Journal of COMPUTER AND FORENSIC SCIENCES

CFS

ISSN 2956-087X
e-ISSN 2956-0799

Volume 2 • Issue 1 • 2023

UNIVERSITY OF CRIMINAL INVESTIGATION AND POLICE STUDIES
Belgrade

UNIVERSITY OF CRIMINAL INVESTIGATION AND POLICE STUDIES, BELGRADE – THE REPUBLIC OF SERBIA
KRIMINALISTIČKO-POLICIJSKI UNIVERZITET, BEOGRAD – REPUBLIKA SRBIJA

CFS

CFS. JOURNAL OF COMPUTER AND FORENSIC SCIENCES

Belgrade, 2023

PUBLISHER

University of Criminal Investigation and Police Studies, Cara Dušana 196, 11080 Belgrade, Serbia

English Language Editor and Proofreader

Jelena Pandža, University of Criminal Investigation and Police Studies, Belgrade, Serbia

Journal Manager

Jelena Pandža, University of Criminal Investigation and Police Studies, Belgrade, Serbia

Typesetting

Jovan Pavlović, University of Criminal Investigation and Police Studies, Belgrade, Serbia

Printed by

Službeni glasnik, Belgrade

Impression

100 copies

TABLE OF CONTENTS

EDITORIAL

Journal of
COMPUTER AND
FORENSIC SCIENCES

CFS 2023, Vol. 2, Issue 1, pp. 3–4
Editorial

EDITORIAL

I Hope to Read Your Own Contribution Soon

Milan Gnjatović
Editor-in-Chief
University of Criminal Investigation and Police Studies, Belgrade
milan.gnjatovic@kpu.edu.rs

Published: July 13, 2023

Starting from a somewhat bold assumption that everyone reads editorials once in a while, I am pleased to present the second issue of the Journal of Computer and Forensic Sciences and to thank the authors, reviewers and editorial team members for their essential effort in preparing this contribution.

The issue brings five original research articles in the field of computer science:

The first article [1] proposes a declarative computational model of a sequential trading system and demonstrates how some of its fundamental properties can be formally proved.

The second article [2] introduces and experimentally evaluates an approach to speech enhancement based on a cycle generative adversarial network, a feature map regularization technique, and augmentation techniques.

The third article [3] presents a comprehensive architectural design overview of a blockchain-based online system for news verification and reputation management.

The fourth article [4] discusses a system architecture for preventing social engineering email attacks.

The last article [5] reports on the generation and evaluation of a set of machine learning models for anomaly detection in online advertising.

Thank you very much for reading the Journal of Computer and Forensic Sciences. I hope to read your own contribution soon!

REFERENCES

[1] D. Žunić, “A Verifiable Model of a Minimal Market Operating Sequentially, with Price and Time Discrete,” *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 7–17, 2023.

[2] B. Popović and M. Janev, “Speech Enhancement by CycleGAN Using Feature Map Regularization,” *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 19–28, 2023.

[3] A. Miljković, M. Čabarkapa, F. Miljković and Lj. Bojić, “iTrust News Certificate: A Blockchain-Based Solution for News Verification and Reputation Management,” *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 29–41, 2023.

[4] M. Brkić, A. Maksimović, and A. Miljković, “A System Architecture for Preventing Social Engineering Attacks via E-mail,” *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 43–52, 2023.

[5] M. Živanović, S. Štrbac-Savić, and Z. Minchev, “Application of Machine Learning Methods for Anomaly Detection in Internet Advertising,” *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 53–61, 2023.

ORIGINAL RESEARCH PAPERS

Journal of COMPUTER AND FORENSIC SCIENCES

CFS 2023, Vol. 2, Issue 1, pp. 7–17
https://doi.org/10.5937/jcfs2-44362
Original research paper

A Verifiable Model of a Minimal Market Operating Sequentially, with Price and Time Discrete

Dragiša D. Žunić[*1]

[1] The Institute for Artificial Intelligence R&D of Serbia

Corresponding author: dragisa.zunic@ivi.ac.rs

Received: April 9, 2023 • Accepted: April 28, 2023 • Published: June 19, 2023

Abstract: This research presents a minimal computational market model, i.e., a model of a trading venue, with sequential order matching, in a declarative style, and proceeds to demonstrate how some fundamental properties can be formally proved. It is a challenging task to formally certify the properties of a fundamental system in any realm of human endeavor, especially of systems with infinite state space. With the recent development of theoretical frameworks based on formal logic, it is now possible (albeit very difficult) to both formalize and reason about an object system in the same language. This research derives from the previous research presented in [1], and represents a simplification to obtain a minimal model. The computational model of a minimal market, presented here in a declarative style, is important from the perspective of both market design and verification.

Keywords: formal logic; market design; financial exchanges; automated reasoning; logical frameworks; computational models.

1. INTRODUCTION

A financial exchange is a platform where buyers and sellers can come together to trade various financial instruments. These instruments include stocks, bonds, commodities, currencies, and derivatives. Financial exchanges provide a central marketplace where market participants can buy and sell financial instruments with each other based on their respective prices. The computational core of a financial exchange is the order matching engine, which handles interaction between buy and sell flows of orders. In order to guarantee trading fairness, exchanges must meet the requirements of regulatory bodies, in addition to numerous general requirements. However, both specifications and requirements are presented in natural language, which hardly qualifies as an adequate method for such a task.

Based on the previous experience, violations frequently originate either from interactions between order types, or from the way matching logic is specified and implemented [2, 3]. Formalization and formal reasoning can play a big role in mitigating these problems. They provide methods to verify properties of complex and infinite state space systems with certainty and have already been applied in fields ranging from hardware design to flight safety and financial contracts [4, 5], with trading systems being considered recently as well [6, 7].

Citation: D. Žunić, "A Verifiable Model of a Minimal Market Operating Sequentially, with Price and Time Discrete," *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 7–17, 2023, https://doi.org/10.5937/jcfs2-44362.

A general sequential order matching core has been formalized in [1], followed by proving properties such as: the trade always takes place at either bid or ask; the market is never in a locked or crossed state; and order priority is never violated. Everything is based on being able to declaratively represent an archetypal sequential (matching in one-by-one fashion) trading system, using only symbols and symbol manipulation, which provides a setting for verification, i.e., some form of semi-automated reasoning about the system's properties.

2. LINEAR LOGIC AND LOGICAL FRAMEWORKS

Linear logic (LL) is a logical system that was introduced by Girard in the 1980s [8]. One of the key features of linear logic is that it has a notion of resources, which allows it to capture more accurately certain aspects of computation and reasoning. In other words, linear logic is a resource conscious logic, and formulas are consumed when used to prove a statement. This is achieved by having structural rules of contraction and weakening explicit in the system, and being able to selectively mark formulas intended to be an unbounded resource by the exponential operator !. Intuitionistic linear logic (ILL) is the restriction of linear logic to the intuitionistic fragment. Formulas in (propositional) ILL are composed of the following connectives: $\otimes$ and 1 (multiplicative conjunction and its neutral element), & and $\top$ (additive conjunction and its neutral element), $\oplus$ and 0 (disjunction and its neutral element), $\multimap$ (linear implication), $\rightarrow$ (intuitionistic implication), ! (exponential).

The logical framework CLF (Concurrent Linear Framework) [9] is based on a fragment of intuitionistic linear logic. It extends the traditional LF [10] framework with the linear connectives $\multimap$, &, $\top$, $\oplus$, 1 and ! to obtain a resource-aware framework with a satisfactory representation of concurrency. The rules of the system impose a discipline on when the (less deterministic) connectives $\otimes$, 1 and ! are decomposed, thus still retaining enough determinism to allow for the implementation of a logical framework. For simplicity, we present only the logical fragment of CLF needed for our encoding. This is only a small fragment of the logical framework, but the need for $\otimes$ in our encodings indicates that anything less than CLF (e.g., LLF) would be less suitable.

The majority of the trading system encoding involves clauses in the following shape (for atomic p_i and q_i): $p_1 \otimes \dots \otimes p_n \multimap \{q_1 \otimes \dots \otimes q_m\}$. We used an implementation of this framework called Celf[1]. Following the tool's convention, variable names start with an upper-case letter.

3. ELECTRONIC TRADING SYSTEMS

As mentioned in the introduction, real life trading systems, both public exchanges and alternative trading systems, differ slightly in the way they manage orders. However, there is a certain common core that guides all those trading systems and embodies the market logic of trading on an exchange. A detailed account for formalization of a general sequential trading core, in a declarative style, is presented in [1].

[1] https://clf.github.io/celf/

The default mode of operation of electronic trading systems globally is the *price/time priority*, although other modes of operation exist. Before we explain the mode of operation and different types of orders, let us introduce basic notions.

An *order* is an investor's instruction to a broker to buy or sell securities (or any asset type traded on a financial exchange), thus we have *buy orders* and *sell orders*. There are two basic types of orders: limit, which is short for limit-price, and market.

A *limit order* has a specific limit price at which it is willing to trade, meaning that it will trade at that price or better. In the case of a limit order to sell, a price limit price p means that the security will be sold at the best available price in the market, but no less than p. And symmetrically for buy limit orders.

A *market order* does not specify the price at which it is willing to trade, and will be immediately (if possible) matched against the best available price for this security.

Besides price, orders have a *quantity*, the amount of securities they are willing to trade. An order is identified by a timestamp, which records the time it enters the trading system. Orders, regardless of the type, are filled eagerly.

Outstanding orders, i.e., those that are waiting to be filled in the trading system, are called *resident orders*. There are two sorted lists that keep track of *active prices*, which are those for which there exists at least one resident limit order, namely *active buy prices* and *active sell prices*. Of particular importance is the maximum value of the active buy prices, called *bid*, and minimal value of active sell prices, called *ask*. The difference between those two prices is the *bid-ask spread*, or simply *spread*.

For each active price, there is a queue of resident orders, sorted by time of arrival (which identifies them uniquely): the order that arrived first is at the front of the queue whilst the last one is last in the queue.

There are numerous models of trading venues, however, there are some standard order types, such as limit, market, and immediate or cancel (IOC) orders, and basic matching rules. The current state of the art in trading venue design (somewhat surprisingly) assumes that orders for a given security enter the trading venue sequentially, one at a time, and they are executed sequentially. An order is filled, or exchanged, when it is successfully matched, regarding the price, against an opposite order (or orders), provided that the quantity of opposite orders was sufficient.

The standard mode of operation is *price/time priority*, which determines how orders are prioritized for execution. Orders are first ranked according to their price; orders of the same price are then ranked depending on when they entered the venue. Other than price/time priority, the most common is the pro-rata matching algorithm, which takes into account the overall volume of the incoming order as well as resident orders at a considered price, thus making the timestamp less important.

Some of the standard regulatory requirements for real world financial trading systems are: order priority is always respected, the system will not illegally prohibit any two orders from trading with each other, no crossed and locked markets (maximum buy price must remain strictly less than the minimum sell price), and transitivity of order ranking (order priority is transitive). It is a problem for financial companies that run trading platforms to guarantee these properties.

4. FORMALIZING A MINIMAL MARKET CORE

We present a computational model for a minimal market operating sequentially (orders are entering the market and are processed in one-by-one fashion). Unlike the general model presented in [1], we consider the following:

- All orders are limit-price[2],
- All orders have unit size quantities.

In this paper, we focus on the computational steps when buy orders are entering. The rules for sell orders are symmetric.

4.1. Properties of the System by Design

The paradigm of logical frameworks enables us to think about the object system design, and with that in mind, we can formally prove the desirable properties. However, the object system formalization must be done a priori with those properties in mind.

Core properties. The computational market model presented here is designed to have the following required properties:

- The market is never in a crossed or locked state (at any given moment, bid is strictly less than ask);
- If the trade occurred, it happened at either bid or ask price (no trade occurs at a price other than the current bid or ask);
- Order priority is always respected (no computation step ever violates the order priority; given arbitrary two orders o_1, o_2, the following holds: if o_1 has a higher priority than o_2, then o_1 is filled before o_2);
- Order priority is transitive (given any three limit orders o_1, o_2 and o_3, the following holds: if o_1 has higher priority than o_2 and o_2 has a higher priority than o_3, then o_2 has higher priority than o_3).

Notice that the properties do not speak about players' strategies. These properties are core in the sense that they refer to the rules of the game, that is, the game itself. Once we extend the model to include the players participating, then these properties extend to that model as well. Namely, they can be stated in the form "Regardless of the players' strategies, the following holds...".

Additional properties. If we transition to a model that matches orders in batches, using a single market clearing price for each batch, and even allow more parallelism in the matching procedure[3], we find ourselves in a world that, besides the core properties, has numerous additional advantages from the perspective of economics and actual financial

[2] Limit orders constitute the market in the price/time priority mode. Unlike, for example, market orders, which cannot become resident orders as they are either filled or canceled. Once we have the fundamentals, it is not difficult to extend the model with different order types, such as market orders, immediate-or-cancel, fill-or-kill, etc.

[3] At the level of fundamental design, the key is to have some parallelism, together with discrete time and price.

markets. Namely, the following additional properties – as presented in the research introducing the frequent batch auctions model by Budish et al. in [2, 3]) – are satisfied:

- Competition on speed transformed into competition on price (pure time priority at entry is now closer to price/time priority);
- Race to the bottom in speed-advantage eliminated;
- Sniping stopped (a tax on a liquidity provision, hurting investors);
- Enhanced liquidity;
- Narrower bid-ask spread;
- A computationally simpler market (structured computational trace);
- Reestablished some form of the efficient market hypothesis.

4.2. Formalization: a minimal market core with unit-size orders

The big picture is provided by a typical depth chart representing the current state of the resident market, where we may notice L, M, and R segments. The two piles of resident orders are those orders that, at the time of entry, were not marketable and thus were stored in the market and not executed (filled). The pile on the left are instructions to buy, whereas the pile on the right are instructions to sell. The most competitive buy price, at this particular moment in time, is denoted as B (bid price), whereas the most competitive sell price is S (ask price). B and S divide the market into three segments, namely L, M and R, in that order. Segment M is the segment corresponding to the bid-ask spread, which in some cases may be non-existent, namely when bid and ask prices touch. See Image 1.

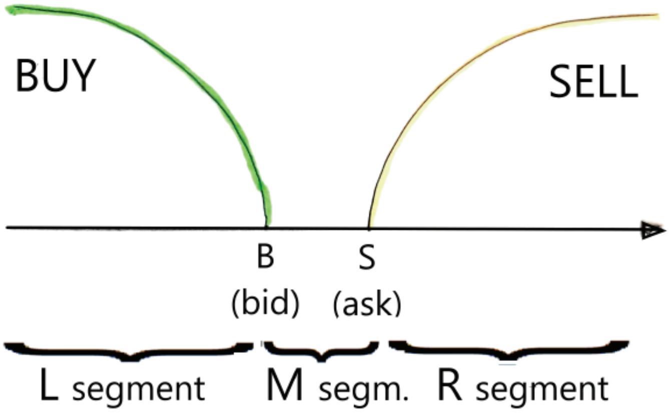


Image 1. *Market view (depth chart). Resident buy and sell orders are displayed. Market's bid and ask and consequently L, M, and R market segments are presented.*

The formalization initiates the model via the begin fact, creating facts actPrices(buy, nil) and actPrices(sell, nil) which, in what follows, will keep track of active prices (active in a sense that there is at least one resident order stored on that price). A fact time(z) is also initiated to keep track of the system time, with each computation step increasing the time counter.

$$\text{begin} \multimap \{\text{actPrices}(\text{buy}, \text{nil}) \otimes \text{actPrices}(\text{sell}, \text{nil}) \otimes \text{time}(\text{z})\}$$

The trading system is represented by the following linear predicates: priceQ(A, P, Q), actPrices(A, L), time(T), where action A can be buy or sell. In this paper, we focus more on buy limit orders.

For an action A and a price P, the queue Q in priceQ(A, P, Q) contains all resident orders with those attributes. Due to how orders are processed, the queue is sorted in ascending order of timestamp. Price queues are never empty, we only maintain price queues for active prices. For an action A, the list L in actPrices(A, L) contains the active prices available in the market, i.e., all the prices at which there is something offered. Note that the bid price is the maximum of L when A is buy and the ask price is the minimum when A is sell. The time is represented by the fact time(T) and increases as the state changes.

Storing orders. The computation when storing takes place (adding to the resident market) is presented in Fig. 1. An order is stored when its limit price P is such that it cannot be exchanged against an opposite resident orders. Namely, when P < ask in the case of a buy order, and when P > bid in the case of a sell order. The rules distinguish whether there are pre-existing resident orders at that price in the market or not.

(L) limit/empty:

$$\text{order(limit, buy, P, ID, N = 1, T)} \otimes \text{actPrices(buy, L)} \otimes$$

$$\text{maxP(L, B)} \otimes \text{less-or-eq(P, B)} \otimes \text{notInList(L, P)} \otimes \text{insert(L, P, LP)} \otimes \text{time(T)}$$

$$\multimap \quad \{\text{priceQ(buy, P, expListP([ID, N = 1, T], nilP))} \otimes \text{actPrices(A, LP)} \otimes \text{time(s(T))}\}$$

(L) limit/queue:

$$\text{order(limit, buy, P, ID, N = 1, T)} \otimes \text{actPrices(buy, L)} \otimes$$

$$\text{maxP(L, B)} \otimes \text{less-or-eq(P, B)} \otimes \text{inList(L, P)} \otimes \text{priceQ(buy, P, PQ)} \otimes$$

$$\text{expListP(PQ, [ID, N = 1, T], PQ}') \otimes \text{time(T)}$$

$$\multimap \quad \{\text{actPrices(buy, L)} \otimes \text{priceQ(buy, P, PQ}') \otimes \text{time(s(T))}\}$$

(M) limit/empty:

$$\text{order(limit, buy, P, ID, N = 1, T)} \otimes \text{actPrices(buy, L)} \otimes \text{actPrices(sell, L}') \otimes$$

$$\text{maxP(L, B)} \otimes \text{minP(L}', \text{S)} \otimes \text{nat-great(P, L)} \otimes \text{nat-less(P, S)}\otimes$$

$$\text{insert(L, P, LP)} \otimes \text{time(T)}$$

$$\multimap \quad \{\text{priceQ(buy, P, expListP([ID, N = 1, T], nilP))} \otimes \text{actPrices(buy, LP)} \otimes \text{time(s(T))}\}$$

Figure 1. *Storing of the incoming unit-size buy limit-orders in the market (in segments L and M).*

The first two rules describe how the incoming buy limit-order (when it cannot be filled) is stored, depending on whether the pre-existing queue exists (in which case that price is already active) or not. In the latter case, a new queue is created (and the corresponding price is activated). The third rule describes storing of orders in the middle segment (between bid and ask), which is by definition empty and therefore there is never a pre-existing queue. So the incoming order will be stored, and the corresponding price added to the list of active prices. By construction, this price always updates the bid or ask, depending onwhether the incoming order was buy or sell order.

Filling orders. The computation when filling takes place (executing an incoming order against the most competitive opposite resident order) is presented in Fig. 2. A limit order is filled (against the most competitive opposite order) when its limit price P satisfies P ≤ bid, in the case of sell orders, or P ≥ ask for buy orders.

(R) limit/1:

$$\text{order}(\text{limit}, \text{buy}, P, ID, N = 1, T\,) \otimes \text{actPrices}(\text{sell}, L') \otimes$$

$$\text{minP}(L', S) \otimes \text{great-or-eq}(P, S) \otimes \text{priceQ}(\text{sell}, S, \text{consP}([ID', N' = 1, T'], \text{nilP}))\otimes$$

$$\text{remove}(L', S, L'') \otimes \text{nat-equal}(N = 1, N' = 1) \otimes \text{time}(T\,)$$

$$\multimap \quad \{\text{actPrices}(\text{sell}, L\,) \otimes \text{time}(s(T\,))\}$$

(R) limit/2:

$$\text{order}(\text{limit}, \text{buy}, P, ID, N = 1, T\,) \otimes \text{actPrices}(\text{sell}, L') \otimes$$

$$\text{minP}(L', S) \otimes \text{great-or-eq}(P, S)\otimes$$

$$\text{priceQ}(\text{sell}, S, \text{consP}([ID', N' = 1, T'], \text{consP}([ID1, N\,1 = 1, T\,1], L))) \otimes$$

$$\text{nat-equal}(N = 1, N' = 1) \otimes \text{time}(T\,)$$

$$\multimap \quad \{\text{actPrices}(\text{sell}, L') \otimes \text{priceQ}(\text{sell}, S, \text{consP}([ID1, N\,1 = 1, T\,1], L)) \otimes \text{time}(s(T\,))\}$$

Figure 2. *Filling of incoming unit-size buy limit-orders against an opposite resident order (computation in segment R).*

The first rule covers the case when the most competitive opposite order in the market is the last in the price-queue (most competitive by definition always sits at ask, i.e., S). Thus, we need to remove the current ask price from the active price list (this effectively defines the new ask price). The second rule covers the case when the most competitive opposite is not the last in that price-queue, and there is no need to update the current ask price or the active price list.

5. TOWARDS THE VERIFICATION OF THE TRADING SYSTEM PROPERTIES

As we have seen in the previous section, the market (exchange, or trading venue) is a dynamic system consisting of two incoming flows of buy and sell orders, which interact with the opposite pool of resident orders – changing the current state of the system. This interaction gives birth to the computation of the financial exchange.

Using a declarative style formalization, we are able to check that this combination of matching rules does not violate desired trading system properties. In particular, we show that the system is never in a crossed or locked market state. A trading system enters a crossed or locked state if a bid price (the most competitive buy price) becomes greater or equal to ask price (the current most competitive sell price), respectively. Intuitively, bid and ask prices are oscillating, i.e., increase or decrease a bit. If a system is designed correctly, it will be impossible to have a state where bid = ask, or bid > ask. Clearly, this is because if there was a chance to fill an incoming order, it would have been done a priori, at the time of entry (an incoming order that can be matched is executed immediately upon arrival, it is not stored).

The minimal market is specified via five state transition rules for incoming buy orders, and symmetrically five rules for incoming sell orders. To prove this property, we need to inspect the rules that change the list of active prices in the way to expand them[4].

Other than that, a fundamental property that could be proved is that the trade, at any given moment, takes place exclusively at either bid or ask. The current version of Celf does not yet support automated meta-reasoning, so the proof is developed by hand. We provide a rough sketch of the formal proof with the intent to demonstrate how this relies on simple induction, when in real life – based on standard testing – it is next to impossible to certify an object system for these properties.

Definition 5.1 (No crossed or locked market) We say that the system satisfies “No crossed or locked property” if, at all times, the maximum buy price in the market is strictly less than the minimum sell price.

5.1. Proving that the Market is Never in a Crossed or Locked State

To prove that *no crossed or locked market* property is maintained, we need to show that the maximum of the list of active prices (for resident buy orders) is less than the minimum of active prices for resident sell orders. This is shown by induction on the reachable states; for each relevant state change, we check if the property is maintained. The bid and ask prices are potentially updated only if a new active price is added to the list of buy or sell active prices. Thus, we need to show that whenever this addition takes place, the resulting lists do not violate the property.

Note that if an order is added to the market, there are no matching orders that it could have been exchanged with. Note also that we only need to worry about those rules that rewrite L into some L′. By analyzing those, we observe that the new L′ is computed by the predicate insert(L, X, L′).

[4] If a rule contracts the list of active prices, then bid and ask are moving away from each other, which does not lead to a potential locked or crossed scenario.

Theorem 5.1 The *no crossed or locked market* property holds in all states.

For every state that is characterized by actPrices(buy, BP), actPrices(sell, SP), maxP(BP, X) and minP(SP, Y), it is the case that $X < Y$.

Proof. The proof goes by induction and case analysis of the state transitions.

Base case: The system comes to life via the beginning fact which generates the initial state. The facts actPrices(buy, L) and actPrices(sell, L) are generated initially using nilN for L. Since maxP(nilN, infinity) and minP(nilN, z), and $z <$ infinity, we have that the property holds for the initial state.

IH: We assume that for a given list of active prices BP, SP, it holds that if actPrices(buy, BP), actPrices(sell, SP), maxP(BP, X) and minP(SP, Y), then $X < Y$.

Using this hypothesis, we proceed to prove the property.

The predicate actPrices(buy, BP) is rewritten to actPrices(buy, BP′), where insert(BP, P, BP′). This happens as a rule when a new order at price P is added to the market. By construction, this rule is only triggered if P is smaller than X (otherwise, that limit order would have already been exchanged). Since the list BP′ is BP extended with a limit price P, we have two cases:

Case 1. If $P \leq X$, X remains the maximum value of BP′, so in the new state we will have actPrices(buy, BP′) and maxP(BP′, X), and therefore $X < Y$ still holds. See rule (L) limit/empty in Figure 1.

Case 2. If $X < P < Y$, P is in the middle region, between bid and ask, and therefore is the maximum value of newly formed BP′, so in the new state we will have actPrices(buy, BP′) and maxP(BP′, P). But by the reasoning above, $P < Y$, the new bid is smaller than the ask, and the property still holds. See rule (M) limit/empty in Figure 1.

With this, we are done with the proof.

6. CONCLUSION

We have presented a declarative representation of a sequentially operating archetypal order/matching system. Having in mind that it can be extremely difficult to verify properties of fundamental systems using standard techniques of testing, especially if the system is of the infinite state-space nature, we show that this is straightforwardly done using methods based on formal logic and inductive reasoning. The challenge, however, is in being able to create a model capturing the nature of an object system at the fundamental level, and at the right level of abstraction for a given challenge.

Trading systems are considered as safety-critical systems which must comply with various criteria, including the regulatory requirements. Thus, having method and tools to execute on this is of great importance. We showed here how to formally certify that the system satisfy one of the most important fundamental properties, namely that the market is never in a locked or crossed state, i.e., that during the computation bid remains strictly smaller than ask. This property is obtained by design, therefore, it is important that the systems are carefully implemented.

An interesting challenge for future work is combining this paradigm to detect and prevent prohibited (and disruptive) trading practices and, related to this, critical states such as flash-crashes or, symmetrically, market-bubbles. This may be the path towards the elements of financial forensics, but we need to clearly understand the method to coherently combine machine learning, as an empirical method, with symbolic AI as a theoretical and qualitative method.

Coming from the computational perspective, one of the key future directions is the design of a market model featuring parallelism and concurrency, together with discrete time and price at the level of fundamental design, which represents a step forward in the quest for the right (computationally speaking) market model.

FUNDING:

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT:

Not applicable.

INFORMED CONSENT STATEMENT:

Not applicable.

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

REFERENCES

[1] I. Cervesato, S. Khan, G. Reis, and D. Žunić, "Formalization of automated trading systems in a concurrent linear framework," In Proceeding of Linearity and TLLA, Oxford UK, 2019, pp. 1–15.

[2] E. Budish, P. Crampton, and J. Shim, "Implementation details for frequent batch auctions: slowing down markets to the blink of an eye," *American Economic Review Papers and Proceedings*, Vol. 104, No. 5, pp. 418–424, 2014.

[3] E. Budish, P. Crampton, and J. Shim, "The high-frequency trading arms race: frequent batch auctions as a market design response," *Quarterly Journal of Economics*, Vol. 130, No. 4, pp. 1547–1621, 2015.

[4] P. Bahr, J. Berthold, and M. Elsman, "Certified symbolic management of financial multi-party contracts," In ICFP 2015, Vancouver, Canada, 2015, pp. 315–327.

[5] S. P. Jones, J. M. Eber, and J. Seward, "Composing Contracts: An Adventure in Financial Engineering (Functional Pearl)," In ICFP 2000, 2000, pp. 280–292.

[6] G. O. Passmore and D. Ignatovich, “Formal Verification of Financial Algorithms,” In CADE 26, Gothenburg, Sweden, 2017, pp. 26–41.

[7] D. Ignatovich and G. O. Passmore, Case Study: 2015 SEC Fine Against UBS ATS, Aesthetic Integration, Ltd., Technical Whitepaper, 2015.

[8] J. Y. Girard, “Linear Logic,” *Theoretical Computer Science*, Vol. 50, pp. 1–102, 1987.

[9] I. Cervesato, K. Watkins, F. Pfenning, and D. Walker, “A Concurrent Logical Framework I: Judgments and Properties,” Technical Report CMU-CS-02-101, CMU Pittsburgh, 2003.

[10] R. Harper, F. Honsell, and G. Plotkin, “A Framework for Defining Logics,” *J. ACM*, Vol. 50, pp. 143–184, 1993.

Journal of
COMPUTER AND
FORENSIC SCIENCES

CFS 2023, Vol. 2, Issue 1, pp. 19–28
https://doi.org/10.5937/jcfs2-45100
Original research paper

Speech Enhancement by CycleGAN Using Feature Map Regularization

Branislav Popović[1*] and Marko Janev[2]

[1] University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia; bpopovic@uns.ac.rs

[2] Serbian Academy of Sciences and Arts, Institute of Mathematics, Kneza Mihaila 36, 11000 Belgrade, Serbia; markojan@uns.ac.rs

* Corresponding author: bpopovic@uns.ac.rs

Received: June 19, 2023 • Accepted: July 3, 2023 • Published: July 10, 2023

Abstract: The state-of-the-art results of single-channel speech enhancement were recently obtained by applying the unpaired dataset CycleGAN network approach, which is comparable to the paired dataset neural network approach. As only a relatively small amount of noisy speech data is usually available in applications, an augmented, semi-supervised CycleGAN is proposed. Recently, the feature map regularized CycleGAN approach was proposed and applied to the image transfer task, obtaining significant improvements on several standard image domain transfer databases. In this paper, we use a feature map regularized CycleGAN and combine it with the augmented semi-supervised approach in order to further improve CycleGAN Speech enhancement. Significant improvements in the speech enhancement task by means of several standard measures are obtained by using the proposed approach in comparison to baseline CycleGAN as well as the augmented CycleGAN approach.

Keywords: feature map regularization; data augmentation; CycleGAN; speech enhancement.

1. INTRODUCTION

Speech enhancement (SE) of the perturbed speech is a very important front-end pre-processing stage component, as it is used as a preprocessing component for the main speech processing systems, such as Speaker Recognition/Verification [1, 2] or Automatic Speech Recognition (ASR) systems [3–7] deployed in noisy conditions. Recently, the Deep Neural Network (DNN) approach showed significant improvement in the task of single channel SE. One of the first improvements in that direction was reported in [8] and expended in [9, 10], where deep neural networks are used to estimate a mask in the Mel frequency domain by using a set of time-frequency unit level features. A supervised learning approach is reported and developed as a regression task in various papers (see [8–18]), but it requires a large amount of data in order to avoid over-fitting, i.e., the network does not generalize well to the unseen data. In order to overcome the mentioned problems, an unsupervised

Citation: B. Popović and M. Janev, “Speech Enhancement by CycleGAN Using Feature Map Regularization,” *Journal of Computer and Forensic Sciences*, vol. 2 no. 1, pp. 19–28, 2023, https://doi.org/10.5937/jcfs2-45100.

approach in the form of cycle-consistent adversarial networks CycleGAN is proposed in [19]. The main idea was to use the unsupervised Generative Adversarial Network GAN architecture, proposed in [20], but in both directions (mapping from noisy to clean domain, which is the appreciated result, but also mapping from clean to noisy domain in order to regularize direct mapping, where one tends to make direct mapping "close" to bijective), where the algorithm is learned on two pools of unlabeled, i.e., unpaired images. It gained great success in the area of image translation (see [19, 24–26]). It has also been applied in the area of speech processing (see [27–36]), with an accent on applying the CycleGAN approach in [31–36]. For example, in [32], the authors proposed a front-end based on Cycle-Consistent Generative Adversarial Network (CycleGAN), which transforms naturally perturbed speech into normal speech and hence improves the robustness of an ASR system. In [34], the authors proposed a non-parallel voice conversion (VC) method that can learn a mapping from source to target speech without relying on parallel data, based on the CycleGAN approach, and they expended on that in [35] and [36]. Nevertheless, in speech processing applications, it is hard to collect a sufficient amount of data on the noisy part of the speech (on the clan part of the speech, a sufficient amount of data is usually available, as there is large number of ASR data bases with clean speech). Thus, motivated by the paper [21], which elaborates on the general problem of using CycleGAN on an asymmetric scarce data problem, applied to an image transfer task, a speech enhancement using augmented semi-supervised learning (SSL) CycleGAN approach is proposed in a task of a single-channel SE [22], obtaining improvements in comparison to the baseline CycleGAN SE method as measured by several well-established speech quality measures. Recently, a CycleGAN regularized by similarity between feature maps corresponding to coder and decoder networks, respectively, reported in [23], obtained significant improvement in comparison to the baseline CycleGAN in the task of image domain translation.

In this paper, we explore the approach in [23] in order to improve the quality of speech in SE task, i.e., in the task of translation of noisy to clean speech. Namely, we apply the feature map regularization technique reported in [23], combined with the augmentation approach reported in [21] and [22], in order to further improve the results of SE in the scarce domain (lack of data on the noisy domain side). As can be seen in the section with experimental results, we obtained improvements in comparison to the baseline system reported in [22].

2. PREVIOUS WORK

In this section, we elaborate on some existing approaches, mainly baseline CycleGAN [19], as well as the semi-supervised augmentation approach SSL CycleGAN [21, 22] and the regularized feature map approach [23].

2.1 Baseline CycleGAN

We first mention the Generative Adversarial Network (GAN), which is the generative network model that aims to model the particular distribution represented by the observations in the given pool of data. We start with the discriminator network D, which is designed to discriminate between the samples generated by the generator network G and the ground truth observations. On the other hand, the generator G models the true data distribution by learning to confuse the discriminator, thus competing in order to reach the Nash equilibrium expressed by the mini-max loss of the training procedure, where the optimization problem is given by

$$\min_{G} \max_{D} E_x[\ln D(x)] - E_z[1 - \ln D(G(z))], \tag{1}$$

where x is distributed according to $p(x)$, while the hidden variable z is distributed according to $p(z)$. On the other hand, when applying GAN architecture in image (or any other) domain translation task (translating from the "left" domain X, to the "right" domain Y), there is the problem of ill-posedness where there are "large" regions in the X domain that are mapped to the same point in the Y domain. In order to regularize the mentioned problem and avoid such situation, the cycle consistency loss is invoked in the overall loss function [19], using two domain translators $L : X \to Y$, $M : Y \to X$ in the form of GAN architectures, realized in mutually opposite directions, enforcing the bijectivity of both L by adding the additional penal in the loss function (1). Namely, the mentioned "close to bijective" could be stated as $M(L(x)) \approx x$, $x \in X$ and $L(M(y)) \approx y$, $y \in Y$, thus obtaining the following additional penal to the loss function:

$$\Omega_{cyc}(L, M) = E_x \| M(L(x) - x \|_1 + E_y \| M(L(y) - y \|_1, \tag{2}$$

thus making the overall loss function in the following form:

$$\Omega(L, M, D_X, D_Y) = \Omega_{adv}(L, M, D_X, D_Y) + \lambda_{cyc}\Omega_{cyc}(L, M), \tag{3}$$

for some $\lambda_{cyc} > 0$ where the adversarial loss Ω_{adv}, according to the concept in (1), is defined as

$$\Omega_{adv}(L, M, D_X, D_Y) = \Omega_{adv}(L, D_Y) + \Omega_{adv}(M, D_X), \tag{4}$$

with

$$\begin{aligned} \Omega_{adv}(L, D_Y) &= E_y[\ln D_Y(y)] - E_x[1 - \ln D_Y(L(x))] \\ \Omega_{adv}(M, D_X) &= E_x[\ln D_X(x)] - E_y[1 - \ln D_X(M(y))] \end{aligned} \tag{5}$$

and $\lambda_{cyc} > 0$ being regularization coefficient, controlling the amount of regularization.

2.2 Augmented SSL CycleGAN

In [21], the authors tackled the general problem of asymmetric scarce data problem, i.e., the scarce domain problem applied in the image transfer task, and applied the proposed approach in the SE task in [22], which invoked the Augmented (Bootstrapped) Semi-Supervised CycleGAN (BTS SSL CycleGAN) in order to overcome the mentioned problem, by applying two strategies. First, by using the relatively small amount of available labeled training data from the scarce domain and SSL approach, overfitting of the discriminator of the scarce domain is prevented, since the amount of data in the scarce domain is limited.

Then, after the initial learning, augmentation of the scarce domain is introduced, by periodically adding the artificially generated examples provided by the GAN network mapping from the regular to the scarce domain. The SSL part of the cost function is given by

$$\Omega_{SSL}(L,M)=\frac{1}{M}\left\{\sum_{i=1}^{M}\left\|L(x_i)-y_i\right\|_1+\sum_{i=1}^{M}\left\|M(y_i)-x_i\right\|_1\right\} \tag{6}$$

thus obtaining the overall cost as

$$\Omega(L,M,D_X,D_Y)=\Omega_{adv}(L,M,D_X,D_Y)+\lambda_{id}\Omega_{id}(L,M)+\lambda_{cyc}\Omega_{cyc}(L,M)+\lambda_{SSL}\Omega_{SSL}(L,M)$$

For some $\lambda_{SSL}>0$ where M is the number of paired examples $(x_i,y_i)\in X\times Y$, $i=1,\dots,M$, available for training, while all other training examples in the data base are unpaired. Concerning the augmentation invoked in [21] and applied in [22], augmentation of the discriminator corresponding to the scarce noisy speech domain is performed by adding the samples obtained by the inverse network mapping from the full to the scarce domain after a number of initial iterations, where D_X is forced not to overfit, by applying the SSL strategy, i.e., by using cost (6) and thus initially training the discriminator D_X as well as the generator G_X that corresponds to the scarce domain. Then, the augmentation is performed as follows: samples generated by the generator M are periodically added to the pool of data corresponding to the scarce domain, thus modifying its statistics, i.e., pdf. Actually, in every k-th training iteration, additional L samples generated by the network $M^{(k)}$ (the current state of the network that maps from the full domain Y to the scarce domain X) are added to the pool of the discriminator D_X. Also, additional well-known identity penal $\Omega_{id}(L,M)$ is added in order to further regularize (see [22]), defined by

$$\Omega_{id}(L,M)=E_y\{L(y)-y\}+E_x\{M(x)-x\} \tag{7}$$

2.3 Feature Map Regularized CycleGAN

Although CycleGAN uses cycle consistency loss defined by (2) in order to enforce "close to the bijectivity" of the direct and the inverse mappings L and M, in [23], a regularized feature map approach is proposed by introducing the cycle consistency type of loss that involves feature maps, where several distances that measure similarity between those are invoked. Thus, the original CycleGAN is additionally regularized. The loss function is then expended as

$$\Omega(L,M,D_X,D_Y)=\Omega_{adv}(L,M,D_X,D_Y)+\lambda_{cyc}\Omega_{cyc}(L,M)+\lambda_{FMPcyc}\Omega_{FMPcyc}(L,M)\,. \tag{8}$$

The regularizer Ω_{FMPcyc} is obtained as follows: Let $F_L^{(f)\tilde{x}}\in R^{m_f\times n_f\times d}$ and $F_L^{(l),\tilde{x}}\in R^{m_l\times n_l\times d}$ be the first and last feature map tensor of network implementing L. If one considers unwrapped tensors along the third dimension, i.e., matrices $\hat{F}_L^{(l)\tilde{x}}=unroled(F_L^{(l)\tilde{x}})\in R^{m_l n_f\times d}$, $\hat{F}_L^{(l)\tilde{x}}=unroled(F_L^{(l)\tilde{x}})\in R^{m_l n_f\times d}$ and considers those to be matrices of $m_f n_f$ observations in R^d and $m_l n_l$ observations in R^d and the same for $F_M^{(f)\tilde{y}}\in R^{m_f\times n_f\times d}$ and $F_M^{(l),\tilde{y}}\in R^{m_l\times n_l\times d}$ be the first and last feature map tensor of network implementing M, we obtain the ML estimates of the mean and the covariance of the probability density function (PDF) describing the statistics that renders $\hat{F}_L^{(f)\tilde{x}}$ and $\hat{F}_M^{(f)\tilde{y}}$ as well as $\hat{F}_M^{(f)\tilde{y}}$ and $\hat{F}_M^{(l),\tilde{y}}$ (the assumption is that the PDFs are multivariate Gaussians), thus enabling to introduce $\Omega_{FMPcyc}(L,M)$ in the form of various informational as well as Riemannian distance/similarity measures between the mentioned Gaussian PDFs G_1, G_2, i.e., parameters of those (see [23]), such as for example those given by simple

$$d_1(G_1,G_2) = \left\|\Sigma_1 - \Sigma_2\right\|_1 + \left\|\mu_1 - \mu_2\right\|_1 \tag{9}$$

$$d_2(G_1,G_2) = \left\|\Sigma_1 - \Sigma_2\right\|_2 + \left\|\mu_1 - \mu_2\right\|_2 \tag{10}$$

or based on informational similarity measure such as, for example, KL divergence between G_1 and G_2

$$d_{KL}(G_1,G_2) = KL(G_1 \,\|\, G_2), \tag{11}$$

with $KL(G_1 \,\|\, G_2)$ given by

$$KL(G_1 \,\|\, G_2) = \frac{1}{2}(\mu_1 - \mu_1)^T \Sigma_1^{-1}(\mu_1 - \mu_1) + \frac{1}{2} tr\left(\Sigma_1^{-1}\Sigma_2\right) - d \tag{12}$$

or, for example, Riemannian distances (see [23]), which we do not give here. Thus, we obtain the reguralizer $\Omega_{FMPcyc}(L,M)$ as

$$\Omega_{FMPcyc}(L,M) = E_{\tilde{x}}\left\{d_{gd}(G_L^{(f),\tilde{x}}, G_L^{(l),\tilde{x}})\right\} + E_{\tilde{y}}\left\{d_{gd}(G_M^{(f),\tilde{y}}, G_M^{(l),\tilde{y}})\right\} \tag{13}$$

where we set $gd \in \{1,2,KL\}$, and $G_L^{(f),\tilde{x}}$ and $G_L^{(l)\,\tilde{x}}$ are Gaussians with parameters obtained by ML estimates from observations in $\hat{F}_L^{(f),\tilde{x}}$ and $\hat{F}_L^{(l)\,\tilde{x}}$ respectively, while $G_M^{(f),\tilde{y}}$ and $G_M^{(l),\tilde{y}}$ are Gaussians with parameters obtained by ML estimates from observations in $\hat{F}_M^{(f)\,\tilde{y}}$ and $\hat{F}_M^{(l)\,\tilde{y}}$ respectively. We refer to the previously mentioned feature map reguralized CycleGAN as FMR-CycleGAN.

3. APPLICATION OF BTS SSL FMR-CYCLEGAN TO SPEECH ENHANCEMENT

In this paper, we combine the approach of Augmented (Bootstrapped) BTS-SSL-CycleGAN with the FMR-CycleGAN approach, both presented in Subsections 2.2 and 2.3, respectively, in order to obtain better speech enhancement results in comparison to the baseline BTS-SSL-CycleGAN approach reported in [22]. We actually additionally regularize speech enhancement BTS-SSL-CycleGAN designed to operate in conditions of low amounts of noisy speech training data (scarce domain) by applying the FMR approach. For simplicity, in order to lower the training time, we use a simpler form of feature map regularization given by

$$\Omega_{FMPcyc}(L,M) = E_{\tilde{x}}\left\{d_{gd}(\hat{F}_L^{(f),\tilde{x}}, \hat{F}_L^{(l),\tilde{x}})\right\} + E_{\tilde{y}}\left\{d_{gd}(\hat{F}_M^{(f),\tilde{y}}, \hat{F}_M^{(l),\tilde{y}})\right\}, \tag{14}$$

with

$$d_{gd}(\hat{F}_L^{(f),\tilde{x}}, \hat{F}_L^{(l),\tilde{x}}) = KL(vect(\hat{F}_L^{(f),\tilde{x}}), vect(\hat{F}_L^{(l),\tilde{x}})), \tag{15}$$

$$d_{gd}(\hat{F}_M^{(f),\tilde{x}}, \hat{F}_M^{(l),\tilde{x}}) = KL(vect(\hat{F}_M^{(f),\tilde{y}}), vect(\hat{F}_M^{(l),\tilde{y}})), \tag{16}$$

where we consider $vect(\hat{F}_L^{(f),\tilde{x}})$, $vect(\hat{F}_L^{(l),\tilde{x}})$, $vect(\hat{F}_M^{(f),\tilde{y}})$, $vect(\hat{F}_M^{(l),\tilde{y}})$ as Euclidian vectors and use KL divergence defined as

$$KL(p,q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i},\ p,q \in R^n. \tag{17}$$

3. NETWORK ARCHITECTURE

In the implementation of the proposed approach, we use the network architecture also used in [22], proposed in [35], which is again the modified CycleGAN architecture proposed in [26], based on the gated CNN proposed in [36], which is the state-of-the-art speech processing architecture. The hyper-parameters are set as follows: The number of epochs was set to 5000, the mini batch size to 1, the generator learning rate to $\eta_G = 0.0002$, the generator learning rate decay set to $\nu_G = \eta/(2\cdot 10^6)$, the discriminator learning rate to $\eta_D = 0.0001$, and the discriminator learning rate decay to $\nu_D = \nu_G$. Also, mel-cepstral coefficients, logarithmic fundamental frequency and aperiodicities are extracted every 5 ms from a randomly chosen fixed-length segment of 128 frames. One-dimensional CNN is used as a generator to capture the relationship among features while preserving the temporal structure.

4. EXPERIMENTAL RESULTS

In this section, we present the experimental results of the proposed FMR regularized BTS-SSL-CycleGAN (with regularization performed as in Section 3) in comparison to baseline CycleGAN as well as BTS-SSL-CycleGAN in a SE task, i.e., in the task of noise speech to clean speech translation. The database contains 200 noisy and clean speech utterances (used as reference signals in order to assert the objective voice quality measures, so it is actually database of paired examples, although it is used in SSL manner, as described in previous sections) produced by 10 male and 10 female speakers (10 utterances per each). The database is also divided in two by means of clean/noisy speech, i.e., 100 clean and 100 noisy speech utterances. As far as the type of noise, it is a mixture of stationary Gaussian noise and various types of non-stationary noise components, including traffic and office noise, crackling, creaking, etc.

In order to assess the objective voice quality of the resulted denoising procedure, we utilize the family of PESQ standards used by phone manufacturers, network equipment (ITU-T P.862 standard recommendation). Namely, we utilize PEQMOS as well as MOSLQO measures. Also, we use a signal-based spectro-temporal measure of similarity between referent and degraded human speech that models human speech, i.e., Virtual Speech Quality Objective Listener (ViSQOL), particularly designed for Voice over IP (VoIP) transmissions.

Using ViSQOL, two other measures are obtained, VISQOL and NSIM. In Tables 1 and 2, the experimental results of the SE task on the described database, expressed in the PEQMOS and MOSLQO measures, are given for the proposed FMR regularized BTS-SSL-CycleGAN, in comparison to the baseline CycleGAN as well as BTS-SSL-CycleGAN. Also, in Tables 2 and 3, the experimental results of the SE task on the same database and for the same algorithms are presented, but expressed in the VISQOL and NSIM measures, respectively. It can be seen that the proposed FMR regularized BTS-SSL-CycleGAN obtains better results on the used database in comparison to the mentioned baseline methods when considered in the context of all objective voice quality measures used. Percentage in the first column of all tables presents the percentage of paired examples used in the actual SSLtype learning of BTS-SSL-CycleGAN as well as FMR regularized BTS-SSL-CycleGAN.

Table 1: PEQMOS: noisy to clean speech task.

[%]	CycleGAN	BTS-SSL	FMR BTS-SSL
25	-	0.837	0.845
50	-	0.857	0.863
100	0.828	0.867	0.874

Table 2: MOSLQO: noisy to clean speech task.

[%]	CycleGAN	BTS-SSL	FMR BTS-SSL
25	-	1.178	1.183
50	-	1.191	1.215
100	1.178	1.238	1.317

Table 3: VISQOL: noisy to clean speech task.

[%]	CycleGAN	BTS-SSL	FMR BTS-SSL
25	-	1.392	1.412
50	-	1.425	1.510
100	1.369	1.452	1.573

Table 3: NSIM: noisy to clean speech task.

[%]	CycleGAN	BTS-SSL	FMR BTS-SSL
25	-	0.579	0.584
50	-	0.581	0.612
100	0.569	0.585	0.615

FUNDING:

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT:

Not applicable.

INFORMED CONSENT STATEMENT:

Not applicable.

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

REFERENCES

[1] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, "An Overview of Noise Robust Automatic Speech Recognition," *IEEEACMTransASLP*, vol. 22, no. 4, Apr., pp. 745–777, 2014.

[2] J. Li, L. Deng, R. Haeb-Umbach, and Y. Gong, *Robust Automatic Speech Recognition: A Bridge to Practical Applications*. Academic Press, 2015.

[3] G. Hinton, L. Deng, D. Yu et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[4] N. Jaitly, P. Nguyen, A. Senior, and V. Vanhoucke, "Application of Pretrained Deep Neural Networks to Large Vocabulary Speech Recognition," In Interspeech 2012: Proceedings of the 13th Annual Conference of the International Speech Communication Association, 2012, pp. 2578–2581.

[5] T. Sainath, B. Kingsbury, B. Ramabhadran et al., "Making Deep Belief Networks Effective for Large Vocabulary Continuous Speech Recognition," In Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop, 2011, pp. 30–35.

[6] L. Deng, J. Li, J. T. Huang et al., "Recent Advances in Deep Learning for Speech Research at Microsoft," In Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013.

[7] D. Yu and J. Li, "Recent Progresses in Deep Learning Based Acoustic Models," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 3, pp. 396–409, 2017.

[8] A. Narayanan and D. Wang, "Ideal Ratio Mask Estimation Using Deep Neural Networks for Robust Speech Recognition," in Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 7092–7096.

[9] DeLiang Wang and Jitong Chen, "Supervised Speech Separation Based on Deep Learning: An Overview," arXiv:1708.07524, 2017.

[10] Bo Li and Khe Chai Sim, "A Spectral Masking Approach to Noise-robust Speech Recognition Using Deep Neural Networks," *IEEE/ACM Transactions on ASLP*, vol. 22, no. 8, pp. 1296–1305, 2014.

[11] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7–19, 2015.

[12] A. L. Maas, Q. V. Le, T. M. O'Neil, O. Vinyals, P. Nguyen, and A. Y. Ng, "Recurrent Neural Networks for Noise Reduction in Robust ASR," In Proc. Interspeech, 2012.

[13] Z. Chen, Y. Huang, J. Li, and Y. Gong, "Improving Mask Learning Based Speech Enhancement System with Restoration Layers and Residual Connection," In Proc. Interspeech, 2017.

[14] Y. Wang, A. Narayanan, and D. Wang, "On Training Targets for Supervised Speech Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1849–1858, 2014.

[15] F. Weninger, H. Erdogan, S. Watanabe et al., "Speech Enhancement with LSTM Recurrent Neural Networks and Its Application to Noise-robust ASR," In International Conference on Latent Variable Analysis and Signal Separation, 2015.

[16] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech Enhancement Based on Deep Denoising Autoencoder," In Interspeech, 2013, pp. 436–440.

[17] X. Feng, Y. Zhang, and J. Glass, "Speech Feature Denoising and Dereverberation Via Deep Autoencoders for Noisy Reverberant Speech Recognition," in Proc. ICASSP. IEEE, 2014.

[18] F. Weninger, F. Eyben, and B. Schuller, "Single-channel Speech Separation with Memory-enhanced Recurrent Neural Networks," In Proc. ICASSP. IEEE, 2014, pp. 3709–3713.

[19] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 27, pp. 2672–2680, 2014.

[21] L. Krstanovic, B. Popovic, M. Janev, and B. Brkljac, "Bootstrapped SSL CycleGAN for Asymmetric Domain Transfer," *Applied Science*, vol. 12, no. 7, 3411, 2022, https://doi.org/10.3390/app12073411

[22] B. Popovic, L. Krstanovic, M. Janev, and S. Suzic, "Speech Enhancement Using Augmented SSL CycleGAN," In Proc. 30th European Signal Processing Conference (EUSIPCO 2022), 2022, pp. 1155–1159.

[23] L. Krstanovic, B. Popovic, and M. Janev, “Feature Map Regularized CycleGAN for Domain Transfer,” *Mathematics*, vol 11, no. 2, 2023, https://doi.org/10.3390/math11020372

[24] C. Wang, H. Zheng, Z. Yu, Z. Zheng, Z. Gu, and B. Zheng, “Discriminative Region Proposal Adversarial Networks for High-quality Image-to-Image Translation,” in Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[25] B. AlBahar and J.-B. Huang, “Guided Image-to-Image Translation with Bi-directional Feature Transformation,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.

[26] J. Zhu, T. Park, P. Isola, A.A. Efros, “Unpaired Image-to-Image Translation Using Cycle-consistent Adversarial Networks,” In: IEEE International Conference on Computer Vision, 2017, pp. 2242–2251.

[27] S. Pascual, A. Bonafonte, and J. Serra, “SEGAN: Speech Enhancement Generative Adversarial Network,” In Interspeech, 2017.

[28] C. Donahue, B. Li, and R. Prabhavalkar, “Exploring Speech Enhancement with Generative Adversarial Networks for Robust Speech Recognition,” arXiv preprint arXiv:1711.05747, 2017.

[29] M. Mimura, S. Sakai, and T. Kawahara, “Cross-Domain Speech Recognition Using Nonparallel Corpora with Cycle-consistent Adversarial Networks,” in Proc. ASRU, 2017, pp. 134–140.

[30] Z. Meng, J. Li, Y. Gong, and B.-H. F. Juang, “Adversarial Feature-mapping for Speech Enhancement,” in Proc. Interspeech, 2018.

[31] Z. Meng, J. Li, Y. Gong, and B. Juang, “Cycle-consistent Speech Enhancement,” In Proc. Interspeech 2018, DOI 10.21437/Interspeech.2018-2409

[32] S. H. Dumpala, R. Chakraborty, S. K. Kopparapu, and I. Sheikh, “Improving ASR Robustness to Perturbed Speech Using Cycle-consistent Generative Adversarial Networks,” In Proc. ICASSP 2019, DOI 10.1109/ICASSP.2019.8683793

[33] T. Kaneko and H. Kameoka, “CycleGAN-VC: Non-parallel Voice Conversion Using Cycle-consistent Adversarial Networks,” In Proc. EUSIPCO, 2018, pp. 2114–2118.

[34] T. Kaneko, H. Kameoka, K. Tanaka, and N. Hojo, “CycleGAN-VC2: Improved CycleGAN-Based Non-Parallel Voice Conversion,” in Proc. ICASSP, 2019, pp. 6820–6824.

[35] T. Kaneko, H. Kameoka, K. Tanaka, and N. Hojo, “CycleGAN-VC3: Examining and Improving CycleGAN-VCs for Mel-spectrogram Conversion,” In Proc. Interspeech 2020, DOI:10.21437/interspeech.2020-2280

[36] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, “Language Modeling with Gated Convolutional Networks,” In Proc. ICML, 2017, pp. 933–941.

Journal of COMPUTER AND FORENSIC SCIENCES

CFS 2023, Vol. 2, Issue 1, pp. 29–41
https://doi.org/10.5937/jcfs2-44736
Original research paper

iTrust News Certificate: A Blockchain-Based Solution for News Verification and Reputation Management

Aleksandar Miljković[*], Milan Čabarkapa[2], Filip Miljković[1], and Ljubiša Bojić[3]

[1] Ministry of Interior of the Republic of Serbia & University of Criminal and Police Studies, Belgrade, Serbia; aleksandar.miljkovic@mup.gov.rs; filip.miljkovic@mup.gov.rs

[2] Faculty of Engineering, University of Kragujevac, Serbia; mcabarkapa@kg.ac.rs

[3] Institute for Philosophy and Social Theory, University of Belgrade, Serbia; ljubisa.bojic@instifdt.bg.ac.rs

* Corresponding author: aleksandar.miljkovic@mup.gov.rs

Received: May 30, 2023 • Accepted: June 20, 2023 • Published: July 7, 2023

Abstract: The proliferation of fake news and misinformation in the digital era poses a significant challenge to news organizations and content creators. In this paper, we introduce the iTrust News Certificate, the architecture of an online blockchain-based solution designed to combat fake news, enhance news verification, and maintain reputation within the media ecosystem. Unlike previous attempts, iTrust News Certificate focuses on user-friendly features while ensuring transparency and reliability. By leveraging blockchain technology, iTrust News Certificate establishes a decentralized and immutable ledger for storing news-related metadata. This ledger ensures the integrity and traceability of news articles, making it extremely difficult for malicious actors to tamper with or propagate false information.

Keywords: fake news; blockchain; smart contracts.

1. INTRODUCTION

The proliferation of digital misinformation and fake news has raised concerns globally, as highlighted by the 2022 Edelman Trust Barometer [1]. The survey revealed widespread uncertainty regarding the reliability of information in the media, with a significant percentage of respondents expressing doubts about their ability to differentiate factual news from falsehoods. [2] examines the diffusion of true and false news stories on social media and highlights the rapid spread of false information. Additionally, the potential weaponization of fake news further amplifies the need for effective solutions to address this issue. [3, 4, 5, 6] examine the exposure to untrustworthy websites during the 2016 U.S. election and its implications for belief in false information.

Blockchain technology has emerged as a promising tool to authenticate and verify content, including news and videos, as seen in [7, 8, 9, 10, 11]. Avivah Litan, an American analyst and researcher in the field of cybersecurity and fraud prevention, predicts that by

Citation: A. Miljković, M. Čabarkapa, F. Miljković, and Lj. Bojić, "iTrust News Certificate: A Blockchain-Based Solution for News Verification and Reputation Management," *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 29–41, 2023, https://doi.org/10.5937/jcfs2-44736.

2023, up to 30% of world news and video content will be authenticated using blockchain technology. This development aligns with the concerns expressed in [12], which emphasizes the need for addressing the threats and limitations posed by the misuse of internet technologies, specifically concerning democratic processes and universal freedoms.

However, the absence of widely accepted standards for identifying, labeling, tracking, and responding to digital misinformation presents a significant challenge, as seen in [13, 14, 15, 16]. Without a universally recognized standard, the potential impact of blockchain technology on combating fake news remains limited. Previous initiatives have faced hurdles due to their lack of user-friendliness and the absence of adequate incentives for participation, contributing to their failure.

To address these issues, we propose the iTrust News Certificate, an architecture of a user-friendly online blockchain-based solution designed to provide news organizations and content creators with a reliable and transparent platform. This solution offers a rating tool for users to label potentially fake content, which is then subject to verification by registered users and independent fact-checkers. The incorporation of various post labels based on user perception and the additional feature of media profiling enhance the overall user experience and participation.

iTrust News Certificate revolutionizes the way users interact with news content by introducing a rating tool embedded at the bottom of each post. Unregistered users can easily label posts as potentially fake, triggering further scrutiny by registered users and independent fact-checkers. Only when a post is independently verified as fake, it can affect the post's Score, which is represented by a color-changing system ranging from various shades of green to orange and eventually to red.

Moreover, the iTrust News Certificate goes beyond the binary labeling of fake news. Each post is categorized based on user perception, providing labels such as "educational," "fun," "useful," "emotional," "adult," "polarizing," "opinionated," and "negative news." To address the lack of incentives for user participation in previous initiatives, iTrust News Certificate aims to introduce an engaging and interactive element through media profiling. As an additional feature of the platform, media profiling provides information about news organizations to anyone seeking to assess their credibility and trustworthiness. This feature is designed to be enjoyable for users, encouraging their active involvement in the verification process.

The second section presents materials and methods used in researching the defined problem. The third section proposes a solution and gives a broad overview of the architecture. The discussion and limitations of the proposed solution are presented in the fourth section. The fifth section concludes this work.

2. MATERIALS AND METHODS

To investigate the effectiveness of utilizing blockchain technology in combating fake news, we adopted a multi-faceted approach encompassing previous European projects, initiatives, and startup endeavors. In this section, we describe the materials and methods employed in each direction, including AI algorithm training, deep fake detection, incentivizing high-quality content, and maintaining online identity and reputation.

The first direction we explored involved training AI algorithms to recognize fake news. We examined the methodologies employed by previous European projects, such as AI-4Media and Fandango, the recent "AI to fight disinformation (RIA)" call, as well as other approaches [17, 18, 19, 20] that use AI algorithms with NLP approaches. It is important to note that the current challenge lies in the algorithm's inability to decisively determine the authenticity of information. However, it can assess the level of suspicion associated with a piece of content. The algorithms label information as suspicious, providing a valuable indicator for further investigation.

Addressing the issue of manipulated photos and videos, we investigated the utilization of blockchain technology for deep fake detection. One notable initiative in this field is the New York Times News Provenance Project [21], which was launched in 2019. The project aimed to establish a registry of all published images from various media outlets, accompanied by meta information such as captions, locations, consent, and copyright details. This comprehensive dataset was made verifiable by anyone, facilitating the detection of deep fakes. Additionally, Facebook initiated its Deep Fake Detection Challenge [22] to encourage the development of effective deep fake detection tools.

Another avenue we explored was the concept of incentivizing high-quality content creation using blockchain technology. Civil Media [23], a blockchain startup, proposed a new economy for journalism in which users could reward media outlets for their exceptional work. However, this vision faced challenges and failed to attract sufficient buyers. Other startups, including Poet and Nwzer [24], have also explored similar possibilities. It is important to note that the realization of this idea requires several intermediate steps, allowing smaller content providers to participate and be rewarded gradually.

The final direction we investigated involved maintaining online identity and reputation through blockchain technology. Mavin [25], a startup, initiated this endeavor by developing a browser plugin that enables readers to rate each piece of content based on their assessment. This approach empowers readers to discern the trustworthiness of the content they encounter and engage with. Moreover, users are assigned reputation scores based on factors such as their identity and expertise, influencing the weight of their votes and evaluations.

Throughout our exploration, we remained mindful of the ethical implications surrounding the use of AI algorithms, blockchain technology, and reputation systems. We considered issues related to privacy, data security, algorithmic bias, and the potential for unintended consequences. Our analysis acknowledges the importance of a balanced approach that upholds user trust and protects individual rights while addressing the challenges posed by fake news.

In this section, we outlined the materials and methods employed in our investigation of utilizing blockchain technology to combat fake news. We examined the AI algorithm training approaches, deep fake detection initiatives, and the concept of incentivizing high-quality content and maintaining online identity and reputation. Our analysis serves as a foundation for the subsequent sections, where we present our approach and findings as well as discuss the implications.

3. PROPOSED ARCHITECTURE

In this section, we propose an architecture for a blockchain-based smart contract solution for fake news detection iTrust News Certificate.

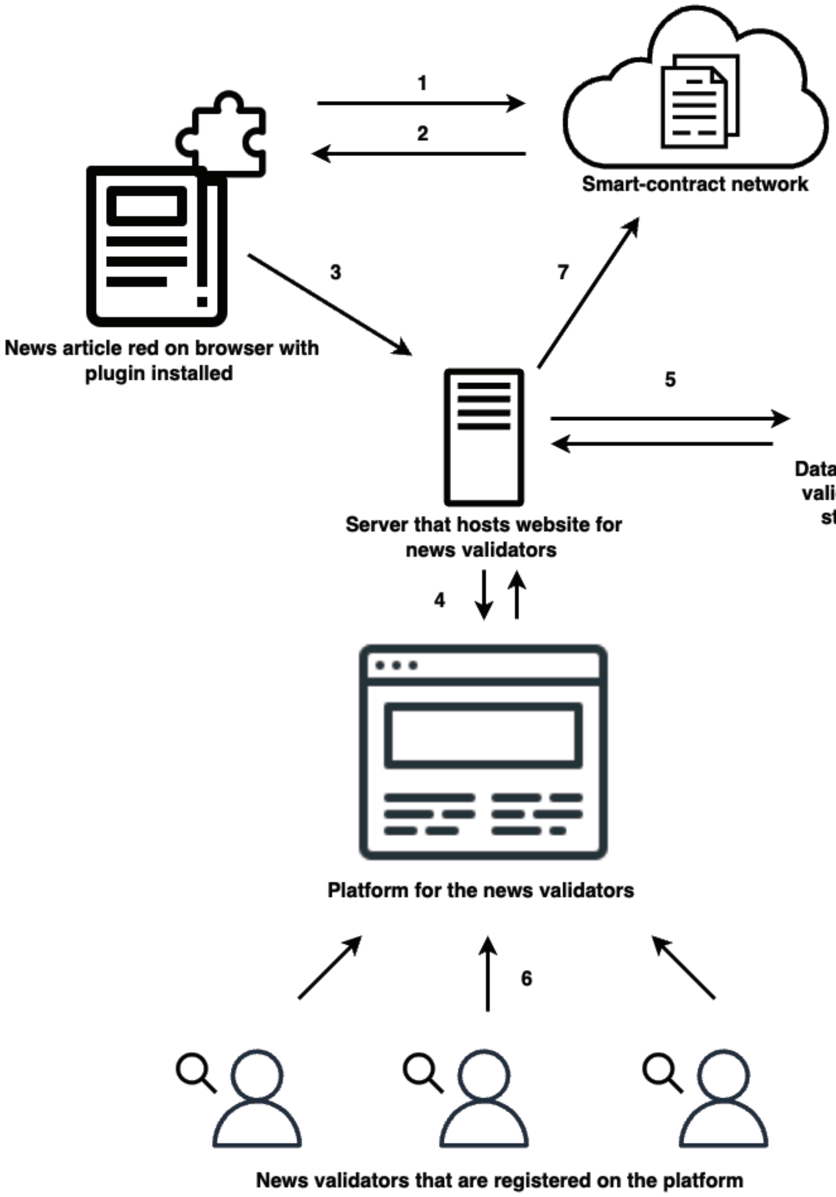

1
2
Smart-contract network
3
7
News article red on browser with plugin installed
5
Server that hosts website for news validators
Database that stores validators data and statistical data
4
Platform for the news validators
6
News validators that are registered on the platform

Figure 1. *The architecture of the iTrust platform.*

As we can see in the architectural diagram shown in Figure 1, the proposed platform consists of several components. The first one is a web browser plugin that monitors the web page that the user is visiting and communicating with the smart contract network directly using web3 technologies and with the platform's web server. The next and most important component is a smart contract network that stores information on reviewed news articles. Apart from that, there is a server that hosts part of the platform that enables news validators to do their part in the system and a database that stores information for registered validators and statistical data.

When a human user of a web browser plugin visits a news article, the dedicated plugin checks whether the article is present in the network. This is done by using the URL of the news article (step 1 in Figure 1). If the URL of the news article is present in the network, the data from that article are retrieved and its rating is displayed (step 2). When the URL is not present, a notification is sent to news validators that there is a new article that is yet to be verified (step 3). The news validator uses the platform's website to access the article by using its credential that is stored in the database and rate the article (steps 4, 5, and 6). An article review is then stored on the smart contract network for future usage.

3.1. Smart Contract

The first component of the architecture involves the creation of a smart contract using the Solidity programming language and the Remix development environment. The smart contract establishes the rules and conditions for user validation and the storage of relevant information, such as news URLs, ratings, user identities, and publishers. Creating a smart contract using the Solidity programming language and the Remix development environment involves several steps, has various use cases, and comes with certain limitations. Let's explore each aspect in detail:

3.1.1 Steps to Create a Smart Contract Using Solidity and Remix

This subsection presents a guide for setting up the Remix IDE, creating and writing Solidity code, compiling the smart contract, deploying it on a blockchain network, and interacting with the deployed smart contract. The Remix IDE is utilized as an integrated development environment to streamline the development process and facilitate the deployment of the smart contract. This subsection serves as a practical resource for researchers and developers aiming to implement and deploy their smart contracts.

To initiate the development process, the Remix IDE is launched through a web browser. This step establishes the foundation for creating, editing, compiling, and deploying smart contracts. Within the Remix IDE, a new Solidity file is generated by clicking on the "+" button and assigning it a .sol extension. This file serves as the container for the smart contract code. Leveraging the Solidity programming language, the structure, functions, and variables of the smart contract are outlined within the newly created Solidity file. Remix's built-in editor enhances the coding experience by providing syntax highlighting and auto-completion functionalities. Navigating to the "Solidity Compiler" tab within the Remix IDE, the desired version of Solidity is selected. Subsequently, the "Compile" button is clicked to initiate the compilation process, which includes error checking and bytecode generation. Within the "Deploy & Run Transactions" tab, the target blockchain network for deploying the smart contract is chosen. Options range from local development environments to various testnets. By clicking the "Deploy" button, the compiled smart contract is deployed onto the selected blockchain network. Successful deployment necessitates the presence of a connected Ethereum wallet. Once the smart contract is deployed, Remix provides an intuitive interface for interacting with its functions and variables. Researchers and developers can conveniently invoke functions, send transactions, and observe state changes within the contract. By leveraging Remix's capabilities, researchers and developers can streamline the implementation of their smart contracts and leverage the Ethereum blockchain's potential.

It is important to note that Solidity and Remix are just two options for creating smart contracts. There are other programming languages (e.g., Vyper) and development environments available as well.

3.1.2 Use Cases of Smart Contracts

Smart contracts play a pivotal role in powering a diverse range of decentralized applications (DApps), eliminating the need for intermediaries, and enabling transparent and trustless interactions among users, as shown in [26]. Beyond their application in DApps, smart contracts offer extensive capabilities in various domains. They facilitate tokenization and crowdfunding efforts, providing a seamless mechanism for the creation and management of tokens, enabling processes such as Initial Coin Offerings (ICOs), crowdfunding campaigns, and the tokenization of assets. Moreover, in supply chain management, smart contracts offer invaluable benefits by tracking and validating the movement of goods throughout the supply chain, thereby enhancing transparency and mitigating the risks of fraud, as shown in [27]. Another practical use case for smart contracts is their ability to function as automated escrow services [28]. By holding funds until predetermined conditions are met, they ensure secure and equitable transactions. Furthermore, smart contracts offer a robust solution for voting systems, enabling the development of transparent and auditable voting mechanisms [29, 30]. By leveraging smart contracts, the possibility of tampering or manipulation within voting systems can be significantly reduced, reinforcing the integrity of democratic processes.

3.1.3 Limitations of Smart Contracts

There are some limitations when developing smart contracts:

- Immutability: A smart contract's code cannot be modified once deployed, which means any bugs or vulnerabilities discovered after deployment may be difficult to rectify.
- Lack of External Data Sources: Smart contracts have limited access to external data sources, making it challenging to incorporate real-time data or interact with external APIs directly.
- Scalability: Ethereum, the most popular blockchain for smart contracts, has scalability limitations, leading to potential congestion and higher transaction fees during periods of high network usage.

- Complexity and Security Risks: Writing secure smart contracts requires expertise, as incorrect code can lead to unintended consequences or vulnerabilities that may be exploited.
- Legal and Regulatory Concerns: The legal and regulatory frameworks surrounding smart contracts are still evolving, which may pose challenges in certain jurisdictions.

3.2. Testing Smart Contract Using Truffle

To ensure the reliability and integrity of the developed smart contract, a comprehensive testing phase is essential. This section explores the utilization of Truffle tools for testing smart contracts, emphasizing the significance of thorough testing before deployment to the network.

Truffle is a popular development framework and a suite of tools used for testing, compiling, and deploying smart contracts on the Ethereum blockchain. It provides a comprehensive environment for building decentralized applications (DApps) and simplifies the process of testing smart contracts. The Truffle Framework offers a development environment with built-in capabilities for smart contract compilation, deployment, and testing. It streamlines the workflow by providing a standardized structure for organizing one's project and managing dependencies. Truffle integrates the Mocha testing framework and the Chai assertion library to facilitate writing and executing tests for smart contracts. Mocha provides a testing framework structure, and Chai offers various assertion styles to make test assertions more expressive and readable. Truffle follows a specific directory structure for organizing test files. By default, test files are stored in the test directory within one's Truffle project. Truffle automatically discovers and runs the tests defined in these files. Truffle allows one to write tests for smart contracts using JavaScript. One can use the Truffle API and web3.js (a JavaScript library for interacting with Ethereum) to interact with deployed contracts, simulate transactions, and verify contract behavior. Truffle provides a migration system that helps manage the deployment of smart contracts to various Ethereum networks. Migrations are written in JavaScript and enable one to specify the sequence of contract deployments and any required initialization steps. Truffle supports automated testing through scripts. One can define scripts that set up the testing environment, deploy contracts, and run tests automatically. This is particularly useful for continuous integration (CI) and continuous deployment (CD) workflows. Truffle can generate code coverage reports to assess the extent to which one's tests cover the smart contract code. This feature helps identify areas of the codebase that lack test coverage and ensures a higher degree of reliability and security. Truffle facilitates testing contracts on different Ethereum networks, such as development, testnet, or mainnet. It provides network configuration options, allowing one to define multiple networks and switch between them easily during testing. Truffle integrates seamlessly with Ganache, a personal Ethereum blockchain for development and testing purposes. Ganache provides a local blockchain environment, allowing one to run tests quickly and simulate various scenarios.

Truffle's testing tools and features greatly simplify the process of writing and executing tests for smart contracts. They enable developers to ensure their contracts' correctness, reliability, and security before deploying them to production networks.

3.3. Deploying Smart Contract to the Ethereum Network

Once the smart contract has been thoroughly tested, it is deployed on the Ethereum network, enabling decentralized and immutable execution of the application. The deployment process, including gas optimization and contract address management, is discussed in this section.

Deployment of the contract can be done directly from the Remix IDE by connecting it to an Ethereum network. To deploy a contract, an Ethereum wallet is needed. The wallet holds the necessary funds to pay for the deployment transaction fees. There is a need for wallets like MetaMask, Mist, or any other Ethereum wallet that supports interacting with smart contracts. We should determine the Ethereum network on which one wants to deploy their contract. One can choose the Ethereum mainnet, a testnet like Ropsten, Rinkeby, or Kovan, or even a local development network like Ganache. In addition, the network may be selected and the contract bytecode and constructor arguments may be specified.

Once the contract is successfully deployed, it can be accessed via its contract address. One can send transactions to execute contract functions, read the contract state, and interact with the contract's public interface. After the deployment of the contract, the deployment transaction needs to be confirmed by the Ethereum network. The confirmation time varies depending on the network's congestion and the gas price one set for the transaction. It is important to note that deploying a smart contract incurs transaction fees (gas costs) for the deployment transaction and any subsequent contract interactions. The gas cost is paid in Ether (ETH) and depends on the complexity of the contract and the network conditions.

3.4. User-friendly Site for News Verification

Creating a user-friendly site is crucial for promoting the widespread adoption of the application. This section focuses on the implementation of a web interface using web3 technology, allowing validated users to view and rate news articles. User authentication and access control mechanisms are integrated to ensure the integrity of the rating system.

To enhance accessibility and engagement, a browser plugin is developed, enabling users to register, rate news articles, and receive notifications for validating new users. The storage of relevant information, such as news URLs, providers, ratings, and user identities, is needed, along with the potential use of hash values for efficient data storage.

3.5. Browser Plugin for Users

Design and development of the user interface for components of the plugin are made using HTML, CSS, and JavaScript. This includes creating the popup when the article exists in the smart contract network, the options page, and other UI elements that the plugin requires. A script runs in the background of the browser and handles the core functionality and interactions of the plugin. This script communicates with the browser's extension APIs and manages events and data. Asynchronous API calls are made to allow the user to browse the page until all the data has been retrieved.

3.6. Analytics Tools for Validators and Users

Beyond user validation, the proposed application incorporates NLP techniques to provide insights into the evaluated news articles. This section explores the generation of basic NLP statistics, such as word frequency analysis, to identify commonly used words and their impact on the overall rating. The potential for creating a valuable dataset for future scientific research is highlighted. There are several natural language processing (NLP) techniques that can be used to provide insights into evaluated news articles.

Text classification is a fundamental NLP technique used to categorize text into predefined categories. It can be applied to news articles to identify the topic or subject matter of the article, such as politics, sports, finance, or entertainment. This classification can provide insights into the distribution of news topics and help in organizing and filtering articles based on their relevance to specific categories.

Sentiment analysis aims to determine the sentiment or opinion expressed in a piece of text. By applying sentiment analysis to news articles, one can gain insights into the overall sentiment towards specific topics, individuals, companies, or events. This information can be useful for understanding public opinion, analyzing market sentiment, or detecting trends.

Named Entity Recognition (NER) is a technique used to identify and classify named entities such as persons, organizations, locations, or dates within a text. By applying NER to news articles, key entities and gain insights into the entities mentioned in the news, their relationships, and their prominence, can be extracted. This can help in identifying important individuals, organizations, or locations associated with specific news topics.

Topic modeling is a statistical technique used to discover abstract topics or themes within a collection of documents. By applying topic modeling to a set of news articles, the underlying topics or themes present in the articles cane be identified. This can provide insights into the distribution of different topics, the emergence of new trends, or the prevalence of certain subjects within the news.

Text summarization techniques can be used to automatically generate concise summaries of news articles. These summaries can provide a quick overview of the main points or key information in the article, allowing users to quickly grasp the content without reading the entire article. This can be particularly useful for news aggregators or applications that provide curated news content.

Information extraction techniques aim to identify and extract structured information from unstructured text. By applying information extraction to news articles, specific pieces of information, such as events, dates, locations, or relationships between entities, can be extracted. This can provide insights into the factual information contained within the articles.

These NLP techniques can be applied individually or in combination to analyze and gain insights from news articles. They enable automated processing, organization, and understanding of large volumes of textual data, facilitating effective information retrieval and analysis in the field of news and media.

4. DISCUSSION

To validate the effectiveness of the iTrust News Certificate, we propose the development of a working prototype that will undergo rigorous testing within the media ecosystem in South East Europe. Furthermore, we invite other media organizations interested in utilizing the iTrust News Solution free of charge to participate in the testing phase. The successful implementation of this prototype will serve as a catalyst to attract additional funding and expand our reach, with the ultimate goal of establishing iTrust News as a European standard for reputation management and the fight against fake news.

By developing a prototype that will be tested within the media ecosystem in South East Europe, as well as inviting other media organizations from other parts of Europe and the whole world to utilize the iTrust News Solution, we aim to establish a robust and effective standard for maintaining the reputation and combating fake news. The successful implementation of the iTrust News Certificate has the potential to attract further funding and grow into a European benchmark, contributing to the restoration of trust and accuracy in news reporting while safeguarding democratic processes and universal freedoms in the digital era.

To support our recommendations, we will conduct statistical analyses on relevant datasets, including the performance of AI algorithms in identifying suspicious content, the detection accuracy of deep fake detection tools, and the impact of reputation scores on content evaluation. These analyses aim to provide quantitative insights into the strengths and weaknesses of the proposed solution.

The development of an adequate UI/UX design poses challenges in ensuring user satisfaction and engagement. The testing phase requires careful organization and precision to guarantee the reliability of the application. Furthermore, real-time monitoring of news updates and the accurate detection of meaningful changes present additional challenges.

It is important to acknowledge certain limitations in our work. The rapidly evolving nature of technology and the limited availability of comprehensive datasets posed challenges in conducting a comprehensive evaluation. Moreover, the feasibility and scalability of implementing blockchain-based solutions require further exploration and real-world testing.

Implementing and maintaining a distributed blockchain system requires significant technical expertise. Developing secure and robust smart contracts, setting up the necessary infrastructure, and ensuring network consensus can be complex and costly. The need for ongoing updates and improvements to address emerging vulnerabilities further adds to the technical complexity.

While blockchain technology ensures transparency and immutability of the recorded reviews, it may pose challenges to reviewers' privacy. The public nature of the blockchain means that reviews can be traced back to individual reviewers, potentially compromising their anonymity. Ensuring the confidentiality of reviewers while maintaining the integrity of the blockchain is a complex task.

5. CONCLUSIONS

This paper presents a comprehensive architectural design overview of an innovative application for user-validated news rating and analysis. By leveraging smart contracts, blockchain technology, and NLP techniques, the application aims to enhance the credibility and evaluation of news articles. The proposed architecture provides a foundation for developing a reliable and user-friendly solution that addresses the challenges associated with misinformation in the digital age.

By leveraging blockchain technology, user ratings, and independent fact-checking processes, iTrust News Certificate presents a user-friendly solution that empowers individuals to effectively identify and combat fake news. With the potential to reshape the media landscape, this innovative platform holds promise for fostering trust, transparency, and accuracy within the realm of news reporting.

The successful implementation of the iTrust News Certificate will contribute to the ongoing efforts to combat fake news and improve the quality of news dissemination. Moreover, it has the potential to encourage media organizations and journalists to prioritize accuracy and accountability in their reporting practices. Ultimately, this innovative platform can significantly foster a more trustworthy and reliable news ecosystem. To the best knowledge of the authors, this is the pioneering architectural blockchain design work in this direction and could be expanded and applicable in different contexts of this topic.

The next step in our future work will be the full implementation of the iTrust News Certificate platform designed during our previous work, deploying it on the real blockchain network and/or blockchain cloud, and putting it into real-life operation.

FUNDING:

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT:

Not applicable.

INFORMED CONSENT STATEMENT:

Not applicable.

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

REFERENCES

[1] Edelman. (2022). Edelman Trust Barometer 2022 [Online]. Available: https://www.edelman.com/trust/2022-trust-barometer [Accessed: June 20, 2023].

[2] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, Mar., pp. 1146–1151, 2018.

[3] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, vol. 31, no. 2, Spring, pp. 211–236, 2017.

[4] S. Lewandowsky, U. K. H. Ecker, and J. Cook, "Beyond Misinformation: Understanding and Coping with the 'Post-Truth' Era," *Journal of Applied Research in Memory and Cognition*, vol. 6, no. 4, Dec., pp. 353–369, 2017.

[5] G. Pennycook and D. G. Rand, "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings," Management Science, vol. 67, no. 11, Nov., pp. 4944–4957, 2019.

[6] A. Guess, B. Nyhan, and J. Reifler, "Exposure to Untrustworthy Websites in the 2016 U.S. Election," *Nature Human Behaviour*, vol. 4, no. 5, May, pp. 472–480, 2020.

[7] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *International Journal of Web and Grid Services*, vol. 14, no. 4, Nov., pp. 352–375, 2018.

[8] S. Zannettou, T. Caulfield, E. De Cristofaro, and M. Sirivianos, "Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web," *ACM Transactions on the Web*, vol. 13, no. 3, Aug., Article 17, 2019.

[9] D. O'Loughlin, P. Matthews, and I. U. Rehman, "An exploration of blockchain-based anti-fake news systems," *Information Processing and Management*, vol. 57, no. 2, Mar., Article 102082, 2020.

[10] Y. Zhang, E. Chang, and G. Li, "Blockchain-based trust management for fake news detection in social media," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, Mar., pp. 51–63, 2020.

[11] Y. Zhang, H. Gao, X. Fan, and E. Chang, "BlockFake: Towards blockchain-based automated fake news detection model for social media," *Future Generation Computer Systems*, vol. 115, May, pp. 495–505, 2021.

[12] Next Generation Internet. (2017, October). HUB4NGI D2.1: NGI Architecture Definition – Version 1.0. [Online]. Available: https://www.ngi.eu/wp-content/uploads/sites/48/2017/10/hub4ngi_d2.1_v1.0.pdf [Accessed: June 20, 2023].

[13] D.M.J. Lazer, et al., "The science of fake news," *Science*, vol. 359, no. 6380, Mar., pp. 1094–1096, 2018.

[14] Y. Benkler, R. Faris, and H. Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press, 2018.

[15] L.M. Neudert, et al., "A Longitudinal Measurement Study of 4chan's Politically Incorrect Forum and its Effect on the Web," ArXiv preprint arXiv:1908.08313, Aug. 2019.

[16] G. Pennycook and D.G. Rand, "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings," *Management Science*, vol. 66, no. 11, Nov., pp. 4944–4957, 2019.

[17] N. Ruchansky, et al., "CsiNet: A Convolutional Neural Network Approach for Fake News Detection," arXiv preprint arXiv:1709.09064, Sept. 2017.

[18] K. Popat, et al., "Deception Detection in News Articles Using Headline Features," In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 2018, pp. 3057–3067.

[19] D.T. Vo, et al., "Combining Graph Convolutional Networks and LSTMs for Fake News Detection," In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, pp. 3442–3452.

[20] M. Potthast, et al., "A Stylometric Inquiry into Hyperpartisan and Fake News," In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, Denmark, 2017, pp. 2319–2324.

[21] News Provenance Project. [Online]. Available: https://www.newsprovenanceproject.com/ [Accessed: June 20, 2023].

[22] AI Research at Facebook. (n.d.). DFDC – Deepfake Detection Challenge Dataset. [Online]. Available: https://ai.facebook.com/datasets/dfdc/ [Accessed: June 20, 2023].

[23] Civil. [Online]. Available: https://civil.co/ [Accessed: June 20, 2023].

[24] NWZER. [Online]. Available: https://nwzer.com/ [Accessed: June 20, 2023].

[25] Fundsup. "Featured on Fundsup: Mavin.org – Content Integrity Movement," Fundsup.co [Online]. Available: https://fundsup.co/featured-on-fundsup-mavin-org-content-integrity-movement/ [Accessed: June 20, 2023].

[26] C. Li, et al., "Smart Contract-Based Crowdfunding for Decentralized Applications," In Proceedings of the 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Prague, Czech Republic, 2017, pp. 456–459.

[27] K. Wang, et al., "Smart Contract-Based Supply Chain Traceability System," In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 3531–3536.

[28] J. Li, et al., "Smart Contract-Based Escrow Service for Secure and Trustworthy E-Commerce Transactions," In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5120–5125.

[29] A. Kiayias, et al., "On Blockchain and Voting," In Proceedings of the 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Paris, France, 2017, pp. 839–847.

[30] M. -V. Vladucu, Z. Dong, J. Medina and R. Rojas-Cessa, "E-Voting Meets Blockchain: A Survey," *IEEE Access*, vol. 11, pp. 23293–23308, 2023, DOI 10.1109/ACCESS.2023.3253682

Journal of COMPUTER AND FORENSIC SCIENCES

CFS 2023, Vol. 2, Issue 1, pp. 43–52
https://doi.org/10.5937/jcfs2-44741
Original research paper

A System Architecture for Preventing Social Engineering Attacks via E-mail

Milan Brkić[1], Aleksa Maksimović[2], and Aleksandar Miljković[3]

[1] Ministry of Interior of the Republic of Serbia; milan.brkic@mup.gov.rs

[2] Ministry of Interior of the Republic of Serbia; aleksa.maksimovic@mup.gov.rs

[3] Ministry of Interior of the Republic of Serbia; aleksandar.miljkovic@mup.gov.rs

* Corresponding author: milan.brkic@mup.gov.rs

Received: May 30, 2023 • Accepted: June 30, 2023 • Published: July 13, 2023

Abstract: Modern business and the expansion of internet technology have caused a great growth in communication via electronic mail. Bearing in mind that the weakest link in any system is the human, the greatest danger of unauthorized access to ICT resources is recognized in this section of the system [6]. For this reason, the greatest attention regarding the protection of ICT systems should be focused on the users and preventive response to phishing campaigns as the most common form of cyber attack. This paper will present the system architecture for preventive response to phishing campaigns. The architecture itself, which will be explained in the continuation of the text, consists of several different modules integrated into a whole. First, a sender analysis module, which would be based on the blacklist principle; next, an email attachment analysis module, which would perform the functions of static and dynamic analysis of potentially malicious attachments; a link analysis module, which would include the application of Cortex, an open source intelligence service; and finally, a text analysis module, based on statistical models.

Keywords: phishing; Cortex; analysis; cyber attack.

1. INTRODUCTION

Phishing attacks have become one of the most prevalent and insidious forms of cybercrime. Phishing is a social engineering attack that tricks individuals into divulging sensitive information or downloading malware through fraudulent emails, websites, or messages. The increasing frequency and sophistication of these attacks pose a significant threat to individuals, organizations, and society as a whole.

This paper focuses on the prevention of phishing attacks through the use of a novel system. To better understand this system, it is important to first examine the nature of social engineering attacks and phishing in general. Social engineering is a tactic used by cybercriminals to manipulate individuals into divulging sensitive information or performing actions that can lead to unauthorized access or data breaches. Phishing is the most prevalent type of social engineering, which uses deceptive emails, websites, or messages to steal perso-

Citation: M. Brkić, A. Maksimović, and A. Miljković, “A System Architecture for Preventing Social Engineering Attacks via E-mail,” *Journal of Computer and Forensic Sciences*, vol. 2, no. 1, pp. 43–52, 2023, https://doi.org/10.5937/jcfs2-44741.

nal or financial information. Phishing attacks can also be used to install malware or ransomware onto a victim's computer, causing significant damage to both individuals and organizations. There are several methods to defend against phishing attacks. One approach is to train individuals to recognize and avoid phishing attempts. This involves educating them about the common tactics used by attackers and how to identify suspicious emails or websites. Another approach is to implement technical defenses, such as spam filters, web filters, and email authentication protocols, which can help detect and block malicious messages or websites. In the second chapter, we are going to talk in more detail about social engineering and phishing, as well as ways to defend against phishing and what is being done in this field.

The novel system we propose uses multiple approaches in combination with machine learning algorithms to analyze incoming emails and identify potential phishing attempts. The system then sends a warning message to the user, advising them to avoid the email or website. The proposed system also includes a reporting mechanism, allowing users to report suspicious emails or websites to IT security personnel. This system is explained in detail in the third chapter. The fourth chapter is reserved for discussion. At the end, we conclude this paper.

2. MATERIALS AND METHODS

2.1. Phishing and Social Engineering

When talking about social engineering and its application from the cyber security point of view, as well as from the data and information security point of view, one must first understand what is meant by social engineering, when it occurs, and where and how it is applied in the everyday lives of the individual and the community. Once we understand these basic concepts related to social engineering, we can try to understand in what way the "bad guys" in the field of cyber security use these techniques that lead to unauthorized access to data and information that can be misused.

Social engineering is a technique or method of human manipulation that has been named in the world of computers and cyber security social engineering, but before the emergence of the cyber world, it existed and was used in various fields, such as marketing, trade, espionage, and everyday life. The very name of this method tells us a lot about it. The two words that describe this method, social and engineering, can be used for the most appropriate definition of it.

The word social makes it clear that we are talking about everyday human life, in a private or professional sense. Society is a collection of living beings connected by the same way of life, an accidental or intentional gathering of several persons, or an association. In this sense, society is defined as the totality of social phenomena, processes, and relationships. This definition of the first word of the name of this method tells us that it is based on man and is closely related to society, at all levels.

The second word in the name of this method is engineering. We can best understand this as a defined way of acting in the execution of a task through certain steps in order to reach the goal. In other words, it is usually possible to represent the method through some kind of algorithm. However, this method is considered a non-technical type of hacker attack, and can be considered a form of art.

The social engineering process typically includes the following phases:

- Examination – The scanning phase is part of the reconnaissance phase, which comes before the examination phase in the social engineering process. During the reconnaissance phase, the attacker gathers information about the target organization, including identifying potential vulnerabilities and weaknesses.
- Development phase of a social engineering attack involves selecting individual targets within the organization being attacked and forming a relationship with the selected targets. However, the claim that attackers usually select people who show signs of being frustrated, overconfident, arrogant, or somehow easy to extract information from is not entirely accurate. While attackers may target individuals who have access to the desired information or object, they may also target individuals who are perceived as helpful or trusting. Attackers may use a variety of tactics, such as building rapport, flattery, or creating a sense of urgency, to manipulate their targets into providing the desired information or taking an unauthorized action. In the development phase, the attacker creates a plan of attack tailored to the selected targets, which may involve creating a pretext, preparing the necessary tools, and practicing the attack.
- Exploitation phase – Attackers use a closer relationship technique with the aim of extracting information, gaining access to a certain ICT system, or achieving the goals they have in mind. However, the claim that if the exploit is successful, the only thing left is to get things done without raising suspicion is not entirely true. After the exploitation phase, attackers may need to maintain access to the compromised system or continue to extract information. Attackers may need to cover their tracks to avoid detection by the target organization's security measures. This may include erasing digital prints, ensuring no items or information are left behind, and taking steps to prevent being identified as an attacker. However, it is important to note that covering one's tracks is not always possible, and there is always a risk of being caught or identified during or after an attack.

To obtain information and fulfill their goal, attackers use several types of techniques to motivate employees to provide information, and those are: authority, social proof, urgency, and scarcity. Attackers are different, and they may have different motives, attributes, and attack characteristics. Some attackers may be motivated by financial gain, while others may be motivated by ideology or revenge. Attackers may also have different skill levels, resources, and access to tools and techniques. There are several common types of phishing attacks [1]:

- Whaling – Whaling is a form of social engineering attack that targets senior executives and high-profile individuals in organizations. It is a type of phishing attack that uses personalized messages and tactics to trick the target into divulging sensitive information or taking an unauthorized action. Whaling attacks may involve impersonating a trusted individual, such as a CEO, to gain the target's trust and obtain sensitive information or access to the organization's systems. Whaling attacks often use sophisticated techniques, such as spear-phishing and social engineering, to create a sense of urgency and manipulate the target into taking the desired action. Whaling attacks can have serious consequences for the targeted organization, including data breaches, financial loss, and damage to the organization's reputation.

- Vishing – Vishing is a type of social engineering attack that uses Voice over IP (VoIP) technology to trick victims into divulging sensitive information or taking an unauthorized action. The term "vishing" is a combination of "voice" and "phishing". In a vishing attack, the attacker typically poses as a trusted authority, such as a bank representative or IT support technician, and uses a recorded or live voice message to create a sense of urgency or fear in the victim. The message may instruct the victim to call a phone number or visit a website to verify their account information or resolve a security issue. When the victim responds, they are prompted to enter their personal or financial information, which the attacker can then use for fraudulent purposes. Vishing attacks can be difficult to detect as they often appear to be legitimate messages from a trusted source. To avoid falling victim to a vishing attack, it is important to be cautious of unexpected or suspicious phone calls and to verify the authenticity of the message before sharing any sensitive information.
- SMS phishing – SMS phishing, also known as smishing, is a type of phishing attack that uses text messages to trick the victim into divulging sensitive information or taking an unauthorized action. The attacker sends a text message with a message that appears to be urgent or important, such as a security alert or account notification, to create a sense of urgency and encourage the victim to act quickly. The message may contain a link or phone number that the victim is instructed to click on or call, which leads to a fake website or automated voice system that prompts the victim to enter their personal or financial information. Once the attacker has obtained this information, they can use it for fraudulent purposes, such as identity theft or financial fraud. SMS phishing attacks can be difficult to detect as they often appear to be legitimate messages from a trusted source. It is important to be cautious of any unexpected or suspicious text messages and to verify the authenticity of the message before taking any action.

Social engineering attacks via email (or phishing emails) are becoming increasingly common and can have devastating consequences for both individuals and organizations. In the case of this method, the attacker usually writes an email, which should convince the victim to access a malicious link, download a file or a document. Such e-mails are sent to many e-mail addresses, and whoever clicks on that link and downloads the file or document is considered infected. One of the possible techniques used by attackers is the farming technique, in which the attacker executes malicious programs on the computer so that all URL traffic is redirected to the attacker's malicious website, thereby enabling the theft of credentials or obtaining some benefit.

In cases where a specific company or network is targeted, the email addresses of its members are targeted. If one takes into account that most companies assign e-mail addresses to their employees based on the first name.surname principle @name_of_the_company.com, the attacker has no problem attacking if they know the names of the workers.

Unlike an ordinary phishing attack, spear phishing attacks are not based on probability but are carefully constructed to be as believable as possible to the victim. This type of attack depends to a large extent on the information gathering stage, as well as on the bait being good enough for the victim to fall for the attack. Email spoofing and website cloning are two common techniques used by scammers to trick victims into divulging sensitive information or taking an unauthorized action. In email spoofing, the attacker falsifies the "From" section of the email to make it appear as if the message is coming from a trusted source, such as that of a bank or a colleague. The message may contain a sense of urgency or a request for personal information, such as login credentials or financial details. If the victim falls for the scam and responds to the email, the attacker can use the information for fraudulent purposes. In website cloning, the attacker creates a fake website that looks identical to a legitimate one, such as a bank or online retailer. The victim is directed to the fake website through a link or email and may be prompted to enter their personal information.

Some of the most common emails that are sent are usually promises of a big prize, or some kind of discount, and all the victims must do is fill out a prompt or enter a website. In these situations, the attacker can be very creative. Events are also known when the attacker allegedly sends a summons to the court via email and asks the user to enter the page to confirm that they received the email. These are some of the more sophisticated attacks. Such attacks are mostly automated due to the huge number of attacks that target large masses of people based on probability theory. This type of attack must be supported from the technical side. The existence of a virus, malware, or malicious site is required. For this type of attack, as well as for most others, a group of tools is most often used called The Social-Engineer Toolkit (SET) created by TrustedSec which is an open-source project. This type of tool allows attackers to carry out social engineering attacks without much technical knowledge.

2.2. Defending against Phishing Attacks

In order to prevent these types of attacks, it is important to have a system in place that can detect and prevent them before they do any damage. Here are some strategies that can be used to prevent social engineering attacks via email.

Employee education: One of the most effective ways to prevent social engineering attacks via email is to educate employees on the various types of scams and how to identify them. This can include providing training on phishing emails, suspicious links, and other common tactics used by scammers.

Employees should be trained to handle emails carefully [5]:

- To verify that a link within an email points to the correct URL, one can hover their mouse over the link without clicking it. This will display the URL destination in the bottom left corner of the browser window. If the URL looks suspicious or does not match the expected website, one should not click on the link. Instead, one should type the website's address directly into the browser or contact the sender to verify the legitimacy of the email. It is also a good practice to avoid clicking on links within unsolicited emails or emails from unknown sources.
- While it is true that clicking on links in emails can be a risky activity, it is not always necessary to avoid them altogether. However, it is important to exercise caution and verify the legitimacy of the email and the link destination before clicking on any links. One should never write down or share passwords under any circumstances.

- Training users not to give their passwords to anyone is an important aspect of cybersecurity awareness. This is because password sharing can lead to unauthorized access to sensitive information and increase the risk of data breaches.
- One should not open an attachment that seems suspicious; it should be sent to the CERT team for analysis.

Spam filtering: Email spam filters can be used to prevent many social engineering attacks from ever reaching an employee's inbox. This can include filtering out messages from suspicious or untrusted sources or blocking messages with certain keywords or phrases.

Two-factor authentication: Implementing two-factor authentication can help prevent attackers from gaining access to email accounts by requiring a secondary form of identification beyond just a password.

Email encryption: Encrypted email systems can help prevent attackers from intercepting and reading sensitive email communications. This is why it is important to use encryption protocols such as Transport Layer Security (TLS) or Secure Sockets Layer (SSL) to secure email communications. These protocols encrypt the data in transit, making it difficult for attackers to intercept and read the messages.

Regular updates and patches: Keeping email systems up-to-date with the latest security patches and software updates can help prevent attackers from exploiting known vulnerabilities.

Security awareness testing: Regular security awareness testing can help organizations identify areas where employees may be vulnerable to social engineering attacks via email and provide targeted training to help them improve their security posture. One of the better ways to check training success is the Gophish tool [4].

Gophish is an open-source phishing simulation tool that allows organizations to test and improve their employees' awareness of phishing attacks. It can be used to create and send simulated phishing emails to employees, track their responses, and provide training and education to help them recognize and avoid phishing attacks. With Gophish, organizations can create realistic phishing campaigns that mimic common attack techniques, such as social engineering, spear phishing, and phishing emails with malicious attachments or links. The tool includes customizable templates for emails and landing pages, as well as detailed analytics and reporting to track the success of the campaigns and identify areas for improvement. Gophish is designed to be user-friendly and easy to set up, with a web-based interface that allows administrators to manage campaigns, track results, and generate reports. It is available for free and can be downloaded from the Gophish website or from GitHub.

By implementing these strategies, organizations can significantly reduce the risk of falling victim to social engineering attacks via email. It is important to note, however, that no system is foolproof, and it is always important to remain vigilant and report any suspicious activity to IT or security personnel.

3. PROPOSED SYSTEM

The system is divided into three key segments:

- Applications for sending emails for analysis;
- System for processing and classifying emails; and
- Systems for training, updating, and improving models.

Each of the segments is described in more detail in the rest of the document and is viewed as a separate system unit that communicates through web services in a controlled network environment.

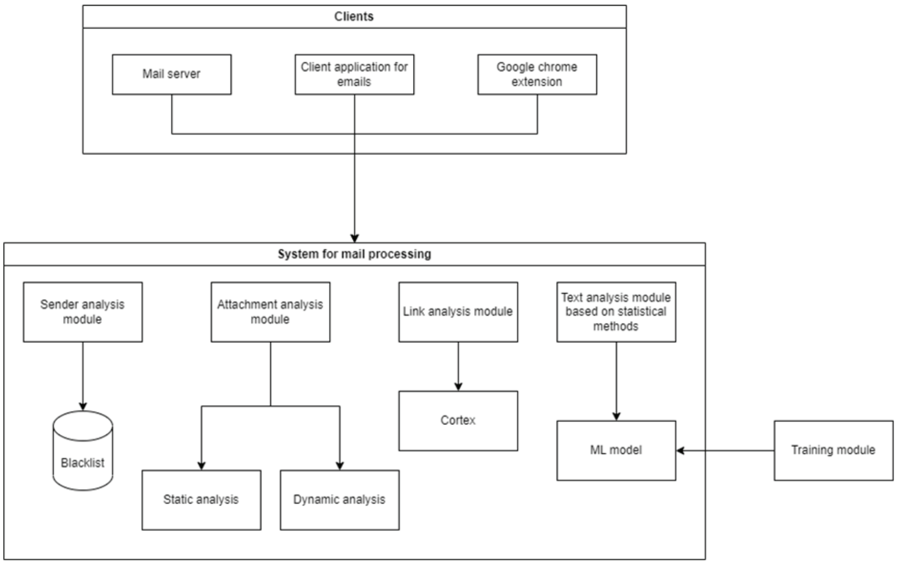


Figure 1. *Architecture of the proposed system.*

3.1. Applications for Sending Emails for Analysis

In order to send an email to the system for analysis, it is necessary to provide a mechanism that will allow the user to log in to this protection system, that is, to be logged in by the system maintainer.

- Browser extensions – A solution exclusively for using a web client for emails. The user decides which email they want to be sent for analysis. This solution can be demanding on the client-side. It also offers the possibility of analyzing emails on the client machine, but there is a security risk if the computer is compromised, that is, if the attacker comes up with statistical models that classify the email, they can modify the attacking email to bypass these rules.
- Server application that automatically analyzes e-mails – An application that automatically listens to the e-mails of users who sign up for this service and sends each e-mail for analysis.

- •A client desktop application that does the same thing as a server application but only on the client-side.
- •An email address to which clients could send emails they consider risky. This is a situation when the user decides which email they want to be sent for analysis by forwarding it to an email address that listens and forwards incoming emails for analysis.

3.2. System for Processing and Classifying Emails

The email processing system is conceived as a series of services that would be divided into several segments, each of which would be in charge of one part of the system.

- •Sender analysis module – List of senders known to have already sent malicious emails, (i.e., black list). It is regularly updated by the system owner.
- •Attachment analysis module – The attachment in the email can be analyzed dynamically and statically. Static could be both hash value analysis and file content analysis, while dynamic would represent the launch of attachments in a sandbox environment and monitoring of work.
- •Module for link analysis – Cortex solution offers analysis of hyperlinks on the web.
- •Text analysis module – The text analysis module would process the text of the email and perform a classification based on a statistical model trained on previously detected spam emails.

In the next part of the text, we would like to emphasize the importance of implementing the Cortex module when responding to phishing emails for analyzing and correlating security data from multiple sources. Cortex is an open-source intelligence platform. It is designed to help security teams automate the analysis and response to security incidents and threats, reducing the time and effort required to investigate and mitigate security incidents. By using Cortex, one gets the possibility to check whether a specific link is on the black list, i.e., the IP address of the server from which the email arrived, as well as other useful information that can be checked using the Cortex module in order to check for malicious content.

Cortex can help security teams automate and streamline their incident response and threat intelligence processes, reducing the time and effort required to investigate and respond to security incidents. It allows users to ingest and correlate data from multiple sources, such as SIEMs, threat intelligence feeds, and security tools, and uses a range of built-in and custom analyzers to perform automated analysis and trigger responses as needed [3].

3.3. Systems for Training, Updating, and Improving Models

This system should allow updating the configuration of each of the models as well as their maintenance and retraining of statistical models:

- •Based on the sender analysis, blacklist is updated;
- •Based on the attachment analysis, the lists of hash values and links to sandboxes are updated;

- To analyze the link path to Cortex;
- For text analysis, training a new statistical model, and updating the existing one.

4. DISCUSSION

There are several problems that the development of such a system would face. Some of the key issues are:

- Training a word processing module is only as good as the training set is representative. The implementation of such a solution would first involve training the model, so that phishing emails that are detected on one's system are examined, using a certain set of words that are repeated in phishing emails, which finally results in a model that is adapted to one's system and needs. Such an approach provides the possibility of a preventive response and reduces the possibility of false positives, because the existing models available on the internet are not sufficiently adapted to the specificities of certain speech areas. Therefore, the goal of this paper and future work is to encourage cyber experts to potentially create a model that will overcome the problems of specific dialects and work on creating a global model that will detect phishing campaigns.
- For dynamic analysis, it is necessary to set up a number of sandboxes that could serve the system. The use of sandbox solutions in this system raises several questions, but one of the most important is the time it takes for the sandbox to analyze a potentially malicious file found in a phishing email, because responding to phishing e-mails requires a prompt reaction in order to prevent data leaks and unauthorized access to the e-mail system.
- Regular training of the text analysis system is required, as well as regular updating of blacklists of malicious senders and hash values of malicious files.
- By regularly updating the model, as well as the database that the system uses to check the status of the email sender's IP address, and the hash value of the file contained in the phishing email, it is possible to receive timely information about current phishing campaigns that aim to harm information systems.

5. CONCLUSION

With the expansion of internet technologies, we are increasingly exposed to cyber attacks. The most frequent type of attack on ICT systems is phishing because it is based on human error, and we all know that humans are the weakest link in the ICT system. Due to all of the above, there was a need to develop a system to protect against phishing attacks.

This paper presents a proposal for system architecture for preventive response to phishing campaigns. As explained in more detail in the "Proposed System" section, the architecture itself is divided into two parts. On the client-side, it would represent an add-on for the client application for working with emails, and an extension for the internet browser, in this case the Google Chrome extension. The architecture also includes a system for processing emails, which would be made up of several modules.

First, a sender analysis module would be based on the blacklist principle. Next, an email attachment analysis module would perform the functions of static and dynamic analysis of potentially malicious attachments. A link analysis module would include the application of Cortex, an open source intelligence service, and finally, a text analysis module would be based on statistical models.

In conclusion, phishing attacks continue to be a significant threat to individuals and organizations worldwide. This paper has outlined the nature of social engineering attacks and the various ways to defend against them. It has also proposed a novel system for preventing phishing attacks, which can help organizations better protect their sensitive information and assets. By implementing a comprehensive approach to preventing phishing attacks, organizations can reduce their risk of falling victim to these insidious cybercrimes.

FUNDING:

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT:

Not applicable.

INFORMED CONSENT STATEMENT:

Not applicable.

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

REFERENCES

[1] G. Weidman, Penetration testing: a hands-on introduction to hacking. San Francisco, CA: No Starch Press, 2014.

[2] C. Hadnagy and M. Fincher, Phishing Dark Waters: The Offensive and Defensive Sides of Malicious E-mails. Hoboken, NJ: Wiley, 2015.

[3] TheHive-Project/Cortex. [Online]. Available: https://github.com/thehive-project/Cortex/. [Accessed: June 15, 2023].

[4] Gophish [Online]. Available: https://github.com/gophish/gophish. [Accessed: June 15, 2023].

[5] N. M. Shekokar, C. Shah, M. Mahajan, and S. Rachh, “An Ideal Approach for Detection and Prevention of Phishing Attacks,” Procedia Computer Science, vol. 49, pp. 82–91, 2015.

[6] G. Nikhita Reddy, G.J. Ugander Reddy, “A study of cyber security challenges and its emerging trends on latest technologies,” ArXiv preprint arXiv:1402.1842, 2014.

Journal of COMPUTER AND FORENSIC SCIENCES

CFS 2023, Vol. 2, Issue 1, pp. 53–61
https://doi.org/10.5937/jcfs2-45169
Original research paper

Application of Machine Learning Methods for Anomaly Detection in Internet Advertising

Marko Živanović[1*], Svetlana Štrbac-Savić[2], and Zlatogor Minchev[3]

[1] Academy of Technical and Art Applied Studies, School of Electrical and Computer Engineering, Belgrade, Serbia; markozivanovic998@gmail.com

[2] Academy of Technical and Art Applied Studies, School of Electrical and Computer Engineering, Belgrade, Serbia; svetlana.strbac@viser.edu.rs

[3] Joint Training Simulation and Analysis Center, Institute of ICT, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences; zlatogor@bas.bg

* Corresponding author: markozivanovic998@gmail.com

Received: June 24, 2023 • Accepted: July 3, 2023 • Published: July 7, 2023

Abstract: This research deals certain with issues regarding downloading data from the Internet, i.e., Internet page advertising, and certain mechanisms to take care of the integrity of the data that is put into the dedicated processing context afterwards. The work also relates to e-commerce, as some advertising scenarios provide high error rates with pricing, which may be unacceptable in various scenarios, such as renting or selling a home. This paper presents a brief overview of the outlier detection methods and machine learning-based classifiers that are used to determine the number of anomalies in the analyzed dataset. This work contributes to the operation of organizations that deal with data accuracy and integrity, such as home rental or selling agencies.

Keywords: data retrieval; machine learning; anomaly detection; e-commerce.

1. INTRODUCTION

The development of the Internet and Internet services provided significant improvements in electronic business and marketing. Sellers who sell goods in such applications often strive to ensure as much profit as possible or to have their product highlighted in the best possible way in order to sell and market the product faster. Through the Internet portal, customers have a detailed overview of products in a certain part of the city or country. However, when entering data by staff, it often happens that the data is not entered correctly, that is, the type of data entered is appropriate, but the values entered for that data in the system are not expected and are inadequate for end users. For example, in the information system for the sale of apartments in the city center, the price per square meter is extremely higher than on the outskirts of the city, in another residential building, or in a building within the same residential unit. Such information systems do not have great accuracy and seriousness, so users usually avoid them and they become less frequented over time. However, a data tracker cannot independently keep track of tens or thousands of records in a database. So, it is obvious that it is necessary to provide automated puzzles of software that employ certain machine learning methods as well as other artificial intelligence algorithms in order to distinguish the data in the ads that is classified as anomaly from those classified as regular data. Various commercial websites that are the most popular on the real estate sales market in the Republic of Serbia were used to check the classifier. Although there are numerous data classification techniques, it is necessary to use the best possible classifiers for these types of data. In machine learning and artificial intelligence, the rule is that one solution is not fully usable in another system. Models for classifying anomalies in data related to real estate prices are not the same as models for classifying anomalies in information system failures per unit of time. Therefore, it is necessary to find a solution that is good enough or acceptable with a certain degree of accuracy.

This paper presents research on machine learning-based anomaly detection classifiers, which are a very good basis for further investigation in the field of anomaly detection in both commercial and non-commercial data sets. Future information systems are expected to have exceptional accuracy and precision, as well as avoid errors that occur during data entry. Precision and accuracy can be achieved by detecting data errors and increasing the severity of various portals and services for the sale and placement of products.

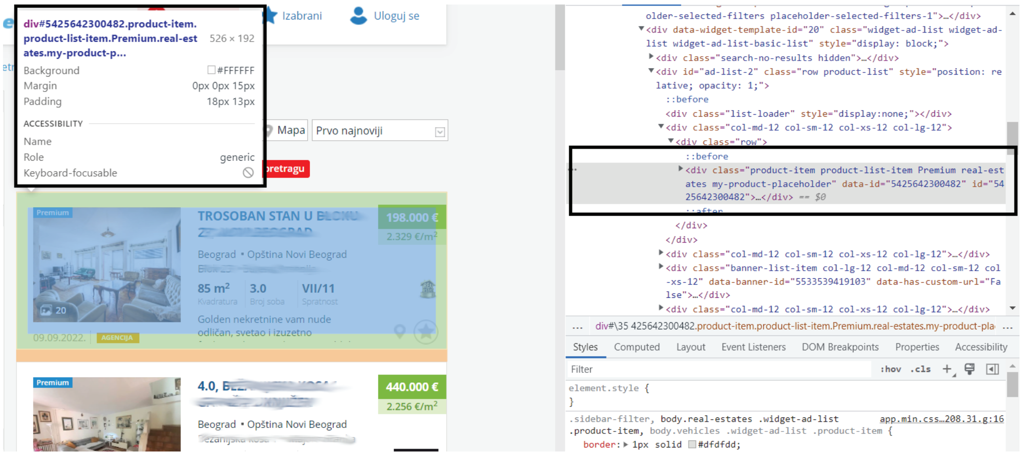
2. RELATED WORK

According to a vast amount of scientific papers published in the previous twenty years, anomaly detection has become one of the major issues in intrusion detection systems. These systems try to seek out deviations from normal behavior that indicate various attack scenarios (both ones on purpose and ones unintended), faults, system and network defects, resource overloading of any suspicious kind (such as providing huge traffic inside the company after business hours, for example), and other anomalous behavior, e.g., dropping a support for email or data service. This paper provides the reader with a brief review of possible (un)supervised learning algorithms for anomaly detection. The references cited within will cover some basic theoretical issues, guiding the researcher further [1]. Vignotto and Engelke [2] proposed two new algorithms for anomaly detection that rely on approximations from the theory of extreme values, which are more robust in such cases. Algorithms employ hypotheses that test points that are extremely far away from the points belonging to the training classes that are more likely to be anomalous objects. Although it may be correct in theory, in practice it depends on how clearly the normal behavior is absorbed – how sterile was the normal behavior training environment acquisition, i.e., how distant were the objects belonging to the normal class from anomalies, and how successful was the noise removal. However, the authors provided the results motivated by the univariate theory of extreme values that make aforementioned theory relatively precise. Rao et al. [3] demonstrated the effectiveness of classifiers in simulations and on real datasets. In this paper, the authors presented cascade grouping of K-means clustering and decision tree ID3 methods to seek out computer network anomalies. One of the IoT's main tasks is to seek out collected data alternations automatically, which includes data pattern deviation techniques. One specific set of algorithms, nowadays known as called deep learning (which actually originated some time ago, but there was no commercial hardware at that time to support it), can search for a specific relationship in billions of corporate IoT data, analyze them, classify them, and provide the right decisions [4]. Azavedo and Hoegner [5] examined the predictability of 299 capital market anomalies enhanced by 30 machine learning approaches and over 250 models in a dataset with more than 500 million firm-month anomaly observations.

3. MATERIALS AND METHODS

Downloading data from Internet ads is a process that involves the use of automata, or "robots" for extracting content from Web pages. Unlike the so-called "scraping" (Web scraping – downloading data from Internet services using automated scripts) of the screen that extracts the pixels of the image and saves them, Web scraping refers to downloading data from a web page and therefore also downloads data that has been generated from a database. All content is mapped from a database that is independent of the organization (relational or non-relational) to another database without the need for manual rewriting of data. Web scraping is used by various e-commerce services and product marketing sites that rely on permitted data collection from Internet sites that support Internet data scraping [6].

Web scraping is used illegally when it comes to stealing the copyright found on a page on the Internet. Also, it is used for the purposes of illegal price reductions on the stock market. Companies dealing with competitive businesses such as shopping facilities, supply chains, content distribution, and the like are particularly vulnerable.



div#5425642300482.product-item.product-list-item.Premium.real-estates.my-product-p... 526 × 192

Background □#FFFFFF
Margin 0px 0px 15px
Padding 18px 13px

ACCESSIBILITY
Name
Role generic
Keyboard-focusable ⃠

```
older-selected-filters placeholder-selected-filters-1">…</div>
▼<div data-widget-template-id="20" class="widget-ad-list widget-ad-
list widget-ad-list-basic-list" style="display: block;">
 ▶<div class="search-no-results hidden">…</div>
 ▼<div id="ad-list-2" class="row product-list" style="position: re
  lative; opacity: 1;">
    ::before
    <div class="list-loader" style="display:none;"></div>
   ▼<div class="col-md-12 col-sm-12 col-xs-12 col-lg-12">
     ▼<div class="row">
        ::before
       ▶<div class="product-item product-list-item Premium real-est
        ates my-product-placeholder" data-id="5425642300482" id="54
        25642300482">…</div> == $0
        ::after
      </div>
    </div>
   ▶<div class="col-md-12 col-sm-12 col-xs-12 col-lg-12">…</div>
   ▶<div class="banner-list-item col-lg-12 col-md-12 col-sm-12 col
    -xs-12" data-banner-id="5533539419103" data-has-custom-url="Fa
    lse">…</div>
   ▶<div class="col-md-12 col-sm-12 col-xs-12 col-lg-12">…</div>
…  div#\35 425642300482.product-item.product-list-item.Premium.real-estates.my-product-pla  …
Styles  Computed  Layout  Event Listeners  DOM Breakpoints  Properties  Accessibility
Filter                                                        :hov  .cls  +
element.style {
}
.sidebar-filter, body.real-estates .widget-ad-list       app.min.css…208.31.g:16
.product-item, body.vehicles .widget-ad-list .product-item {
  border: ▶ 1px solid □#dfdfdd;
```

Figure 1. *An example of an information system with HTML tags.*

There are three sets of data in the paper. The first set consists of 1067 data taken from the website. Each record has its name in text format, square footage as an integer, number of rooms as an integer, and apartment prices in decimal format.

This applies to other sets as well. The second one consists of 750 data taken from the website, while the third one consists of 1050 real data instances. Therefore, the data are not artificially generated but are real and taken from the system for selling apartments.

We have used the data from the Web sites of apartment sale or renting agencies with the sole purpose of preventing malicious financial gain for these agencies. Several Web sites have been scraped, and we have chosen three for this research. The main goal is to download the data on prices and potentially eliminate the competitors' adversarial behavior (leading to could-have-been-done social engineering-based attacks) in the labor market. The so-called attacks are taking place in an industry where price plays a major role in customers' decisions and where customers are spending money. Most often, these are travel agencies, computer equipment stores, etc. For example, a computer store monitors other competitors and analyzes the market. A permanent scraping system also tracks prices that change in real time. Then Internet browsers rank higher the Web sites that offer lower prices than other competitors.

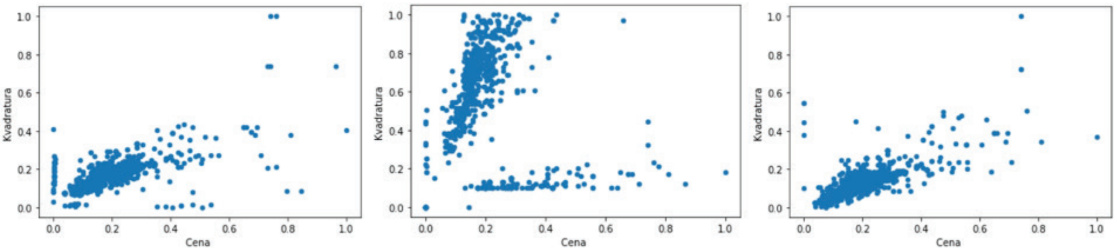


Figure 2. *Visual display of data from commercial Web systems. Label "Cena" on the x-axis on all three diagrams represents the price of the apartment, while label "Kvadratura" on the y-axis represents the size of the flat. Both images are exported from the Python programming language and are x- and y-axis scaled to the range [0, 1] with the purpose of better presentation to the reader.*

The methods that were used during the realization of the task are enlisted below:

- Angle-based Outlier Detector (ABOD) – a class for detecting deviations from the pattern based on the angle formed between different features. There are two versions of ABOD supported – fast, based on KNNs, and original one, with O(n^3) time complexity [7];
- Cluster-based Local Outlier Factor (CBLOF), which is also a clustering-based algorithm that measures the distance to the nearest large cluster center (normal data cluster) [8];
- Histogram-based outlier detection (HBOS) – with static bin numbers and the Birge-Rosenblatz automated bin number method [9];
- The Isolation Forest detection (IF);
- Local Correlation Integral (LOCI);
- Minimum covariance determinant (MCD);
- Multiple Objective Generative Adversarial Active Learning (MO-GAAL);
- Single-Objective Generative Adversarial Active Learning (SO-GAAL);
- Principal Component Analysis (PCA);
- Rotation-based Outlier Detector (ROD);

- Auto-encoder (AE) – a ANN for unsupervised useful data representations learning;
- Unsupervised anomaly detection using Empirical Cumulative Distribution Functions (ECOD), which is parameter-free;
- Isolation-based anomaly detection using Nearest-Neighbor Ensembles (INNE);
- Lightweight On-line Detector of Anomalies (LODA) [10].

Readers may consult [7–10] for details on the methods and algorithms employed, including some others. Acronyms will be used sub-sequentially.

4. RESULTS

According to the results obtained for the first data set, we can conclude that the most anomalies are concentrated around 102 to 109, so for this data set, the algorithms marked with a symbol † (in Tables 1–3) provide the best accuracy, and if we compare the raw data, we will see that they are significantly close to the accuracy.

Table 1. *Results obtained from the data in the first dataset (1067 in total).*

Algorithm	Anomalies	Regular data
ABOD	0	1067
CBLOF	53	1014
HBOS	50	1017
IF	54	1013
KNN	45	1022
Average KNN	34	1033
LOCI †	105	962
MCD †	109	958
MO-GAAL †	106	961
One-class SVM detector †	107	960
PCA †	107	960
ROD †	107	960
SP †	107	960
SO-GAAL †	107	960
AE †	107	960
CBLOF †	102	956
ECOD †	107	960
GMM †	105	962
INNE †	107	960
KDE †	107	960
Lightweight on-line detector of anomalies	82	985

According to the obtained results for the second data set, we can conclude that the most anomalies are concentrated around 72 to 75, so for this data set, the colored algorithms have the best accuracy, and if we compare the raw data, we will see that they are significantly close to the accuracy.

Table 2. *Results obtained from the data in the second dataset (750 in total).*

Algorithm	Anomaly	Regular data
ABOD	0	750
CBLOF	38	712
HBOS	34	716
IF	38	718
KNN	29	721
Average KNN	29	721
LOCI †	75	675
MCD †	74	676
MO-GAAL †	75	675
One-class SVM detector †	75	675
PCA †	72	678
ROD †	75	675
SP †	75	675
SO-GAAL †	75	675
AE †	74	676
CBLOF †	75	675
ECOD †	75	675
GMM †	75	675
INNE †	74	676
KDE †	75	675
Lightweight on-line detector of anomalies	61	689

According to the results obtained for the last data set, we can conclude that the most anomalies are concentrated around 97 to 105, so for this data set, the colored algorithms have the best accuracy, and if we compare the raw data, we will see that they are significantly close to accuracy.

Table 3. *Results obtained from the data in the third dataset (1050 in total).*

Algorithm	Anomaly	Regular data
ABOD	0	1050
CBLOF	51	999
HBOS	50	1000
IF	49	1001
KNN	47	1003
Average KNN	34	1006
LOCI †	105	945
MCD †	105	945
MO-GAAL †	104	946
One-class SVM detector †	104	946
PCA †	105	945
ROD †	104	946

SP †	104	946
SO-GAAL †	104	946
AE †	105	945
CBLOF †	95	953
ECOD †	105	945
GMM †	105	945
INNE †	105	945
KDE †	105	945
Lightweight on-line detector of anomalies	86	946

For the purpose of a brief visual presentation, we have put together a couple of graphics representing the results of LOCI, MCD, and MO-GAAL's results on all three datasets.

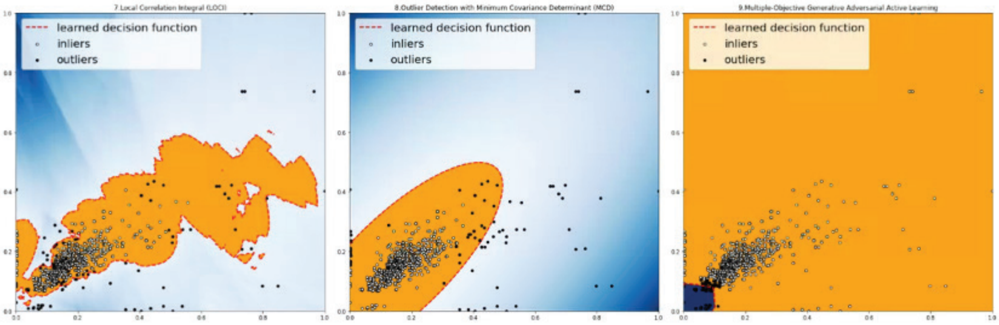


Figure 3. *LOCI, MCD, and MO-GAAL's results on the first dataset.*

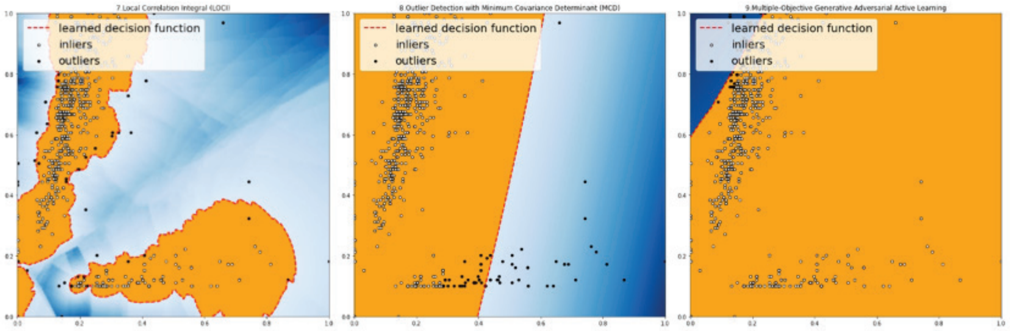


Figure 4. *LOCI, MCD, and MO-GAAL's results on the second dataset.*

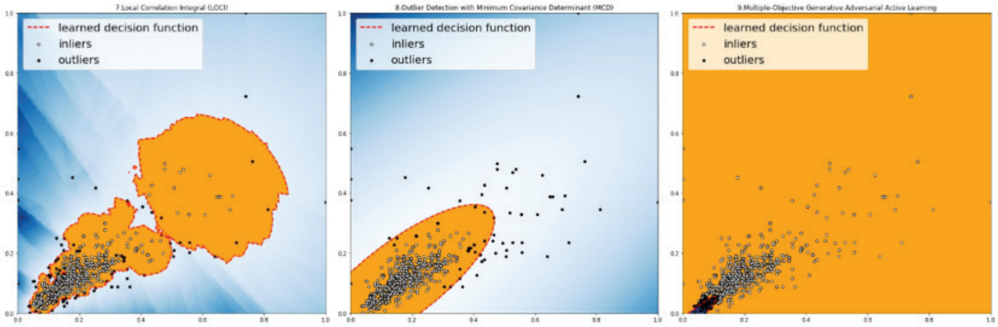


Figure 5. *LOCI, MCD, and MO-GAAL's results on the third dataset.*

7. DISCUSSION AND CONCLUSION

This paper presented the methods of collecting and visualizing the obtained data. Also, it was noted how the data was implemented in different ways within the Web pages. We chose the ratio of attributes between which the highest degree of data correlation occurred, as was the case with the ratio of apartment price and square footage. Using the principles of machine learning in anomaly theory and anomaly detection methods, we presented different types of anomalies and classifiers, as well as a graphical representation of the data. Data were successfully downloaded from three different sites for the purpose of better determination of anomalies. The idea was to prove that even if the data is different and downloaded from different websites, they have some similarities. We still performed the classification on the basis of two attributes (characteristics): of apartment price and square footage. Some anomalies are also obvious to humans, for example, there cannot be a physical quantity that is equal to 0, and its value is up to several thousand euros. However, not all anomalies are apparent in the system at first glance. Already, when we have more samples, we have a better overview of the exceptions in the data. A description of significant classifiers for anomaly detection and their implementation in different datasets is given. When detecting anomalies, exceptions and regular data are separately classified, and the number of them is stated. The number of exceptions is significant due to the assessment and evaluation of machine learning models. Systems that have approximately similar or the same solutions give us a better value for the model. We have over twenty different classifiers, and each of them performs with almost similar accuracy on all three datasets. Each of the data sets also has its own visual representation, which is very significant and differs from the data set and the use case of the algorithm. Furthermore, space is left for the implementation of such a system in business intelligence and systems that will automatically detect anomalies, exclude such anomalies, and upload classified, accurate data to Web services.

FUNDING:

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT:

Not applicable.

INFORMED CONSENT STATEMENT:

Not applicable.

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

REFERENCES

[1] S. Omar, A. Ngadi, and H. H. Jebur, “Machine Learning Techniques for Anomaly Detection: An Overview,” *International Journal of Computer Applications*, vol. 79, no. 2, Oct., pp. 33–41, 2013.

[2] E. Vignotto and S. Engelke, “Extreme Value Theory for Anomaly Detection – The GPD Classifier,” *Extremes*, vol. 23, no 4, pp. 501–520, 2020.

[3] K. H. Rao, G. Srinivas, A. Damodhar, and M. V. Krishna, “Implementation of Anomaly Detection Technique Using Machine Learning Algorithms,” *International Journal of Computer Science and Telecommunications*, vol. 2, no. 3, June, pp. 25–31, 2011.

[4] T. Çavdar, N. Ebrahimpour, M. T. Kakız, and F. B. Günay, “Decision-making for the Anomalies in IIoTs Based on 1D Convolutional Neural Networks and Dempster-Shafer Theory (DS-1DCNN),” *The Journal of Supercomputing*, vol. 79, no. 2, pp. 1683–1704, 2023.

[5] V. Azevedo and C. Hoegnerm, “Enhancing Stock Market Anomalies with Machine Learning,” *Review of Quantitative Finance and Accounting*, vol. 60, no. 1, pp. 195–230, 2023.

[6] R. Mitchell, *Web scraping with Python: Collecting More Data from The Modern Web*. O’Reilly Media, Inc, 2018.

[7] N. Silva, J. Soares, V. Shah, M. Y. Santos, and H. Rodrigues, “Anomaly Detection in Roads with a Data Mining Approach,” *Procedia Computer Science*, vol. 121, pp. 415–422, 2017.

[8] Y. Zhao, Z. Nasrullah, and Z. Li, “PyOD: A Python Toolbox for Scalable Outlier Detection,” *Journal of Machine Learning Research*, vol. 20, no. 96, pp. 1–7, 2019.

[9] M. Ahmed, N. Choudhury, and S. Uddin, “Anomaly Detection on Big Data in Financial Markets,” In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2017, pp. 998–1001.

[10] pyod 1.0.5 documentation. [Online] Available: https://pyod.readthedocs.io/. [Accessed: May 29, 2023].

Journal of
COMPUTER AND
FORENSIC SCIENCES
CFS

Instructions and Information for Authors

Thank you for considering to submit a manuscript to the Journal of Computer and Forensic Sciences. The points below provide general instructions and information for authors. If you have any questions, please contact us at **comput.forensic.sci@kpu.edu.rs**.

Submission Checklist

- Ensure that your manuscript fits the **Aims and Scope** of the Journal.
- Use the **Microsoft Word template** or the **LibreOffice template** to prepare your manuscript.
- Ensure that your manuscript *complies with our* **research** *and* **publishing ethics** guidelines.
- Ensure that all authors have signed the **Author Statement**.

Aims and Scope

Journal of Computer and Forensic Sciences covers advanced and innovative research across the fields of computer and forensic sciences. More information about the Aims and Scope is available **here**.

Open Access

The Journal of Computer and Forensic Sciences is an open access, peer-reviewed scientific journal. All accepted manuscripts are made freely and permanently available online immediately upon publication, without subscription charges.

No Article Processing Charges (APC)

The journal does not have *submission charges or article processing charges.*

Manuscript Types

The journal publishes:

- Original research papers – recent research results in computer and forensic sciences,
- Review articles (solicited reviews) – comprehensive and up-to-date systematic review of a specific area,
- Case reports – describing interesting and exceptional cases, providing new information to the readership.

Research Ethics

Manuscripts reporting on research involving human subjects, animals, cell lines, and plants will be scrutinized by the editorial office. Editors may ask the authors for documentary evidence or reject any submission that does not meet the research ethics requirements.

Please note the following research ethics guidelines:

- Research involving human subjects must be carried out following the rules of **the Declaration of Helsinki**[1] of 1975, revised in 2013.
- For research on animals, authors should ensure that their research complies with the principles of the 3Rs (i.e., **Replacement, Reduction and Refinement**[2]) which provide a framework for ethical decision making in the use of animals in research and teaching.
- For research involving cell lines, the origin of any cell line should be stated.

Publication Ethics

We adhere to the Core Practices and Guidelines of the **Committee on Publication Ethics**[3], and expect authors to comply with its best ethical publication practices. Plagiarism, data fabrication, and image manipulation are not acceptable. If evidence of misconduct is found, appropriate action will be taken to correct or retract the publication.

Manuscript Submission

The authors may submit a manuscript that has not been published before, that is not under consideration for publication elsewhere, and that has been approved by all co-authors. All manuscripts should be submitted through **the online submission system.**

Templates

Please use the **Microsoft Word template** or the **LibreOffice template** to prepare your manuscript.

Language

All manuscripts should be written in English. Please note the following:

- Our reviewers are advised to distinguish between the quality of writing and the quality of ideas. However, authors are strongly encouraged to carefully edit and proofread their manuscripts.
- All accepted manuscripts undergo professional English editing (free of charge), and proofreading by the authors.
- Authors are also strongly urged to avoid using language or examples that may be perceived as discriminatory.

Manuscript Length

Different types of manuscripts require more or less space. Therefore, we imply no restrictions on the length of manuscripts, provided that the text is concise and comprehensive.

Manuscript Structure

All manuscripts should consist of three main parts: the front matter, the main body, and the back matter.

The front matter should include:

- **Title**: The manuscript title should be specific and relevant.
- **Author list**, **affiliations,** and **email addresses**: At least one author should be named as the corresponding author.
- **Abstract**: The a*bstract should be a single paragraph and must not exceed* 200 words. It should express the purpose of the study, indicate the main methods applied, and summarize the main findings.
- **Keywords**: Three to five specific keywords should be added.

The main body structure depends on the type of manuscript.

- In original research articles, it should include the following sections: **Introduction**, **Materials and Methods**, **Results**, **Discussion**, and **Conclusions** (authors can make appropriate minor modifications to this section structure).
- In review articles, it should consist of literature review sections.
- In case studies, it should consist of sections describing and discussing the case study.

The back matter should include the following sections:

- **Funding: Authors should disclose a**ll sources of funding for their research. For research that did not receive external funding, please add “This research received no external funding”.
- **Acknowledgments (optional):** The authors may acknowledge any support that contributed to their manuscript, which is not included in the funding section.
- **Author Contributions (optional):** For manuscripts with several authors, their individual contributions can be specified.
- **Institutional Review Board Statement: For studies involving humans or animals, p**lease add “The study was conducted according to the guidelines of the Declaration of Helsinki” and a**dd the Institutional Review Board Statement and approval number. If ethical review and approval were waived, the authors are required to provide a detailed justification. For studies not involving humans or animals, please add “Not applicable”.**
- **Informed Consent Statement: For studies involving humans, p**lease add “Informed consent was obtained from all subjects involved in the study”**. If informed consent was waived, the authors are required to provide a detailed justification. For studies not involving humans, please add “Not applicable”.**
- **Conflicts of Interest:** Authors must disclose all conflicts of interest that may directly or potentially influence or impart bias on the work. Examples of potential conflicts of interest include but are not limited to: research grants, honoraria, financial support, employment, consultancies, affiliations, intellectual property rights, financial relationships, personal or professional relationships, and personal beliefs. If there is no conflict of interest, please add “The authors declare no conflict of interest.”
- **References:** The Reference section must provide a numbered list of references, as recommended by the **IEEE Citation Guidelines**[4]. The list is comprised of the sequential enumerated citations. A number enclosed in square brackets, placed in the text of the report, indicates the specific reference. Citations are numbered in the order in which they appear.

Figures, Tables, and Equations

- All figures and tables should be inserted into the main body of the manuscript (preferably close to their first citation) and numbered following their number of appearance.
- All figures and tables should have an explanatory caption.
- All figures should be at a sufficiently high resolution (i.e., 300 dpi or higher) and provided in a single zip archive. Preferable formats are TIFF, JPEG, and EPS.
- All equations should be numbered following their number of appearance.
- All equations should be editable by the editorial office (i.e., not provided in a picture format).

Citation Policy

If a manuscript includes material (e.g., figures, tables, text passages, etc.) taken from other sources, its source must be clearly cited. When appropriate, the authors should obtain permission from the copyright owner(s) and include evidence that such permission has been granted when submitting their manuscripts.

References cited in the text must appear in the References list, and vice versa. Personal communications and classical works are cited in text only and are not included in the References list.

Authors should not engage in citation manipulation, including but not limiting to excessive self-citation and "honorary" citations. Authors should not cite advertisements or advertorial material.

Editorial Procedures and Peer-Review

- **Initial checks:** All submitted manuscripts are first checked whether they fit the aims and scope of the Journal and meet its standards. At this stage, your manuscript may be rejected before peer-review or returned to the authors for revision and resubmission.
- **Peer review:** Once a manuscript passes the initial checks, it is assigned to at least two independent experts for peer-review. A double-blind review is applied. The guidelines for reviewers are available **here**.

Editorial decision and revision: The decision on a manuscript is one of the following:

- Accept in present form,
- Minor revision,
- Major revision,
- Reject.

Author appeals: In a response to the reviewers, the authors should address all reviewers' comments. The response should be organized by presenting reviewers' comments one by one, followed by the authors' response. Authors may appeal a rejection by sending an e-mail including a detailed justification to the Editorial Office.

Guidelines for Reviewers

Thank you for considering reviewing a manuscript for the Journal of Computer and Forensic Sciences. We rely upon the knowledge and commitment of our peer reviewers to ensure the academic integrity of our Journal. The points below provide general reviewing guidelines. If you have any questions, please contact us at **comput.forensic.sci@kpu.edu.rs**.

Before Reviewing

Before you accept our invitation to review a manuscript, please consider the following:

- **Timeliness:** Please try to submit your reviews on time. If you cannot meet a given deadline, please let the editor know.
- **Reviewer qualifications:** You have been invited to review the manuscript because the editor believes that your expertise covers the topic of the manuscript. However, if the manuscript is outside your expertise, you should decline to review it. If the manuscript is *generally within your expertise but you do not feel confident assessing certain parts of it, please notify the editor.*
- **Conflicts of interest:** You should disclose potential conflicts of interest. If you recognize the author's work, have a financial or commercial conflict of interest related to the reported results, or have strong feelings about a controversial question considered in the manuscript, you should disqualify yourself. If you are unsure whether you have a conflict of interest, discuss your concerns with the editor.
- **Confidentiality:** You should keep the content of the manuscript confidential. If you want to involve your students or postdocs in your review, you must obtain permission from the editor. If permitted, your assistants must be informed of the confidentiality requirement.
- **Anonymity:** Our journal operates double-blind peer review, which means that the reviewers and authors are unaware of each other's identities. You must not reveal your identity to the authors (e.g., in your comments, in metadata of submitted files, etc.).
- **Interactions:** There is no open interaction between reviewers and no public commenting during formal peer review. After you have submitted your report, you will have access to other reviewers' reports. We will also inform you on the final editorial decision of the paper. The reviewer reports, the author responses to reviews, and the editor decision letters are not published.
- **Language:** All review reports must be written in English.
- **Reviewer acknowledgment:** Once a year, we recognize our reviewers with annual listings in the journal. If you do not wish to have your name included in this list, please let us know.
- **Ethical guidelines:** Please note that all reviewers for our Journal are expected to follow **the COPE Ethical Guidelines for Peer Reviewers**[5].

Evaluating Manuscripts

Regarding your comments for authors

Start your review report by writing a paragraph or two in which you summarize the manuscript, emphasize its main contributions, and list its strengths and weaknesses. Then continue with the assessment of the individual sections of the manuscript. You should consider questions such as:

- Is the manuscript relevant for the field and suitable for the Journal?
- Is the research question original and well-defined?
- Is the manuscript clear and well-structured? Does it contain all of the sections you would expect?
- Are the cited references relevant and complete?
- Is the methodology clearly explained? Are the methods appropriately selected?
- Are the data underlying the research representative and balanced?
- Is the manuscript scientifically and technically sound? Is the experimental design appropriate? Are the reported results reproducible?
- Are the results analyzed and interpreted correctly? Are the conclusions supported by evidence? Is the manuscript statistically sound?
- Does the theory fit the data?
- Are the figures, tables, source codes, etc. appropriate?
- Is the manuscript of interest to the scientific community and the Journal's audience?
- Do you think that the reported results may advance the field?
- Is the English language of sufficient quality?

When preparing your review report, you should:

- **Ensure that your identity is not disclosed;**
- Be objective and constructive;
- Be detailed; your feedback should help the authors improve their manuscript;
- **Number each comment;**
- **Cite page numbers when referring to specific parts of the manuscript;**
- Scrutinize the manuscript, not the authors; avoid any derogatory personal comments or unfounded accusations;
- Make sure to distinguish between the quality of writing and the quality of ideas, especially for authors whose first language may not be English;
- Immediately report any suspected breaches of ethics, including scientific misconduct, fraud, and plagiarism.

Regarding your comments for editors

Your comments to the editors will not be revealed to the authors or other reviewers. These comments are optional. However, if provided, they should be consistent with your comments to the authors.

Regarding your final recommendation

To make a final recommendation on a manuscript, please choose one of the following options:

- **Accept in present form:** The manuscript fulfills all of the requirements described above, although some small fixes may be required (e.g., typos or grammatical corrections, etc.). No additional action by the review is required.
- **Minor revision:** The manuscript requires a small number of easily correctable errors or minor content correction or clarification. The article is in principle accepted after revision based on the reviewer's comments.
- **Major revision:** The manuscript offers relevance or value but contains significant deficiencies and requires a major rework. The acceptance of the manuscript depends on the revisions.
- **Reject:** The manuscript has serious flaws or does not offer relevance or value.

Your final recommendation should match your comments for the authors. Please note that the final recommendation will be visible to editors and other reviewers, but not to the authors.

Acknowledgment to Reviewers for this Issue

The editorial team of the Journal of Computer and Forensic Sciences wishes to thank the following reviewers, who have performed an essential role in ensuring the academic integrity of this publication:

- Dragiša Mišković, PhD, Institute for Artificial Intelligence Research and Development of Serbia;
- Prof. Srđan Savić, PhD, Faculty of Technical Sciences, University of Novi Sad, Serbia;
- Prof. Muzafer Saračević, PhD, Department of Computer Science, University of Novi Pazar, Serbia;
- Nenad Korolija, PhD, School of Electrical Engineering, University of Belgrade;
- Prof. Ivan Tot, PhD, University of Defence, Military Academy, Belgrade, Serbia;
- Prof. Ivan Pavkov, PhD, Alfa BK University, Belgrade;
- Prof. Dušan Joksimović, PhD, University of Criminal Investigation and Police Studies, Belgrade; and
- Prof. Vojkan Nikolić, PhD, University of Criminal Investigation and Police Studies, Belgrade.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

343.9

JOURNAL of Computer and Forensic Sciences / editor in chief Milan Gnjatović. - [Štampano izd.]. - Vol. 1, no. 1 (2022)- . - Belgrade : University of Criminal Investigation and Police Studies, 2022-
(Beograd : Službeni glasnik). - 30 cm

Polugodišnje. - Drugo izdanje na drugom medijumu:
Journal of Computer and Forensic Sciences (Online) = ISSN 2956-0799
ISSN 2956-087X = Journal of Computer and Forensic Sciences (Štampano izd.)
COBISS.SR-ID 85369353