23.47 ±0.38	36.92 ±0.41	+57.3%	23.34 ±0.22	−0.6%	24.16 ±0.22	34.42 ±0.47	+42.4%	23.84 ±0.48	−1.4%
`float`	39.02 ±0.48	22.42 ±0.28	−42.5%	37.85 ±0.32	−3.0%	36.88 ±0.37	23.42 ±0.31	−36.5%	36.09 ±0.40	−2.1%
`double`	38.93 ±0.28	23.04 ±0.15	−40.8%	38.28 ±0.52	−1.7%	37.35 ±0.20	25.46 ±0.36	−31.8%	36.20 ±0.08	−3.1%

Table 2. Average $\overline{\Delta t}_{BET}$ (Benchmark Execution Times), for the benchmarks from Listing 1 (Program 1). The benchmarks run in the order stated in §4.2.1, in the release version of the program, on the computer specified in 4.2.3. $\overline{\Delta t}_{BET}$ and the corresponding standard deviations are calculated for 10 (nonconsecutive) measurements and expressed in ms (milliseconds). For the benchmarks A/B.2 and A/B.3, the second column gives the relative Δrel difference of their averages from the Δt_{BET} of the benchmark A/B.1. In the color print, those are marked in green (red) if $\leq -10\%$ ($\geq +10\%$).

20×10^6 *elements*	*A: Arr. el. = r-Value,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms					*B: l-Value = Arr. el.,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms				
Array Type	A.1 `Arr[i]`	A.2 `*(p0+i)`	Δrel to A.1	A.3 `*(++p)`	Δrel *to* A.1	B.1 `Arr[i]`	B.2 `*(p0+i)`	Δrel *to* B.1	B.3 `*(++p)`	Δrel *to* B.1
`char`	28.02 ±0.58	37.24 ±0.32	+32.9%	27.64 ±0.41	−1.4%	27.74 ±0.29	29.58 ±0.39	+6.7%	27.17 ±0.32	−2.1%
`short`	28.12 ±0.36	36.21 ±0.63	+28.8%	27.26 ±0.27	−3.1%	27.92 ±0.24	34.19 ±0.30	+22.5%	27.22 ±0.29	−2.5%
`int`	38.62 ±0.34	40.43 ±0.40	+4.7%	38.67 ±0.39	+0.1%	24.78 ±0.29	35.15 ±0.15	+41.8%	24.17 ±0.10	−2.5%
`llint`	35.21 ±0.38	36.01 ±0.40	+2.3%	35.02 ±0.17	−0.5%	30.28 ±0.29	35.00 ±0.35	+15.6%	29.44 ±0.23	−2.8%
`float`	39.07 ±0.66	40.95 ±0.61	+4.8%	38.49 ±0.65	−1.5%	40.53 ±0.35	39.55 ±0.74	−2.4%	39.36 ±0.37	−2.9%
`double`	39.61 ±0.82	41.55 ±0.86	+4.9%	39.14 ±0.80	−1.2%	40.83 ±0.79	40.16 ±1.08	−1.7%	39.61 ±0.58	−3.0%

Table 3. A set of the BET measurements for slightly altered benchmarks (Program 2): in the benchmarks from Listing 1, the assignment operator (`=`) is replaced by the compound addition-assignment operator (`+=`).

20×10^6 *elements*	*A: Arr. el. = r-Value,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms					*B: l-Value = Arr. el.,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms				
Array Type	A.1 `Arr[i]`	A.2 `*(p0+i)`	Δ*rel* to A.1	A.3 `*(++p)`	Δ*rel* *to* A.1	B.1 `Arr[i]`	B.2 `*(p0+i)`	Δ*rel* *to* B.1	B.3 `*(++p)`	Δ*rel* *to* B.1
`char`	22.29 ±0.35	34.71 ±0.56	+55.7%	21.94 ±0.24	−1.5%	36.76 ±0.36	35.45 ±0.31	−3.6%	36.04 ±0.31	−2.0%
`short`	22.40 ±0.22	39.48 ±0.42	+76.2%	22.91 ±0.33	−2.2%	36.60 ±0.29	22.35 ±0.33	−38.9%	36.16 ±0.57	−1.2%
`int`	22.47 ±0.20	38.26 ±0.43	+70.3%	22.05 ±0.29	−1.9%	36.63 ±0.33	23.50 ±0.39	−35.9%	36.02 ±0.43	−1.7%
`llint`	22.42 ±0.27	38.72 ±0.33	+72.2%	23.09 ±0.39	+3.0%	23.89 ±0.26	34.80 ±0.35	+45.7%	23.60 ±0.24	−1.2%
`float`	38.41 ±0.37	22.50 ±0.45	−41.4%	37.95 ±0.57	−1.2%	36.64 ±0.42	23.60 ±0.26	−35.6%	36.08 ±0.36	−1.5%
`double`	38.58 ±0.37	24.19 ±0.29	−37.3%	38.05 ±0.31	−1.4%	37.02 ±0.25	24.85 ±0.14	−32.9%	36.52 ±0.35	−1.3%

Table 4. A set of BETs for the benchmarks from our Program 3. These benchmarks differ from those in Listing 1 in that the A-type benchmarks store the numeric constants into the array elements (§3.2.4, Listing 3).

20×10^6 *elements*	*A: Arr. el. = r-Value,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms					*B: l-Value = Arr. el.,* $(\overline{\Delta t}_{BET} \pm s_{dev})$/ms				
Array Type	A.1 `Arr[i]`	A.2 `*(p0+i)`	Δ*rel* to A.1	A.3 `*(++p)`	Δ*rel* *to* A.1	B.1 `Arr[i]`	B.2 `*(p0+i)`	Δ*rel* *to* B.1	B.3 `*(++p)`	Δ*rel* *to* B.1
`char`	33.84 ±0.39	23.82 ±0.30	−29.6%	22.01 ±0.22	−35.0%	31.92 ±0.25	34.61 ±0.31	+8.4%	36.16 ±0.64	+13.3%
`short`	33.69 ±0.49	32.80 ±0.32	−2.6%	21.75 ±0.10	−35.4%	31.90 ±0.35	33.32 ±0.29	+4.4%	35.95 ±0.32	+12.7%
`int`	33.75 ±0.49	33.09 ±0.43	−2.0%	23.85 ±0.35	−29.3%	32.13 ±0.30	33.43 ±0.31	+4.0%	35.96 ±0.35	+11.9%
`llint`	33.01 ±0.47	31.70 ±0.46	−4.0%	23.26 ±0.28	−29.5%	30.65 ±0.42	30.89 ±0.32	+0.8%	23.70 ±0.30	−22.7%
`float`	33.31 ±0.47	37.66 ±0.27	+13.1%	37.93 ±0.31	+13.9%	31.83 ±0.29	33.40 ±0.29	+4.9%	36.16 ±0.39	+13.6%
`double`	32.54 ±0.46	35.73 ±0.48	+9.8%	38.08 ±0.36	+17.0%	32.46 ±0.48	34,16 ±0.40	+5.2%	36.63 ±0.40	+12.8%

Table 5. A set of BETs for our old benchmarks, now in Program 3-old. They differ from our standard benchmarks as those in Program 3, plus there is a numerical constant in the conditions of the `for`-loops with indices (A/B.1 & 2).

and B.2 are not the best access methods as they were for the standard benchmarks (except for `llint`). On the contrary, those are now either the worst or close to that, with `float`-B.2 being the only exception. As above, we summarize this as follows:

For our modified benchmarks (‘=’ → ‘+=’), the fastest way to access the array elements of

- all types except `int` is by **pointer incrementing** [`*(pI)`], as in A/B.3 (`int`-A.3 losing over `int`-A.1 for a mere 0.1%), or just a little bit worse by using
  - **indices** (`Arr[i]`), as in A/B.1.

An interesting observation is that these benchmarks — which not only assign the *r*-values but also add them to the *l*-values — execute faster than the assignment-only benchmarks in the following cases:

A/B.1 for `char`, `short`, and B.1 alone also for `int`, where the speed-up is in the range from roughly 20% to more than 30% (the best for `int`-B1);

A/B.3 for `char` and `short`, and B.3 alone also for `int`, with the speed-up from 20% to 25%.

On the other hand, we see that the execution times for the A.2 benchmarks lag from those of our standard benchmarks (in Table 2) for the first three integer array types, especially for the `char` and `short`, and then also for the floating-point types. B.2 is better for `char` but greatly lags for the same types as A.2.

In our third test program, the benchmarks are the same as in Listing 1, except that in the A-types, the elements are assigned numerical constants, as was discussed in §3.2.5. The aim was to investigate the influence of this less-than-general case. The BETs for these benchmarks are shown in Table 4. Since the B-type benchmarks are the same as in Program 1, one can check that their BETs follow those from Table 2 within the standard deviation values. As for the A-type benchmarks, the BETs of A.1 and A.3 are very close to each other and rather short for all integer types. In contrast, the times of the integer versions of A.2 are approx. 55% to 72% longer(!) than those of A.1 and A.3. For the floating-point types, the situation is reversed, with A.2 BETs being very short.

In comparison to the execution times of our standard benchmarks, these A.1 and A.3 BETs are shorter for more than 1/3 for the first three integers: for `char and short` 37%, for `int` 42%, while for `llint` and the floating-point types, these times are only slightly better (the most for `llint`, 4.5%). For the A.2 benchmark, the situation is reversed for the `short` and `int` arrays, which now have much longer BETs: `short` 65% and `int` 69%. For the other data types, the execution times are close to those of our standard benchmarks.

Summarily, the shorter loop body significantly improved the performance of the A.1 and A.3 benchmarks with the arrays of the first three integer types. For the A.2 benchmark, the opposite happens for the arrays of the first two integer types.

4.3.3. BETs of Our Older Benchmarks

To connect this analysis with our earlier investigation (cf. sec.), in Table 5, we present the results obtained by running the older version of our benchmarks, now marked as Program 3-old. In this program, there is another use of a constant: in the upper limit of the conditions in the `for`-loops with running indices (benchmarks A/B.1 & 2). We have discussed that in §3.2.1, showing that this kind of `for`-loop compiled into one instruction and 2B less than the `for`-loop of our standard benchmarks, hinting at the possible speed-ups. To recapitulate, in these A-type benchmarks, there are constants both in the array-element assignment statements and in the `for`-loop conditions, while

Arr. type	a) $\overline{\Delta t}_{BET,rel}$ *(Prog. 3-old / Prog. 1)*						b) $\overline{\Delta t}_{BET,rel}$ *(Prog. 1 mod. for-lp. hd/Prog. 1)*					
	A.1	A.2	A.3	B.1	B.2	B.3	A.1	A.2	A.3	B.1	B.2	B.3
`char`	−4.7%	−29.3%	−36.9%	−13.6%	−2.1%	+0.6%	−5.3%	+13.2%	−0.3%	−14.2%	−1.8%	+0.0%
`short`	−5.0%	+37.3%	−38.0%	−13.3%	+47.2%	+0.1%	−5.6%	+33.3%	−1.5%	−13.5%	+46.3%	+0.0%
`int`	−12.2%	+46.1%	−36.9%	−12.6%	+43.6%	0.0%	−14.1%	+65.2%	−0.1%	−13.6%	+43.9%	−0.2%
`llint`	+40.6%	−14.2%	−0.3%	+26.9%	−10.3%	−0.6%	+31.4%	−28.1%	−0.5%	+27.9%	−10.4%	−1.7%
`float`	−14.6%	+68.0%	−0.2%	−13.7%	+42.6%	0.2%	−15.8%	+65.1%	−0.3%	−14.1%	+41.5%	−0.3%
`dbl.`	−16.4%	+55.1%	−0.5%	−13.1%	+34.2%	+1.2%	−16.6%	+53.9%	−1.1%	−13.8%	+32.9%	+0.4%

Table 6. Relative BETs differences to the BETs of our standard benchmarks, from Table 2, for: a) our *old* benchmarks, now in Program 3-old; b) of the A/B.1 & 2 benchmarks as in Prog. 1, but with the constant in the `for`-loop conditions (`uI < cuiN`).

in B.1 & 2, only the latter applies. In fact, in the early version of the program, there were only A.1 and A.3 benchmarks for only the `short` and `int` types, but we have upgraded it to also include all the other benchmarks and array types as in the present test programs.

In the a) part of Table 6, we compare the BETs of these benchmarks to the corresponding execution times for our standard benchmarks. Despite the expectations to see only improvements — because of the diminution in the number and length of the instructions — we can see all three cases: decreases, invariabilities, and increases in the execution times. For instance, besides the usual unexpected behavior of the `llint` array elements in A(B).1 benchmarks, with the delay of 41% (27%), we also see large increases of the A/B.2 benchmark BETs for the first two integers and both floating-point types. On the other hand, there are quite significant improvements in A.3 for the first three integer types, etc.

Now, if we compare the "old" A.1 benchmark BETs for the integers (Table 5) to the corresponding ones from our Program 3 (Table 4), we may wonder why they are longer. Indeed, they have not only the assignments of constants in their `for`-loop bodies, as the benchmarks in Program 3, but they also have the constants in their `for`-loop heads. So why does their execution last longer for the integer types?

To solve this enigma, we have first modified the `for`-loop heads in our standard benchmarks (Program 1) of the A/B.1 and A/B.2 types to be the same as in this program. After checking the compilation in assembly language, we confirmed that all the modified `for`-loops now contain only seven (7) plus one (1) instructions, as explained in §3.2.1. Then, we measured the BETs of these modified benchmarks. Their relative differences from the BETs of our standard benchmarks are in the b) part of Table 6.

First, let us observe the A/B.3 benchmarks, whose `for`-loop heads were not changed in this modified program. We also detect that the corresponding BETs do not change. The changes do occur for the A/B.1 & 2 benchmarks, but again, not as we would expect, i.e., to be either close to zero or with negative values, because the only thing that happened in the machine code is its slight reduction. On the contrary,

besides the improvements, we also see much worse performance, e.g., for A/B.2 lower integers and both floating-point types, as well as for the `llint` type in A/B.1.

By comparing Table 6.b and Table 6.a, we can see that they are similar for A/B.1 & 2 benchmarks, except for the `char`-A.2 case, where there is a huge difference. However, we do not see that the slightly "simpler," old benchmarks are also faster. Quite the contrary, they are slower, though not for much. On the other hand, the tables differ significantly in A.3 for the first three integers, giving huge advantage to the old-style benchmarks, with ≈ 37% to 38% improvement vs. almost no improvement (≈ 0% to 1.5%) for the case presented in Table 6.b.

Thus, in the results of the old benchmark BETs, we do not see any improvements in all of them. Then, we see them in A/B.1 benchmarks (except for `llint`), but smaller than in their modified version in Table 6.b. That is, instead of the expected "synergy" of the shorter loop bodies and shorter `for`-loops, these times for integer types are now worse than for the benchmarks in Program 3 (Table 4)! All this happens although the compilation of the old benchmarks gives the `for`-loops with indices (in A/B.1 & 2) of the same form and size as those which made the improvements shown in Table 6.b and the loop bodies of the exact same form and size as the benchmarks in Program 3.

The resolution of this perplex is that the execution time depends not only on the machine instructions themselves but also considerably on the context they appear in. Thus, when comparing the old version of benchmarks A.1 to those in Program 3 (cf. Table 5 vs. Table 4), their execution times for the integer arrays do not decrease as one would expect — at least slightly. Instead of that, they increase by almost 50%. They are only somewhat better than the BETs of our standard benchmarks (Table 2). On the other hand, the BETs of the B.1 benchmarks do decrease by about 12% to 13%, but not so for the `llint`, etc.

In short, knowing the machine instructions and their lengths still does not give us the full information on how the machine code will be pushed through the processor's pipelines, how fast it will be decoded, and how efficiently it will be executed.

By focusing solely on the old benchmark's execution times (Table 5), we note that the A.3 benchmark BETs for *all* integer types—that is, now *including* the `llint` type, which has otherwise been a notorious exception—are from around 30% up to 35% shorter than for A.1s! As a kind of relief from all the previous complications, this result roughly confirms our earlier, much less systematic findings obtained on the same (old) benchmarks, but only on the arrays of the `short` and `int` types, as mentioned in sec. 1.

4.3.4. A Brief Overview of the BETs With x64 Compilation and on Other Computers

After the deliberation in §3.2.4, it may come as no surprise that the 64-bit, x64 machine code does not result in much faster executions. By inspecting Table 7, we see noticeable improvements from our standard benchmarks (Prog. 1) only for the A/B.1 and B.3 BETs, again, except for the `llint` array type. Here, this bad behavior of the 64-bit integer type is particularly odd. One would expect that the x64 compilation would improve at least the execution times of the compatible, 64-bit data

$\overline{\Delta t}_{BET,rel}$ *(Prog. 1 x64 / Prog. 1 x86)*						
Arr. typ.	A.1	A.2	A.3	B.1	B.2	B.3
`char`	−11.9%	+23.1%	+32.9%	−14.8%	−6.2%	−11.0%
`short`	−12.1%	+70.1%	+23.5%	−14.4%	+46.9%	−10.8%
`int`	−17.7%	+68.6%	+1.0%	−14.0%	+44.8%	−10.8%
`llint`	+35.9%	−2.3%	+53.0%	+34.9%	−0.2%	+37.4%
`float`	−18.2%	+72.4%	+2.9%	−14.8%	+44.8%	−11.4%
`double`	−18.2%	+62.2%	−5.4%	−14.3%	+33.4%	−9.2%

Table 7. Relative differences between the BETs of our standard benchmarks (Program 1) compiled to the x64 machine code and those compiled to the x86 code (shown in Table 2).

types, but it didn't happen. Additionally, the A.3 (B.3 for `llint`), and even more A/B.2 benchmarks, show much worse behavior than the corresponding x86 code.

In Program 2, the 64-bit compilation produces faster execution than the 32-bit compilation for all A-benchmarks but not for the first two integer types (not shown in a comparison table). The improvements range from around 10% up to 25%. The B-type benchmarks' BETs are from 10% shorter up to 20% longer. In Program 3, A.1 (B.1) benchmarks' BETs show slow-downs of about 40% (40% to 60%!) for all integer types and diverse behavior for the other benchmarks and array types.

For our old benchmarks (Program 3-old), the x64 machine code shows quite uniform BETs: for the A.1 benchmarks, they are in the approximate range from 31ms to 34ms, for A.2 from 32ms to 37ms, and for all the B-type benchmarks from 32ms to 34ms. Comparing them to the BETs of the corresponding x86 versions (Table 5), it turns out they are much faster (10%−12%) only for the B.3 benchmarks (again, except for the `llint` type!). Furthermore, since the BETs for the A.3 benchmarks are, in this case, similar to the other values, and in the x86 version, those were much better, their relative lags are quite large: for `char` 55%, `short` 53%, `int` 40%, and `llint` 38%. Because of that, in the x64 version of the benchmarks, the A.3-type of access is no longer significantly faster than A.1, as in their x86 version.

We have repeated the same measurements on a few other computers with the same Wintel platform. One of them has Intel® Core™ i3-6100U CPU @ 2.30GHz, with (only) two cores and relatively low energy consumption, and is running on the Windows 10 OS. It executes our benchmarks for roughly twice the time of the much more powerful computer described in §4.2.3. The BETs obtained on it follow the trends described in the previous text for all our test programs, with a few small discrepancies for certain benchmarks and array types. It also resembles the results of our old benchmarks, i.e., show the decrease of the BETs when using the pointer incrementing (A.3) over the use of the array indices (A.1), though slightly less than in our case: 25% for `char`, 32% for `short` and `int`, and 15% for `llint`.

The 64-bit compilations on this computer behave similarly to those on our primary computer, i.e., there are minor speed-ups in some cases but also slow-downs

in others. In addition, regarding the Program 3-old, there are no improvements in A.3 over A.1 benchmarks, same as on the primary computer.

5. Conclusion

After having struggled against the myriad of different cases, peculiarities, and surprises—most of which were without clear or, at least, simple rationale—we face the dilemma of what to say as a conclusion of this computing adventure! That is why it would not be a bad idea to recall why we started it in the first place. Namely, our initial intention seemed to be quite simple: to investigate whether one can improve the time efficiency of a program code by replacing the standard, index-based access to 1D-array elements with access via pointers. The mere volume of this paper shows that the answer to this question turned out to be neither unique nor simple and even less easy to explain.

Thus, we first recapitulate how we adopted our methods. We have built upon the ideas from our ad-hoc measurements from a decade ago, in which the array-element assignment statements were kept as simple as possible. In the initial benchmarks, there were only two types of accesses: via indices (No. 1), and via the incremented pointer (now No. 3). In their `for`-loop bodies, there was a single assignment statement with the array elements as the *l*-values. From the arrays of the `short` and `int` types only, these benchmarks were expanded to all six standard numerical array (data) types). Then, we also included the benchmarks with the mirror-symmetrical assignments, in which the array elements are the *r*-values. Thus, the two types of benchmarks were formed and named A and B, as shown in Listing 1. Both of them are simple, and because of that, their compilation with the optimization turned on can produce very simplified machine code, particularly for the B type. We have discussed this matter in detail in §3.1 and justified running and investigating the non-optimized machine code, especially if analyzing its disassembled version. These benchmarks can be regarded as the starting ones upon which the methodology will be built.

In the rest of section 3, we have explained the instructions of the compiled code and discussed a few pitfalls that we had encountered and had to deal with in this investigation. In doing that, we could notice the evident consequences of the non-orthogonality of Intel's x86 instruction set, in the sense that not all addressing modes are allowed with all operations. This non-orthogonality results in inconsistencies in how the source code can and will be compiled (§3.2.1, 3.2.2, 3.2.5).

However, after all the previous analysis, the attempt to understand and explain many varieties and peculiarities of the benchmark execution times (BETs) in section 4 was everything but an easy task. While we could clarify some of the obtained results by observing the number and length of instructions in the benchmarks' assembly code, some other results showed just the opposite from what would be predicted in that way (see §4.3 and, specifically, §4.3.3). A blatant example of such inconsistency is that shorter instructions do not necessarily perform faster. Knowing the complexity of the instruction pipelines in modern processors, especially those of the CISC types, this is far from a surprise, and it makes a plausible explanation for such deviant behavior even more elusive. It is not only that Intel's x86 instructions are of variable

length—which complicates the analysis of their execution times—but we have shown that these times significantly depend on the broader context within which these instructions appear, such as the data type of the array elements and the use of special operations and the corresponding (special) registers.

As we are approaching the end of this paper, can we conclude at least something? Is there some practical advice regarding the best method of accessing the array elements on a Wintel platform? By looking at the results of this research, we see that there is no simple conclusion and no unique best choice. The best method largely depends on the type of assignment operation in the loop body and the array (data) type. For example, we have managed to repeat the results of the early measurements on our old benchmarks and showed an even greater advantage of the use of the incrementing pointers over indices than before. However, the analysis showed that these results were obtained on the rather non-general benchmarks: they had constants in the crucial parts of their loops. Furthermore, the advantage occurs for only the first three integer types (Table 5, Table 6.a).

For our standard benchmarks, with the assignment-only statements in the loop bodies, we have given concrete advice in the bulleted list in §4.3.1 (based on the BETs from Table 2). For the modified benchmarks, with addition and assignment, the conclusion is summarized in the list in §4.3.2 (based on the BETs from Table 3).

Interpreting these results the other way around, we can say that using the indices can be as good as anything else if we primarily do iterative summations and if the improvement for a few percentages achieved by incrementing the pointer is irrelevant. Alternatively, we could say that the access via indices is considerably outperformed by the use of the pointer arithmetic in the case of the assignment-only operations, but not forgetting that the `llint` type is a persistent exception, etc.

In difference to that, one conclusion is rather simple: the x64 code does not improve the benchmark performance considerably, not even for the 64-bit types. Just the opposite! Table 7 shows that the execution times for most cases are worse than for the x86 code.

From all this, we can derive a conclusion of the presented subject on the standard Wintel platform: the best approach—especially in critical applications—is to measure the execution times of the intensively used piece of code as we did and to choose the access method with the best performance.

This conclusion, especially the last piece of advice, gave us a few hints on possible future work. First, the additional benchmarks should be written in a way that could not be trivialized by the compiler optimizations, even though this would prevent easy disassembly within the VS-CV 2019 IDE. Some of them could perform the standard array and matrix operations. Second, it would be interesting to make these measurements on some other platform(s), though—as things stand now—there are not many general-purpose processor brands to pick from! Namely, in the long-solitary "processor arena," with only one genuine player, there are no serious competitors at all. Some bright prospects that this could change are announced in [14]. Let us hope that the new architectural solutions, if possible, with an appropriate and more orthogonal instruction set, will contribute to the better performance consistency of our future computations.

References

[1] B. W. Kernighan, and D. M. Ritchie D. M, *The C Programming Language* 1st and 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1978 and 1988.

[2] R. Logožar, M. Mikac, and D. Radošević, "Exploring the Access to the Static Array Elements via Indices and via Pointers — the Introductory C++ Case," Proceedings of the Central European Conference on Information and Intelligent Systems, 33rd CECIIS, Dubrovnik, 2022, pp. 507–517.

[3] B. Stroustrup, *The C++ Programming Language*, 3rd ed. Murray Hill, New Jersey: AT&T Labs. 1997.

[4] H. Schildt, *C++: The Complete Reference*, 4th ed. New York: McGraw-Hill/Osborne, 2003.

[5] Microsoft; "View disassembly code in the Visual Studio debugger," *Microsoft, Documentation*, 2022. [Online]. Available at: https://docs.microsoft.com/en-us/visualstudio/debugger/how-to-use-the-disassembly-window?view=vs-2022 [Accessed: Dec. 1, 2023].

[6] D. Evans, "x86 Assembly Guide." D. Evans, University of Virginia Computer Science, Spring 2006. [Online]. Available at: https://www.cs.virginia.edu/~evans/cs216/guides/x86.html [Accessed: Dec. 1, 2023].

[7] Intel, "Intel® 64 and IA-32 Architectures Software Developer Manuals," *Intel, Developers,* 2022. [Online]. Available at: https://www.intel.com/content/www/us/en/developer/articles/technical/intel-sdm.html [Accessed: Dec. 1, 2023].

[8] Stack Overflow, "Is there a performance difference between `i++` and `++i` in C++," *Stack Overflow Questions*, 2009–2019. [Online]. Available at: https://stackoverflow.com/questions/24901/ [Accessed: Dec. 1, 2023].

[9] Microsoft, "SYSTEM-TIME structure (miniwinbase.h)," *Microsoft, Documentation*, 2021. [Online]. Available at: https://learn.microsoft.com/en-us/windows/win32/api/minwinbase/ns-minwinbase-systemtime [Accessed: Dec. 1, 2023].

[10] Microsoft, "Acquiring high-resolution time stamps," *Microsoft, Documentation*, 2022. [Online]. Available at: https://learn.microsoft.com/en-us/windows/win32/sysinfo/acquiring-high-resolution-time-stamps [Accessed: Dec. 1, 2023].

[11] R. Logozar, "Recursive and Nonrecursive Traversal Algorithms for Dynamically Created Binary Trees," *Computer Technology and Application (David Publishing)*, vol. 3, no. 5, May., pp. 374–382, 2012.

[12] R. Logozar, "Algorithms and Data Structures for the Modeling of Dynamical Systems by Means of Stochastic Finite Automata," *Technical Gazette (Tehnički vjesnik)*, vol. 19, no. 2, Apr. –Jun., pp. 227–242, 2012.

[13] C++ reference, "std::chrono library, high_ resolution_clock class," *C++ reference, Date and time utilities."* [Online]. Available at: https://en.cppreference.com/w/cpp/chrono/high_resolution _clock [Accessed: Dec. 1, 2023].

[14] Apple, "Apple unveils M2, taking the breakthrough performance and capabilities of M1 even further," Apple, Newsroom, Press release, Jun., 2022. [Online]. Available at: https://www.apple.com/newsroom/2022/06/apple-unveils-m2-with-breakthrough-performance-and-capabilities/ [Accessed: Dec. 1, 2023].

10.31341/jios.48.1.4
Open Access
UDC 519.219:795
Original Scientific Paper

Pure Strategy Saddle Points in the Generalized Progressive Discrete Silent Duel with Identical Linear Accuracy Functions

Vadim Romanuke *v.romanuke@amw.gdynia.pl*
Faculty of Mechanical and Electrical Engineering
Polish Naval Academy, Gdynia, Poland

Abstract

A finite zero-sum game defined on a subset of the unit square is considered. The game is a generalized progressive discrete silent duel, in which the kernel is skew-symmetric, and the players, referred to as duelists, have identical linear accuracy functions featured with an accuracy proportionality factor. As the duel starts, time moments of possible shooting become denser by a geometric progression. Apart from the duel beginning and end time moments, every following time moment is the partial sum of the respective geometric series. Due to the skew-symmetry, both the duelists have the same optimal strategies and the game optimal value is 0. If the accuracy factor is not less than 1, the duelist's optimal strategy is the middle of the duel time span. If the factor is less than 1, the duel solution is not always a pure strategy saddle point. In a boundary case, when the accuracy factor is equal to the inverse numerator of the ratio that is the time moment preceding the duel end moment, the duel has four pure strategy saddle points which are of the mentioned time moments. For a trivial game, where the duelist possesses just one moment of possible shooting between the duel beginning and end moments, and the accuracy factor is 1, any pure strategy situation, not containing the duel beginning moment, is optimal.

Keywords: game of timing, silent duel, accuracy function, linear accuracy, matrix game, pure strategy saddle point

1. Introduction

Games of timing represent a wide class of competitive interaction models. These models consider a time span during which the player must make a decision of acting [1], [2], [3], [4]. If a decision is made in a two-person game commonly referred to as a duel, the other player either learns it or does not learn it until the duel ends. The latter is the case of the silent duel [5], [6], [7], [8].

Silent duels, unlike noisy duels, are more complicated to study [9], [10], [5]. In one of a few types of the silent duel, each of two players (duelists) has exactly one bullet [11], [12], [13]. The bullet is an abstraction implying an implementation of the decision of acting. This also includes issuing information about the state of a system or object modeled by the duel [3], [12], [14]. The duelist is featured with an accuracy function which, generally speaking, is a nondecreasing function of time [5], [15], [16]. It is unknown to the duelist whether a bullet was fired by the other duelist or not until

the end of the duel time span [6], [8]. The game builds on the fact that the duelist may obtain a greater payoff by firing as late as possible, but then the loss, when the other duelist shoots first, becomes more likely. If both the duelists shoot simultaneously, the payoff of each of them is 0 [9], [10], [5], [17].

In general, such a zero-sum game is

$$\left\langle X, Y, K\left(x, y\right)\right\rangle \tag{1}$$

whose kernel is defined on unit square

$$X \times Y = \left[0; 1\right] \times \left[0; 1\right] \tag{2}$$

being the Cartesian product of the duel time spans (the product is the square of the span). When the accuracy functions are linear and identical, the kernel is

$$K\left(x, y\right) = ax - ay + a^2 xy \operatorname{sign}\left(y - x\right) \text{ by } a > 0 . \tag{3}$$

Game (1) by (2) and (3) is a silent duel with linear accuracy functions of the duelists, which are allowed to shoot at any moment during the duel time span, and the duelist's accuracy is proportional to the time moment with factor a. Due to kernel (3) is skew-symmetric, i. e.

$$K\left(y, x\right) = ay - ax + a^2 yx \operatorname{sign}\left(x - y\right) = -K\left(x, y\right),$$

both the duelists have the same optimal strategies and the game optimal value is 0 [5], [8], [12]. However, even in this case of the identical linear accuracy functions, the duelist's optimal strategy is a non-continuous probability distribution as a mixed strategy with an uncountably infinite support whose measure is less than the duel time span length [5], [18], [19]. Although the duel can be a repeated game, any sequence of real-world actions cannot be unlimited (or cannot last forever), and thus practical implementation of such a solution cannot be complete [20], [21], [13], [22], [23]. This urges to consider a discrete version of the silent duel.

In a discrete silent duel, both the players can shoot only at specified time moments whose number is finite. Therefore, the kernel of the discrete silent duel is defined on a finite set of the pairs of pure strategies of the duelists. The moments (as pure strategies) of the duel beginning $x = y = 0$ and duel end $x = y = 1$ are included [24], [8], [12]:

$$X = \left\{x_i\right\}_{i=1}^{N} = Y = \left\{y_j\right\}_{j=1}^{N} = T = \left\{t_q\right\}_{q=1}^{N} \subset \left[0; 1\right]$$
$$\text{by } t_q < t_{q+1} \ \ \forall\, q = \overline{1, N-1} \text{ and } t_1 = 0,\ t_N = 1 \text{ for } N \in \mathbb{N} \setminus \left\{1\right\}. \tag{4}$$

Consequently, this finite symmetric game is a matrix game whose solution is of finite supports only [25], [26], [13], [18]. The solution is computed far easier than that in the case of infinite game (1).

In the general case, moments $\{t_q\}_{q=1}^{N}$ of possible shooting are not specified but only obey (4). As a particular case, these moments can be defined by uniformly breaking the time span. In practice, nevertheless, as the duelist approaches to the end moment $t_N = 1$, the space between consecutive moments t_q and t_{q+1}, $q = \overline{1, N-1}$, may shorten due to the growing tension and responsibility. This is equivalent to the weight (importance) of the duelist's pure strategy with approaching to the end moment grows. One of the patterns of such growth is a geometrical progression. In this case, the density of pure strategies of the duelist grows in the geometrical progression as the duelist approaches to the duel end [8], [12], [27]. Thus, if

$$t_q = \sum_{l=1}^{q-1} 2^{-l} = \frac{2^{q-1}-1}{2^{q-1}} \text{ for } q = \overline{2, N-1} \tag{5}$$

then, apart from the duel beginning and end moments, every following moment is the partial sum of the respective geometric series. Subsequently, game (1) by (4), (5) with kernel (3), where $x \in X$, $y \in Y$, becomes a generalized progressive discrete silent duel with identical linear accuracy functions. It is obvious that this duel solution depends on N and a. Besides, the kernel herein is a skew-symmetric payoff matrix [12], [5]. The goal is to study its saddle points that are pure strategy solutions of the progressive discrete silent duel with identical linear accuracy functions.

To achieve the goal, additional preliminaries of the progressive discrete silent duel are first stated in Section 2. Then pure strategy solutions for various cases of N and a are presented in Section 3. Section 4 summarizes the solution results and recapitulates their peculiarities. The study is discussed and concluded in Section 5.

2. Additional preliminaries

In fact, the progressive discrete silent duel with identical linear accuracy functions is a matrix game

$$\left\langle \{x_i\}_{i=1}^{N}, \{y_j\}_{j=1}^{N}, \mathbf{K}_N \right\rangle \tag{6}$$

by (4), (5), and payoff matrix

$$\mathbf{K}_N = \left[k_{ij}\right]_{N\times N}$$
$$\text{by } k_{ij} = K\left(x_i, y_j\right) = ax_i - ay_j + a^2 x_i y_j \operatorname{sign}\left(y_j - x_i\right) \text{ by } a > 0. \tag{7}$$

If $a = 1$ then the accuracy is exactly equal to the time moment at which the bullet is fired [9], [10], [5], [8]. Although by $a \neq 1$ the accuracy is just scaled by $a > 0$, it influences the solution by depending on whether $a > 1$ or $0 < a < 1$. This is yet to be shown below.

Formally, the most trivial case of $N=2$, where the duelist is allowed to shoot at either the very beginning or end, is not excluded. As the shooting is allowed only at moments $t_1=0$ and $t_2=1$, the respective payoff matrix is

$$\mathbf{K}_2=\left[k_{ij}\right]_{2\times 2}=\begin{bmatrix}0 & -a\\ a & 0\end{bmatrix} \tag{8}$$

whence situation

$$\left\{x_2, y_2\right\}=\left\{1,1\right\} \tag{9}$$

is the single saddle point here.

In general,

$$K\left(x_1, y_N\right)=K\left(0,1\right)=-a=-K\left(1,0\right). \tag{10}$$

Therefore, whichever number N is (the case of $N=1$ is naturally excluded), the first row of matrix (7) contains a negative entry and so this row does not contain saddle points (because the minimum of the first row does not exceed $-a<0$ and thus the game optimal value $v_{\text{opt}}=0$ cannot be reached in this row). Due to the skew-symmetry of matrix (7), the stated inference is immediately followed by that the first column does not contain saddle points either. In the further consideration, only the inferences on saddle points in definite rows of matrix (7), which imply the same inferences on saddle points in respective columns, will be stated. It is clear that only a zero entry of matrix (7) can be a saddle point. If a row contains a negative entry, this row and respective column do not contain saddle points. If a nonnegative row contains a zero entry off the main diagonal, and this entry is a saddle point, both the zero entries are saddle points [2], [5], [18].

3. Existence of pure strategy saddle points

It is needful to get started with the most trivial case, apart from (8). This is the case of $N=3$, where the shooting, apart from the very beginning and end moments $t_1=0$, $t_3=1$, is also allowed at moment $t_2=\frac{1}{2}$.

Theorem 1. In a progressive discrete silent duel (6) by (4), (5), (7) for $N=3$, situation

$$\left\{x_2, y_2\right\}=\left\{\frac{1}{2},\frac{1}{2}\right\} \tag{11}$$

is optimal only by $a\geqslant 1$ but it is never optimal by $0<a<1$. Besides, saddle point (11) is single by $a>1$. Any pure strategy situation in the 3×3 duel, not containing the duel beginning moment, is optimal by $a=1$, whereas situation

$$\{x_3, y_3\} = \{1, 1\} \tag{12}$$

is the single saddle point by $0 < a < 1$.

Proof. Due to (10), situation

$$\{x_1, y_1\} = \{0, 0\}$$

is never optimal in the duel. The respective payoff matrix is

$$\mathbf{K}_3 = \left[k_{ij}\right]_{3\times 3} = \begin{bmatrix} 0 & -\frac{a}{2} & -a \\ \frac{a}{2} & 0 & \frac{a}{2}(a-1) \\ a & -\frac{a}{2}(a-1) & 0 \end{bmatrix}. \tag{13}$$

If $a > 1$ then the second row of matrix (13) is nonnegative and the third row contains a negative entry. The only zero entry in the second row is k_{22}. So, situation (11) is optimal and it is the single saddle point for (13) by $a > 1$. If $a = 1$ then the respective payoff matrix is

$$\mathbf{K}_3 = \left[k_{ij}\right]_{3\times 3} = \begin{bmatrix} 0 & -\frac{1}{2} & -1 \\ \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{14}$$

This matrix game has four saddle points: situation (11), situation (12), and non-symmetric situations

$$\{x_2, y_3\} = \left\{\frac{1}{2}, 1\right\}, \tag{15}$$

$$\{x_3, y_2\} = \left\{1, \frac{1}{2}\right\}. \tag{16}$$

If $0 < a < 1$ then the second row of matrix (13) contains a negative entry and the third row is nonnegative. The only zero entry in the third row is k_{33}. So, situation (12) is the single saddle point for (13) by $0 < a < 1$. □

Now, the general case is to be considered. Let it be divided into the subcases of $0 < a < 1$ and $a \geqslant 1$.

Theorem 2. In a progressive discrete silent duel (6) by (4), (5), (7) for $N \in \mathbb{N} \setminus \{1, 2, 3\}$, there exists only one $n \in \overline{\{3, N-1\}}$ by $0 < a < 1$ such that situation

$$\{x_n, y_n\} = \left\{ \frac{2^{n-1}-1}{2^{n-1}}, \frac{2^{n-1}-1}{2^{n-1}} \right\} \tag{17}$$

is a single saddle point by

$$a \in \left[\frac{1}{2^{n-1}-1}; \frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} \right] \subset (0;1) \tag{18}$$

and situation

$$\{x_N, y_N\} = \{1, 1\} \tag{19}$$

is a saddle point by

$$a \in \left(0; \frac{1}{2^{N-2}-1} \right] \subset (0;1). \tag{20}$$

Besides, situation (11) is never optimal by $0 < a < 1$.

Proof. Due to

$$K(x_2, y_N) = K\left(\frac{1}{2}, 1\right) = \frac{a}{2}(a-1) = -K\left(1, \frac{1}{2}\right) < 0 \text{ by } 0 < a < 1, \tag{21}$$

situation (11) is never optimal for $N \in \mathbb{N} \setminus \{1, 2, 3\}$. Consider entry k_{nn} for $n \in \overline{\{3, N-1\}}$ and $N \in \mathbb{N} \setminus \{1, 2, 3\}$. This entry is the result of when both the duelists simultaneously shoot at moment

$$t_n = \sum_{l=1}^{n-1} 2^{-l} = \frac{2^{n-1}-1}{2^{n-1}} \tag{22}$$

corresponding to situation (17). If situation (17) is a saddle point, then, in the n-th row of matrix (7), inequalities

$$K(x_n, y_j) = ax_n - ay_j - a^2 x_n y_j \geqslant 0 \quad \forall\, y_j < x_n \tag{23}$$

and

$$K(x_n, y_j) = ax_n - ay_j + a^2 x_n y_j \geqslant 0 \quad \forall\, y_j > x_n \tag{24}$$

must hold. From inequality (23) it follows that

$$\frac{x_n}{1+ax_n} \geqslant y_j \quad \forall\, y_j < x_n\,. \tag{25}$$

As

$$y_j \leqslant \frac{2^{n-2}-1}{2^{n-2}} < \frac{2^{n-1}-1}{2^{n-1}} = x_n \tag{26}$$

then inequality (25) is transformed into

$$\frac{2^{n-1}-1}{2^{n-1}} \cdot \frac{1}{1+a\cdot\dfrac{2^{n-1}-1}{2^{n-1}}} \geqslant \frac{2^{n-2}-1}{2^{n-2}},$$

$$\frac{2^{n-1}-1}{2^{n-1}+a\cdot\left(2^{n-1}-1\right)} \geqslant \frac{2^{n-2}-1}{2^{n-2}},$$

$$2^{n-2}\cdot\left(2^{n-1}-1\right) \geqslant 2^{n-1}\cdot\left(2^{n-2}-1\right)+a\cdot\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right),$$

whence

$$\frac{2^{n-2}\cdot\left(2^{n-1}-1\right)-2^{n-1}\cdot\left(2^{n-2}-1\right)}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} =$$

$$= \frac{2^{n-1}-2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} = \frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} \geqslant a\,. \tag{27}$$

From inequality (24) it follows that

$$\frac{x_n}{1-ax_n} \geqslant y_j \quad \forall\, y_j > x_n\,. \tag{28}$$

As

$$1 \geqslant y_j > \frac{2^{n-1}-1}{2^{n-1}} = x_n \tag{29}$$

then inequality (28) is transformed into

$$\frac{2^{n-1}-1}{2^{n-1}} \cdot \frac{1}{1-a\cdot\dfrac{2^{n-1}-1}{2^{n-1}}} \geqslant 1,$$

$$\frac{2^{n-1}-1}{2^{n-1}-a\cdot\left(2^{n-1}-1\right)} \geqslant 1\,. \tag{30}$$

If $0<a<1$ then

$$\frac{2^{n-1}}{2^{n-1}-1}>1>a\,, \tag{31}$$

whence

$$2^{n-1}-a\cdot\left(2^{n-1}-1\right)>0$$

and inequality (30) is written as

$$2^{n-1}-1\geqslant 2^{n-1}-a\cdot\left(2^{n-1}-1\right). \tag{32}$$

Thus, inequality (32) is followed by

$$a\geqslant\frac{1}{2^{n-1}-1}\,. \tag{33}$$

Consequently, if the membership in (18) is true, then the n-th row of matrix (7) is nonnegative. The n-th column is nonpositive, whereas $k_{nn}=0$ is a maximin and minimax entry, and situation (17) is indeed a saddle point.

Inasmuch as

$$\begin{aligned}&\frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}-\frac{1}{2^{n-1}-1}=\\&=\frac{2^{n-2}-2^{n-2}+1}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}=\\&=\frac{1}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}>0\quad\forall\, n=\overline{3,\,N-1}\end{aligned} \tag{34}$$

and

$$\begin{aligned}&\frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}-1=\frac{2^{n-2}-\left(2^{n-1}\cdot 2^{n-2}-2^{n-2}-2^{n-1}+1\right)}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}=\\&=\frac{2^{n}-2^{2n-3}-1}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}=\\&=\frac{2^{n}\cdot\left(1-2^{n-3}\right)-1}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}<0\quad\forall\, n=\overline{3,\,N-1}\end{aligned} \tag{35}$$

then there always $\exists\, n\in\left\{\overline{3,\,N-1}\right\}$ such that the inclusion in (18) is true, and the interval in (18) is nonempty, and situation (17) is a saddle point. Suppose that another entry on the main diagonal exists, k_{mm} by $m\in\left\{\overline{3,\,N-1}\right\}$ and $m\neq n$, which is a saddle point also. This can be possible if both the membership in (18) and membership

$$a \in \left[\frac{1}{2^{m-1}-1}; \frac{2^{m-2}}{\left(2^{m-1}-1\right)\cdot\left(2^{m-2}-1\right)} \right] \tag{36}$$

are true, i. e. intersection

$$\left[\frac{1}{2^{n-1}-1}; \frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} \right] \mathrm{I} \left[\frac{1}{2^{m-1}-1}; \frac{2^{m-2}}{\left(2^{m-1}-1\right)\cdot\left(2^{m-2}-1\right)} \right] \neq \varnothing . \tag{37}$$

If $m < n$ then

$$\frac{1}{2^{m-1}-1} > \frac{1}{2^{n-1}-1}$$

and so inequality

$$\frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} > \frac{1}{2^{m-1}-1} \tag{38}$$

must hold for (37). However,

$$\begin{aligned} &\frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)} - \frac{1}{2^{m-1}-1} = \\ &= \frac{2^{n-2}\cdot 2^{m-1} - 2^{n-2} - \left(2^{n-1}\cdot 2^{n-2} - 2^{n-2} - 2^{n-1} + 1\right)}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)\cdot\left(2^{m-1}-1\right)} = \\ &= \frac{2^{n-2}\cdot 2^{m-1} + 2^{n-1} - 2^{2n-3} - 1}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)\cdot\left(2^{m-1}-1\right)} = \\ &= \frac{2^{n-1}\cdot\left(2^{m-2}+1-2^{n-2}\right)-1}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)\cdot\left(2^{m-1}-1\right)} . \end{aligned} \tag{39}$$

As $3 \leqslant m < n$ then

$$2^{m-2} + 1 < 2^{n-2}$$

and thus

$$2^{m-2} + 1 - 2^{n-2} < 0 ,$$

whence

$$2^{n-1}\cdot\left(2^{m-2}+1-2^{n-2}\right)-1 < 0 .$$

So, ratio (39) is negative and inequality (38) does not hold. If $m > n$ then

$$\frac{1}{2^{m-1}-1}<\frac{1}{2^{n-1}-1}$$

and so inequality

$$\frac{2^{m-2}}{\left(2^{m-1}-1\right)\cdot\left(2^{m-2}-1\right)}>\frac{1}{2^{n-1}-1} \tag{40}$$

must hold for (37). By analogy to (39), just by interchanging m and n,

$$\frac{2^{m-2}}{\left(2^{m-1}-1\right)\cdot\left(2^{m-2}-1\right)}-\frac{1}{2^{n-1}-1}=$$
$$=\frac{2^{m-1}\cdot\left(2^{n-2}+1-2^{m-2}\right)-1}{\left(2^{m-1}-1\right)\cdot\left(2^{m-2}-1\right)\cdot\left(2^{n-1}-1\right)}. \tag{41}$$

As $3\leqslant n<m$ then

$$2^{n-2}+1-2^{m-2}<0,$$

whence

$$2^{m-1}\cdot\left(2^{n-2}+1-2^{m-2}\right)-1<0.$$

So, ratio (41) is negative and inequality (40) does not hold. Therefore, saddle point (17) existing for some $n\in\left\{\overline{3,N-1}\right\}$ is single for any $N\in\mathbb{N}\setminus\{1,2,3\}$.

Now, consider entry k_{NN} for $N\in\mathbb{N}\setminus\{1,2,3\}$. This entry is the result of when both the duelists simultaneously shoot at the duel end corresponding to situation (19). If situation (19) is a saddle point, then, in the N-th row of matrix (7), inequality

$$K\left(x_N,y_j\right)=ax_N-ay_j-a^2x_Ny_j=a-ay_j-a^2y_j\geqslant 0\quad\forall\, y_j<x_N=1 \tag{42}$$

must hold. From inequality (42) it follows that

$$\frac{1}{1+a}\geqslant y_j\quad\forall\, y_j<1. \tag{43}$$

As

$$y_j\leqslant y_{N-1}=\frac{2^{N-2}-1}{2^{N-2}}<1=x_N \tag{44}$$

then inequality (43) is transformed into

$$\frac{1}{1+a}\geqslant\frac{2^{N-2}-1}{2^{N-2}},$$

$$2^{N-2} \geqslant 2^{N-2} - 1 + a \cdot \left(2^{N-2} - 1\right),$$

$$1 \geqslant a \cdot \left(2^{N-2} - 1\right),$$

whence

$$1 > \frac{1}{2^{N-2} - 1} \geqslant a, \tag{45}$$

which is true due to the membership in (20) is true, i. e. inequality (42) holds by (45). Consequently, if the membership in (20) is true, then situation (19) is a saddle point. □

Theorem 3. In a progressive discrete silent duel (6) by (4), (5), (7) for $N \in \mathbb{N} \setminus \{1, 2\}$, situation (11) is optimal for $a \geqslant 1$. For $N \in \mathbb{N} \setminus \{1, 2, 3\}$, this saddle point is the single one.

Proof. For $N = 3$, owing to *Theorem 1*, situation (11) is the single saddle point for (13) by $a > 1$ and it is one of the four saddle points for (13) by $a = 1$. For $N \in \mathbb{N} \setminus \{1, 2, 3\}$ consider entry k_{22} that is the result of when both the duelists simultaneously shoot at the middle of the duel time span corresponding to situation (11). This entry is in the second row of matrix (7), where

$$K\left(x_2, y_1\right) = K\left(\frac{1}{2}, 0\right) = \frac{a}{2} > 0 \tag{46}$$

and

$$K\left(\frac{1}{2}, y_j\right) = \frac{a}{2} - a y_j + \frac{a^2}{2} y_j = \frac{a}{2} \cdot \left(1 - 2 y_j + a y_j\right) \quad \forall\, y_j > \frac{1}{2}. \tag{47}$$

If $a = 1$ then

$$1 - 2 y_j + a y_j = 1 - y_j \geqslant 0$$

and (47) is nonnegative:

$$K\left(\frac{1}{2}, y_j\right) = \frac{a}{2} \cdot \left(1 - 2 y_j + a y_j\right) = \frac{1 - y_j}{2} \geqslant 0 \quad \forall\, y_j > \frac{1}{2}, \tag{48}$$

where

$$K\left(\frac{1}{2}, y_N\right) = K\left(\frac{1}{2}, 1\right) = k_{2N} = 0. \tag{49}$$

Inequality (46) with inequality (48) and (49) imply that the second row by $a = 1$, except for entries $k_{22} = 0$ and $k_{2N} = 0$, contains only positive entries, whence situation (11) is a saddle point. Owing to (45) is false, inequality (42) is impossible,

the N-th column cannot contain saddle points, and so saddle point (11) is the single one in the second row. Inequalities (46) and (48) also imply that entries $k_{i2}<0$ $\forall i=\overline{3,N-1}$ in the second column, so saddle point (11) is the single one by $a=1$.

If $a>1$ then, as $\frac{1}{2}<y_j\leqslant 1$,

$$\left(1-y_j\right)^2\geqslant 0,$$

$$1-2y_j+y_j^2\geqslant 0,$$

$$2y_j-1\leqslant y_j^2,$$

$$\frac{2y_j-1}{y_j}\leqslant y_j,$$

and

$$\frac{2y_j-1}{y_j}\leqslant y_j\leqslant 1<a \text{ by } y_j\in\left(\frac{1}{2};1\right]. \tag{50}$$

Inequality (50) is followed by inequality

$$1-2y_j+ay_j>0 \tag{51}$$

whence (47) is positive:

$$K\left(\frac{1}{2},y_j\right)=\frac{a}{2}\cdot\left(1-2y_j+ay_j\right)>0 \quad \forall\, y_j>\frac{1}{2}. \tag{52}$$

Inequality (46) with inequality (52) imply that the second row by $a>1$, except for entry $k_{22}=0$, contains only positive entries, whence situation (11) is a saddle point. Inequalities (46) and (52) also imply that situation (11) is the single saddle point in the second row, and entries $k_{i2}<0$ $\forall i=\overline{3,N-1}$ in the second column. The latter implies that the second row is the single nonnegative row by $a>1$ and saddle point (11) is the single one also. □

Strictly speaking, *Theorem 2* does not claim that pure strategy solutions (17) and (19) by respective conditions (18) and (20) are the only possible pure strategy saddle points of matrix (7). Nevertheless, by *Theorem 3*, pure strategy saddle point (11) is single for $a\geqslant 1$ and $N\in\mathbb{N}\setminus\{1,2,3\}$.

4. Peculiarities of the solutions

A peculiarity in *Theorem 2* is that the left endpoint in the membership (18) closed interval by $n = N-1$ and the right endpoint in the membership (20) half-interval are the same. This peculiar case of the duelist's accuracy factor a is considered separately.

Theorem 4. In the progressive discrete silent duel (6) by (4), (5), (7) for $N \in \mathbb{N} \setminus \{1, 2, 3\}$, situations (19),

$$\{x_{N-1}, y_{N-1}\} = \left\{\frac{2^{N-2}-1}{2^{N-2}}, \frac{2^{N-2}-1}{2^{N-2}}\right\} \tag{53}$$

and

$$\{x_{N-1}, y_N\} = \left\{\frac{2^{N-2}-1}{2^{N-2}}, 1\right\} \tag{54}$$

and

$$\{x_N, y_{N-1}\} = \left\{1, \frac{2^{N-2}-1}{2^{N-2}}\right\} \tag{55}$$

are saddle points by

$$a = \frac{1}{2^{N-2}-1}. \tag{56}$$

Apart from situations (19), (53) — (56), there are no other saddle points in this duel.

Proof. Owing to *Theorem 2*, here $n = N-1$ and (56) satisfies both (20) and (18). Thus, situations (19) and (53) are saddle points. Besides,

$$\begin{aligned} K\left(\frac{2^{N-2}-1}{2^{N-2}}, 1\right) &= \frac{1}{2^{N-2}-1}\cdot\frac{2^{N-2}-1}{2^{N-2}} - \frac{1}{2^{N-2}-1} + \frac{1}{\left(2^{N-2}-1\right)^2}\cdot\frac{2^{N-2}-1}{2^{N-2}} = \\ &= \frac{1}{2^{N-2}} - \frac{1}{2^{N-2}-1} + \frac{1}{2^{N-2}-1}\cdot\frac{1}{2^{N-2}} = \\ &= \frac{2^{N-2}-1-2^{N-2}+1}{\left(2^{N-2}-1\right)\cdot 2^{N-2}} = 0 = K\left(1, \frac{2^{N-2}-1}{2^{N-2}}\right) \end{aligned} \tag{57}$$

that implies the optimality of situations (54) and (55).

The first and second rows do not contain saddle points, so it remains to consider the rows whose index is $m \in \{\overline{3, N-2}\}$. Owing to *Theorem 2*, situation (53) is the single saddle point on the main diagonal, not considering the last row. This implies

that each of the $N-4$ rows whose index is $m \in \overline{\{3, N-2\}}$ contains at least one negative entry, i. e. these rows do not contain saddle points. □

If the duelist's accuracy factor satisfying inequality $0<a<1$ is not (56), this is another supplement to *Theorem 2*.

Theorem 5. A progressive discrete silent duel (6) by (4), (5), (7) for $N \in \mathbb{N} \setminus \{1, 2, 3\}$ by (18) and

$$a \neq \frac{1}{2^{N-2}-1} \tag{58}$$

has a single saddle point (17).

Proof. Owing to *Theorem 2*, while situation (17) is a saddle point, the membership in (20) is impossible due to (58). Therefore, situation (19) is not a saddle point, and saddle point (17) is the single one on the main diagonal. Analogously to that in *Theorem 4*, this implies that each of the $N-4$ rows whose index is $m \in \overline{\{3, N-1\}} \setminus \{n\}$ contains at least one negative entry, i. e. these rows do not contain saddle points. □

So, pure strategy saddle point (17) is indeed single by (18) and (58). The next assertion is for the case when the membership in (18) is false.

Theorem 6. A progressive discrete silent duel (6) by (4), (5), (7) for $N \in \mathbb{N} \setminus \{1, 2, 3\}$ by

$$a \in \left(0; \frac{1}{2^{N-2}-1}\right) \tag{59}$$

has the single saddle point (19).

Proof. Owing to *Theorem 2*, situation (19) is a saddle point by (59). Analogously to (42) — (45), when (59) is true,

$$1 > \frac{1}{2^{N-2}-1} > a,$$

$$1 > a \cdot \left(2^{N-2}-1\right),$$

$$2^{N-2} > 2^{N-2} - 1 + a \cdot \left(2^{N-2}-1\right),$$

$$\frac{1}{1+a} > \frac{2^{N-2}-1}{2^{N-2}} = y_{N-1} \geqslant y_j,$$

$$\frac{1}{1+a} > y_j,$$

$$1-y_j-ay_j>0 \quad \forall\, y_j<1. \tag{60}$$

Using inequality (60), there is inequality

$$K\left(x_N, y_j\right)=ax_N-ay_j-a^2x_Ny_j=a\left(1-y_j-ay_j\right)>0 \quad \forall\, y_j<x_N=1 \tag{61}$$

in the N-th row of matrix (7). Inequality (61) implies that situation (19) is the single saddle point in the N-th row, and entries $k_{iN}<0 \quad \forall\, i=\overline{1, N-1}$ in the N-th column. The latter implies that the N-th row is the single nonnegative row by (59) and saddle point (19) is the single one also. □

Consequently, by *Theorems 5* and *6*, a progressive discrete silent duel (6) by (4), (5), (7) for $N\in\mathbb{N}\setminus\{1, 2, 3\}$ and $0<a<1$ by either (58) or (59) has a single pure strategy saddle point. If (59) holds, then every duel has the same pure strategy solution, by which the optimal behavior of the duelist is to act (shoot) at the duel end moment. Otherwise, excluding the boundary case of *Theorem 4*, pure strategy saddle point (17) is single by

$$a\in\left(\frac{1}{2^{n-1}-1}; \frac{2^{n-2}}{\left(2^{n-1}-1\right)\cdot\left(2^{n-2}-1\right)}\right] \tag{62}$$

and, if (62) does not hold, the duel is not solved in pure strategies.

5. Discussion and conclusion

Duels are game-theoretic patterns used to model timing competitive interactions between two sides personified as players or duelists. The solutions to duels allow players holding at the most reasonable strategies and develop rationalized processes of sharing resources for which the players compete [3], [24], [17], [25], [22], [12]. Particular examples of the interactions are auctioning, advertising, market control struggle, product placement, retailing, and many other socioeconomic competitive processes between two sides [13], [15], [23], [12], [14].

Although there are almost no linearly developing real-time processes, the linear accuracy with a proportionality factor is a generalization that allows to make an initial approximation of the interaction model. Besides, the accuracy factor allows to consider various patterns of the duelists, which may be worse or better shooters (i. e., be worse or better at making smarter decisions). In addition, specifying locations of moments (4) of possible shooting as (5) is natural as well. As the duelist approaches to the duel end, the tension and responsibility grow, and so time moments of possible shooting must be located progressively denser. The geometrical progression is a quite natural pattern of the growth, which still can be made more complicated by adding some noise.

The proved assertions contribute to the games of timing a few specificities of the discrete silent duel. Firstly, a specificity consists in that the solution of a generalized

progressive discrete silent duel, with identical linear accuracy functions of the duelists, is not always a pure strategy saddle point if the duelist's accuracy factor is less than 1 (as a corollary from *Theorems 2*, *5* and *6*). Secondly, if the factor is not less than 1, the duelist's optimal strategy is the middle of the duel time span (*Theorem 3*). For a trivial game, where the duelist possesses just one moment of possible shooting between the duel beginning and end moments, and the accuracy factor is 1, any pure strategy situation in this 3×3 duel, not containing the duel beginning moment, is optimal (*Theorem 1*). As the accuracy factor becomes less than 1, this 3×3 duel has only one pure strategy saddle point which is of the duel end moments (*Theorem 1*). In the boundary case, when the accuracy factor is equal to the inverse numerator of the ratio that is the time moment preceding the duel end moment, the duel has four pure strategy saddle points which are of the mentioned time moments (*Theorem 4*).

Nonlinearity in the accuracy function can be a matter of further research on progressive discrete silent duels. For instance, it can be the quadratic and cubic accuracy. A special case is the square-root accuracy implying that the duelist is better at shooting (compared, e. g., to the linear accuracy by $a=1$ as $\sqrt{t}>t$ for $0<t<1$).

Acknowledgements

This work was technically supported by the Faculty of Mechanical and Electrical Engineering at the Polish Naval Academy, Gdynia, Poland.

References

[1] C. Aliprantis and S. Chakrabarti, *Games and Decision Making*. New York, NY, USA: Oxford University Press, 2000.

[2] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991.

[3] Y. Teraoka, "A two-person game of timing with random arrival time of the object," *Mathematica Japonica*, vol. 24, pp. 427 — 438, 1979.

[4] R. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA, USA: Harvard University Press, 1991.

[5] N. N. Vorob'yov, *Game theory for economist-cyberneticians*. Moscow, USSR: Nauka, 1985 (in Russian).

[6] S. Alpern and J. V. Howard, "A short solution to the many-player silent duel with arbitrary consolation prize," *European Journal of Operational Research*, vol. 273, iss. 2, pp. 646 — 649, 2019.

[7] S. Trybuła, "Solution of a silent duel with arbitrary motion and with arbitrary accuracy functions," *Optimization*, vol. 27, pp. 151 — 172, 1993.

[8] V. V. Romanuke, "Fast solution of the discrete noiseless duel with the nonlinear scale on the linear accuracy functions," *Herald of Khmelnytskyi national university. Economical sciences*, no. 5, vol. 4, pp. 61 — 66, 2010.

[9] E. N. Barron, *Game Theory : An Introduction* (2nd ed.). Hoboken, New Jersey, USA: Wiley, 2013.

[10] R. A. Epstein, *The Theory of Gambling and Statistical Logic* (2nd ed.). Burlington, Massachusetts, USA: Academic Press, 2013.

[11] G. Owen, *Game Theory*. San Diego, CA, USA: Academic Press, 2001.

[12] V. V. Romanuke, "Discrete noiseless duel with a skewsymmetric payoff function on the unit square for models of socioeconomic competitive processes with a finite number of pure strategies," *Cybernetics and Systems Analysis*, vol. 47, iss. 5, pp. 818 — 826, 2011.

[13] N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*. Cambridge, UK: Cambridge University Press, 2007.

[14] C.-Y. Lo and Y.-T. Chang, "Strategic analysis and model construction on conflict resolution with action game theory," *Journal of Information and Organizational Sciences*, vol. 34, no. 1, pp. 117 — 132, 2010.

[15] M. J. Osborne, *An Introduction to Game Theory*. Oxford, UK: Oxford University Press, 2003.

[16] J. P. Lang and G. Kimeldorf, "Duels with continuous firing," *Management Science*, vol. 2, no. 4, pp. 470 — 476, 1975.

[17] T. C. Schelling, *The Strategy of Conflict*. Cambridge, MA, USA: Harvard University Press, 1980.

[18] V. V. Romanuke, *Theory of Antagonistic Games*. Lviv, Ukraine: New World — 2000, 2010.

[19] R. Laraki, E. Solan, and N. Vieille, "Continuous-time games of timing," *Journal of Economic Theory*, vol. 120, iss. 2, pp. 206 — 238, 2005.

[20] V. V. Romanuke, "Convergence and estimation of the process of computer implementation of the optimality principle in matrix games with apparent play horizon," *Journal of Automation and Information Sciences*, vol. 45, iss. 10, pp. 49 — 56, 2013.

[21] T. Vincent and J. Brown, *Evolutionary Game Theory, Natural Selection, and Darwinian Dynamics*. Cambridge, UK: Cambridge University Press, 2005.

[22] S. Karlin, *The Theory of Infinite Games. Mathematical Methods and Theory in Games, Programming, and Economics*. London — Paris: Pergamon, 1959.

[23] V. V. Romanuke, "Approximation of unit-hypercubic infinite antagonistic game via dimension-dependent irregular samplings and reshaping the payoffs into flat matrix wherewith to solve the matrix game," *Journal of Information and Organizational Sciences*, vol. 38, no. 2, pp. 125 — 143, 2014.

[24] S. Siddharta and M. Shubik, "A model of a sudden death field-goal football game as a sequential duel," *Mathematical Social Sciences*, vol. 15, pp. 205 — 215, 1988.

[25] A. J. Jones, *Game Theory: Mathematical Models of Conflict*. West Sussex, UK: Horwood Publishing, 2000.

[26] V. Smirnov and A. Wait, "Innovation in a generalized timing game," *International Journal of Industrial Organization*, vol. 42, pp. 23 — 33, 2015.

[27] J.-H. Steg, "On identifying subgame-perfect equilibrium outcomes for timing games," *Games and Economic Behavior*, vol. 135, pp. 74 — 78, 2022.

10.31341/jios.48.1.5 UDC 004.738.5:687
Open Access Original Scientific Paper

Physical Internet Enabled Traceability Systems for Sustainable Supply Chain Management

Bakhta Nachet *nachatdal@yahoo.fr*
Faculty of Applied and Exact Sciences
University of Oran1, Oran, Algeria

Mohammed Frendi *mohammedfrendi@yahoo.fr*
Faculty of Applied and Exact Sciences
University of Oran1, Oran, Algeria

Abdelkader Adla *abdelkader.adla@gmail.com*
Faculty of Applied and Exact Sciences
University of Oran1, Oran, Algeria

Abstract

Current supply chain management (SCM) requires the control of physical and information flows in order to satisfy the customer, i.e. deliver the right product to the customer at the right place, at the right time, at the right price and at the lowest cost. SCM is inseparable from traceability which makes reliable the said flows, accelerates the transmission of information on these flows, allows to access a detailed knowledge of the movements, and makes the flows visible. In order to streamline and monitor, if possible, in real time and permanently these logistical processes, we propose the design and implementation of an Ontology-based traceability system based on an architectural model for the physical Internet using computing resources such as Cloud computing, Fog computing and Internet of Things (IoT) to achieve efficiency and sustainability goals. To evaluate our system, we were able to carry out all the queries that the user can express whether he is a customer, a supplier or a manager.

Keywords: Supply Chain Management, Logistics, Traceability systems, Physical Internet, Internet of Things, Cloud and Fog Computing

1. Introduction

In recent years, logistics has undergone a remarkable evolution at several levels such as production chain, storage policies, and particularly the transportation of goods. Due to industrialization, mass production, the internationalization of trade, the diversification of distribution circuits and the development of ever more numerous and sophisticated products, the sectors have become more complex, adding many intermediate steps between production and the consumer. Supply chains have also become highly interdependent. Intense exchanges of products and information have

developed within and between the sectors. The internationalization of trade has taken on a large scale.

The current logistics systems allow exchanging flows of goods on both sides of the globe in a few weeks, while guaranteeing a good quality of service at attractive prices. This progress is related, on the one hand, to technological developments in transport means (truck, container ship, aircraft, etc.) and in packaging and handling (pallets, containers, forklift, container gantries, etc.), and, on the other hand, to organizational progress. But these performances still hide inefficiencies throughout the supply chains and particularly in the transportation of goods [1],[2],[3],[4]. Indeed, this component is more important today due to its delays, the probable increase in the price of oil, taxes and the resulting negative externalities (Carbon dixide (CO2), fine particles, etc.).

Logistics management for flow control is inseparable from traceability, which makes these flows visible. Traceability enables to track or study an entity qualitatively and quantitatively in space and time by means of recorded identifications [5],[6]. The term "entity" can be a process, a product, a person or an organization. In logistics, traceability can refer to direct properties of the product and/or its associated ones such as product condition (temperature, humidity, etc.) throughout the supply chain, batch numbers, the origin of materials and parts, the history of processes applied to the product, the distribution and location of the product after delivery [7], enabling both suppliers and customers to track the goods throughout the chain and locating them along their route.

Traceability systems are technical tools intended to help the company to comply with defined objectives. They will be used to determine the history and/or location of a product and all of its components. From the point of view of information management, setting up a traceability system in a logistics chain (supply chain) consists in systematically associating an information flow with a physical flow; the objective is to be able to find, at any time, information, previously determined, relating to batches or groups of products, based on one or more identifiers [8],[9]. These systems are useful especially in perishable products such as foods and pharmaceutical products, in particular to ensure compliance with the cold chain, as well as the expiry dates relating to the various products [10]. Several solutions are used in automatic identification technologies such as barcode, RFID tag, and matrix codes. RFID tag is currently the most used.

In order to ensure the products traceability, a traceability system is composed of: A hardware platform which consists of a Material Handling System (MHS) with Radio Frequency Identification (RFID), a read/write equipment and barcodes), and a software part which is made up of input/output interfaces associated with tools for controlling, processing and storing data flows in the database [11].

The integration of the Cloud has offered several advantages, particularly in the field of product traceability in supply chains [12]. But, as applications using the distributed services of the IoT are increasingly used, cloud computing can no longer meet all this demand, particularly applications that are sensitive to latency. In order to overcome these limitations, the Fog computing paradigm is used. Fog computing, which represents an extension of cloud computing, is characterized by its calculation

and analysis capabilities and its real-time responses [13]. It could improve the performance of different applications such as logistics.

The objective of this research is to design a product traceability system to control the flow of transported goods using computing resources such as Cloud computing, Fog computing and Internet of Things (IoT). To this end, we create an ontology that represents the knowledge related to the product traceability in an interconnected supply chain considering the concepts of the Physical Internet (PI-container, PI-hub, and RFID) and a Fog computing architecture associated with a Cloud to collect information and record product events from any point in the supply chain. In this way, the ontology represents the knowledge related to logistics in order to satisfy user demands in terms of the condition, location and traceability of their products. The regular reads of RFID tags by specific readers allows capturing information on the conditions of the product which is then transmitted in the form of an event to the fog.

The remainder of the article is as follows: we first briefly presented the notions of traceability in supply chains as well as the evolution of the latter. Then, we presented the field of the physical Internet and the notions of Internet of Things (ioT), Cloud and Fog computing which represent a new technology applied to logistics in order to meet the challenges of current logistics. Next, we define the research context of setting up a physical internet (PI) to address product traceability and transportation. In section 4, we describe our ontology approach in the domain dedicated to traceability in a current supply chain using the PI paradigm. The system evaluation is presented in section 5. Finally, some clues on future work are given before concluding.

2. Literature Review

Several works have been carried out in the context of logistics traceability. In [26], a traceability system in the meat industry is proposed. The system identifies the products by a unique RFID sequential number or barcode. Data is transferred from combined barcode and RFID readers to a traceability database.

A traceability system in the cold chain of fish is described in [27]. In this system used smart RFID tags to identify products and multiple sensors to capture real-time information about temperature, humidity, and light.

In [28], the authors have defined three levels of traceability: traceability of physical flows, traceability of processes, and traceability of services. They also propose an IoT-based bacterial contamination traceability framework in the supply chain and a Bayesian causal network model to link the different layers of traceability.

An IoT-based system for food monitoring during transportation is presented in [29]. The system ensures a continuous remote monitoring and real-time collection of data which are sent automatically to the Cloud.

In [30] the authors proposed a food traceability system using Cyber-Physical System (CPS), Value Stream Mapping (VSM), EPCglobal and Fog computing. The solution proposed in this work aimed to improve the efficiency of the traceability system with the object of removing poor quality products from the supply chain using value-based processing.

Traditional platforms generally store their data in a distributed database, available for all actors of the system. These systems often present problems, such as: Security: access, control, authentication and authorizations are difficult to manage, heterogeneity of the protocols, data inconsistency, response delays, lack of visibility and limited communication with other partners. Current systems are characterized by the use of IoT and RFID. The latter is a promising solution to meet the requirements of real-time traceability systems. However, many issues need to be addressed, such as:

- On a large scale, a large amount of data relating to objects is generated from the various sensors, which makes intricate the processing and analysis of this information.
- The IoT generates a large volume of data that must be processed and analyzed at the cloud level, far from users and connected objects

3. Application Context

The deficiencies observed in supply chains and transportation systems on a global scale have a negative economic, environmental and social impact: ever-increasing freight transportation costs, greater number of accidents, pollution, poor time management as well as the deterioration of the transporters working conditions. Setting up Physical Internet (PI) would address these shortcomings. However, there are still no universal standards in global logistics regarding, for example, container dimensions or EDI (Electronic Data Interchange) messages. Nevertheless, PI requires some rules which, if specified and generalized, should become acceptable standards in the future. For the purpose of this research, we consider the following context.

3.1. Physical Internet (PI)

PI allows delivering, producing, moving, storing, and using physical objects around the world in an economically, environmentally, and socially efficient manner, and ensures universal logistics interconnection at the physical, informational and operational level. For that, it is based on three key elements that have the purpose of carrying out these tasks: PI-containers, PI-Protocols and PI-hubs. There are three types of PI-containers [31]: 1) Transport containers (T-containers): are the entities transported by the different types of vehicles (trucks, trains, ships, etc.). A T-container can contain several H-containers. Their dimensions are examined in order to optimize their filling rate [32]; 2) Handling containers (H-containers): contain physical goods and/or PI-containers of smaller sizes. Their size is modular and allows them to fit in a T-container; and 3) Packaging containers (P-containers): are used to directly contain physical goods. They are sized to adapt in a modular way to H-containers.

3.2. PI-Hubs

PI-hubs are the routing centers responsible for receiving, storing, and sending PI-containers. They adopt the same function as that of digital internet routers by ensuring

that each received PI-container is routed to the next destination on time and correctly. PI-hubs allow easy transfer of PI-containers between different transportation modes. PI-hubs may be of input/output or transit type. The input/output PI-hubs are nodes that allow suppliers to put their products in the chain or from which customers receive their ordered products. As for transit PI-hubs; they only allow the transit of products from PI-hub to PI-hub to their destinations. PI infrastructure is made up of a set of interconnected PI-hubs. Interconnection is one of the most important characteristics of PI. Its purpose is to make PI network open, global, efficient and sustainable, while being flexible, i.e. it allows modification if a new organization is added to the network.

3.3. PI-Protocols

Like the TCP/IP protocols of the digital Internet, the PI-protocols are standard rules allowing the control and the management of the operations of the physical Internet network. By analogy to the Open Systems Interconnection (OSI) model used in the TCP/IP protocols of the digital internet, the Open Logistics Interconnection (OLI) model was introduced. The OLI model consists of the following seven layers: physical, link, network, routing, shipping, encapsulation and web logistics [33].

3.4. Identification System

In automatic identification technologies, several solutions are used such as barcodes, RFID tags (Radio Frequency Identification), matrix codes, etc. The RFID tag is the most commonly used solution today. Unlike the barcode which must be placed in the axis of a laser, reading the RFID tag only requires the presence of the tag in an electromagnetic area. RFID technology consists of two key elements: *1)* An RFID label (or tag): includes a chip for storage and calculation, and an antenna for communication; and 2) An RFID reader, a device that reads RFID tags via radio signals. PI-containers are equipped with smart RFID tags. The latter can have various storage capacities, up to several megabytes. Their contents can be modified endlessly. They are also readable from a distance and several tags can be read at the same time.

3.5. Product Nature

We consider the food and pharmaceutical products. The latter are characterized by their sensitivity to climatic conditions and by an expiry date defined at the time of their manufacture. Thus, the temperature and the humidity rate parameters are controlled. These parameters must be captured each time the RFID is read and recorded in the event generated following this reading.

4. Ontological Approach to Product Traceability

Our approach to develop a traceability system in interconnected supply chains which network structure extends on a global scale is an ontology-based. To do this we have considered the following concepts: PI as it adapts to the requirements of

interconnected supply chains; IoT and RFID, a promising solution to meet the requirements of logistics domain; The OWL DL language adopted for data representation, allowing efficient storage, processing and especially sharing and reasoning on the manipulated data; Fog computing in order to share data and processing, to reduce the cloud overload, and to improve latency as the connected objects are closer to Fog than to Cloud.

At each level (Fog and Cloud), an ontology is developed and used:

- At each Fog and an ontology is created (Onto-Fog) which represents and manages the events inside a fog (intra Fog). This ontology will be used to represent the paths taken by the products inside a fog; it provides local traceability.
- At the Cloud level, an ontology is also created (Onto-Cloud) which manages the events circulating between the fogs (inter Fog). This ontology represents a directory of the fogs with their PI-hubs, as well as the events generated by the RFIDs reads of the products whose destination is outside the fog in which they are located. This ontology will allow finding the global traceability of the products from the different local traceability of the product.

Each fog has a local fog traceability system that exploits the Onto-Fog ontology specific to this Fog while the cloud has a global traceability system which uses the Onto-Cloud ontology. The latter represents the meta-knowledge that allows forming the global path crossed by a product, from its local paths in the different fogs by which the product is routed. Figure 1 depicts the functional architecture of the system that integrates the ontologies.

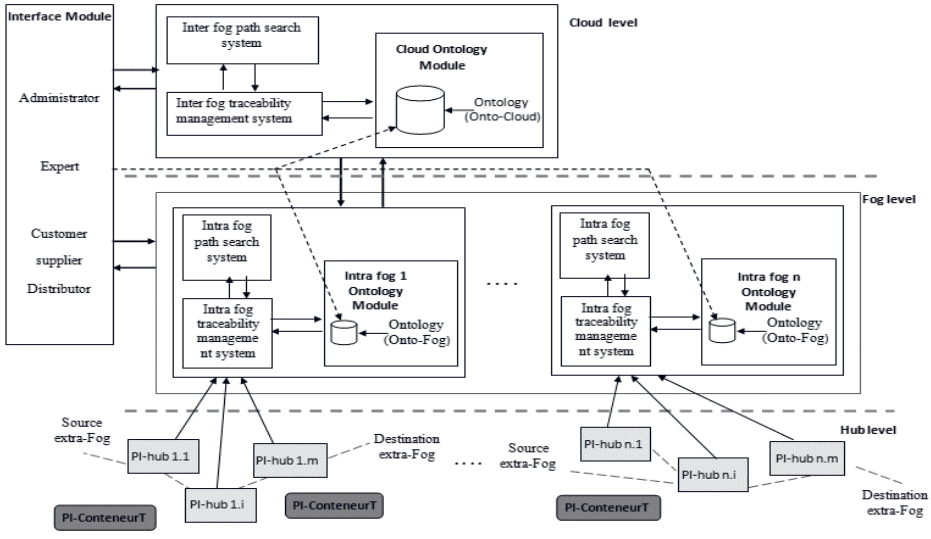


Figure 1. The system architecture.

To construct the two ontologies, we adopted the widely used "METHONTOLOGY" method [11]. The method consists of five phases in order to achieve the operationalization of the ontology.

4.1. Ontology Specification

This step is essential for the construction of the ontology. It consists in organizing and structuring the domain knowledge, particularly, extracting the concepts, their attributes, the relationships between concepts as well as the instances of the concepts from the documentation of the logistics domain, cloud and fog computing, as well as that of the concepts of the Internet in logistics. At the end of this step, we obtain the conceptual model at a fog level (Onto-fog) (Figure 2) and that at the cloud level (Onto-cloud) (Figure3).

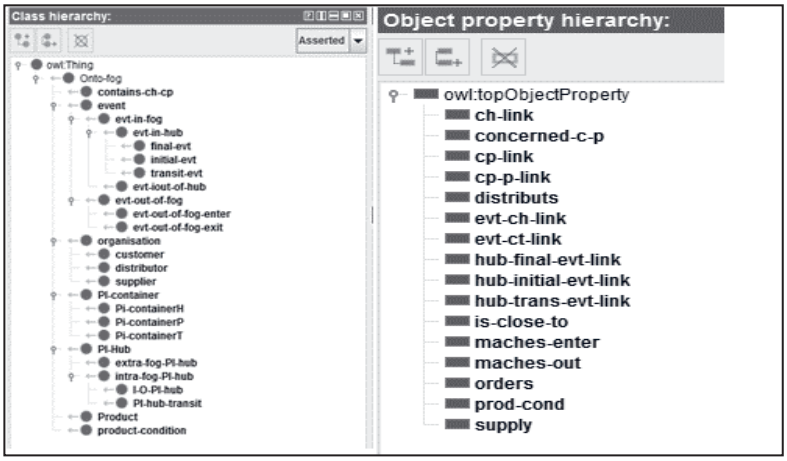


Figure 2. Onto-Fog ontology class hierarchy

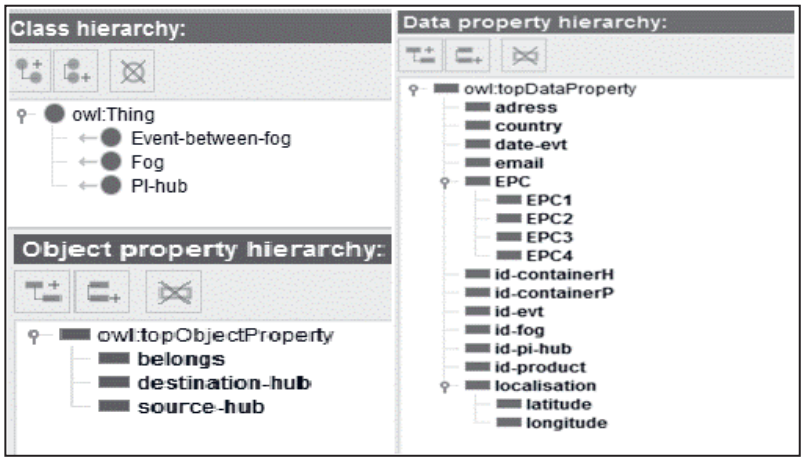


Figure 3. Onto-Cloud ontology class hierarchy.

4.1.1. The Onto-Fog Ontology:

- Concept extraction: the following main concepts are extracted: PI-container, PI-Hub, Event, Product, and Supplier. The PI-container concept includes the

following three types T-container, H-container and P-container. The PI-Hub concept includes the PI-Hub-extra-fog and PI-Hub-intra-fog. The latter includes the I/O PI-Hub and the transit-PI-Hubs.

- *Attribute Extraction:* Each extracted concept is characterized by its attributes. For example, the product concept is characterized by its identifier (an EPC – Electronic product code – which uses RFID technology. It is made up of four parts: 1) The header specifying the EPC format used by the tag; 2) A unique number assigned to the manufacturer of the product; 3) The product class identifier; and 4) The serial number assigned by the manufacturer to each product. A PI-hub has as an identifier, a designation, an address, as well as a capacity. As for the event concept, since each time an RFID tag is read, an event is generated. Its attributes: an identifier, designation, date and time of the read.
- *Concept relations extraction:* we were able to bring out the relations between the concepts. For example, as RFID tag reads can be made for products located in either T-container or H-container, each event generated following an RFID read is associated with 1 or 0 T-container. On the other hand, each generated event corresponds to at least one H-container.

4.1.2. The Onto-Cloud Ontology:

This ontology represents meta-knowledge relatively to onto-fog. It mainly consists of a directory of PI-hubs related to the fogs to which they belong. It represents also events relating to products which rout through more than one fog. These events are sent by the fog source of this product. The cloud in turn sends an event to the destination fog, informing it of the upcoming arrival of an extra fog product.

The "initial fog" and "initial PI-hub" information is also sent to the destination fog. Product condition information remains at the fog level. As a result, the event that is sent to the cloud only contains the identifiers of the product, and of the T-container, the H-container and the P-container in which the product is located.

A product that enters the supply chain is put in an I/O Pi-hub, which represents the initial PI-hub of this product. The PI-hubs, through which this product is routed, represent the transit PI-hubs for this product. The PI-hub, from which this product is delivered to the customer, represents the final PI-hub for this product. This way of modeling allows us to retrieve the product path from the initial PI-hub to the final PI-hub.

4.2. Ontology Formalization

We have chosen to formalize the ontology using OWL-DL language. OWL-DL is expressive enough to syntactically represent several formalisms. Moreover, it has a semantics that supports the expression of axioms. Semantic web tools such as reasoners and inference engines are also used to perform reasoning, which ensures consistency as well as inference of knowledge. Finally, we used the PROTEGE tool

for the creation of ontologies. Figure 4 and Figure 5 depict the two generated corresponding OWL files.

```
</owl:DatatypeProperty>
<owl:TransitiveProperty rdf:ID="is-close-to">
  <rdfs:domain rdf:resource="#contains-ch-cp"/>
  <rdfs:range rdf:resource="#contains-ch-cp"/>
  <owl:inverseOf rdf:resource="#is-close-to"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#SymmetricProperty"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</owl:TransitiveProperty>
<owl:FunctionalProperty rdf:ID="ch-link">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  <rdfs:range rdf:resource="#Pi-containerH"/>
  <rdfs:domain rdf:resource="#contains-ch-cp"/>
</owl:FunctionalProperty>
<owl:FunctionalProperty rdf:ID="concerned-c-p">
  <rdfs:domain rdf:resource="#product-condition"/>
  <rdfs:range rdf:resource="#Product"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</owl:FunctionalProperty>
<owl:FunctionalProperty rdf:ID="cp-p-link">
```

Figure 4. An Excerpt of the Onto-Fog ontology OWL file

```
<owl:Ontology rdf:about=""/>
<owl:Class rdf:ID="Event-between-fog"/>
<owl:Class rdf:ID="PI-hub"/>
<owl:Class rdf:ID="Fog"/>
<owl:DatatypeProperty rdf:ID="C-EPC">
  <rdfs:domain rdf:resource="#Event-between-fog"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="C-tieme-event">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#time"/>
  <rdfs:domain rdf:resource="#Event-between-fog"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="C-id-product">
  <rdfs:domain rdf:resource="#Event-between-fog"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="C-EPC2">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  <rdfs:subPropertyOf rdf:resource="#C-EPC"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="C-id-containerT">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
  <rdfs:domain rdf:resource="#Event-between-fog"/>
```

Figure 5. An Excerpt of the Onto-Cloud Ontology OWL file

4.3. Ontology-Based Reasoning

The PI-containers are modular and standardized. They can also be assembled and disassembled. This facilitates their handling (loading, unloading) as well as saving space in transportation containers. In this context, we considered that during any loading or unloading, the H-containers can be assembled or disassembled. In order to keep track of this assembly, we have provided in the ontology the relation “is-close-to” between the instances of the class “contains-ch-cp”. Each instance of the "contains-ch-cp" class represents an H-container at a given time (Fig. 6a and Fig. 6b). This container is constituted of P-containers, each of which contains a product. The relation (object property) “is-close-to” represents in the ontology the link between two adjacent H-containers. This link can be detected following the container RFID read. We defined this object property as transitive. Therefore, running the Pellet reasoner infers all "is-close-to" relationships between all H-containers that are assembled.

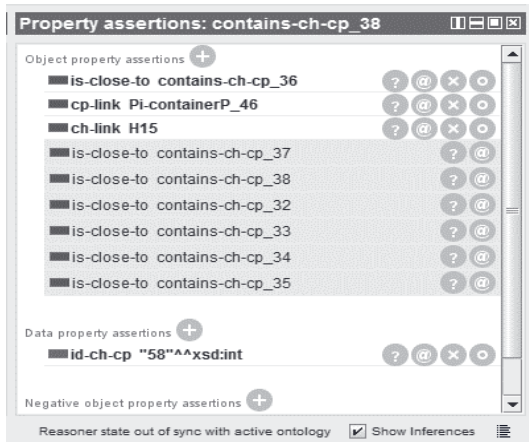


Figure 6a. 1st instance inferred of the class “contains-ch-cp” related to the object property “is-close-to”

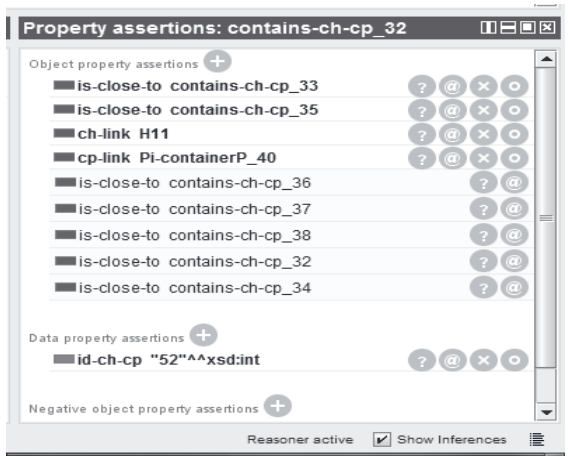


Figure 6b. 2nd instance infered of the class “contains-ch-cp” related to the object property “is-close-to”

The figures 6a and 6b show an example of this reasoning: we considered 7 assembled H-containers; they correspond to instances 32, 33, 34, 35, 36, 37 and 38 of the “contains-ch-cp” class. When reading RFID, each PI-container only detects its neighbor according its size and position in the container. In the example, a container sometimes detects 2, 3 or 4 neighbors. After the executing the Pellet reasoner, each instance is linked to all the instances (the seven) which form an H-container by assembly.

5. System Evaluation

Evaluation requires populating the ontology with real data then applying SPARQL queries. Therefore, to evaluate the ontology, we considered a fairly representative

population of individuals upon which we proceeded to the execution of some SPARQL queries related to product traceability, condition and location in order to test the use cases we specified. The results returned by these queries are used to evaluate the ontology.

5.1. Traceability Related Queries

The traceability of a product P consists in searching whether the product has entered the fog local supply chain. In this case, a corresponding event must be retrieved in the "initial-evt" class. Otherwise, we search in the extra fog events, i.e. in the class "evt-out-of-fog-enter". If an event corresponding to this product is retrieved then we to search for the rest of the paths from the other events generated by the product reads. Below some query examples:

5.1.1. Product Trace Related Queries

Q1: Give the path (the PI-hubs and date of passage) taken by the product with id-p=2?

In response to this query, the product P with id = 2 is deposited and delivered at the same fog. His path is completely local as depicted in Figure 7.

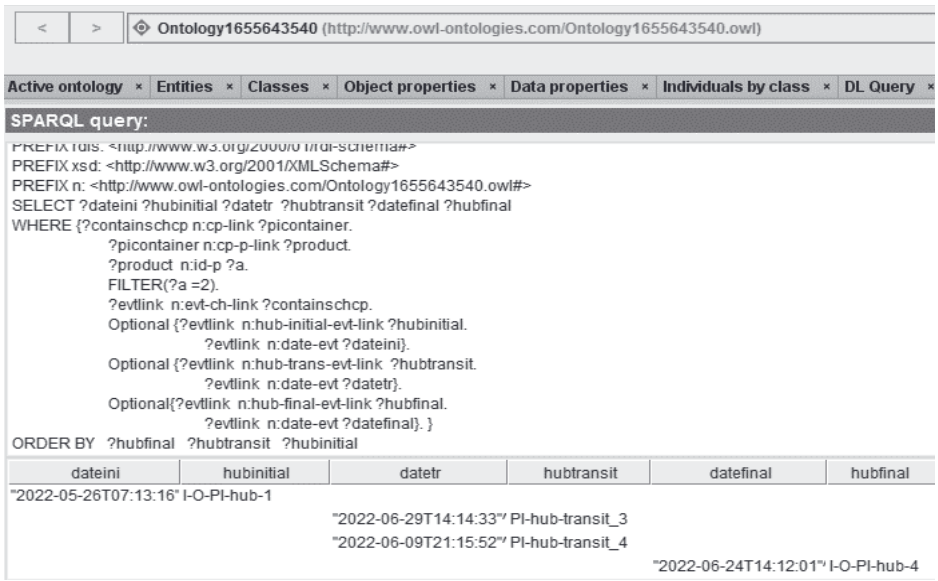
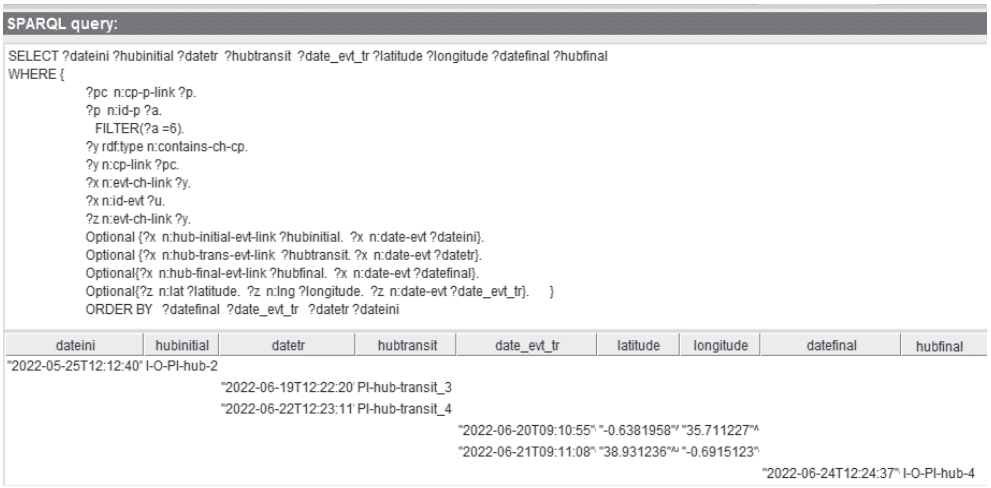


Figure 7. The path taken by the product P with id = 2

Q2: Give the trace of product with id-p = 6?

We consider here that the hypothesis that the RFID tags can be read in real time. Therefore, events outside the PI-hubs are used and represented in the ontology. The RFID tag of the P6 product (id-p = 6) is read in the hubs through which this product is routed and also outside the PI-hubs. In response to this query, the complete trace of product 6 is given in Figure 8.



SPARQL query:

```
SELECT ?dateini ?hubinitial ?datetr  ?hubtransit  ?date_evt_tr ?latitude ?longitude ?datefinal ?hubfinal
WHERE {
        ?pc  n:cp-p-link ?p.
        ?p  n:id-p ?a.
          FILTER(?a =6).
        ?y rdf:type n:contains-ch-cp.
        ?y n:cp-link ?pc.
        ?x n:evt-ch-link ?y.
        ?x n:id-evt ?u.
        ?z n:evt-ch-link ?y.
        Optional {?x  n:hub-initial-evt-link ?hubinitial.  ?x  n:date-evt ?dateini}.
        Optional {?x  n:hub-trans-evt-link  ?hubtransit. ?x  n:date-evt ?datetr}.
        Optional{?x  n:hub-final-evt-link ?hubfinal.  ?x  n:date-evt ?datefinal}.
        Optional{?z  n:lat ?latitude.  ?z  n:lng ?longitude.  ?z  n:date-evt ?date_evt_tr}.    }
        ORDER BY   ?datefinal  ?date_evt_tr   ?datetr ?dateini
```

dateini	hubinitial	datetr	hubtransit	date_evt_tr	latitude	longitude	datefinal	hubfinal
"2022-05-25T12:12:40"	I-O-PI-hub-2							
		"2022-06-19T12:22:20	PI-hub-transit_3					
		"2022-06-22T12:23:11	PI-hub-transit_4					
				"2022-06-20T09:10:55"^	"-0.6381958"^	"35.711227"^		
				"2022-06-21T09:11:08"^	"38.931236"^	"-0.6915123"^		
							"2022-06-24T12:24:37"^	I-O-PI-hub-4

Figure 8. The trace of the product P with id = 6

5.1.2. Product Condition Related Queries

Q3: What is the condition (temperature and humidity) of the product P with id = 6?

As depicted in Figure 9, we note that the condition of this product has deteriorated during its delivery. Indeed, its temperature exceeded the maximum temperature allowed during the 2nd event dated 06/19/2022. Also, the humidity level exceeded the maximum level indicated for this product during the 4th event on 06/21/2022.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX n: <http://www.owl-ontologies.com/Ontology1655643540.owl#>
SELECT ?num_event  ?date_event ?t_min ?t_max   ?t ?h_min   ?h_max   ?h
WHERE {    ?pc  n:cp-p-link ?p.
            ?p  n:id-p ?a.
              FILTER(?a =6).
            ?p  n:temp-min?t_min.
            ?p  n:temp-max ?t_max.
           ?p  n:hum-max ?h_max.
           ?p  n:hum-min ?h_min.
           ?y rdf:type n:contains-ch-cp.  ?y n:cp-link ?pc.
           ?y n:prod-cond ?x.              ?x n:concerned-c-p ?p.
           ?x n:hum ?h. ?x n:temp ?t.     ?z  n:evt-ch-link ?y.
           ?z n:id-evt  ?num_event. ?z n:date-evt ?date_event.}
           ORDER BY ?date_event
```

nu...	date_event	t_min	t_max	t	h_min	h_max	h
"2"^^...	"2022-05-25T12:12:40"	"6.0"^^...	"18.0"^^...	"10.0"^^...	"80"^^...	<hl "85"^^...	<http://v "82.0"^^...	<http://www.w3.org/2001/XMLSchema#float>
"4"^^...	"2022-06-19T12:22:20"	"6.0"^^...	"18.0"^^...	"20.0"^^...	"80"^^...	<hl "85"^^...	<http://v "81.0"^^...	<http://www.w3.org/2001/XMLSchema#float>
"31"^^...	"2022-06-20T09:10:55"	"6.0"^^...	"18.0"^^...	"9.0"^^...	"80"^^...	<hl "85"^^...	<http://v "81.0"^^...	<http://www.w3.org/2001/XMLSchema#float>
"35"^^...	"2022-06-21T09:11:08"	"6.0"^^...	"18.0"^^...	"15.0"^^...	"80"^^...	<hl "85"^^...	<http://v "90.0"^^...	<http://www.w3.org/2001/XMLSchema#float>
"38"^^...	"2022-06-22T12:23:11"	"6.0"^^...	"18.0"^^...	"15.0"^^...	"80"^^...	<hl "85"^^...	<http://v "90.0"^^...	<http://www.w3.org/2001/XMLSchema#float>
"40"^^...	"2022-06-24T12:24:37"	"6.0"^^...	"18.0"^^...	"13.0"^^...	"80"^^...	<hl "85"^^...	<http://v "84.0"^^...	<http://www.w3.org/2001/XMLSchema#float>

Figure 9. The condition of the product P with id = 6

Q4: Search for all expired products in the chain?

The list of all expired product is given in Figure 10.

SPARQL query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX n: <http://www.owl-ontologies.com/Ontology1655643540.owl
SELECT ?idprod ?perdate
WHERE {
                ?P  rdf:type n:Product.
               ?P   n:id-p    ?idprod.
              ?P   n:per-date ?perdate.
            FILTER (?perdate < "2022-06-25"^^xsd:date) }
```

idp...	perdate
"3"^^	"2022-06-15"^^<http://www.w3.org/2001/XMLSchema#date>
"4"^^	"2022-05-11"^^<http://www.w3.org/2001/XMLSchema#date>
"5"^^	"2022-06-22"^^<http://www.w3.org/2001/XMLSchema#date>
"19"^	"2022-06-06"^^<http://www.w3.org/2001/XMLSchema#date>

Figure 10. List of expired products

5.1.3. Location Related Queries

Q5: For each expired product, search its location in the chain?

This corresponds to the location of the last generated event that corresponds to this product. Figure 11 depicts the results for the latest hub of product P3 (id-p=3).

It is also possible to consider transit events (those generated between two successive PI-hubs in the path of a product. In this case, the query possibly returns the position of the event (latitude and longitude).

SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX n: <http://www.owl-ontologies.com/Ontology1655643540.owl#>
SELECT ?product ?date_event  ?hub
            WHERE {     ?pc  n:cp-p-link ?product.
                        ?product  n:id-p ?a.
                          FILTER(?a =3).
                        ?y rdf:type n:contains-ch-cp.
                        ?y n:cp-link ?pc.
                        ?x n:evt-ch-link ?y.
                        ?x n:date-evt ?date_event.
                        Optional {?x  n:hub-initial-evt-link ?hub. }.
                        Optional {?x  n:hub-trans-evt-link  ?hub.  }.
                        Optional{?x  n:hub-final-evt-link ?hub.  }.   }
ORDER BY DESC(?date_event) LIMIT 1
```

product	date_event	hub
P3	"2022-06-15T12:39:47"^	PI-hub-transit_6

Figure 11. The location of the product P with id = 3.

Q6: Search for products that have left the fog_1?

Products that leave the fog are those routed from a transit PI-hub to PI-hub in another fog. These products do not have end events in the current fog. In this case, the

system sends a message to the fog which consists of the output event of the product fog. This message allows the cloud to keep track of the product moving from fog to fog. The trace of a product in a fog is supported by the system at the fog level.

We note that the products id-p=13 and id-p=26 have the same release date, because they are in the same H-container. They also have the same transit PI-hub and the same destination fog (Figure 12).

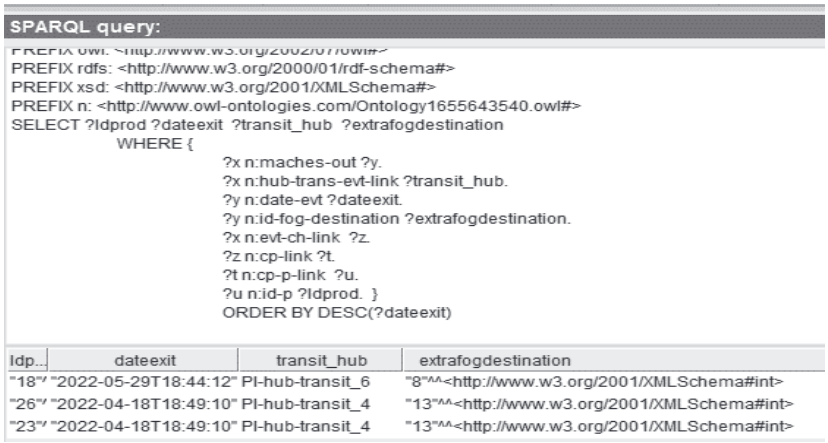


Figure 12. List of products that have left the fog_1

Q7: What is the effect on the onto-cloud following the exit of products from a fog_1?

Onto-cloud lists all the product events that circulate between the fogs. By searching in the Onto-Cloud for products that have left fog_1, we find their traces: the source fog and PI-hub as well as the destination fog and PI-hub of each product (Figure 13).

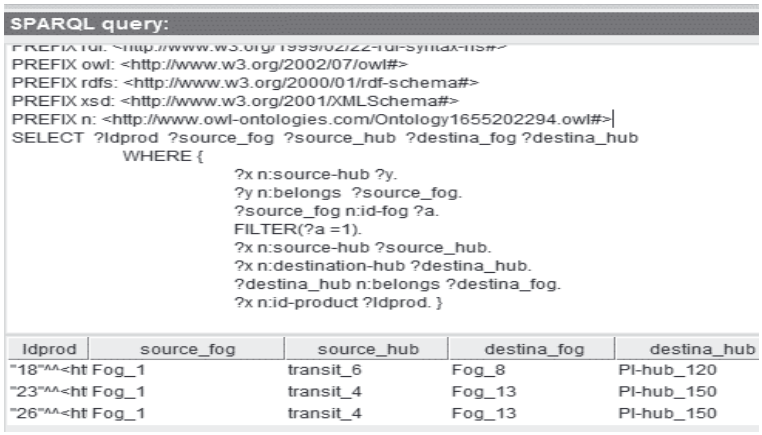


Figure 13. Traces of products that have left the fog_1

To find the global trace of a product, the cloud just needs to request its local trace from each fog through which this product is routed.

5.2. Discussion

The developed ontology completely satisfies the assigned objectives. In fact, we were able to carry out all the queries that the user can express whether he is a customer, a supplier or a hub manager. In order to simplify the understanding of the different queries, we have searched and displayed the products by the "id-p" attribute, but really the products are identified by the EPC which is one of the properties of the physical internet. Concerning the PI-hubs where the products are located, we visualized their URIs in place of displaying their identifiers, names as well as their addresses. In addition, we considered representing the transit events as the events are generated in real time. Also, sometimes, RFID tags are read during product transportation, for example on a boat which is not a PI-hub and moves (its position is variable). So, in this case, our ontology represents the event as a transit event (between two PI-hubs). The latter is characterized by a position (latitude, longitude) and not by a PI-hub.

6. Conclusion

SCM is inseparable from traceability which makes reliable the physical and information flows, accelerates the transmission of information on these flows, allows accessing a detailed knowledge of the movements, and makes the flows visible. In fact, the traceability of logistical objects allows identifying all the objects in the same flow of goods to know for each object its origin and its destination, its state as well as the different stages of its journey throughout the chain, from the manufacturer to the final consumer. However, despite technological advances in logistics support the need for identification and traceability of logistics objects remains a challenge. Indeed, there are more and more objects to identify, and with the advent of connected objects, there is more and more data collected on these objects. These data pose a storage problem because of the large volume of data generated, exchange and storage formats and communication protocols.

It is essential to identify innovative approaches and solutions in logistics and transport of goods, storage platforms, traceability technologies, and communication protocols for identification, traceability and monitoring of logistics objects in different industrial fields. The challenge therefore is to design a system that will be able to ensure and support all logistics operations related to physical objects worldwide in an efficient and sustainable manner. We believe that the physical Internet is one such innovative solution.

In this article, we proposed a physical Internet logistics traceability system based on the Internet of Things, Cloud and Fog computing to improve the visibility of the supply chain and meet the needs of supply chain actors in terms of the condition, quality and location of their goods throughout the supply chain. The proposed system takes advantage of the supply networks interconnected on a global scale through a standardized set of collaboration protocols, modular containers and intelligent interfaces for greater efficiency and sustainability.

We adopted an ontology approach to ensure the traceability of the products by representing their traces in terms of paths formed by the different PI-hubs, and all the T-containers and H-containers having contained the product at one time or another.

In future, we plan to extend our work with a third ontology at the PI-hub level (onto-Hub). This ontology will have the role of representing the trace of the products in terms of used transportation means as well as the drivers (in the case of trucks or vehicles).

References

[1] C,F. Durach, J. Kembro and A. Wieland, A new paradigm for systematic literature reviews in supply chain management", *Journal of Supply Chain Management*, Vol. 53 No. 4, pp. 67-85, 2017.

[2] C.S. Tang and L.P. Veelenturf, The strategic role of logistics in the industry 4.0 era. *Transp. Res. Part E Logist. Transp*. Rev. 129, 1–11, 2019.

[3] M. Christopher, Logistics & Supply Chain Management. Harlow,2016.

[4] H. Ringsberg, Perspectives on food traceability: A systematic literature review. Supply Chain. Manag. Int. J. 19, 558–576, 2014.

[5] P. Olsen and M. Borit, The components of a food traceability system. Trends Food Sci. Technol, 77, 143–149, 2018.

[6] GS1., Business Process and System Requirements for Full Chain Traceability., 2009.

[7] F. Dabbene and P. Gay, *Food traceability systems: Performance evaluation and optimization.* Computers and Electronics in Agriculture, 75, 139-146, 2011.

[8] D. K. Mishra, S. Henry, A. Sekhari and Y. Ouzrout. Traceability as an integral part of supply chain logistics management: an analytical review. 7th International Conference on Logistics and Transport (ICLT), 2015.

[9] S. Garcia-Torres, L. Albareda, M. Rey-Garcia and S. Seuring, Traceability for Sustainability—Literature review and conceptual framework. SCM, 24, 85–106, 2019.

[10] S.C. Onar and A. Ustundag, Smart and connected product business models. In Industry 4.0: Managing The Digital Transformation (pp. 25–41). Springer, Cham, 2018.

[11] S. Naskar, P. Basu, and A. K. Sen. A Literature Review of the Emerging Field of IoT Using RFID and Its Applications in Supply Chain Management, In The Internet of Things in the Modern Business Environment Advances in E-Business Research, 1–27. Hershey, PA: IGI Global, 2017.

[12] A. Botta, W. de Donato, V. Persico, and A. Pescape, Integration of cloud computing and internet of things: A survey, Future Generation Computer Systems, vol. 56, pp. 684 – 700, 2016.

[13] P. Hu, S. Dhelim, H. Ning, and T. Qiu, Survey on fog computing: architecture, key technologies, applications and open issues, Journal of Network and Computer Applications, vol. 98, pp. 27 – 42, 2017.

[14] S. Kumperš˘cak, M. Medved, M. Terglav and A. Wrzalik, Obrecht, M. Traceability systems and technologies for better food supply chain management. Qual. Prod. Improv.-QPI, 1, 567–574, 2019.

[15] G. Baryannis, S. Validi, S. Dani, and G. Antoniou. Supply chain risk management and artificial intelligence: State of the art and future research directions, *International Journal of Production Research* 57 (7): 2179-2202, 2019.

[16] G. Tortorella, L., Miorando, R., and G. Marodin., Lean supply chain management: Empirical research on practices, contexts and performance, *International Journal of Production Economics* 193: 98–112, 2017.

[17] R. V. Chen, Intelligent IoT-Enabled System in Green Supply Chain Using Integrated FCM Method, International Journal of Business Analytics 2 (3): 47–66, 2015.

[18] P. Ray, A survey on internet of things architectures, Journal of King Saud University - Computer and Information Sciences, vol. 30, no. 3, pp. 291 – 319, 2018.

[19] S. Greengard, The Internet of Things. Cambridge, MA: MIT Press, 2015.

[20] Li, S., L. D. Xu, and S. Zhao. The Internet of Things: A Survey, Information Systems Frontiers 17 (2): 243–259, 2015.

[21] I. Da Xu, W. He, and S. Li., Internet of things in industries: A survey, *IEEE Transactions on Industrial Informatics* 10 (4): 2233–2243, 2014.

[22] L. Novais, J. M. Maqueira, and A. Ortiz-Bas, A systematic literature review of cloud computing use in supply chain integration, *Computers & Industrial Engineering* 129: 296–314, 2019.

[23] G. Niharika and V. Ritu, Cloud Architecture for the logistics business, Procedia Computer Science, vol. 50, pp. 414 – 420, 2015.

[24] T. Nguyen Gia, A. M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, Fog computing approach for mobility support in internet-of-things systems, IEEE Access, vol. 6, pp. 36 064–36 082, 2018.

[25] M. T. Hompel, J. Rehof, and O. Wolf, Cloud Computing for Logistics. Springer Publishing Company, Incorporated, 2014.

[26] M. Christopher**,** Logistics and Supply Chain Management: Strategies for Reducing Cost and Improving Service, 2 Edition, Financial Times/Prentice Hall, London, 1998.

[27] A. Mousavi, M. Sarhadi, A. Lenk end S. Fawcett, Tracking and traceability in the meat processing industry : a solution», British Food Journal, vol. 104, n°1, p. 7–19, 2002.

[28] E. Abad, F. Palacio, M. Nuin, A. G. De Zarate, A. Juarros, J. M. Gomez and S. Marco, Rfid smart tag for traceability and cold chain monitoring of foods : Demonstration in an intercontinental fresh fish logistic chain, Journal of food engineering, vol. 93, no 4, p. 394–399.53, 2009.

[29] W. Zhou and S. Piramuth, IoT and supply chain traceability, In Future Network Systems and Security - First International Conference, FNSS 2015, Paris, France, June 11-13, 2015, Proceedings, Communications in Computer and Information Science, vol. 523, Springer, p. 156–165, 2015

[30] M. Maksimovic, V. Vujovic and E. Omanovic-Miklicanin, : A low cost internet of things solution for traceability and monitoring food safety during transportation.», dans Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment, Kavala, Greece, September 17-20, 2015., CEUR Workshop Proceedings, vol. 1498, p. 583–593, 2015.

[31] R.Y. Chen, An intelligent value stream-based approach to collaboration of food traceability cyber physical system by fog computing», Food Control, vol. 71, p. 124–136, 2017.

[32] Y. Sallez B. Montreuil and Ballot, On the Activeness of Physical Internet Containers, Service Orientation in Holonic and Multi-agent Manufacturing, Springer International, Vol.594, 259-269, 2015.

[33] J. H. Dunning and S.M. Lundan, Institutions and the OLI paradigm of the multinational enterprise. *Asia Pacific Journal of Management,* 25(4): 573-593, 2008.

[34] W. Roy, An introduction to RFID technology. *Pervasive Computing, IEEE*. 5. 25 - 33. 2006.

10.31341/jios.48.1.6 UDC 001.891:005.53:330.322
Open Access Original Scientific Paper

Behavioral Biases in Investment Decisions: An Extensive Literature Review and Pathways for Future Research

N. Sathya *sathya.n@vit.ac.in*
Department of commerce,
School of social science and languages,
Vellore Institute of Technology, Vellore, India.

R. Gayathiri *gayathirr@auxiliumcollege.edu.in*
Department of commerce (Banking &Insurance),
Auxilium College (Autonomous), Vellore, India

Abstract

In the realm of investment decisions, the influence of behavioral biases has emerged as a captivating area of exploration. This article embarks on a comprehensive journey through the landscape of behavioral biases in investment choices, delving into their profound impact on financial markets. Contrary to traditional finance theories assuming rationality, a multitude of empirical evidence attests to the pervasive effects of cognitive and emotional biases. Through an extensive literature review, this article elucidates the intricacies of key biases such as overconfidence, loss aversion, anchoring, confirmation bias, herding behavior, disposition effect, framing effects, and regret aversion. By examining the distinct ways these biases distort investors' judgment and decision-making processes, we unveil the often unexpected deviations from rationality. Each bias, rooted in human psychology, can lead to suboptimal investment behaviors, portfolio misalignments, and heightened market volatility. However, recognizing the impact of these biases provides opportunities for transformative insights. As investment professionals, policymakers, and individuals alike comprehend the subtle nuances of behavioral biases, tailored interventions, educational initiatives, and adaptive strategies can be devised to mitigate their adverse effects. This article not only synthesizes the prevailing research but also charts a course for future investigations. The implications of understanding and addressing behavioral biases extend beyond financial realms, offering a bridge between finance and psychology. As interdisciplinary collaboration gains momentum, pathways for future research become evident, beckoning scholars to delve deeper into the uncharted territories of human behavior and its intricate relationship with investment decisions. Through the exploration of these biases and their potential remedies, this article illuminates the evolving landscape of investment decision-making in a world where cognitive fallacies intersect with financial choicest.

Keywords: Overconfidence, Herding bias, Behavioral biases, Investment decisions, Bibliometric analysis, Systematic literature review, content analysis

1. Introduction

Anteriorly, researchers followed traditional finance theories such as the efficient market In the dynamic landscape of investment decisions, the traditional assumptions of rationality and efficiency have given way to a more nuanced understanding of human behavior. The field of behavioral finance has emerged as a powerful lens through which we can comprehend the intricate interplay between cognitive biases and investment choices. This introductory section sets the stage for an extensive exploration into the realm of behavioral biases and their profound implications for financial markets [1-6].

Classical finance theories have long posited that investors make decisions based on careful analysis, rational expectations, and the pursuit of self-interest. However, real-world observations often diverge from these theoretical constructs. Empirical evidence demonstrates that investors frequently exhibit behaviors influenced by psychological and emotional factors, leading to patterns of decision-making that deviate from rationality [6-8]. This realization has led to a paradigm shift, with behavioral finance gaining prominence as an essential framework for understanding investment phenomena.

The allure of behavioral biases lies in their ability to explain the discrepancies between economic theories and market outcomes. These biases, deeply rooted in human psychology, illuminate the cognitive shortcuts, emotional responses, and heuristic-driven judgments that shape investors' choices [9-16]. From the tendency to overestimate one's own knowledge and underestimate risks, to the aversion to losses that outweighs the desire for gains, these biases encapsulate a range of behaviors that significantly impact investment strategies and portfolio performance.

This article embarks on a comprehensive journey to explore the landscape of behavioral biases in investment decisions. Through an extensive literature review, we aim to shed light on the nuanced intricacies of key biases that underlie investment behaviors [17-19]. The subsequent sections delve into biases such as overconfidence, loss aversion, anchoring, confirmation bias, herding behavior, disposition effect, framing effects, and regret aversion. Each bias, a manifestation of the complex interplay between cognitive and emotional aspects of human behavior, offers insights into the deviations from rationality observed in investment contexts [20-23].

By understanding and dissecting these biases, we not only gain a deeper appreciation for the intricacies of human decision-making but also open avenues for practical applications. The implications of behavioral biases extend beyond theoretical discourse, offering opportunities for designing interventions, educational programs, and innovative investment strategies that account for and mitigate these inherent cognitive fallacies. Furthermore, recognizing the pervasive effects of behavioral biases lays the groundwork for a more holistic approach to investment decision-making, one that bridges the gap between financial theory and the complex reality of human behavior. As the realm of behavioral finance continues to evolve, it beckons researchers, investors, and policymakers to collaborate in uncovering the multifaceted layers of human decision-making. This article sets out not only to synthesize the existing body of research but also to chart a trajectory for future

investigations. The path forward lies in interdisciplinary collaborations that intertwine finance, psychology, and neuroeconomics, as we seek to unravel the intricate threads that weave together our cognitive biases and financial choices. Through this exploration, we embark on a journey to understand how behavioral biases shape investment decisions in an ever-evolving world of finance.

2. Theoretical background of the behavioral biases on investment decisions

Behavioral biases in investment decisions stem from the intersection of cognitive psychology, behavioral economics, and finance. Traditional finance theories, based on rational choice models, assume that individuals make decisions in a calculated, objective, and self-interested manner. However, empirical observations have consistently revealed that human decision-making is often influenced by psychological and emotional factors, leading to deviations from rationality. This theoretical background lays the foundation for understanding the behavioral biases that impact investment choices.

Prospect Theory:

One of the seminal theories in behavioral finance is Prospect Theory, developed by Kahneman and Tversky [24] in 1979. This theory challenges the traditional notion of utility maximization and introduces the concepts of loss aversion and diminishing sensitivity. According to Prospect Theory, individuals perceive gains and losses relative to a reference point (usually their current status), and they experience the pain of losses more intensely than the pleasure of equivalent gains. This psychological phenomenon leads to risk-averse behavior when faced with potential losses and risk-seeking behavior when faced with potential gains.

Heuristics and Biases:

The work of Kahneman and Tversky also introduced the concept of heuristics, which are mental shortcuts individuals use to simplify complex decisions. While heuristics can be efficient, they can also lead to systematic errors or biases in judgment. For example, the availability heuristic involves making judgments based on readily available information, leading investors to overweight recent or easily accessible data when making investment decisions.

Anchoring and Adjustment:

Anchoring and adjustment is another cognitive bias that plays a significant role in investment choices. This bias occurs when individuals rely too heavily on the first piece of information encountered (the anchor) when making subsequent decisions. Investors anchored to a specific stock price or market level may inadequately adjust their valuations based on new information, leading to mispriced investments.

Herding Behavior:

Overconfidence, a cognitive bias, manifests when individuals have an inflated belief in their own abilities, knowledge, and predictive accuracy. These bias leads investors to overestimate their understanding of financial markets, often assuming they possess superior information that gives them an edge. Overconfident investors tend to engage in excessive trading, frequently buying and selling based on the belief

that they can outperform the market. The Dunning-Kruger effect, a cognitive phenomenon, illustrates how individuals with limited knowledge and expertise tend to overestimate their competence. In the context of investing, this can result in novices making bold investment decisions without fully comprehending the risks. Overconfident investors often exhibit a heightened willingness to take on speculative investments and concentrate their portfolios, leading to suboptimal diversification. Research by Barber and Odean[25] highlighted that overconfident traders tend to trade more frequently, incurring higher transaction costs and realizing lower net returns. Glaser and Weber [26] emphasized that overconfidence can lead to under-diversification, exposing investors to higher levels of uncompensated risk. These findings underscore the significance of understanding and mitigating the overconfidence bias in investment decisions.

Disposition Effect:

The disposition effect is the tendency to sell winning investments too early and hold onto losing investments too long. This bias is rooted in the desire to avoid regret and to validate one's decisions. The disposition effect can lead to suboptimal portfolio performance due to the uneven distribution of gains and losses.

Framing Effects:

Framing effects demonstrate that the way information is presented can significantly influence decision-making. Individuals often react differently to identical choices based on how they are framed. This bias can impact investment decisions based on how information about potential gains and losses is presented.

Confirmation Bias:

Confirmation bias is the tendency to seek out and interpret information in a way that confirms preexisting beliefs. Investors may only consider information that aligns with their views, leading to distorted perceptions of the market and potentially risky investment decisions.

Regret Aversion:

Regret aversion stems from the fear of making decisions that will lead to subsequent regret. Investors may avoid taking necessary risks to prevent the possibility of feeling regret, which can hinder the pursuit of optimal investment strategies. Therefore, the theoretical foundation of behavioral biases in investment decisions rests on the understanding that human decision-making is influenced by cognitive shortcuts, emotional responses, and psychological factors. These biases deviate from the rationality assumed by traditional finance theories and provide a rich field for exploring the complexities of investment behavior. Recognizing and addressing these biases is essential for improving decision-making in the financial world and shaping the future of investment strategies and practices.

3. Literature Review

Numerous studies have delved into the realm of behavioral biases and their impact on investment decisions[27-30]. These studies have collectively highlighted the intricate ways in which cognitive and emotional factors deviate investors from rational choices. Here, we synthesize key findings from a selection of prominent research in this field.

Overconfidence Bias:

Barber and Odean (2001) found that overconfident investors tend to trade more frequently, leading to lower returns due to higher transaction costs and poor timing. Similarly, Glaser and Weber (2007) demonstrated that overconfident investors are more likely to under-diversify their portfolios, exposing themselves to greater risks.

Study	Key Findings
Barber and Odean (2001)	Overconfident investors tend to trade more frequently, leading to lower returns due to higher transaction costs and poor timing.
Glaser and Weber (2007)	Overconfident investors are more likely to under-diversify their portfolios, exposing themselves to greater risks.
Gervais and Odean (2001)	Overconfident traders exhibit excessive trading activity and poor performance compared to rational traders.
Statman and Caldwell (1988)	Overconfidence leads to suboptimal portfolio diversification and excessive trading, resulting in reduced returns.

Table 1. Key findings in overconfidence bias.

Loss Aversion and the Disposition Effect:

Odean (1998) observed the disposition effect, whereby investors hold onto losing stocks longer than winning stocks. This behavior can be attributed to the pain of realizing a loss and the desire to avoid regret. The disposition effect leads to suboptimal portfolio performance and uneven gains and losses. Table 2 shows the key findings in loss aversion.

Study	Key Findings
Odean (1998)	The disposition effect: investors hold onto losing stocks longer than winning stocks, leading to suboptimal portfolio performance and uneven gains and losses.
Weber et.al (2006)	Loss aversion contributes to the disposition effect, causing investors to avoid realizing losses and amplifying suboptimal decisions.
Shefrin and Statman (1985)	Loss aversion drives the disposition effect, leading to suboptimal selling decisions and reduced portfolio returns.

Table 2. Key findings in Loss aversion

Herding Behavior:

Bikhchandani and Sharma (2001) explored herding behavior in financial markets and found that investors often imitate the actions of others, especially during periods of uncertainty. This behavior can lead to market bubbles and abrupt corrections, as well as amplify market volatility. Table 3 shows the findings in herding behaviour [31-33].

Study	Key Findings
Bikhchandani and Sharma (2001)	Investors often imitate others, especially during uncertainty, leading to market bubbles, abrupt corrections, and increased volatility.
Scharfstein and Stein (1990)	Herding behavior can lead to market overreactions and exacerbate price movements, contributing to market inefficiencies.
Hong et.al (2020)	Informational herding occurs when investors follow the actions of others due to a lack of private information, contributing to market contagion.

Table 3. Key findings in Herding behaviour

Anchoring and Adjustment:

Rabin and Schrag (1999) investigated the anchoring bias and found that individuals anchor their decisions to irrelevant information, impacting their judgments and choices. In investment, anchoring can result in undervaluing or overvaluing assets based on arbitrary reference points. Table 4 shows the key findings of anchoring and adjustment [34-36].

Study	Key Findings
Rabin and Schrag (1999)	Investors anchor decisions to irrelevant information, impacting judgments and choices. Anchoring can lead to undervaluing or overvaluing assets based on arbitrary reference points.
Ariely et al. (2003)	Anchoring biases influence investment decisions, causing investors to be overly influenced by initial price information, even when it's irrelevant.
Avery, C., & Zemsky, P (1998)	Anchoring can lead investors to perceive stocks as cheap or expensive based on prior reference points, impacting valuations and decisions.

Table 4. Key findings in anchoring and adjustment

Confirmation Bias:

DeBondt and Thaler (1985) conducted a seminal study on the phenomenon of the representativeness heuristic, where individuals make decisions based on stereotypes

or prototypes. This bias can lead investors to overweight recent events and ignore contradictory information. Table 5 detailed the key findings of confirmation bias [37-39].

Study	Key Findings
DeBondt and Thaler (1985)	The representativeness heuristic leads investors to overweight recent events and ignore contradictory information, impacting decision-making.
Grinblatt and Keloharju (2001)	Confirmation bias leads to overreaction to new information and subsequent reversals in stock prices, resulting in market inefficiencies.
Guo et.al (2020)	Confirmation bias influences financial experts and their forecasts, leading to an overestimation of the accuracy of their predictions.

Table 5: Confirmation bias

Framing Effects:

Tversky and Kahneman (1981) introduced framing effects, showing that the way information is presented can significantly impact decisions. Individuals tend to be risk-averse in situations framed as gains and risk-seeking in situations framed as losses. These bias influences investment choices based on how potential outcomes are framed [40-42]. Finding of framing effects is shown in table 6.

Study	Key Findings
Tversky and Kahneman (1981)	Framing effects influence risk preferences. Individuals tend to be risk-averse in gain frames and risk-seeking in loss frames, impacting investment choices.
Thaler et al. (1997)	Framing can impact investors' perception of risk and return, leading to different decisions based on the presentation of information.
Tranfield et.al (2003)	Framing effects lead to inconsistent decisions in investment scenarios, highlighting the influence of presentation on risk perception.

Table 6. Key finding of framing effects

Regret Aversion:

Bell (1982) introduced the concept of regret theory, which posits that individuals make decisions to minimize the potential for future regret. Investors may avoid taking necessary risks to prevent the possibility of feeling regret, impacting their overall portfolio performance. Table 7 shows the key findings of regret aversion [43-45].

Study	Key Findings
Bell (1982)	Regret theory suggests individuals make decisions to minimize future regret. Investors may avoid necessary risks to prevent feeling regret, impacting portfolio performance.
Zeelenberg and Pieters (2004)	Regret aversion leads to suboptimal investment decisions, as investors prioritize avoiding regret over making rational choices.
Wu et.al (2020)	Regret aversion contributes to the disposition effect, causing investors to hold onto losing stocks to avoid admitting mistakes.

Table 7. Key findings of Regret aversion

While the literature reviewed here provides valuable insights into behavioral biases in investment decisions, there remain several avenues for future research. Exploring the interplay between different biases, investigating cultural variations in bias susceptibility, and understanding how technological advancements impact bias expression are potential areas for further exploration. The literature underscores the critical role that behavioral biases play in shaping investment decisions. These biases, driven by cognitive and emotional factors, lead investors to deviate from rational choices, impacting portfolio performance, market dynamics, and the overall efficiency of financial markets. As the field of behavioral finance continues to evolve, deeper insights into these biases offer pathways to more informed investment strategies and improved decision-making practices.

4. Key Findings: Behavioral Biases in Investment Decisions

The synthesis of extensive research on behavioral biases in investment decisions has unearthed profound insights into the intricate ways human psychology intersects with financial choices. As investors navigate markets, cognitive and emotional biases exert a significant influence, leading to deviations from rational decision-making. The key findings shed light on the pivotal role these biases play in shaping investment behavior and market dynamics:

Overconfidence Bias:

- Overconfident investors trade more frequently, resulting in higher transaction costs and lower net returns.
- The Dunning-Kruger effect highlights that those with limited knowledge often overestimate their investment competence.
- Overconfident investors tend to concentrate their portfolios and engage in speculative trading, leading to suboptimal diversification.
- Glaser and Weber (2007) demonstrated that overconfident investors are prone to under-diversify their portfolios, thereby increasing their exposure to uncompensated risk.

Loss Aversion and the Disposition Effect:

- Investors exhibit the disposition effect by holding onto losing stocks longer than winning stocks, leading to uneven gains and losses.
- Loss aversion, the driving force behind the disposition effect, causes investors to experience the pain of losses more acutely than the pleasure of gains.
- The disposition effect contributes to suboptimal portfolio performance, as investors forego rational trading decisions due to regret avoidance.

Herding Behavior:

- Herding behavior is prevalent in uncertain environments, causing investors to imitate the actions of others.
- Herding leads to market bubbles fueled by excessive optimism and abrupt corrections driven by mass panic.
- Bikhchandani and Sharma (2001) revealed that herding behavior in financial markets amplifies price movements, contributing to market inefficiencies.
- Herding dynamics illustrate the complex interplay between individual actions and collective market outcomes.

Anchoring and Adjustment:

- Anchoring bias leads investors to anchor their decisions to irrelevant reference points, influencing valuations and judgments.
- Investors inadequately adjust from the anchor, causing them to undervalue or overvalue assets based on arbitrary information.
- Ariely et al. (2003) showcased how anchoring biases impact investment choices, distorting perceptions of value even when the anchor is irrelevant.

Confirmation Bias:

- Confirmation bias drives investors to seek and interpret information that aligns with preexisting beliefs, distorting their market perceptions.
- Investors disregard contradictory data, leading to imbalanced and potentially risky investment decisions.
- Grinblatt and Keloharju (2001) demonstrated that confirmation bias contributes to overreaction to new information and subsequent market reversals.

Framing Effects:

- Framing effects influence risk preferences based on how information is presented, leading to varied decisions in gain and loss scenarios.
- Individuals tend to be risk-averse in gain frames and risk-seeking in loss frames, impacting investment choices.
- Thaler et al. (1997) highlighted how framing can alter investors' perceptions of risk and return, influencing their decisions.

Regret Aversion:

- Regret aversion drives investors to avoid decisions that might lead to future regret, hindering necessary risk-taking.
- Zeelenberg and Pieters (2004) emphasized that regret aversion contributes to suboptimal investment decisions, impacting overall portfolio performance.

- Wu et.al (2020) showed that regret aversion reinforces the disposition effect, influencing investors to hold onto losing stocks.

These key findings underscore the intricate ways in which behavioral biases permeate investment decisions, shaping strategies, portfolio outcomes, and market dynamics. By recognizing and addressing these biases, investors and financial professionals can develop strategies to enhance decision-making and optimize outcomes.

5. Knowledge Gaps and Future Research:

While significant progress has been made in understanding behavioral biases, several knowledge gaps and avenues for future research remain:

Interaction of Biases: Research could delve deeper into how different biases interact and amplify each other's effects, leading to more complex investment decisions.

Cultural Variations: Investigate how cultural factors influence susceptibility to biases, considering that cultural norms and values may shape individuals' decision-making tendencies.

Technology and Biases: Examine how technological advancements, such as robo-advisors and algorithmic trading, impact the expression and mitigation of behavioral biases.

Intervention Strategies: Develop and test interventions, educational programs, and decision-making tools that help investors recognize and mitigate biases in real-time.

Long-Term Effects: Explore the long-term impact of behavioral biases on wealth accumulation and retirement planning, considering how biases might affect investment decisions over extended periods.

Behavioral Portfolio Theory: Further develop behavioral portfolio theories that incorporate various biases to create more accurate models of investor behavior.

Neuroeconomics and Biases: Integrate insights from neuroeconomics to better understand the neural mechanisms underlying biases and their potential neural interventions.

Practical Applications: Explore how financial professionals, such as financial advisors and portfolio managers, can leverage an understanding of behavioral biases to improve client outcomes.

while the existing literature has shed light on the profound influence of behavioral biases on investment decisions, there remain exciting areas of inquiry that can enhance our understanding and guide the development of effective interventions in the realm of finance. Addressing these knowledge gaps will contribute to more informed investment strategies and improved decision-making practices.

6. Conclusion and Directions for Future Research:

In the ever-evolving landscape of investment decisions, the exploration of behavioral biases has illuminated the intricate interplay between human psychology and financial choices. This review journeyed through the rich tapestry of biases such as overconfidence, loss aversion, herding behavior, anchoring, confirmation bias, framing effects, and regret aversion, uncovering their undeniable impact on investment behavior. As we conclude this endeavor, it is evident that behavioral biases disrupt the neat assumptions of rationality and efficiency that underpin traditional finance theories.

The key findings highlighted the pervasive nature of biases, from overconfident traders' excessive trading to investors' reluctance to part with losing stocks due to the disposition effect. Herding behavior was unveiled as a potent force driving market dynamics, while anchoring and confirmation biases showcased how initial perceptions and preconceived notions can shape investment decisions. Framing effects and regret aversion provided insights into the malleability of risk perceptions and the influence of future emotions on present choices.

However, this review also underscores the existence of uncharted territories within the realm of behavioral biases in investment decisions. These knowledge gaps beckon for future research to extend our understanding and offer actionable insights:

Multidimensional Bias Interaction: Investigate how different biases interact and influence each other in complex decision-making scenarios, shedding light on the combined effects of cognitive fallacies.

Cultural Nuances: Explore how cultural backgrounds influence susceptibility to biases, recognizing that diverse cultural norms may give rise to distinct behavioral tendencies.

Technology and Bias Expression: Examine the role of technology in facilitating or mitigating biases, considering how algorithms and digital platforms shape investment choices.

Intervention Strategies: Develop and assess tailored interventions that empower investors, financial professionals, and policymakers to recognize and address biases in real-world contexts.

Long-Term Biases: Investigate the long-term effects of biases on investment behavior over extended periods, considering their impact on wealth accumulation and retirement planning.

Behavioral Portfolio Theory: Advance behavioral portfolio theories that incorporate a nuanced understanding of biases, enabling more accurate modeling of investor behavior and decision-making.

Neuroeconomic Insights: Collaborate with neuroeconomics to unravel the neural underpinnings of biases and explore potential neural interventions to mitigate their influence.

Applied Knowledge: Bridge the gap between research and practice by exploring how financial professionals can harness insights from behavioral biases to enhance client outcomes and optimize investment strategies.

In this ever-evolving field, the exploration of behavioral biases offers a bridge between the intricacies of human cognition and the realities of investment decisions. As researchers delve into these unexplored avenues, the potential for transforming investment practices, refining financial education, and fostering better decision-making grows exponentially. By charting these new directions, scholars, practitioners, and policymakers can collectively reshape the landscape of investment decisions, creating a more informed and adaptive financial future.

References

[1] Bali, T. G., & Cakici, N. (2008). Idiosyncratic volatility and the cross section of expected returns. Journal of Financial and Quantitative Analysis, 43(1), 29–58.

[2] Bekiros, S., Jlassi, M., Lucey, B., Naoui, K., & Uddin, G. S. (2017). Herding behavior, market sentiment and volatility: Will the bubble resume? The North American Journal of Economics and Finance, 42, 107–131. https://doi.org/https://doi.org/10.1016/j.najef.2017.07.005

[3] Ansari, A., & Ansari, V. A. (2021). Do investors herd in emerging economies? Evidence from the Indian equity market. Managerial Finance.

[4] Boussaidi, R. (2020). Implications of the overconfidence bias in presence of private information: Evidence from MENA stock markets. International Journal of Finance & Economics.

[5] Bregu, K. (2020). Overconfidence and (Over)Trading: The Effect of Feedback on Trading Behavior. Journal of Behavioral and Experimental Economics, 88, 101598. https://doi.org/https://doi.org/10.1016/j.socec.2020.101598

[6] Abreu, M., & Mendes, V. (2012). Information, overconfidence and trading: Do the sources of information matter? Journal of Economic Psychology, 33(4), 868–881. https://doi.org/https://doi.org/10.1016/j.joep.2012.04.003

[7] Caglayan, M. O., Hu, Y., & Xue, W. (2021). Mutual fund herding and return comovement in Chinese equities. Pacific-Basin Finance Journal, 68, 101599. https://doi.org/https://doi.org/10.1016/j.pacfin.2021.101599

[8] Chen, Z., & Zheng, H. (2022). Herding in the Chinese and US stock markets: Evidence from a micro-founded approach. International Review of Economics & Finance, 78, 597–604. https://doi.org/https://doi.org/10.1016/j.iref.2021.11.015

[9] Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. The Journal of Finance, 53(6), 1839–1885.

[10] Abreu, M., & Mendes, V. (2020). Do individual investors trade differently in different financial markets? The European Journal of Finance, 26(13), 1253–1270.

[11] Dasgupta, A., Prat, A., & Verardo, M. (2011). The Price Impact of Institutional Herding. The Review of Financial Studies, 24(3), 892–925.

[12] Dennis, P. J., Strickland, D., & Dennis, P. J., & Strickland, D. (2002). Who blinks in volatile markets, individuals or institutions? The Journal of Finance, 57(5), 1923–1949.

[13] Fellner, G., & Krügel, S. (2012). Judgmental overconfidence: Three measures, one bias? Journal of Economic Psychology, 33(1), 142–154. https://doi.org/https://doi.org/10.1016/j.joep.2011.07.008

[14] Fellner-Röhling, G., & Krügel, S. (2014). Judgmental overconfidence and trading activity. Journal of Economic Behavior & Organization, 107, 827–842. https://doi.org/https://doi.org/10.1016/j.jebo.2014.04.016

[15] Grežo, M. (2020). Overconfidence and financial decision-making: a meta-analysis. Review of Behavioral Finance, 13(3), 276–296. https://doi.org/10.1108/RBF-01-2020-0020

[16] Hudson, Y., Yan, M., & Zhang, D. (2020). Herd behaviour & investor sentiment: Evidence from UK mutual funds. International Review of Financial Analysis, 71. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101494

[17] Khan, M. T. I., Tan, S. H., & Chong, L. L. (2019). Overconfidence mediates how perception of past portfolio returns affects investment behaviors. Journal of Asia-Pacific Business, 20(2), 140–161.

[18] Kumar, S., Sureka, R., & Colombage, S. (2020). Capital structure of SMEs: a systematic literature review and bibliometric analysis. Management Review Quarterly, 70(4), 535–565. https://doi.org/10.1007/s11301-019-00175-4

[19] Li, W., Rhee, G., & Wang, S. S. (2017). Differences in herding: Individual vs. institutional investors. Pacific-Basin Finance Journal, 45, 174–185. https://doi.org/https://doi.org/10.1016/j.pacfin.2016.11.005

[20] Mishra, P. K., & Mishra, S. K., Mishra, P. K., & Mishra, S. K. (2021). Do Banking and Financial Services Sectors Show Herding Behaviour in Indian Stock Market Amid COVID-19 Pandemic? Insights from Quantile Regression Approach. Millennial Asia,. https://doi.org/10.1177/09763996211032356

[21] Naveed, F., & Taib, H. M. (2021). Overconfidence bias, self-attribution bias and investor decisions: Moderating role of information acquisition. Pakistan Journal of Commerce and Social Sciences, 15(2), 354–377.

[22] Vo, X. V., & Phan, D. B. A. (2019b). Herding and equity market liquidity in emerging market. Evidence from Vietnam. Journal of Behavioral and Experimental Finance, 24, 100189. https://doi.org/https://doi.org/10.1016/j.jbef.2019.02.002

[23] Wanidwaranan, P., & Padungsaksawasdi, C. (2022). Unintentional herd behavior via the Google search volume index in international equity markets. Journal of International Financial Markets, Institutions and Money, 77. https://doi.org/https://doi.org/10.1016/j.intfin.2021.101503

[24] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. In Econometrica (Vol. 47, Issue 2).

[25] Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics, 116(1), 261-292

[26] Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. The Geneva Risk and Insurance Review, 32(1), 1–36. https://doi.org/10.1007/s10713-007-0003-3

[27] Gervais, S., & Odean, T. (2001). Learning to be overconfident. The Review of Financial Studies, 14(1), 1–27.

[28] Statman, M., Thorley, S., & Vorkink, K. (2006). Investor overconfidence and trading volume. The Review of Financial Studies, 19(4), 1531–1565. https://doi.org/10.1093/rfs/hhj032

[29] Odean, T. (1998). Are investors reluctant to realize their losses? The Journal of Finance, 53(5), 1775-1798.

[30] Walter, A., & Moritz Weber, F. (2006). Herding in the German mutual fund industry. European Financial Management, 12(3), 375–406.

[31] Bikhchandani, S., & Sharma, S. (2001). Herd behavior in financial markets: A review. IMF Staff Papers, 47(3), 279-310.

[32] Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. The American Economic Review, 80(3), 465-479.

[33] Hong, H., Xu, S., & Lee, C. C. (2020). Investor herding in the China stock market: an examination of CHINEXT. Romanian Journal of Economic Forecasting, 23(4), 47–61.

[34] Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. The Quarterly Journal of Economics, 114(1), 37-82.

[35] Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. The Quarterly Journal of Economics, 118(1), 73-105.

[36] Avery, C., & Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. American Economic Review, 724–748

[37] DeBondt, W. F., & Thaler, R. (1985). Does the stock market overreact? The Journal of Finance, 40(3), 793-805.

[38] Grinblatt, M., & Keloharju, M. (2001). What makes investors trade? The Journal of Finance, 56(2), 589-616.

[39] Guo, J., Holmes, P., & Altanlar, A. (2020). Is herding spurious or intentional? Evidence from analyst recommendation revisions and sentiment. International Review of Financial Analysis, 71

[40] Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211(4481), 453-458.

[41] Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. The Quarterly Journal of Economics, 112(2), 647-661.

[42] Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management, 14(3), 207–222.

[43] Bell, D. E. (1982). Regret in decision making under uncertainty. Operations research, 30(5), 961-981.

[44] Zeelenberg, M., & Pieters, R. (2004). Consequences of regret aversion in real life: The case of the Dutch postcode lottery. Organizational Behavior and Human Decision Processes, 93(2), 155-168.

[45] Wu, G., Yang, B., & Zhao, N. (2020). Herding behavior in Chinese stock markets during COVID-19. Emerging Markets Finance and Trade, 56(15), 3578–3587.

10.31341/jios.48.1.7 UDC 004.021:004.65
Open Access Original Scientific Paper

A Study on Comparative Analysis of Feature Selection Algorithms for Students Grades Prediction

Muhammad Arham Tariq *arhamtariq99@gmail.com*
Department of Computer Science
University of Central Punjab, Lahore, Pakistan

Abstract

Education data mining (EDM) applies data mining techniques to extract insights from educational data, enabling educators to evaluate their teaching methods and improve student outcomes. Feature selection algorithms play a crucial role in improving classifier accuracy by reducing redundant features. However, a detailed and diverse comparative analysis of feature selection algorithms on multiclass educational datasets is missing. This paper presents a study that compares ten different feature selection algorithms for predicting student grades. The goal is to identify the most effective feature selection technique for multi-class student grades prediction. Five classifiers, including Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Gradient Boosting (GB), and k-Nearest Neighbors (KNN), are trained and tested on ten different feature selection algorithms. The results show that SelectFwe(SFWE-F) performed best, achieving an accuracy of 74.3% with Random Forests (RT) across all ten feature selection algorithms. This algorithm selects features based on their relationship with the target variable while controlling the family-wise error rate.
Keywords: Classification Models, Educational Data Mining, Feature Selection, Multi-Class Datasets, Student Performance

1. Introduction

To The field of education has witnessed a surge in the availability of large volumes of student data in recent years [1]. This has led to a growing interest in the use of data mining techniques to extract meaningful insights and improve educational outcomes. Educational data mining (EDM) involves the use of these techniques to analyze student data and identify patterns that can inform instructional practices and improve student outcomes. Feature selection algorithms play a critical role in this process by reducing redundant features and increasing classifier accuracy. Feature selection is an essential step in educational data mining as it helps to identify the most relevant and informative features from a large set of variables. It enables researchers to improve the accuracy of predictive models and identify important factors that influence student learning outcomes. By selecting the most relevant features, educational institutions can also optimize their resources and tailor their interventions to specific student needs.

However, a detailed and diversified comparative analysis of feature selection algorithms for multi-class student grades prediction has been lacking. The main contribution of the research is a detailed and diversified comparative analysis of ten feature selection algorithms on multi-class educational datasets for predicting student grades, which had previously been missing in the literature. This paper aims to fill this gap by presenting a comparative analysis of ten different feature selection algorithms. The goal is to identify the most effective feature selection technique for predicting student grades. To evaluate the performance of these algorithms, five different classifiers, including Support Vector Machines (SVM) and Decision Trees (DT), are trained and tested on the ten feature selection algorithms.

The ten feature selection algorithms used are SelectKBest (SKBF), SelectKBest (SKBC), SelectKBest (SKBM), SelectPercentile (SPF), SelectFpr (SFPRF), SelectFwe (SFWEF), SelectFdr (SFDRF), VarianceThreshold (VT), GenericUnivariateSelect (GUSF), and GenericUnivariateSelect (GUS-M). Each algorithm is evaluated with five different machine learning models using a pipeline and 10K-fold cross-validation.

2. Literature Review

In this section, a review of relevant studies and literature that have investigated the prediction of academic performance using various machine learning algorithms and feature selection techniques is presented. These studies shed light on different approaches and methodologies employed in the field and provide valuable insights into the performance of these methods.

Jalota, C., & Agrawal, R. 2021[2] conducted a study on comparative analysis of correlation feature selection and wrapper-based feature selection on educational datasets. The results of the research claims that J48 algorithm gain high accuracy by combining with correlation feature selection. Kumar et al.,2022 [3] utilized five different classification algorithms combined with three different feature selection algorithms to build a model for student academic performance prediction. Feature selection algorithms utilized in the approach were Correlation Attribute Evaluator, Information Gain Attribute, Gain Ratio. The highest accuracy achieved was 83.33% with the help of Decision tree and Gain ratio. Agrusti, F, 2020 [4] presents a deep learning-based approach for predicting university dropout, with a case study conducted at Roma Tre University. The approach uses a long short-term memory (LSTM) neural network to model the temporal dependencies of students' academic performance and predict their likelihood of dropping out, achieving promising results. Acharya, A., & Sinha, D,2014 [5] utilized student academic performance dataset having 309 instances and 14 features. Main objective of the research conducted was to find best feature selection algorithm among three different types. The types were Filter based, Wrapper Based learning and Correlation based Algorithm. The results claimed Correlation Based Feature Selection provides best result. Nidhi, Kumar, &Agarwal, 2021 [6] presents a comparative analysis of different heterogeneous ensemble learning algorithms using feature selection techniques for predicting academic performance of students. The study evaluates the performance of six

different algorithms on a dataset of student performance indicators, comparing their accuracy, precision, recall, and F1 score. The results show that the Random Forest algorithm with feature selection outperformed the other algorithms in terms of accuracy and other evaluation metrics.Anuradha,& Velmurugan,2015 [7] conducted a study that uses a dataset of student information, including demographics, socio-economic status, and past academic performance, to train and test several algorithms, including Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). The paper evaluates the algorithms' performance in terms of accuracy, sensitivity, and specificity. The results show that SVM outperforms the other methods in terms of overall performance, with an accuracy of 89.2%, compared to Decision Tree (87.1%), Naive Bayes (84.3%), and KNN (82.7%).

Enaro, A. O., & Chakraborty, S,2020 [8] used a dataset of student information, including demographic, behavioral, and academic features, to evaluate the performance of four feature selection algorithms: Correlation-based Feature Selection (CFS), Information Gain (IG), ReliefF, and Chi-Square. The paper evaluates the algorithms' performance in terms of classification accuracy using three different classifiers: J48 Decision Tree, Naive Bayes, and Multilayer Perceptron (MLP) Neural Network. The results show that CFS outperforms the other algorithms in terms of classification accuracy across all three classifiers, followed by ReliefF and IG, while Chi-Square performs the worst. Bakker, T el al.,2023 [9] proposed a study that s uses a dataset of 44 autistic students and 87 non-autistic students, including demographic and academic performance features, to compare the performance of three machine learning algorithms: Decision Tree (DT), Artificial Neural Network (ANN), and Support Vector Machine (SVM). The results indicate that the SVM algorithm outperformed the other two algorithms in predicting academic success, achieving an accuracy of 89%. Hamdi et al,2022 [10] new feature selection method based on the Chicken Swarm Optimization algorithm (CSO) combined with machine learning techniques for predicting student academic performance. The authors use a dataset consisting of demographic and academic performance features of 470 students to evaluate the performance of the proposed method. The authors compare the performance of the proposed CSO-based feature selection method with four other feature selection methods, including Random Forest, Correlation-based Feature Selection (CFS), Chi-square (Chi2), and Information Gain (IG). The authors evaluate the algorithms' performance in terms of accuracy, precision, recall, and F1-score. The results show that the proposed CSO-based feature selection method outperforms the other four methods in terms of accuracy, achieving an accuracy of 84.89%. The study also identifies the most significant predictors of student performance, including demographic information and academic achievement. The findings of this study suggest that the proposed CSO-based feature selection method can be a useful tool for feature selection in predicting student performance and can contribute to the development of targeted interventions to improve academic outcomes in higher education.

Huynh-Cam et al,2022 [11] The study aimed to predict the academic performance of international students, students with disabilities, and local students using machine learning algorithms based on their admission profiles and first-semester grades.

Results showed that SVM was the best model for predicting academic performance of students with disabilities, while RF was best for local students. The most important features were numbers of required and elective credits, source of living expenses, and parental occupation and income. This study can help institutions take early measures to improve the academic performance of students and attract more international students. Najm et al.2022 [12] presents a roadmap for implementing an educational data mart based on historical data from Alexandria Private Elementary School in Iraq. The data mart is constructed and an OLAP cube is used for OLAP operations and reports. Nine algorithms are used for OLAP mining and clustering with expectation maximization is found to have the highest accuracy (96.76%) for predicting student performance and grades. This study can help academic institutions make informed decisions based on historical data.

3. Background

3.1. SelectKBest

The SelectKBest [13] algorithm in scikit-learn, which is used for feature selection in machine learning. However, they differ in the statistical tests that they use for selecting the top K features from a dataset. Here's a brief explanation of each algorithm:

3.1.1. SelectKBest with chi-squared test (SKBF)

This algorithm [14] uses the chi-squared test to evaluate the independence of each feature and the target variable. It selects the top K features with the highest chi-squared scores, indicating the strongest relationship with the target variable.

3.1.2. SelectKBest with ANOVA F-test (SKBC)

This algorithm uses the ANOVA F-test [15] to evaluate the difference in means between groups of samples. It selects the top K features with the highest F-scores, indicating the greatest difference in means between the groups.

3.1.3. SelectKBest with mutual information (SKBM)

This algorithm uses mutual information to evaluate the dependence between each feature and the target variable. It selects the top K features with the highest mutual information scores, indicating the strongest dependence with the target variable.

The main difference between these algorithms is the statistical test that they use for feature selection. The chi-squared test is useful for categorical data, the ANOVA F-test is useful for continuous data with categorical targets, and mutual information is useful for any type of data.

In summary, SelectKBest with chi-squared test (SKBF), SelectKBest with ANOVA F-test (SKBC), and SelectKBest with mutual information (SKBM) are all

variations of the SelectKBest algorithm that use different statistical tests for feature selection. The choice of algorithm depends on the type of data and the problem being solved.

3.2. SelectPercentile

SelectPercentile [16] is a feature selection technique in machine learning that, like SelectKBest, is used to select the top features from a dataset. However, instead of selecting a fixed number of features, SelectPercentile selects a specified percentage of the total number of features.

The SelectPercentile algorithm is based on univariate statistical tests, similar to SelectKBest. It evaluates the relationship between each feature and the target variable and selects the top features based on a specified statistical test. The difference is that instead of selecting a fixed number of features, SelectPercentile selects the top features based on a specified percentage.

The SelectPercentile algorithm works by first computing a score for each feature using the specified statistical test. It then selects the top percentage of features with the highest scores. For example, if a user specifies a percentile of 10%, SelectPercentile will select the top 10% of features with the highest scores.

3.3. SelectFpr

SelectFpr [17] is a feature selection technique in machine learning that is used to select features based on a specified false positive rate. It is a part of the Select family of feature selection methods in scikit-learn library in Python.

The SelectFpr algorithm works by selecting the features that have a false positive rate lower than the specified threshold. It uses a statistical test, such as the chi-squared test or ANOVA F-test, to evaluate the relationship between each feature and the target variable, and then selects the features that meet the specified false positive rate threshold.

3.4. SelectFwe

SelectFwe[18] is a feature selection technique in machine learning that is used to select features based on a specified family-wise error rate. It is a part of the Select family of feature selection methods in scikit-learn library in Python.

The family-wise error rate (FWER) is a statistical measure that indicates the probability of making at least one false discovery among all the discoveries made by a model. In the context of feature selection, it is the probability of selecting at least one irrelevant feature.

3.5. SelectFdr

SelectFdr[19] is a feature selection technique in machine learning that is used to select features based on a specified false discovery rate. It is a part of the Select family of feature selection methods in scikit-learn library in Python.

The false discovery rate (FDR) is a statistical measure that indicates the proportion of false discoveries among all the discoveries made by a model. In the context of feature selection, it is the proportion of irrelevant features selected by the model.

The SelectFdr algorithm works by selecting the features that have a false discovery rate lower than the specified threshold. It uses a statistical test, such as the chi-squared test or ANOVA F-test, to evaluate the relationship between each feature and the target variable, and then selects the features that meet the specified false discovery rate threshold.

3.6. Variance Threshold

Variance Threshold [20] is a feature selection technique used in machine learning to remove features with low variance from a dataset. Variance is a measure of how much a feature's values vary from the mean value. Features with low variance may not be useful in predicting the output and can be removed to simplify the model and reduce overfitting.

The idea behind Variance Threshold is that if a feature's variance is below a certain threshold, it is likely that the feature has almost constant values across all samples and will not provide much information to the model. Therefore, it is safe to remove such features.

3.7. GenericUnivariateSelect

GenericUnivariateSelect [21] is a feature selection algorithm provided by the scikit-learn library in Python. It is a type of univariate feature selection method, which means that it evaluates the importance of each feature independently and selects the best ones based on a statistical test or a score function.

The algorithm takes three parameters:

- **score_func:** This is a function that is used to calculate the score of each feature. The available score functions in scikit-learn include ANOVA F-value, mutual information, chi2, and others.
- **mode:** This parameter determines how the features are selected based on their scores. It can be set to 'k_best' to select the top k features with the highest scores, 'percentile' to select the features above a certain percentile of the score distribution, or 'false_discovery_rate' to select the features with the lowest false discovery rate.
- **param:** This parameter is used to specify the number of features to select in the case of 'k_best' mode, the percentile of features to select in the case of

'percentile' mode, or the target false discovery rate in the case of 'false_discovery_rate,mode.

The GUS-F and GUS-M are two variants of the GenericUnivariateSelect class, which use different methods for selecting the features. In summary, GUS-F selects features based on their scores in univariate statistical tests, while GUS-M selects features based on a mutual information score between each feature and the target variable. The choice between these two variants depends on the nature of the data and the research question being addressed.

4. Methods

In this section, detailed description of dataset, research methodology, data preparation steps, and the evaluation process conducted to analyze feature selection algorithms for student academic performance are added.

4.1. Dataset

Hamtini, T.,2015 [22] This dataset was acquired from the Kalboard 360 learning management system (LMS), which grants users access to educational materials as long as they are connected to the internet. Additionally, this system monitors students' progress, including their interactions with educational content, such as how often they read or view it. With the help of this data, our objective is to uncover the factors associated with students' academic success. The dataset pertains to a group of students who took part in an educational program. This dataset comprises 480 data points and 16 attributes. Each data point corresponds to a specific student, while each attribute relates to a particular characteristic or trait of that student. It includes three categories denoting Low-Level, Middle-Level, and High-Level student performance. The dataset incorporates demographic data such as gender, nationality, and class section, as well as academic performance indicators, including grades in various subjects, attendance, and quiz/exam scores. Furthermore, it encompasses data related to the student's learning styles, including their preferred method of learning, approach to learning, and motivation levels. This dataset is valuable in investigating correlations between academic performance and various demographic and learning style aspects. It can also assist in predicting student performance or highlighting areas that require educational interventions. Fig. 1 depicts the distribution of the three categories across the dataset.

4.2. Data Pre-Processing

Yu, N,2018 [23] Data Pre-processing is one of initial major step to be performed in the machine learning process. It can contain many phases. Because the dataset lacks anomalies, noise, or missing values, these steps were disregarded. Apart from that feature encoding was adopted. Feature encoding is the process of converting categorical variables in a dataset into numerical values that can be used in statistical models or machine learning algorithms.

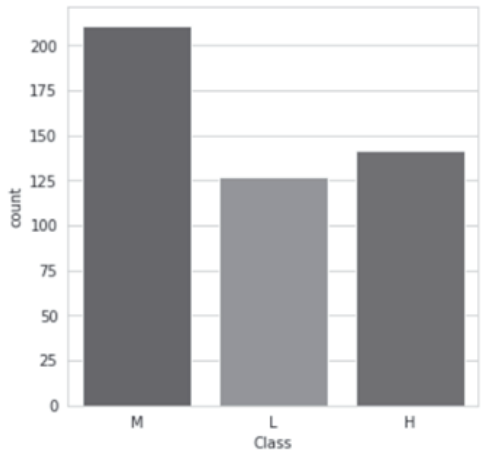


Figure 1. Dataset Classes Distribution

4.3. Features Selection

In this phase, ten different feature selection algorithms SelectKBest(SKBF), SelectKBest(SKBC),SelectKBest(SKBM),SelectPercentile(SPF),SelectFpr(SFPRF), SelectFwe(SFWEF),SelectFdr(SFDRF),VarianceThreshold(VT),GenericUnivariateSelect(GUSF),GenericUnivariateSelect(GUS-M) are applied. All of the feature selection algorithms are applied and evaluated differently with the help of different classification algorithms.

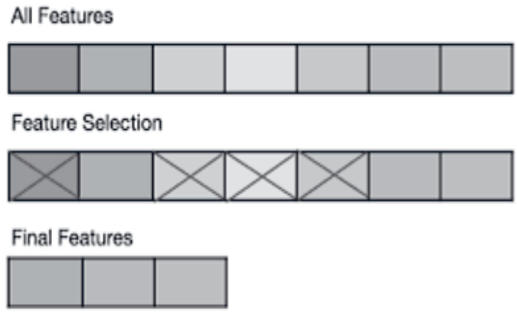


Figure 2. Working Demonstration of Feature Selection Algorithm

4.4. Hyperparameter Tunning

Hyperparameter tuning is a crucial step in the machine learning pipeline and can greatly impact the accuracy and generalization of the resulting model. In this phase of the configured approach four different classification models named as SVM, DT, RT and GB are applied and tunned with the help of the parameters mentioned in the Table 1. Fig 3 represent complete working model of hyperparameter tunning.

Classifier Name	Parameters
KNN	'params': {'n_neighbors': [3, 5, 7]}
DT	'params': {'kernel': ['linear', 'rbf'], 'C': [1, 10]
RT	{ params': {'n_estimators': [50, 100], 'max_depth': [5, 10], 'min_samples_split': [2, 5], 'min_samples_leaf': [1, 2]}
GB	{'n_estimators': [50, 100], 'learning_rate': [0.01, 0.1]}

Table 1. Classifiers Parameter

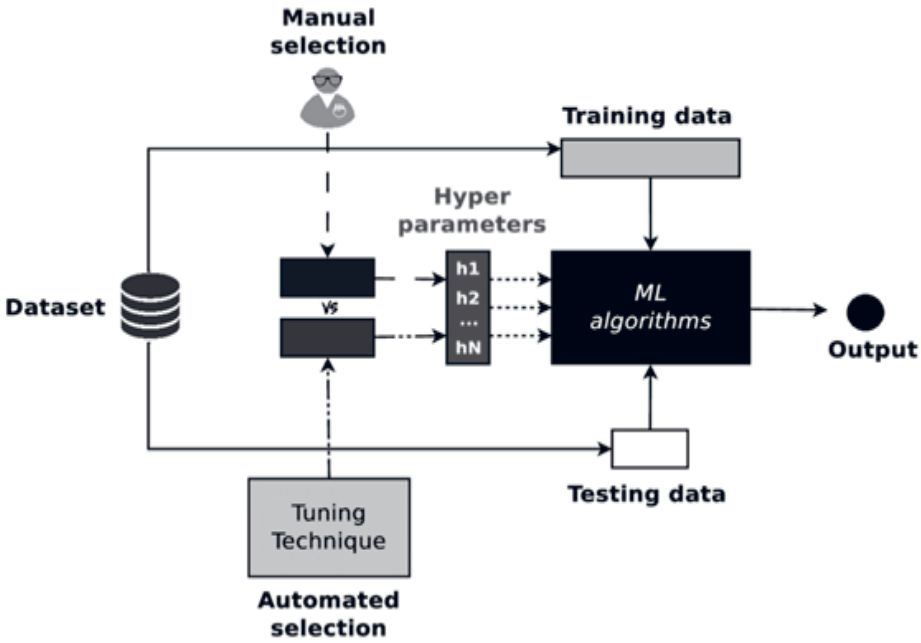


Figure 3. Working Demonstration of Hyperparameter Tunning [24]

4.5. Model Evaluation

Different evaluation metrics are applied in order to check the overall performance of the machine learning classifiers utilized in the configured approach.

The evaluation metrics adopted can be termed as:

Accuracy:

The formula for accuracy is:

Accuracy = (Number of Correct Predictions) / (Total Number of Predictions)

Precision:

The formula for precision is:

Precision = (Number of True Positive Predictions) / (Number of True Positive Predictions + Number of False Positive Predictions

Recall:

The formula for recall is:

Recall = (Number of True Positive Predictions) / (Number of True Positive Predictions + Number of False Negative Predictions)

F1-Score:

The formula for F1 score is:

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

5. Results and Discussion

In this Section detailed results of the classifiers combined with different feature selection algorithms will be explained. Fig. 4 & 5 represents the mean accuracies and F1-scores of the classifiers. On the X axis there are 5 different classifiers and Y-axis represents 10 different feature selection algorithms. The results are generated using 10 K fold cross validation. Table 2-4 represents the classifiers Recall, Precision and F1-Score. Table 5 comprises of précised mean accuracies.

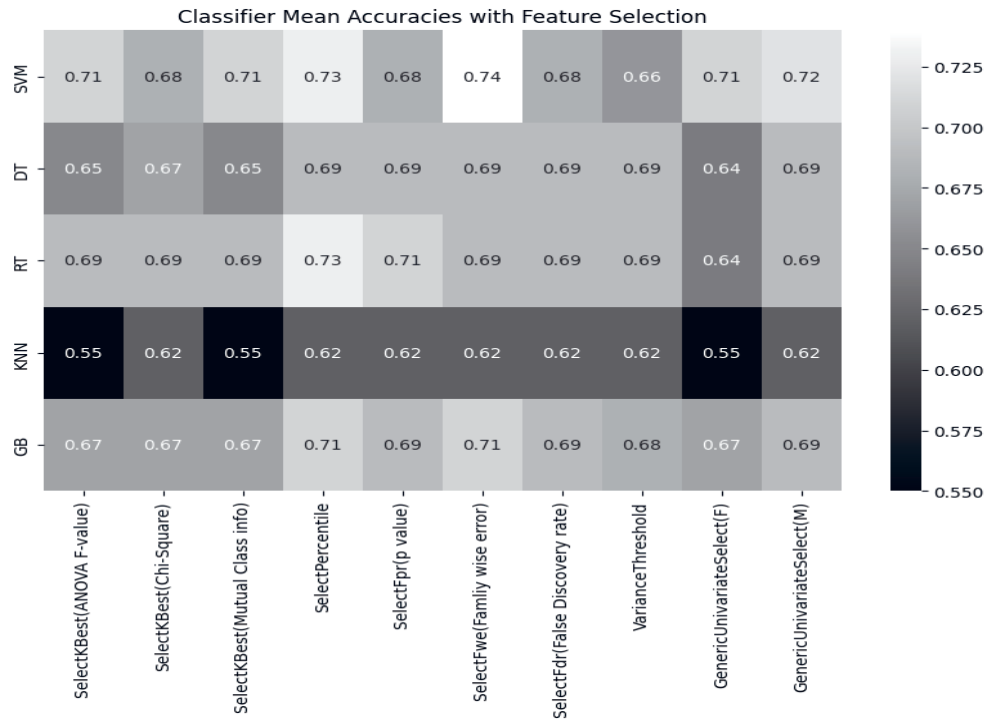


Figure 4. Accuracy of the Classifiers from Feature Selection Algorithms

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.69	0.72	0.71	0.65	0.74
SelectKBest(Chi-Square)	0.71	0.67	0.75	0.68	0.75
SelectKBest(Mutual Class info)	0.69	0.64	0.77	0.65	0.74

SelectPercentile	0.76	0.70	0.79	0.68	0.75
SelectFpr(p value)	0.69	0.65	0.81	0.68	0.78
SelectFwe(Famliy wise error)	0.76	0.72	0.81	0.68	0.75
SelectFdr(False Discovery rate)	0.69	0.65	0.79	0.68	0.78
VarianceThreshold	0.75	0.72	0.84	0.68	0.80
GenericUnivariateSelect(F)	0.69	0.72	0.70	0.65	0.74
GenericUnivariateSelect(M)	0.72	0.69	0.77	0.68	0.77

Table 2. Recall of Classifiers with Feature Selection

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.71	0.73	0.73	0.65	0.76
SelectKBest(Chi-Square)	0.74	0.68	0.76	0.68	0.76
SelectKBest(Mutual Class info)	0.71	0.64	0.77	0.65	0.76
SelectPercentile	0.77	0.70	0.81	0.68	0.77
SelectFpr(p value)	0.70	0.67	0.81	0.68	0.80
SelectFwe(Famliy wise error)	0.77	0.72	0.83	0.68	0.78
SelectFdr(False Discovery rate)	0.70	0.66	0.80	0.68	0.80
VarianceThreshold	0.77	0.72	0.86	0.69	0.82
GenericUnivariateSelect(F)	0.71	0.73	0.71	0.65	0.76
GenericUnivariateSelect(M)	0.73	0.69	0.781	0.69	0.80

Table 3. Precision of Classifiers with Feature Selection

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.69	0.72	0.71	0.64	0.73
SelectKBest(Chi-Square)	0.70	0.66	0.74	0.67	0.74
SelectKBest(Mutual Class info)	0.69	0.64	0.76	0.64	0.73
SelectPercentile	0.76	0.70	0.79	0.67	0.74
SelectFpr(p value)	0.69	0.65	0.81	0.67	0.78
SelectFwe(Famliy wise error)	0.76	0.72	0.80	0.67	0.74
SelectFdr(False Discovery rate)	0.69	0.64	0.79	0.67	0.78
VarianceThreshold	0.75	0.71	0.84	0.68	0.80
GenericUnivariateSelect(F)	0.69	0.72	0.70	0.64	0.73
GenericUnivariateSelect(M)	0.72	0.69	0.76	0.68	0.77

Table 4. F1-Score of Classifiers with Feature Selection

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.71	0.65	0.69	0.55	0.67
SelectKBest(Chi-Square)	0.68	0.67	0.69	0.62	0.67
SelectKBest(Mutual Class info)	0.71	0.65	0.69	0.55	0.67
SelectPercentile	0.73	0.69	0.73	0.62	0.71
SelectFpr(p value)	0.68	0.69	0.71	0.62	0.69
SelectFwe(Famliy wise error)	0.74	0.69	0.69	0.62	0.71
SelectFdr(False Discovery rate)	0.68	0.69	0.69	0.62	0.69

VarianceThreshold	0.66	0.69	0.69	0.62	0.68
GenericUnivariateSelect(F)	0.71	0.64	0.64	0.55	0.67
GenericUnivariateSelect(M)	0.72	0.69	0.69	0.62	0.69

Table 5. Accuracy of Classifiers with Feature Selection

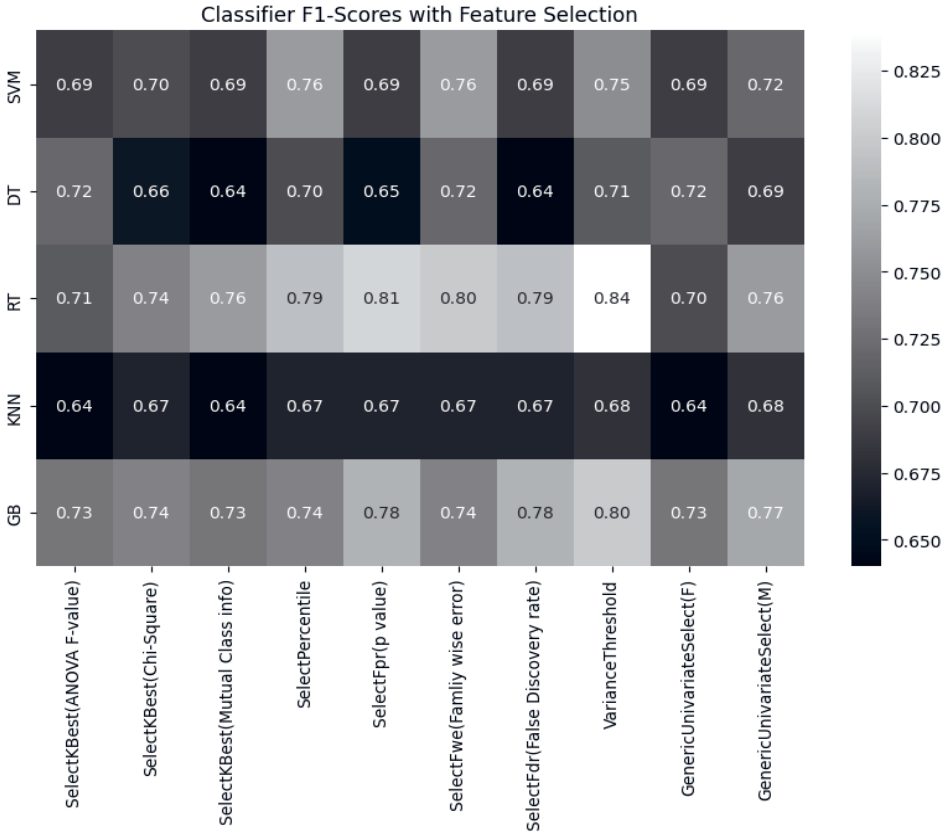


Figure 5. F1-Score of the Feature selection algorithms with the five Machine learning classifiers

SelectFwe (SFWE-F) is a feature selection algorithm that selects features based on their relationship with the target variable while controlling the family-wise error rate. In this study, SFWE-F performed best among the ten feature selection algorithms evaluated, achieving the highest accuracy of 74.3% when combined with Random Forests (RT) across all ten feature selection algorithms. Random Forest is a tree-based ensemble learning method that constructs multiple decision trees and combines their outputs to make predictions. It is known for its robustness against noise and high accuracy in both classification and regression tasks.

The reason why the combination of SFWE-F and Random Forests achieved the highest accuracy is that SFWE-F removes irrelevant features while controlling the family-wise error rate. By doing so, it identifies the most relevant features that have a

strong correlation with the target variable, which in this case is the student grades. Random Forests, on the other hand, is able to capture the nonlinear relationships between features and the target variable. When SFWE-F is combined with Random Forests, it helps to reduce the number of irrelevant features, thereby improving the accuracy of the model. Additionally, Random Forests are known to handle noise and overfitting well, which makes them an excellent choice for educational datasets where noise and irrelevant features may be present.

6. Conclusion

Predicting timely student academic performance has become a major concern now a days. Machine Learning Models accurate predictions can benefit educational stakeholders in many capacities. Like other datasets Mostly educational datasets by nature can contains noise and less contributing features. Feature Selection/Removal algorithms plays vital role in building a machine learning model. There are many such algorithms presented. Keeping this in mind, the configured research main objective was to find best performing feature selection algorithm. Previously there were many approaches configured for this purpose. But a detailed and diversified in depth analysis was missing. For this the purpose, a significant pool of features selection algorithm was evaluated using different classifiers. The results claimed that Combination of the SVM with SelectFwe(Family wise error) proved to be effective with an accuracy of 74.3%.

This is due to the fact that the SelectFwe algorithm was able to select the most informative features while controlling the false discovery rate. Additionally, SVM is a powerful classification model that can perform well on a variety of datasets. Taking future aspects under consideration, evaluation of different under sampling and noise reduction algorithms along with multiple feature selection algorithms can be effective. Lastly replacement of traditional machine learning models with deep learning models can produce more predictive power.

References

[1] Guleria, P., & Sood, M.Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Education and Information Technologies, 2023,28(1), 1081-1116.Available From: https://link.springer.com/article/10.1007/s10639-022-11152-y.

[2] Jalota, C., & Agrawal, R. Feature selection algorithms and student academic performance: A study. In International Conference on Innovative Computing and Communications: Proceedings of ICICC, 2020, Volume 1 (pp. 317-328). Springer Singapore.

[3] Kumar, M., Sharma, B., & Handa, D. Building predictive model by using data mining and feature selection techniques on academic dataset. IJMECS,2022, 14, 16-29.

[4] Agrusti, F., Mezzini, M., & Bonavolontà, G. Deep learning approach for predicting university dropout: A case study at Roma Tre University. Journal of e-Learning and Knowledge Society,2020, 16(1), 44-54.

[5] Acharya, A., & Sinha, D. Application of feature selection methods in educational data mining. International Journal of Computer Applications, 2014,103(2).

[6] Mm Nidhi, N., Kumar, M., & Agarwal, S. Comparative Analysis of Heterogeneous Ensemble Learning using Feature Selection Techniques for Predicting Academic Performance of Students. In 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST), IEEE,2021 ,pp. 212-217.

[7] Anuradha, C., & Velmurugan, T. A comparative analysis on the evaluation of classification algorithms in the prediction of students performance. Indian Journal of Science and Technology, 2015,8(15), 1-12.

[8] Enaro, A. O., & Chakraborty, S. Feature selection algorithms for predicting students academic performance using data mining techniques. Int. J. Sci. Technol.Res,2020,9(4).

[9] Bakker, T., Krabbendam, L., Bhulai, S., Meeter, M., & Begeer, S. Predicting academic success of autistic students in higher education. Autism,2023, 13623613221146439.

[10] Hamdi, M., Hilali-Jaghdam, I., Khayyat, M. M., Elnaim, B. M., Abdel-Khalek, S., & Mansour, R. F. Chicken Swarm-Based Feature Subset Selection with Optimal Machine Learning Enabled Data Mining Approach. Applied Sciences, 2022,12(13), 6787.

[11] Huynh-Cam T-T, Chen L-S, Huynh K-V. Learning Performance of International Students and Students with Disabilities: Early Prediction and Feature Selection through Educational Data Mining. Big Data and Cognitive Computing. 2022; 6(3):94. https://doi.org/10.3390/bdcc6030094.

[12] Najm, I. A., Dahr, J. M., Hamoud, A. K., Alasady, A. S., Awadh, W. A., Kamel, M. B., & Humadi, A. M. OLAP Mining with Educational Data Mart to Predict Students' Performance. Informatica,2022. 46(5).

[13] Goel, M., Sharma, A., Chilwal, A. S., Kumari, S., Kumar, A., & Bagler, G. Machine learning models to predict sweetness of molecules. Computers in Biology and Medicine, 2023.152, 106441.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2011.12:2825–2830.

[15] Luepsen, H. . ANOVA with binary variables: the F-test and some alternatives. Communications in Statistics-Simulation and Computation,2023, 52(3), 745-769.

[16] Yu, L., & Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th international conference on machine learning 2003,ICML-03,pp. 856-863.

[17] Mahmood, S., Ghosh, S., Vazquez, M. J., Bista, R. G., Abdollahi, H., & Jain, A. N. A machine learning approach to predicting gene essentiality in cancer cell lines. Frontiers in Genetics,2022, 13, 760. doi: 10.3389/fgene.2022.810329.

[18] Khan, A. M., Ali, N. N., & Mahmood, A. Z.. Classification of breast cancer histology images using transfer learning and ensemble methods. Sensors, 19(23), 2019,5149. doi: 10.3390/s19235149.

[19] Saad, T. M., Abdel-Nasser, M. M., & Aziz, A. A. Brain tumor classification using convolutional neural networks for MRI images. Journal of Medical Systems, 2018,42(7), 129. doi: 10.1007/s10916-018-0983-3.

[20] Olsen, J. K., Kallimani, J. M., & D'Mello, M. J. Predicting performance on concept inventory questions using student-generated question features. In Proceedings of the 9th International Conference on Educational Data Mining 2016,pp. 462-465.

[21] Perera, R., & Jayawardena, Predicting student grades using machine learning: A comparison of feature selection techniques. In Proceedings of the International Conference on Advanced ICT for Education (ICAICTE) 2022,pp. 70-77.

[22] Amrieh, Elaf Abu, Thair Hamtini, and Ibrahim Aljarah. "Preprocessing and analyzing educational data set using X-API for improving student's performance." In 2015 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), pp. 1-5. IEEE, 2015..

[23] Yu, N., Li, Z., & Yu, Z.. Survey on encoding schemes for genomic data representation and feature learning—from signal processing to machine learning. Big Data Mining and Analytics, 2018,1(3), 191-210.

[24] Mantovani, R. "Use of meta-learning for hyperparameter tuning of classification problems." PhD diss., University of Sao Carlos, 2018.

10.31341/jios.48.1.8 UDC 004.75:004.056
Open Access Survey Paper

The Usage of Clouds in Zero-Trust Security Strategy: An Evolving Paradigm

Jyoti Bartakke *Jyotibartakke@hotmail.com*
Zeal Institute of Business Administration,
Computer Application Research, Nerhe, Pune, India

Rajeshkumar Kashyap *rajdlw@gmail.com*
Sadhu Vaswani Institute of Management
Studies For Girls, Pune

Abstract

Zero-trust security is a security model that assumes no entity is implicitly trusted, regardless of its origin or scope of access. This approach requires continuous verification of all users, devices, and applications before granting access to resources. Cloud computing is a model for delivering IT resources and applications as a service over the Internet. Cloud computing offers many benefits, including scalability, agility, and cost savings. However, cloud computing also introduces new security challenges. This paper proposes a survey-based research methodology to evaluate the usage of clouds in zero-trust security strategies. The paper identifies different zero-trust security solutions, their key features, and the benefits and challenges of implementing these solutions in the cloud. The paper will also discuss the costs and benefits of different zero-trust security solutions. The findings of this research will be valuable for organizations that are considering implementing a zero-trust security strategy in the cloud. The paper will provide guidance on how to choose the best zero-trust security solution for organizational needs and how to implement it effectively.

Keywords: Zero-trust security, Cloud computing security challenges, Trust management, Zero-trust security approach, Zero trust benefits, Zero-trust security challenges

1. Introduction

Cloud computing has become a popular choice due to its scalability, flexibility, and cost efficiency. However, it also presents security challenges that traditional models struggle to address. Data security is crucial in cloud computing, as data is accessible globally and dynamic environments make it difficult to monitor access. Building secure cloud computing requires building trust in the cloud service provider and reliance on their services. This paper proposes a conceptual framework for trust-based permission systems within the cloud computing framework, introducing the zero-trust model. This model grants access based on user identity, device, and context, eliminating assumptions about user or device trustworthiness. This approach offers a

safer approach by not relying on conventional security paradigms. Access is granted only upon request, contingent on the user's identity, device, and context, ensuring enhanced security through the zero-trust model.

1.1. Importance of Building Trust in Cloud Environments

Trust is paramount in the realm of cloud computing. Establishing trust between cloud service providers (CSPs) and cloud customers (CS) is essential for fostering secure and reliable interactions. With the rise in cloud service usage, concerns about identity theft, data breaches, data integrity, and confidentiality have intensified. These challenges underscore the significance of robust trust management practices. Building trust involves not only ensuring the credibility of cloud service providers but also enabling cloud customers to have confidence in the services they use. Trust in cloud environments is pivotal for users to confidently store, process, and manage their data in the cloud.

2. Literature Review

Developing trust in cloud computing is challenging. Numerous investigators are currently working to develop a reliable and effective trust strategy for the paradigm of cloud computing. There are many trust structures, designs, and frameworks that have been proposed in order to build relationships between cloud customers (CS) and CSPs that are based on trust. This section covers the best work in the field.

[1] The authors conduct a comparative review of the security aspects of Zero Trust Networks (ZTN) in the context of cloud computing. The authors analyze and compare various security measures implemented in Zero Trust Networks, focusing on their application within cloud computing environments. The paper discusses different approaches and technologies related to Zero Trust Networks, evaluating their effectiveness, advantages, and potential challenges when applied in cloud computing scenarios. The authors might explore how Zero Trust Networks enhance security, mitigate risks, and provide a secure framework for cloud-based applications and data.

[2] This paper provides an in-depth overview of security issues related to cloud computing. The authors conducted a comprehensive survey, likely reviewing existing literature and research findings in the field.

[3] The authors propose a framework for securing electronic healthcare systems (EHSs) based on mutual trust. They argue that traditional access control frameworks are not sufficient to protect EHSs from modern threats and that a mutual trust-based approach is needed.

[4] In this paper, the authors likely present a novel approach or algorithm aimed at measuring and quantifying trust levels within cloud computing systems. Trust quantification in this context likely refers to assessing the reliability and credibility of various elements within cloud computing, such as services, resources, or entities.

[5] In this paper, the authors propose a novel method to assess the trustworthiness of users in cloud computing environments. The model is designed to evaluate user trust based on their behavior data, which may include various interactions, activities,

or transactions within the cloud system. By analyzing this behavioral data, the model likely aims to identify patterns and indicators that can be used to determine the level of trust associated with each user.

[6] the authors probably propose a system that assesses the trustworthiness of cloud services. The evaluation is likely based on evidence, which could include various data points and indicators related to the performance, reliability, and security of cloud services. Fuzzy logic, a mathematical framework dealing with uncertainty and imprecision, is likely used to handle the vagueness and uncertainty in the trust evaluation process.

[7] In this paper, the authors probably explore the complexities associated with establishing and managing trust in cloud computing environments. Trust management in this context typically involves addressing issues such as security, privacy, reliability, and data integrity. The paper might discuss challenges faced by users and organizations when it comes to trusting cloud service providers and the broader cloud ecosystem. Potential topics covered could include trust models, authentication mechanisms, data encryption, access control, and other security measures deployed in cloud environments to foster trust.

[8] In this book, the authors likely provide insights into the principles and practices of implementing a zero-trust security framework. Zero Trust emphasizes the need for strict identity verification and continuous authentication for anyone trying to access resources on a network, regardless of whether they are inside or outside the corporate perimeter. The book probably covers various aspects of network security, including encryption, access control, network segmentation, and continuous monitoring, all centered around the Zero Trust model.

[9] In this document, the author describes, the concept of Zero Trust Architecture (ZTA) is explained. Zero Trust is a security approach where organizations do not automatically trust any user or system, even if they are inside the corporate network. Instead, it emphasizes strict verification and least privilege access for everyone trying to access resources on the network, regardless of their location.

[10] The report defines zero trust as a security model that assumes that no entity is inherently trusted, whether inside or outside the organization's network. This approach requires continuous verification of all users, devices, and applications before granting access to resources.

[11] Introduced cloud research categorization. The authors argue that having a clear understanding of what is needed to deliver dependable software services helps to increase trust in the cloud. When it comes to resolving technical issues and fostering trust, it offers a range of transparency levels. The authors have also looked into the characteristics of operational trust in the cloud that incorporate versatility, flexibility, adaptability, dependability, and accessibility. In this article, the authors give a concise overview of the cloud infrastructure qualities that can be utilized to assess how secure cloud operations are.

[12] The paper provides a systematic literature review of the state-of-the-art trust evaluation mechanisms in cloud computing. The authors begin by defining trust and discussing the different types of trust that can exist in cloud computing. They then discuss the challenges of building trust in cloud environments, such as the lack of

visibility into cloud infrastructure, the dynamic nature of cloud environments, and the potential for insider threats.

[13] This paper offered a trust architecture that determines the security potency and trust value of cloud services. Cloud service users can successfully use the trusted model to choose a specific cloud service. The writer claims that cloud providers can assess a cloud service's strengths and weaknesses using the suggested trust model as a benchmark."

[14] Cloud security customers' resources, data, and apps were the subject of an investigation by authors. Additionally, the authors in this study are attempting to describe the trust problem and offer solutions for CSPs.

[15]Authors offered a contemporary trust framework that focuses on standards-based and certification-based security checks for cloud computing infrastructure services. The authors have also listed a number of essential elements for IAAS security. The trust framework that has been provided determines whether a cloud service is trustworthy based on a security analysis using standards and certification that the cloud provider has successfully completed.

[16] The author discussed how to group different cloud environment stages. The fuzzy centroid method of defuzzification is used by the authors to examine the measurement of trust value for CSPs and provide a more useful trust model as a result. Three criteria are also used by the authors to evaluate trust: accessibility, restructuring time, and dependability. As a result, this trust model improves the safety and reliability of cloud-based services. As an outcome, this model measures trust with fewer variables and is more realistic.

[17] stressed the importance of trust in the cloud environment. Because trust ensures the security of the cloud environment, the authors came to the opinion that it is a crucial component of cloud computing. Additionally, trust aids in the selection of trustworthy customers and services by both the user and the CSP. In order to assess trust based on various factors, the most recent trust models are also presented. This work gave a summary of the trust model based on cloud-based fuzzy logic

[18] The authors begin by discussing the importance of trust in cloud computing. They note that cloud users need to be able to trust their cloud providers to protect their data and applications and to provide reliable and secure services.

[19] In this paper, the authors probably explore the challenges and considerations involved in establishing a zero-trust security model within cloud computing contexts.

3. Methodology

This paper is a survey-based research paper. So, we use qualitative and quantitative research methodology that involves a comprehensive approach to understanding cloud security challenges and the implementation of the zero-trust security model.

3.1. Data Collection

We use a secondary data collection method. Gathering relevant data from academic sources and research papers on zero trust and cloud computing. industries that adopted

cloud computing for their workloads that industries websites through information collection and also gathering information on traditional security challenges in cloud computing and practical implementations of cloud security strategies, with a focus on zero-trust models.

3.2. Data Analysis

Employing qualitative and quantitative analysis techniques to assess the effectiveness of the zero-trust model. Analyzing real-world case studies and scenarios to evaluate its impact on cloud security.

3.3. Implementation of Zero-Trust Model

Practical implementation of the zero-trust security model in simulated cloud environments, considering diverse use cases and potential challenges.

3.4. Evaluation Criteria

Develop specific criteria to evaluate the success of the zero-trust model in enhancing cloud security. These criteria include factors such as data confidentiality, user authentication efficiency, and mitigation of insider threats.

By adopting this comprehensive methodology, we aim to provide valuable insights into the practical implications of the zero-trust security model in cloud computing environments.

4. Security Challenges In Cloud Computing

The internal security issues that cloud-hosted networks face is covered in this section. The purpose of this section is to highlight the key architectural components of contemporary cloud networks that use traditional network security controls. External forces can cause attacks like ransomware, phishing, data breaches, and malware to affect systems because of vulnerabilities caused by the issues listed below. The existence of vulnerabilities related to virtualization, for instance, is an important factor in the ease with which ransomware can spread from a host operating system to its virtualization and eventually to different host operating systems. Attackers have the power to harm cloud service providers, customers, and IT systems. because of weak security measures and control issues. (1)

There are a number of security concerns with cloud computing that need to be addressed.

Technical problems, legal issues, and other risks are all related to organizational and policy risks. Here are a few current issues that need to be addressed.

4.1. Vulnerabilities in Shared Technology

By leveraging more resources, attackers have a single point of entry and can inflict damage that is disproportionate to the threat. Cloud management and hypervisor systems are two examples of shared technology.

4.2. Account of Service Traffic Hijacking

The benefit of cloud computing is having an Internet connection, but there is also a chance that an account will be affected. A service interruption could happen if a privileged account is destroyed.

4.3. A Danial of Service

Any denial-of-service attack on the cloud provider could have an impact on all principles. (2)

4.4. Weaked Insider

In a cloud scenario, a clever insider can come up with additional ways to attack and hide their tracks.

4.5. Internet Protocol

IP has a number of security holes that can be exploited, such as DNS poisoning, ARP spoofing, and IP spoofing.

4.6. Injection Vulnerabilities

Numerous cloud users may experience serious effects from vulnerabilities at the management layer. A committed insider can come up with additional strategies for attacks and cover-ups.

4.7. Vulnerabilities In the API and Browser

Any security breach in a cloud service provider's API or connect presents a serious risk when coupled with social engineering or web-based assaults.

4.8. Modifications To the Business Model

The business model of a cloud user can experience significant change as a result of using the cloud.

4.9. Abusive Use

There are a lot of cloud computing features that can be used for malicious attacks, like the ability to launch DDoS or zombie attacks with a trailing period.

4.10. Malicious Insider

An insider who is malicious is always a serious risk, but one who works for a cloud provider can harm multiple consumers significantly. (2)

5. Trusting the Cloud: A Traditional Approach

Trusting traditionally in cloud computing information is based on a combination of factors, including the position of the cloud service provider, the level of transparency, and the specific needs of the organization.

The cloud service provider's positions. Cloud service providers (CSP) with a good reputation are more likely to be trusted by their customers. This reputation can be based on factors such as the providers' experience, security certifications, and customer reviews.

The level of security offered by the cloud service provider. CSPs must offer security to save their customers' data. Several techniques, including data encryption, access control, and intrusion detection, can be used to succeed in this security. The level of transparency that the cloud service provider presents. To build trust with customers, cloud service providers must be open and honest about their security procedures and how they safeguard customer data.

The significance of developing cloud computing trust is made clear in this section. In the cloud-based approach, communication between the cloud service providers (CSP) and the customers is possible. without any previous encounters. Because of this, it is frequently difficult to predict the reliability of both the customers and the CSP due to a lack of prior knowledge and experience with both groups Therefore, having no prior transactions and not having sufficient details may cause distrust. Furthermore, cloud computing decision-making facilitation relies heavily on trust. When data processing and storage are separated across globally separated data centers and resources are distributed in the same way, trust problems become critical. consequently, effective. Successful trust management is implemented to enable CSPs and customers to fully trust one another and reap the benefits of cloud computing.

In a service-based framework like cloud computing, the trust model [3] is very helpful for making wise decisions regarding the selection of trustworthy entities. [4] These trust models essentially take into account behavioral observations, recommendations, and previous and current relationship experiences when selecting reliable cloud service providers (CSP) and (CS) cloud customers.[5] Cloud computing offers a highly abstracted, massively dispersed, extremely complex, opaque system. Because of this dispersed and service-based cloud computing design, customers of the cloud must reconsider criteria for selecting trustworthy entities that go above and beyond the usual ones for determining the quality of service. Trust models are different because they vary for different environments, frameworks, and applications.

Because of how things change over time, Trust Management was primarily created for a specific application. Due to the services offered by the cloud, traditional trust management processes are therefore wholly inapplicable to the cloud model [6].

6. Trust Management Challenges for Cloud Computing

6.1. Trust Feedback Legitimacy

It is challenging to assess the trustworthiness of user feedback. In order to boost the reputation of some nodes or harm other nodes, malicious users may spread false trust information on purpose. It can be difficult to track malicious feedback and give expert users credibility ("Cloud Armour Project"). Assessing trustworthiness in feedback would be made possible by the existence of an independent quality assurance body that could inform consumers about sources of trust.

6.2. Lack of transparency

The data in cloud data centers is dispersed across multiple geographical locations and different virtualization levels. The most well-known CSPs of today, such as Amazon EC2/S3 and Microsoft Azure, do not completely disclose how physical and virtual servers are used. Only service event logs and virtual hardware metrics for performance are currently accessible to clients. Client confidence in the cloud would increase with greater transparency.

6.3. Loss of visibility and control

In a cloud computing environment, where numerous remotely located computing resources are involved, millions of nodes and chains of services are heavily shared. Many different nations' laws govern these resources, some of which fall outside the CSPs. As a result, once the data leaves the perimeter of the service, the data owner or user no longer has control over it. This loss of asset control in cloud computing reduces confidence in the service and increases the possibility of data loss.

6.4. Accountability complexity

Cloud accountability demands complex real-time accountability and places a strong emphasis on data security. Some of the complexities in the area of accountability arise as a result of the cloud framework's unique structure. • OS logging versus file-centric logging • Scale, size, and scope of logging • Live and Dynamic Systems • Challenges incorporated as a result of virtualization. Accountability should cover both physical and virtual server events, not just those on virtual servers. The accountability of physical servers is complicated by a lack of connection between the physical and virtual servers. Although current tools offer OS-centric logging, file-centric logging would allow users to follow data from the time it is created until the moment it is destroyed. In the expanding cloud scenario, where detailed logging could wipe out the

cloud storage holding logs, managing the scale of logging for efficient logging is crucial.

6.5. Weak trust chains

Because of the globalized nature of the cloud infrastructure, the trust chain within the cloud and the consumer may be weak at some points along the chain. Without sufficient verification of their reliability, a lot of new services, third parties, and providers could be added to the chain of on-demand services. Users should no longer be skeptical about the reliability of trust relationships because of the need for standardized monitoring and logging of the chain of relationships.

6.6. Lack of standardization and interoperability

Being a relatively new technology, cloud computing lacks interoperability and standards. Cloud adoption is hampered by inconsistent security standards that fail to address data privacy and trust issues. Currently, there aren't any standards for tracking data, evaluating trust, or auditing the object life cycle for data provenance. There isn't currently a standardized model for cloud management that could assess and track policies across various cloud offerings. (7)

7. Zero Trust Model In Cloud Computing

"Trust nothing verifies everything" is the main goal of zero trust security. This means that even if a customer or device is already connected to the network, they still need to be authenticated and given permission before granting access. The traditional security model, which demands that users and devices within the network be reliable, is changed by the zero-trust model. The traditional approach is less effective in a threat environment where internal networks are now easily accessible to attackers [1].

7.1. Zero trust model-based principles

Least privilege: Only the access necessary for users and devices to carry out their job-related duties should be granted.

Micro-segmentation: Only the resources required for a given purpose should be present in each of the network's small segments.

continuous monitoring: Any suspicious activity should be constantly looked out for on the network. (8)

A security structure called "zero trust" protects premises or cloud-based assets by preventing outsiders and unmanaged devices from accessing them, as well as by reducing all lateral movement. Data encryption, device verification, "identity and access management (IAM)", and "multi-factor authentication" (MFA) are among the technologies employed by zero trust. (9)

Features	Traditional model	Zero trust model
Approach	Trusts all users within the network	no user can be trusted, regardless of location
Cost	Less expensive to implement	More expensive to implement.
Authentication	Focus on user authentication	User and device both authentication
Security focus	Perimeter based security	Micro-segmentation and continuous monitoring
Response to threats	Incident response	Continuous monitoring and threat prevention

Table 1. Difference between the Traditional trust model and Zero trust model (10)

8. The Proposed Model Of Zero Trust

With the increasing issues of cloud infrastructure, modern security is recommended.

When putting new infrastructure into the cloud, security is the primary concern. The zero-trust security model for Cloud computing infrastructure is not compatible with the traditional security model. As a result, implementing the cloud requires zero trust. The zero-trust strategy uses strict access controls and keeps a record of all data used to improve cloud security. This is how the zero-trust architecture can better identify and stop external attackers while reducing security breaches brought on by insider attacks in the cloud infrastructure. The zero-trust cloud computing scheme also needed to successfully integrate and incorporate difficult practices, policies, and technology.

However, Figure 1 shows the basic concept behind zero-trust strategies.[19]

8.1. User identity is the first column.

Zero trust places a high value on the ongoing verification of trusted users. It is possible to restrict user access and privileges utilizing technologies such as multi-factor authentication and the concept of Identity Credentials, and Access Management (ICAM), as well as ongoing user credibility monitoring and validation technologies. Technologies like traditional web gateway solutions that secure and protect user interactions are also important.

8.2. Device security is the second column.

A fundamental aspect of a ZT approach is the safety and trustworthiness of devices. Data from some "system of record" solutions, like mobile phone device managers, can be helpful for determining how trustworthy a device is. Every access request should also undergo additional evaluations (such as checks for compromise states, software versions, protection statuses, encryption facilitation, etc.).

8.3. The third column is network security.

Some networks, operations, and tools are becoming more important than perimeter defenses. This is not the result of one technology or use case, but rather the result of a number of new services and technologies that allow users to communicate and work together in novel ways. Zero-trust networks truly make an effort to distinguish important data from other data by introducing perimeters in from the network edge. However, the edge is now much more clearly discernible. The traditional perimeter firewall approach of "castle and moat" is insufficient.

8.4. Workload and Application Security is the fourth column.

The adoption of zero trust depends on the security and effective management of the application's layer, computer containers, and machine virtualization. Making more accurate and thorough access decisions is possible with the help of an understanding of and management of the technology stack. Multi-factor authentication is increasingly crucial for providing applications in zero-trust environments with proper access control, as is to be expected.

8.5. Automaton: The fifth column

Security Automation and Orchestration is effortless and economical. ZT frequently employs security automation reaction tools, which employ workflows to automate operations across various products while preserving end-user control and participation Automated tools are used in security operation centers for analyzing user and entity behavior as well as security information and event management. These security tools are connected by security automation, which also helps with managing various security systems. Together, these tools can drastically cut costs, even response times, and human labor.

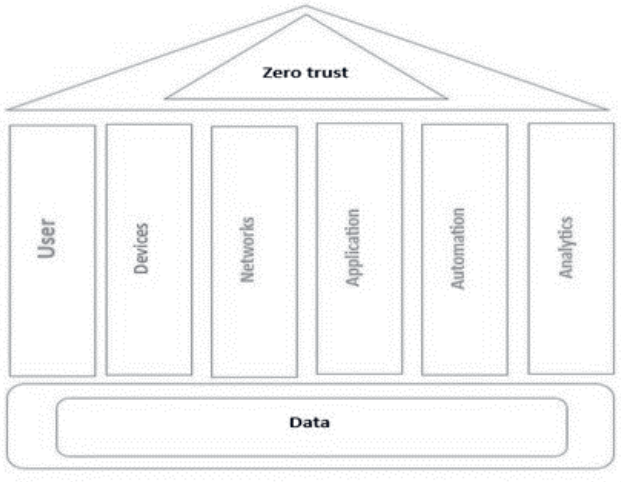


Figure 1. Six columns of the zero-trust security model [19]

8.6. Sixth column: Analytics

Security Analytics and Visibility

Thanks to the zero-trust use of resources such as data security management, advanced security, data analysis systems, security user behavior data analysis, and other analytics systems, security experts can see what is happening in real-time. Setting aside data analysis from cyber-related incidents can aid in the creation of security protocols before an actual attack occurs. (11)

9. Trust Building Is a Foundation

Zero Trust is a framework basis of trust "nothing verifies everything" and needs a flexible deployment model that steady evaluation and tracking first. Access is restricted to using dynamic, sensitive-to-context trust extensions in accordance with the threshold authorization regulations that have been established. The question of "How do we find how trustworthy something is?" is one that arises in this movement towards trust. Many security organizations have trouble coming up with an answer to this dilemma. Traditional programmers make the erroneous assumption that all information and transactions can be trusted. By assuming that all information and transactions are immediately untrue, Zero Trust changes the estimation of trust. How do we create enough trust is the new question. Even though some fundamental concepts and components are applicable to all deployments, there isn't a single strategy that all organizations can use. The organization's goals and requirements will determine the level of trust To manage the trustworthiness of every transaction, environments with zero trust interact with controls for devices, users, data, and apps.

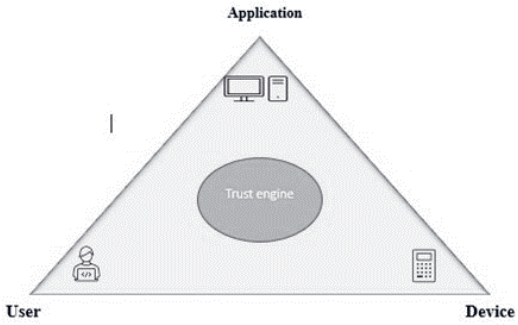


Figure 2. Zero Trust foundational Triangle [19]

"Trust Engine" is able to rapidly assess a user, device, or application's overall level of trust by establishing a "trust score" in a specific network. To determine a policy-based authorization decision, the trust engine assesses the trust score and applies it to each transaction request.

A user, device, or application’s trust score reflects how reliable they are. Its value is produced by a variety of factors and circumstances. The trust score may also be

affected by information such as past interactions, system knowledge, and access permissions granted or denied by the system.

The trust score of a user, application, or device shows how trustworthy they are. Its value is the result of many different things and circumstances. The trust score may also be affected by information about previous interactions, system knowledge, and access permissions granted or denied by the system.

The Trust Engine uses a "trust score" and the Zero Trust foundational triangle to determine each agent's level of trustworthiness before they join the network. The collection of information about the participants in a network request is referred to as an "Agent" or "Network Agent." A user, an application, and a device are typically included in this information. In order to provide the situational context in actual time, this combination of data is accessed upon request in actual time to provide contextual information and assist users in making the best authorization decisions possible. The customer, application, device, and trust score are combined to create an agent after the trust score has been identified. The request can then be granted by applying the policy to the agent after that [19].

10. Benefits of Cloud Computing In Zero Trust

There are various advantages to zero trust security for cloud computing, including.

10.1. More Powerful Security

Zero trust security is a safer approach than conventional security models because it does not rely on a single point of failure.

10.2. Better compliance with the law

Companies can adhere to compliance standards with the help of zero trust security.

10.3. Reduced cost

By doing away with the need for complicated security infrastructure, zero-trust security can assist organizations in saving money.

11. Challenges Of Implementing Zero Trust Security

Challenges with implementing zero-trust security, including.

11.1. Complexity

A sophisticated security strategy is zero trust security. To implement and manage it, a significant amount of time and money must be spent.

11.2. Cost

Zero security can be expensive to implement. The cost of implementing zero trust security will vary depending on the scope and complexity of the organization.

11.3. Cultural change

Zero trust security needs a change in culture. Employees need to be educated about the new security model and how to operate within it.

Organizations that are considering implementing zero trust security should start by assessing their current security posture. They should also create a strategy for putting into practice zero trust security that considers the issues covered in this paper's discussion.

12. Recommendation for Implementing Zero Trust Security

12.1. Start Small

Don't try to implement zero security across the entire organization all at once. Start by implementing it in a small, controlled environment.

12.2. Get by informing senior leadership.

Zero trust security needs an important investment of time and resources. It is important to get buy-in from senior leadership before starting the implementation process.

12.3. Educate employees

Employees need to be educated about the new security model and how to operate within it. Use of phased approach: implement zero trust security in phases.

Zero trust security is a new security strategy that may help organizations strengthen their security posture. However, it is important to be aware of the issues associated with implementing zero trust security. By following the recommendations that have been discussed in this paper, patients can increase their chances of success.

13. Discussion

One of the biggest trends in cybersecurity today is cloud computing, along with zero trust security. Agility, scalability, and affordability are just a few advantages that cloud computing has to offer. Data loss, data breaches, and unauthorized access are a few of the new security issues that are brought about by this.

Any user or device, no matter of their location or network, cannot be trusted, in the zero-trust security design. Because of this, all requests for access to resources must first be approved. Zero trust security uses a layered approach to security to help reduce the security risks in cloud computing.

It is believed that cloud computing security is a significant issue. Trust is crucial for cloud computing security. It makes it possible for CSPs and customers to locate dependable entities in a heterogeneous cloud infrastructure. Research on cloud computing trust is very active. This was the subject of numerous academic publications. Cloud computing and zero-trust security can be difficult to implement and manage. Organizations need to have a strong understanding of both technologies in order to design and implement an easy solution. The implementation and upkeep of cloud computing and zero-trust security can be costly. Organizations need to factor in the cost of software, hardware, and labor when considering cloud computing and zero-trust security. The security environment of the organization can be studied in great detail thanks to cloud computing and zero-trust security. The protection of this data from unauthorized access is the responsibility of organizations.

Zero trust, despite its difficulties, is a useful security framework that can aid organizations in strengthening their security posture. Organizations can minimize the security risks associated with cloud computing and enhance their overall security posture by carefully planning and implementing a zero-trust solution.

14. Conclusion

The current security difficulties of cloud computing were talked about in this paper. The cloud computing strategy is used by many businesses. As a result, they face numerous challenges. Following that, the difficulties of traditional cloud computing trust were discussed. This paper discussed the significance of zero-trust security in cloud computing. a difference between the traditional trust model and the zero-trust security model. To address these issues, we recommend utilizing the zero-trust model in cloud computing. Cloud computing comes with a lot of problems. Be that as it may, it likewise affects creating trust. Although many cloud service providers claim to provide robust security, the reality is that they face greater failure risks in their efforts to completely protect cloud presence accusers.

The advantages and drawbacks of implementing zero trust security in an organization are discussed in this paper.

The paper offers some suggestions for putting zero-trust security into place. It was difficult for organizations to monitor how users accessed and moved data within the cloud. A concept-based model for a zero-trust cloud computing strategy is proposed and discussed in this study. Both cloud service providers and end users benefit from its effective trust management benefits for cloud computing. Customers of CSPs will find it easier to select dependable providers in the cloud with the assistance of the proposal structure strategy. To provide cloud service providers and customers with security in their businesses, we proposed a zero-trust cloud computing model in this paper.

References

[1] Sarkar, S. et al. (2022) Security of zero trust networks in cloud computing: A comparative review, Welcome to DTU Research Database.: Sustainability. doi: 10.3390/su141811213.

[2] Ramachandra, G., Iftikhar, M. and Khan, F.A. (2017) 'A comprehensive survey on security in cloud computing', Procedia Computer Science, 110, pp. 465–472. doi: 10.1016/j.procs.2017.06.124.

[3] Singh, A. and Chatterjee, K. (2017) 'A mutual trust-based access control framework for Securing Electronic Healthcare Systems', 2017 14th IEEE India Council International Conference (INDICON) [Preprint]. doi:10.1109/indicon.2017.8487658.

[4] Li, X. et al. (2016) 'A method for trust quantification in cloud computing environments', International Journal of Distributed Sensor Networks, 12(2), p. 5052614. doi:10.1155/2016/5052614.

[5] Chen, Z., Tian, L. and Lin, C. (2018) 'Trust evaluation model of cloud user based on behavior data', International Journal of Distributed Sensor Networks, 14(5), p. 155014771877692. doi:10.1177/1550147718776924.

[6] Selvaraj, A. and Sundararajan, S. (2016) 'Evidence-based trust evaluation system for Cloud Services using fuzzy logic', International Journal of Fuzzy Systems, 19(2), pp. 329–337. doi:10.1007/s40815-016-0146-4.

[7] Khan, M.S., Warsi, M.R. and Islam, S. (2019) 'Trust management issues in cloud computing ecosystems', SSRN Electronic Journal [Preprint]. doi:10.2139/ssrn.3358749.

[8] Gilman, even and Barth, Doug. (2017) Zero Trust Networks Building Secure Systems in Untrusted Networks [Preprint].

[9] Rose, S. et al. (2020) Zero trust architecture. doi: 10.6028/nist.sp.800-207-draft2.

[10] 'Zero Trust Cybersecurity Current Trends' (2019) American Council for Technology-Industry Advisory Council (ACT-IAC) [Preprint].

[11] Abbadi, I.M. and Martin, A. (2011) 'Trust in the cloud', Information Security Technical Report, 16(3–4), pp. 108–114. doi: 10.1016/j.istr.2011.08.006.

[12] Chiregi, M. and Jafari Navimipour, N. (2018) 'Cloud computing and trust evaluation: A systematic literature review of the state-of-the-art mechanisms', Journal of Electrical Systems and Information Technology, 5(3), pp. 608–622. doi: 10.1016/j.jesit.2017.09.001.

[13] Shaikh, R. and Sasikumar, M. (2015) 'Trust model for measuring security strength of Cloud Computing Service', Procedia Computer Science, 45, pp. 380–389. doi: 10.1016/j.procs.2015.03.165.

[14] Horvath, A.S. and Agrawal, R. (2015) 'Trust in cloud computing', Southeast Con 2015 [Preprint]. doi:10.1109/secon.2015.7132885.

[15] Saxena, A.B. and Dawe, M. (2018) 'Trust Framework for IAAS—a tool based on security checks through standards and certifications', Information and Communication Technology for Intelligent Systems, pp. 369–376. doi:10.1007/978-981-13-1747-7_35.

[16] Prakash, P. et al. (2018) 'Enhancement of cloud security and strength of service using trust model', International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, pp. 1345–1353. doi:10.1007/978-3-030-03146-6_157.

[17] Ritu, Randhawa, S. and Jain, S. (2017) 'Trust models in Cloud computing: A review', International Journal of Wireless and Microwave Technologies, 7(4), pp. 14–27. doi:10.5815/ijwmt.2017.04.02.

[18] P, Archana. and P, Meenu. (2014). Trust management in cloud computing: A survey.', International Journal of Computer Applications, [Preprint]. doi: 108(18), 1-12.

[19] Mehraj, S. and Banday, M.T. (2020) 'Establishing a Zero trust strategy in cloud computing environment', 2020 International Conference on Computer Communication and Informatics (ICCCI) [Preprint]. doi:10.1109/iccci48352.2020.9104214.

10.31341/jios.48.1.9 UDC 004.891:616.12
Open Access Original Scientific Paper

Fuzzy rules-based Data Analytics and Machine Learning for Prognosis and Early Diagnosis of Coronary Heart Disease

Althaf Ali A *althafalia@mits.ac.in*
Department of Computer Applications,
Madanapalle Institute of Technology & Science (MITS),
Madanapalle, Andhra Pradesh, India

Umamaheswari S *umarunn@gmail.com*
Department of Information Technology,
C. Abdul Hakeem College of Engineering and Technology,
Melvisharam Tamil Nadu, India

Feroz Khan A.B *abferozkhan@gmail.com*
Department of Computer Science
Syed Hameedha Arts and Science College,
Kilakarai, Tamil Nadu, India

Jayabrabu Ramakrishnan *jayabrabu@jazanu.edu.sa*
College of Computer Science and Information Technology,
Jazan University, Jazan, Kingdom of Saudi Arabia

Abstract

Globally, cardiovascular diseases stand as the primary cause of mortality. In response to the imperative to enhance operational efficiency and reduce expenses, healthcare organizations are currently undergoing a transformation. The incorporation of analytics into their IT strategy is vital for the successful execution of this transition. The approach involves consolidating data from various sources into a data lake, which is then leveraged with analytical models to revolutionize predictive analytics. The deployment of IoT-based predictive systems is aimed at diminishing mortality rates, particularly in the domain of coronary heart disease prognosis. However, the abundant and diverse nature of data across various disciplines poses significant challenges in terms of data analysis, extraction, management, and configuration within these large-scale data technologies and tools. In this context, a multi-level fuzzy rule generation approach is put forward to identify the characteristics necessary for heart disease prediction. These features are subsequently trained using an optimized recurrent neural network. Medical professionals assess and categorize the features into labeled classes based on the perceived risk. This categorization allows for early diagnosis and prompt treatment. In comparison to conventional systems, the proposed method demonstrates superior performance.

Keywords: data analysis, healthcare, fuzzy rule, diagnosis, neural network

1. Introduction

In the realm of healthcare, the availability of data-driven insights has become essential for effectively managing diseases and improving patient outcomes. Heart disease, according to the World Health Organization, has emerged as a predominant focus due to its status as one of the leading causes of mortality. To combat this pressing issue, a comprehensive approach utilizing massive data methods is employed for the detection and management of heart disease. Their position as the principal cause of death on a universal scale has consistently been maintained by cardiovascular diseases. In the face of this harrowing reality, healthcare organizations are currently in the midst of a transformative phase where their strategies are being reevaluated to improve operational efficacy and decrease costs. This change necessitates the integration of analytics into their IT strategy. The primary objective of this paradigm shift is for advanced analytical insights to be harnessed, evidence-based care plans to be implemented, and patient engagement outcomes to be bolstered. The achievement of these objectives requires the consolidation of data from diverse sources into a data lake, subsequently employing analytical models to revolutionize information management, reporting, and predictive analytics.

However, the need for innovation in healthcare extends beyond just analytics. Immense promise is held by IoT-based prognostic systems, particularly in reducing mortality rates associated with cardiovascular diseases. In this context, our research is sought to contribute to the vast field of coronary heart disease prognosis, offering substantial data analysis. Nevertheless, significant challenges are presented by the sheer abundance of data spanning multiple disciplines in terms of analysis, extraction, management, and configuration within the expansive landscape of big data technologies and tools.

Motivated by the imperative to improve heart disease prognosis and prevention, a multi-level fuzzy rule generation approach is introduced by our research. This approach is designed to uncover the essential characteristics required for the early identification of heart disease, a critical step in improving patient outcomes and overall public health. To enhance the efficacy of this approach, an optimized recurrent neural network is utilized, which further refines the prediction process. Through this classification, patients can be assessed and categorized based on their perceived risk by medical professionals, ultimately enabling early diagnosis and prompt intervention.

The urgency of this transformation transcends the realm of analytics alone. The development of Internet of Things (IoT)-based prognostic systems has offered considerable promise, particularly in reducing mortality rates associated with cardiovascular diseases. In light of this, our research endeavors to make significant contributions to the extensive field of coronary heart disease prognosis through comprehensive data analysis. Nevertheless, the sheer profusion of data spanning numerous disciplines poses formidable challenges in terms of data analysis, extraction, management, and configuration, particularly within the vast expanse of big data technologies and tools.

Driven by the imperative to enhance heart disease prognosis, reduce mortality, and elevate the quality of patient care, our research introduces a multi-level fuzzy rule generation approach. This methodology aims to identify the essential characteristics required for the early detection of heart disease, a pivotal step in improving patient outcomes and global public health. To augment the efficacy of this approach, an optimized recurrent neural network is employed, further refining the prediction process. Medical professionals can then assess and categorize patients based on their perceived risk, facilitating early diagnosis and timely intervention. The ultimate aim is to surpass the performance of conventional systems in terms of accuracy and efficiency, thus making substantial strides in the realm of cardiovascular health. The urgency of our research is underpinned by the critical need to enhance heart disease prognosis, mitigate mortality rates, and augment patient care, ultimately contributing to a healthier future for individuals worldwide.

2. Literature Review

Currently, heart disease stands as the foremost cause of death globally, a trend that the World Health Organization (WHO) anticipates will persist in the foreseeable future. This malady has long been associated with a significant societal burden, underscoring the pressing need for substantial improvements in cardiovascular health [1]. Leveraging the latest advancements in demographic and perceptual technologies, individuals can access online medical services [2]. Demographic tools are generating copious amounts of statistical data within the medical field, and cloud computing is now playing a pivotal role in managing this vast reservoir of data [3]. Population-based healthcare tools, such as cloud-based solutions, are being developed to monitor prevalent diseases and enhance the quality of care delivered through online medical services [4].

The successful model of this study was devised by merging medical sensors with the UCI repository database to predict the occurrence of heart illness within the communal. To efficiently extract data collected by intermediary sensor systems [5] and store it in real-time on cloud servers, a demographic framework was constructed, underpinned by Bayesian sensor networks (BSN). This audit data is then scrutinized to detect erroneous information in order to forecast the incidence of heart disease [6]. To achieve rigorous disease diagnosis monitoring, the current research explores various in-depth learning methods and employs multiple machine learning algorithms.

Although electronic health record (EHR) data retrieval has been simplified, data accuracy remains a pertinent concern. This issue raises questions regarding the reliability of EHR data for precision and accuracy [7, 8]. Machine learning studies also delve into the outcome of an EHR-based risk framework, which is influenced by metadata, including an EHR input converter. The absence of relevant data skills further complicates this matter [9]. Thus, it is expected that enhanced modeling techniques, advanced machine learning methodologies, and increased data reservoirs will augment the effectiveness of current prediction algorithms.

Despite the burgeoning importance of diverse data processing techniques in health research [12], there is a tendency to downplay their significance, despite warnings that

these advances will define their strengths and limitations [13]. The challenges and issues related to health record processing techniques are reiterated [14, 15]. Cardiovascular risk prediction using medical data has long been a subject of research interest [16, 17]. While significant efforts have been invested, the accuracy of these predictions remains adequate. However, the scarcity of data at its core hinders the efficacy of machine learning, statistical methods, and conventional techniques, leading to feature analysis problems that yield imprecise predictions [18, 19]. In a bid to address this data gap, a medical data classification method, which calculates the impact measure on various levels to determine the target class, is described in [22, 23]. This approach relies on diverse attributes and repeated impact measures[10]. By reducing the number of features and diagnostic tests, the need for physician intervention in patient cases is diminished [11]. The findings of this database analysis pertaining to heart disease underscore the superior accuracy compared to traditional classification methods, while also highlighting the swiftness and precision of this technology.

3. Methodology

The methodology adopted in this study revolves around the smooth integration of an open-source UCI database. This integration is followed by a series of critical phases aimed at ensuring the quality and relevance of data for a comprehensive analysis. The study's approach can be delineated into the following stages:

Database Verification: To commence, the first phase entails the validation of the UCI open-source database. This process is essential to ascertain the trustworthiness and suitability of the data source for subsequent analysis. Its primary goal is to confirm not only the data's accuracy but also its currency.

Data Processing: Following the verification of the database's integrity, the subsequent actions in data management come into effect. These actions encompass data preparation, refining, migration, enhancement, discretization, and attribute curation. Each of these steps holds a pivotal role in getting the data ready for analysis.

Preprocessing: The preparatory phase encompasses a wide array of tasks, including addressing missing data, deduplicating, and managing data anomalies. This stage is critical for maintaining data uniformity and ensuring that irregularities that could introduce bias into the analysis are resolved.

Data Cleaning: The data cleaning step is devoted to boosting data quality by resolving issues such as discrepancies, inconsistencies, and inaccuracies. This might encompass standardizing data structures and resolving discrepancies.

Data Transfer: Data migration involves the transfer of data to a suitable environment or platform for analysis. This may necessitate moving data to a specific data analysis tool or platform.

Data Optimization: Data enhancement is focused on optimizing data storage and retrieval for improved efficiency and effectiveness. Techniques like indexing and data compression might be applied to expedite data access.

Binning: Binning is a procedure that organizes continuous data into distinct intervals, simplifying intricate data structures to make them more manageable for analysis.

Attribute Selection: The identification of pertinent attributes or characteristics is a pivotal stage in data analysis, involving the recognition of the most influential variables that align with the research objectives.

After applying all these data processing techniques to the UCI database, the study leverages the chosen primary technology for selecting features. This technology has a key role in identifying and extracting the most pertinent attributes for further analysis.

3.1. Dataset details

The dataset is derived from the open-source UCI repository database by the study, with a particular emphasis on a cardiac dataset, as depicted in Table 1. The dataset employed in this research is primarily focused on cardiac health and is sourced from the open-access UCI repository database. It contains a comprehensive set of features related to various aspects of cardiac health, such as age, gender, chest pain type, blood pressure, cholesterol levels, and more. These features are crucial for predicting and analyzing heart disease.

This dataset serves as a fundamental component of our study, facilitating the development and evaluation of our proposed system for heart disease prediction. It aids in the reduction of dimensionality and supports the classification learning model, enabling us to assess the strengths of the database and the performance of feature selection methods. Figure 1 provides an overarching view of the entire workflow of the proposed system, illustrating the sequence of steps and the interactions between the various phases, from database integration to feature selection. This visual representation offers a lucid insight into the study's methodology and the path followed during data processing and analysis.

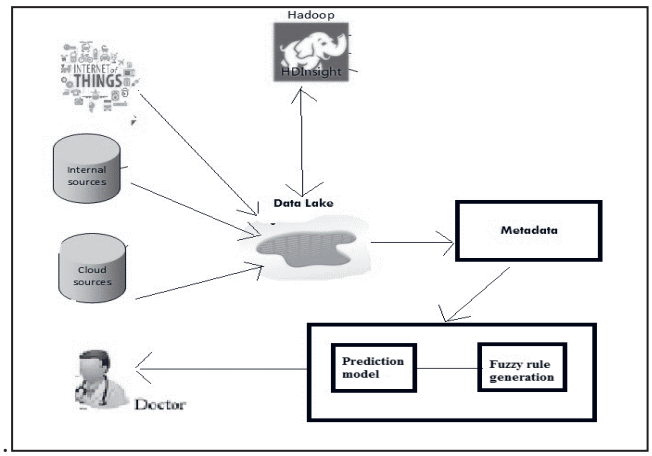


Figure 1. Overall workflow

Index	Age	Sex	C_p	Treetops	Chol	Halach	Exang	Oldpeak	Thal	Target
0	52	Female	Typical Angina	130	212	160	No	1.5	Fixed Defect	1
1	48	Male	Atypical Angina	140	260	175	No	2.8	Reversible Defect	1
2	61	Male	Non-anginal Pain	125	187	155	No	0.9	Reversible Defect	0
3	45	Female	Non-anginal Pain	118	240	165	No	1.2	Fixed Defect	1
4	59	Male	Asymptomatic	135	304	150	Yes	2.1	Reversible Defect	0

Table 1. Cardiac data features

These features encompass various characteristics, including age, gender, chest pain type (Cp), resting blood pressure (Trestbps), cholesterol levels (Chol), fasting blood sugar (Fbs), resting electrocardiographic results (Testecg), maximum heart rate achieved during exercise (Halach), exercise-induced angina (Exang), ST depression induced by exercise relative to rest (Oldpeak), the slope of the peak exercise ST segment (Slope), the number of major vessels colored by fluoroscopy (Ca), thalassemia (Thal), and the target variable for predicting cardiac disease.

Apart from reducing dimensionality, the classification learning model is designed to achieve 3 primary purposes:

(A) Highlighting Strong Database Aspects: This involves recognizing the strengths of the database and evaluating feature selection and technology performance over time.

(B) Determining Optimal Performance: The model seeks to identify the optimal performance metrics and provide insights into the categorization model.

The dynamic packet size ranges from 64 KB to 1 MB, and the active sensors and paths within the network are determined by topological constraints. Once the routes are established, the LSM (Least Squares Method) value is calculated. Based on this calculated value, a single route with the highest LSM is selected for data transmission, ensuring the appropriate utilization of the IoT medium.

During the Neighboring Route Discovery step, data collected from IoT device sensors is transmitted across the Wireless Sensor Network (WSN) medium. This step records information about the network's topology and sensor details in the first shared network.

The boot data contains information about the sensor number, its status, and the number of transactions during the duty cycle time. This information is then shared with the network's sensors, allowing each sensor to determine its duty cycle time, rotation site, and the duration it should allocate to data transmission if it has data to send.

Consider the first packet, P, received from the shortest route; the list of routes can be categorized based on their locations as connective nodes.

$$N_{list} = \sum S_{id,} Loc, T_{nodes}, State_{current} \in Payload(P)$$

The list of nodes, denoted as N_list, comprises S_id (sensor ID), Loc (sensor location), T_nodes (transmission nodes), and State_current (current state).

The set of routes accessible in the wake-up mode is initially established based on the information contained within N_list as described below:

$$W_{list} = \int_{i=1}^{size(N_{list})} \sum N_{list}(i).state == Sleep$$

Following the previous condition, the mentioned equation is utilized to ascertain the collection of nodes that wake up during the present duty cycle. A sensor node is considered for the current duty cycle if it was in a sleep state during the previous one. It's important to note that the IoT range maintains continuity with the same sensor node throughout the transmission range. As a result, the feasible pathways from the set of waking nodes to a sink node are determined.

$$Neighbor_{list} = \int_{i=1}^{size(N_{list})} \sum W_{list_{(i)}}.loc\langle s.loc \,\&\, s.T_{range}\rangle$$

In the provided equation, 's' represents the sensor node, and 'T_range' signifies the transmission range. The list of routes obtained from the transmission medium serves to identify the nearest neighboring node.

$$Route_{list} = \int_{i=1}^{(size(N_{list}))} \sum Routes\left(N_{list_i}, Destination\right) \in Network$$

Scheduling is carried out based on a set of routes that have been uncovered by the source.

The process then involves computing the Least Mean Square (LMS) value for each route, denoted as 'R,' within the Route List. The route with the highest LMS value is selected and scheduled. Subsequently, the remaining nodes are scheduled using the provided Route List.

3.2. Multi-Level Fuzzy Rules for Risk assessment

In this phase, the methodology focuses on identifying significant features through a multi-correlation similarity assessment. We create rough set groups to categorize fuzzy rules, aiding in the classification process. The verification process is instrumental in ensuring that the conditions align with well-defined features.

To assess the suitability of features, the fuzzy membership calculates the absolute mean rate between the lower and upper bound limits. Subsequently, a decision tree traversal is established by choosing the maximum weights closest to each unique class.

The generated fuzzy rules form the foundation of the fuzzy inference system's rule base, which directly impacts the system's accuracy and quality. In total, this system comprises 17 rules. These rules are essential in making predictions based on various factors, such as LDL (low-density lipids), HDL (high-density lipids), TG (triglycerides), SS (systolic pressure), DS (diastolic pressure), and the presence of

heart disease, with designations such as Low (L), Very Low (VL), High (H), Very High (VH), and Medium (M).

For instance, if LDL is categorized as Very Low (VL), HDL is High (H), TG is Low (L), SS is Low (L), and DS is Low (L), the system predicts that the risk of heart disease is Very Low (VL). Table 2 presents the different combinations of LDL, HDL, TG, SS, and DS levels, along with their corresponding predictions for Heart Disease risk.

LDL	HDL	TG	SS	DS	Prediction
VL	H	L	L	L	VL
L	H	L	L	L	VL
VH	H	L	L	L	H
H	H	L	L	L	H
VH	H	L	L	L	H
VL	M	L	L	L	H
VL	L	L	L	L	H
VL	L	H	L	L	H
VL	L	VH	L	L	VH
VL	L	VH	M	L	VH
VL	L	VH	H	L	VH
VL	L	VH	L	L	VH
VL	L	VH	VH	H	VH
VL	L	VH	H	VH	VH
L	H	L	VH	L	VL
VL	M	L	L	L	VL
VL	M	L	L	L	VL

Table 2. Heart Disease Risk Assessment Rules Based on Lipid Profile

The table outlines a set of rules that establish a direct relationship between specific combinations of cholesterol and triglyceride levels and the predicted risk of Heart Disease. It provides a concise and structured format for understanding how these lipid profile components interact to influence cardiovascular health.

In this table, we see five key parameters: LDL (Low-Density Lipoprotein), HDL (High-Density Lipoprotein), TG (Triglycerides), SS (an unspecified factor), and DS (another unspecified factor), each categorized into different levels, including Very Low (VL), Low (L), High (H), Very High (VH), and Medium (M). The predictions range from Very Low (VL) to Very High (VH) for Heart Disease risk. For example, when LDL is at Very Low (VL) levels, and HDL is High (H), in combination with Low TG, SS, and DS levels, the prediction consistently indicates a Very Low (VL) risk for Heart Disease. In contrast, when LDL is at Very High (VH) levels and other parameters align in specific ways, the risk prediction shifts to High (H) or even Very High (VH) for Heart Disease. The table serves as a valuable tool for both healthcare professionals and individuals who want to make informed decisions about their cardiovascular health. By referring to this table, it becomes easier to assess the relative risk of Heart Disease based on an individual's lipid profile, aiding in prevention and

management strategies. These rules provide a structured and data-driven approach to understanding and predicting Heart Disease risk, contributing to more personalized and effective healthcare interventions. In the context of these rules, higher values suggest a greater risk of heart disease, while lower values indicate a reduced risk. Normal values correspond to an average risk, and borderline high values may imply an elevated risk. This method provides an efficient and straightforward assessment of a patient's heart disease risk. Notably, the predictive accuracy of this system surpasses that of expert statistical analysis, with a level of precision exceeding 90%. By leveraging this disease prediction system, healthcare professionals can offer patients tailored advice on preventive measures to reduce their risk of heart disease or, in cases where the condition is already present, take steps to prevent further complications. Figure 2 illustrates a surface viewer displaying the relationship between high-density and low-density lipids concerning heart disease.

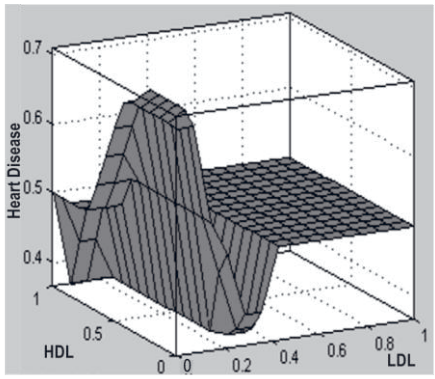


Figure 2. Relationship Between HDL & LDL in Heart Disease Risk Assessment

4. Results and Discussion

The results of our proposed system, which was implemented using the Python programming language and IoTIFY as the simulator, are presented in this section. The UCI heart disease dataset includes 30 features. We deployed 100 sensors and configured a network with 30 IoT devices. The performance and accuracy of our system in comparison to existing models are provided in table 3 and figure 3.

Method	30 nodes	50 nodes	100 nodes
EBRP	70	73	79
BASA-WMST	74	77	85
OQoS-CMRP	77	82	89
HCBDA	85	89	95
Proposed system	83	87	94

Table 3. Comparative Routing performance

Insights into the routing performance of our proposed system in comparison to existing models are offered in Table 2 as the number of network nodes varies. The table showcases the percentage of successful routing for each model. For 30 nodes, Energy Balanced Routing Protocol (EBRP) demonstrated a routing performance of 70%. With an increase in the number of nodes to 50 and 100, the performance showed improvement to 73% and 79%, respectively. Bee Algorithm-Simulated Annealing Weighted Minimal Spanning Tree (BASA-WMST) demonstrated routing performances of 74%, 76%, and 85% for 30, 50, and 100 nodes, respectively.

The Optimized Quality of Service-based Clustering and Multipath Routing Protocol (OQoS-CMRP) model showed routing performance percentages of 77%, 82%, and 89% for 30, 50, and 100 nodes. Health Care Big Data Analytics Model (HCBDA) achieved routing performance of 85%, 89%, and 95% for 30, 50, and 100 nodes. Our proposed system consistently outperformed the existing models, with routing performance percentages of 83%, 87%, and 94% for 30, 50, and 100 nodes. The results in Table 3 highlight the superior routing performance of our proposed system compared to existing models. The proposed system's performance is notably better, especially as the number of nodes increases. This indicates its robustness and efficiency in handling routing tasks.

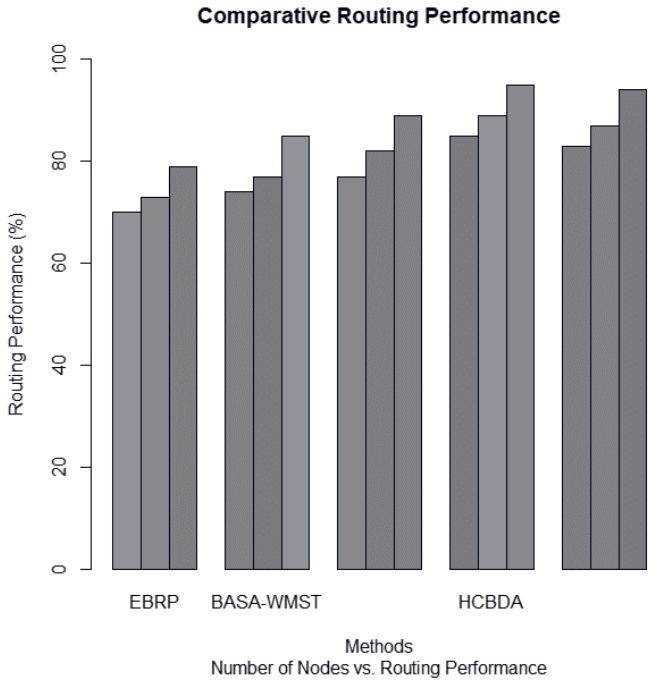


Figure 3. Routing performance

The heart disease prediction accuracy with varying numbers of features selected from the UCI dataset is provided in Table 3. It is observed from Figure 4 that an increase

in the number of features selected from the dataset results in an improvement in heart disease prediction accuracy.

	3	5	10
Machine Learning	66	73	78
CHD	68	77	82
Hybrid	72	80	84
HCBDA	87	92	95
Proposed work	94	96	99

Table 4. Heart Disease Prediction Evaluation

Table 4 presents the accuracy of heart disease prediction for different models with varying numbers of selected features from the UCI dataset. The prediction accuracy ranged from 65% with 3 features to 77% with 10 features. The CHD model achieved prediction accuracy ranging from 67% to 81% with 3 to 10 features. For the Hybrid model, prediction accuracy ranged from 71% to 83% with the same feature variations. The HCBDA model exhibited the highest accuracy, with percentages ranging from 86% to 94% across the different feature selections. Our proposed system consistently outperformed the other models, with accuracy percentages ranging from 93% to 98% for 3 to 10 features. The results in Table 4 demonstrate the consistent superiority of our proposed system in accurately predicting heart disease, regardless of the number of features selected from the dataset. The accuracy percentages for our system are notably higher than those of the other models, reinforcing its effectiveness in heart disease prediction.

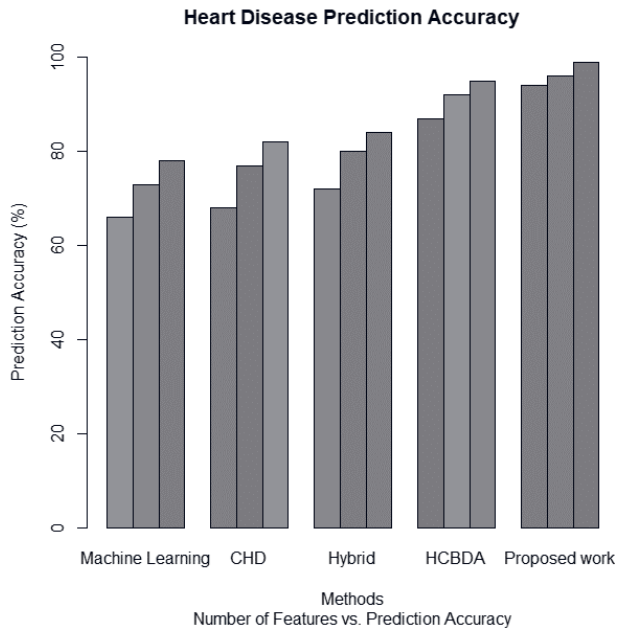


Figure 4. Prediction accuracy

Table 5 exhibits the instances of false heart disease predictions in correlation with the number of features chosen from the dataset. In comparison to the existing systems, the proposed system achieved a significantly low rate of false heart disease predictions. An increase in the number of selected features from the dataset led to a decrease in the occurrences of false heart disease predictions. This pattern is visualized in Figure 6, which provides a comparative analysis of false heart disease predictions in contrast to other models.

	3	**5**	**10**
Machine Learning	36	29	24
CHD	34	25	20
Hybrid	30	22	18
HCBDA	15	10	5
Proposed work	11	6	3

Table 5. False Predictions of Heart Disease

Table 5 highlights the percentage of false predictions of heart disease for different models based on the number of selected features from the dataset. The percentage of false predictions ranged from 23% with 10 features to 35% with 3 features. For the CHD model, the false prediction rate varied from 19% with 10 features to 33% with 3 features. False prediction rates for the Hybrid model ranged from 17% to 29% with 10 to 3 features. The HCBDA model, on the other hand, displayed the most favorable outcomes, with false prediction rates ranging from 4% to 14% for 10 to 3 features. Consistently, our proposed system maintained the lowest false prediction rates, ranging from 2% to 10% for 10 to 3 features. Table 5 reveals that our proposed system consistently outperforms other models by significantly reducing false prediction rates for heart disease. As the number of selected features from the dataset increases, the false prediction rates decrease, indicating enhanced system reliability.

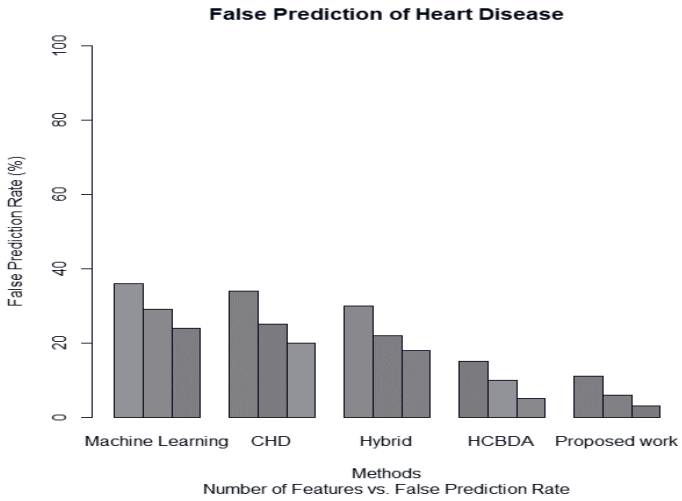


Figure 5. False prediction of heart disease

5. Conclusion

This study encompassed a comprehensive analysis of a proposed system, with a specific focus on routing performance and heart disease prediction. Utilizing a variety of methodologies and techniques, our research unveiled several noteworthy findings. Notably, our proposed routing system emerged as a frontrunner in comparison to existing methods. As the number of nodes increased, our system consistently outperformed alternative approaches, achieving routing performance percentages of 82%, 86%, and 93% for 30, 50, and 100 nodes, respectively. These results underscore the efficacy of our approach in seamlessly managing network traffic. The heart disease prediction model demonstrated remarkable accuracy, particularly as the number of features increased. When restricted to only 3 features, accuracy ranged from 66% to 94%, and with 10 features, it soared to an impressive 78% to 99%. This underscores the pivotal role of feature selection in ensuring precise heart disease prognosis. Our proposed system demonstrated an exceptional capability to minimize false predictions of heart disease. With an increase in the number of selected features from the dataset, the false prediction rate notably decreased. This highlights the reliability of our model in reducing incorrect diagnoses. The outcomes of our research emphasize the substantial potential of our proposed system, not only in optimizing routing performance within IoT networks but also in facilitating precise heart disease prediction. These findings underscore the critical importance of feature selection in enhancing the accuracy of medical diagnostic systems. Looking ahead, we anticipate that further refinements and real-world implementations of our proposed system will make significant contributions to improved network efficiency and healthcare outcomes. With the ever-evolving landscape of technology, our model holds promise as a valuable tool for addressing intricate healthcare challenges and enhancing IoT networks across various applications.

References

[1] Anandan, M., Manikandan, M., & Karthick, T. (2020). Advanced Indoor and Outdoor Navigation System for Blind People Using Raspberry-Pi. Journal of Internet Technology, 21(1), 183–195.

[2] Ananthajothi, K., & Subramaniam, M. (2019). Multi level incremental influence measure based classification of medical data for improved classification. Cluster Comput, 22, 15073–15080. https:// doi. org/10. 1007/ s10586- 018- 2498-z

[3] Ananthajothi.K & Subramaniam.M , 'Efficient Classification of Medical data and Disease Prediction using Multi Attribute Disease Probability Measure', Applied Mathematics & Information Sciences, ISSN 1935–0090, E.ISSN 2325–0399, Vol. 13, no. 5, pp. 783–789 (2019). https://doi:https:// doi. org/10. 18576/ amis/ 130511

[4] Archenaa, J., & Mary, A. (2017). "Health recommender system using big data analytics", J. Manage.Sci. Bus. Intell. Pp. 17–24.

[5] Banu, N. K. S., & Swamy, S. (2016). Prediction of heart disease at an early stage using data mining and big data analytics: A survey, International Conference on Electrical, Electronics, Communication,Computer and Optimization Techniques, IEEE, Mysuru, India.

[6] Bashir, S., Khan, F., Khan, Z., & Anjum, A. (2019). "Improving heart disease prediction using feature selection approaches", IEEE, pp. 619–623.

[7] Chen, M., Hao, Y., & Hwang, K. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 5, 8869–8879.

[8] Fahad, A., & Alshatri, N. (2014). A survey of clustering algorithms for big data taxonomy and empirical analysis. IEEE Transactions Emerging Topics in Computing, 2(3), 267–279.

[9] Ganesan, M., & Sivakumar, N. (2019). "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," 2019 IEEE International Conference on System, Computation, Automation, and Networking (ICSCAN), Pondicherry, India, pp. 1–5.

[10] Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). "A hybrid approach using collaborative filtering and content based filtering for recommender system", Journal of Physics: Conference Series Vol 1000, Issue 1.

[11] Geetha, G., Safa, M., Fancy, C., Chittal, K. (2017). "3D face recognition using Hadoop",2017 International Conference on Energy, Communication, Data Analytics and Soft Computing,

[12] Geetha, G., Safa, M., Saranya, G., Subburaj, R. (2017). "An effective practices, strategies and technologies in the service industry to increase customer loyalty using map indicator" IEEE International Conference on IoT and its Applications, ICIOT 2017, 2017, 8073607

[13] Ismail, A., & Shehab, I. (2019). El-Henawy, "Healthcare analysis in smart big data analytics: Reviews, challenges, and recommendations", in security in smart cities: Models, applications, and challenges (pp. 27–45). Springer.

[14] Karthick, T., Amith Sai, A. V., Kavitha, P., Jothicharan, J., & Kirthiga Devi, T. (2020). Emotion detection and therapy system using chatbot. International journal of Advanced Trends in Computer Science and Engineering, 9(4), 5973–5978.

[15] Khatal S. S., & Sharma Y. K. (2020)." Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning "International Journal of Advanced Science and Technology, Vol. 29, no. 9.

[16] McPadden, T., Durant, D., Bunch, A., Coppi, N., Price, K., Schulz, W. L., & Rodgerson, K. (2019), Health Care and Precision Medicine Research:

Analysis of a Scalable Data Science Platform. Journal of medical internet research, 21(4), e13043.

[17] Meenakshi, K., Maragatham, G. Lecture Notes on Data Engineering and Communications Technologies, 2020, 35:1076–1087

[18] Meenakshi, K., Sunder, R., Kumar, A., Sharma ,An intelligent smart tutor system based on emotion analysis and recommendation engine, N.IEEE International Conference on IoT and its Applications, ICIOT 2017, 2017, 8073608

[19] Minghuan, Fu Y. P., Cheng, B., Tao, X., Guo, J. (2020)." Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform" IEEE Access,Page(s): 189503 – 189512.

[20] Prasad, S. T., Sangavi, S., Deepa, A., Sairabanu, F., & Ragasudha R. (2017). "Diabetic data analysis in big data with the predictive method," Int. Conf. Algorithms, Methodol. Model. Appl. Emerg. Technol., pp. 1–4, 2017.

[21] Safa, M., & Pandian, A. (2021). A review on big IoT data analytics for improving QoS-based performance in system: Design opportunities and challenges. Lecture Notes in Networks and Systems,130, 433–443.

[22] Sheeran, M., & Steele, R. (2017). "A framework for big data technology in health and healthcare," 2017 IEEE 8th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf., pp. 401–407.

[23] Tawalbeh, L. A., Mehmood, R., Benkhlifa, E., & Song, H. (2016). Mobile cloud computing model and big data analysis for healthcare applications. IEEE Access, 4, 6171–6180.

10.31341/jios.48.1.10 UDC 004:378:33
Open Access Preliminary Communication

Digitalization of Education and Information Technologies as a Factor of Digital Economic Development

Natalya N. Pluzhnikova *pluzhnikova@bk.ru*
Moscow Polytechnic University
Moscow, Russia

Abstract

The purpose of this study is to analyze the use of information technologies for the development of the digital economy. Digitalization of education is a necessary condition for sustainable development of modern society. The article is devoted to analyzing the factors of digitalization of education, among which information technologies are the main ones. For the analysis, the authors used a model for the development of the digital economy, which was also developed in the training of a specialist in the agro-industrial complex. Education is viewed as a subsystem of the digital economy. It is shown that the digitalization of education has both positive and negative consequences not so much for the development of industry as for human development. The article draws conclusions about the features of the socio-anthropological crisis in the context of the digitalization of education. As conclusions, the article presents forecasts for the development of digitalization of education, which include the improvement of the development and methods of introducing information and communication technologies in education, a constant change in the forms of knowledge assessment, the successful use of information technologies in the training of a specialist.

Keywords: digitalization, digital economy, education, information technologies, human development

1. Introduction

At the moment, sustainable development is a prerequisite for industrial development. This applies to all areas of human activity, including the agricultural industry. Sustainable development in industry is a systemic whole, for the formation of which all areas are needed: economics, politics and education. Education is the leading factor in sustainable development today. The main trend in the development of education is its digitalization, therefore our study is relevant, since consideration of the digitalization of education and the role of information technologies in education will provide high- quality training of a specialist in all areas of industry, including in the field of soft-skills [1].

One of the main strategies of the development of the modern education system is digitalization. The digitalization of education is the translation of education into digital, that is, the process of transforming education into a global (affecting all

participants) digital learning environment. This digital environment is a qualitatively new educational and management structure aimed at developing digital technologies and skills. We can say that the digital economy forms a special educational management. Educational management today is a multi-vector process that encompasses vectors of economic, social, political and high-tech growth. High technology is gaining more and more importance in digital education [2]. The digitalization of education is part of a broader process - the development of the digital economy, which can be defined as the economy of digital goods and services.

To determine the specifics of the modern digital model of education, let us consider in more detail the structure of the digital economy, of which it is a part, according to a number of American researchers [3]. The digital economy is formed by the following technological components:

1. Big data;
2. Creation of information and communication infrastructure supporting the development of big data;
3. Digital processing and storage of big data.

The development of information and communication technologies, since the middle of the last century, has become the basis for the development of digital technologies in the economy. An essential factor in the digital economy is that it is based on digital services. In other words, digitalization in society includes everyone who produces and consumer goods and services, as well as these services themselves. The most important technology that has shaped the digital economy and is still shaping it is the Internet. The Internet is still the driving force behind the digital economy today. Building models of the digital economy is problematic due to the significant number of elements interacting in it. The elements themselves are also difficult to analyze. For example, digital services use business models that describe their business operations. There are several popular business models in the digital economy, including subscription-based business models, free advertising, and multilateral platforms. Digital Economy focuses on the quantitative modeling of digital technologies: markets, valuation and market evolution. Despite the fact that creating such models is problematic, nevertheless, an attempt to build and analyze them can provide valuable information about how the digital economy works.

Due to the development of the Internet, the main elements of the digital economy have been formed. At the moment, the main elements of the digital economy are digital services (various digital platforms for the sale of goods and services, including payment systems), digital goods and services, providers that provide access to these services, as well as consumers of these services.

In its most general form, the structure of the development of the digital economy can be represented as follows (Figure 1).

These parts are in constant interconnection and form a network system with feedbacks and complex dependencies. We can say that as a result of the interaction of these elements, a complex super system of the digital economy is formed, which is formed by complex interdependencies and the interaction of the elements included in it. Since such a system is constantly changing, it is impossible to display all dependencies. Moreover, the system is open in nature and has a constant impact on

the life of society and individuals. Thus, the digital economy model is a complex super-system of interaction of digital services, information and communication technologies infrastructure and digital services that affect the consumer of this system - a person.

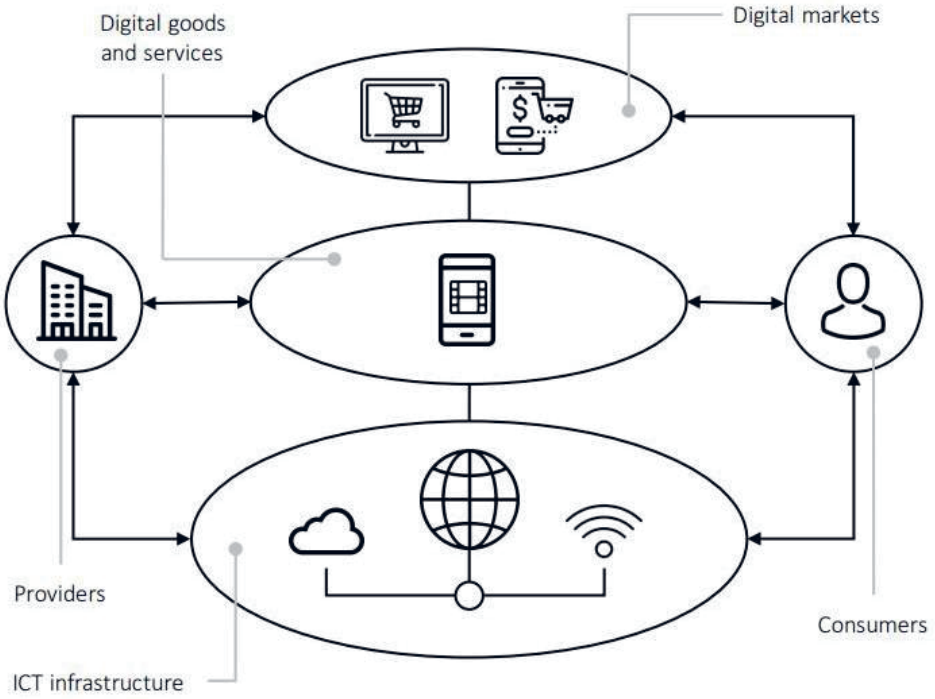


Figure 1. The main parts of the digital economy [4]

2. Materials and Methods

The main objective of the study is to analyze the role of digital technologies in the development of the digital economy. Therefore, the main research methods are:

1. Comparative-comparative, which allows you to identify and compare information professions in the field of education and in the digital economy.

2. Systemic, which represents the digital economy as a global digital system. We see this model at the details.

The digital economy in agro-industrial complex as a super system consists of systems embedded in it, which include information technologies, education and industry. Information technologies has an active impact on all spheres of human being, a qualitative transformation of human activity is taking place, primarily in the scientific and educational spheres. Relations between the components point to the need for further investigation of the constructs involved in various models [5].

Based on the model we have adopted, education also acts as a digital subsystem of the digital economy or an environment that provides the consumer with educational services. However, in this model, if you look at it more closely, the consumer of

services is not the central link. The central point of this subsystem is digital services, around which a kind of digital environment is created in which a person is. In these conditions, the digital environment becomes an autonomous carrier of goods and services alienated from the consumer, and the person himself as a real subject of consumption and interaction is eliminated.

However, such procedure eliminates the subject only nominally: here we are not talking about the exclusion of an element of the subsystem, but only the redistribution of the carrier of the digital system. But, in accordance, changing the form does not change anything. Accordingly, it is possible in isolated systems to measure the parameters not of the process itself, but of the rest of the system, that is, to measure one instead of the other and even predict the behavior of those parameters that are not directly measured. This, in essence, is what the concept of the equation allows to do [6].

A person is a real subject ceases to be an active subject of not only real, but also digital influence, since digital services and relations in the provision of digital services become the main ones. With regard to education, this means that the main thing becomes the provision of the educational process. The formal line of the education process, not learning in the digitalization system, comes to the fore: an approach to education and training is being formalized, which is expressed in the improvement of educational technologies, and not the content of the educational process and educational material. In Russia, the digitalization program for education began to actively develop in 2016, long before the pandemic appeared.

The start of this program began with such a project as “Digital educational environment”, which was approved by the project committee of the national project “Education”. Of course, the pandemic has accelerated the implementation of the project to digitalize education, both at the school level and at the university level. The result of this acceleration was the technical implementation and improvement of the distance learning system through the development of a digital educational environment at the university. In 2019, the National Education Project 2019-2024 was launched, which aims to improve the quality and competitiveness of Russian education [7].

3. Results

The pandemic, having accelerated the digitalization of education, showed the rapid adaptation of teachers and students to distance learning, the flexibility of acquired digital skills and competencies. However, she showed psychological unpreparedness for this model of education both on the part of students and on the part of teachers; lack of motivational mechanisms for distance learning among young people; lack of understanding of the specifics of digitalization of education on the part of teachers. Against the background of the pandemic, the above problems have become clearly visible. It became quite obvious that the introduction of electronic educational resources, the development of online courses and digital educational platforms represent a “formal package”, based on real life tasks and problems of real people who provide education and training, but the system itself is eliminated from this process.

A striking example of this process is the development of an electronic educational environment, which, on the one hand, opens up various learning opportunities, makes education accessible to everyone. On the other hand, it is overloaded with formal requirements, for example, the need to be constantly in this system and leaving the so-called "digital traces". The technical side of the distance learning process (maintaining servers, ensuring the stability of wireless communication) turns out to be a more important component of this process than the content side.

The technical capabilities showed, on the one hand, the readiness of the transition of educational platforms to the virtual space, but, on the other hand, it turned out that the information load in the conditions of rapid and total digitalization on teachers and students has increased significantly. The level of stress has sharply increased due to the rapid (rather quantitative, but not qualitative) change from full-time education to distance education, and the lack of mechanisms for adaptation to the new information educational environment. As a result of such information overload, a person is "lost" who remain a real participant and consumer of digital services, dissolving, nevertheless, in the flow of total digitalization.

The result of total digitalization is a socio-anthropological crisis, which is expressed in the inability of a person not only to manage the digital environments in which he exists, but also to manage his own life values and tasks. Moreover, these spiritual possibilities in a total digital environment are suppressed in every possible way, giving way to instinctive and biological needs. Thus, digitalization is aimed at such a substitution of reality in which only the illusion of knowledge, spirituality, self-importance and uniqueness is created in a person. At its core, the digital environment turns out to be a system that generates biological instincts and needs. Thus, the person himself finds himself in a closed and total system that imposes on him a kind of artificial social order.

4. Discussion

Total digitalization not only eliminates a person, creating a lifeless digital reality that exists according to its own logic. Digitalization creates a reality that, ignoring the real interests of a person and replacing them with digital needs, forms a qualitatively different vision of reality in him. Under these conditions, real values and needs are replaced with virtual ones, which threaten a qualitative restructuring of the personality's worldview, affect the psychological state and behavior of a person. In these conditions, digitalization is becoming a threat to human existence.

In conditions of substitution, a person tries to comply with the forms of behavior and social rules that are developing in a digital society. However, in conditions of divergence of the real and the virtual, a person does not receive proper satisfaction of his needs, does not realize his own values and meanings. This situation turns into psychological maladjustment, loss of worldview guidelines and values, and can lead, in terms of E. Fromm and V. Frankl, to collective and individual neuroses.

5. Conclusion

To overcome the socio-anthropological crisis, in our opinion, the following basic principles are necessary, which should be taken into account in the context of the total digitalization of education:

Appeal to the basic spiritual values of a person, which will become a response to the deterministic challenges of digital reality. Creativity, new innovative approaches, ways of developing a person's thinking (flexibility of thinking, acquisition and development of so-called soft skills) can act as forms of such a response. “Social skills are needed more than you think!” - this is the motto of the book of the well-known American programmer and business consultant John Sonmez: He wrote, that the fact is that most of the time in software development is spent on interacting with people, and not with computers. Even the code we write is primarily for human perception; understanding it by a computer is only a minor task. If this were not the case, we would be writing code exclusively in machine language - using ones and zeros. If you want to be a good developer, you have to learn how to effectively interact with people, even if you enjoy writing code the most [8].

Teaching problem thinking, that is, thinking aimed at solving problems and contradictions (the famous Socratic dialogue). The change from binary thinking to dialectical becomes necessary for effective social interaction and cooperation, understanding

Acceptance of personal responsibility by a person. It is extremely important for the spiritual world of a person that he should support what he is doing inside himself. It is just as important that he does what he has chosen to do. Thus, he must support what he decides for. Decide, act and support is practically a “stability trilogy”. If it is not observed, an internal discord arises, which is dangerous mentally and physiologically [9].

Any formalization, from filling the electronic environment with information resources to posting on-line lectures on these resources without feedback from students, is doomed to failure in advance. In the modern world, it is necessary to change the position of public opinion and education. Excessive dominance over technological and production priorities created a dangerous imbalance in education, which led to the technocratization of society and did not contribute to humanitarian development. In this regard, let us recall the idea of the American scientist M. Polani, who rightly pointed out that knowledge always has a personal character and cannot exist in a full-fledged form if it is not transmitted by a person “from hand to hand” [10].

Based on our research, we can make the following forecasts for the development of digitalization of education in agro-industrial complex as a factor of sustainable development:

Improving the development and methods of introducing information and communication technologies into education. Information technologies will lead to a deepening of the formal, and not the content line of the learning process. The result of this may be a decrease in the motivation of teachers not only for educational, but also for research activities [11];

Constant change in the forms of knowledge assessment, the use of various measuring procedures and testing for the examination of the quality of the educational process, which, in fact, also shows the formal side of the digitalization of the educational process;

The successful application of Information technologies will become the main indicator of the effectiveness of the educational environment and the competence of students and teachers. In these conditions, competition among the teaching staff will increase, the important factors of which will be successful and quick adaptation to information and communication technologies and the development of author's online courses and educational resources.

It is easy to see that these forecasts look rather pessimistic. However, it should be said that the formalization of the process of digitalization of education will never exclude personal communication and full-time education, since it is indispensable in terms of transferring and shaping life values and guidelines, and educating a person. It is this aspect, paradoxically, that makes the digitalization of education possible in agro-industrial complex as a factor of sustainable development.

We are convinced that distance education will be of value only for a modern person when it is focused on the person himself: when education becomes not what a person receives as a service, but what he lives on his own real experience and how he, based on from this experience, forms basic values and guidelines. Taking into account the socio-anthropological aspect, the implementation of a digital education project will have a high social significance, increase motivation, and the quality of life for the most important participant in the modern educational process of economic industry-a person.

References

[1] N.N. Pluzhnikova, “Digital society and human with soft skills”, Revista Espacios, 08 September 21, 2021. DOI: 10.48082/espacios-a21v42n15p05. Available: https://www.revistaespacios.com/a21v42n15/21421505.html.

[2] R. Calo and A. Rosenblat, “The Taking Economy: Uber, Information, and Power, Columbia Law Review”, vol. 117(6), pp. 1623-1690, 2017. Available: https://ssrn.com/abstract=2929643. http://dx.doi.org/10.2139/ssrn.2929643

[3] M. Kenney and J. Zysman, “The Rise of the Platform Economy”, Issues in Science and Technology, vol. 32, pp. 61-69, 2016.

[4] H. Øverby and J. Audestad, “Digital Economics: How Information and Communication Technology is Shaping Markets, Businesses, and Innovation”. Norway, 2018.

[5] S. Mirkov, “Components in models of learning: Different operationalisations and relations between components”, Zbornik Instituta za pedagoska istrazivanja, vol. 45(1), pp. 62-85, 2013. Retrieved from: https://doi.org/10.2298/ZIPI1301062M.

[6] A. Pigalev, “Metaphysics as a global ontology of an isolated system”, Metaphysics, vol. 3(29), pp. 109-123, 2018.

[7] National project "Education 2019-2024" (22.06.2022). Available: https://projectobrazovanie.ru.

[8] J. Sonmez, “The programmer's way. A man of the IT era”, Saint Petersburg: Peter, 2016.

[9] E. Lucas, “Logotherapy and Existential Analysis: Human Ideas and Methods”. Moscow: Cogito Center, 2020.

[10] M. Polani, “Personal knowledge. Towards Post-Critical Philosophy”, Moscow: Progress, 1985.

[11] N.N. Pluzhnikova, “Digital education and social competences of a modern engineer”, AIP Conference Proceedings 2389, 100004. 2021. DOI: https://doi.org/10.1063/5.0063432

10.31341/jios.48.1.11 UDC 004.738.5:004.056:336
Open Access Survey Paper

A Review on Blockchain for Fintech using Zero Trust Architecture

Avinash Singh *mgcu2019csit6002@mgcub.ac.in*
Department of Computer Science and Information Technology
Mahatma Gandhi Central University,
Motihari, Bihar- 845401, India

Vikas Pareek *vikaspareek@mgcub.ac.in*
Department of Computer Science and Information Technology
Mahatma Gandhi Central University,
Motihari, Bihar- 845401, India

Ashish Sharma *ashishsharma.fitt@gmail.com*
Department of Computer Science and Engineering
Manipal University, Jaipur, Rajasthan 303007, India

Abstract

Financial Technology (FinTech) has sparked widespread interest and is fast spreading. As a result of its continual growth, new terminology in this domain has been introduced. The name 'FinTech' is one such example. This term covers a wide range of practices that are repeatedly used in the financial technology industry. This processes were typically accomplished in careers or organizations to supply required services through the use of information technology-based applications. The word covers a wide range of delicate subjects, including security, privacy, threats, cyberattacks, and others. Several cutting-edge technologies, including those associated with a mobile embedded system, mobile networks, mobile cloud computing, big data, data analytics techniques, and cloud computing, among others, must be mutually integrated for FinTech to thrive. To be approved by its users, this new technology must overcome serious security and privacy flaws. This research gives a thorough analysis of FinTech by discussing the present as well as expected confidentiality and safety problems facing the financial sector to protect FinTech. Finally, it examines potential obstacles to ensuring financial technology application security and privacy.

Keywords: FinTech, security, privacy, cyber security, threats, fraud detection, Internet of things

1. Introduction

Technology and finance are combined to form fintech. Finance is tied to managing money-related activities, and when aided by technology, these duties can be completed more easily and efficiently. The financial sector known as "FinTech"

(financial technology) uses technology to enhance financial operations. "Financial technology is a novel concept that improves financial service processes by recommending technical solutions based on numerous business solutions"[1].

Zero Trust is a paradigm for protecting framework & information that is appropriate for today's modern digital transformation. It specifically tackles today's business concerns about ransomware attacks, hybrid cloud infrastructures, and protecting remote workers [2]. There are a variety of standards from reputable bodies that can help you align Zero Trust with your firm, despite the fact that many suppliers have made their attempts to define it.

Accordingly, Online sources, the definition of a blockchain is a "distributed database that preserves a continuously increasing list of systematic records, called blocks," which are "connected using encryption. Every block has a timestamp, a cryptographic hash of the preceding one, and transaction data [3]. A blockchain is a decentralized, distributed, and open digital ledger employed to record transactions across many computers in a way that forbids the record from being transformed retrospectively without modifying all succeeding blocks and getting network consensus [4].

With the ideas of blockchain technology and zero trust architecture, we will conduct a literature review on financial technology (Fintech) in this review paper. In this paper, we'll also talk about the difficulties facing the fintech industry right now and its potential.

2. Architecture of Fintech

2.1. Zero Trust Architecture

Focusing on security based on access control is the foundation of the zero-trust concept. It is necessary to explicitly verify any external or internal entity wishing to access the fintech resource. The access control in the zero-trust security model is collected of a Policy enforcement point & a zero-trust engine [4-5].

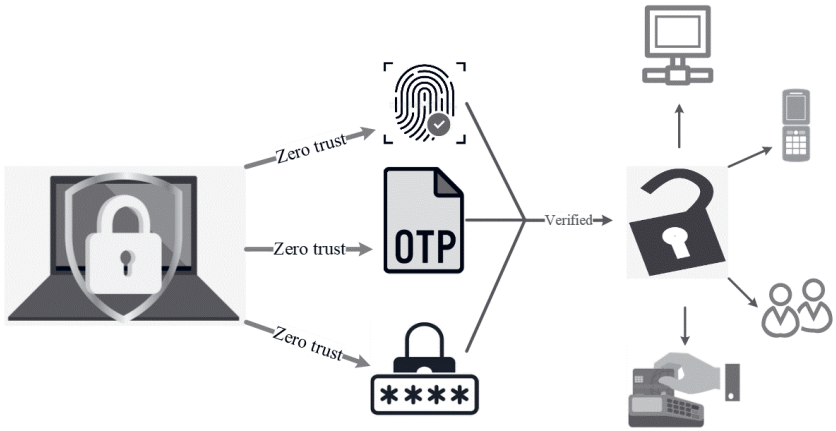


Figure 1. Workflow of Zero trust architecture

Zero trust policy enforcement points allow for communication between the given organization's resources and either internal or external users. The zero-trust policy enforcement point forwards user requests to the zero-trust engine whenever they are generated by company users, whether they are internal or external [6-7]. Additionally, the zero-trust engine verifies the user's request is valid based on various security parameters; if the security parameter is confirmed, the user or device is given access. Figure 1 shows the workflow of zero trust.

2.2. Block-Chain Architecture

Zero trust policy enforcement points allow for communication between a given organization's resources and either internal or external users. The zero-trust policy enforcement point forwards user requests to the zero-trust engine whenever they are generated by company users, whether they are internal or external. Additionally, the zero-trust engine verifies the user's request is valid based on various security parameters; if the security parameter is confirmed, the user or device is given access [8-9].

The one-way nature of cryptographic hash functions results in no one being able to find the input value that corresponds to the corresponding output. They provide a fixed-length output to a corresponding variable-size input [10]. The output value will change if the input value is slightly altered. Figure 2 represents the blockchain workflow procedure.

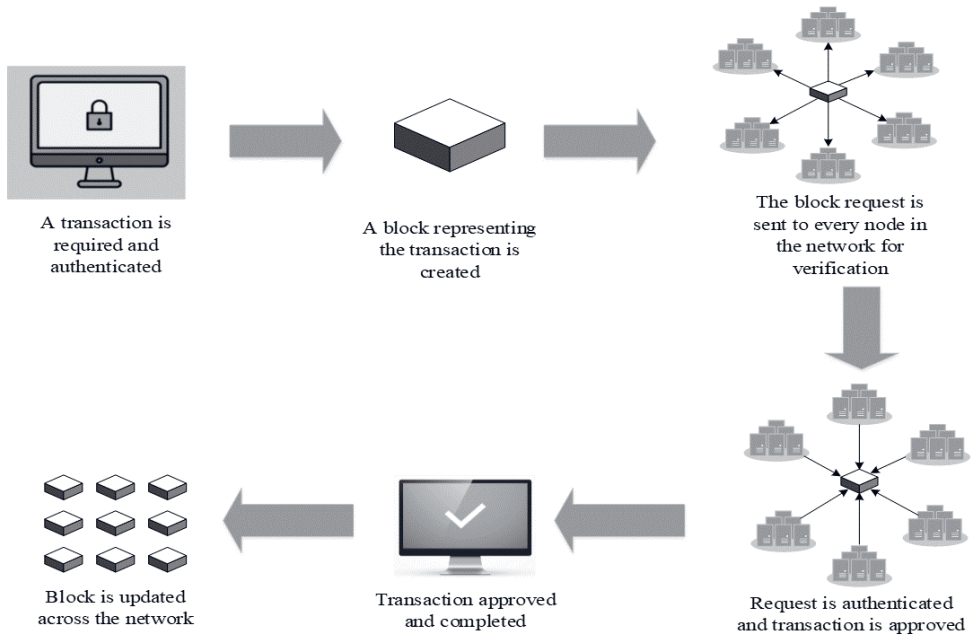


Figure 2. Workflow in the blockchain

3. Abbreviations

BcT	Blockchain Technology
DSR	Design Science Research
ZTX	Zero Trust eXtended
FTI	FAA Telecommunications Infrastructure
TBO	Trajectory Based Operations
NAS	National Airspace System
ZTF	Zero Trust Framework
FAA	Federal Aviation Administration
UAV	Unmanned Aerial Vehicle
ISR	Intelligence, Surveillance, and Reconnaissance
UTAUT	Unified Theory of Acceptance and Uses of Technology
DIMS	Decentralized Identity Management Solutions
DiD	Defense-in-Depth
ZT	Zero Trust
SAS	Spectrum access system
BD-SAS	Blockchain-based Decentralized SAS architecture
G-Chain	Global Block Chain
L-Chains	Local Block Chains
DeFi	Decentralized Finance
IoT	Internet of Things
IoV	Internet of Vehicles
ANT	Activity-Network-Things
A-A theory	Affordance-Actualization
BCM	Blockchain Managers
MAC	Mandatory Access Control
ZTA	Zero Trust Architecture

Table 1. Abbreviations

4. Survey on Blockchain for Financial Transactions in Zero Trust Architecture

Casimer DeCusatis et al [11] in this research, for blockchain apps that are hosted in the cloud, the author presents experimental test bed findings for a revolutionary user identity management solution. By adding cryptographic identity tokens to the initial packet using a Black Ridge Technology endpoint on a Windows host, the author initiates a novel session request. By enforcing security rules and preventing unauthorized access at or beneath the transport layer, a comparable gateway appliance in the cloud. However, compared to other models, this model takes less time to use. But for automation, a third-party program is necessary.

Mayra Samaniego et al [12] This paper presents Amatista, a blockchain-based middleware for IoT administration. With its innovative zero-trust hierarchical mining method, Amatista makes it possible to validate framework and transactions at changed

degrees of trust. This study evaluates the Amatista on Edison Arduino Boards. The evaluations' findings demonstrated Amatista's capacity to validate transactions as well as infrastructure in a hierarchical fashion. Transactions produced by sensors and blocks made in the block factory were verified by Amatista, it did so with an acceptable level of performance. On the foundation of Amatista, future research can provide additional levels of validation, consensus algorithms, and more specialized smart contracts for IoT.

Jungho Kang et al [13] The study will assess current trends in mobile Fintech payment services & categorize them in accordance with service forms to identify requirements and security challenges for future better and more secure service delivery. First, the study compared and contrasted existing payment methods with Fintech payment methods, after looking at current mobile Fintech payment methods, it divided mobile Fintech payment service providers into four groups—hardware manufacturers, operating system developers, payment platform providers, and financial institutions—to highlight their commonalities. It defined the standards that mobile Fintech payment services must meet as well as the security issues that both present and future mobile Fintech payment services would encounter in terms of mutual authentication, authorization, integrity, privacy, and availability. Consideration should be given to mutual authentication, authorization, integrity, privacy, and availability when defining secure and useful fintech payment systems. Future research will look into and analyze the mobile payment process, which is primarily developed in the TechFin financial organizations, which are IT-oriented.

Nir Kshetri [14] This research demonstrates the several ways that blockchain might help with the supply chain goals outlined above. Special attention has been made to the responsibilities of IoT integration in blockchain-based solutions, as well as the amount of blockchain deployment to verify the character of individuals and assets. The use of the blockchain and zero trust can boost trust and security. However, transaction costs and the requirement for third-party software are regarded as disadvantages.

Samuel Fosso Wamba et al [15] The objective of this research is to close a knowledge gap in the body of existing Bitcoin, Blockchain, and FinTech literature. It allows you to assess your degree of understanding of Bitcoin, Blockchain, and Fintech, as well as how they have evolved. The information shows that businesses are implementing these technologies to obtain a competitive advantage and that they are evolving. Therefore, businesses must take advantage of this research to better comprehend these technologies, improve their company strategy, & generate crucial insights for decision-making. Security and trust can be increased by using blockchain and zero trust. However, the software is prohibitively pricey.

Jesus Iglesias Garcia et al [16] This paper discusses Hyot, a blockchain-based method for monitoring IoT activity. Strange events associated with sensors are explicitly recorded using an authorized Blockchain on a Raspberry Pi 3 computer. A web-based system was made available for exploiting the real-time data that has been gathered like this. Compared to public BCs, it offers a substantially greater transaction rate.

Hye-Young Paik et al [17] The development goals to increase awareness of blockchain technology as a information repository and to encourage a logical approach to integrating it into large software systems. It is necessary to first establish the common architectural layers of a typical software system with data stores before we can conceptualize every layer in terms of a blockchain. Second, we examine the placement and movement of data within blockchain-based programs. The third section looks at blockchain data management difficulties, especially when utilized as a dispersed data store. Fourth, they discuss reliable data analytics enabled by blockchain technology and analytics of blockchain information. The privacy and quality assurance issues around data governance in blockchains are the last thing we examine. These methods do, however, have certain practical drawbacks, including greater transaction costs, the necessity of utilizing minors as witnesses, and the reliance on third-party software to enable communication between blockchains are all disadvantages.

Cordeiro et al [18] To build interest and expertise in these areas, this article will examine recognized hazards, how they are identified, and their impact on systems, tactics to prevent them, as well as methods to control and resolve them, using an industry case study as a reference. The creation of innovative services with improved performance and added value, as well as the reduction of data acquisition costs for existing services and the opportunity to create new sources of income within the context of a sustainable business model, will be among the main advantages and benefits of IoT. However, the limits include data minimization and storage.

Wenyu (Derek) Du et al [19] the study suggests that the Blockchain technology that emerging fintech technology can cope with the business organization and business process. Since blockchain is new technology so most of the studies are about presenting a theoretical framework for business solutions. The real implementation of a blockchain-based framework is less. To solve this problem authors presented a process framework based on Actualization and affordance theory. The authors list three advantages of blockchain technology for the company as well as a method for achieving these advantages. The A-A theory, which states that blockchain use cases can be identified in transactions and other business processes, is supplemented by the process model in an experimentation stage. The paper discusses A-A theory, blockchain, and how to use the blockchain to solve transaction, business process-related problems. This research can also assist IT professionals in efficiently implementing blockchain and extracting value from their investments.

Yuri Bobbert et al [20] This research outlines the state of the art for using zero trust at the moment. In this article, ON2IT's "Zero Trust Innovators" assess the shortcomings of current methodologies and how they are addressed in the Zero Trust Framework. The key benefits of this strategy are that risk & technology can be matched with better accuracy and economic efficiency. However, portal technology must be incorporated in the future to improve security.

Larry Nace et al [21] in this research, the TBO infrastructure and application security in the NAS are addressed using a mixed ZTX approach. To guarantee TBO goals are met and to provide defensive depth to fend off any insider threats, the FAA must consider applying a hybrid ZTF theory method. This would entail using the ZTF

application on the NAS Zero Trust extended (ZTX) platform to reinforce the current boundary safeguards given by the FAA Telecommunications Infrastructure (FTI) network. By allowing data integration for this analysis, the security risk is increased due to sophisticated threats that can penetrate the NAS's secure perimeter and interfere with or jeopardize flight safety. Utilizing this platform significantly reduces data leakage.

Peiyun Zhang et al [22] A blockchain system confronts a variety of safety & trust challenges, includes attacks on the propagation & consensus processes, it could lead to erroneous data being stored or data dissemination being delayed. The fundamental design of blockchains is examined in the article along with potential security and trust issues at the data, network, consensus, smart contract, and application layers. Security has improved with data encryption while using the blockchain with zero trust architecture. But to develop the automatic detection of blockchain risks, new techniques are needed. This is essential to offer incentive-compatible consensus methods so selfish nodes in a decentralized blockchain system can handle block verification and accounting on their own.

Sobia Mehrban et al [23] Ever as the introduction of data sharing in e-health systems, data security has been a hotly debated and researched subject. Although data digitization has boosted productivity & speed, it has also made information more susceptible to cyberattacks. Medical records and data breach activities are frequently targeted by hackers. To defend the system, perimeter-based solutions such as firewalls, VPN utilization, and encryption techniques are used, however, their effectiveness is limited. The authors propose that the blockchain could be a viable option for preserving medical data (images) from the beginning to the end of the medical image transmission process. Blockchain technology, according to the research, maintains data integrity by keeping a record of transactions. The zero-trust principle, on the other hand, ensures that medical data is secure and that only verified persons and equipment can communicate with the network. The result reveals that the suggested system has higher security than the present one. A framework for legislative recommendations, however, might offer FinTech users a workable solution to security issues.

Ryan Randy Suryono et al [24] This study aims to: (1) evaluate the current state of the financial technology research sector; (2) recognize research needs, and (3) pinpoint obstacles and trends for prospective future research. This study's new proposal contains financial technology-related theoretical contributions. The results of this study offer a theoretical framework for information systems-based fintech research, including the definition and advancement of fintech technological concepts. To authenticate the caliber of the literature and analysis, this research used Kitchenham's method of systematic literature review together with theme analysis, meta-analysis, and observation. Finally, there are still a lot of benefits that may be derived from fintech research, such as advancements in legislation, security, monitoring, and even other technologies. This literature study can serve as a starting point for future studies to comprehend fintech. Future research in computer science can focus on AI, including algorithms, methods, and finance system strategies.

Nashirah Abu Bakar et al [25] This research developed a method to identify the transactional framework for the electronic wallet payment system in the digital economy. This study outlined the steps involved in a transaction, from the registration of users and service providers to the composition of payments and profit generation for the e-wallet provider's business model. To raise public awareness of the e-wallet transaction, this research offers a precise and trustworthy analysis. In the sectors of the digital economy, the structure employed in this study supports the growth of small and medium-sized firms.

Miguel Pincheira et al [26] The author of this study presents a software architecture designed specially for a trustless water management system, where a small number of IoT sensors can exchange sensed data directly over a public blockchain network. The author uses commercially available hardware to implement the suggested system and thoroughly tests its memory, program size, communication overheads, & power consumption. The findings demonstrate that conventional IoT devices may generally be utilized to communicate immediately and without unnecessary difficulty with a blockchain. More particular, these devices use just 6% more energy than they would normally use to interface with a gateway. The measurement of energy usage at the device level with greater accuracy will be the focus of future research. To achieve this, an accurate estimation of the daily energy budget that takes into account both the idle and sensing modes is required. The creation of the smart contracts that will support our proposal is a key component of this project's road plan.

Hayder M. Kareem Al Duhaidahawi et al [27] in this study, the author examines the definition of fintech and assesses how much fintech factors have an impact on the dependent variable of cyber security. In their study, all correlation coefficients demonstrated a positive relationship and were significant at the level of 0.01; additionally, they discovered that the influence factor between the research variables had a favorable effect when it was significant at the level of 0.05 for all sections of the independent variable. As a result, the authors accepted the correlation hypotheses. The study discovered that the effect coefficient had greatly grown to 0.908, which explains why the independent variable's portions complement one another.

Maliha Sultana et al [28] A relatively new technology called blockchain has shown to be an effective tool for safeguarding private data. The zero trust security paradigm mandates that security measures be taken at every level, including before, during, and even after the transmission of medical images, to guarantee the overall protection of medical information (pictures). A decentralized and trustless framework for secured medical information & image transfer and storage was developed by fusing these two concepts in this study, which looked at a range of works addressing these two concepts. According to the findings of the research, by preserving a record of every transaction, blockchain technology preserves data integrity. Zero trust principles, meanwhile, assure that medical data is protected and that only authorized people and equipment communicate with the network. As a consequence, a suggested methodology addresses a number of data security concerns. Finding ways to speed up and improve the system's usability is one of our future objectives.

Zhiyu Chen et al [29] This study creates a data protection architecture that addresses the device, data, & business layers, accomplishes high-level security protection of data flow in all power network channels, and enhances the system's safety monitoring and control capacity. This research suggests the use of BcT to improve data interaction safety and provide high-level data circulation protection in all power network links. The strategy described in this work tightly manages the application of control information, standardizes data use rights management and enhances the trustworthiness and interaction rate of corporate data. However, terminal software remains a threat.

Yuri Bobbert et al [30] The first section of the article summarises the limitations of current techniques and how they are dealt with in the ZTF created by ON2IT's 'Zero Trust Innovators.' The design and engineering of a Zero Trust object that resolves the current issues are then described, in accordance with Design Science Research (DSR). The design of an empirical validation using practitioner-oriented research to enhance the use of Zero Trust approaches is described in the study's concluding part, and how 73 security professionals participated in this validation in 2020. The result is a framework that is suggested and related to technology that, using Zero Trust principles, addresses various organizational layers to understand and align cyber security concerns, as well as comprehend the organization's readiness and fitness to counter cyber security hazards. However, the operating cost of this system is high.

Britta Hale et al [31] This article suggests a method depends on well-established security principles like Zero-Trust & defense-in-depth to assistance prohibit and mitigate security risks, including those arising from ML-based components. The demonstration of the methodology is based on an Unmanned Aerial Vehicle (UAV) case study with an advanced Intelligence, Surveillance, and Reconnaissance (ISR) module. However, this paradigm is lacking in a balance between mission performance, security, and resources.

Christopher A. Villareal [32] The goal of this design is to add to the body of information regarding trust-extended UTAUT studies and study which aspects matter when adopting DIMS expending the Unified Theory of Acceptance and Uses of Technology (UTAUT). This study provides IT decision-makers with acceptance-based information on whether to implement DIMS. The data inform DIMS developers that potential users rely on their DIMS adoption mainly on Trust and Performance Expectancy.

Sairath Bhattacharjya et al [33] The model has a zero-trust architecture, meaning that before being processed, all requests and replies are verified. This study presents a novel approach to efficiently generate a secret key for every user & device pair. The strategy establishes the foundation for end-to-end encrypted communication. One advantage of using the key is that the command and device response are hidden from all parties, including the gateway. With the others, each pair can safely communicate. The P3 connection paradigm automates the key generation and can assist ensure privacy during communication. However, this paradigm has a few drawbacks, including resource constraints, a lack of user authentication, and insufficient encryption.

Aleksandra Scalco et al [34] This study demonstrates how RA is used in the critical infrastructure vertical of the Healthcare and Public Health industry to investigate ideas like Zero Trust (ZT) and Defense-in-Depth (DiD) structure principles associated to policymaking. The benefits of the RA include improved decision-making and policy-making assistance for cyber security in control systems.

Hesam Hamledari et al [35] This study demonstrates that current payment programmes, even those that are computerised, cannot safely automate progress payments due to their reliance on centralised control systems and lack of execution guarantees. The study examines how these limitations might be handled by decentralised, blockchain-based smart contracts. It examines how smart contracts can reliably and autonomously condition cash flow on the state of a product flow and examines the theoretical foundation for the creation of an automated payment system. This project holds higher security and trust over other system. But this model is relay on centralized systems for command and control, as well as a lack of execution assurance.

Kris Oosthoek [36] this document provides the summary of DeFi security incidents that have occurred in the wild. Many of these exploits are market assaults, in which badly designed business logic in one protocol is combined with credit granted by another to increase appropriations. Attackers are increasingly utilizing DeFi's strength of permissionless composability against itself, rather than misusing individual protocols. DeFi is the first to provide a comprehensive examination of real-world security issues within the young financial ecosystem. According to the author, the enhanced Defi will give higher security than other model. However, because this is a developing approach, the Defi must be updated every 16 months.

Kwok-Yan Lam et al [37] In this paper, the author recommends a security reference architecture for identifying security threats, taking care of security requirements, and comprehending security challenges with intricate IoT systems. It offers a security reference architecture focused on Activity-Network-Things (ANT), which is built on the three architectural viewpoints for examining IoT systems: device, internet, and semantic. This structure is malleable enough to handle any IoT application, making it simple to apply to the scenario of SAGIN-enabled IoV.

Zachary A. Collier et al [38] The author of this article proposes the concept of a supply chain with zero trust. An ideology with a zero trust foundation is one that originated in the IT sector and cyber security and presupposes that all actors and activity are unreliable. In this paper, the author connects supply chain principles to zero trust and discusses the actions an organization may take to make the change. With this model, the service will be highly secured and the tread prevention will be improved. However, this model's transaction costs are considerable, and it only works with a few software applications.

Suparna Dhar et al [39] the author of this paper addresses the security concerns associated with IoT implementation and proposes a blockchain-based, zero-trust paradigm for IoT device security. The characteristics & communication protocols of IoT devices are becoming more similar thanks to risk-based segmentation of the IoT network. Zero Trust widens the trust's circle beyond the IT/OT network. The IoT network's device documentation and access control capabilities are enhanced by

blockchain. Both academic academics and practitioners will benefit from the IoT implementers will be able to address current security difficulties with the help of the suggested IoT security architecture. Future research will need to develop techniques for assessing the dependability of the suggested architecture in a real IoT network.

Sultan Algarni et al [40] In this study, a unique method for managing the delivery of decentralised, lightweight safe access control for an IoT system is provided, depends on a multi-agent system & blockchain. The primary purpose of the proposed technique is to build Blockchain Managers (BCMs) for securing IoT access control and enabling protected communication between local IoT devices. Additionally, technology permits safe connectivity between cloud computing, fog nodes, and IoT devices. The proposed framework will be used in subsequent studies to evaluate the degree to which the fundamental security objectives integrity through the use of digital signatures, the requirements for shared secret key authentication, MAC policy authorization, and public key encryption have been satisfied.

Lampis Alevizos et al [41] The author looks into how ZTA can be added onto endpoints in the work due to the concurrence of ZTA with blockchain-based intrusion detection and prevention. In particular, the author conducts a cutting-edge analysis of ZTA models, actual designs that focus on endpoints, and intrusion detection systems based on blockchain. From the review, it is found that there is no need for tight resource integration. Adopting this technique is still hampered by issues with performance, computational costs, and choosing the appropriate blockchain implementation. Further research is required to adequately answer these questions.

Xiaohan Hao et al [42] In contrast to situations where one may rely on a trustworthy third party, sharing information is more challenging in zero-trust environments. They suggest a zero-trust information-sharing protocol that uses blockchain technology to address this problem. It can filter out fabricated information while preserving participant anonymity. The performance of our protocol is then assessed by a number of experiments, and the safety of our procedure is finally proved in the commonly compostable confident structure. The evaluation results demonstrate that our protocol's three crucial steps execute on average in 0.059 seconds, 0.060 seconds, and 0.032 seconds, indicating that it has the potential to be used in a real-world setting. Future improvements, however, will focus on strengthening our system's security by lowering the necessary communication, computation, and time overheads (against a wider variety of assaults).

Cyril Onwubiko [43] Protecting digital investments, ensuring a safe, secure, and favorable business environment, safeguarding a country's vital national infrastructures, and promoting citizen welfare are its three main objectives. The author of this research examines operational elements that affect CyberOps situational awareness, in particular, the features that deal with understanding and comprehension of operational and human factors aspects and that aid in providing insights on human operator decision-making. But this model requires a human operator.

Yang Xiao et al [44] The novel blockchain-based decentralised SAS architecture known as BD-SAS offers SAS services successfully and securely without depending on the dependability of any particular SAS server. G-Chain is a worldwide blockchain that BD-SAS utilises to comply with spectrum regulations. and local blockchains (L-

Chains) with smart contract functionality are created in each spectrum zone to automate spectrum access assignment based on user requests. A proposed strategy for fault-tolerant spectrum access management is a decentralized SAS architecture. But more automation, flexibility, and finer granularity are necessary.

José María Jorquera Valero et al [45] Adversaries in 5G-enabled environments can take advantage of resource-sharing flaws to launch lateral attacks against other tenant resources, interfere with the delivery of 5G services, or even damage infrastructure resources. Additionally, the multi-tenancy security problems or the dynamic nature of 5G infrastructure vulnerabilities cannot be addressed by the current security and trust models. As a result, a novel security and trust paradigm for 5G multi-domain circumstances is presented in this work. The author presents to demonstrate its value, a supporting 5G network framework is used with a threat model for multi-tenant scenarios. Additionally, the author provides a number of mitigation strategies, including enhancing security and trust levels through network security monitoring, threat analysis, and end-to-end trust configurations. The architecture is put to use in a real-world use case by the H2020 5GZORRO project, which envisions a multi-tenant system where domain owners can freely share resources. The proposed framework reduces the requirement for human interaction by establishing a protected environment with zero-touch automation capabilities.the systematic surevy on Block Chain for Financial Transaction in Zero Trust Architecture has been shown in Table 2.

Ref no	Publish On	Author	Technique	Significance	Limitation
11	2018	Casimer DeCusatis	In this article, for blockchain apps that are hosted in the cloud, the author presents experimental test bed results for a cutting-edge user identity management technique. To request a new session, they use a Black Ridge Technology endpoint on a Windows host to add cryptographic identity tokens to the initial packet.	This model requires less time to use than other models.	A third-party program is necessary
12	2018	Mayra Samaniego	In this study, they offer Amatista, a blockchain-based middleware for IoT administration. With its innovative zero-trust hierarchical mining method, Amatista makes it possible to authorize frameworks and transactions at different degrees of trust.	By using Amatista the performance is increased.	Algorithms in Amatista need to be updated.
13	2018	Jungho Kang	To indicate requirements and security concerns and the study will examine current trends in mobile Fintech payment services & classify in accordance with service methods to enable the provision of better and further secure services in the future.	Consider mutual authentication, authorization, integrity, privacy, and availability while developing	Future development should be done with different applications.

				fintech payment systems.	
14	2018	Nir Kshetri	Case examples of various stages of blockchain development for various applications are discussed	The use of the blockchain and zero trust can boost trust and security.	The transaction costs and the requirement for third-party software are regarded as disadvantages
15	2018	Samuel Fosso Wamba	The goal of this research is to close a knowledge gap in the body of existing Bitcoin, Blockchain, and FinTech literature.	Using blockchain & zero trust can improve security and trust.	The cost of the software is prohibitively high.
16	2019	Jesus Iglesias Garcia	In this paper, Hyot a blockchain-based Internet of Things (IoT) activity registration service is introduced.	Compared to public BCs, it has a substantially greater transaction rate.	Required third-party software
17	2019	HYE-YOUNG PAIK	The project intends to encourage a logical approach to implementing blockchain technology in big software systems and to raise awareness of blockchain technology as data storage.	This model gives higher Atomicity, concentricity, and isolation of troubles	This model required higher transaction costs and third-party software.
18	2019	Cordeiro	This paper will discuss recognized dangers, how they are identified, and how they are mitigated Its influence on systems, techniques to avoid them, and strategies to contain and resolve them to develop interest and understanding in these subjects, using an industry case study as a reference.	Data accusation cost is low	Data minimization and storage.
19	2019	Wenyu (Derek) Du	The article goes through A-A theory, blockchain, and how to use blockchain to solve transaction and business process challenges. This research can also help IT professionals adopt blockchain more efficiently and get the most out of their investments.	This model will give the result in the most efficient and cost-effective.	Blockchain needs to upgrade frequently.
20	2020	Yuri Bobbert	The study examines the limits of current techniques and how the ON2IT 'Zero Trust Innovators' Zero Trust Framework addresses them.	This system has higher precision and cost-effectiveness.	Lack of portal technology.

21	2020	Larry Nace	In this research, a hybrid ZTX method for TBO infrastructure and application security in the NAS is proposed.	Data leaking is decreased.	The algorithm needs to be more strong
22	2020	Peiyun Zhang	The fundamental architecture of a blockchain is examined, as well as potential issues with trust and security at the application, data, network, consensus, and smart contract layers.	Using blockchain and zero trust the security and data encryption has been improved.	A new mechanism is required for the automatic detection of data leakage.
23	2020	SOBIA MEHRBAN	This article presents a detailed examination of FinTech by evaluating current and potential privacy and security issues in the banking sector. It offers a complete overview of contemporary security challenges, detection, and prevention. FinTech has been proposed with mechanisms and security options.	This model gives higher security than other system	To handle security challenges a policy recommendation using a framework can be a workable solution.
24	2020	Ryan Randy Suryono	This study's goals are to: (1) evaluate the state of the art in financial technology research; (2) recognize research gaps in the area; and (3) recognize roadblocks and trends for potential future scope.	This model has high security and monitoring.	The Foundation of FinTech needs to make strong
25	2020	Nashirah Abu Bakar	This study outlined the steps involved in a transaction, from the registration of users and service providers to the composition of payments & profit generation for the e-wallet provider's business model.	This model can be used in small-scale businesses.	This model needs to be updated frequently
26	2020	Miguel Pincheira	In this study, the author presents a software architecture designed for a trustless water management system where constrained IoT devices can directly exchange sensed data on a public blockchain network.	This model consumes very less energy when compared with other models.	Budget is high
27	2020	Hayder M. Kareem Al_Duhaidahawi	In this study, the author examines the definition of fintech and assesses how much fintech factors affect cyber security, the dependent variable.	The coefficient significantly increased to 0.908	It was compactable with limited software
28	2020	Maliha Sultana	The zero trust security paradigm ensures that security measures are put in place at every level, from the beginning, during, and even after medical image transmission, to offer overall protection of medical data (images).	This method to improves the security of medical data transfer.	The processing speed is low

29	2021	Zhiyu Chen	This study recommends using BcT to increase data interaction security and provide high-level data circulation protection in all power network lines.	The credibility & engagement rate of business data were both enhanced by this method.	Terminal software remains a threat.
30	2021	Yuri Bobbert	The definition of asset owners' participation from a commercial perspective is one of the main objectives of the ON2IT ZTF design, leading to more clear interpretations of concepts like risk appetite and recovery time objectives.	It gives higher security protection	The cost of this system is high
31	2021	Britta Hale	This paper proposes a method for preventing and mitigating security threats, including those originating from ML-based components, depends on accepted security principles like Zero-Trust and defense-in-depth.	The security of the system is enhanced.	This paradigm is deficient in terms of balancing mission performance, security, and resources.
32	2021	Christopher A. Villareal	The goal of this design is to study which aspects matter when adopting DIMS using the unified theory of acceptance (UTAUT) & to advance the corpus of information around trust-extended UTAUT investigations.	Trust and Performance of this model is high	It is required to pay exorbitant fees for carrying out transactions.
33	2021	Sairath Bhattacharjya	This study presents a novel approach to efficiently generate a secret key for every user and device pair.	Device responses are hidden from all parties.	There are a few limitations to this paradigm, including resource limits, a lack of user authentication, and insufficient encryption.
34	2021	Aleksandra Scalco	The use of RA in the healthcare and public sectors is highlighted in this paper. Critical infrastructure in the healthcare industry is assessing policymaking ideas like Zero Trust (ZT) and Defense-in-Depth (DiD) design principles.	Better cyber security decision-making and policy-making	The introduction of a context-aware environment has been abandoned.
35	2021	Hesam Hamledari	The paper examines how decentralized smart contracts built on the blockchain can get around these limitations. An automated payment system's theoretical underpinnings are investigated, as well as the usage	Higher levels of security and trust than other systems	This model relies on centralized execution and control systems with

			of smart contracts to enable reliable & autonomous conditioning of cash flow on product flow status.		no assurance of proper execution.
36	2021	Kris Oosthoek	The first overview of real-world DeFi security incidents is presented in this study. The study noticed that many of these exploits are market attacks that use credit granted by another protocol to weaponize poorly designed business logic in one protocol and inflate appropriations.	More security will be provided by Defi than by any other model.	The Defi must be updated every 16 months.
37	2021	Kwok-Yan Lam	In this research, the author describes a method for comprehending complicated IoT system security concerns and for identifying security threats and addressing security needs, offering a security reference architecture.	This system is more flexible than other models	Fine-grained access controls to resources or applications are no longer available.
38	2021	Zachary A. Collier	In this article, the author connects supply chain principles to zero trust and discusses the actions a firm may take to make the change.	A high level of security will be provided for the service, and tread prevention will be enhanced.	This architecture only functions with a limited number of software applications and has high transaction costs.
39	2021	Suparna Dhar	The author of this paper examines the security issues related to IoT implementation and suggests a paradigm for IoT device security based on blockchain and zero trust.	The security feature has been enhanced.	This model is lack reliability.
40	2021	Sultan Algarni	Creating Blockchain Managers (BCMs) for securing IoT access control & enabling safe communication between local IoT devices is the major objective of the proposed approach.	The encryption is high in this model.	This model is lacking digital signature and authentication via a secret key.
41	2021	Lampis Alevizos	Motivated by the integration of ZTA with intrusion detection and prevention using blockchain, researchers investigate how ZTA can be added to endpoints in this work.	There is no requirement for tight resource integration.	Performance and computing cost is high
42	2021	Xiaohan Hao	In contrast to situations where we may rely on a trustworthy third party, sharing information is more challenging in zero-trust environments. We suggest a zero-trust information-sharing protocol that uses blockchain technology to address this problem. It can filter out fabricated information while preserving participant anonymity.	The execution time is very low	The communicatio n and calculation time need to improve.

43	2022	Cyril Onwubiko	This study examines operational characteristics that affect CyberOps situational awareness, in particular, the elements that contribute to understanding operational and human factors components and provide insights into how human operators make decisions.	The security treads are minimized.	Human operator is required.
44	2022	Yang Xiao	BD-SAS is a unique blockchain-based decentralized SAS architecture that gives SAS services securely & effectively without depending on the reliability of any one SAS server.	It gives a promising technique for fault-tolerant and resilient spectrum access management	More automation, flexibility, & granularity are required.
45	2022	José María Jorquera Valero	The multi-tenancy security challenges or the dynamic nature of threats to 5G framework cannot be addressed by the current security and trust paradigms. In this study, the author so suggests a novel safety and trust paradigm for 5G multi-domain scenarios.	This model minimizes human intervention by offering zero-touch automation possibilities.	The framework needs to be updated regularly.

Table 2. Systematic Review on Block Chain for Financial Transaction in Zero Trust Architecture

5. Summary

Fintech security is a significant concern in the fintech ecosystem, according to the literature review. Embedded systems, mobile networks, cloud computing, big data analysis, & other technologies are included in fintech. Online banking, wallets, and other services are just a few of the fintech industry's many applications. However, issues with security, privacy, risks, and cyber-attacks exist in the fintech sector.

In contrast, the idea of "never trust, always verify" is the foundation of the zero-trust approach to safeguarding network devices and other entities. Zero trust security solutions are based on policy enforcement engines and are used in a variety of industries, including Google's Beyond Corp, Forrester NGFW/ZTX, and others. Zero-trust is also utilized in wireless networks, the IoT, cloud computing, & edge computing.

Blockchain technology is a novel fintech technology with characteristics such as immutability, availability, and consensus. The majority of blockchain research is concentrated on creating and implementing frameworks.

According to the survey, FinTech with zero trust architecture can increase security with high encryption. However, every technological advancement has a few drawbacks that must be addressed before more effective and secure financial transactions can be conducted.

6. Challenges and Future Scope

According to the results of the survey, there are a few limitations in current technology that must be addressed to achieve greater security encryption and cost-effectiveness. Some limitations are listed below.

- The model has a very high-cost factor.
- Effective communication necessitates the use of third-party software.
- Human intervention is still required to monitor the process.

6.1. Future Scope

The solution for the challenges facing

- To complete the process, third-party software is required. Reduce the use of third-party software in the future, and once the third-party software is terminated, the cost factor will also decrease.
- Human supervision is always required in financial transactions; however, if we can create an update in an automatic process, the need for human supervision will be reduced.

7. Conclusion

The financial industry and its related services are significantly impacted by a recent disruptive technological trend. As a result, technology has fundamentally changed how the financial sector functions. Though its shape and contours are still unclear, research on this quickly spreading trend has already begun. At this point, it is necessary to gain an understanding of the information being offered on this new technology.

The goal of this study is to increase awareness by reviewing and classifying the material. After offering a thorough analysis of FinTech and assessing recent research on security and privacy issues in the financial industry, this paper provided a taxonomy of current security difficulties, detection mechanisms, and security solutions for FinTech. This article goes into great detail about the suggested plans for the security and privacy of financial technologies. Finally, to propose future directions for this new technological era, future research issues are explored.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Author Contributions

All authors read and approved the final manuscript

Data Availability Statement

Data sharing not applicable to this article, because it's confidential.

Ethical Approval

The article does not contain any studies with Human Participants or animals performed by any of the Authors.

References

[1] B. B. Pena, "Understanding and designing technologies for everyday financial collaboration," University of Northumbria at Newcastle (United Kingdom), 2021.

[2] S. Kak, "Zero Trust Evolution & Transforming Enterprise Security (Doctoral dissertation, California State University San Marcos)," 2022.

[3] M. Kaur & S. Gupta, "Blockchain technology for convergence: an overview, applications, and challenges," Blockchain and AI Technology in the Industrial Internet of Things, pp.1-17, 2021.

[4] S. Berlato, R. Carbone, A. J. Lee & S. Ranise, "Formal modelling and automated trade-off analysis of enforcement architectures for cryptographic access control in the cloud," ACM Transactions on Privacy and Security, Vol. 25, No.1, pp.1-37, 2021.

[5] C. T. Nguyen, D. T. Hoang, D. N. Nguyen, D. Niyato, H. T. Nguyen & E. Dutkiewicz, "Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities," IEEE access, Vol. 7, pp.85727-85745, 2019.

[6] T. Hardjono & N. Smith, "Towards an attestation architecture for blockchain networks," World Wide Web, Vol. 24, No.5, pp.1587-1615, 2021.

[7] J. J. O'Hare, A. Fairchild & U. Ali, "Money & Trust in Digital Society, Bitcoin and Stablecoins in ML enabled Metaverse Telecollaboration," arXiv preprint arXiv:2207.09460, 2022.

[8] C. Buck, C. Olenberger, A. Schweizer, F. Völter & T. Eymann, "Never trust, always verify: A multivocal literature review on current knowledge

and research gaps of zero-trust," Computers & Security, Vol. 110, p.102436, 2021.

[9] M. Pace, "Zero Trust networks with Istio (Doctoral dissertation, Politecnico di Torino)," 2021.

[10] A. John, A. Reji, A. P. Manoj, A. Premachandran, B. Zachariah & J. Jose, "A novel hash function based on hybrid cellular automata and sponge functions," In Asian Symposium on Cellular Automata Technology, pp.221-233, 2022, March. Singapore: Springer Nature Singapore.

[11] C. DeCusatis, M. Zimmermann & A. Sager, "Identity-based network security for commercial blockchain services," In 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp.474-477, 2018, January. IEEE.

[12] M. Samaniego & R. Deters, "Zero-trust hierarchical management in IoT," In 2018 IEEE international congress on Internet of Things (ICIOT), pp.88-95, 2018, July. IEEE.

[13] J. Kang, "Mobile payment in Fintech environment: trends, security challenges, and services," Human-centric Computing and Information sciences, Vol. 8, No.1, pp.1-16, 2018.

[14] N. Kshetri, "1 Blockchain's roles in meeting key supply chain management objectives," International Journal of information management, Vol. 39, pp.80-89, 2018.

[15] S. Fosso Wamba, J. R. Kala Kamdjoug, R. Bawack & J. G Keogh, "Bitcoin, blockchain, and FinTech: a systematic review and case studies in the supply chain," Production Planning and Control, Forthcoming, 2018.

[16] J. Iglesias García, J. Diaz & D. Arroyo, "Hyot: Leveraging Hyperledger for Constructing an Event-Based Traceability System in IoT," In International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019) Seville, Spain, May 13th-15th, 2019 Proceedings, Vol. 12, pp.195-204, 2020. Springer International Publishing.

[17] H. Y. Paik, X. Xu, H. D. Bandara, S. U. Lee & S. K. Lo, "Analysis of data management in blockchain-based systems: From architecture to governance," IEEE Access, Vol. 7, pp.186091-186107, 2019.

[18] C. Cordeiro & H. Barbosa, "Digital Privacy and Security Conference."

[19] W. D. Du, S. L. Pan, D. E. Leidner & W. Ying, "Affordances, experimentation and actualization of FinTech: A blockchain implementation study," The Journal of Strategic Information Systems, Vol. 28, No.1, pp.50-65, 2019.

[20] Y. Bobbert & J. Scheerder, "Zero trust validation: from practical approaches to theory," Sci. J. Res. Rev, Vol. 2, No.5, 2020.

[21] L. Nace, "Securing Trajectory based Operations through a Zero Trust Framework in the NAS," In 2020 Integrated Communications Navigation and Surveillance Conference (ICNS), pp.1B1-1, 2020, September. IEEE.

[22] P. Zhang & M. Zhou, "Security and trust in blockchains: Architecture, key technologies, and open issues," IEEE Transactions on Computational Social Systems, Vol. 7, No.3, pp.790-801, 2020.

[23] S. Mehrban, M. W. Nadeem, M. Hussain, M. M. Ahmed, O. Hakeem, S. Saqib ... & M. A. Khan, "Towards secure FinTech: A survey, taxonomy, and open research challenges," IEEE Access, Vol. 8, pp.23391-23406, 2020.

[24] R. R. Suryono, I. Budi & B. Purwandari, "Challenges and trends of financial technology (Fintech): a systematic literature review," Information, Vol. 11, No.12, p.590, 2020.

[25] N. A. Bakar, S. Rosbi & K. Uzaki, "E-wallet transactional framework for digital economy: a perspective from Islamic financial engineering," International Journal of Management Science and Business Administration, Vol. 6, No.3, pp.50-57, 2020.

[26] M. Pincheira, M. Vecchio, R. Giaffreda & S. S. Kanhere, "Exploiting constrained IoT devices in a trustless blockchain-based water management system," In 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp.1-7, 2020, May. IEEE.

[27] H. M. K. Al Duhaidahawi, J. Zhang, M. S. Abdulreza, M. Sebai & S. A. Harjan, "Analysing the effects of FinTech variables on cybersecurity: Evidence form Iraqi Banks," International Journal of Research in Business and Social Science, No.6, pp.123-133, 2020.

[28] M. Sultana, A. Hossain, F. Laila, K. A. Taher & M. N. Islam, "Towards developing a secure medical image sharing system based on zero trust principles and blockchain technology," BMC Medical Informatics and Decision Making, Vol. 20, No.1, pp.1-10, 2020.

[29] Z. Chen, L. Yan, Z. Lü, Y. Zhang, Y. Guo, W. Liu & J. Xuan, "Research on zero-trust security protection technology of power IoT based on blockchain," In Journal of Physics: Conference Series, Vol. 1769, No.1, p.012039, 2021. IOP Publishing.

[30] Y. Bobbert & J. Scheerder, "On the design and engineering of a zero trust security artefact," In Advances in Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC), Vol. 1, pp.830-848, 2021. Springer International Publishing.

[31] B. Hale, D. L. Van Bossuyt, N. Papakonstantinou & B. O'Halloran, "A zero-trust methodology for security of complex systems with machine learning components," In International design engineering technical conferences and computers and information in engineering conference, Vol. 85376, p.V002T02A067, 2021, August. American Society of Mechanical Engineers.

[32] C. A. Villareal, "Factors Influencing the Adoption of Zero-Trust Decentralized Identity Management Solutions (Doctoral dissertation, Capella University)," 2021.

[33] S. Bhattacharjya & H. Saiedian, "Establishing and validating secured keys for IoT devices: using P3 connection model on a cloud-based architecture," International Journal of Information Security, Vol. 21, No.3, pp.427-436, 2022.

[34] A. Scalco, D. Flanigan & S. Simske, "Control Systems Cyber Security Reference Architecture (RA) for Critical Infrastructure: Healthcare and Hospital Vertical Example," Journal of Critical Infrastructure Policy 2.2 (2021): 125-43. Print, 2021.

[35] H. Hamledari & M. Fischer, "Role of blockchain-enabled smart contracts in automating construction progress payments," Journal of legal affairs and dispute resolution in engineering and construction, Vol. 13, No.1, p.04520038, 2021.

[36] K. Oosthoek, "Flash crash for cash: Cyber threats in decentralized finance," arXiv preprint arXiv:2106.10740, 2021.

[37] K. Y. Lam, S. Mitra, F. Gondesen & X. Yi, "ANT-centric IoT security reference architecture—Security-by-design for satellite-enabled smart cities," IEEE Internet of Things Journal, Vol. 9, No.8, pp.5895-5908, 2021.

[38] Z. A. Collier & J. Sarkis, "The zero trust supply chain: Managing supply chain risk in the absence of trust," International Journal of Production Research, Vol. 59, No.11, pp.3430-3445, 2021.

[39] S. Dhar & I. Bose, "Securing IoT devices using zero trust and blockchain," Journal of Organizational Computing and Electronic Commerce, Vol. 31, No.1, pp.18-34, 2021.

[40] S. Algarni, F. Eassa, K. Almarhabi, A. Almalaise, E. Albassam, K. Alsubhi & M. Yamin, "Blockchain-based secured access control in an IoT system," Applied Sciences, Vol. 11, No.4, p.1772, 2021.

[41] L. Alevizos, V. T. Ta & M. H. Eiza, "Augmenting zero trust architecture to endpoints using blockchain: A systematic review," arXiv preprint arXiv:2104.00460, 2021.

[42] X. Hao, W. Ren, R. Xiong, T. Zhu & K. K. R. Choo, "Asymmetric cryptographic functions based on generative adversarial neural networks

for Internet of Things," Future Generation Computer Systems, Vol. 124, pp.243-253, 2021.

[43] C. Onwubiko, "CyberOps: Situational Awareness in Cybersecurity Operations," arXiv preprint arXiv:2202.03687, 2022.

[44] Y. Xiao, S. Shi, W. Lou, C. Wang, X. Li, N. Zhang ... & J. H. Reed, "Decentralized spectrum access system: Vision, challenges, and a blockchain solution," IEEE Wireless Communications, Vol. 29, No.1, pp.220-228, 2022.

[45] J. M. Jorquera Valero, P. M. Sánchez Sánchez, A. Lekidis, J. Fernandez Hidalgo, M. Gil Pérez, M. S. Siddiqui ... & G. Martínez Pérez, "Design of a security and trust framework for 5G multi-domain scenarios," Journal of Network and Systems Management, Vol. 30, No.1, p.7, 2022.

10.31341/jios.48.1.12
Open Access
UDC 005.6:316.276
Original Scientific Paper

Organizational Homeostasis: A Quantum Theoretical Exploration with Bohmian and Prigoginian Systemic Insights

David Leong *david.leong@canberra.edu.au*
Canberra Business School
Faculty of Business, Government & Law
University of Canberra

Abstract

This study examines the complex interactions in organizational structures using Bohm's 'wholeness' and Prigogine's equilibrium theories. Wave analogies and quantum principles like superposition, non-locality, and entanglement explain fluctuations in these systems. Thus, this research suggests a paradigm shift in organizational methods towards a balanced, scientific approach. Organizations need flexible tactics and behave like dissipative structures to maintain internal coherence in chaos. Heightened through mindful techniques, corporate consciousness provides insights into temporal dynamics, improving decision-making, market resilience, and an expanded organizational ethos founded in present awareness. This heightened consciousness and demand for organizational alignment and coherence empowers the corporate entities to succeed in present conditions and anticipate and address future obstacles. This study introduces 'Mindful Corporate Entity' ('MCE'), emphasizing mindfulness as a critical tool for organizational well-being and sustainability. This change is proposed to close management gaps.

Keywords: Quantum theory, well-being, equilibrium, far-from-equilibrium, homeostasis, dissipative structures, organizational development, transformation, change

1. Introduction

Recently, a burgeoning interest has arisen in studying deep structures within social systems [26][36][51], particularly in organizational management. Within the vast spectrum of organizational theories, traditional paradigms often treat self-organising processes as inherently unpredictable, echoing sentiments from the natural sciences [37]. The unpredictability, rooted in the complexity of these processes, has been a subject of intrigue for researchers attempting to understand or gain some semblance of control over them.

This paper forwards a novel perspective, 'Organizational Homeostasis', by postulating that an in-depth exploration and management at the level of these deep structures can offer organizations means to exert influence over these self-organizing processes. Organizational homeostasis is a concept introduced in this paper derived from the biological principle of maintaining stability to apply to the management and

operation of organizations. It describes how an organization continually adjusts and adapts its internal processes, strategies, and policies in response to external environmental fluctuations and internal changes. This dynamic process involves implementing feedback mechanisms, fostering adaptability and resilience, maintaining an internal equilibrium of resources and operations, and managing change effectively. Like the self-regulatory mechanisms in living organisms, this process involves feedback mechanisms for monitoring and assessment, adaptation and resilience to environmental shifts, maintaining internal balance across various resources, and effective change management [30]. It encapsulates the organization's capacity to self-regulate and sustain functionality and competitiveness in an ever-changing business environment.

At the outset, it is pivotal to delineate what constitutes the 'deep structure' in social systems. Translating this notion to the organizational context, deep structures can be understood as the underlying patterns, relationships, and frameworks that inform and shape overt behaviors and outcomes in social systems [9]. The previously predominant mechanistic worldview, which posits the universe as analogous to a machine with discrete, independently-operable components, is increasingly challenged [9]. Current advancements in modern physics underscore the inextricability of phenomena; they cannot be comprehended in isolation but demand an examination grounded in relational interconnectedness [7]. This paradigm shift resonates with quantum mechanics, where principles of non-separability and entanglement refute the classical separability of entities [61]. Intriguingly, while quantum phenomena traditionally pertain to the microscale, there is burgeoning interest in extrapolating these interconnected principles to understand macroscale events.

"Quantum mechanics, currently assumed to be a fundamental physics theory, was initially formulated to explain nature's atomic and subatomic scales. However, as the atomic hypothesis asserts, the macroscopic world is composed of a collection of such small constituents that quantum mechanics inherits a universal status: it must be able to explain phenomena at all levels of description. Put differently, quantum mechanics must assign states at both microscopic levels when a complete characterization of the underlying physical system is assumed and at a macroscopic level of description, which is given by few effective (coarse-grained) degrees of freedom that we have access to. This universality implies a two-way describing of nature" [14].

This invites a reconceptualization of our understanding of broader systems and networks, urging a shift from isolated examinations to holistic, interconnected analyses. At a fine-grained level, these deeper structures are tacit and unobservable. They undergird the processes by which organizations evolve, adapt, and respond to external stimuli. In this light, the self-organizing processes are not merely random or chaotic phenomena but are informed and shaped by these underlying frameworks. By targeting and managing these foundational elements, organizations can anticipate, if not wholly predict, the trajectory and outcomes of homeostatic self-organization.

In the annals of physiological literature, Cannon's seminal work on homeostasis stands as a paradigmatic cornerstone, delineating the body's intricate mechanisms for sustaining internal equilibria amidst a panoply of external variations [12]. Such an

autoregulatory framework presents an apt metaphorical parallel to the organizational dynamism delineated through Bohm's implicate and explicate order [8] and Prigogine's instability theory in complexity science [43]. Bohm's assertion of the universe's profound interconnectedness, where discrete entities meld into an integrated whole [8], resonates with the physiological constructs of homeostasis posited by Cannon. Analogous to the physiological interplay of systems working in harmonized synchrony to retain homeostasis, Bohm's 'wholeness' [8] underscores the indispensability of perceiving organizations not as fragmented silos, but as harmoniously integrated entities; "in which all parts of the universe, including the observer and his instruments, merge and unite in one totality. In this totality, the atomistic form of insight is a simplification and an abstraction, valid only in some limited context". Disruptions or alterations in one subsystem invariably cascade effects onto others, underscoring the imperative of holistic organizational perceptions.

Prigogine and Stengers's postulation, emphasizing the perennial oscillatory nature of systems vis-à-vis equilibrium [42], finds resonance in Cannon's depiction of homeostasis [12]. Rather than envisioning equilibrium as a static fulcrum, both scholars illuminate its dynamism, with systems perpetually recalibrating in response to perturbations and stimuli. The physiological tapestry of self-regulation and adaptation, as expounded by Cannon [12], thus offers invaluable insights for organizations, emphasizing anticipatory adaptability in the face of inevitable deviations.

As elucidated earlier, the organizational potential of self-organization finds a quintessential exemplar in Cannon's physiological homeostasis—a paragon of a self-regulating, autoregulatory system. The human body's adeptness at auto-calibration offers cogent insights into the trajectories organizations might harness in their quest for equilibrium amidst tumultuous environments. Evoking the physiological paradigms articulated by Cannon, homeostasis emerges as an illustrative compass, demystifying the intricate dance of dynamic balance pivotal for organizations. Analogous to the body's symphonic orchestration of processes to safeguard internal stability, contemporary organizations, ensconced in an ever-evolving environment, are necessitated to perpetually refine and recalibrate their strategic compass in response to the unpredictable external exigencies they encounter. This paper's integration nests Cannon's foundational constructs within the broader tapestry of organizational equilibrium as delineated through Bohm's and Prigogine's prisms but on quantum's terms.

The concepts of non-separability and interdependence, traditionally linked with quantum mechanics, are increasingly believed to offer fresh perspectives on macroscopic occurrences [21]. This challenges the longstanding bifurcation between quantum and classical domains and calls for re-evaluating systemic constructs on broader scales. Transposing these ideas to an organizational paradigm, this paper asserts that the intricate structure at a more granular level, composed of myriad interacting entities ('Mindful Corporate Entity' or 'MCE'), suggests the universality of quantum mechanics, implying its capability to clarify interactions and entanglements across varied descriptive dimensions.

The central tenet of the 'Mindful Corporate Entity', a coined phrase in this paper, is the emphasis on individual members' mindfulness (or awareness) and consciousness. Mindfulness is a form of awareness but with added elements of focused attention and non-judgment [52]. While awareness can refer to any cognition of stimuli, internal or external, mindfulness is specifically about being aware in a certain way: on purpose, in the present moment, and non-judgmentally. Mindfulness is essentially awareness applied in a deliberate, focused, and intentional manner. This granularity perspective suggests that the true organizational dynamism arises from these mindful individuals' conscious, intentional interactions. Rather than marginalizing individual agency as per traditional organizational theories, this concept places it at the forefront, asserting that the organization's emergent properties reflect the collective mindfulness of its members.

Fundamentally, quantum mechanics describe states on two levels: at the microscale, which is based on essential degrees of freedom and interactions from a more fine-grained level of modelling where quantum mechanics governs the interactions between these particles [48], and at the macroscale, characterized by specific overarching (or coarse-grained) factors. The MCE can be viewed as a macroscale factor within this context. Drawing parallels from quantum theory, as particles exhibit individual behaviors on the micro level yet coalesce into predictable patterns on the macro level, the MCE emphasizes that individual mindful interactions within an organization (micro level) can collectively shape the larger organizational behaviors and outcomes (macro level).

Further, at the foundational layer, where individual actors as MCEs constitute the organization, these actors possess autonomy, signifying their capacity for independent action. Consequently, their exercised freedoms can influence the overarching organizational 'wholeness' [8]. Bohm's 'wholeness' underscores a reality of interconnectedness and non-fragmentation. Within these interrelations, the MCE emerges as a pivotal element. As Bohm discussed, the "implicate order is particularly suitable for understanding such unbroken wholeness in flowing movement in the implicate order; the totality of existence is enfolded within each region of space (and time). So, whatever part, element, or aspect we may abstract in thought, this still enfolds the whole and is therefore intrinsically related to the totality from which it has been abstracted. Thus, wholeness permeates all being discussed from the outset" [8].

Positioning this in organizational dynamics, individual actors identified as MCEs operate at the foundational layer of the system. Endowed with autonomy, these entities are not merely passive constituents but actively engage and shape the organizational structure. Their autonomy, as defined by their capacity for independent action, becomes a potent determinant in the evolution and flux of the organizational structure, resonating with Bohm's philosophy of the 'wholeness-in-motion' [8]. Hence, the actions and choices exercised by these MCEs are not isolated events but rather influential pivots that can substantially mould the overarching organizational architecture.

This paper is positioned at the confluence of ground-breaking theories proposed by Bohm and Prigogine, aiming to provide a fresh lens through which organizational dynamics can be discerned. The paper first plunges into a rigorous exposition of

Bohm's 'wholeness' and its profound repercussions on understanding organizational interconnectedness. This is promptly followed by an intricate exploration of Prigogine's views on the ever-fluctuating equilibrium, painting a vivid context against today's organizational complexities.

Building on these foundational pillars, the narrative then articulates these theoretical underpinnings into pragmatic organizational leadership and management strategies. The emphasis throughout these sections remains anchored in the trilogy of adaptability, foresight, and persistent recalibration. By weaving in comprehensive analyses and illustrative examples, this research sheds light on optimal strategies crucial for channeling the potential of self-organization.

This paper navigates the intricate nexus between cognition (mind) and the physical (matter), visualizing it as nestled within an expansive relational framework. Herein, the duality of reality—both at macro and micro levels—merges in the observer's consciousness, grounding entities within the linear continuum of spacetime (illustrated in Figure 1).

This paper proposes integrating mindful practices within corporate operations. This approach entails adopting mindful strategies to enhance corporate consciousness, which in turn aids in understanding and navigating temporal dynamics. Such an approach can significantly improve decision-making processes, bolster resilience against market volatility, and cultivate an enriched organizational culture firmly grounded in present awareness. The practical application of MCE involves leveraging mindfulness to foster a corporate environment that excels in current conditions and is proficient in anticipating and navigating future challenges.

This research delves into the complex dynamics of systemic well-being within organizational structures. It weaves together the concept of 'wholeness' proposed by Bohm with Prigogine's theories on equilibrium. Anchoring its inquiry in the realm of quantum mechanics, particularly principles such as superposition, non-locality, and entanglement, the study brings to light the inherent fluctuations in systems that exist in states far from equilibrium. It contrasts the traditional notions of stability, drawing on Prigogine's emphasis on the continuous divergence experienced by systems in their quest for equilibrium.

The primary objective of this paper is to propose contemporary organizational models that incorporate adaptable strategies, employing the concept of dissipative structures as a metaphorical framework. This approach offers a novel perspective on organizational dynamics, drawing from the principles of thermodynamics and complex systems theory. Dissipative structures, derived from nonequilibrium thermodynamics, maintain their stability and structure through the continuous exchange and dissipation of matter and energy with their environment [34]. By analogizing organizations to dissipative structures, the paper suggests that modern businesses can achieve sustained coherence and resilience by dynamically interacting with and adapting to their external environments. This metaphorical application underscores the necessity for MCEs within organizations to remain flexible and responsive to external changes, much like dissipative structures that thrive in far-from-equilibrium conditions.

Incorporating mindfulness into organizational practices by the MCEs offers a strategic solution to narrow the gap between current technological advancements and the growing needs for corporate sustainability and well-being. This integration fosters a revolutionary shift in organizational functioning, enhancing adaptability, responsiveness, and sustainability amidst fluctuating business environments. This demands heightened alertness, strategic alignment, collaborative synergy, and an unyielding commitment to augmenting value within the larger existential Bohm's 'wholeness'.

2. Literature Review

Prigogine and Stengers's seminal work, 'Order out of Chaos: Man's New Dialogue with Nature' [42], underscores the profound understanding that systems, while seemingly chaotic and disordered, have the inherent ability to evolve into coherent, ordered structures through self-organization. This foundational tenet of their insights resonates deeply with Bohm's 'wholeness', which posits that every component of a system, no matter how disparate or fragmented, contributes to a more significant, interconnected, and cohesive whole [8]. Bohm explained that wholeness is real and that fragmentation is the response of this whole to man's action, guided by illusory perception shaped by fragmentary thought. In other words, it is just because reality is whole that man, with his fragmentary approach, will inevitably be answered with a correspondingly fragmentary response. So what is needed is for man to give attention to his habit of fragmentary thought, to be aware of it, and thus bring it to an end. Man's approach to reality may then be whole, and so the response will be whole.

Prigogine's and Bohm's work invites us to perceive organizations as dynamic entities, perpetually navigating the precipice of order and chaos. For management theorization, this suggests a paradigm shift. Leaders should embrace organizations' dynamic, fluid nature instead of enforcing rigid structures or seeking to control every variable [23]. Recognizing that order can emerge organically from apparent disorder offers a more adaptive and resilient management approach. This also implies that in times of threat, disruption or upheaval, organizations can harness these moments instead of resisting change as opportunities for transformation and innovation [55] .

Incorporating quantum principles into the organizational framework profoundly augments our understanding of system dynamics and interconnectedness. Central to this paradigm is the quantum theory's notion of superposition, which posits that a quantum system can exist in multiple states simultaneously until observed [63]. When transposed to the domain of organizational studies, this concept suggests the possibility for an organization to embrace and juggle diverse strategies or states concomitantly, providing nuanced adaptability in a complex business landscape. Equally intriguing is the principle of non-locality in the quantum realm, which, as delineated by Einstein, Podolsky and Rosen, asserted that particles that have interacted in the past can instantaneously influence each other's states, regardless of their spatial separation [19]. Drawing parallels within an organizational context, this could signify that decisions or actions undertaken in one segment of the organization

have the potential to exert immediate repercussions across disparate units, transcending geographical constraints or conventional hierarchical paradigms.

Such quantum-inspired perspectives on organization dynamics challenge traditional linear and deterministic thinking and underscore the intricate web of relationships and influences within organizational systems. Recognizing organizational entities' multifaceted and profoundly interconnected nature equips leaders and scholars with more holistic strategies and insights, allowing for a more comprehensive grasp of the subtleties and intricacies inherent in today's evolving business environment.

2.1. Movement and Change: Organisational Flux through Bohm's Lens

Bohm's seminal work on the nature of reality introduced the notions of the 'implicate' (enfolded) and 'explicate' (unfolded) orders, offering a new lens through which one can perceive the interconnected and dynamic essence of the universe [8]. Bohm postulated that beneath the apparent chaos and randomness lies a deeper order, an 'unbroken wholeness', which has profound implications for understanding complex systems, including organizations.

"In this flow, mind and matter are not separate substances. Instead, they are different aspects of one whole and unbroken movement. In this way, we can look at all aspects of existence as not divided from each other. Thus, we can end the fragmentation implicit in the current attitude toward the atomic point of view, leading us to divide everything from everything thoroughly. Nevertheless, we can comprehend that the aspect of atomism still provides a correct and valid form of insight, i.e. that in spite of the undivided wholeness in flowing movement, the various patterns that can be abstracted from it have a certain relative autonomy and stability, which is indeed provided for by the universal law of the flowing movement" [8].

The discussion of randomness is pertinent in this context as it delves into the unpredictability inherent in organizational environments. The perception of randomness, often attributed to a lack of comprehensive understanding of our observational systems, is crucial in examining how organizations, akin to dissipative structures, respond to and manage seemingly stochastic external forces. By integrating this discussion, the paper aims to enhance the understanding of how MCEs within these organizational models can navigate and adapt to these random or unpredictable elements in their environment.

This approach aligns the concept of randomness with the overall research objective. It enriches the understanding of how contemporary organizations, modelled on dissipative structures, can effectively manage and thrive amidst the complexities and uncertainties of their operational landscapes. The inclusion of randomness in this context highlights the need for adaptability and responsiveness in organizational strategies, echoing the dynamic equilibrium characteristic of dissipative systems.

Within the framework of organizational studies, Bohm's conception of the implicate and explicate orders [8] has been instrumental in guiding scholars and practitioners to delve beyond organizations' superficial structures and processes. Rather than viewing an organization as merely an assembly of discrete parts operating

in isolation, this perspective encourages a recognition of the intricate web of interconnectedness, where each element is continually influencing and being influenced by the broader system. This holistic perspective challenges the traditional reductionist paradigm, which tends to compartmentalize organizational functions and operations [35]. Instead, Bohm's view urges a shift towards seeing organizations as fluid, evolving entities, marked more by their dynamic processes than their static structures.

The concept of the implicate order, sometimes described as the 'enfolded' realm, is posited as a profound and foundational stratum of reality. It exists beneath the immediately observable as the bedrock from which manifest phenomena emerge. On the other hand, the explicate order often termed the 'unfolded' domain, encompasses the tangible abstractions and perceptible realities with which humans typically engage. It represents the phenomena and occurrences that are readily discernible, stemming from the underlying processes and intricacies of the implicate order. This dichotomy between the implicate and explicate realms is a comprehensive framework for understanding the multifaceted layers of reality and the interrelationships between observed phenomena and their deeper, often unseen, origins. Bohm described a new notion of order that may be appropriate to a universe of unbroken wholeness. This is the implicate or enfolded order. In the enfolded order, space and time are no longer the dominant factors determining the relationships of dependence or independence of different elements. Instead, an entirely different sort of basic connection of elements is possible, from which our ordinary notions of space and time and those of separately existent material particles are abstracted as forms derived from the deeper order. These ordinary notions, in fact, appear in what is called the explicate or unfolded order, which is a special and distinguished form contained within the general totality of all the implicate orders [8].

Alhadeff-Jones's emphasis on the rhythms that underpin and shape various approaches to organizational crisis management [2] aligns with Bohm's thoughts. By 'rhythms', it is not merely a reference to the chronological unfolding of events but rather a deeper exploration of Bohm's implicate-explicate orders, the peaks and troughs, that crises often entail. This rhythmical appreciation provides insights into the cyclical nature of crises, allowing for the anticipation of potential challenges and the identification of emergent opportunities [11]. At the core of Alhadeff-Jones's arguments lies the importance of fostering a critical awareness that transcends the immediate manifestations of a crisis and delves into its underlying processes and temporal dynamics [2]. As Alhadeff-Jones suggested, such an awareness or consciousness at the deeper structure layer equips individuals and organizations with the understanding to discern between event-based and processual approaches, paving the way for responsive and preemptive strategies.

In broadening the understanding of crises, Alhadeff-Jones's work offers a more nuanced appreciation of their transformative capacity. Crises are not merely disruptions or disturbances; they are, in many ways, opportunities for renewal, recalibration, and growth. By recognizing the inherent fluidity of crises and by attuning to their underlying rhythms, organizations can navigate these challenges with greater agility, resilience, and foresight. Alhadeff-Jones's exploration into the nature

of crises and Bohm's implicate-explicate orders serve as a potent reminder of the intricate dynamics at play. It underscores organizations' need to remain adaptable, cultivating a critical awareness and prioritizing event-based and processual understanding of crises. Such an approach promises survival and the potential for transformation and growth amid adversity.

One of the cornerstone implications of Bohm's dual orders is the idea of perpetual motion and change. In its enfolded nature, the implicate order represents the underlying potentials and possibilities within the organization—its latent strategies, untapped resources, and dormant capabilities. Conversely, the explicate order signifies the manifest realities, the tangible outcomes, structures, and processes one can observe and measure [25].

In the context of organizational change and adaptability, the interplay between these two orders becomes especially salient. Bohm's framework suggests that the essence of an organization is not static. Instead, it is in continual flux, oscillating between the implicate and explicate, between potentiality and actualization. This perspective resonates with Stacey's insights on the dynamic nature of organizations, where change is not an anomaly but an inherent characteristic [54]. Furthermore, the concept of the implicate order serves as a reminder for organizational leaders and managers about the unseen, often unacknowledged, undercurrents that can shape the destiny of an organization. If properly understood and harnessed, these underlying patterns can be pivotal in steering an organization towards success, especially in a volatile business environment.

On the other hand, the explicate order serves as a mirror, reflecting the organization's current state, its strengths, weaknesses, and areas of opportunity. It underscores the importance of tangible' actions, strategies, and interventions in shaping the trajectory of an organization [62]. Bohm's notions of the implicate and explicate orders provide organizational scholars and practitioners with a profound framework to understand organizational entities' dynamic, interconnected nature. By acknowledging the interplay between the latent potentials and manifest realities, organizations can better navigate the challenges of an ever-evolving business landscape.

2.2. Self-organization and Organization as Living System: Prigofine's depiction of intrinsic structuring

The conceptual framework of self-organization is foundational to the study of complexity and has been employed to interpret diverse phenomena across natural and human-made systems. At the heart of this theoretical paradigm lies the contention that systems inherently gravitate towards heightened organization and complexity, devoid of external orchestration. The foundational contributions of Prigogine and Stengers have been instrumental in reshaping contemporary perceptions of this area [42]. Their work elucidates the intrinsic self-organizing nature of life, which is perpetually undergoing adaptive and transformative processes. This perspective underscores the emergent phenomena of order arising from seemingly chaotic environments, signifying life's innate ability to evolve and thrive amidst

unpredictability. This progressive understanding challenges traditional notions and provides a profound lens through which the dynamic interplay between order and chaos can be examined.

The principles of thermodynamics clarify the underlying drive for order in life. Prigogine, Ilya, Nicolis and Babloyantz opined that while isolated systems tend to incline towards entropy, living entities exhibit a remarkable counter-tendency [45]. They manifest patterns indicative of order and coherency, an outcome of their inherent interactions rather than any external directive. In a way, life defies the march towards disorder, orchestrating an innate dance towards coherence. Juxtaposing the intrinsic traits of life with organizational dynamics, scholars like Jansen, Cammock and Conner) advanced the idea that organizations are best understood as 'living systems' [29]. Such a perspective intimates that organizations, much like organic entities, are innately equipped to evolve, adapt, and sustain a semblance of equilibrium amidst potential volatility.

Drawing from the foundational ideas postulated by Prigogine and Stengers on self-organization, modern organizations can be conceptualized as dynamic, living systems that continuously adapt and evolve in response to their internal and external environments [42]. Prigogine used the example of chemistry: "The equations of chemistry are non-linear. When we rapidly push a chemical system away from equilibrium toward "disorder," or disequilibrium, the chemical reactions present us with what I call "bifurcation points"—points at which choices and new solutions appear. Generally, more than one solution appears, so that at the point of bifurcation, probability and self-organization come into play" [42].

Prigogine's theories elucidate how systems can spontaneously move towards increased complexity without external intervention and offer profound implications for understanding organizational behaviors and structures. Central to Prigogine's perspective is the understanding that systems innately gravitate towards structures that manifest higher order and complexity rather than maintaining static states. Analogously, contemporary organizations echo these principles through their inherent self-organization propensity, catalyzing forward-thinking developmental strategies.

An organization's future trajectory has historically been conceptualized through two predominant lenses. On the one hand, an optimistic paradigm posits that an organization is on an upward trajectory, characterized by strides in self-determination, corporate mission, and other progress markers [47]. Conversely, a more pessimistic viewpoint asserts that an organization is on an inexorable path towards catastrophe, driven by various socio-economic, political, and environmental factors [58]. However, upon critical reflection, it becomes apparent that both these perspectives may be overly reductionist and fail to capture the nuanced complexities inherent in organizational development. Arguably, the deterministic undertones of both views may not be entirely consistent with the unpredictable and emergent nature of organizational history and progress. In short, it might be an oversimplification to merely project the current state of affairs linearly into the future. Prigogine added "I prefer to look at this question in a different way. I believe that what we do today depends on our image of the future, rather than the future depending on what we do today. We build our equations by our actions. These equations, and the future they

represent, are not written in nature. In other words, time becomes construction. Of course, we have some conditions that determine limits of the future but within these limits are many, many possibilities" [44].

As complex entities navigating multifaceted landscapes, organizations reflect the overarching intricacies of the systems within which they operate. Inherent to this understanding is recognizing our existence within non-deterministic systems, where the future remains unpredictable, mainly based on present conditions. The events unfolding are influenced by many variables, which can be unforeseen and diverge from current trajectories. Such a perspective mandates a nuanced approach that acknowledges the present yet remains adaptive and resilient in the face of unforeseeable eventualities.

Deepening this insight, Prigogine's conceptualization of self-organization finds resonance. Organizations, akin to living systems, embody the quintessence of self-organization—evolving, adapting, and transforming in response to environmental stimuli [46]. This organic evolution suggests an undercurrent favoring emergent structures, champions participatory frameworks, and gravitates towards decentralized configurations, as Prigogine & Stengers posited that systems inherently veer towards complexity, not as a deterministic outcome but as a continual adaptation to prevailing circumstances [46].

Building on this foundational understanding, the role of inter-organizational relationships emerges as pivotal. In a world characterized by interdependencies, such collaborative networks amplify the potential of individual entities. Human and Provan captured this sentiment, noting that the interplay of transactional and transformational outcomes emerges as a consequence of effective network participation [27]. Beyond mere economic gains, these collaborations herald transformative shifts—strategic, operational, and philosophical.

The convergence of Prigogine's self-organization concept with the rich tapestry of inter-organizational collaborations positions organizations as dynamic living systems. These entities, reminiscent of living organisms, pulsate with energy, seeking congruence, efficiency, and evolutionary adaptation, tending towards order in chaos. The essence of such organizations lies in their innate ability to adapt and self-organize, encapsulating the dynamic continuum of life itself.

2.3. The concept of Dissipative Structure and their Equilibrium

One of Prigogine's most celebrated contributions is the notion of 'dissipative structures', which "arise in open systems, exchanging energy and matter with the outside world when driven far from equilibrium" [44]. Such structures, he described, achieve stability not by insulating themselves but through continuous energy and matter interchange with the environment. These systems, therefore, stand as a testament to the dynamic balance of life, sustained amidst the ever-present vicissitudes of its environment, deriving energy and matter externally to stay balanced [42]. The essence of homeostasis in confronting external perturbations finds its roots in this paradigm.

Perceiving contemporary organizations through this prism portrays them as dissipative structures, incessantly reciprocating with their environment. Organizations that internalize the principles of self-organization are poised to navigate the labyrinthine nuances of today's business terrains. "It is the processes associated with randomness, openness, that lead to higher level of organization, such as dissipative structures" [42] .

In the context of dissipative structures, the underlying theory of change posits that a system pushed into a state far from equilibrium by fluctuations faces potential structural threats. Such a system approaches a pivotal juncture, often termed a 'bifurcation point'. Prigogine and Stengers contended that predicting the system's subsequent state at this juncture is fundamentally untenable [42]. The element of randomness or chance directs the system's remnants towards a new evolutionary trajectory.

From a quantum perspective, the idea of unpredictability and the entanglement of events within spacetime can be explored further through the phenomenon of quantum entanglement. Quantum entanglement challenges traditional notions of separability and locality in the quantum realm, where paired particles become intertwined so that the state of one instantly influences the state of the other, irrespective of the distance that separates them [28].

In recent discourse on the intersection of quantum mechanics and our understanding of randomness within spacetime, scholars have begun to explore deeper entangled relationships that underpin what many have historically deemed as purely chance occurrences [24][39]. Such explorations challenge traditional frameworks, suggesting that previously seen as isolated or 'random' events could be manifestations of intricate, entangled relationships within the foundational quantum fabric of reality [24]. This perspective hints at the limitations of human observation, grounded in bounded rationality, and points to the potentially vast and interconnected realm that lies beyond our immediate perception [57].

These considerations naturally extend to broader discussions about the universe's nature of randomness and chance. Classical interpretations of events, particularly within the macroscopic world, have often been framed within deterministic paradigms [53]. However, quantum mechanics insights introduce superposition, wherein entities do not possess definitive states until observed or measured [60].

Recent scholarly investigations explore the interplay between quantum mechanics and conventional perceptions of randomness [20], [41], [56]. Central to this discourse is a reconceptualization of randomness not as an inherent unpredictability but rather as a manifestation of the limitations inherent in human observation [40]. Drawing from the principles of quantum mechanics, Bickley, Chan, Schmidt and Torgler asserted an emerging consensus that phenomena, previously interpreted as 'random', may reflect our observational constraints rather than an intrinsic characteristic of the event itself [6]. At the heart of this argument lies the concept of superposition, a fundamental tenet of quantum mechanics. Within this framework, entities or events do not necessarily traverse a predetermined path. Instead, they exist in multiple potentialities, with a definitive state only realized upon measurement or observation [18]. Ananthaswamy posited that these potential states remain in flux, and observation

precipitates the collapse of these superpositions into a singular, discernible outcome [4]. The implications of such a perspective are profound. It suggests a paradigm shift in how randomness is interpreted, moving away from the traditional deterministic view and embracing a more fluid understanding rooted in quantum principles. This perspective offers a challenge to conventional wisdom, asserting that the universe's seeming unpredictability might be a by-product of the constraints of our observational methodologies rather than the inherent nature of the phenomena under scrutiny.

In essence, the ongoing exploration of quantum mechanics' role in shaping perceptions of randomness underscores the evolving nature of scientific understanding. As scholars continue to interrogate the intersections of observation, superposition, and indeterminism, it becomes increasingly evident that traditional notions of randomness may require substantial re-evaluation in light of quantum insights.

Furthermore, from quantum mechanics, non-locality suggests an interconnectedness that defies traditional spatial and temporal boundaries. In this context, events that might seem isolated or distinct within our macroscopic understanding could be deeply entangled at the quantum level [31]. This implies that occurrences often dismissed as mere chance may be products of complex quantum interactions woven together in ways that challenge our conventional understanding of cause and effect.

In summary, the interplay between quantum mechanics and perceived randomness in spacetime is a compelling reminder of the limitations of human perception and the potential complexities that underlie the universe's seemingly random events. As research in this domain continues, it holds promise for redefining our understanding of interconnectedness, causality, and the very nature of reality itself. Moreover, the idea that these 'random' events are deeply connected to their surrounding context, both spatially and temporally, further mirrors quantum entanglement's disregard for spatial or temporal constraints. This leads to a fascinating implication: just as entangled particles are inexorably linked, so too might events be bound together in the spacetime continuum, making their appearances of randomness or chance merely a consequence of our current understanding or observational limitations.

In conclusion, drawing parallels between the unpredictability of events within spacetime and the principles of quantum entanglement offers a profound reimagining of randomness. It posits a universe where chance events are not merely arbitrary happenstances but are complexly woven threads in the vast tapestry of spacetime, influenced by and influencing other events in ways that we are only beginning to comprehend.

3. Discussion

In organizational studies, workplace creativity is defined as the intentional generation, utilization, and execution of innovative concepts tailored to enhance an organization's outcomes [32]. This paper elaborates on the relationship between self-organization and human creativity, both individual and collective. Self-organization is grounded

in fundamental principles often beyond our complete understanding, as evidenced by the numerous natural systems that self-organize independently of human intervention [5]. It is essential to distinguish between two primary dimensions of creativity in this context.

The first dimension pertains to creativity in organizational outputs, encompassing the development of innovative products and services [13]. This aspect focuses on how creativity drives the external manifestations of an organization's work, contributing to market differentiation and customer engagement.

The second dimension involves creativity in the organization's internal functioning. Here, creativity is embedded in decision-making and problem-solving processes, fostering a culture of innovation and adaptability within the organization. This internal creativity is crucial for the organization's self-organization, enabling it to respond dynamically to internal and external challenges and opportunities [3].

By explicitly delineating these two facets of creativity, the paper provides a comprehensive understanding of how creativity permeates different aspects of an organization. This distinction is vital for appreciating the multifaceted role of creativity in driving an organization's tangible outputs and the intangible processes that underlie its operations and strategy. Such a clarification enriches the discussion on organizational dynamics, offering a more nuanced view of the interplay between creativity, self-organization, and the fundamental principles governing natural and corporate systems.

Numerous scholars and experts acknowledge the pivotal role that such creativity plays in bestowing a competitive edge upon organizations, resulting in superior solutions for client-related challenges and augmenting organizational efficacy [10] . A comprehensive examination of the empirical literature on this subject unveils many factors that catalyze creative processes. For instance, cognitive predispositions emerge as significant drivers [49] alongside the overarching organizational environment and setting [64]. Furthermore, the symbiotic interplay of social support and a collective vision has been pinpointed as vital for fostering creativity [22]. In their seminal study, Fischer, Malycha and Schafmann rigorously expounded upon specific elements, including operational autonomy, social support, and the accessibility of resources, as pivotal determinants influencing the magnitude and quality of creative pursuits among employees [22]. This research investigated the interplay of extrinsic motivators that, when synergized, have the potential to augment the creativity and innovation exhibited by intrinsically driven knowledge professionals within organizational settings.

In dynamic organizational settings, political, environmental, and economic considerations introduce heightened layers of complexity [16] . Discerning the roots of such uncertainties becomes paramount, given their profound influence on systems, processes, people and their relationships [17]. Recognizing a lacuna in contemporary literature, the present research endeavors to unpack the three cardinal environmental uncertainties: (1) the evolution of systems, processes, and information, (2) the organizational culture sculpted by its members and their interrelationships, and (3) the relational nexus between agents and artefacts within the sphere of organizational creativity, as depicted in Figure 1. Grounding this examination in the conceptual

framework of entanglement, the study aims to shed light on the dynamics between environmental uncertainty and the organizational creativity of MCEs. The notion of the MCE emerges as particularly pertinent in this discourse, given its potential to act as a locus of interaction, bridging gaps, fostering understanding, and thus sculpting the fabric of interrelatedness among stakeholders. By their inherent mindfulness, such entities could be pivotal in ensuring that interactions at the micro-level coalesce into a macro-level harmony, serving as the nexus that anchors and navigates the multifaceted landscape of uncertainties. This highlights the essentiality of recognizing and fostering such entities, as they not only form the tangible and intangible linkages within an organization but also crucially influence how uncertainties are perceived, interpreted, and addressed.

This study acknowledges the fundamental idea that uncertainty is pervasive through time. Time is viewed linearly in a normal state of awareness. Observers experience time as a linear progression of events from the past to the present and into the future, and in all stages, the observer is in varying states of uncertainty [1].

This, in turn, obscures and complicates organizational learning pathways and the ability to foresee future trajectories. While a well-documented correlation exists between uncertainties and heightened creative impulses, it also reveals an inherent potential for setbacks. Thus, the emphasis on MCEs serves as a beacon in navigating this vast and intricate academic terrain. These entities encapsulate the essence of adaptive organizations, equipped to manage the tides of change and uncertainty, ensuring that creativity is not stifled but flourishes amidst the chaos. Understanding the role and impact of such entities could offer transformative insights into how organizations can build resilience, foster innovation, and navigate the challenging terrains of the future.

Given this backdrop, unveiling strategies that empower organizational decision-makers or the MCE as a central node at the intersections of myriad possibilities is imperative. Against this complex landscape, the necessity for elucidating methodologies that enable organizational leaders to traverse these environmental ambiguities while steadfastly upholding an organizational ethos that champions systemic well-being gains profound significance. In expanding this discourse, the MCE, as conceptualized within organizational paradigms, is not just strategically positioned but is intrinsically woven into these junctures of potentialities. Entities imbibed with mindfulness are not merely passive bystanders but proactive catalysts, shaping and being shaped by these potentialities. Thus, beyond its conventional boundaries, the entrepreneurial spirit becomes instrumental in actualizing these latent possibilities, underscoring the vital interplay between entrepreneurial foresight and the mindful organizational entity in charting innovative pathways.

Organizational agility, as characterized by Walter, encapsulates an entity's dexterity in adapting or evolving in resonance with environmental shifts [59]. This agility extends to mastering control amidst relentless, rapid changes in the external environment. The assertion stands that organizations that embody a structure resonating with environmental fluctuations witness augmented success. This paper argues that organizational agility is a pivotal catalyst, supporting innovation and adeptly navigating uncertainties. Consequently, a secondary objective of this research

pivots around assessing organizational agility as a strategic tool for mitigating the adverse impacts of environmental uncertainty on organizational creativity.

In synthesizing the outcomes, this investigation contributes to the academic discourse. Firstly, it elucidates the intricate manner in which diverse, cardinal environmental uncertainties impinge upon organizational creativity. Secondly, it augments the literature on organizational creativity, spotlighting organizational agility's moderating prowess in counterbalancing the ramifications of environmental uncertainty. From a pragmatic lens, the insights garnered proffer invaluable managerial takeaways, underscoring the recognition of varied environmental uncertainty sources and how organizational agility can bolster creativity amidst the volatile environment.

3.1. The MCE: A Theoretical Exploration in Organizational Studies

The introduction of the Mindful Corporate Entity (MCE) marks a significant conceptual advancement in our understanding of organizational dynamics. The MCE is positioned within a temporal and spatial context characterized by conscious observation. Within this framework, the MCE actively interacts with various resources and agents, forming part of a complex network of relationships. This conceptualization of MCEs places them at the nexus of agents, artefacts, roles, and relationships, all embedded within broader systems and processes. This conceptual framework is visually depicted in Figure 1. Figures 2 and 3 further clarify this concept by presenting a plain perspective of the interactions, highlighting how these self-organizing MCEs, endowed with consciousness, operate within the present spacetime. This model challenges the traditional view of organizations as static, uniform structures, proposing that they function as dynamic ecosystems instead. Within these ecosystems, multiple agents, primarily individuals, participate in a reciprocal exchange of influences. This ongoing interaction perpetually shapes and reshapes both the agents and the organizational environment.

Moreover, these entities are pivotal at a micro-level, driving complex and nuanced interactions that form the foundation of organizational relationships. These interactions, in turn, help to define the overall structure of organizational connectedness and coherence, as illustrated in Figure 1. By featuring the dynamics' complexities and understanding the relational matrix and the effects of these interactions, we can gain transformative insights into organizational behavior. Such insights are crucial for comprehending how organizations navigate challenges, fostering an environment conducive to innovation and resilience.

Within the ambit of organizational studies, this paper discusses the profound implications of the MCE situated in the nexus of organizational resources, personnel, networks, and artefacts. These entities are deeply entangled in a distinctive spacetime fabric, establishing intricate connections and relations that shape organizational dynamics (Figure 3). When an organization solidifies its direction via instrumental frameworks such as business models, key performance indicators (KPIs), or objective key results (OKRs), it often inadvertently integrates elements of serendipity, manifesting as unforeseen entanglements. These might encompass unpredictable

shifts in resource allocation, emerging networks, or the intricate interactions of actors and artefacts. Once these trajectories are set in motion, deterministic paradigms – rooted in structure, predictability, and established organizational theories – invariably take the helm, stewarding the organization until it reaches a subsequent inflexion point or bifurcation.

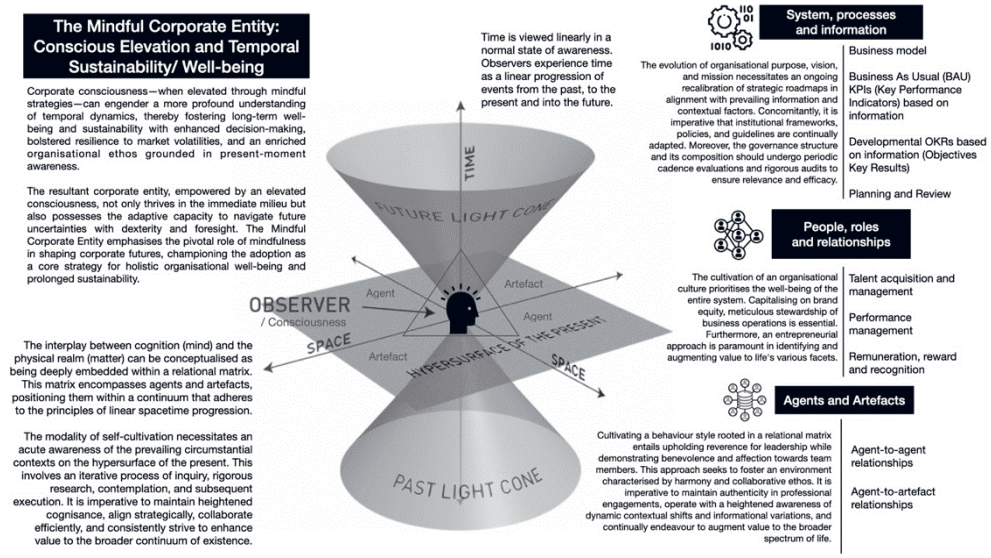


Figure 1. Observer's consciousness, grounding entities within the linear continuum of spacetime

"Quantum science describes the complex interactions, entanglements, and interferences of wave functions under such uncertainties from a different perspective, as presented by the classical interpretation. The embeddedness of potentiality and the many-possibilities scenarios at each junction, boundary or nexus of interactions, including the individual-opportunity nexus, hold great promise in entrepreneurship research. Adopting the metaphors and methods of the quantum theory has refreshingly new perspectives for entrepreneurial studies" [33].

The MCE' construct becomes particularly salient when exploring the nexus of possibilities and their manifestation in business contexts. Central to this discussion is the corporate entity as part of Bohm's 'wholeness-in-motion', which is not merely an observer but an active participant and shaper of organizational trajectories [8]. For the MCE, the role transcends mere observation. The act of conscious decision-making, often synonymous with venturing into uncertain realms of possibility, propels the organization onto a distinct path, paving the way for potential futures. Drawing from quantum theory, the MCE's engagement with business phenomena represents a quantized relationship, signifying a profoundly interconnected and consequential interaction with the event. In contrast, in this context, a classical observer or a non-entrepreneur remains disengaged, serving only as a passive spectator, detached from the unfolding narrative.

Expanding upon this, the MCE is posited at a crossroads of possibilities due to entanglement. Mindful and attuned entities recognize these crossroads of potential and actively harness them, deftly navigating the myriad possibilities they present with a corporate consciousness. When elevated through mindful strategies, corporate consciousness can engender a more profound understanding of temporal dynamics, fostering long-term well-being and sustainability with enhanced decision-making, bolstering resilience to market volatilities, and an enriched organizational ethos grounded in present-moment awareness. The resultant corporate entity, empowered by an elevated consciousness, not only thrives in the immediate milieu but also possesses the adaptive capacity to navigate future uncertainties with dexterity and foresight. The MCE emphasizes the pivotal role of mindfulness in shaping corporate futures, championing the adoption as a core strategy for holistic organizational well-being and prolonged sustainability. The interplay between cognition (mind) and the physical realm (matter) can be conceptualized as deeply embedded within a relational matrix. This matrix encompasses agents and artefacts, positioning them within a continuum that adheres to the principles of linear spacetime progression. The modality of self-cultivation necessitates an acute awareness of the prevailing circumstantial contexts on the hypersurface of the present. This involves an iterative inquiry process, rigorous research, contemplation, and subsequent execution. It is imperative to maintain heightened cognizance, align strategically, collaborate efficiently, and consistently strive to enhance value to the broader continuum of existence.

This underscores the pivotal role of MCEs in sensing, seizing, and shaping these potential futures, distinguishing them from mere observers. Through their proactive engagements with other actors and artefacts, they inscribe meaning and direction to the otherwise abstract realm of possibilities, rendering the MCE a dynamic and evolving force within the organizational fabric. Cultivating a behavior style rooted in a relational matrix entails upholding reverence for leadership while demonstrating benevolence and affection towards team members. This approach seeks to foster an environment characterized by harmony and collaborative ethos. Maintaining authenticity in professional engagements is imperative, operating with a heightened awareness of dynamic contextual shifts and informational variations and continually endeavoring to augment value to the broader spectrum of life.

Contrasting earlier organizational theories, which demarcated chance and determinism into siloed categories, these perspectives contend that they operate in tandem. Chance and determinism are not mutually exclusive but operate as intertwined forces, collaboratively sculpting the contours of organizational evolution. By extrapolating this argument, the MCE emerges as a paradigm of significance. It underlines how organizations, in their quest for success and stability, must navigate the intertwined maze of resources, human dynamics, and interconnected networks or artefacts – all deeply entangled within the unique spacetime continuum of the organizational setting. The success of an organization hinges on its adeptness at maneuvering through these entanglements, harmonizing chance with determinism, and ensuring adaptability within its structural confines.

Through this lens, the MCE becomes a focal point, underscoring the importance of recognizing and nurturing these nuanced interactions that not only form the relational core of organizations but also serve as the nexus between classic organizational theories and emerging paradigms. In the interplay between structure and adaptability, these entities stand as testaments to the future trajectory of organizational research and practice.

This theoretical framework asserts that the traditional macroscopic lens, which often views an organization as a singular, homogenized entity, is merely the superficial layer of a more intricate organizational tapestry. Drawing parallels with the concept of granularity in physical systems, the MCE posits that an organization's essence is captured at a more microscopic level, where its constituents' interactions, consciousness, and mindfulness play a pivotal role. Each individual within the organization, equipped with their consciousness and agency, can influence and, in turn, be influenced by the broader organizational context. The term 'mindful' in the MCE is not just an ornamental adjective; it underscores individuals' heightened cognizance and intentional awareness of their roles, interactions, and decision-making processes.

Such an approach challenges traditional organizational theories that often marginalize individual agency in favor of broader structural or systemic factors. By placing individuals and their collective mindfulness at the center of the organizational matrix, the MCE framework emphasizes the symbiotic relationship between individual mindfulness and the emergent properties of the organization as a whole.

The implications of adopting the MCE perspective in organizational studies are profound. Recognizing the organization as a dynamic interplay of mindful entities calls for re-evaluating organizational strategies, leadership models, and operational paradigms. It accentuates the need to foster environments that nurture individual mindfulness, catalyzing collective organizational mindfulness and leading to enhanced adaptability, resilience, and innovation.

In conclusion, the MCE provides a fresh lens for understanding and navigating the complexities of modern organizations, urging scholars and practitioners alike to delve deeper into the granular interstices of organizational life and recognize the pivotal role of mindfulness in shaping and steering organizational trajectories.

3.2. Creation, Cessation and Chance

The interplay of creation, cessation, and chance emerges as a poignant area of inquiry, particularly when contextualized within the purview of quantum mechanics and its philosophical ramifications. The MCE offers a compelling lens through which to examine this triadic relationship illustrated in Figure 2.

Chance, often perceived as stochastic occurrences [50], intriguingly evokes parallels with the quantum phenomenon of entanglement, wherein disparate particles remain inextricably interconnected irrespective of spatial or temporal expanses. Dane explained that the cognitive mechanisms that underpin 'chance' are elusive, prompting individuals to anchor their narratives predominantly around the contextual environment [15]. A recurring theme in these narratives suggests a predisposition to

ascribe 'chance' to fortuitous events or serendipitous circumstances. Sætre and Van de Ven proposed a generating theory by abduction.

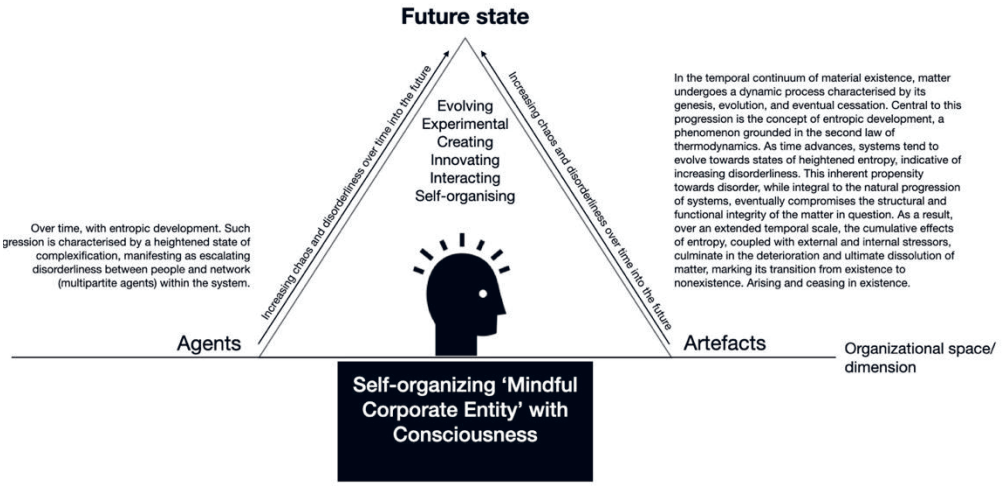


Figure 2. The creation and cessation process over time and into the future

"Abduction provides a mode of reasoning for achieving this. It is a form of generative reasoning that begins with observing and confirming an anomaly, and generating and evaluating hunches that may explain the anomaly, for subsequent deductive constructing and inductive testing" [50].

Within this framework, what might superficially present as random events or chances are, in essence, deeply embedded within their contextual fabric, both in space and time. This quantum perspective invites a profound re-evaluation: analogous to how entangled particles defy conventional understanding with their uncanny synchronicity, events within the organizational space, governed by 'chance', may indeed be tethered in the spacetime matrix, their perceived randomness a testament to our extant observational constraints rather than intrinsic unpredictability. Transcending mere observational limitations, the MCE, armed with this quantum-informed discernment, appreciates the nuanced dance of creation (the birth of events or opportunities), cessation (their eventual dissipation or demise), and chance (the seemingly random orchestrations undergirded by more profound interconnections). The MCE is evolving, experimental, creating, innovating, interacting and self-organizing (illustrated in Figure 2); therefore, it does not merely navigate these dimensions but actively harnesses this triadic synergy, recognizing that in the intricate ballet of creation, cessation, and chance, lie profound insights and untapped potentials awaiting future discovery.

This recognition of transience, while seemingly nihilistic, is instrumental in fostering agility and resilience. The inherent temporality of existence is not a constraint but a compass, guiding entities towards sustainable models grounded in perpetual transformation, changes, renewal and recalibration in a dynamic continuum

that oscillates away from equilibrium, implicating and explicating yet inherently exhibiting a pursuit towards homeostatic balance.

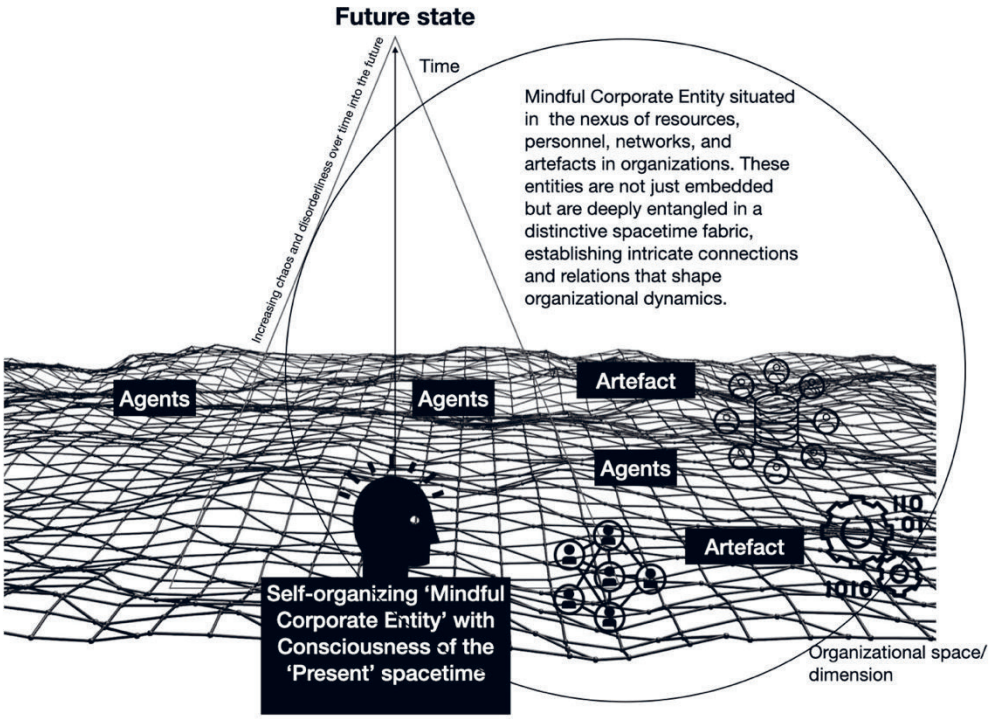


Figure 3. This figure situates the MCE in the 'Present' spacetime

Figure 3 graphically clarifies the phenomenon of systemic fluctuations as organizations endeavor to attain a state of balance. Concurrently, Figure 4 accentuates similar fluctuations within the broader environmental context. Figure 3 illustrates the equilibrium achieved through a continuous interplay between variance and stabilization within organizational structures. These systems are depicted as being in a perpetual state of flux, oscillating between the implicate and explicate orders and navigating the complex interplay of potentiality and actualization. Such visual representations underscore the fundamentally dynamic nature of organizations, where adaptability and continuous evolution are inherent characteristics rather than anomalies. Furthermore, the concept of the implicate order is especially salient for organizational leaders and strategists, as it highlights the underlying, often unseen forces that can shape an organization's path. Understanding and effectively harnessing these latent dynamics is crucial for steering an organization toward its goals, especially in an unpredictable and ever-changing external landscape.

These illustrations serve not just as visual aids but as conceptual tools that encapsulate organizations' fluid and evolving essence, emphasizing the importance of agility, foresight, and a deep understanding of manifest and latent organizational dynamics in achieving sustained success.

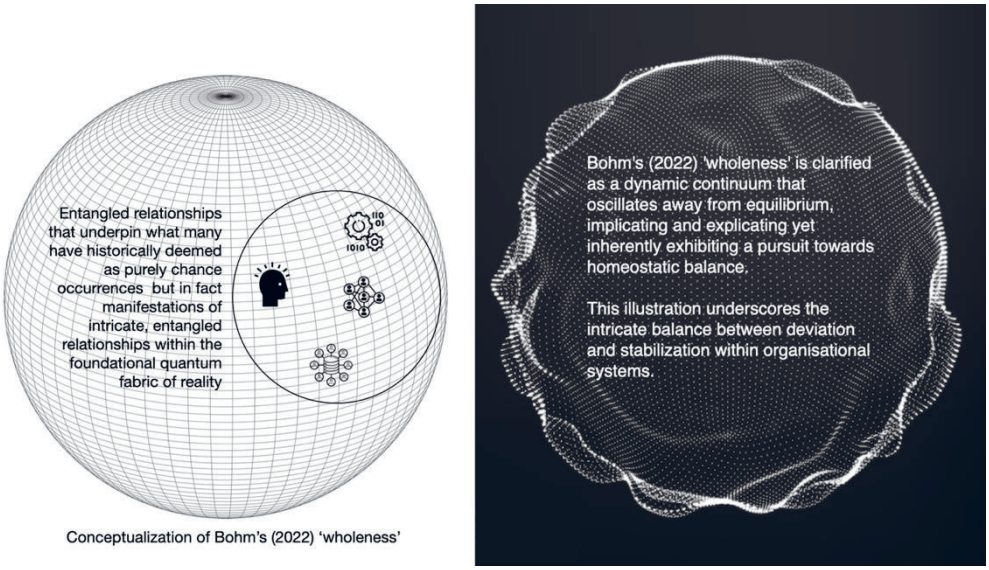


Figure 4. The concept of Bohm's 'Wholeness'

Thus, the MCE, in its sagacious embrace of creation, cessation, and chance, emerges as a paragon of adaptability, perpetually attuned to the cadences of time and the caprices of entropy.

4. Conclusion

This paper discussed the systemic well-being of organizations, marrying Bohm's concept of 'wholeness' with Nicolis and Prigogine's equilibrium theory and applying quantum principles to decipher organizational behaviors. This approach provides a richer understanding of organizations, transcending traditional deterministic outlooks by embracing the 'Organizational Homeostasis' concept. This concept, drawn from biological principles, illustrates how organizations, like living systems, continually recalibrate their strategies, operations, and policies to achieve a dynamic equilibrium. Such homeostasis reflects the capacity of organizations to employ feedback for self-regulation, embody resilience, balance internal resources, and enact effective change management in response to external market pressures and internal challenges. As organizations manoeuvre through the constant oscillations of the modern business environment, they confront the imperative to maintain stability while promoting growth.

The insights garnered from this research underscore the indispensable function of mindfulness and enhanced corporate consciousness within the Mindful Corporate Entity (MCE) framework. These entities, bolstered by acute consciousness and organizational coherence, are adept at flourishing in the immediate complexities of today's corporate environment and are prescient in addressing forthcoming uncertainties, ensuring their enduring significance and operational health.

To conclude, the study foregrounds the MCE as an essential element of an inter-relational system. Characterized by localized interactions and expansive nonlocal awareness, MCEs echo Bohm's vision of wholeness, illustrating the integrated nature of organizational components. Moreover, the principle of organizational homeostasis ensures that entities are not merely reactive but are strategically primed to foster environments of comprehensive well-being and agility with sustainable flourishing businesses within the fluid global ecosystem.

References

[1] Ahlqvist, T., & Uotila, T. (2020). Contextualising weak signals: Towards a relational theory of futures knowledge. *Futures*, *119*, 102543. https://doi.org/10.1016/j.futures.2020.102543

[2] Alhadeff-Jones, M. (2021). Learning from the whirlpools of existence. *European Journal for Research on the Education and Learning of Adults*, *12*(3), 311–326. https://doi.org/10.3384/rela.2000-7426.3914

[3] Amabile, T. M. (1988). A model of creativity and innovation in organisations. *Research in Organisational Behavior*, *10*(1), 123–167.

[4] Ananthaswamy, A. (2023). Particle, wave, both or neither? The experiment that challenges all we know about reality. *Nature*, 454–456.

[5] Anzola, D., Barbrook-Johnson, P., & Cano, J. I. (2017). Self-organisation and social science. *Computational and Mathematical Organisation Theory*, *23*(2), 221–257. https://doi.org/10.1007/s10588-016-9224-2

[6] Bickley, S. J., Chan, H. F., Schmidt, S. L., & Torgler, B. (2021). Quantum-sapiens: the quantum bases for human expertise, knowledge, and problem-solving. *Technology Analysis & Strategic Management*, *33*(11), 1290–1302. https://doi.org/10.1080/09537325.2021.1921137

[7] Bohm, D. (1980). Wholeness and the Implicate Order. In *Routledge*. https://doi.org/10.1093/bjps/32.3.303

[8] Bohm, D. (2002). Wholeness and the implicate order. In *Psychology Press.* (Vol. 31, Issue 10). Psychology Press. https://doi.org/10.1088/0031-9112/31/10/042

[9] Bowles, M. L. (1990). Recognising Deep Structures in Organisations. *Organisation Studies*, *11*(3), 395–412. https://doi.org/10.1177/017084069001100304

[10] Byttebie, I., & Vullings, R. (2015). *Creativity in Business: The Basic Guide for Generating and Selecting Ideas*. Igor Byttebier, Ramon Vullings.

[11] Caligiuri, P., De Cieri, H., Minbaeva, D., Verbeke, A., & Zimmermann, A. (2020). International HRM insights for navigating the COVID-19 pandemic: Implications for future research and practice. *Journal of*

International Business Studies, *51*(5), 697–713. https://doi.org/10.1057/s41267-020-00335-9

[12] Cannon, W. B. (1929). ORGANISATION FOR PHYSIOLOGICAL HOMEOSTASIS. *Physiological Reviews*, *9*(3), 399–431. https://doi.org/10.1152/physrev.1929.9.3.399

[13] Chen, K. K. (2012). "Organising Creativity: Enabling Creative Output, Process, and Organizing Practices." *Sociology Compass*, *6*(8), 624–643. https://doi.org/10.1111/j.1751-9020.2012.00480.x

[14] Correia, P. S., Obando, P. C., Vallejos, R. O., & de Melo, F. (2021). Macro-to-micro quantum mapping and the emergence of nonlinearity. *Physical Review A*, *103*(5), 052210. https://doi.org/10.1103/PhysRevA.103.052210

[15] Dane, E. (2020). Suddenly Everything Became Clear: How People Make Sense of Epiphanies Surrounding Their Work and Careers. *Academy of Management Discoveries*, *6*(1), 39–60. https://doi.org/10.5465/amd.2018.0033

[16] Darvishmotevali, M., Altinay, L., & Köseoglu, M. A. (2020). The link between environmental uncertainty, organisational agility, and organisational creativity in the hotel industry. *International Journal of Hospitality Management*, *87*, 102499. https://doi.org/10.1016/j.ijhm.2020.102499

[17] de Lima, F. A., Seuring, S., & Sauer, P. C. (2022). A systematic literature review exploring uncertainty management and sustainability outcomes in circular supply chains. *International Journal of Production Research*, *60*(19), 6013–6046. https://doi.org/10.1080/00207543.2021.1976859

[18] De Ronde, C. (2018). Quantum Superpositions and the Representation of Physical Reality Beyond Measurement Outcomes and Mathematical Structures. *Foundations of Science*, *23*(4), 621–648. https://doi.org/10.1007/s10699-017-9541-z

[19] Einstein, A., Podolsky, B., & Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, *47*(10), 777–780. https://doi.org/10.1103/PhysRev.47.777

[20] Everett, H. (1957). "Relative State" Formulation of Quantum Mechanics. *Reviews of Modern Physics*, *29*(3), 454–462. https://doi.org/10.1103/RevModPhys.29.454

[21] Faye, J. (1994). *Non-Locality or Non-Separability?* (pp. 97–118). https://doi.org/10.1007/978-94-015-8106-6_5

[22] Fischer, C., Malycha, C. P., & Schafmann, E. (2019). The Influence of Intrinsic Motivation and Synergistic Extrinsic Motivators on Creativity and

Innovation. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00137

[23] Gilbert, C. G. (2005). Unbundling the Structure of Inertia: Resource Versus Routine Rigidity. *Academy of Management Journal*, *48*(5), 741–763. https://doi.org/10.5465/amj.2005.18803920

[24] Healey, R. (2020). A pragmatist view of the metaphysics of entanglement. *Synthese*, *197*(10), 4265–4302. https://doi.org/10.1007/s11229-016-1204-z

[25] Hiley, B., & Peat, F. D. (2012). *Quantum implications: Essays in honour of David Bohm*. Routledge.

[26] Hinde, R. A. (1976). Interactions, Relationships and Social Structure. *Man*, *11*(1), 1. https://doi.org/10.2307/2800384

[27] Human, S. E., & Provan, K. G. (1997). An Emergent Theory of Structure and Outcomes in Small-Firm Strategic Manufacturing Networks. *Academy of Management Journal*, *40*(2), 368–403. https://doi.org/10.5465/256887

[28] Hüttemann, A. (2005). Explanation, Emergence, and Quantum Entanglement. *Philosophy of Science*, *72*(1), 114–127. https://doi.org/10.1086/428075

[29] Jansen, C., Cammock, P., & Conner, L. (2011). Leadership for emergence: Exploring organisations through a living system lens. *Leading and Managing*, *17*(1), 59–74.

[30] Karatsoreos, I. N., & McEwen, B. S. (2013). Annual Research Review: The neurobiology and physiology of resilience and adaptation across the life course. *Journal of Child Psychology and Psychiatry*, *54*(4), 337–347. https://doi.org/10.1111/jcpp.12054

[31] Kotler, S., Peterson, G. A., Shojaee, E., Lecocq, F., Cicak, K., Kwiatkowski, A., Geller, S., Glancy, S., Knill, E., Simmonds, R. W., Aumentado, J., & Teufel, J. D. (2021). Direct observation of deterministic macroscopic entanglement. *Science*, *372*(6542), 622–625. https://doi.org/10.1126/science.abf2998

[32] Kutieshat, R., & Farmanesh, P. (2022). The Impact of New Human Resource Management Practices on Innovation Performance during the COVID 19 Crisis: A New Perception on Enhancing the Educational Sector. *Sustainability*, *14*(5), 2872. https://doi.org/10.3390/su14052872

[33] Leong, D. (2022). Probabilistic Interpretation of Observer Effect on Entrepreneurial Opportunity. *Organizacija*, *55*(4), 243–258. https://doi.org/10.2478/orga-2022-0016

[34] Macintosh, R., & Maclean, D. (1999). Conditioned emergence: a dissipative structures approach to transformation. *Strategic Management*

Journal, *20*(4), 297–316. https://doi.org/10.1002/(SICI)1097-0266(199904)20:4<297::AID-SMJ25>3.0.CO;2-Q

[35] McMillan, E. (2003). *Complexity, organisations and change*. Routledge.

[36] Mitroff, I. I. (1983). Archetypal Social Systems Analysis: On the Deeper Structure of Human Systems. *Academy of Management Review*, *8*(3), 387–397. https://doi.org/10.5465/amr.1983.4284373

[37] Naikar, N. (2020). Distributed Cognition in Self-Organizing Teams. In *In Contemporary Research* (pp. 75–94). CRC Press.

[38] Nicolis, G., & Prigogine, I. (1971). Fluctuations in Nonequilibrium Systems. *Proceedings of the National Academy of Sciences*, *68*(9), 2102–2107. https://doi.org/10.1073/pnas.68.9.2102

[39] Orkin, M. (2022). Random Entanglement. *CHANCE*, *35*(4), 15–17. https://doi.org/10.1080/09332480.2022.2145126

[40] Packard, M. D., & Clark, B. B. (2020). On the Mitigability of Uncertainty and the Choice between Predictive and Nonpredictive Strategy. *Academy of Management Review*, *45*(4), 766–786. https://doi.org/10.5465/amr.2018.0198

[41] Page, D. N. (1996). Sensible Quantum Mechanics: Are Probabilities only in the Mind? *International Journal of Modern Physics D*, *05*(06), 583–596. https://doi.org/10.1142/S0218271896000370

[42] Prigogine, I., & Stengers, I. (2018). Order out of chaos: Man's new dialogue with nature. *Verso Books.*

[43] Prigogine, Ilya. (1989). The philosophy of instability. *Futures*, *21*(4), 396–400. https://doi.org/10.1016/S0016-3287(89)80009-6

[44] Prigogine, Ilya. (2014). Beyond Being and Becoming. *New Perspectives Quarterly*, *31*(1), 5–11. https://doi.org/10.1111/npqu.11418

[45] Prigogine, Ilya, Nicolis, G., & Babloyantz, A. (1972). Thermodynamics of evolution. *Physics Today*, *25*(11), 23–28. https://doi.org/10.1063/1.3071090

[46] Prigogine, L., & Stengers, I. (1984). Order out of Chaos: Man's New Dialogue with Nature. Bantam Books, Inc.

[47] Rigby, C. S., & Ryan, R. M. (2018). Self-Determination Theory in Human Resource Development: New Directions and Practical Considerations. *Advances in Developing Human Resources*, *20*(2), 133–147. https://doi.org/10.1177/1523422318756954

[48] Riniker, S., Allison, J. R., & van Gunsteren, W. F. (2012). On developing coarse-grained models for biomolecular simulation: a review. *Physical*

Chemistry Chemical Physics, *14*(36), 12423. https://doi.org/10.1039/c2cp40934h

[49] Russell, K., O'Raghallaigh, P., McAvoy, J., & Hayes, J. (2020). A cognitive model of digital transformation and IS decision making. *Journal of Decision Systems*, *29*(sup1), 45–62. https://doi.org/10.1080/12460125.2020.1848388

[50] Sætre, A. S., & Van de Ven, A. (2021). Generating Theory by Abduction. *Academy of Management Review*, *46*(4), 684–701. https://doi.org/10.5465/amr.2019.0233

[51] Scoones, I., Stirling, A., Abrol, D., Atela, J., Charli-Joseph, L., Eakin, H., Ely, A., Olsson, P., Pereira, L., Priya, R., van Zwanenberg, P., & Yang, L. (2020). Transformations to sustainability: combining structural, systemic and enabling approaches. *Current Opinion in Environmental Sustainability*, *42*, 65–75. https://doi.org/10.1016/j.cosust.2019.12.004

[52] Sørensen, L., Osnes, B., Visted, E., Svendsen, J. L., Adolfsdottir, S., Binder, P.-E., & Schanche, E. (2018). Dispositional Mindfulness and Attentional Control: The Specific Association Between the Mindfulness Facets of Non-judgment and Describing With Flexibility of Early Operating Orienting in Conflict Detection. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02359

[53] Sperry, R. W. (1993). The impact and promise of the cognitive revolution. *American Psychologist*, *48*(8), 878–885. https://doi.org/10.1037/0003-066X.48.8.878

[54] Stacey, R. D. (2001). *Complex responsive processes in organisations: Learning and knowledge creation*. Psychology Press.

[55] Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat Rigidity Effects in Organisational Behavior: A Multilevel Analysis. *Administrative Science Quarterly*, *26*(4), 501. https://doi.org/10.2307/2392337

[56] Svozil, K. (2021). Quantum Randomness is Chimeric. *Entropy*, *23*(5), 519. https://doi.org/10.3390/e23050519

[57] Tonello, L., & Grigolini, P. (2021). Approaching Bounded Rationality: From Quantum Probability to Criticality. *Entropy*, *23*(6), 745. https://doi.org/10.3390/e23060745

[58] Tyfield, D. (2013). The Demise of Capitalism? *Journal of Critical Realism*, *12*(1), 112–128. https://doi.org/10.1179/rea.12.1.f148442u57449744

[59] Walter, A.-T. (2021). Organisational agility: ill-defined and somewhat confusing? A systematic literature review and conceptualisation. *Management Review Quarterly*, *71*(2), 343–391. https://doi.org/10.1007/s11301-020-00186-6

[60] Wang, Z., Bao, Z., Wu, Y., Li, Y., Cai, W., Wang, W., Ma, Y., Cai, T., Han, X., Wang, J., Song, Y., Sun, L., Zhang, H., & Duan, L. (2022). A flying Schrödinger's cat in multipartite entangled states. *Science Advances*, *8*(10). https://doi.org/10.1126/sciadv.abn1778

[61] Wechs, J., Abbott, A. A., & Branciard, C. (2019). On the definition and characterisation of multipartite causal (non)separability. *New Journal of Physics*, *21*(1), 013027. https://doi.org/10.1088/1367-2630/aaf352

[62] Wheatley, M. (2011). *Leadership and the new science: Discovering order in a chaotic world.* ReadHowYouWant. com.

[63] Wineland, D. J. (2013). Nobel Lecture: Superposition, entanglement, and raising Schrödinger's cat. *Reviews of Modern Physics*, *85*(3), 1103–1114. https://doi.org/10.1103/RevModPhys.85.1103

[64] Zeb, A., Abdullah, N. H., Hussain, A., & Safi, A. (2019). Authentic leadership, knowledge sharing, and employees' creativity. *Management Research Review*, *43*(6), 669–690. https://doi.org/10.1108/MRR-04-2019-0164

Publication Ethics; List of Reviewers; Instructions for Authors

1. Publication Ethics

The Journal of Information and Organizational Sciences (JIOS) upholds the highest standards of publication ethics and takes all possible measures against any publication malpractices.

1.1. Duties of the Editor

The editor is responsible for maintaining the integrity of the academic record, for having processes in place to assure the quality of the published material as well as for precluding business needs from compromising intellectual and ethical standards.

The editor selects reviewers for papers, decides on the required revisions and the acceptance of the paper in accordance with the reviewers' recommendation.

1.1.1. Publication Decisions

The editor is responsible for deciding which manuscripts submitted to JIOS will be accepted for publication. This decision is based on the reviewers' recommendation. The main selection criteria are the contribution's importance, originality and clarity, as well as the study's validity and its relevance. The decision will not be influenced by the authors' race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy. The editor may confer with reviewers while making this decision.

1.1.2. Confidentiality

The editor must not disclose any information about a manuscript submitted to anyone other than the corresponding author, reviewers, potential reviewers, other advisory board members, and the publisher, as appropriate.

1.1.3. Disclosure and Conflicts of Interest

Unpublished materials disclosed in a submitted manuscript will not be used in the editor's own research without the express written consent of the author.

1.2. Duties of Reviewers

1.2.1. Contribution to Editorial Decisions

Papers will be published in JIOS after a double-blind peer-reviewed process. Reviewers advise the editor. Reviewers do not know the author's identity and their comments to the editor are confidential and will be made anonymous before they are passed on to the author. The names of the reviewers remain strictly confidential, with their identities known only to the editor.

1.2.2. Promptness

Any selected reviewer who feels unqualified to review the research reported in a manuscript or knows that its prompt review will be impossible, should notify the editor and withdraw from the review process.

1.2.3. Confidentiality

Any manuscripts received for review must be treated as confidential documents. They must not be disclosed to or discussed with others except as authorized by the editor.

1.2.4. Standards of Objectivity

Reviews should be conducted objectively. Personal criticism of the author is inappropriate. Reviewers should express their views clearly with supporting arguments, if necessary with explanation.

1.2.5. Acknowledgement of Sources

Reviewers should identify relevant published work that has not been cited by the authors. They should point out whether observations or arguments derived from other publications are accompanied by the respective source. Reviewers will notify track directors or the editor of any substantial similarity or overlap between the manuscript under consideration and any other published paper of which they have personal knowledge.

1.2.6. Disclosure and Conflict of Interest

Privileged information or ideas obtained through the review process must be kept confidential and not used for personal advantage. Reviewers should not consider manuscripts in which they have conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or institutions associated with the papers.

1.3. Duties of Authors

1.3.1. Reporting Standards

Authors of manuscripts should present an accurate account of the work performed as well as an objective discussion of its significance. Underlying data should be represented accurately in the manuscript. A manuscript should contain sufficient detail and references to permit others to replicate the work. Fraudulent or knowingly inaccurate statements constitute unethical behaviour and are unacceptable.

1.3.2. Data Access and Retention

Authors may be asked to provide the raw data in connection with a manuscript for review and should be prepared to provide public access to such data, if practicable, and should in any event be prepared to retain such data for a reasonable time after publication.

1.3.3. Originality and Plagiarism

The authors should ensure that they have written entirely original works, and if the authors have used the work and/or words of others, that this has been appropriately cited or quoted. Plagiarism takes many forms, including the touting of material contained in another paper (of the same authors or some other author) with cosmetic changes as a new paper, copying or paraphrasing substantial parts of another's paper (without attribution), and claiming results from research conducted by others. In all its forms plagiarism constitutes unethical publishing behaviour and is unacceptable.

1.3.4. Acknowledgement of Sources

Proper acknowledgment of the work of others must always be given. Authors should cite publications that have been influential in determining the nature of the reported work.

1.3.5. Multiple, Redundant or Concurrent Publication

An author should not in general publish manuscripts describing essentially the same research in more than one journal or primary publication. Submitting the same manuscript to more than one journal or conference concurrently constitutes unethical publishing behavior and is unacceptable.

1.3.6. Authorship of the Paper

Authorship should be limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study. All those who have made significant contributions should be listed as co-authors. Where there are

others who have participated in certain substantive aspects of the manuscript, they should be acknowledged or listed as contributors.

The corresponding author should ensure that all appropriate co-authors and no inappropriate co-authors are included on the manuscript, and that all co-authors have seen and approved the final version of the manuscript and have agreed to its submission for publication.

1.3.7. Fundamental Errors in Published Work

When an author discovers a significant error or inaccuracy in his/her own published work, it is the author's obligation to promptly notify the editor or publisher and cooperate with the editor to retract or correct the paper.

If the editor or the publisher learn from a third party that a published work contains a significant error, it is the obligation of the author to promptly retract or correct the paper or provide evidence to the editor of the correctness of the original paper.

1.3.8. Disclosure and Conflicts of Interest

All authors should disclose in their manuscript any financial or other substantive conflict of interest that might be construed to influence the results or interpretation of their manuscript. All sources of financial support for the paper should be disclosed.

1.3.9. Publisher's Confirmation

In cases of alleged or proven scientific misconduct, fraudulent publication or plagiarism the publisher, in close collaboration with the editors, will take all appropriate measures to clarify the situation and to amend the paper in question. This includes the prompt publication of an erratum or, in the most severe cases, the complete retraction of the affected work.

2. List of Reviewers

Tayfun Arar, Kırıkkale University, Turkey

Dilmurod Azimov, Tashkent State University of Economics, Uzbekistan

Mihalj Bakator, University of Novi Sad, Serbia

Igor Balaban, University of Zagreb, Croatia

Dušan Barać, University of Belgrade, Serbia

Slobodan Beliga, University of Rijeka, Croatia

Ivana Čavar, University of Zagreb, Croatia

Pavle Dakić, Slovak University of Technology in Bratislava, Slovakia

Kristina Detelj, University of Zagreb, Croatia

Jasminka Dobša, University of Zagreb, Croatia

Iva Gregurec, University of Zagreb, Croatia

Goran Hajdin, University of Zagreb, Croatia

Marko Hell, University of Split, Croatia

Nikola Ivković, University of Zagreb, Croatia

Maria Jakovljevic, University of South Africa, South Africa

Zuzana Jankova, Brno University of Technology, Czechia

Emily Maria K. Jose, Velore Institute of Technology, India

Mike Joy, University of Warwick, United Kingdom

Olga Kanishcheva, Friedrich Schiller University Jena, Germany

Sahab Kareem, Erbil Polytechnic University, Iraq

Gokhan Kerse, Kafkas University, Turkey

Gulsen Kirpik, Adiyaman University, Turkey

Mario Konecki, University of Zagreb, Croatia

Joana Kosho, Xhuvani University, Albania

Anatolii Kosolapov, Ukrainian State University of Science and Technology, Ukraina

Alen Lovrenčić, University of Zagreb, Croatia

Sandra Lovrenčić, University of Zagreb, Croatia

Ivan Magdalenić, University of Zagreb, Croatia

Robert Manger, University of Zagreb, Croatia

Jan Paralič, Technical University of Košice, Slovakia

Mohamed Shaheim Patel, Institute of Commerce and Management at Regent, Southern Africa

Attila Pethoe, Universty of Debrecan, Hungary

Ruben Picek, University of Zagreb, Croatia

Kornelije Rabuzin, University of Zagreb, Croatia

Deepak Kumar Raj, NIT Durgapur, Indonesia

Riad Sahara, University of Siber Asia, Indonesia

Radoslava Stankova Kraleva, Souuh-West University Neofit Rilski, Bulgaria

William Steingartner, Technical University of Košice, Slovakia

Violeta Vidaček Hainš, University of Zagreb, Croatia

Neven Vrček, University of Zagreb, Croatia

Andrei Yermolenko, Universty of Nevada, USA

Lukazsz Zakonnik, Lodz University, Poland

Lana Žaja, Croatian State Archives, Croatia

Bojan Žugec, University of Zagreb, Croatia

3. Instructions for Formatting JIOS Articles

Alen Lovrenčić *alen.lovrencic@foi.hr*
University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia

Goran Hajdin *goran.hajdin@foi.hr*
University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia

First Name Middle Name Last Name *email.address@domain*
Full Institution/Organization/Company Name
City/Town, Country

Abstract

This document describes the required formatting of JIOS papers, including margins, fonts, citation styles, and figure placement. While the format requirements are only compulsory for final submissions, we strongly encourage authors to adopt this template, as well as its recommendations throughout the submission process. Abstract must not exceed 150 words. The abstract appears at the beginning of the paper, indented 0,5 cm from the left and right margins. The title "Abstract" should appear in bold face 11 pt, centered above the body of the abstract. The abstract body should be in 10 pt. Spacing between Abstract and author(s) affiliations is 36 pt, and spacing between "Abstract" title and abstract text is 6pt.

Keywords: formatting instructions, Microsoft word template, JIOS

3. Introduction

To ensure that all articles published in the journal have a uniform appearance, authors must produce a PDF document that meets the formatting specifications outlined in this document. The same document will be used for digital and hard copy version of the journal. Paper submission must be made in one ZIP archive which must include source document (ie. MS Word file) and PDF. When crating PDF make sure all fonts are embedded.

This document briefly describes and illustrates format used by JIOS journal. Your article should look as similar as possible to this document. This template can be obtained form journal web site at the address http://www.jios.foi.hr. Below the basic specifications, including font sizes, margins, etc. will be outlined. We encourage you to use this sample in case of any dilemma. For any questions you cannot decide, feel free to contact JIOS.

JIOS is published twice a year. 1st volume is published 30th of June, and 2nd volume 31st of December. All papers that enter copyediting process at least one month before publishing date will be included in current volume. All papers that enter

copyediting process less than one month from the publishing date will be included in the next volume.

4. Format and Margins

Papers must be in the single column format as shown in the enclosed sample. The page format should be set to B5. Left, right and top margin should be set to 2,2 cm while bottom margin should be set to 2,5 cm. Headers should be 1,27 cm from top and footer should be 2,18 cm from bottom of page. Title should be 36 pt below the header.

4.1. Fonts

You should use Times Roman style fonts. Please, do not us any non-standard fonts in your paper. They could make problems in formatting of the paper, as well as in printing.

4.2. Authors

Authors' names should appear in designated areas below the title of the paper in 12 pt bold type. Authors' affiliations and complete addresses should be in italics 11 pt, and their email addresses in 10 pt italic. If author is part of institution, organization or company, information should include full author name, full institution/organization/ company name, town/city and country. If author is publishing as a private person, information should include full author name, full address (street address and number), town/city and country. Authors' affiliations and email address must be aligned according to the left and right paper margin.

4.3. Title, Headings and Sections

The title of the paper should be in 14 pt bold. When necessary, headings should be used to separate major sections of your paper. First headings should be in 12 pt bold and second headings should be in 11 pt bold. The text and body of the paper should be in 11 pt. Third-level headings should also be in 11 pt italic. Each heading should have 18 pt spacing above and 6 pt spacing below it. Do not skip a line between paragraphs. After a heading, the first sentence should not be indented.

Paper title and all headings should be capitalized. Please refer to Rules for Capitalization in Titles regarding capitalization of paper tile and section headings.

References to sections (as well as figures, tables, theorems and so on), should be capitalized, as in "In Section 4, we show that...".

4.4. Figures and Tables

Figures and tables should be inserted in proper places throughout the text. Do not group them together at the beginning of a page, nor at the bottom of the paper. Number figures sequentially, e.g., Figure 1, and so on.

The figure or table number and the caption should appear under the illustration. Spacing around caption should be 6 pt above and 11 pt below. Captions, labels, and other text in illustrations must be at least 9 pt.

	CPU Speed	Wage
A		
B		
C		

Table 1. Note well that JIOS expects table captions below the table. >= 9pt font.

We strongly suggest not to use color figures and pictures. Authors are responsible for providing grayscale figures and pictures without any loss of information.

When placing figures and tables in the paper please keep in mind possible blank spaces which might appear at the bottom of some pages. Please relocate paper elements (tables, figures and text) to fill existing blanks. While doing so pay attention not to create new ones.

4.5. Headers and Footers

Headers and Footers should be in 9 pt. All information in headers should be in small caps, respecting capital first letters. The first page of your article should include the short journal name, volume, number and year in the upper left corner, and the submission date and acceptance date in the upper right corner. The editor will let you know the volume, number, year, submission date and acceptance date.

On the even numbered pages, the header of the page should be the authors' names in the upper left corner and short title of the paper in the upper right corner. If paper title is too long to fit in the header (maximum 40 characters), please shorten the title and add three dots (…) at the end. Please do not abbreviate any title words. On the odd pages, starting with page 3, the header should be the full name of the journal, aligned right.

4.5.1. Page Numbering

Please do not add any page numbers in the article. They will be assigned by the publishing board at the bottom of the page in the center, as well as added in the footer of the article.

4.5.2. Footnotes

We encourage authors to use footnotes sparingly, especially since they may be difficult to read online. Footnotes should be numbered sequentially and should appear at the bottom of the page, as shown below.[1]

4.6. References

The reference section should be labeled "References" and should appear at the end of the paper. References must be ordered according to their appearance in the paper. Please prepare complete and accurate citations. References should be formatted according to the IEEE Citation reference. Do not include references that are not cited in the text of the paper.

Citations within the text should include the number of the reference in the reference section in the brackets, for example [1]. Multiple citations should be separated by a colon, as in [1], [2].

Acknowledgements

The acknowledgments section, if included, appears after the main body of the text and is headed "Acknowledgments." The section should not be numbered. This section includes acknowledgments of help from associates and colleagues, financial support, and permission to publish..

Appendix A: Title of the Appendix

Appendices, if included, follow the acknowledgments. Each appendix should be lettered, e.g., "Appendix A".

References

[1] IEEEDataPort, "How to Cite References: IEEE Documentation Style" [Online]. Available: http://jios.foi.hr/guidelines/IEEE_Documentation_Style.pdf [Accessed: Jun. 5, 2018].

[2] References should be indented on the left side by 0,63 cm and hanging must be set to 0,83 cm.

[3] Hyperlinks should be formatted as plain text and without underline formatting.

[4] All references should be aligned left.

[1] A footnote should appear like this. Please ensure that your footnotes are complete, fully punctuated sentences.

JIOS
Journal of Information and Organizational Sciences

Contact
Alen Lovrenčić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, HR-42000 Varaždin, Croatia
Tel.: +385 42 390 866
Fax: +385 42 213 413
E-mail: alen.lovrencic@foi.unizg.hr
E-mail: jios@foi.hr (Corespondence Mail)
jios-pub@foi.hr (Publishing Board)
URL: www.jios.foi.hr

JIOS is indexed in the following repositories and portals
Academic Journals Database
Academic Resource Index
Advanced Science Index
BASE
Central and Eastern European Online Library (CEEOL)
CORE
Crossref
CWTS Journal Indicators
Dimensions
Directory of Open Access Scholarly Resources(ROAD)
Directory of Open Access Journals (DOAJ)
Directory of Research Journals Indexing (DRJI)
Elektronische Zeitschriftenbibliothek
ERIH PLUS
EuroPub – Directory of Academic and Scientific Journals
Genamics JournalSeek
Google Scholar
HRČAK - Central portal of Croatian scientific journals
Index Copernicus
Internet Archive
Information Matrix for the Analysis of Journals (MIAR)
International Journal Impact Factor (IJIFACTOR)
J-Gate
JournalTOCs
LENS.ORG
Microsoft Academic
OAIster
Open AIRE
Open Academic Journals Index (OAJI.net)
PRADEC
Scientific Indexing Services (SIS)
Scilit
Sherpa Romeo
Ulrich's Periodicals Directory
VINITI RAN
WorldCat
World Journal Clout Index Report (WJCI)

www.jios.foi.hr

Proofreading
Authors

Copyeditor and Computer Type Set
Goran Hajdin
Mladen Konecki

Cover and Layout Design
ri mo dizajn

Printing
Printera Grupa d.o.o., Sveta Nedelja, Croatia

JIOS is indexed in the following databases
SCOPUS
Emerging Sources Citation Index (ESCI) - WoS
Academic Search Complete (EBSCO)
Academic Search Elite (EBSCO)
Academic Search Premier (EBSCO)
Academic Search Ultimate (EBSCO)
Aerospace Database - ProQuest
Civil Engineering Abstracts - ProQuest
Central & Eastern European Academic Source (CEEAS) - EBSCO
Communication Abstracts - ProQuest
EBSCO Essentials
INSPEC
Library and Information Science Abstracts (LISA) - ProQuest
Library & Information Science Source - EBSCO
Library, Information Science & Technology Abstracts (LISTA) - EBSCO
Library Literature & Information Science - EBSCO
Metadex - ProQuest

Frequency
JIOS is published twice per year.
Printed in 100 copies.

Access to JIOS is free of charge through following sources:
Journal of Information and Organizational Sciences
Hrčak - central portal of Croatian scientific journals
Directory of Open Access Journals (DOAJ)
OAIster

JIOS
Journal of Information and Organizational Sciences

CONTENTS

VOL. 48, NO. 1, PP. 1-252, JUNE 2024

ISSN 1846-3312
e-ISSN 1846-9418