UNIVERZA NA PRIMORSKEM

FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN INFORMACIJSKE TEHNOLOGIJE

DOKTORSKA DISERTACIJA
(DOCTORAL THESIS)

EMPIRIČNA ŠTUDIJA UPORABE TEHNOLOGIJE VERIŽENJA BLOKOV V OBSTOJEČIH SISTEMIH IN ARHITEKTURAH

(TRADEOFFS IN USING BLOCKCHAIN TECHNOLOGY FOR SECURITY, PRIVACY, AND DECENTRALIZATION:THEORETICAL AND EMPIRICAL PERSPECTIVES)

ALEKSANDAR TOŠIĆ

KOPER, 2022

UNIVERZA NA PRIMORSKEM

FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN INFORMACIJSKE TEHNOLOGIJE

DOKTORSKA DISERTACIJA
(DOCTORAL THESIS)

EMPIRIČNA ŠTUDIJA UPORABE TEHNOLOGIJE VERIŽENJA BLOKOV V OBSTOJEČIH SISTEMIH IN ARHITEKTURAH (TRADEOFFS IN USING BLOCKCHAIN TECHNOLOGY FOR SECURITY, PRIVACY, AND DECENTRALIZATION:THEORETICAL AND EMPIRICAL PERSPECTIVES)

ALEKSANDAR TOŠIĆ

KOPER, 2022

MENTOR: IZR. PROF. DR. JERNEJ VIČIČ

SOMENTOR: DOC. DR. MICHAEL DAVID BURNARD

Acknowledgements

I would like to thank my supervisors Assoc. Prof Jernej Vičič, and Assist. Prof. Michael Burnard for their continuous support and guidance during my PhD study. Special thanks to Tine Šukljan for helping me test, develop, and improve the countless ideas I continuously shared with him. His ability to quickly understand, simplify problems, and identify potential solutions were very influential in shaping my research. My thanks and appreciations also go to fellow co-workers for all the stimulating discussions that helped me overcome many obstacles.

This endeavor would not have been possible without the continuous support of my family. Their moral support was relentless despite the countless times my attention shifted away from them in favor of my study. Most importantly, they were the only constant motivation I had.

Special thanks to both institutions, University of Primorska Faculty of Mathematics, Natural Sciences and Information Technologies and Innorenew CoE for all the support.

Above all, I would like to thank the love of my life and mother of my children Olga. She was my muse, my inspiration, and a companion for more then a decade. I owe everything to her.

I dedicate this work to my son Ian, and my daughter Ema, hoping it will set a milestone for them to surpass.

Abstract

In these thesis we identify four selected topics in which blockchain technology can have a positive or transformative effect on existing solutions. We propose new protocols, which change the current standards to add functionality, improve performance or overcome limitations of existing blockchain networks. We identify four distinct topics and make contributions to each topic. Our results range from full protocol specification, and implementation, empirical study based simulations to data analysis.

On the topic of decentralized orchestration on the edge, we propose a blockchain protocol coupled with a full implementation and test results. In the protocol, we propose a change in block structure, which is more closely related to a decentralized state machine. The protocol has a unique lightweight consensus mechanism based on verifiable delay functions suitable for edge devices. The protocol implements a deterministic orchestrator responsible for migrating applications in an effort to balance resource consumption across the network.

On the topic of analysis, and anomaly detection in transaction networks, we test existing cryptocurrency transaction network for conformity to Benford's law. We establish that generally, such networks conform to Benford's law, and identify issues related to those that do not. We further show that the method can be reliably used on networks with a temporal component further extending it's usefulness.

On the topic of peer to peer gaming, we review existing state of the art protocols and identify the lack of Sybil resistance and collusion prevention. We propose a modified blockchain protocol that addresses these issues using a decentralized trusted source of randomness. We show how our protocol can use the existing cheat detection mechanisms and at the same time prevent collusion among players.

On the topic of privacy preservation in sensor networks, we propose a protocol for querying sensors in a privacy preserving way. The main contribution is the design of the framework, which transfers computation to the sensors instead of transferring data to a centralized entity. Using multilayered encryption, we show how both data and computation can be concealed from external adversary. We further improve on the security by decoupling the underlying sensor network from the users with a blockchain based role based access control.

Povzetek

V tej disertaciji identificiramo štiri izbrane teme, pri katerih lahko tehnologija veriženja blokov pozitivno ali transformativno vpliva na obstoječe rešitve. Predlagamo nove protokole, ki spreminjajo trenutne standarde, da bi dodali funkcionalnost, izboljšali zmogljivost ali presegli omejitve obstoječih omrežij blockchain. Identificiramo štiri različne teme in prispevamo k vsaki temi. Naši rezultati segajo od celotne specifikacije protokola in implementacije, simulacij na podlagi empiričnih študij do analize podatkov.

Na temo decentralizirane orkestracije na robu predlagamo protokol blockchain skupaj s popolno implementacijo in rezultati testiranja. V protokolu predlagamo spremembo bločne strukture, ki je tesneje povezana z decentraliziranim končnim avtomatom. Protokol ima edinstven nezahteven mehanizem soglasja, ki temelji na preverljivih funkcijah zakasnitve (Verifiable Delay Functions – VDF), primernih za robne naprave. Protokol implementira deterministič nega orkestratorja, ki je odgovoren za selitev aplikacij v prizadevanju za uravnoteženje porabe virov v omrežju.

Na temo analize in odkrivanja anomalij v transakcijskih omrežjih preizkušamo obstoječe transakcijsko omrežje kriptovalut glede skladnosti z Benfordovim zakonom. Ugotavljamo, da so takšna omrežja na splošno v skladu z Benfordovim zakonom, in identificiramo težave v omrežjih, ki niso skladna z BZ. Nadalje pokažemo, da je metodo mogoče zanesljivo uporabiti v omrežjih s časovno komponento, ki dodatno razširi njeno uporabnost.

Na temo peer to peer igranja iger pregledamo obstoječe najsodobnejše protokole in ugotovimo pomanjkanje odpornosti proti Sybil napadu in preprečevanje tajnega dogovarjanja. Predlagamo spremenjen protokol verige blokov, ki obravnava te težave z uporabo decentraliziranega vira naključnosti. Pokažemo, kako lahko naš protokol uporabi obstoječe mehanizme za odkrivanje goljufanja in hkrati prepreči tajno dogovarjanje med igralci.

Na temo ohranjanja zasebnosti v senzorskih omrežjih predlagamo protokol za poizvedovanje senzorjev na način, ki ohranja zasebnost. Glavni prispevek je zasnova ogrodja, ki prenaša izračun na senzorje namesto prenosa podatkov v centralizirano entiteto. Z uporabo večplastnega šifriranja pokažemo, kako je mogoče podatke in izračun prikriti

pred zunanjim nasprotnikom. Varnost dodatno izboljšujemo tako, da osnovno senzorsko omrežje ločimo od uporabnikov z nadzorom dostopa na podlagi vlog na podlagi verige blokov.

Contents

Chapter 1

Introduction

More than a decade has passed since the invention of the Bitcoin protocol, which at present, is the largest and most used P2P network in history. Since then, many new protocols were developed in an attempt to improve upon the Bitcoin protocol [15] or introduce new concepts, which leverage the unique properties of blockchain networks. Blockchain protocols are used to build networks in which nodes are assumed to not be honest. The nodes in the network keep a fully replicated local copy of a ledger stored as a chain of blocks. The chain is linked so that every block contains a hash of its predecessor. In order to add a block, nodes in the network form consensus on what the next block should be. From the initial Proof of Work, or Nakamoto consensus [15], in which nodes compete to solve a mathematical problem where the only way to solve it is to guess the result, thereby proving the computation was done. Other consensus mechanisms have been proposed such as Proof of Stake(PoS) [17], Proof of Authority(PoA), etc. Consensus mechanisms make trade-offs between speed, security, and decentralization commonly referred as the blockchain trilemma, which is analogue to the CAP theorem [8]. The resulting system can be viewed as an immutable ledger that is both transparent, and verifiable. Perhaps the biggest achievement of blockchains is that the protocol makes no assumption about the honesty of participating nodes. As such, it can be viewed as a trust machine used by potentially untrustworthy parties to transact in an open, verifiable and transparent way.

The most successful application of blockchains are arguably cryptocurrencies with Bitcoin leading as the largest, and most used peer to peer(P2P) network in history [16]. This thesis does not shy away from the fact all successful implementations make use of a native currency by which protocols can be secured through economic incentives. Instead it attempts to see through the controversy by providing tools, improvements, and use-cases where blockchains are applicable.

Applying blockchain technology to other systems and developing new concepts around blockchains has produced a plethora of innovations and growth in research

activities in the area. Fields such as healthcare [1], supply chain management [10, 12], and fin-tech [7] have received a lot of attention due to their palpable potential benefits. Our research widens the search for possible benefits of blockchain technology in systems where their benefit is less perceptible.

We identify four topics in which blockchain technology can provide useful effects, improvements, or systemic transformations of existing solutions. While the fields of application may be distinct, they are joined by contributions to the use and adoption of blockchain technology.

The overall objective of this study is to investigate the suitability of blockchain networks to resolve problems and improve system performance in several domains.

1.1 Towards decentralized edge computing

In recent years, cloud computing became a commonly used architecture for most applications. The shift of the geography of computation was incentivized by many factors ranging from ease of software maintenance [4], reliable quality of service(QoS) [13], hardware flexibility, and cost (CapEx to OpEx) [2], etc However, with the expected growth of data generation and consumption and storage and service provisioning in cloud computing environments, the architecture is pushing network bandwidth requirements to the limit [19]. Edge computing in it's simplest form can be defined as an architecture in which computation is moved to the edge of the network in order to make use of the geographic proximity to decrease latency and improve bandwidth. This recent paradigm shift attempts to address the overly geographically-centralized cloud architecture. However, distributing services to the edge introduces new challenges such as resource allocation, service and application migration, trust, etc.. Blockchain technology may be used to address some of the issues. It can serve as a layer of trust between the system, and the end user by providing a verifiable and transparent ledger of the state of the system. To achieve this, a new protocol is required that would overcome the latency constraint, decentralized resource allocation, and real-time container migrations [24].

1.2 Towards a robust analysis of cryptocurrency transaction networks

Since the inception of Bitcoin, many alternative systems were developed. Some remain blockchain based, where transactions are stored and consequently timestamped in blocks to create a canonical chain through consensus. Others employ a directed

acyclic graph based data structures, where there is no single canonical chain. Instead, transactions reference and confirm previous transactions in order to increase the system's throughput by sacrificing some security features. Moreover, the transaction structure can be changed to achieve privacy, i.e., by using ring signatures in Monero [18]. Forum [14] predicts 10% of the global domestic product to be stored on blockchain based public ledgers by year 2025. The growing interest inspired many developers, researchers, and innovators to dedicate their time in an effort to improve existing systems. The effects can be observed through the thousands of cryptocurrencies, and networks that exist presently. The growing velocity of these networks further increases the risk for regulators to protect the consumer, and the stability of the financial system. Assuming frauds grow in parallel with the velocity and total value locked in the underlying network, a method for fast, and efficient anomaly detection is paramount. However, with the growth of innovation in this space, the techniques employed must search for a generic solution that makes little or no assumptions about the underlying network.

There are clear benefits in providing a technology agnostic tool to analyse open ledgers to raise alarms about suspicious behaviour which requires further, more fine-grained analysis. Although more than a decade has passed from the first transaction of the first cryptocurrency - Bitcoin (BTC) [15], only the last few years have seen a large enough number of transactions over a long enough time frame that some statistical analysis can reliably be carried out. The potential of using Benford's law [5], a law of anomalous numbers in a non-altered form for discovering fraudulent, or at least suspicious, activity on cryptocurrencies in the same way it is used in standard financial forensics could be beneficial for the ecosystem [25]. Many networks including cryptocurrency transaction networks include a temporal component. The applicability of Benford's law on temporal data further extends it's usefulness in detecting anomalies [23].

1.3 Towards decentralized multiplayer game architectures

The gaming industry was estimated to be worth nearly 135 billion in 2019 with an estimated growth of 10% per year [9]. The recent trends toward multiplayer games have been very successful with games like Fortnight earning more than 2.4 billion in revenue in 2018 alone [20]. Steam, the biggest game distribution platform reported it serves as many as 18.5 million clients concurrently. This scale of demand requires cloud computing enabled servers need to be migratable in real time. Additionally, network latency

was reduced due to localisation approaches where servers are spawned geographically close to clients if possible. However, maintaining a player base of thousands or even millions together with the hardware and software infrastructure is both very expensive and difficult to maintain [26]. The recent idea of a "sharing economy" can be applied in tandem with the paradigm shift to edge computing. More specifically, clients on the edge of the system can profit from sharing resources, such as bandwidth and computing power, thereby releasing the burden on centralized servers.

This can be achieved by using a peer to peer (P2P) architecture. P2P gaming architectures have been studied extensively but have not been widely adopted [26]. The main issues are closely related to the lack of authority and trust. Centralized architectures solve these issues with authoritative servers. The server's tasks are to simulate gameplay, validate and resolve conflict in the simulation, and store the game state. P2P multiplayer architectures were previously able to address some of the cheating vectors but required some level of centralization. Recent developments in blockchain protocols could address the aforementioned drawbacks by circumventing known cheats while maintaining high decentralization [22].

1.4 Towards trustless and privacy preserving indoor location (WSNs)

An indoor location system may be one of many location aware applications in fields of medicine, robotics, industrial optimisation, psychology, security, etc.. Most current solutions require knowledge about the position of occupants within the building at any given time. Existing approaches to on-site location data collection suffer from both usability issues and technological obstacles. Typical implementations include but are not limited to wearable devices [3] (i.e., location aware bracelets) that can be discarded by unaware users, or require frequent battery charging, on-site support, and maintenance. Sensor networks that do not rely on wearable devices usually include cameras and microphones coupled with automatic face detection software that have a psychological impact on occupants and raise privacy concerns.

Even though encryption effectively provides data privacy, monitoring indoor activities by relying on wireless IoT devices could disclose contextual information on data transmission [11], not only posing risks to the privacy of individuals, but also compromising building security.

A privacy-aware IoT and blockchain-based indoor location solution is particularly suitable for application in medical facilities, public buildings, and residential homes as a framework for privacy-aware indoor location monitoring. It could be applied for

structural health monitoring, studying behavioral patterns of a building's occupants and health-related issues such as locating lost patients with memory and orientation disorders, fall detection, and also identifying violations of social distancing, counting the number of persons in a room, and determining when and which room needs surface disinfection due to over-utilization, etc. To unlock the full potential of indoor location systems, a framework combining hardware and software for privacy preserving computation is required [21].

1.5 Research Questions and Hypothesis

RQ-1: Is it possible to implement a decentralized orchestrator for real time application migration?
H-1: The joint use of experimental Checkpoint/Restore In Userspace (CRIU) and blockchain with a scalable consensus protocol can be used to implement decentralized orchestration without a single point of failure (SPOF).
RQ-2: Can Benford's law be applied to fraud detection in cryptocurrency transaction networks?
H-2: Cryptocurrency transaction networks conform to Benford's law and can be used to detect potential anomalous behaviour.
RQ-3: Is it possible to address Sybil resistance in existing P2P multiplayer game architectures?
H-3: A blockchain protocol with secure decentralized randomness can be used to assign players to games, and store states. A randomly selected quorum of referees can provide sufficient resistance to Sybil based cheats.
RQ-4: Can efficient indoor location-based computation be achieved in a decentralized way without violating privacy of data and computation?
H-4: A sensor network for indoor location can be queried by wrapping computation tasks in multi-layered encryption eliminating the need for sensitive data transfer. To preserve the privacy of sensors, blockchain based smart contracts can perform the wrapping of computation tasks, thus decoupling users from the underlying sensor network.

Chapter 2

Published Papers

2.1 Paper 1

Title: A Blockchain-based Edge Computing Architecture for the Internet of Things
Authors: Aleksandar Tošić, Jernej Vičič, Michael David Burnard, Michael Mrissa
Year: 2022
Journal: Sensors (pending peer review)
DOI: 10.20944/preprints202111.0489.v1
Link:https://www.preprints.org/manuscript/202111.0489/v1

MDPI

Article

A Blockchain Protocol for Real-time Application Migration on the Edge

Aleksandar Tošić [1,†,‡] [3,*], Jernej Vičič [1,‡] [2], Michael Burnard [4,5], Michael Mrissa [6,7]

[1] University of Primorska Faculty of Mathematics, Natural Sciences and Information Technologies; aleksandar.tosic@upr.si
[2] University of Primorska Faculty of Mathematics, Natural Sciences and Information Technologies; jernej.vicic@upr.si
[3] InnoRenew CoE; Livade 6 6310 Izola, Slovenia; aleksandar.tosic@innorenew.eu
[4] InnoRenew CoE; Livade 6 6310 Izola, Slovenia; mike.burnard@innorenew.eu
[5] Institute Andrej Marušič; Muzejski trg 2, 6000 Koper, Slovenia; mike.burnard@iam.upr.si
[6] InnoRenew CoE; Livade 6 6310 Izola, Slovenia; michael.mrissa@innorenew.eu
[7] University of Primorska Faculty of Mathematics, Natural Sciences and Information Technologies; michael.mrissa@upr.si
* Correspondence: aleksandar.tosic@upr.si;
† Current address: Glagoljaška 8, 6000 Koper, Slovenia
‡ These authors contributed equally to this work.

Version August 11, 2022 submitted to Journal Not Specified

Abstract: The Internet of Things (IoT) is experiencing widespread adoption across industry sectors ranging from supply chain management to smart cities, buildings, and health monitoring. However, most software architectures for IoT deployment rely on centralized cloud computing infrastructures to provide storage and computing power, as cloud providers have high economic incentives to organize their infrastructure into clusters. Despite these incentives, there has been a recent shift from centralized to decentralized architecture that harnesses the potential of edge devices, reduces network latency, and lowers infrastructure cost to support IoT applications. This shift has resulted in new edge computing architectures, but many still rely on centralized solutions for managing applications. A truly decentralized approach would offer interesting properties required for IoT use cases. In this paper, we introduce a decentralized architecture tailored for large scale deployments of peer-to-peer IoT sensor networks and capable of run-time application migration. We propose a leader election consensus protocol for permissioned distributed networks that only requires one series of messages in order to commit to a change. The solution combines a blockchain consensus protocol using Verifiable Delay Functions (VDF) used for decentralized randomness, fault tolerance, transparency, and no single point of failure. We validate our solution by testing, and analyzing the performance of our reference implementation. Our results show that nodes are able to reach consensus consistently, and the Verifiable Delay Function proofs can be used as an entropy pool for decentralized randomness. We show our system can perform autonomous real-time application migrations. Finally, we conclude that the implementation is scalable by testing it on 100 consensus nodes running 200 applications.

Keywords: Fault Tolerance; Blockchain; Internet of Things; Edge Computing; Peer-to-Peer; Decentralized; Sensor Networks; Verifiable Delay Functions

1. Introduction

Cloud computing solutions have driven the centralization of computing, process control (e.g., business information, manufacturing, distributed systems, IoT management), and data storage to data centres. Existing cloud-based solutions have few incentives, aside from reducing network latency,

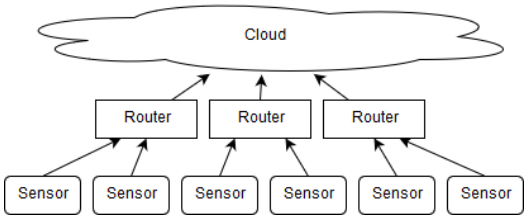

Figure 1. Standard sensor network architecture.

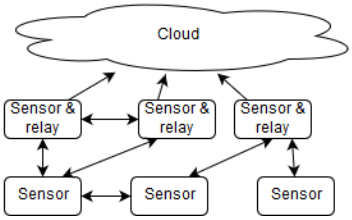

Figure 2. Mesh sensor network architecture.

to distribute computing and storage resources. There are many reasons why centralization is more appealing. These range from legislative reasons, tax policies, availability and affordability of high speed internet and electrical power, reduction of maintenance costs, and even climate preservation [1]. However, cloud computing solutions are struggling to address the specific challenges of emerging IoT, and edge computing use cases.

The ever-growing number of devices on the edge causes scalability challenges for centralized architectures such as cloud-based ones. Edge devices tend to be heterogeneous and existing IoT platforms remain isolated and unable fully exploit their potential. Moreover, these devices have considerable computing resources, which for the most part remain underutilized as most applications perform computation on the cloud. A major challenge in this area is supporting homogeneous usage of edge devices, which requires applications to migrate at run-time from an overloaded device to a more available one. Currently, there is no standardized platform for general purpose computing that supports such run-time application migration. Another limitation to large scale deployment of sensor networks is the infrastructural investment needed to support the network, as typical architectures require a middle layer infrastructure that enables access to the cloud (Fig. 1 and [2]).

We believe these challenges can be overcome as recent technological advances have provided partial solutions and have presented new opportunities. These advances have paved the way for the recent paradigm shift from centralized to decentralized architectures for IoT [3]. First, as edge devices are becoming more powerful and capable of running complex software, they provide a huge pool of available, yet underutilized, computing resources. Second, containerization solutions (as opposed to virtualization) have been gaining momentum to overcome heterogeneity problems while preserving acceptable performance. Containerization software (e.g., Docker) provides software abstraction that enables general purpose computing on edge devices. Third, with the growth of edge devices capable of direct wireless communication, a mesh network approach has become worth exploring as a solution to reduce or eliminate the middle layer infrastructure needed for devices to connect to each other and the cloud (Fig. 2).

The need for edge computing is well illustrated with scenarios related to ad-hoc networks [4], and especially with peer-to-peer wireless sensor networks. The paradigm shift towards decentralization is relevant to numerous application domains such as smart building monitoring, structural health monitoring, self-driving vehicles, micro service architectures, mobile devices, etc.. In our work, we have experimented with a cultural heritage building located in Bled, Slovenia. We deployed several sensors to monitor the building state for maintenance purposes and air quality to provide safety for visitors. In the case where buildings are located in remote areas, as in our use case, edge devices

must self-regulate and optimize their behaviour at run-time. They also must have the capacity to scale up as the number of devices grows (scalability), to adjust when dysfunctions occur, for example when devices leave the network (experience byzantine behaviour), and the operation of all devices should be recorded safely for later analysis (transparency). In a cloud-based environment, edge devices send data to the cloud where computation occurs. However, issues such as poor network coverage, frequent disconnection, cost of infrastructural investment, inadequate dependability, and security concerns remain unaddressed [5], [6], and [7]. Edge computing solutions attempt to reduce network latency, increase fault tolerance, dependability and security, and reduce the cost of infrastructural investment needed to provide network coverage. They also operate independently of an external network connection.

To address these issues, we propose an architecture based on an innovative combination of existing technologies. Specifically, our architecture provides a general purpose computation model allowing large scale sensor networks to distribute the computational load among edge devices (sensors, controllers, etc.). Using containerization, applications can be built using any programming language or stack, containerization also serves as an abstraction layer between the application requirements, and the host hardware. The decision making process for resource allocation is made by a decentralized orchestrator implemented as a consensus protocol that outputs a migration strategy, which is in turn stored on the blockchain [1]. It features high fault tolerance, full transparency, reduced network infrastructure cost, and no single point of failure. The network layer uses decentralized randomness to constantly change the network topology to allow efficient propagation of information pertaining to resource utilization of nodes.

The rest of this paper is structured as follows: Section 2 provides the necessary background knowledge and overviews the most relevant related works to highlight the originality of our proposal. Section 3 details our architecture and its operation. Section 4 describes our evaluation environment Section 5 summarizes the results and gives guidelines for future work.

2. Background knowledge and related work

The most critical unmet monitoring challenges according to [8] are: mobility management, scalability and resource availability at the edge of the network, coordinated decentralization, interoperability and avoiding vendor lock-in, optimal resource scheduling among edge nodes, and fault tolerance. No widely-used cloud monitoring tool for edge computing fully addresses these challenges. some requirement remain unmet by any existing solution, as many system aspects including container, end-to-end network quality are not adequately addressed [8]. The EU project RECAP [9] presents a vision of the next generation of intelligent, self-managed, and self-remediated cloud computing systems (i.e., a system that can monitor and relocate resources to achieve Quality of Service - QoS). The project also describes models intended to be integrated in network topology-aware application orchestration and resource management systems from an edge computing perspective [10]. Another solution, AutoMigrate [11], incorporates a selection algorithm to determine which services should be migrated to optimize availability. Although this system has solutions for most of the problems we address, it does not resolve the Single Point Of Failure (SPOF) issue because it relies on a central service to orchestrate migrations. Our decentralized implementation eliminates the SPOF issue.

2.1. Orchestration solutions for edge computing

By definition, orchestration denotes control by a single entity over many. This differs from choreography, which is more collaborative and allows each involved party to describe its part in the interaction [12]. We have identified the most successful orchestration solution to be Kubernetes [13],

[1] A blockchain is a growing list of records called blocks, linked together using cryptography, and the nodes follow a shared consensus protocol to validate new blocks.

the most used and most feature-rich orchestration tool [14], Docker Swarm[2], Amazon Web Service Elastic Container Service (AWS ECS) [15], the Distributed Cloud Operating System[3], and Nomad[4].

The Decenter EU project[5] proposes decentralized orchestration technologies for fog-to-edge computing. Although the project does support decentralized orchestration between multiple domains and records service level agreements and violations to the blockchain, the solution is designed as a federated approach where a multi-domain orchestrator overviews several domains, that in turn are driven from local orchestrators [16]. The project also implements a blockchain to act as a brokerage platform where smart contracts guarantee resource sharing across domains [17]. In contrast to a federated approach, our implementation is fully decentralized with a randomly selected orchestrator at each interval, thus avoiding the SPOF problem and not relying on a trusted third party.

All of the architectures discussed above have a common flaw: the SPOF problem. In each case, the flaw is characterised by a single orchestration entity. Most solutions also lack support for edge devices. Our proposed solution addresses these shortcomings, while providing full transparency, variability of the system, completely decentralized operation backed with a strongly secure, scalable, and efficient consensus mechanism.

Recently, a decentralized protocol for orchestration of containers named Caravela was proposed [18]. The solution relies on a Chord for resource discovery, and employs a volunteer system in which nodes are categorized as suppliers (supplying resources), buyers(searching for resources), and traders(mediating supply/search for offers). The authors show that their solution can scale using a random migration algorithm, but fails fulfill deployment requests. It also is not able to fulfill the global binpack scheduling policy due to a lack of global shared state.

2.2. Container platforms

We are using containers as a primary execution environment. Containers, as used in this paper, are a group of namespaced processes run within an operating system. Docker is the most widely used platform according to [19] and one of the few platforms that can migrate apps at run-time and enable easy communication. For these reasons it was used as the main testing platform.

2.3. Available blockchain solutions

The proposed solution makes use of a blockchain to store the state transitions of the network in a verifiable, and transparent way. Unlike existing blockchains which either use an account based model [20], or an UTXO model [21], our blocks do not store transactions or account states. The block structure is tailored to accommodate application migration and verifiability of migrations. Hence, the blocks are snapshots of the state of the system containing information about available resource, and required resources of applications managed by the system.

A survey of the most notable readily available blockchain solutions for private network yielded three candidates:

- Implementation of a private Ethereum network, although the implementation is fairly simple [22], the available consensus mechanisms include PoW, which is not secure for networks with no value, and proof of authority (PoA), which limits the consensus nodes to a subset of trusted nodes thereby decreasing decentralization, and security.
- Implementation of a HyperLedger blockchain in all configurations requires notable CPU burdens [23]. As the number of nodes in the network grows, the system requirements scale far beyond what can be considered sustainable for edge devices.

---

2 https://github.com/docker/swarm
3 https://dcos.io/
4 https://www.hashicorp.com/products/nomad
5 Decenter project homepage: https://www.decenter-project.eu

- Multichain [6] also presents a viable alternative for a private blockchain network [24], again not suitable for edge devices [24]. Moreover, it is primarily focused on facilitating transactions of cryptocurrency, and assets.
- Solana [25] similarly uses verifiable delay functions as a source of entropy for their leader rotation algorithm. However, their VDF implementation requires thousands of graphical processing units to meet the speed requirements, which is not suitable for edge devices.

All the presented available off-the-shelf solutions satisfy most of the criteria posed by the research experiment, but they all rely heavily on the computation power which makes them unsuitable for edge devices. Further, the required block structure and changes on the protocol would outweigh the benefits and accumulate technical debt.

2.4. Decentralized self-managing IoT architectures

A survey of the scientific literature shows multiple solutions that address decentralized self-managing architectures for the IoT. The most notable examples are:

- Maior et al. [26] present a theoretical description of a decentralized solution for energy management in IoT architectures. The solution is aimed at smart power grids. They present 4 algorithms with analyses of correctness in order to describe the behavior of self-governing objects.
- Higgins et al. [27] propose a distributed IoT approach for electrical power demand management.
- Suzdalenko and Galkin [28] extend the approach by Higgins et al. [27] by allowing users to individually join, and depart the environment at run-time.
- Niyato et al. [29] propose a system that addresses home energy management system wheres devices communicate directly among themselves.
- dSUMO [30] address the synchronization bottleneck by proposing a distributed and decentralized microscopic simulation (the focus is on data throughput and not fault tolerance; throughput is increased using a decentralised setting).
- Al-Madani et al. [31] address indoor localization utilizing Wireless Sensor Networks (WSNs) relaying on publish/subscribe messaging model. The results show that the Really Simple Syndication (RSS) [32] format achieves acceptable accuracy for multiple types of applications.

Our proposed solution differs from the previous contributions in two ways.

- other solutions typically focus on a single problem presenting an optimal solution for it, we argue that an IoT architecture requires multiple optimisation criteria. We consider multiple criteria and include a framework to add more criteria in the future.
- our protocol is highly decentralized as it allows all nodes to participate in the consensus, while maintaining low hardware requirements fit for edge devices

A related approach by Samaniego and Deters [33] suggests using virtual resources in combination with a permission-based blockchain for provisioning IoT services on edge hosts. They use blockchain to manage permissions only, and therefore provide security using blockchain. In contrast, our approach uses blockchain to store all information about service choreography which makes it verifiable over time, while still providing security.

The main contribution of this paper is a light-weight blockchain protocols, which can achieve high decentralization and low hardware requirements typically found in edge devices. The proposed protocol inherits ideas from Ethereum 2.0 but replacing the source of entropy needed for consensus with a VDF function. Moreover, the structure of the block carries the state transition information, and unlike existing blockchains does not have the concept of account, balances, and transactions.

---

[6] MultiChain Open source blockchain platform: https://www.multichain.com/

3. Proposed decentralised architecture

In this section, we provide a general description of our architecture [34] and highlight its main components. The main purpose of our architecture is to enable verifiable and decentralized management of applications on the edge. In our vision, applications can be built as containers, and submitted to the network by reaching any node via an API. We use containerization to decouple the host running the application from the application and address the issue of hardware and software heterogeneity. This allows the protocol to assume an application can be run on all nodes in the network. A randomly selected and decentralized orchestrator on the network would then be able to choreograph the execution of the application, and migrate applications between hosts at run-time. As our architecture is fully decentralized, each node is locally driven by a protocol that participates in establishing the global state of the network via a specially built consensus mechanism. Nodes in the network reach consensus on a migration plan in an effort to improve the resource allocation of running applications. A migration plan is viewed as a state transition, which is stored on the blockchain formed by the participating nodes. We implement a choreographed solution, which is a collaborative, rather than a directed approach (as opposed to orchestration). Choreographed systems define a way for each member to describe its role in the interaction [12]. This collaborative approach avoids the SPOF problem. Despite this advantage, there are no choreography solutions known to address the open problems described in the start of the Section 2.

To provide a global understanding of our fully decentralized architecture, we first describe the architecture of a single node, followed by the interaction protocols between nodes.

All identified orchestration solutions presented in Section 2.1 rely on a primary/replica model. The main service selects the applications that need to be reallocated according to a selected optimization algorithm.

Our migration algorithm is able to:

- pause a container,
- transfer the context to a different host,
- resume the execution given the context.

Additionally, we implemented migrations using checkpoint/restore in userspace, or CRIU, an experimental feature available in Docker [35].

3.1. Overview of node architecture

The node application that we developed is containerized in Docker. As shown in Fig. 3, the internal architecture of a node is composed of the following modules that support application management:

- Networking layer: this layer deals with network communication through deployed APIs.
- Gossip protocol: A rendezvous based gossip protocol is used to build a distributed hash table (DHT) that maps public IP's of nodes to their network address. The messages are encoded using protocol buffers[7], it is the underlying protocol that makes sure all messages reach all nodes in the network while minimising network usage.
- Block propagation protocol: relies on the gossip protocol to spread newly accepted blocks over the network.
- Resource propagation protocol: relies on the network layer to deliver the state of resources (currently CPU, RAM, disk usage, and network utilization of Docker containers) over the network to the receiving node.
- Consensus protocol: ensures all nodes reach consensus in a decentralized way (presented below).

[7] https://developers.google.com/protocol-buffers

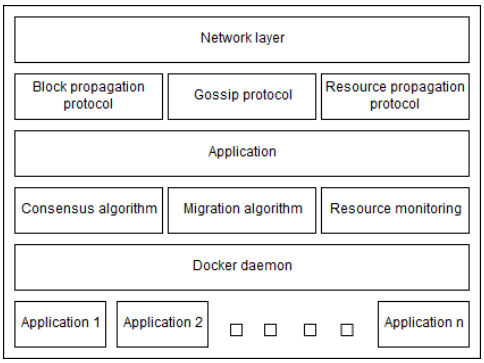

Figure 3. General overview of a node architecture.

- Migration algorithm: guarantees that a migration strategy is reached whenever needed thanks to a deterministic algorithm. This algorithm is executed at each slot until the proposed block is accepted and finalized. The output of the algorithm is included in the block to construct a verifiable, and transparent log of each application's life-cycle.
- Docker daemon: hosts applications and is used for abstracting the underlying heterogeneity between devices, systems, and applications. It provides support to our solution via Docker APIs.
- Resource monitoring: relies on the Docker API to monitor the state, and resource allocations of the hosting device and the applications running on it.

3.2. Storing system states in a blockchain

All nodes share information about their states through a federated type topology obtained by distributed clustering of nodes explained in more detail in Section 3.3. We define a state as a matrix of vectors describing resource consumption associated with each application. A *resource pool* data structure is replicated in all nodes and contains information about all node states. In our use case, we define a vector with the following values:

$$\{app, cpu, ram, disk, network, timestamp\}$$

This provides a time series of system resource utilisation for each application across an operating period. The resources used by applications are obtained through the Docker API and represented in percentages for simplicity. Taken at a specific time interval, the vector is a block that includes a list of per-application resource statistics, as shown in Table 1 where *Node* is a 256bit hash representing the system wide unique ID of the application, RAM, DISK, and CPU are floats representing the portion of node's available resources used by the application. Finally, the average latency is computed as the 30 second moving average of Round-trip delay (RTT) towards randomly selected validators.

Table 1. An example of a data block.

V	Node	RAM	DISK	CPU	Average Latency
v_0	A	50%	23%	90%	23ms
v_1	B	47%	87%	23%	33ms
v_2	C	12%	25%	15%	51ms
v_3	A	35%	14%	56%	101ms
v_4	D	25%	74%	16%	9ms

From a data block, it is then possible to compute a migration plan to optimize the allocation of applications to nodes according to the resource states of all nodes. The migration plan is also included

in the block, which produces a transparent computational log for verifying if the adopted migration plan was actually efficient and fair. The architecture does not enforce any specific migration algorithm. The only constraints are that the algorithm must be deterministic and must rely only on data included in the block (reached by consensus). For the same inputs to a deterministic algorithm, proposed migrations can be verified much like transactions are verified in public blockchains.

In order to provide liveliness and responsiveness, delivering resource consumption statistics to the block producer must be faster then $\frac{2}{3} * slotTime$. Using the gossip protocol produces unwanted latency, and greatly increases resource utilization maintaining the message queue (MQ). To overcome this, we implement a distributed k-means clustering algorithm that requires no communication between nodes to compute. Clustering is used to group nodes in a separate overlay network where statistics are propagated using UDP protocol. The seed used to compute k-means is shared by all nodes, the VDF proof. Cluster representatives are nodes that are responsible for requesting resource utilization statistics from their members, and transmitting them to the block producer. The timing details are strongly intertwined with the synchronicity of the consensus algorithm further explained in chapter 3.4.

This federated overlay topology greatly decreases decentralization and consequently fault tolerance in case a cluster representative exhibits byzantine behaviour. However, this is not concerning considering that a failure to disseminate resource utilization only delays potential migrations for a subset of applications in the current block. Once a new block is accepted, a new overlay topology is computed. Eventually a node experiencing byzantine behaviour is excluded from the validator set as detailed in chapter 3.4.

3.3. Migration algorithm and verifiability

To forge a block, nodes compute a migration plan based on resource statistics in the previous block. The migration plan is executed once the proposed block is accepted. Application migration is realized using Docker commands to pause the application, compress it, and transfer it to the destination node where it is restored. Alternatively, using CRIU, only the state of the running container is extracted, and migrated. All migration plans are securely stored in the blockchain for eventual verification. The time to produce a block is configurable, and largely depends on the requirements for responsiveness, and resource availability, and network size. However, there are some lower bounds set by the consensus protocol (empirically, 5 seconds), under which we experience occasional block propagation issues, vote propagation, and aggregation delays that can cause unplanned soft forks.

Each block contains data that describes the states of nodes and the migration plan resulting from the application of the generation algorithm. It also contains the signature of the previous block to follow the principles of the blockchain, so that all blocks are dependent on the previous blocks, which makes it irreversible. To demonstrate our approach, we relied on the sample algorithm presented in Table 1 to generate migration plans according to the resource pool. Blocks also include meta-data that facilitate their utilization such as block hash, previous block hash, VDF proof, aggregated votes, validator set updates, slot, and epoch.

3.4. Consensus mechanism

A key component of a blockchain is the ability for nodes to reach consensus on the global state of the ledger. With increasing interest in blockchain technology in recent years, many consensus algorithms have built upon basic proof of work[8] concepts. However, most algorithms used in permissionless blockchain implementations rely on basic game theory assumptions, which hold only when the blockchain facilitates value transfers, where we can rely on actors acting according to their own (financial) interests (i.e., the nothing at stake problem). In permissioned networks, where there is

[8] Proof of work [36] is a technique that protects from various attacks by requiring a certain amount of processing power to use a service, which makes a potential attack worthless because it becomes too costly.

Algorithm 1 Deterministic migration plan generation.

Input: BlockData
Output: Generation plan
  $Max \Leftarrow FindMaxLoadedNode(BlockData)$
  $Min \Leftarrow FindMinLoadedNode(BlockData)$
  if !AppQueue.isEmpty() **then**
    while !AppQueue.isEmpty() **do**
      $Min \Leftarrow FindMinLoadedNode(BlockData)$
      $Min.addApp(AppQueue.dequeue())$
    end while
  else
    $AppToMigrate \Leftarrow Max.MaxLoadApp$
    $DeltaScore \Leftarrow (Max.score - Min.score)$
    $NextDeltaScore \Leftarrow (Max.score - AppToMigrate.score) - (Min.score + AppToMigrate.score)$
  end if
  if $Math.abs(DeltaScore > NextDeltaScore)$ **then**
    Migrate($AppToMigrate$, Min)
  end if

usually no monetary value, the consensus algorithms used in monetary blockchain implementations are not appropriate. Instead, a known family of consensus algorithms for permissioned networks can be used based on voting schemes for leader elections like PBFT [37], Proof of Elapsed Time (PoET) [38] or RAFT [39]. However, these algorithms require multiple messages to be sent through the network in order to commit a change.

Our algorithm is based on a random draw that is universally verifiable. To achieve decentralized randomness and verifiability, we make use of Verifiable Delay Functions (VDF) [40]. A VDF is a function that takes a large quantity of non-parallel work to compute, and produces a verifiable proof. More specifically, VDFs are similar to time lock puzzles but require a trusted setup where the verifier prepares each puzzle using its private key. Additionally, a difficulty parameter can be adjusted to increase the amount of sequential work, thereby increasing the delay. We extend our previous consensus algorithm [34] such that nodes first compute a VDF depending on the difficulty assigned for block $n+1$, and desired *slotTime*, which is a configurable parameter of the network. We then use the proof $P_n = VDF((n-1)_{hash}, (n-1)_{difficulty})$ as a decentralized entropy pool for random number generator(RNG) to draw decentralized randomness for a given *slot*. For every slot, nodes are able to self-elect into consensus roles (e.g., *Block Producer*, *Validator*, *Committee Member*) as outlined in Algorithm 2. Due to the seeded RNG, all nodes compute the same assignment of roles for all participating nodes thereby not requiring any message exchange to agree on their roles. Moreover, the canonical nature of the chain provides some security guarantees so that the roles for future block $n+1$ cannot be computed before block n is accepted. Once roles are assigned for a given slot, nodes perform their sub-protocols as follows:

1. *Block Producer* is a singular node elected each slot to produce a candidate block. The candidate block is sent to all committee members. Upon sending, the block producer listens for attestations for $\frac{2}{3} * slotTime$, and aggregates them. The aggregated signature is then included in the block header, and gossiped to the entire network if a sufficient number of votes are received, otherwise a skip block is proposed.
2. *Committee Member* are responsible for attesting to candidate blocks. They verify the block integrity, signatures, and data to produce a Boneh-Lynn-Shacham signature (BLS), then send the signature to the block producer.
3. *Validator* nodes receive a new block, verify the integrity and committee signatures to decide to either accept or reject the block.

The protocol assumes all validating nodes form a validator set, which is shared among all nodes participating in the consensus protocol. The assumption is guaranteed by logging inclusions and exclusions in blocks. To build the validator set, a node builds the chain to the current tip (last block),

and upon verifying each block, executes the validator state transition function to reconstruct the validator set. The state transaction function simply stores changes to the membership of the validator set. Nodes that want to participate in the consensus gossip their signed inclusion request, once included into a block, they are considered in the validator set by all nodes simultaneously, and can begin participating by role self-election. Nodes are excluded from the validator set when they are elected to a role of *Block Producer*, and fail to deliver the candidate block to the committee in time. The Committee will then vote for a skip block, which includes only the exclusion of the *Block Producer*. In a permissioned setting, this is considered sufficient to evaluate future failures in case a node is faulty. The node can rejoin the validator set at any time by gossiping an inclusion request. We define the consensus protocol more formally in Algorithm 3.

Algorithm 2 Role election

Input: Slot, ValidatorSet
Output: Roles[]
$Slot_{seed} \Leftarrow VDF(chain_{(slot-1)}.hash, chain_{(slot-1)}.difficulty)$
$ValidatorSet \Leftarrow Shuffle(ValidatorSet, Slot_{seed})$
$Roles['blockProducer'] \Leftarrow ValidatorSet.subset(0, 1)$
$Roles['committee'] \Leftarrow ValidatorSet.subset(1, committeeSize)$
$Roles['validator'] \Leftarrow ValidatorSet.subset(committeeSize, ValidatorSet.size)$

3.5. Security and fault tolerance considerations

Fault tolerance is an important property of the system. The system must guarantee the liveliness of applications running at any given time. Hence the risk of accidental forks (a split in the blockchain) must be examined. In a permissioned setting, forks are accidental and are a product of node failures or message propagation delays. We provide various scenarios of forks, and show how the fork choice rule addresses them.

1. The proposed block b for the current slot s is not propagated to all committee members in time C. A forked subset $C_f \subset C$ votes, and includes a skip block sb for slot s.
   (i) when $\frac{|C|}{2} > C_f$, block b will pass the majority vote, and the tip of the chain is b. However, C_f tip is sb. In this case C_f will produce different role assignments, and attempt to build on sb. Even if the block producer $bp \in C_f$, a majority vote cannot pass as $\frac{|C|}{2} > C_f$. Therefore C_f will add another sb. Eventually, the real block will reach the forked nodes, and due to a hash mismatch, nodes will initiate the fork resolution protocol.
   (ii) when $\frac{|C|}{2} < C_f$, block b will **not** pass the majority vote, and the tip of the chain is sb. No fork will occur.
2. Alternatively, attestations for b can be aggregated in time, but b fails to propagate to all committee members in time. A subset of committee members may then assume the block producer experienced a fault, and start gossiping sb. A network partition in the validator set occurs due to a race condition. However, eventually b will reach nodes with the tip sb. Due to a hash mismatch, they will initiate the fork resolution protocol.

Fig. 4 illustrates how fork resolution works. At height 2, two different blocks are proposed and accepted. both reference the correct previous block hash at which all nodes agreed on the same block. However, any blocks after height 2, will have a different previous block hash. Eventually, one of the chains has to be dismissed. For each of the aforementioned cases where a fork can occur, this eventually happens. In case of disconnect or high latency, the network eventually reaches higher connectivity as peers build new connections. Moreover, for each slot, nodes take up new roles in the consensus protocol, and the likelihood of effected nodes to maintain the same roles decreases exponentially.

In Nakamoto-style [21] consensus algorithms, the fork choice rule states that the longest chain (most proof of work) is the correct chain. However, a vote/role based consensus reduces variance in

Algorithm 3 Consensus

Input: Role[]
switch ($Roles[nodeId]$)
case blockProducer**:**
$block.migrations \Leftarrow prepareMigrationPlan(containerStats)$
$block.signature \Leftarrow sign(block)$
$broadcast(block, committee)$
$votes \Leftarrow await(\frac{slotTime}{3})$
$block.votes \Leftarrow BLS.aggregate(votes)$
if $hasMajority(block.votes)$ **then**
$gossip(block)$
else
$skipBlock()$
end if
case committee**:**
$candidateBlock \Leftarrow await(\frac{slotTime*2}{3})$
if $candidateBlock == null$ **then**
$skipBlock()$
else
$proof \Leftarrow verify(candidateBlock.proof)$
$migrations \Leftarrow verify(candidateBlock.migrationPlan)$
$signature \Leftarrow verify(candidateBlock.signature)$
if $(proof$ & $migrations$ & $signature)$ **then**
$send(vote, blockProducer)$
else
$skipBlock()$
end if
end if
case validator**:**
$block \Leftarrow await(slotTime)$
if $block == null$ **then**
$skipBlock()$
else
$proof \Leftarrow verify(block.proof)$
$migrations \Leftarrow verify(block.migrationPlan)$
$signature \Leftarrow verify(block.signature)$
$votes \Leftarrow verify(block.votes)$
if $(proof$ & $migrations$ & $signature)$& $votes)$ **then**
$chain \Leftarrow block$
end if
end if
end switch

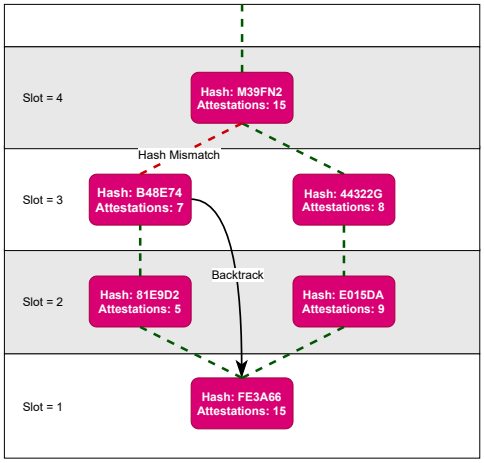

Figure 4. Fork resolution protocol

block time and the forked chain can have an identical block height. Instead, when a node cannot add a new valid block due to a mismatch of previous block hashes it backtracks to rechecking the attestations for each block down to the forked block, and afterwards rebuilds the chain including the blocks by following the chain with most cumulative attestations. Note that in order for a node to receive a valid block with a different previous block hash, the network partitions/high latency had to be resolved for the node to receive the alternative chain.

Another aspect of forks is the fault tolerance related to applications running in the system. Every block includes a migration plan, and in case of a fork, two or more migration plans are created and accepted by two disjoint sets of nodes. A migration plan will include all applications, which guarantees liveliness and variability of computation. Moreover, in case of a chain split, both sets of nodes (A, B) will execute the their respective plans which can unfold in the following two ways:

1. An application λ is planned to migrate from a node in A, to a node in set B or vice versa.
2. An application λ is planned to migrate to another node within sets A or B
3. An application λ does not need to migrate in either chain.

In order for an application to migrate from one node to another, a direct connection between the nodes must be established where one node sends a compressed version of the container to the other. To execute this, both the origin and destination node must agree and run the migration protocol. In case a fork occurred due to high network latency or complete disconnection between the two sets, the migration protocol will attempt to communicate between the sets, resolving the fork as shown in Fig. 4 as long as communication is possible. Whenever a migration plan requires an application to migrate between two conflicting chains, it forces a fork resolution. Moreover, in such cases, only one application is run at the same time. However, in the event application λ is planned to migrate within its originating chain, the migration will not force the network to reconnect. This could be considered a hard fork as there is no connection between the two networks, and results in separate instances of the network. The final example is when application λ is not required to migrate in which case one instance remains running and the system is not affected. The only example where serious faults might occur is when the network is well connected and forks because two nodes drew a winning ticket (same number). However, the migration algorithm is deterministic and the input is in the previous block. This means both block producers will produce the same migration plan even though the block hashes will be different. Although there will be two conflicting chains, the migration plan will be the same.

4. Evaluation and empirical results

To analyze the performance of our implementation, a node was selected to perform logging operations about the state into a time series database. The test environment was built using Docker Swarm to create a cluster. The cluster is comprised of 8 nodes, each with a 16-core (32 thread) Ryzen Threadripper CPU, and 32GB of RAM. The cluster nodes form an overlay network where latencies are almost non existent. Hence, artificial latency was added on individual UDP packets transmitted to create a more realistic environment. To deploy the network, a Docker service was created that runs the containerized node software across the cluster balancing the load across nodes. Each test-net was started with a boostrap setup, whereby the first node is considered the bootstrapping node, and it's public key, and IP address is known to all other nodes. The Docker service starts a new node every 10 seconds to avoid unrealistic network congestion. Each node was limited to 1 CPU core, and 256MB of RAM which exceeds the requirements for running the protocol. Applications were also submitted as Docker images and were able to execute by having each node run a Docker daemon inside their Docker container instance. This two level abstraction allowed the test-net to separate the node resources, and application resources from the host. Figure 5 outlines the architecture used by the cluster.

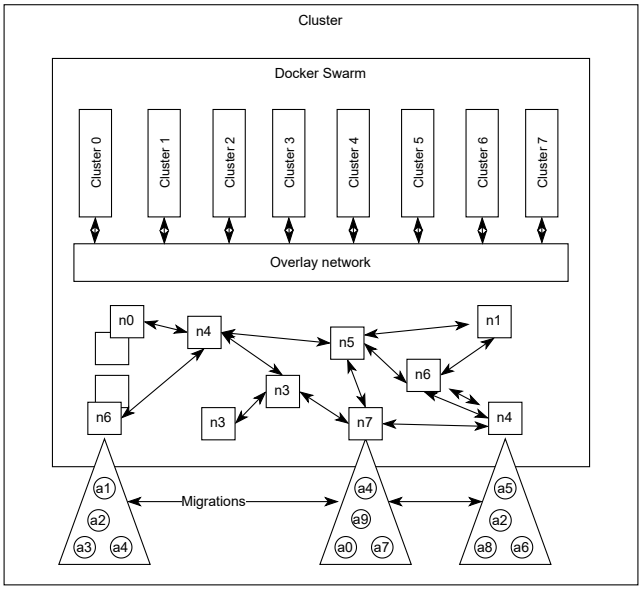

Figure 5. Cluster architecture

4.1. Consensus Layer

To verify that VDF based consensus provides good decentralized randomness, we analyze the distribution of assigned roles. Figure 6 shows the frequency nodes were elected into individual roles. Additionally, since container resource consumption statistics are propagated through a decentralized k-means clustering, cluster representatives are also shown. We observe that nodes have been elected into all roles, while there is some variance in the block producer role, the sample size is only 1000 slots in which only one node is selected as block producer for each slot; a more uniform distribution is expected with a larger sample size. Additionally, nodes joined the network gradually, which also effected the distribution.

To validate the scalability of the consensus layer, we examined a network of 1000 nodes with a committee size of 256 nodes, and a target block time of 16 seconds. Figure 7 shows the block times, and the number of votes per block that were successfully aggregated within the time window for the given slot. We observe that all proposed blocks were accepted as the majority vote threshold was surpassed,

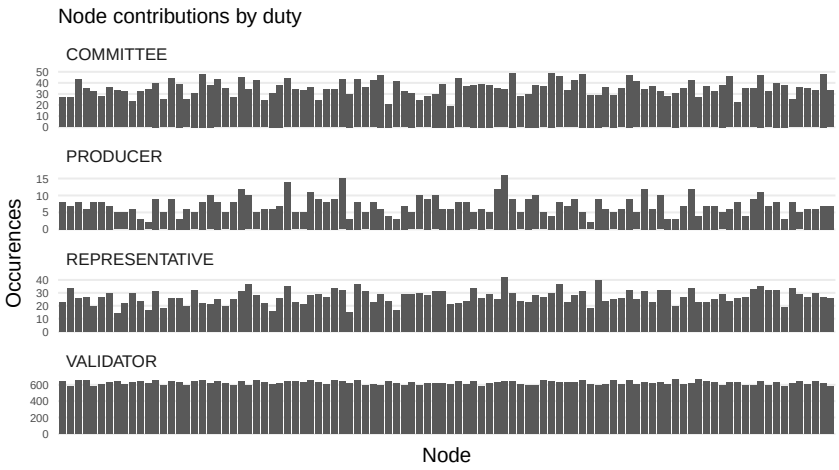

Figure 6. Distribution of roles across all participating nodes.

and no skip blocks were produced despite the low block time, and size of the committee. Moreover, we observe almost no variance in block time indicating that the system had no issue propagating messages.

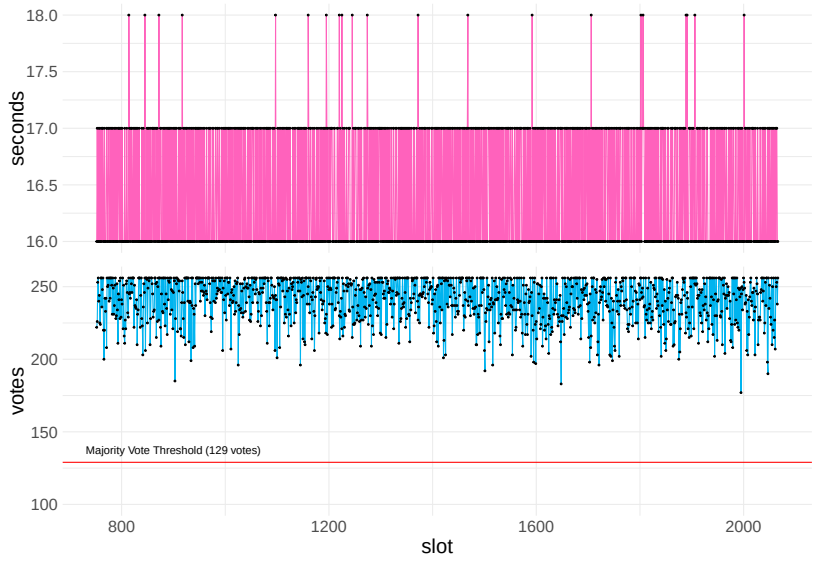

Figure 7. Committee vote aggregation, and block times in a network of 1000 nodes, and 256 committee members.

4.2. Orchestration and Migration

One of the most important features of the system is the ability to migrate applications in a decentralized, transparent, and verifiable way. The decentralized orchestrator aims to distribute load across the network evenly by migrating applications away from nodes with heavy load to those with resources available. To test the performance and efficiency of migrations, we consider the worst case scenario in which all applications were submitted to one node.

Figure 8 illustrates the CPU load of nodes across the last 750 slots because nodes join the network gradually, and affect the early distribution, which skews the observations. We observe that the orchestrator migrated applications away from nodes with high CPU consumption to nodes with more available resources. This resulted in a gradual decline of the mean CPU load across the network.

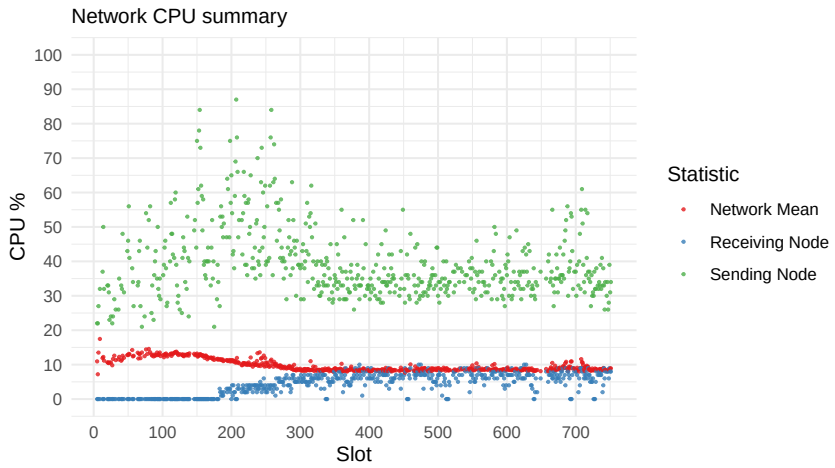

Figure 8. CPU load distribution of the entire network over the last 750 slots.

In Figure 9 we compare both migration times of both test-nets to evaluate the feasibility and performance of CRIU enabled migrations. We break down a migration into 3 steps: *Save*, *Transmit*, and *Resume*. For standard migrations, saving requires pausing the running container, exporting, and compressing it. In CRIU, the container is also paused but instead of exporting it, only the state is extracted and compressed. After transmission, the receiving node must resume the container. In standard migrations, the container is uncompressed, and resumed, while using CRIU, a new container from the same base image is created, and the uncompressed state is injected into it.

Using CRIU, the payload for transmission is much smaller, and hence transmission time is greatly improved over standard. Additionally, compression is a CPU intensive task. Compressing and decompressing only the state of an application instead of the entire container is considerably faster. The median uncompressed exported state of the application using standard migration was 142.2 MB. Using CRIU, the median size of the uncompressed state was 15.2MB. The spikes in standard migrations can be attributed to lack of resources as nodes under heavy stress from running other applications lack the resources needed to perform the compression promptly. Table 2 provides a statistical summary of the observed times in milliseconds. We observe that CRIU enabled migrations are not only faster but also produce more consistent migration times. This can be observed by the considerably lower standard deviation in Table 2.

Type	Segment	Min.	Max.	Med.	Mean	SD
CRIU	Resume	2366	10012	3259	3540	1166
CRIU	Save	1975	8368	2701	3126	1111
CRIU	Transmit	46	833	79	88	61
Standard	Resume	1449	34337	7414	9550	6637
Standard	Save	4010	49080	11231	12875	6942
Standard	Transmit	506	15007	1624	2047	1467

Table 2. Summary of migration times in milliseconds.

Another way to visualise the dynamics of the system is shown in Figure 10. Each polarized bar chart shows nodes on x axis. Stacked bars are used to illustrate the number of applications running

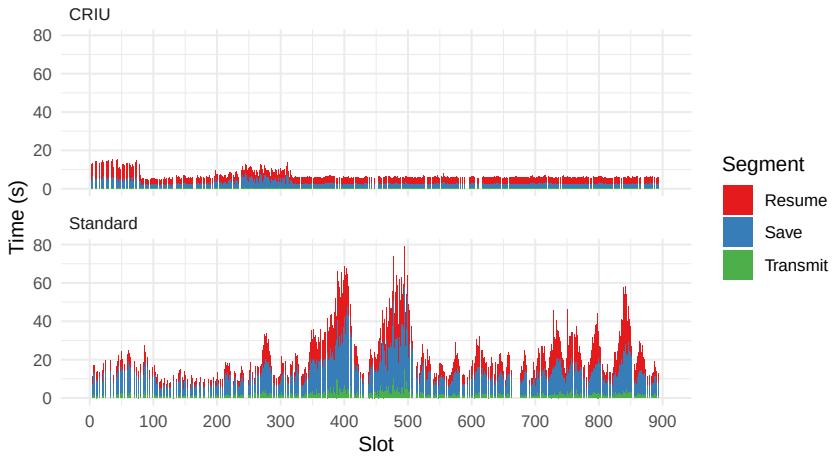

Figure 9. Migration timing comparison between standard and CRIU enabled migrations.

on the node, and their respective CPU consumption in %. We observe that, initially, the application distribution was uneven with a very high CPU load on one node. This is the result of submitting incoming applications to one node. Over time, the system is able to evenly distribute applications across the network.

Figure 10. Discrete time visualisation of applications and their CPU load utilization on participating nodes. Colours indicate individual containers active on each node, and are not necessarily the same container on different nodes.

4.3. Network Clustering

The performance of the orchestrator heavily depends on the propagation speed of resource allocation from all validators. In a clustered network, *validators* report their resources to their cluster *representatives*, which finally send an aggregated report to the block producer. To avoid a potential attack vector on the clustering, the network topology changes every slot. Figure 11 shows the time needed to deliver the resource reports to *representatives*, and finally the *producer*. We observe that in the first few minutes while the nodes are joining the network at a high frequency the propagation times are noticeably slower (still well within the $\frac{1}{block_time}$) but stabilize quickly even with networks of 1000 nodes.

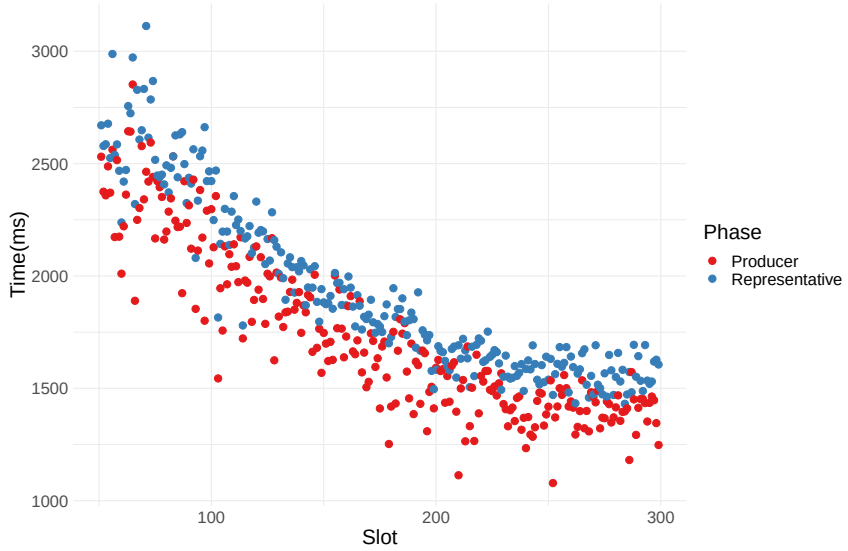

Figure 11. Time distribution of resource propagation in two phases. Initially, validators submit their resource statistics to their respective cluster representatives (*To representative*). After, cluster representatives send collected reports to the block producer (*To Producer*). There were a total of 50 clusters created each slot within a target slot time of 16 seconds, which sets the upper bound for resource propagation at 5.4 seconds.

5. Conclusions

In this paper, we introduced a decentralized architecture capable of run-time application migration for large scale deployments of peer-to-peer IoT sensor networks. We describe three key contributions, namely a scalable consensus protocol layer, an efficient, secure and dynamic topology, and a decentralized orchestrator capable of low latency real time application migrations.

We evaluate each contribution by performing empirical tests with our reference implementation of the protocol. Additionally, we improve migration times by implementing CRIU, an experimental feature of Docker that allows the system to migrate an application's state without effecting its run-time. Using CRIU enabled migrations, we observe considerable reduction (nearly 10-fold) and improved consistency in migration times.

The results of our experiments show that distributed consensus and application management is possible at run-time, thus opening the door to several improvements towards self-managing IoT platforms. The increase in network usage and CPU load has shown to be acceptable when taking into account the scalability, fault tolerance, transparency, and absence of a SPOF that our solution brings. Importantly, we have shown that blockchain overhead is a negligible aspect of the actual cost of application migration as the system is able to finalize blocks with slot times as low as 5 seconds while maintaining higher decentralization than existing platforms such as Multichain, which uses a variation of practical byzantine fault tolerance (PBFT) consensus, and Hyperledger Fabric [41] which uses Raft [39].

As future work, we will explore the limits of our solution with respect to network instability (devices entering and leaving the network) and explore solutions to reduce the required computational power while maintaining optimal application management. Moreover, the algorithm governing the decentralized orchestrator will be extended to allow applications to submit migration policies the orchestrator will respect. As future work, more efficient orchestration algorithms should be explored with emphasis given on performing multiple migrations in the same slot with a non-cycle constraint.

Further, geo-sharding the network must be explored. In a geo-sharded network, nodes participating are assigned into shards based on their geographical location. A weaker consensus within a shard can speed up the state transition by periodically snapshotting sharded states into the main chain. This will enable applications to specify more complex migration policies by limiting a geographical area within which the application may run (geo-fencing). Moreover, a geographically aware system can perform better migrations by migrating applications closer to clients in order to improve network latency. Using ereasure coding, the storage requirements of individual nodes is greatly reduced [42] as full replication is not needed.

Author Contributions: Conceptualization, A.T., J.V. and M.M.; methodology: A.T., J.V., M.M. and M.B.; software, A.T.; validation, A.T., J.V., M.M. and M.B.; formal analysis, A.T., J.V., M.M. and M.B.; investigation, A.T., J.V., and M.M.; funding acquisition and resources, J.V., M.B. and M.M.; data curation, A.T. and M.B; writing–original draft preparation, A.T.; writing–review and editing, A.T., J.V., M.M. and M.B.; visualization, A.T. and M.B.

Funding: This research was funded by H2020 grant numbers 739574 and 857188 and by the Slovenian Research Agency (ARRS) grant number J2-2504.

Acknowledgments: The authors gratefully acknowledge the European Commission for funding the InnoRenew CoE project (H2020 Grant Agreement #739574) and the PHArA-ON project (H2020 Grant Agreement #857188) and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund) as well as the Slovenian Research Agency (ARRS) for supporting project number J2-2504.

Conflicts of Interest: "The authors declare no conflict of interest.'The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

References

1. Jaeger, P.T.; Lin, J.; Grimes, J.M.; Simmons, S.N. Where is the cloud? Geography, economics, environment, and jurisdiction in cloud computing. *First Monday* **2009**, *14*.
2. Khalid, Z.; Fisal, N.; Rozaini, M. A Survey of Middleware for Sensor and Network Virtualization. *Sensors* **2014**, *14*, 24046–24097. doi:10.3390/s141224046.
3. Garcia Lopez, P.; Montresor, A.; Epema, D.; Datta, A.; Higashino, T.; Iamnitchi, A.; Barcellos, M.; Felber, P.; Riviere, E. Edge-centric Computing: Vision and Challenges. *SIGCOMM Comput. Commun. Rev.* **2015**, *45*, 37–42. doi:10.1145/2831347.2831354.
4. Hoebeke, J.; Moerman, I.; Dhoedt, B.; Demeester, P. An overview of mobile ad hoc networks: applications and challenges. *Journal-Communications Network* **2004**, *3*, 60–66.
5. Sha, K.; Wei, W.; Yang, T.A.; Wang, Z.; Shi, W. On security challenges and open issues in Internet of Things. *Future Generation Computer Systems* **2018**, *83*, 326–337.
6. De Souza, L.M.S.; Vogt, H.; Beigl, M. A survey on fault tolerance in wireless sensor networks. *Interner Bericht. Fakultät für Informatik, Universität Karlsruhe* **2007**.
7. Padmavathi, D.G.; Shanmugapriya, M.; others. A survey of attacks, security mechanisms and challenges in wireless sensor networks. *arXiv preprint arXiv:0909.0576* **2009**.
8. Taherizadeh, S.; Jones, A.C.; Taylor, I.; Zhao, Z.; Stankovski, V. Monitoring self-adaptive applications within edge computing frameworks: A state-of-the-art review. *Journal of Systems and Software* **2018**, *136*, 19–38.
9. Östberg, P.O.; Byrne, J.; Casari, P.; Eardley, P.; Anta, A.F.; Forsman, J.; Kennedy, J.; Le Duc, T.; Marino, M.N.; Loomba, R.; others. Reliable capacity provisioning for distributed cloud/edge/fog computing applications. 2017 European conference on networks and communications (EuCNC). IEEE, 2017, pp. 1–6.
10. Le Duc, T.; Oestberg, P.O. Application, Workload, and Infrastructure Models for Virtualized Content Delivery Networks Deployed in Edge Computing Environments. 2018 27th International Conference on Computer Communication and Networks (ICCCN). IEEE, 2018, pp. 1–7.
11. Diallo, M.H.; August, M.; Hallman, R.; Kline, M.; Slayback, S.M.; Graves, C. AutoMigrate: a framework for developing intelligent, self-managing cloud services with maximum availability. *Cluster Computing* **2017**, *20*, 1995–2012.
12. Peltz, C. Web services orchestration and choreography. *Computer* **2003**, *36*, 46–52.
13. Hightower, K.; Burns, B.; Beda, J. *Kubernetes: Up and Running: Dive Into the Future of Infrastructure*; " O'Reilly Media, Inc.", 2017.

548 14. Mercl, L.; Pavlik, J. The comparison of container orchestrators. Third International Congress on Information
549 and Communication Technology. Springer, 2019, pp. 677–685.
550 15. Acuña, P. Amazon EC2 Container Service. In *Deploying Rails with Docker, Kubernetes and ECS*; Springer,
551 2016; pp. 69–98.
552 16. Rathi, V.K.; Chaudhary, V.; Rajput, N.K.; Ahuja, B.; Jaiswal, A.K.; Gupta, D.; Elhoseny, M.; Hammoudeh, M.
553 A Blockchain-Enabled Multi Domain Edge Computing Orchestrator. *IEEE Internet of Things Magazine* **2020**,
554 *3*, 30–36. doi:10.1109/IOTM.0001.1900089.
555 17. Savi, M.; Santoro, D.; Di Meo, K.; Pizzolli, D.; Pincheira, M.; Giaffreda, R.; Cretti, S.; Kum, S.w.;
556 Siracusa, D. A Blockchain-based Brokerage Platform for Fog Computing Resource Federation. 2020
557 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), 2020, pp.
558 147–149. doi:10.1109/ICIN48450.2020.9059337.
559 18. Pires, A.; Simão, J.; Veiga, L. Distributed and Decentralized Orchestration of Containers on Edge Clouds.
560 *Journal of Grid Computing* **2021**, *19*, 1–20.
561 19. Mazzoni, E.; Arezzini, S.; Boccali, T.; Ciampa, A.; Coscetti, S.; Bonacorsi, D. Docker experience at infn-pisa
562 grid data center. Journal of Physics: Conference Series. IOP Publishing, 2015, Vol. 664, p. 022029.
563 20. Buterin, V.; others. Ethereum white paper. *GitHub repository* **2013**, *1*, 22–23.
564 21. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system, 2009.
565 22. Schäffer, M.; di Angelo, M.; Salzer, G. Performance and Scalability of Private Ethereum Blockchains.
566 Business Process Management: Blockchain and Central and Eastern Europe Forum; Di Ciccio, C.;
567 Gabryelczyk, R.; García-Bañuelos, L.; Hernaus, T.; Hull, R.; Indihar Štemberger, M.; Kő, A.; Staples,
568 M., Eds.; Springer International Publishing: Cham, 2019; pp. 103–118.
569 23. Thakkar, P.; Nathan, S.; Viswanathan, B. Performance benchmarking and optimizing hyperledger fabric
570 blockchain platform. 2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of
571 Computer and Telecommunication Systems (MASCOTS). IEEE, 2018, pp. 264–276.
572 24. Ismailisufi, A.; Popović, T.; Gligorić, N.; Radonjic, S.; Šandi, S. A private blockchain implementation using
573 multichain open source platform. 2020 24th International Conference on Information Technology (IT).
574 IEEE, 2020, pp. 1–4.
575 25. Yakovenko, A. Solana: A new architecture for a high performance blockchain v0. 8.13. *Whitepaper* **2018**.
576 26. Maior, H.A.; Rao, S. A self-governing, decentralized, extensible Internet of Things to share electrical power
577 efficiently. 2014 IEEE International Conference on Automation Science and Engineering (CASE). IEEE,
578 2014, pp. 37–43.
579 27. Higgins, N.; Vyatkin, V.; Nair, N.K.C.; Schwarz, K. Distributed power system automation with IEC 61850,
580 IEC 61499, and intelligent control. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications
581 and Reviews)* **2011**, *41*, 81–92.
582 28. Suzdalenko, A.; Galkin, I. Instantaneous, short-term and predictive long-term power balancing techniques
583 in intelligent distribution grids. Doctoral Conference on Computing, Electrical and Industrial Systems.
584 Springer, 2013, pp. 343–350.
585 29. Niyato, D.; Xiao, L.; Wang, P. Machine-to-machine communications for home energy management system
586 in smart grid. *IEEE Communications Magazine* **2011**, *49*, 53–59. doi:10.1109/MCOM.2011.5741146.
587 30. Bragard, Q.; Ventresque, A.; Murphy, L. Self-balancing decentralized distributed platform for urban traffic
588 simulation. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *18*, 1190–1197.
589 31. Al-Madani, B.M.; Shahra, E.Q. An Energy Aware Plateform for IoT Indoor Tracking Based on RTPS.
590 *Procedia computer science* **2018**, *130*, 188–195.
591 32. Teh, P.L.; Ghani, A.A.A.; Chan Yu Huang. Survey on application tools of Really Simple Syndication (RSS):
592 A case study at Klang Valley. 2008 International Symposium on Information Technology, 2008, Vol. 3, pp.
593 1–8. doi:10.1109/ITSIM.2008.4631980.
594 33. Samaniego, M.; Deters, R. Using blockchain to push software-defined IoT components onto edge hosts.
595 Proceedings of the International Conference on Big Data and Advanced Wireless Technologies. ACM, 2016,
596 p. 58.
597 34. Tošić, A.; Vičič, J.; Mrissa, M. A Blockchain-based Decentralized Self-balancing Architecture for the Web of
598 Things. European Conference on Advances in Databases and Information Systems. Springer, 2019, pp.
599 325–336.

35. Bozyigit, M.; Wasiq, M. User-level process checkpoint and restore for migration. *ACM SIGOPS Operating Systems Review* **2001**, *35*, 86–96.
36. Dwork, C.; Naor, M. Pricing via Processing or Combatting Junk Mail. Advances in Cryptology — CRYPTO' 92; Brickell, E.F., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1993; pp. 139–147.
37. Castro, M.; Liskov, B.; others. Practical byzantine fault tolerance. OSDI, 1999, Vol. 99, pp. 173–186.
38. Chen, L.; Xu, L.; Shah, N.; Gao, Z.; Lu, Y.; Shi, W. On security analysis of proof-of-elapsed-time (poet). International Symposium on Stabilization, Safety, and Security of Distributed Systems. Springer, 2017, pp. 282–297.
39. Ongaro, D.; Ousterhout, J. In search of an understandable consensus algorithm. 2014 {USENIX} Annual Technical Conference ({USENIX}{ATC} 14), 2014, pp. 305–319.
40. Boneh, D.; Bonneau, J.; Bünz, B.; Fisch, B. Verifiable delay functions. Annual International Cryptology Conference. Springer, 2018, pp. 757–788.
41. Androulaki, E.; Barger, A.; Bortnikov, V.; Cachin, C.; Christidis, K.; De Caro, A.; Enyeart, D.; Ferris, C.; Laventman, G.; Manevich, Y.; others. Hyperledger fabric: a distributed operating system for permissioned blockchains. Proceedings of the thirteenth EuroSys conference, 2018, pp. 1–15.
42. Perard, D.; Lacan, J.; Bachy, Y.; Detchart, J. Erasure code-based low storage blockchain node. 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2018, pp. 1622–1627.

© 2022 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

2.2 Paper 2

Title: Use of Benford's law on academic publishing networks
Authors: Aleksandar Tošić, Jernej Vičič
Year: 2021
Journal: Journal of Informetrics
DOI: 10.1016/j.joi.2021.101163
Link: https://www.sciencedirect.com/science/article/pii/S1751157721000341

Journal of Informetrics 15 (2021) 101163

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Use of Benford's law on academic publishing networks

Aleksandar Tošić [a,b,*], Jernej Vičič [a,c]

[a] University of Primorska Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia
[b] InnoRenew CoE Livade 6, 6310 Izola, Slovenia
[c] Research Centre of the Slovenian Academy of Sciences and Arts, The Fran Ramovš Institute, Slovenia

ARTICLE INFO

Article history:
Received 4 August 2020
Received in revised form 26 March 2021
Accepted 6 April 2021
Available online 20 April 2021

Keywords:
Benford's law
Citation network
Bibliography

ABSTRACT

Benford's law, also known as the first-digit law, has been widely used to test for anomalies in various data ranging from accounting fraud detection, stock prices, and house prices to electricity bills, population numbers, and death rates. Scientific collaboration graphs have been studied extensively as data availability increased. Most research was oriented towards analysing patterns and typologies of citation graphs and co-authorship graphs. Most countries group publications into categories in an attempt to objectively measure research output. However, the scientific community is complex and heterogeneous. Additionally, scientific fields may have different publishing cultures, which make creating a unified metric for evaluating research output problematic. In complex systems like these, it is important to regularly observe potential anomalies and examine them more carefully in an attempt to either improve the evaluation model or find potential loopholes and misuses. In this paper, we examine the potential application of Benford's law on the official research database of Slovenia. We provide evidence that metrics such as number of papers per researcher conform to Benford's distribution, while the number of authors per paper does not. Additionally, we observe some anomalies and provide potential reasoning behind them.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

This paper describes a new application of Benford's law to the analysis of scientific research collaboration network, focusing on the example of Slovenian scientific publications and authors. It presents scientific foundations for the presented methodology with simplified usage directions. The method was tested on a case study of the Slovenian research collaboration network and shows that Benford's law holds. Further, several scientific and multidisciplinary fields were tested to see if the distribution of first digits, too, obey Benford's law.

The widely known phenomenon called Benford's law (Benford, 1938; Singleton, 2011), also referred to as the first-digit law, is an observation about the frequency distribution of leading digits in many real-life sets of numerical data. It describes the distribution of digits in natural and social processes. The numbers take the form of a logarithmic distribution. Though very old and extensively researched, Benford's law is still an interesting tool for finding anomalies in data. Further, the ever growing amount of data generated calls for simple and effective methods for anomaly detection. While Benford's law has defied many attempts at an easy derivation (Berger & Hill, 2011), many have focused on its application rather then its

* Corresponding author at: University of Primorska Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia.
E-mail addresses: aleksandar.tosic@innorenew.eu (A. Tošić), jernej.vicic@upr.si (J. Vičič).
URLs: https://innorenew.eu (A. Tošić), https://osebje.famnit.upr.si/jernej/ (J. Vičič).

https://doi.org/10.1016/j.joi.2021.101163
1751-1577/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

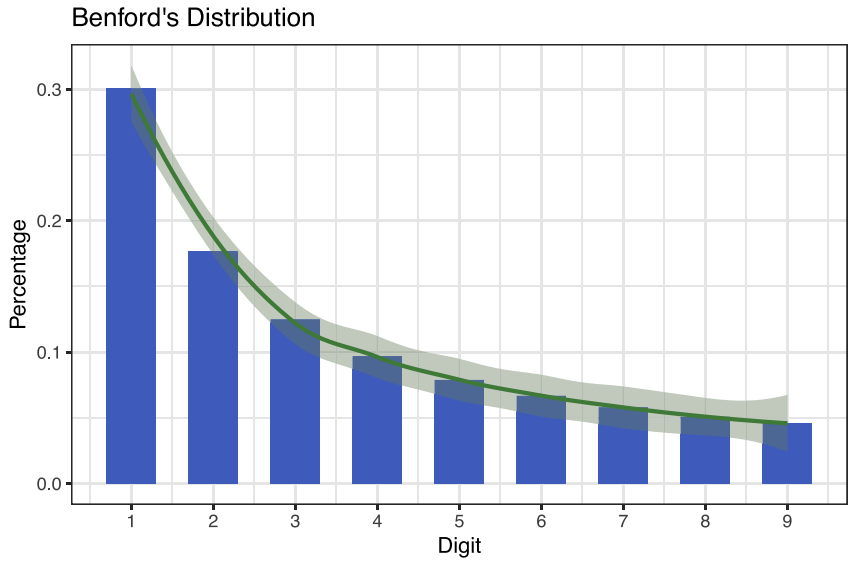

Fig. 1. The distribution of digits in accordance to Benford's law (Singleton, 2011).

theoretical background (Nigrini, 1996). The applications range from election fraud detection, detecting image manipulation, accounting fraud detection (Durtschi, Hillison, & Pacini, 2004), scientific fraud detection (Ranstam et al., 2000), etc. Benford's law has been effectively and frequently used in forensic accounting as presented in Bhattacharya and Kumar (2008) and Nigrini (2012). The same methodology has also been successfully used in other analyses such as the modelling of behavioural features for social network users (Golbeck, 2015) and meteorological events, for example, the travelled distances of tropical cyclones since 1842.

The objective of this paper is to evaluate the potential use of Benford's law on research networks. The method can be used to detect potential discrepancies from a macro level. Upon detection, the method can produce observations that skew the distribution referred to as suspects. These observations are a great entry point for further, more fine grained analysis in an attempt to explain and reason about them. The method is useful to self evaluate research communities as well as the maturity of specific research fields. It can serve as a feedback loop to the regulator in order to improve on the valuation system and respond to potentially unwanted shift in direction. Additionally, we show the method can be applied on temporal data. This is especially useful to examine when discrepancies occurred.

This paper is composed as follows: Section 2 describes Benford's law (Singleton, 2011) and its applications followed by a description of the state of the art in the scientific field with the assumption that the paper is multidisciplinary. Section 3 presents the Slovenian scientific/research network that was used as a case study for the presented methodology followed by methodology and results. The paper concludes with a discussion.

2. Benford's law

The first-digit law is an observation about the frequency distribution of leading digits. It is also known as the Newcomb-Benford law or Benford's law. It has been apparently first discovered by polymath Newcomb and published in Newcomb (1881) and later rediscovered by physicist F. Benford and presented in Benford (1938). The Benford's law (Singleton, 2011) defines a fixed probability distribution for leading digits of any kind of numeric data with the following properties:

- Data with values that are formed through a mathematical combination of numbers from several distributions.
- Data that has a wide variety in the number of figures (e.g., data with plenty of values in the hundreds, thousands, tens of thousands, etc.)
- Data set is fairly large, as a rule of a thumb at least 50–100 observations (Kenny, 2015).
- Data is right skewed (i.e., the mean is greater than the median), and the distribution has a long right-tail rather than being symmetric.
- Data has no predefined maximum or minimum value (with the exception of a zero minimum).

The distribution of digits is presented in Fig. 1; the digit 1 occurs in roughly 30% of the cases, and the other digits follow in a logarithmic curve. It has been shown that this result applies to a wide variety of data sets (Singleton, 2011), including

electricity bills, street addresses, stock prices, house prices, population numbers, death rates, and lengths of rivers. The equation for the distribution of the first digits of observed data is presented in Eq. (1).

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(1 + \frac{1}{d}\right) \tag{1}$$

3. Slovenian research network

In Slovenia, researchers are evaluated by a methodology issued by the Slovenian Research Agency (ARRS). The methodology is a transparent set of rules also implemented in the national research information system SICRIS (Korošec, 2014). Scientific contributions that get added to the database are classified into typologies and then credited with points upon verification. Special emphasis is given to scientific articles that are evaluated based on data from the Web of Science (WoS) with SCIE/SSCI/AHCI journal indexes, Journal of Citation Reports impact factor database (JCR), and WoS citations as well as Scopus Source Normalized Impact per Paper impact factor database (SNIP) for social sciences and humanities and Scopus citations (Falagas, Pitsouni, Malietzis, & Pappas, 2008). Such a robust system is needed in order to provide a holistic method for government funding of young researchers as well as national and foreign project proposals. Moreover, the system is used for progression of academic rank by enforcing minimal requirements from PhD onward. A detailed and systematic overview of Slovenian research network is given by (Curk, 2019).

Journal papers are covered by typologies 1.01, 1.02, and 1.03 and are ranked into four categories (quarters), corresponding to the journal's position in scientific field, that are used to compute the score, which is then divided equally amongst the authors. Additional points are added based on the number of citations to each author equally.

Another important aspect is the minimum requirements for PhD dissertations, which require candidates to publish at least one paper in an SCI indexed journal. (Heneberg, 2016) observed that bibliometric indicators increasingly affect careers, funding, and reputation, creating a new extreme: "where a scientist publishes has become much more important than what is published", which is in agreement with Holub, Tappeiner, and Eberharter (1991) and Shibayama and Baba (2015). While the need for a holistic and robust system for valuating research contributions is arguably necessary, it is equally important for the system to describe specifics for each research field. Some research fields have very different publishing habits and cultures, making it difficult to provide a unified framework of evaluation while maintaining comparability and simplicity. (Larivière, Archambault, Gingras, & Vignola-Gagné, 2006) concluded there are significant differences in publishing habits between social sciences and humanities (SSH) and natural sciences and engineering (NSE), creating a particular problem in the field of bibliometric valuations.

The discrepancy between valuation metrics and publishing habits can create unwanted incentives. Coupled with a systemic point-based funding system, this can create undesired incentives to abuse the valuation methodology in one's favour. Additionally, different research fields have different publishing cultures with respect to the topology of the contribution. It is important to observe, and monitor the publication network from a macro perspective to identify trends, and address potential discrepancies. A holistic view is of great importance both to academia to identify trends, as for the government to adjust the valuation metrics and guide the research towards predefined goals. Moreover, identifying unexpected changes in the publishing network on a macro level calls for further, more fine grained analysis on the micro level.

4. State of the art

Benford's law has been thoroughly researched and its theoretical grounds have been proved in many scientific papers. The phenomenon is discussed in greater detail in Berger and Hill, 2011 the article also provides strengthened versions of, and simplified proofs for, many key results in the literature. Many researchers have verified for themselves that the law is widely obeyed but have also noted that the popular explanations are not completely satisfying (Fewster, 2009). Bibliometric and infometric studies have addressed the issue of academic network and co-authorship network profusely going from global or national views Braun and Glaenzel (1996) or Leydesdorff and Wagner (2008) to the individual level Melin (2000) and Newman (2004). Ariel Xu and Chang (2020) show that the co-authorship network correlates well with the academic performance. There are numerous papers that found a positive relationship between the international collaboration and the research impact such as Narin and Whitlow (1990) or Katz and Hicks (1997). The background, the current status, and trends of academic social networks are researched and finds presented in Kong, Shi, Yu, Liu, and Xia (2019). A study on research collaboration Benavent-Pérez, Gorraiz, Gumpenberger, and de Moya-Anegón (2012) as well as some studies that date a few decades ago, such as Bordons, Gomez, Fernández, Zulueta, and Méndez (1996), focus on geographical impact on international and intra-national research collaboration. Ortega (2014) presents an analysis of the relationship between research impact and the structural properties of co-author networks. The methodology described in our paper is most suitable for the observation of maturity of the research area. There has been some research in this area such as Keathley-Herring et al. (2016); Pelacho, Ruiz, Sanz, Tarancón, and Clemente-Gallardo (2021) presents an analysis of the evolution of a targeted science field. One of our aims is to observe the changes through time in bibliographic network resulting in maturity of the network. (Batagelj & Maltseva, 2020) proposes a method to transform bibliographic networks, using the works' publication year, into corresponding temporal networks based on temporal quantities and then defines interesting temporal properties of nodes, links and their groups thus providing an insight into evolution of bibliographic networks.

The rest of the section presents state of the art on various fields and aspects that are connected with our research.

4.1. Benford's law applications

The application of Benford's law is by far most prevalent in the accountant fraud detection and there has been a lot of research in the area, such as (Drake & Nigrini, 2000) who introduces students to Benford's Law and Digital Analysis (analysis of digit and number patterns of a data set), which can be used as an analytical procedure and fraud detection tool. Nigrini (2017) presents a current literature overview of the area. Durtschi et al. (2004) presents Benford's law as a simple and effective tool for the detection of fraud. The purpose of the paper is to assist auditors in the most effective use of digital analysis based on Benford's law by identifying data sets which can be expected to follow Benford's distribution, and presenting types of frauds that would be "detected/not detected" by such analysis.

Cleary and Thibodeau (2005) however, points out some inherent problems that potentially arise in the use of the Benford's law in the auditing process. The paper compares the merits of Benford's law and typical statistical test-by-test approach.

The simplicity of the Benford's law as a tool allows for a broad range of uses. Hickman and Rice (2010) examined crime statistics at the USA National, State, and local level in order to test for conformity to the Benford distribution. Burke and Kincanon (1991) observe the distribution of initial digits of physical constants, their results are inconclusive, though. One of the more recent researches involving Benford's law is Zhang (2020), where the authors propose a test of the reported number of cases of coronavirus disease 2019 in China with Benford's law and report that the reported numbers of affected people abide to Benford's law.

4.2. Research integrity and unethical behaviour

Benford's law has been most successfully used in accounting fraud detection and quite a few research projects aimed at using the same tool to detect frauds in other fields such as Zhang (2020) in the detection of counterfeiting COVID-19 reports. It comes naturally to take the inspiration from financial area and just shift it to the new domain, but this is not our aim, the use of the presented tool as a fraud detection metric still needs to be explored.

The problem of research integrity and unethical behaviour has spread in new forms and dimensions as observed by many scholars such as Martin (2013) and Bohannon (2013). There is now a growing body of research on scientific integrity and misconduct as well as presentation of guidelines and best practices such as Fanelli (2013) or Peterson (2007). There are whole courses devoted to this issue at the university level using textbooks such as Macrina (2014) and Sponholz (2000). Many research papers and opinions have been published in recent years expressing concerns over research integrity such as Godecharle, Nemery, and Dierickx (2014) and Bernstein (1984), which examines ethical issues raised in one example case. Edwards and Roy (2017) argues that scientists have become increasingly perverse in terms of competition for research funding and development of quantitative metrics to measure performance. The peer-review system's effectiveness as a means of preventing misconduct in science is challenged in papers such as Van der Heyden, van de Derks Ven, and Opthof (2009). Surveys indicate increasing numbers and extremes of misconduct (John, Loewenstein, & Prelec, 2012). Our paper does not try to newly define the delicate subject of academic integrity, nor does it present a tool that pinpoints misconduct of a single author or a single poor research contribution. Instead, the article proposes a methodology for following changes and assessing the maturity of research system.

4.3. The research of the Slovenian research network

Each new bibliographic source, should be tested for its suitability for bibliometric analyses such as the case of Microsoft Academic Search (Ortega, 2014). The Slovenian academic/research network has been examined extensively through different perspectives including a quantitative and qualitative methodological approach to scientific cooperation by Mali, Pustovrh, Cugmas, and Ferligoj (2018), a study of community structures through scientific co-authorship (Cugmas, Ferligoj, & Kronegger, 2019). Pisanski, Pisanski, and Pisanski (2020) present two methods to ease visualization of large networks such as bibliographic networks. They showcase the methods on Slovenian research network. Ferligoj, Kronegger, Mali, Snijders, and Doreian (2015) examine the collaboration structures and dynamics of the co-authorship network of all Slovenian researchers. Its goal is to identify the key factors driving collaboration and the main differences in collaboration behavior. A new measure for interdisciplinarity that takes into account graph content and structure is proposed in Karlovčec and Mladenić (2015). The proposed new measure is applied in exploratory analysis of research community in Slovenia; a commentary to this paper (Rodela, 2016) addresses two shortcomings while still supporting the weight of the paper. Lužar, Levnajić, Povh, and Perc (2014) presents a study of the dynamics of interdisciplinary sciences in the case of Slovenian scientific network.

5. Methodology

As mentioned in Section 1, this paper proposes a methodology for following changes and assessing the maturity of research system. As such, the purpose is to present scientific grounds that allow feasibility and usefulness of the method as well as to propose a set of usage guidelines and a use case where our hypotheses were confirmed. The observation sets need to conform to all the basic prerequisites for Benford's law as described in Section 2. The Slovenian research network

described in Section 3 was used as a use case for a network of scientific publications. The methods and premises are easily applicable to other European and non-European countries. The data identified in our study includes, but it is not limited to, the following examples:

- the number of publications per author,
- the number of co-authors per author,
- authors are usually categorized in one or more scientific fields,
- authors are associated to one or more research institutions that can be further geo-located,
- publications are all tagged with the year they were published.

Another comparison comes naturally when observing the presented data: "the number of authors per publication", but although, in theory, this number is unlimited, in practice they are not, eliminating just a few of the most populated publications. We are left mostly with numbers 1 or low 2-digit numbers. We hypothesise, and confirm the validity in Section 7, that the following comparisons are all subject to Benford's law:

- the number of co-authors per author,
- the number of publications per author.

The highest number of publications per author in our dataset was 8483 and there were more than 0.5% that were higher that 1000, which satisfies (although borderline) the second prerequisite described in Section 2. Each of the assumptions can be further distributed on selected sets:

1. the whole (Slovenian) scientific network,
2. broken into scientific areas (e.g., natural sciences, social sciences, etc.),
3. grouping publications in discrete time periods,
4. grouping authors into geographical areas.[1]

The rationale behind selection of the presented sets follows: #1 all publications, #2 is there a scientific area that has is in starting phases of development or are the numbers of publications and connections skewed by some other property, #3 was there a time period when the observed set did not behave according to Benford's law, #4 are there geographical areas (usually corresponding to national borders – ethnic or institutional perceptions of research publications) where, again, the numbers do not conform to Benford's law.

Testing that data conforms to Benford's distribution has been done with many goodness of fit tests ranging from Pearson's Chi squared, Kolmogorov–Smirnov D statistics, Freedman's modification of Watson U^2 statistics, euclidean distance d statistics, and many others. However, no real data will ever follow the exact distribution; hence, most analysis supplements statistical testing with graphical representations that help in pointing out suspicious patterns in the data for further investigation. Additionally, different tests have different reactions on sample sizes. The Chi square test suffers from an excess power problem in that when the number of observations becomes large (above 5000 records estimated by Nigrini (2012)) it becomes more sensitive to insignificant spikes, leading to the conclusion that the data does not conform. Nigrini (2001) suggested some statistical tests can render misleading results when applied to large number of observations. On the other hand, Druica, Oancea, and Vâlsan (2018) conclude that MAD test is reliably with as low as 200 observations. Alexander (2009) proposed the Mantissa Arc test, which is a very interesting geometrical test. Unfortunately, it tolerates little deviation from Benford's distributions. Nigrini (2012) concluded that the best test is Mean Absolute Deviation (MAD), also setting critical objective scores for conformity (0.000), acceptable conformity (0.006), Marginally acceptable conformity(0.012), and nonconformity (0.015). The adapted MAD is used to measure the average deviation between the heights of the bars and the Benford line. The higher the MAD, the larger the average difference between actual and expected proportions. In our use case, we perform all conformity tests using all three of the aforementioned tests as our sample sizes are well within the acceptable ranges. We supplement the statistical tests with graphical representations; the results are presented in Section 7.

Basic recipe: Select a big enough set of aggregated data that conforms to Benford's law prerequisites. Gather data and count aggregated values. Count leading values and perform Mean Absolute Deviation (MAD) on the gathered data. Plot simple bar charts with the numbers for each leading digit and observe the distribution. If the data does not conform to Benford's law, investigate further.

[1] this set was not addressed in the paper but the authors do not expect any discrepancy from the presented results as long as we keep the number high enough to conform to Benford+s law prerequisites.

5.1. Description of the data set and data acquisition process (Slovenian academic network)

The Slovenian academic network is represented by two public services: Slovenian Current Research Information System (SICRIS) (Curk, 2019), which stores data about scientific research including scientist/researcher education, degree, scientific field, affiliation, type of employment, and national funded research projects, and Cooperative Bibliographic System and Services (COBISS) (Seljak & Bošnjak, 2006), which stores data about all publications (including research publications that are our primary concern in this paper). A research classification scheme is used for unique identification of researchers. Researchers' bibliographies are created in the shared cataloguing process; however, the use of a uniform methodology of documents/works is mandatory for classification of bibliographic items. The two services are seamlessly combined and publicly available through a querying interface. Although the data is publicly available and can be acquired for research purposes, we opted for crawling the available data. In March 2020, a local database was constructed by slowly crawling the service in order to avoid load problems. Each researcher is represented by an unique Researcher ID (mstid) issued by the Slovenian Research Agency. The foreign co-authors were not disambiguated; a new entity was created for each co-author of an observed bibliographic entry. The data gathered in this process is valid as long as we observe Slovenian authors and use foreign co-authorship only in aggregated form. The local database's structure is presented in Fig. 2.

6. Misuse of the proposed metric

This section presents an example where the proposed metric will give misleading results. Benford's law is not applicable for the number of authors per research contribution case as it violates the second constraint outlined in Section 2 that requires a wide variety in the number of figures, ranging into hundreds and thousands. The highest number of coauthors on a single contribution in Slovenian academic network is 40 and there are only 92 contributions with 30 or more co-authors and only 323 with 20 or more. Fig. 3 shows a skewed distribution of first digits. The first digits were analyzed with a sample size of 788,410 observations. The MAD Conformity (Nigrini, 2012) was classified as Nonconformity with all statistical tests having *P*-values near 0. Since the number of authors per paper does not conform to Benford's law.

7. Results and discussion

This section presents the results of the methods applied on identified sets (all described in Section 5. Fig. 4 shows that distribution of first digits for number of contributions per author for the whole dataset (e.g., the whole Slovenian scientific network) conforms to Benford's law. The MAD Conformity according to (Nigrini, 2012) is Acceptable.

Table 2 summarizes the results of testing individual research fields on conference papers and journal papers, separately. The main statistics used to determine if the data conforms to Benford's law was mean absolute deviation (MAD Conformity) (Nigrini, 2012). The distortion factor model indicates whether the digit patterns are over- or understated and extent of the distortion. Additionally, Pearson's Chi-squared and Mantissa Arc tests are given for clarity and confirmation in cases with marginally acceptable conformity. Based on the aforementioned tests, data sets get classified as either nonconformity, marginal, acceptable, or close conformity. We observe that most fields conform very strongly, with exceptions being number of journal papers published in social sciences and conference papers published in humanities and biotechnology. Further analysis can be drawn from the digit distribution charts for individual data set.

Figures 5, and 6 provide more insight into individual research fields, and contribution typologies. Suspects are marked (red) if the squared difference between the observed digit distribution and Benford's distribution is greater then 4. Finding the correct threshold is an empirical process in which most suspect observations can be obtained by starting with the highest threshold (observation with the highest squared difference) and decreasing it in a step by step process to obtain a larger set of observations that skew the fit the most. These observations can provide more insight into the objective MAD conformity results. The threshold was chosen empirically solely for a meaningful graphical representation. Social sciences have the biggest offset in the first three digits, which could be explained by the different publishing culture or an increase in the number of researchers that only publish the single journal paper required for academic title or PhD thesis.

We establish that, in general, the number of published papers per author conforms to Benford's distribution with some variances between different research fields, which could be attributed to different publishing cultures. However, using Benford's law to possibly identify anomalies in publishing is not very useful if performed without taking into account the temporal aspect of the data. We show that Benford's law can be used on shorter time frames to narrow the search considerably. Analyzing temporal data using Benford's distribution has been shown before (Sambridge, Tkalčić, & Jackson, 2010), where samples are divided into time intervals and tested individually. In Fig. 7, we show the digit distribution of the same data split by decade. We observe that from 1960 until 1990 the digit distribution does not conform. However, after the 1990s we see a strong conformation. The data from 1960–1970 only has 38 observations, which is well below the threshold. The contributions published between 1970–1980 only amount to 493 observations, suggesting the minimum sample size could be increased considerably for this particular use case.

Table 1 shows the numerical results for conformity tests. The key observation is that Benford's law can be used on lower time frames as long as the sample size is large enough. Additionally, we observe a sizable increase in the number of papers published between 1990–2000; this is possibly a result of increased Internet access availability, which bolstered international collaboration that resulted in more published papers.

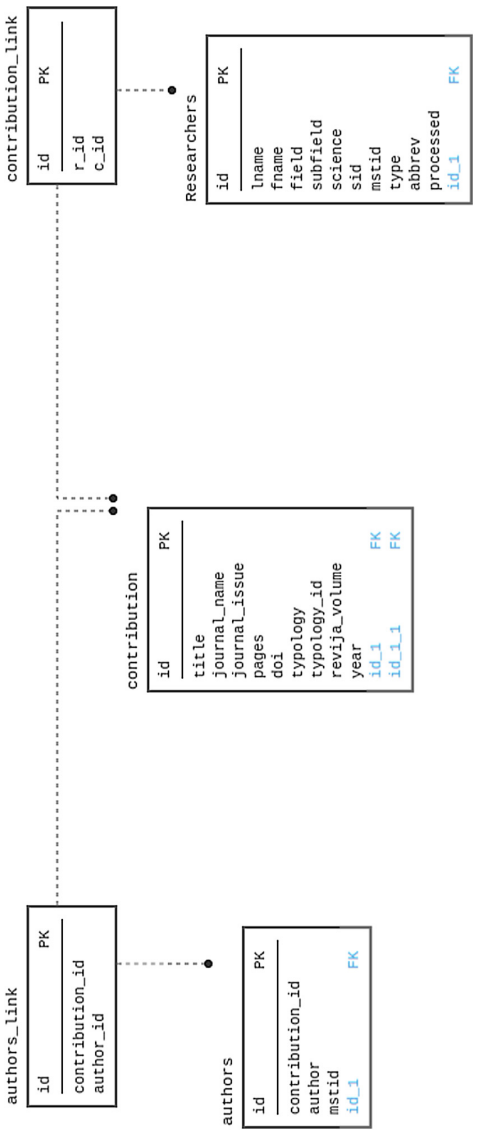

Fig. 2. Structure of the gathered data set.

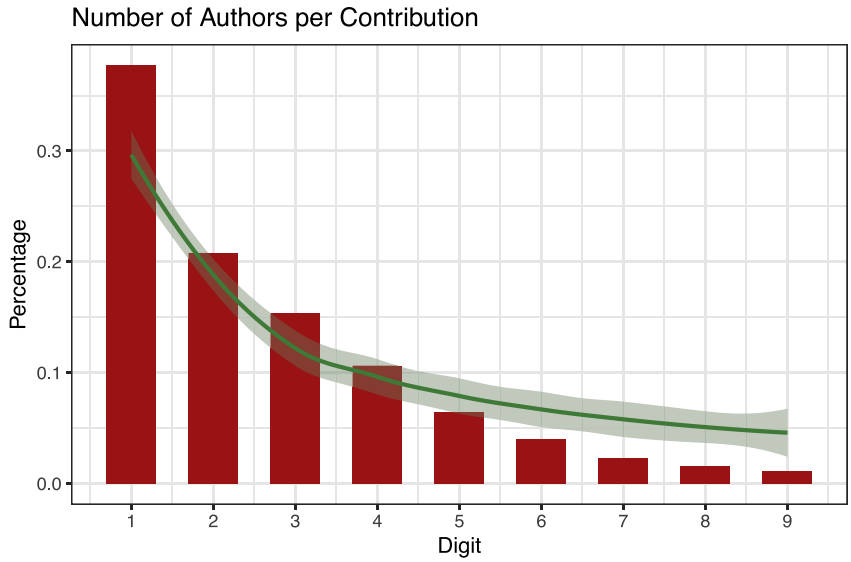

Fig. 3. Example of a misused metric. The graph shows that the observed data do not conform to Benford's law. The data does not conform to Benford's law because the number of authors per publication rarely exceeds 10. The Benford's law is not applicable to the presented data.

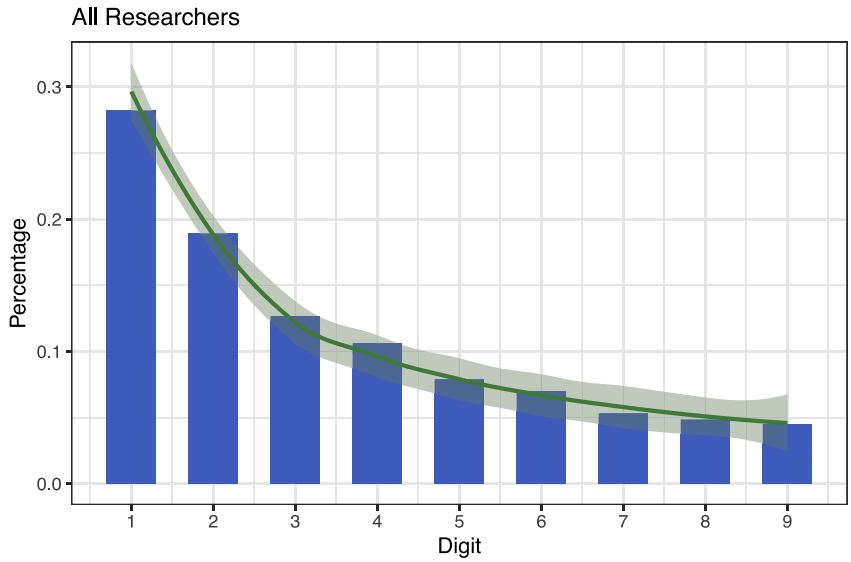

Fig. 4. The distribution of first digits for number of contributions per author for the whole Slovenian scientific network conforms to Benford's law. The MAD Conformity (Nigrini, 2012) is Acceptable.

Fig. 8 shows the results of tests performed on a yearly basis for each individual field. We observe that for the first 20 years the sample sizes were below the threshold $T = 50$, and we have marked those with "Insufficient data" accordingly. We observe that after the 1990s the Slovenian scientific community matured as confirmed by the general conformity of all fields with the exception of humanities, which lagged behind a few years. From these results, we conclude Benford's law can be effectively used on a yearly basis to identify potential discrepancies, which could be attributed to a number of factors ranging from unethical behaviour, different publishing cultures, inefficiencies in the valuation metric, underfunded research fields, etc. A more detailed analysis should be performed on those observations that skew the distribution the most, refered to as suspects.

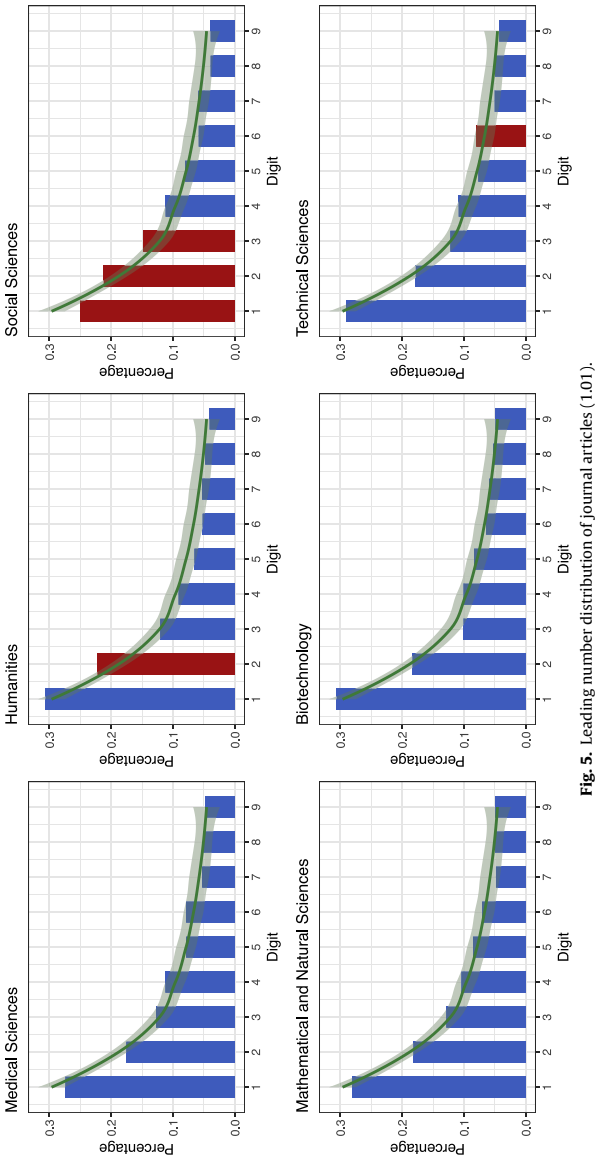

Fig. 5. Leading number distribution of journal articles (1.01).

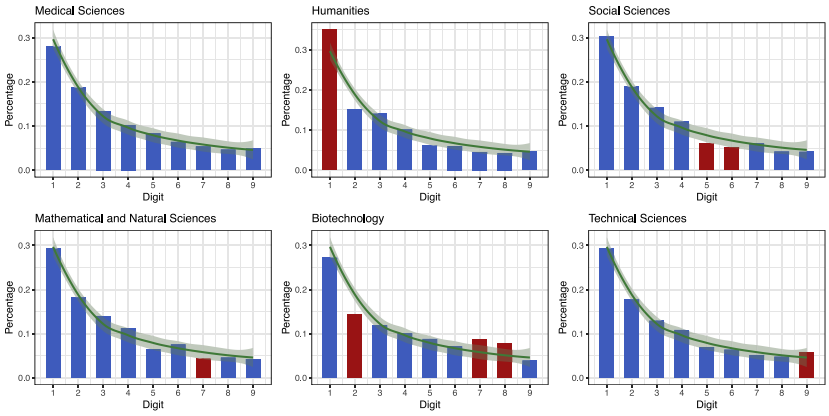

Fig. 6. Leading number distribution of conference articles (1.08).

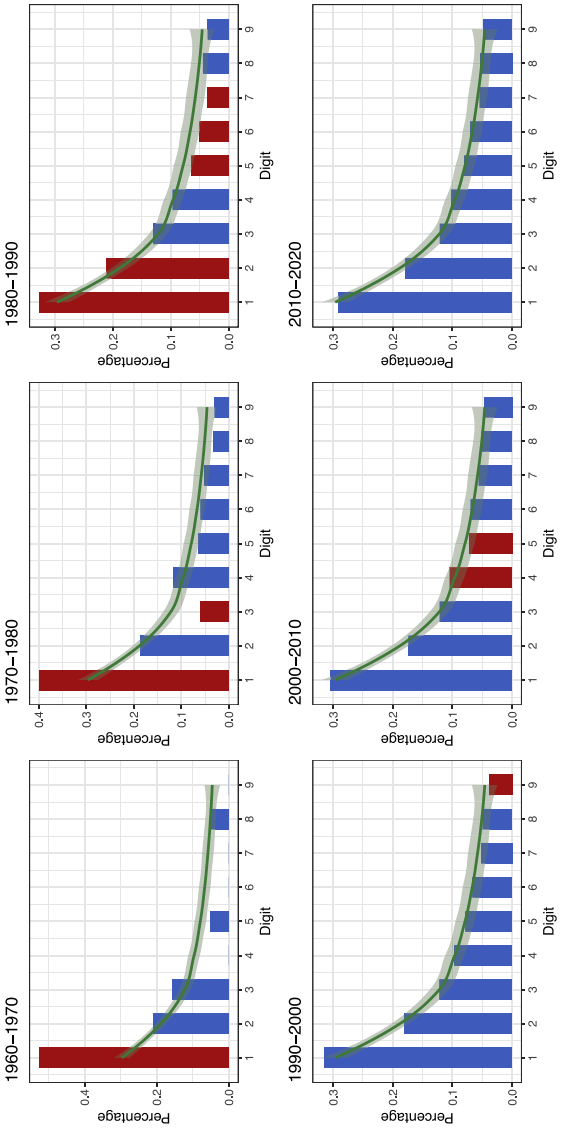

Fig. 7. Number of papers per author grouped by decade.

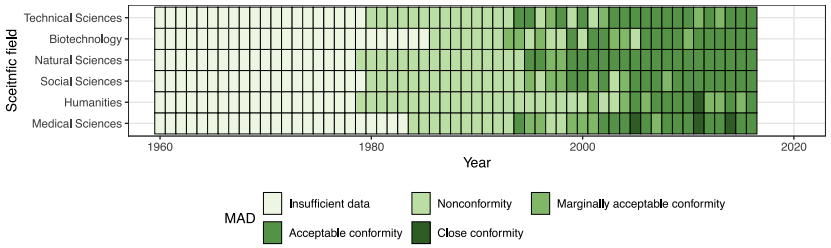

Fig. 8. Benford's conformity for each research field on a yearly basis.

Table 1
Conformity tests for number of contributions per author per decade.

Data set	Observations	Pearson's Chi-squared test		Mantissa Arc Test		MAD	MAD Conformity	Distortion Factor
		X-squared	*P*-value	L2	*P*-value			
1960–1970	38	17.502	0.02529	0.12376	0.00907	0.0653	Nonconformity	−51.41354
1970–1980	403	35.344	2.31E−05	0.046227	8.12E−09	0.0284	Nonconformity	−44.27548
1980–1990	1898	49.522	5.05E−08	0.0047154	0.0001298	0.0149	Marginal conformity	−31.19197
1990–2000	5176	13.556	0.09411	0.0011942	0.002068	0.0042	Close conformity	−12.50242
2000–2010	8819	13.872	0.08515	0.00040233	0.02878	0.0032	Close conformity	−7.73507
2010–2020	9660	13.909	0.08418	0.00024952	0.08978	0.0039	Close conformity	−6.949466

Table 2
Benford's conformity by field and typology. Abbreviations: H = Humanities, SS = Social Sciences, TS = Technical Sciences, B = Biotechnical Sciences, MNS = Mathematics and Natural Sciences, and MS = Medical Sciences.

Data set	Observations	Pearson's Chi-squared test		Mantissa Arc Test		MAD	MAD Conformity	Distortion Factor
		X-squared	*P*-value	L2	*P*-value			
H (1.01)	977	18.843	0.01572	0.002711	0.07074	0.0113	Acceptable conformity	−30.43495
H (1.08)	817	18.406	0.01838	0.0079669	0.00149	0.0156	Nonconformity	−44.13039
SS (1.01)	1401	38.374	6.42E−06	0.0067329	8.01E−05	0.0171	Nonconformity	−15.35622
SS (1.08)	1346	19.807	0.01109	0.0006882	0.396	0.0104	Acceptable conformity	−421.12554
TS (1.08)	2567	19.203	0.01381	0.00001248	0.9685	0.006	Close conformity	−14.40058
BS (1.01)	775	4.5918	0.8002	0.0011387	0.4137	0.005	Close conformity	−10.07059
BS (1.08)	713	29.099	0.0003046	0.012097	0.0001795	0.016	Nonconformity	−2.469761
MNS(1.08)	1477	19.085	0.01441	0.00020869	0.7347	0.01	Acceptable conformity	−19.83444
MNS(1.01)	1919	9.2685	0.3202	0.00064159	0.2919	0.0068	Acceptable conformity	−13.24065
MS (1.01)	1583	12.084	0.1475	0.0015583	0.08485	0.007	Acceptable conformity	−10.55757
MS (1.08)	1388	6.033	0.6435	0.00015432	0.8072	0.007	Acceptable conformity	−19.89672

8. Conclusion and further work

This paper proposes a methodology for following changes and assessing the maturity of research system using Benford's law. The paper identifies the type of data sets that can be used for this purpose. The presented method was evaluated on a real-world test case: the Slovenian research network. Research findings suggest that the method can be used to identify possible non-isolated cases of deviations. The method identifies groups and does not concentrate on individuals. To identify individual cases, a more granular inspection is needed by analysing suspect observations. We identify two approaches for narrowing the search space, namely analyzing individual research fields and temporal analysis. We conclude the method can be reliably used for individual research fields and shows some discrepancies in the social sciences and humanities, which can be attributed to field-specific publishing culture that requires a more fine-grained method to conform. Additionally, temporal analysis confirms that the method can be used reliably on 12-month intervals, provided the sample meets number of observation threshold. Temporal analysis hints at the possibility of using annual research reports to continuously monitor the behaviour of individual research fields.

Additionally, we observe an unexpected increase in the number of papers written between 1990–2000. We identified two possible hypothesis:

- In 1991 Slovenia became the first republic that split from Yugoslavia. The change in political structure had large implications in the funding, promotion, and general structure of the scientific network.
- In the 90s, the Internet became a widely used media of information exchange. This allowed researchers to bridge the physical gaps that existed before, and allowed for more international collaboration.

Additionally, temporal data suggests that in this decade the conformity improved considerably.

The paper also presents a case where the method was intentionally misused, thus producing misleading results. The importance of which is emphasised to avoid false positives. However, we are fully aware of potential shortcomings of using smaller data samples, and suggest the method be verified on countries with larger data sets.

All aggregated (and anonymized) data is available on Zenodo:[dataset] https://doi.org/10.5281/zenodo.3935770 to provide researchers the ability to replicate our experiment and reuse our data for additional research.

In the future, we would like to test the proposed method on a new test case; a selection of possible candidates are the French national repository HAL,[2] Current Research Information System in Norway (Cristin),[3] Hungarian Scientific Bibliography,[4] and Information system for research, experimental development and innovation research for Czech Republic.[5]

The authors also hypothesize that the method would perform as predicted on a group of authors that extends across national borders such as authors in a global scientific fields using methodology similar to (Zdravevski et al., 2019).

Author contributions

Conceived and designed the analysis: Aleksandar Tošić, Jernej Vičič
Collected the data: Aleksandar Tošić
Contributed data or analysis tools: Aleksandar Tošić, Jernej Vičič
Performed the analysis: Aleksandar Tošić, Jernej Vičič
Wrote the paper: Aleksandar Tošić, Jernej Vičič

Acknowledgment

The authors gratefully acknowledge the European Commission for funding the InnoRenew project (Grant Agreement #739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia for investment funding from the Republic of Slovenia and the European Union's European Regional Development Fund.

References

Alexander, J. C. (2009). *Remarks on the use of Benford's law*. Available at SSRN 1505147
Ariel Xu, Q., & Chang, V. (2020). Co-authorship network and the correlation with academic performance. *Internet of Things*, *12*, 100307. https://doi.org/10.1016/j.iot.2020.100307 https://www.sciencedirect.com/science/article/pii/S2542660520301396
Batagelj, V., & Maltseva, D. (2020). Temporal bibliographic networks. *Journal of Informetrics*, *14*, 101006. https://doi.org/10.1016/j.joi.2020.101006 http://www.sciencedirect.com/science/article/pii/S1751157719301439
Benavent-Pérez, M., Gorraiz, J., Gumpenberger, C., & de Moya-Anegón, F. (2012). The different flavors of research collaboration: A case study of their influence on university excellence in four world regions. *Scientometrics*, *93*, 41–58.
Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 551–572.
Berger, A., & Hill, T. P. (2011). A basic theory of Benford's Law. *Probability Surveys*, *8*, 1–126. https://doi.org/10.1214/11-PS175 http://projecteuclid.org/euclid.ps/1311860830
Bernstein, G. S. (1984). Scientific rigor, scientific integrity: A comment on sommer and sommer. *American Psychologist*, *39*, 1316. https://doi.org/10.1037/0003-066X.39.11.1316.a
Bhattacharya, S., & Kumar, K. (2008). Forensic Accounting and Benford's Law [In the Spotlight]. *IEEE Signal Processing Magazine*, *25* https://doi.org/10.1109/MSP.2007.914724, 152–150; http://ieeexplore.ieee.org/document/4472258/
Bohannon, J. (2013). Who's afraid of peer review? *Science*, *342*, 60–65. https://doi.org/10.1126/science.342.6154.60
Bordons, M., Gomez, I., Fernández, M., Zulueta, M., & Méndez, A. (1996). Local, domestic and international scientific collaboration in biomedical research. *Scientometrics*, *37*, 279–295.
Braun, T., & Glaenzel, W. (1996). International collaboration: Will it be keeping alive east European research? *Scientometrics*, *36*, 247–254.
Burke, J., & Kincanon, E. (1991). Benford's law and physical constants: The distribution of initial digits. *American Journal of Physics*, *59*, 952. https://doi.org/10.1119/1.16838
Cleary, R., & Thibodeau, J. C. (2005). Applying digital analysis using Benford's law to detect fraud: The dangers of type i errors. *AUDITING: A Journal of Practice & Theory*, *24*, 77–81. https://doi.org/10.2308/aud.2005.24.1.77
Cugmas, M., Ferligoj, A., & Kronegger, L. (2019). Scientific co-authorship networks. *Advances in Network Clustering and Blockmodeling*, 363–387.
Curk, L. (2019). Implementation of the evaluation of researchers' bibliographies in Slovenia. *Procedia Computer Science*, *146*, 72–83. https://doi.org/10.1016/j.procs.2019.01.082. , 14th International Conference on Current Research Information Systems, CRIS2018, FAIRness of Research Information. http://www.sciencedirect.com/science/article/pii/S1877050919300870
Drake, P. D., & Nigrini, M. J. (2000). Computer assisted analytical procedures using Benford's law. *Journal of Accounting Education*, *18*, 127–146.
Druica, E., Oancea, B., & Vâlsan, C. (2018). Benford's law and the limits of digit analysis. *International Journal of Accounting Information Systems*, *31*, 75–82. https://doi.org/10.1016/j.accinf.2018.09.004 http://www.sciencedirect.com/science/article/pii/S146708951730101X
Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, *5*, 17–34.
Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, *34*, 51–61.
Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of pubmed, scopus, web of science, and google scholar: Strengths and weaknesses. *The FASEB Journal*, *22*, 338–342.
Fanelli, D. (2013). Redefine misconduct as distorted reporting. *Nature*, *494*, 149–149.
Ferligoj, A., Kronegger, L., Mali, F., Snijders, T. A., & Doreian, P. (2015). Scientific collaboration dynamics in a national scientific system. *Scientometrics*, *104*, 985–1012.
Fewster, R. M. (2009). A simple explanation of Benford's law. *The American Statistician*, *63*, 26–32. https://doi.org/10.1198/tast.2009.0005
Godecharle, S., Nemery, B., & Dierickx, K. (2014). Heterogeneity in european research integrity guidance: Relying on values or norms? *Journal of Empirical Research on Human Research Ethics*, *9*, 79–90.
Golbeck, J. (2015). *Benford's law applies to online social networks*.. CoRR abs/1504.0. arXiv:1504.04387
Heneberg, P. (2016). From excessive journal self-cites to citation stacking: Analysis of journal self-citation kinetics in search for journals, which boost their scientometric indicators. *PLOS ONE*, *11*.

[2] HAL: https://hal.archives-ouvertes.fr/.
[3] Cristin: https://www.cristin.no/english/.
[4] MTMT database:https://m2.mtmt.hu/gui2/.
[5] Information system for research, experimental development and innovation research: https://www.rvvi.cz/riv.

Van der Heyden, M. A., van de Derks Ven, T., & Opthof, T. (2009). Fraud and misconduct in science: The stem cell seduction. *Netherlands Heart Journal*, *17*, 25–29.
Hickman, M. J., & Rice, S. K. (2010). Digital analysis of crime statistics: Does crime conform to Benford's law? *Journal of Quantitative Criminology*, *26*, 333–349. https://doi.org/10.1007/s10940-010-9094-6
Holub, H. W., Tappeiner, G., & Eberharter, V. (1991). The iron law of important articles. *Southern Economic Journal*, 317–328.
John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.
Karlovčec, M., & Mladenić, D. (2015). Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring. *Scientometrics*, *102*, 433–454.
Katz, J., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, *40*, 541–554.
Keathley-Herring, H., Van Aken, E., Gonzalez-Aleu, F., Deschamps, F., Letens, G., & Orlandini, P. C. (2016). Assessing the maturity of a research area: Bibliometric review and proposed framework. *Scientometrics*, *109*, 927–951.
Kenny, D. A. (2015). *Measuring model fit*. http://davidakenny.net/cm/fit.htm
Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications*, *132*, 86–103. https://doi.org/10.1016/j.jnca.2019.01.029 https://www.sciencedirect.com/science/article/pii/S1084804519300438
Korošec, A. (2014). Sicris, v3. *Organizacija Znanja*, *19*, 22–26. http://www.dlib.si/?URN=URN:NBN:SI:DOC-KY3LXZHR
Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*, 997–1004.
Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, *2*, 317–325.
Lužar, B., Levnajić, Z., Povh, J., & Perc, M. (2014). Community structure and the evolution of interdisciplinarity in Slovenia's scientific collaboration network. *PLOS ONE*, *9*, e94429.
Macrina, F. L. (2014). *Scientific integrity: Text and cases in responsible conduct of research*. John Wiley & Sons.
Mali, F., Pustovrh, T., Cugmas, M., & Ferligoj, A. (2018). The personal factors in scientific collaboration: Views held by slovenian researchers. *Corvinus Journal of Sociology and Social Policy*, *9*, 3–24.
Martin, B. R. (2013). Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy*, *42*, 1005–1014. https://doi.org/10.1016/j.respol.2013.03.011 http://www.sciencedirect.com/science/article/pii/S004873331300067X
Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*, 31–40.
Narin, F., & Whitlow, E. S. (1990). *Measurement of scientific cooperation and coauthorship in CEC-related areas of science*. Commission of the European Communities Directorate-General.
Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, *4*, 39–40. https://doi.org/10.2307/2369148
Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*, 5200–5205.
Nigrini, M. (2001). Digital analysis using benford's law: Tests and statistics for auditors. *EDPACS*, *28*, 1–2. https://doi.org/10.1201/1079/43266.28.9.20010301/30389.4
Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. *The Journal of the American Taxation Association*, *18*, 72.
Nigrini, M. J. (2012). *Benford's law: Applications for forensic accounting, auditing, and fraud detection*. John Wiley & Sons. https://www:wiley:com/en-us/Benford%27s+Law%3A+Applications+for+Forensic+Accounting%2C+Auditing%2C+and+Fraud+Detection-p-9781118152850
Nigrini, M. J. (2017). Audit sampling using benford's law: A review of the literature with some new perspectives. *Journal of Emerging Technologies in Accounting*, *14*, 29–46.
Ortega, J. L. (2014). Influence of co-authorship networks in the research impact: Ego network analyses from microsoft academic search. *Journal of Informetrics*, *8*, 728–737. https://doi.org/10.1016/j.joi.2014.07.001 https://www.sciencedirect.com/science/article/pii/S1751157714000613
Pelacho, M., Ruiz, G., Sanz, F., Tarancón, A., & Clemente-Gallardo, J. (2021). Analysis of the evolution and collaboration networks of citizen science scientific publications. *Scientometrics*, *126*, 225–257.
Peterson, M. (2007). Best practices for ensuring scientific integrity and preventing misconduct. *Essays on Ethical Standards*, *21*.
Pisanski, T., Pisanski, M., & Pisanski, J. (2020). A novel method for determining research groups from co-authorship network and scientific fields of authors. *Informatica*, *44*.
Ranstam, J., Buyse, M., George, S. L., Evans, S., Geller, N. L., Scherrer, B., et al. (2000). Fraud in medical research: An international survey of biostatisticians. *Contemporary Clinical Trials*, *21*, 415–427.
Rodela, R. (2016). On the use of databases about research performance: Comments on karlovčec and mladenić (2015) and others using the sicris database. *Scientometrics*, *109*, 2151–2157.
Sambridge, M., Tkalčić, H., & Jackson, A. (2010). Benford's law in the natural sciences. *Geophysical Research Letters*, *37*, 1–5.
Seljak, T., & Bošnjak, A. (2006). Researchers' bibliographies in cobiss. si. *Information Services & Use*, *26*, 303–308.
Shibayama, S., & Baba, Y. (2015). Impact-oriented science policies and scientific publication practices: The case of life sciences in Japan. *Research Policy*, *44*, 936–950.
Singleton, T. W. (2011). IT audit basics: Understanding and applying Benford's law. *Isaca Journal*, *3*, 6.
Sponholz, G. (2000). Teaching scientific integrity and research ethics. *Forensic Science International*, *113*, 511–514.
Zdravevski, E., Lameski, P., Trajkovik, V., Chorbev, I., Goleva, R., Pombo, N., et al. (2019). Automation in systematic, scoping and rapid reviews by an nlp toolkit: A case study in enhanced living environments. *Enhanced living environments*, 1–18.
Zhang, J. (2020). *Testing case number of coronavirus disease 2019 in china with Newcomb-Benford law*. arXiv:2002.05695

2.3 Paper 3

Title: Application of Benford's Law on Cryptocurrencies
Authors: Aleksandar Tošić, Jernej Vičič
Year: 2022
Journal: Journal of Theoretical and Applied Electronic Commerce Research
DOI: 10.3390/jtaer17010016
Link: https://www.mdpi.com/0718-1876/17/1/16/htm

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Application of Benford's Law on Cryptocurrencies

Jernej Vičič [1,2,*,†] and Aleksandar Tošić [1,3,†]

[1] Faculty of Mathematics Natural Sciences and Information Technologies, University of Primorska, 6000 Koper, Slovenia; aleksandar.tosic@upr.si
[2] Research Centre of the Slovenian Academy of Sciences and Arts, The Fran Ramovš Institute, 1000 Ljubljana, Slovenia
[3] InnoRenew CoE, 6310 Izola, Slovenia
* Correspondence: jernej.vicic@upr.si
† These authors contributed equally to this work.

Abstract: The manuscript presents a study of the possibility of use of Benford's law conformity test, a well proven tool in the accounting fraud discovery, on a new domain: the discovery of anomalies (possibly fraudulent behaviour) in the the cryptocurrency transactions. Blockchain-based currencies or cryptocurrencies have become a global phenomenon known to most people as a disruptive technology, and a new investment vehicle. However, due to their decentralized nature, regulating these markets has presented regulators with difficulties in finding a balance between nurturing innovation, and protecting consumers. The growing concerns about illicit activity have forced regulators to seek new ways of detecting, analyzing, and ultimately policing public blockchain transactions. Extensive research on machine learning, and transaction graph analysis algorithms has been done to track suspicious behaviour. However, having a macro view of a public ledger is equally important before pursuing a more fine-grained analysis. Benford's law, the law of first digit, has been extensively used as a tool to discover accountant frauds (many other use cases exist). The basic motivation that drove our research presented in this paper was to test the applicability of the well established method to a new domain, in this case the identification of anomalous behavior using Benford's law conformity test to the cryptocurrency domain. The research focused on transaction values in all major cryptocurrencies. A suitable time-period was identified that was long enough to present sufficiently large number of observations for Benford's law conformity tests and was also situated long enough in the past so that the anomalies were identified and well documented. The results show that most of the cryptocurrencies that did not conform to Benford's law had well documented anomalous incidents, the first digits of aggregated transaction values of all well known cryptocurrency projects were conforming to Benford's law. Thus the proposed method is applicable to the new domain.

Keywords: cryptocurrency; Benford's law; anomaly detection; method application

check for updates

Citation: Vičič, J.; Tošić, A. Application of Benford's Law on Cryptocurrencies. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 313–326. https://doi.org/10.3390/jtaer17010016

Academic Editor: Jani Merikivi

Received: 7 November 2021
Accepted: 8 February 2022
Published: 25 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Benford's law [1], also known as the first-digit law, has been widely used as a tool to discover anomalies in various data ranging from accounting fraud detection, stock prices, house prices to electricity bills, population numbers, natural phenomena, death rates and recently so popular COVID-19 cases reports. Cryptocurrencies, also referred to as Blockchain-based currencies or crypto coins, have become a global phenomenon known to most people.Throughout the paper we will rely on the definition presented by [2] (cryptocurrency). A cryptocurrency is in fact quite a narrow, albeit recognizable, description of a subset of an umbrella class of cryptoassets. While still somehow geeky and not understood by most people, banks, governments and many companies are aware of its importance.

Since the inception of Bitcoin, many alternative systems have been developed. Some remain blockchain-based, where transactions are stored and consequently timestamped

in blocks to create a canonical chain through consensus. Others employ a directed acyclic graph based data structures, where there is no single canonical chain. Instead, transactions reference and confirm previous transactions in order to increase the system's throughput by sacrificing some security features. Moreover, transaction structure can be changed to achieve privacy, i.e., using ring signatures in Monero [3]. Regardless of the underlying data structure, consensus mechanism, or network protocol, cryptocurrencies are decentralized and permissionless computer networks that maintain a transparent ledger of transactions. Unlike cryptocurrencies, where a user can have an arbitrary number of wallets (identities), centralized and permissioned systems are easier to monitor, detecting suspicious behaviour or anomalies where approaches are analogue to traditional banking systems, as users are assumed to have a verifiable identity.

A report from The World Economics Forum [4] predicts 10% of the global domestic product to be stored on blockchain based public ledgers. The growing interest has made many developers, research, and innovators dedicate their time in an effort to improve on the existing systems. The effects can be observed through the thousands of cryptocurrencies and networks that exist presently. The growing velocity of these networks further increases the risk for the regulator to protect the consumer and the stability of the financial system. The United Nations Office on Drugs and Crime estimated up to 5% of the global GDP of laundered money [5]. Assuming frauds grow in parallel with the velocity and total value locked in the underlying network, a method for fast and efficient anomaly detection is paramount. However, with the growth of innovation in this space, the techniques employed must search for a generic solution that makes little or no assumptions about the underlying network.

Our approach attempts to provide a technology agnostic tool to analyze open ledgers to alert of potential suspicious behaviour which requires further, more fine-grained analysis. Although more than 12 years have passed since the first transaction of the first cryptocurrency—Bitcoin (BTC) [6]—only the last few years have seen a big enough number of transactions and a large enough time frame for some statistical analysis to be carried out. Our research focused on empirical proof whether Benford law [1], a law of anomalous numbers, could be used in a non-altered form for discovering fraudulent or at least suspicious activity on cryptocurrencies in the same way it is used in standard financial forensics.

Although we could observe the cryptocurrency transactions as just another financial tool that should comply to all the used mechanisms (among them also the Benford law conformity for identifying frauds and other anomalous behavior), there are some properties that must be addressed or at least be observed:

- Mining transactions (mostly with mining pools) for all cryptocurrency assets that are based on the Proof of Work (PoW) [7] consensus mechanism, by which the cryptocurrency blockchain network achieves distributed consensus. Mining pools, where most of the miners are concentrated, pay out rewards to miners based on the computing power contributed. The payouts are mostly scheduled to occur once the miner is owed more than the threshold to save up on transaction fees. As many miners keep the default threshold, many transactions are possibly of the same value;
- Default transaction fees (GAS) are the same. GAS refers to the pricing value required to successfully conduct a transaction or execute a contract on the Ethereum blockchain platform.

The basic idea of the research was to test if Benford's law conformity can be used as a tool to detect anomalies in cryptocurrencies. The paper is structured as follows: Section 2 presents the basic properties of Benford's law and its usages, Section 3 presents the state of the art, followed by Methods and Materials in Section 4. The results are presented in Section 5 and are discussed in Section 6.

2. Benford's Law

Benford's law, also called the Newcomb–Benford law or the first-digit law, is an observation about the frequency distribution of leading digits. The observation was first

J. Theor. Appl. Electron. Commer. Res. **2022**, 17 315

discovered by [8] and later rediscovered by [1]. Benford's law defines a fixed probability distribution for leading digits of any kind of numeric data with the following properties [9]:

- Data with values from several distributions;
- Data that has a wide variety in the number of digits (e.g., data with plenty of values in the hundreds, thousands, tens of thousands, etc.);
- A data set that is fairly large, as a rule of a thumb consisting of at least 50–100 observations [10], although usually thousands of observations;
- Data is right-skewed (i.e., the mean is greater than the median), and the distribution has a long right-tail rather than being symmetric;
- Data has no predefined maximum or minimum value (with the exception of a zero minimum).

The distribution of digits is presented in Figure 1; the digit 1 occurs in roughly 30% of the cases, and the other digits follow in a logarithmic curve. It has been shown that this result applies to a wide variety of data sets [9]. Some examples are presented in Section 3. The equation for the distribution of the first digits of observed data is presented in Equation (1).

$$P(d) = log_{10}(d+1) - log_{10}(d) = log_{10}(1+\frac{1}{d}) \quad (1)$$

The quantity $P(d)$ is proportional to the space between d and $d+1$ on a logarithmic scale. Therefore, this is the distribution expected if the logarithms of the numbers (but not the numbers themselves) are uniformly and randomly distributed.

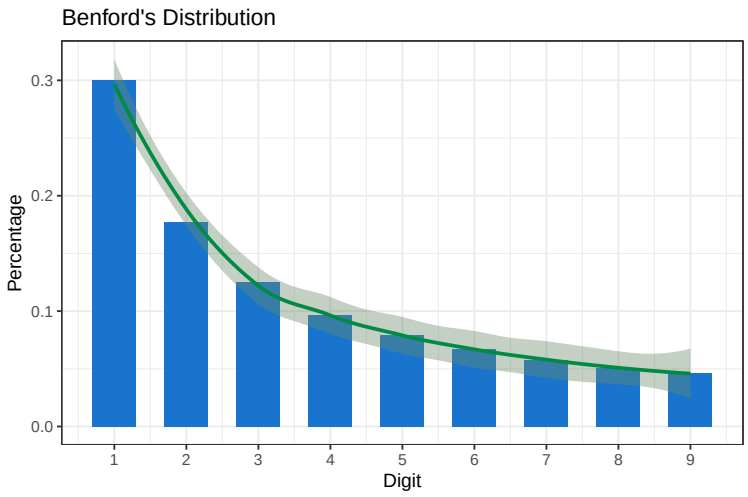

Figure 1. The distribution of digits in accordance to Benford's law [9]. Blue colored bars represent digits that conform to Benford's law.

3. State of the Art

Benford's law has been thoroughly researched and its theoretical grounds have been proved in many scientific papers. The methodology and basic mathematical grounds are discussed in greater detail by [11]. Many researchers have verified for themselves that the law is widely obeyed but have also noted that the popular explanations are not completely satisfying [12]. To the authors' knowledge, there has been no research in using Benford's law as a tool for the detection of anomalies in cryptocurrency transactions.

Benford's law has been extensively used in the accountant fraud detection and prevention, and there has been a lot of research in the area, such as [13,14], who present a literature overview of the area. Ref. [15] introduces Benford's Law and Digital Analysis (analysis of

digit and number patterns of a data set), which can be used as an analytical procedure and fraud detection tool. Ref. [16] presents Benford's law as a simple and effective tool for the detection of fraud. The purpose of the paper is to assist auditors in the most effective use of digital analysis based on Benford's law by identifying data sets, which can be expected to follow Benford's distribution, and presenting types of frauds that would be "detected/not detected" by such analysis. However, there are some research findings that point out some inherent problems that potentially arise in the use of Benford's law in the auditing process such as [17].

The simplicity of Benford's law as a tool allows for a broad range of uses. Ref. [18] examined crime statistics at the USA National, State, and local level in order to test the conformity to Benford's law distribution. Ref. [19] observed the distribution of initial digits of physical constants; however, their results were inconclusive.

One of the more recent researches involving Benford's law is [20]. The authors proposed a test of the reported number of cases of coronavirus disease in 2019 in China with Benford's law and report that the reported numbers of affected people abide to Benford's law.

Ref. [21] presented an overview of identified frauds that can be committed in the cryptocurency paradigm. Identified frauds include Ponzi schemes [22], fake initial coin offering schemes, pump and dump schemes, as well as cryptocurrency theft. Ref. [23] identified the main reasons for frauds and manipulation in cryptocurrencies: lack of consistent regulation, relative anonymity, low barriers of entry, exchange standards, and sophistication. Ref. [24] performed an end-to-end characterization of the counterfeit token in the Ethereum network, targeting Erc20 coins. Ref. [25] aimed to demonstrate that Bitcoin, the most known cryptocurrency, constitutes a substantial danger in terms of criminal enterprise. Ref. [26] presented an economic analysis of money laundering schemes utilizing cryptocurrencies, which aims at providing an answer to the open question of whether cryptocurrencies constitute a driver for money laundering. Ref. [27] proposed an approach to detect illicit accounts on the Ethereum blockchain using well proven machine learning techniques. Recent anomaly detection makes use of machine learning approaches. Support Vector Machines (SVM) were used to detect anomalies in the Bitcoin network [28]. However, the analysis is on the network level, and not on individual transactions. A clustering approach with Random Forest (RF) was used to detect wallets with anomalous behaviour [29]. However, the approach makes assumptions on the underlying structure of transactions to extract the features needed, and thereby lacks generality. A recent study showed that neural networks can be used to detect abnormalities with good stability and effectiveness, but the technique is limited to smart contract platforms, and not general transaction networks. Kamišalić et al. [30] presented a detailed overview of various techniques used for anomaly detection. This highlights the need for a simpler implementation agnostic technique for preliminary screening of public ledgers.

4. Methodology

As mentioned in Section 1, this paper proposes a methodology for identifying out-of-the-ordinary behavior and possibly detect frauds in blockchain-based currency. As such, the purpose is to present scientific grounds that allow feasibility and usefulness of the method as well as to propose a set of usage guidelines and a use case where our hypotheses were confirmed.

Our research experiment started with gathering all transactions on the Ethereum (ETH) network. Ethereum was chosen for these properties: It is one of the biggest cryptocurrencies by market capitalization and number of transactions processed; the network houses multiple cryptocurrencies (tokens) that could be compared directly (this part of the experiment is still open); and it is a well-documented and accessible blockchain. The first preliminary results revealed that transaction values (non-aggregated) of the whole Ethereum network do not conform to Benford's law [1] as is presented in Figure 2. Blue color depicts the leading digits that conform to Benford's law, red color depicts the non-conforming digits. The reasoning is further discussed in Section 4.1.

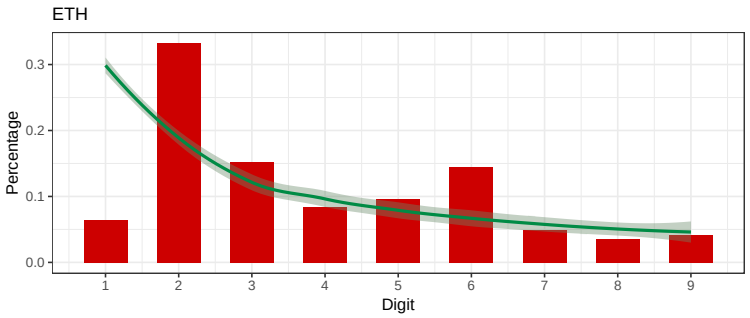

Figure 2. The leading digits of all ETH transaction values do not conform to Benford's law. The daily aggregated values conform to the same metric (see Figure 3), leading to a possible conclusion that there are too many automatic transactions in the network, but the aggregated values avoid this effect.

Although this does not mean that there was any artificial manipulation or any other kind of anomaly, we investigated further. According to [31] Benford's law metric can be used to achieve similar goals on aggregated data. We explored the same phenomenon on aggregated values (number of transactions in an observed period, aggregated transaction values, ...). Most of the aggregated values conform to Benford's law according to goodness of fit chi square (χ^2) test [32], which in most literature, such as [16], is considered as a suitable tool to test Benford's law conformity. We extended our research to all major cryptocurrencies with enough transactions in the selected time-period.

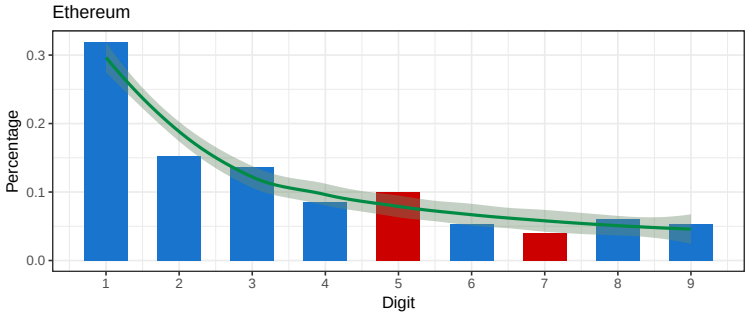

Figure 3. The leading digits of daily aggregated ETH transaction values in USD conform to Benford's law. Blue colored bars represent digits that conform and red colored bars represent digits that do not conform to Benford's law.

4.1. Methods

The observation sets need to conform to all the basic prerequisites for Benford's law as described in Section 2. This is the agenda for the executed research:

- Take all major cryptocurrencies into consideration;
- Express all aggregated daily transactions in one currency—we selected USD ($) as the most used fiat currency in comparisons;
- Select a viable observation period:
  - Starting date for each currency was the date of the first successful transaction;
  - Ending date for the observation period was set long enough into the past so that the frauds or abnormal behavior were well documented (in the forms of law-suits, scandals, vanished cryptocurrencies, well-documented special properties of specific currencies). We selected the year as the end of 2018, almost three years in the past;

- A long enough observation period that makes Benford's law conformity observation feasible (as presented in Section 2). In the body of surveyed literature, the sample size varies from 200 [33] to a few hundred thousand. We opted for doubling the minimum sample size—selecting all cryptocurrencies with 400 or more transaction days;
- Perform the MAD test [34] and classify all the cryptocurrencies according to [35] and visually observe all conformity graphs;
- Perform a literature review for all the currencies that do not conform to Benford's law and establish if there are any abnormalities documented for the selected time frame.

Testing conformity to Benford's law distribution has been done with many goodness of fit tests ranging from Pearson's Chi squared [36], Kolmogorov-Smirnov D statistics [37], Freedman's modification of Watson U^2 statistics [38], euclidean distance d statistics, and many others. However, no real data will ever follow the exact distribution; hence, most analysis supplements statistical testing with graphical representations that help in pointing out suspicious patterns in the data for further investigation. Additionally, different tests have different reactions on sample sizes. The Chi square test suffers from an excess power problem in that when the number of observations becomes large (above 5000 records estimated by [35]) it becomes more sensitive to insignificant spikes, leading to the conclusion that the data does not conform. Ref. [39] suggested that some statistical tests can render misleading results when applied to large number of observations. On the other hand, ref. [40] conclude that the Mean Absolute Deviation MAD test [34] is reliable with as low as 200 observations (as additional safety measure, we opted doubling that value to 400 in our experiment). Ref. [41] proposed the Mantissa Arc test, which is a very interesting geometrical test. Unfortunately, it tolerates little deviation from Benford's distributions.

Ref. [35] concluded that the best test is Mean Absolute Deviation (MAD), and a lot of the state-of-the art literature agrees with this proposal. Ref. [35] also presents a list of thresholds to classify the observed conformity:

- Conformity (0.000);
- Acceptable conformity (0.006);
- Marginally acceptable conformity (0.012);
- Nonconformity (0.015 and above).

The adapted MAD is used to measure the average deviation between the heights of the bars and the Benford line. The higher the MAD, the lower the conformity. We opted to perform conformity tests using all three of the aforementioned tests as our sample sizes are well within the acceptable ranges. All presented statistical tests are also supplemented with graphical representations; the results are presented in Section 5.

The Criteria That the Objects under Scrutiny Must Meet

Select a big enough set of aggregated data that conforms to Benford's law prerequisites described in Section 2. Observing only ledgers, the prerequisites that must be met are:

- The ledger must have support querying for transactions that contain the sending address, receiving address, amount, and timestamp;
- The assets being transferred must be denominated in any universally comparable form (any fiat currency (i.e., US Dollars) meets this criterion) at the time of transfer.

Count leading digits and perform Mean Absolute Deviation (MAD) conformity [35] on the gathered data. Plot simple bar charts with the numbers for each leading digit and visually and manually observe the distribution. If the data does not conform to Benford's law, investigate further.

4.2. Materials

DataHub cryptocurrency datasets (DataHub cryptocurrency datasets: https://datahub.io/cryptocurrency accessed on 1 March 2021) hosts daily aggregated data about all transac-

tions on all crypto coin networks from the first mined block on the Bitcoin network till the end of 2018. As such, it presents the perfect data source for our research. The problem that arises is how to get more recent data. The problem is further discussed in Section 6.

The data that support the findings of this study are openly available on Zenodo (Zenodo: https://zenodo.org/record/4682976 accessed on 1 January 2022, doi:10.5281/zenodo.4682976).

5. Results

This section presents the results of the experiment following the methodology from Section 4. All the figures in this section have the same format: a graph showing the distribution of leading digits. Red colored bars represent suspect values, which skew the distribution the most. Suspects are classified where the mean absolute deviation is above the threshold of 4. The threshold can be adjusted to increase the sensitivity. Suspects are useful as a starting point for further investigation in the case of nonconformity.

The time interval selected was between 2009 and 2018. Most of the cryptocurrencies were in an early development phase without a use-case or product, and consequently the amount of transactions recorded was negligible. Table 1 presents all cryptocurrencies that conformed to the prerequisites presented in Sections 2 and 4. The most discriminating factor in this phase was the minimum number of observations, which was set to 400 days (roughly double the minimal number of observations for Benford's law to be meaningful). This property eliminated all currencies that were started later than the last quarter of 2017. Each cryptocurrency is presented by its name and the ticker, number of observations (equal to the number of days), starting and ending date of the observation period and all the values from Benford's law conformance test. The currencies were grouped into four groups according to [35] and were also sorted according to this grouping from best to worst conformance.

All non-conformant cryptocurrencies were thoroughly observed and a list of publicly announced anomalies and even frauds was compiled for each of these cryptocurrencies. The two best performing and two cryptocurrencies with the biggest market cap were also observed in details. The results are presented in the remainder of the section. All the other cryptocurrencies can be further analyzed using the available accompanying data (Zenodo: https://zenodo.org/record/4682976 accessed on 1 January 2022, doi:10.5281/zenodo.4682976) in the raw aggregated data form, a list of Benford's law conformity values and charts.

Two "best conforming" cryptocurrencies, Ethereum classic (ETC) and Vertcoin (VTC), both still respectable projects, were classified as "Close conformity". The two biggest blockchain platforms regarding market capitalization, Bitcoin (BTC) and Ethereum (ETH), were classified as "Acceptable conformity" and "Marginally acceptable conformity", respectively. Figure 4 shows Benford's law conformance chart for further visual examination for all four cryptocurrencies.

Six of the currencies from Table 1 were classified as "non-conformant" to Benford's law: EOS (EOS), TENX token (TENX), Veritaseum (VERI), Basic Atention Token (BAT), PIVX (PIVX), and Dogecoin (DOGE). Each of the cryptocurrencies from this list will be presented and discussed.

J. Theor. Appl. Electron. Commer. Res. **2022**, *17* 320

Table 1. Conformity tests for all major cryptocurrencies in the observed time-period with more than 400 days of transactions on the blockchain. The records are sorted according to MAD Conformity column, from close conforming to nonconforming.

Currency	Obs.	Pearson's Chi-Squared Test		Mantissa Arc Test		MAD	MAD Conformity	Distortion Factor	Start Date	End Date
		X-Squared	*p*-Value	L2	*p*-Value					
Ethereum Classic (ETC)	750	1.766027	0.9873638	0.0000861	0.9374726	0.00351481	Close	−0.1409321	2015-07-30	2018-08-12
Vertcoin (VTC)	1666	7.30948	0.5036398	0.000165673	0.7588044	0.005795195	Close	−1.621333	2014-01-10	2018-08-12
Metal (MTL)	400	5.15115	0.7413057	0.001522525	0.543889	0.01089584	Acceptable	−0.1257945	2017-06-29	2018-08-12
Status (SNT)	411	7.692396	0.4640798	0.001050824	0.6492816	0.01005221	Acceptable	−1.560401	2017-06-19	2018-08-12
Aragon (ANT)	452	5.696092	0.6812311	0.005388913	0.08752867	0.01078389	Acceptable	3.448495	2017-05-15	2018-08-12
Waves (WAVES)	603	5.14964	0.7414692	0.003501951	0.1210349	0.008216651	Acceptable	−1.721726	2016-06-02	2018-08-12
Iconomi (ICN)	658	10.17673	0.2528404	0.0008252317	0.5810012	0.0104604	Acceptable	0.8235436	2016-09-30	2018-08-12
NEO (NEO)	665	3.823118	0.8727192	0.0008927334	0.5522979	0.006478303	Acceptable	1.316035	2016-09-09	2018-08-12
Lisk (LSK)	811	11.45478	0.1772377	0.001803885	0.231552	0.009606102	Acceptable	3.072645	2016-04-06	2018-08-12
Stellar (XLM)	1009	9.622045	0.2925614	0.002200075	0.1086226	0.007992198	Acceptable	1.221268	2014-08-05	2018-08-12
Verge (XVG)	1387	8.300241	0.4047048	0.002115592	0.05316656	0.007575786	Acceptable	−2.84182	2014-10-09	2018-08-12
MaidSafeCoin (MAID)	1560	10.43771	0.2356377	0.003279288	0.006001835	0.007513696	Acceptable	3.73407	2014-04-22	2018-08-12
Dash (DASH)	1641	5.958045	0.6519316	0.001418531	0.09750916	0.00621291	Acceptable	0.8615983	2014-01-19	2018-08-12
DigiByte (DGB)	1649	25.9	0.00111	0.003.21	0.005	0.01088511	Acceptable	−2.4136	2014-01-10	2018-08-12
Bitcoin (BTC)	1933	30.8193	0.0001512958	0.0006696828	0.2740357	0.01158613	Acceptable	5.881506	2013-04-28	2018-08-12
Gnosis (GNO)	468	8.754344	0.3634412	0.006937894	0.03889326	0.01312756	Marginally acc.	1.135551	2017-04-18	2018-08-12
Golem (GLM)	633	11.07461	0.1975074	0.003690431	0.0967096	0.0129236	Marginally acc.	6.131378	2016-11-11	2018-08-12
Zcash (ZEC)	653	20.82315	0.007632357	0.001029657	0.5104994	0.01293599	Marginally acc.	−0.9372237	2016-10-28	2018-08-12
Decred (DCR)	915	17.6832	0.02373108	0.0005975181	0.5788401	0.01375337	Marginally acc.	−1.586765	2016-02-08	2018-08-12
Ethereum (ETH)	1102	25.77399	0.00115	0.000378	0.658996	0.01482756	Marginally acc.	−0.08431323	2015-08-07	2018-08-12
NEM (XEM)	1230	27.13364	0.0006703807	0.008295528	0.000037	0.01417723	Marginally acc.	3.19854	2015-03-29	2018-08-12
Tether (USDT)	1258	34.91683	0.0000277	0.0138	0.00000003	0.01391653	Marginally acc.	−5.969747	2014-10-06	2018-08-12
EOS (EOS)	401	15.36398	0.05244271	0.003494984	0.2462301	0.0200535	Nonconformity	−2.819878	2017-06-20	2018-08-12
TENX token (TENX)	402	10.5	0.234	0.00808	0.0389	0.01539412	Nonconformity	−7.119347	2017-06-27	2018-08-12
Veritaseum (VERI)	431	11.32151	0.1841391	0.01211339	0.005402612	0.01726905	Nonconformity	−1.603899	2017-04-25	2018-08-12
Basic Atention T. (BAT)	438	19.05523	0.01456707	0.01293943	0.003456598	0.02196946	Nonconformity	0.2319942	2017-05-29	2018-08-12
PIVX (PIVX)	903	28.08438	0.0004584671	0.01199764	0.0000197	0.01890993	Nonconformity	−7.031687	2016-01-30	2018-08-12
Dogecoin (DOGE)	1702	83.1755	1.1210^{-14}	0.02422157	1.2510^{-18}	0.0214206	Nonconformity	−9.527495	2013-12-08	2018-08-12

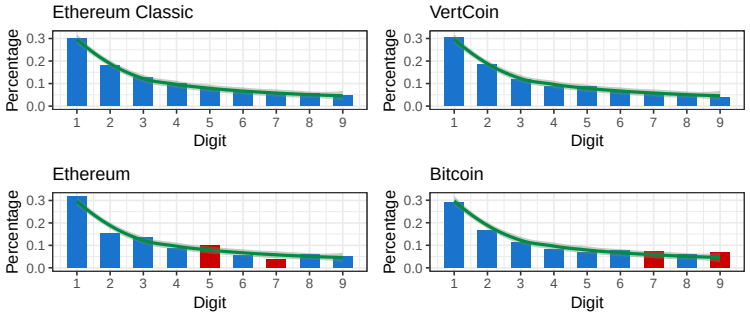

Figure 4. The two best conforming (ETC) and (VTC) currencies with "Close conformity" and the two biggest cryptocurrencies (BTC)—"Acceptable conformity" and (ETH)—"Marginally acceptable conformity" for aggregated value in USD transaction history.

5.1. TENX Token (TENX)

Figure 5 shows the TENX aggregated transactions and the conformance to Benford's law. The MAD value, a well documented Wirecard scandal (Crypto.com, TenX crypto debit cards were frozen following the Wirecard scandal: https://decrypt.co/33695/crypto-debit-cards-frozen-following-wirecard-scandal accessed on 1 March 2021) shows a possible reason for non-conformity.

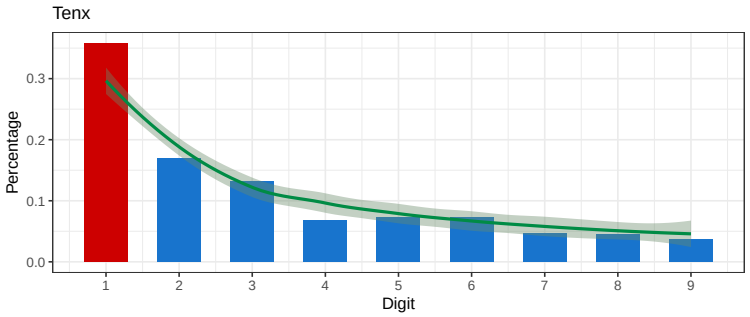

Figure 5. TENX aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 4 (almost) underflows. Overall, the daily aggregated transaction values do not conform.

5.2. Veritaseum (VERI)

Figure 6 shows the VERI aggregated transactions and the conformance to Benford's law.

The U.S. Securities and Exchange Commission (SEC) said it has reached a settlement with Reggie Middleton, organizer of the fraught $14.8 million Veritaseum (VERI) initial coin offering (ICO) (Analysis of the Veritaseum Scam: https://steemit.com/money/@financialcritic/analysis-of-the-veritaseum-scam accessed on 1 March 2021). The case was closed on October 2019, but the frauds were committed well within the observation period of our research.

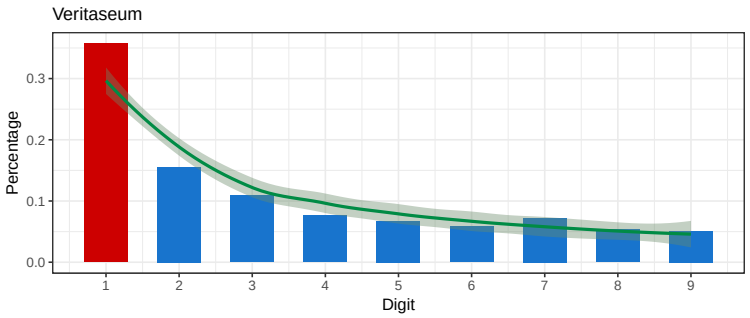

Figure 6. VERI aggregated transactions and the conformance to Benford's law. Digit 1 overflows. Overall the daily aggregated transaction values do not conform.

5.3. Dogecoin (DOGE)

Figure 7 shows the DOGE aggregated transactions and the conformance to Benford's law. The coin was introduced as a satire initially in December 2013 and included an image of the Doge meme as its logo. The author of this coin/crypto currency revealed this motivation publicly. Some properties showing the soundness of our decision are as follows:

- On the 24 September 2018 (a randomly chosen date on a working day at the end of our observation period): the last tweet from the official Tweeter account on 14 July 2018 (80 days) (Dogecoin twitter account: https://twitter.com/Dodgecoin accessed on 1 March 2021);
- Fun and friendly internet currency, the dogecoin logo is a dog from a meme;
- 24 h trading volume on all exchanges according to CoinCodex (Concodex: https://coincodex.com/crypto/XXX/exchanges/ accessed on 1 March 2021) was USD 42.51 million dollars.

In the last years Dogecoin has gained a lot of positive reputation as being a "lost cause" founding platform, and, especially in 2021, the value of the coin has seen a rapid increase in price with the help of celebrity exposure [42]. However, these recent developments were excluded from our analysis as we fixed the observation period from the start of the crypto-assets till the end of 2018.

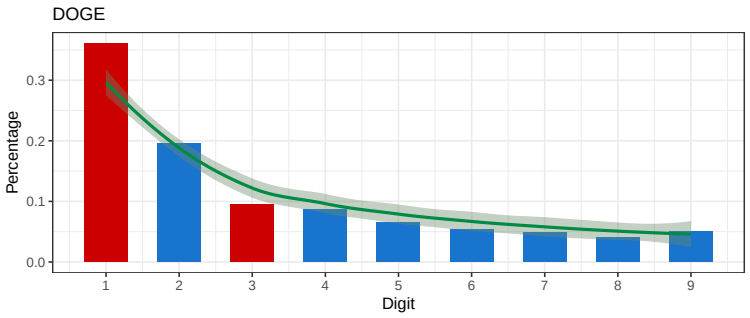

Figure 7. DOGE aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 3 underflows. Overall the daily aggregated transaction values do not conform.

5.4. Basic Attention Token (BAT)

Figure 8 shows the BAT aggregated transactions and the conformance to Benford's law. The transactions of the BAT coin are mostly automatically generated as this coin is the basis of a digital marketing platform that periodically rewards users for participation, and as such break Benford's law prerequisites.

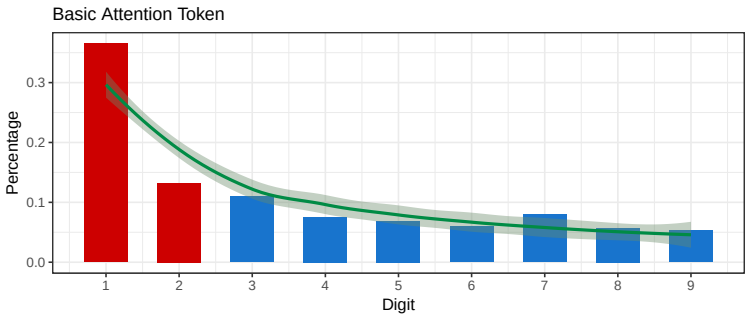

Figure 8. BAT aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 2 underflows, digit 7 (almost) overflows. Overall the daily aggregated transaction values do not conform.

5.5. PIVX (PIVX)

Figure 9 shows the PIVX aggregated transactions and the conformance to Benford's law. There was no scandal reported for the PIVX project in the observation period (in fact, the authors could not find any notable anomaly for this cryptocurrency). The only speculation that the authors could give is that the PIVX network relies on anonymous transactions that could be used to hide anomalies.

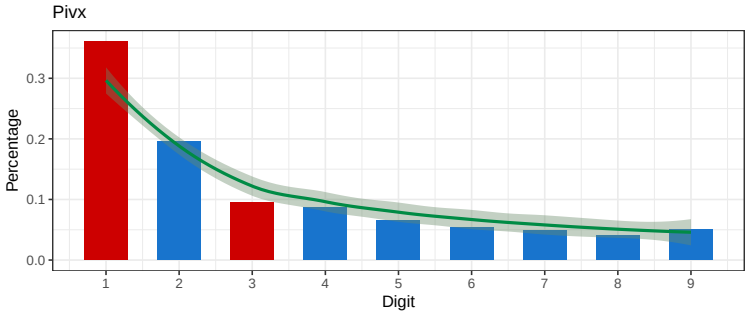

Figure 9. PIVX aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 3 underflows. Overall the daily aggregated transaction values do not conform.

5.6. EOS (EOS)

Figure 10 shows the EOS aggregated transactions and the conformance to Benford's law. EOS is regarded as a valid project and survived until 2021. The only drawback is that in 2018, the project was in the starting phase and the backing capital risen by the backers of the project was an order of magnitude bigger than what the proposed project promised to accomplish ("Why EOS Failed to Kill Ethereum: The Fatal Flaw of Centralization in a Decentralized Market": https://coincodex.com/article/10454/why-eos-failed-to-kill-ethereum-the-fatal-flaw-of-centralization-in-a-decentralized-market/ accessed on 1 March 2021).

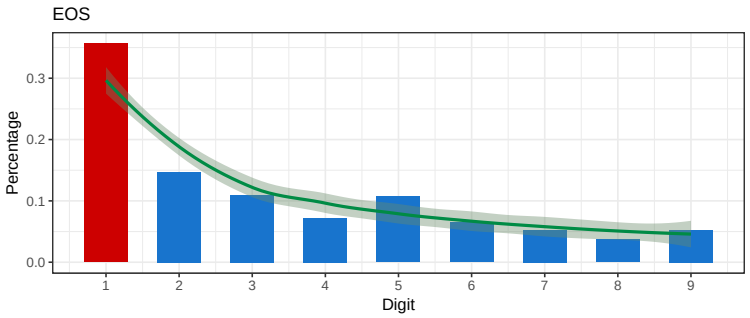

Figure 10. TENX aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digits 2 and 4 (almost) underflow. Overall the daily aggregated transaction values do not conform.

5.7. Additional Currencies

An examination of all remaining cryptocurrencies that did not meet the criteria presented in Section 4, mainly due to the lack of data, show additional cases that support the validity of the presented method. By lowering the requirement for the minimum number of observations to 300 days, we can observe additional cryptocurrencies that do not conform to Benford's law that have documented scams and scandals attributed to the observation period, such as: the Enigma (ENG) (Enigma Ethereum marketplace was hijacked, its investors duped by phishing scam: https://www.zdnet.com/article/enigma-ethereum-marketplace-hijacked-by-attackers/ accessed on 1 March 2021); SALT (SALT) (SALT COIN EXIT SCAM! Massive selloff predicted by Morgan Stanley: https://www.youtube.com/watch?v=E2iNt3Z6qaY accessed on 1 March 2021); and Waltonchain (WTC) (Monumentall stupid tweet blows up in blockchain company's face: https://mashable.com/2018/02/28/waltonchain-twitter-scam-wtc/?europe=true accessed on 1 March 2021).

6. Discussion and Future Work

The main goal of the presented research was to test the applicability of Benford's law to the cryptocurrency transaction networks as a preliminary screening tool. The research focused on some well-documented anomalies and frauds from the past and compared the proposed metric on proven ecosystems that performed normally in the same time period. We focused on the time period between 2009 (time of the first transaction on the Bitcoin network) and 2018, as there were already enough transactions to meet all of Benford's law prerequisites, but also enough time had passed so that the anomalies and frauds had already emerged to the public.

The results show that the proposed method is suitable for the proposed domain. All the big blockchain platforms by market capitalization that were not biased by any big scandal or lawsuit and that are still functioning three years after the observation timeframe, such as Bitcoin (BTC), Ethereum (ETH), or OmiseGo (OMG), conform to Benford's law. However, failing to comfort to Benfords distribution does not necessarily imply fraud. The method can produce false positives in the form of non-conformity of a cryptocurrency and no particular fraudulent reason can be found. This can result from the nature of the transactions of the observed currency. The method does not find the actual anomaly, but it can be used as a preliminary screening that should always lead into fine-grained methods such as Machine Learning methods and graph-based searching. The inspection of the six cryptocurrencies that were classified as non-conforming to Benford's law revealed three currencies with well-documented anomalies: two (TENX and VERI) were tainted by scandals and lawsuits and one (DOGE) was invented as a joke—and in the first years it was regarded so. As an additional observation, Dogecoin is now a respected cryptocurrency and in the last year grew to USD $50B market capitalization. The method is obviously not

suitable to predict the future of an observed cryptocurrency. The transactions of the BAT coin are mostly automatically generated, as this coin is the basis of a digital marketing platform. The two remaining cryptocurrencies that were identified by the method as possible candidates for anomalous behaviour were EOS and PIVX, and although we could speculate to some extension why these two did not conform to Benford's law, the results are inconclusive.

All major cryptocurrencies that existed in the selected time-frame (2009–2018) were tested for the conformity to Benford's law. The data availability statement is presented in Section 4.2.

Future work, which is already underway, will focus on newer data. One such possible source has already been identified: Kaggle (Cryptocurrency Historical Prices: https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory accessed on 1 March 2021). Another open issue that can be tackled with the same methodology is a comparison of all ERC20 tokens [43]. Ethereum-based cryptocurrencies were selected to ensure a common (thus fair) technical basis—all these cryptocurrencies use the same technological platform, so all possible reasons for differences that arise from basic technology are eliminated.

Author Contributions: Conceptualization, A.T. and J.V.; methodology: A.T. and J.V.; software, A.T.; validation, A.T. and J.V.; formal analysis, A.T. and J.V.; investigation, A.T. and J.V.; funding acquisition and resources, J.V.; data curation, A.T.; writing–original draft preparation, A.T. and J.V.; writing–review and editing, A.T. and J.V.; visualization, A.T. and J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by H2020 grant number 739574 and by the Slovenian Research Agency (ARRS) grant number J2-2504.

Institutional Review Board Statement: The data gathering process did not involve the use of human subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available on Zenodo (Zenodo: https://zenodo.org/record/4682976 accessed on 1 January 2022, doi:10.5281/zenodo.4682976).

Acknowledgments: The authors gratefully acknowledge the European Commission for funding the InnoRenew project (Grant Agreement #739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Regional Development Fund). They also acknowledge the Slovenian Research Agency ARRS for funding the project J2-2504.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Benford, F. The law of anomalous numbers. *Proc. Am. Philos. Soc.* **1938**, *78*, 551–572.
2. Lansky, J. Possible state approaches to cryptocurrencies. *J. Syst. Integr.* **2018**, *9*, 19–31. [CrossRef]
3. Noether, S. Ring SIgnature Confidential Transactions for Monero. *IACR Cryptol. ePrint Arch.* **2015**, *2015*, 1098.
4. Mettler, M. Blockchain technology in healthcare: The revolution starts here. In Proceedings of the 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, 14–16 September 2016; pp. 1–3.
5. Campbell-Verduyn, M. Bitcoin, crypto-coins, and global anti-money laundering governance. *Crime Law Soc. Chang.* **2018**, *69*, 283–305. [CrossRef]
6. Nakamoto, S. Bitcoin Whitepaper. Technical Report. 2008. Available online: Bitcoin.org (accessed on 1 March 2021).
7. Jakobsson, M.; Juels, A. Proofs of work and bread pudding protocols. In *Secure Information Networks*; Jakobsson, M., Juels, A., Eds.; Springer: Leuven, Belgium, 1999; pp. 258–272.
8. Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. *Am. J. Math.* **1881**, *4*, 39–40. [CrossRef]
9. Singleton, T.W. IT Audit Basics: Understanding and Applying Benford's Law. *Isaca J.* **2011**, *3*, 6.
10. Kenny, D.A. Measuring Model Fit. 2015. Available online: http://davidakenny.net/cm/fit.htm (accessed on 1 March 2021).
11. Berger, A.; Hill, T.P. A basic theory of Benford's Law. *Probab. Surv.* **2011**, *8*, 1–126. [CrossRef]
12. Fewster, R.M. A Simple Explanation of Benford's Law. *Am. Stat.* **2009**, *63*, 26–32. [CrossRef]
13. Kumar, K.; Bhattacharya, S. Detecting the dubious digits: Benford's law in forensic accounting. *Significance* **2007**, *4*, 81–83. [CrossRef]

14. Nigrini, M.J. Audit sampling using Benford's law: A review of the literature with some new perspectives. *J. Emerg. Technol. Account.* **2017**, *14*, 29–46. [CrossRef]
15. Drake, P.D.; Nigrini, M.J. Computer assisted analytical procedures using Benford's Law. *J. Account. Educ.* **2000**, *18*, 127–146. [CrossRef]
16. Durtschi, C.; Hillison, W.; Pacini, C. The effective use of Benford's law to assist in detecting fraud in accounting data. *J. Forensic Account.* **2004**, *5*, 17–34.
17. Cleary, R.; Thibodeau, J.C. Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors. *Audit. J. Pract. Theory* **2005**, *24*, 77–81. [CrossRef]
18. Hickman, M.J.; Rice, S.K. Digital Analysis of Crime Statistics: Does Crime Conform to Benford's Law? *J. Quant. Criminol.* **2010**, *26*, 333–349. [CrossRef]
19. Burke, J.; Kincanon, E. Benford's law and physical constants: The distribution of initial digits. *Am. J. Phys.* **1991**, *59*, 952. [CrossRef]
20. Zhang, J. Testing Case Number of Coronavirus Disease 2019 in China with Newcomb-Benford Law. *arXiv* **2020**, arXiv:2002.05695.
21. Baum, S.C. Cryptocurrency Fraud: A Look into the Frontier of Fraud. Ph.D. Thesis, Georgia Southern University, Statesboro, GA, USA, 2018.
22. Zuckoff, M. *Ponzi's Scheme: The True Story of a Financial Legend*; Random House Incorporated: New York, NY, USA, 2006.
23. Twomey, D.; Mann, A. Fraud and manipulation within cryptocurrency markets. In *Corruption and Fraud in Financial Markets: Malpractice, Misconduct and Manipulation*; Alexander, C., Cumming, D., Eds.; Wiley: Hoboken, NJ, USA, 2020; pp. 205–250.
24. Gao, B.; Wang, H.; Xia, P.; Wu, S.; Zhou, Y.; Luo, X.; Tyson, G. Tracking Counterfeit Cryptocurrency End-to-end. *Proc. ACM Meas. Anal. Comput. Syst.* **2020**, *4*, 1–28. [CrossRef]
25. Brown, S.D. Cryptocurrency and criminality: The Bitcoin opportunity. *Police J.* **2016**, *89*, 327–339. [CrossRef]
26. Brenig, C.; Müller, G. *Economic Analysis of Cryptocurrency Backed Money Laundering*; ECIS 2015 Completed Research Papers; Association for Information Systems: Atlanta, GA, USA, 2015; pp. 1–18.
27. Farrugia, S.; Ellul, J.; Azzopardi, G. Detection of illicit accounts over the Ethereum blockchain. *Expert Syst. Appl.* **2020**, *150*, 113318. [CrossRef]
28. Sayadi, S.; ben Rejeb, S.; Choukair, Z. Anomaly Detection Model Over Blockchain Electronic Transactions. In Proceedings of the 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; pp. 895–900. [CrossRef]
29. Baek, H.; Oh, J.; Kim, C.Y.; Lee, K. A Model for Detecting Cryptocurrency Transactions with Discernible Purpose. In Proceedings of the 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Zagreb, Croatia, 2–5 July 2019; pp. 713–717. [CrossRef]
30. Kamišalić, A.; Kramberger, R.; Fister, I. Synergy of Blockchain Technology and Data Mining Techniques for Anomaly Detection. *Appl. Sci.* **2021**, *11*, 7987. [CrossRef]
31. Shi, J.; Ausloos, M.; Zhu, T. Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries. *Phys. A Stat. Mech. Its Appl.* **2018**, *492*, 878–888. [CrossRef]
32. Fischer, R.A. *Statistical Methods for Research Workers*; Oliver and Boyd: Edinburgh, UK, 1925.
33. Carslaw, C.A. Anomalies in income numbers: Evidence of goal oriented behavior. *Account. Rev.* **1988**, *63*, 321–327.
34. Gorard, S. Revisiting a 90-year-old debate: The advantages of the mean deviation. *Br. J. Educ. Stud.* **2005**, *53*, 417–430. [CrossRef]
35. Nigrini, M.J.M.J. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*; Wiley: Hoboken, NJ, USA, 2012; p. 352.
36. Pearson, K.X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [CrossRef]
37. Berger, V.W.; Zhou, Y. *Kolmogorov–Smirnov Test: Overview*; Wiley Statsref: Statistics Reference Online: Hoboken, NJ, USA, 2014.
38. Freedman, L.S. Watson's UN2 statistic for a discrete distribution. *Biometrika* **1981**, *68*, 708–711. [CrossRef]
39. Nigrini, M. Digital Analysis Using Benford's Law: Tests and Statistics for Auditors. *EDPACS* **2001**, *28*, 1–2. [CrossRef]
40. Druică, E.; Oancea, B.; Vâlsan, C. Benford's law and the limits of digit analysis. *Int. J. Account. Inf. Syst.* **2018**, *31*, 75–82. [CrossRef]
41. Alexander, J.C. Remarks on the Use of Benford's Law. 2009. Available online: http://dx.doi.org/10.2139/ssrn.1505147 (accessed on 1 March 2021).
42. Livni, E. Serious money is flowing to the joke cryptocurrency Dogecoin. *New York Times*, 2 August 2021; pp. 1–2.
43. Somin, S.; Gordon, G.; Altshuler, Y. Network analysis of erc20 tokens trading on ethereum blockchain. In *International Conference on Complex Systems*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 439–450.

2.4 Paper 4

Title: A Decentralized Authoritative Multiplayer Architecture for Games on the Edge
Authors: Aleksandar Tošić, Jernej Vičič
Year: 2021
Journal: Computing and Informatics
DOI: 10.31577/cai_2021_3_522
Link: http://147.213.75.17/ojs/index.php/cai/article/view/2021_3_522

Computing and Informatics, Vol. 40, 2021, 522–542, doi: 10.31577/cai_2021_3_522

A DECENTRALIZED AUTHORITATIVE MULTIPLAYER ARCHITECTURE FOR GAMES ON THE EDGE

Aleksandar Tošić

University of Primorska, The Andrej Marušič Institute
Muzejski Trg 2, 6000 Koper, Slovenia
&
Innorenew CoE
e-mail: aleksandar.tosic@upr.si

Jernej Vičič

University of Primorska, The Andrej Marušič Institute
Muzejski Trg 2, 6000 Koper, Slovenia
&
Research Centre of the Slovenian Academy of Sciences and Arts
The Fran Ramovš Institute
e-mail: jernej.vicic@upr.si

Abstract. With the ever growing number of edge devices, the idea of resource sharing systems is becoming more appealing. Multiplayer games are a growing area of interest due to the scalability issues of current client-server architectures. A paradigm shift from centralized to decentralized architectures that would allow greater scalability has gained a lot of interest within the industry and academic community. Research on peer to peer network protocols for multiplayer games was mainly focused on cheat detection. Previously proposed solutions address the cheat detection issues on a protocol level but do not provide a holistic solution for the architecture. Additionally, existing solutions introduce some level of centralization, which inherently introduces single point of failures. We propose a blockchain-based, completely decentralized architecture for edge devices with no single point of failure. Our solution relies on an innovative consensus mechanism based on verifiable delay functions that additionally allows the network to derive verifiable randomness.

We present simulation results that show the assignment of players and referees to instances is pseudo-random, which inherently prevents collusion-based cheats and vulnerabilities.

Keywords: Edge computing, consensus, peer to peer, network protocol, multiplayer games, blockchain

Mathematics Subject Classification 2010: 68T50

1 INTRODUCTION

The gaming industry is worth almost 135 billion at the time of writing [7]. The same source predicts a steady 10 % growth in the next 2 years, reaching 180 billion by the end of 2021. The recent trends toward multiplayer games have been very successful with games like Fortnight earning more than 2.4 billion in revenue in 2018 alone [26]. Steam, the biggest game distribution platform reported it serves as much as 18.5 million clients concurrently. Cloud computing enabled servers need to be migrated real time in order to meet the demand of clients. Additionally, network latency was reduced due to localisation approaches where servers are spawned geographically close to clients if possible. However, maintaining a player base of thousands or even millions together with the hardware and software infrastructure is both very expensive and difficult to maintain [29]. The recent idea of a "sharing economy" can be applied in tandem with the paradigm shift to edge computing. More specifically, clients on the edge of the system can profit from sharing resources, such as bandwidth and computing power, thereby releasing the burden on centralized servers.

This can be achieved by using a peer to peer (P2P) architecture. P2P gaming architectures have been studied extensively but have not been widely adopted [29]. The main issues are closely related to the lack of authority and trust. Centralized architectures solve these issues with authoritative servers. The server's tasks are to simulate game play, validate and resolve conflict in the simulation, and store the game state. P2P multiplayer architectures were previously able to address some of the cheating vectors but required some level of centralization.

More recently, blockchain technology has gained a significantly large interest. Research in cryptography and fault tolerant consensus mechanisms has been driving the evolution of decentralized P2P systems.

The already available schemes that, at least theoretically, address most of the identified cheats in distributed gaming architectures RACS [28] and Goodman [12] still retain a central authority either to store the game state or as a refereeing authority and, thus, still retain the Single Point Of Failure – (SPOF) [8] property. Our research presented in this paper mostly focuses in the elimination of the SPOF but still being able to successfully address the same set of cheats. We were able to

achieve the set goals and were even able to partially address the Collusion cheat, as described in Section 5.7.

2 STATE OF THE ART

Baughman et al. [3] propose an improvement of their lock-step protocol [2] Asynchronous Synchronization (AS), the first protocol for providing cheat-proof and fair play-out of centralized and distributed network games. The protocol also provides implicit robustness in the face of packet loss. At proving the correctness of their approach, they make a number of assumptions: there exists a reliable channel between all players; all players know of all other players; players are able to authenticate messages from each other player; and all players wait only a finite time before making decisions and revealing commitments. Their approach can be implemented in a true peer-to-peer fashion, thus eliminating the SPOF, but, as it can be seen from Table 1, the approach is not immune to Replay/Spoof cheat [28].

GauthierDickey et al. [11] present a protocol designed to improve on lock-step protocol [2] by reducing latency while continuing to prevent cheating. They achieve this by adding a voting mechanism to compensate for packet loss in the environment. They call this protocol New Event Ordering (NEO).

Corman et al. [6] present SEA protocol and argue that it outperforms NEO algorithm in all cheat prevention properties; further, they present three possible cheats that the NEO protocol fails to address: Attacker can replay updates for another player. Attacker can construct messages with any previously seen votes attached. Since the votes are signed, the messages will appear to come from another player. Attacker can send different updates to different opponents.

Cronin et al. [14] present SP protocol which addresses the late-commit cheat and presents a performance improvement on the existing protocols (lock-step).

Goodman [12] proposes IRS hybrid C/S – P2P design; it operates by routing request messages through a centralized server and relaying them to proxy clients, a secure method by which it is certain that the requesting and proxy clients received the same message. The proxy clients perform calculations for others, relieving the server of the calculation burden. The code of the IRS approach relies on identifying malicious clients. This can be done with a certain probability and can still lead to cheat exploits.

Pellegrino and Dovrolis [20] propose a change from Client-Server architecture to Peer-to-Peer with Central Arbiter architecture (PP-CA) that contains server bandwidth requirements when increasing number of players, effectively solving the biggest scalability problem. The system still retains the SPOF in the form of the centralized arbiter. The paper focuses entirely on the elimination of the bandwidth problem and does not deal with any cheats; actually, it introduces a new form of cheating (e.g., blind opponent – BO, discussed in Section 3).

Webb et al. [28] compare all algorithms and show that all the previous distributed protocols and schemes are vulnerable to several cheats. Their proposed

scheme (RACS), which extends PP-CA [20], solves most of the problems but still has the SPOF in the form of the Identity server and Referee: a process running on a trusted host that has authority over the game state.

Most of the presented protocols are SPOF-free as they address only the P2P communication protocol but are vulnerable to cheats, as can be seen on Table 1. The RACS and IRS address most of the cheats but reintroduce the SPOF. Our scheme eliminates the SPOF problem and still retains all the properties described in RACS [28] and at least partially deals with the Collusion cheat that, at least in our opinion, cannot be eliminated by means of protocols and technology.

3 CHEAT TAXONOMY

In this article, we use the definition of Yan and Randel [30] for online game cheating: "Any behaviour that a player uses to gain an advantage over his peer players or achieve a target in an online game is cheating if, according to the game rules or at the discretion of the game operator (i.e., the game service provider, who is not necessarily the developer of the game), the advantage or the target is one that he is not supposed to have achieved." Cheaters try to gain unfair advantage over other players. This can totally destroy the in-game economics of an online game or simply ruin the gaming experience. Grievers, as the name implies, are players with the sole intention of hurting other players' experience as much as possible. When this behaviour adheres to game rules, it is technically not a form of cheating and is out of scope of this paper. Both groups exploit the same set of cheats.

The taxonomy presented in Table 1 and accompanying text and later used in this paper follows the taxonomy presented in Web et al. [28] and Yahyavi et al. [29], we added 3 additional entities, 2 were not addressed by the previous research (Robustness and User data privacy), the last one (Lack of Devices Situation) is a consequence of our approach and not applicable to other architectures. It is addressed later in this section. Multiple authors addressed the issue of systematic classification of cheating in online games, such as Yan and Randel [30] who present a taxonomy of 15 types of cheats and just by comparing the number of entries we could assume that the later introduces additional forms. We argue that the set of cheats presented in our paper fully covers the whole set presented in Yan and Randel [30] with the addition of two new entries that are discussed separately. The translations are presented in Table 1 in the second column. The "Cheating by compromising passwords" can be classified as "Social engineering" class and as such omitted. We argue that these two entries cannot be successfully addressed by the game architecture, they must be addressed mostly by informing the players.

1. *Bug* – bugs in games can lead to potential misuse by the players. No scheme directly addresses this problem, it is assumed that the bugs will be fixed by software developers.
2. *IE, IC* – the goal of IE (Information Exposure) is to obtain secret information to which the cheater is not entitled, thus gaining an unfair advantage in selecting

	Yan	**Problem/Cheat**	**RACS**	**IRS**	**C/S**	**AS**	**NEO**	**Damage**
1	L	Bug	√	√	√	√	√	√
2	A, F, H	IE, IC	√	√	√	×	×	√
3		Bots	×	×	×	×	×	×
4	A, C, F	Supp. update, TS, FD	√	√	√	√	√	√
5	A, F, H	Replay, Spoofing	√	√	√	×	√	√
6	A, D, F, H	Undo	n/a	n/a	n/a	√	×	n/a
7	A, C, F, H	BO	√	√	n/a	n/a	n/a	√
8	G	DDoS	√	√	√	√	√	√
9	B	Collusion	×	×	×	×	×	√[1]
10	M,	Robustness	×	×	×	×	×	√
11	J	User data privacy	√	√	√	×	×	√
12		Lack of Devices Situation	n/a	n/a	n/a	n/a	n/a	√
13	E	Exploiting AI	n/a	n/a	n/a	n/a	n/a	n/a
14	I, O	Social Engineering	n/a	n/a	n/a	n/a	n/a	n/a^*

Table 1. Chart presenting all identified cheats and schemes for detection and removal. The n/a is given to a scheme that does not have to deal with the observed cheat for implicit reasons (mostly architectural).

the optimal action. The IC (Invalid Command) cheat occurs when an application or data files are modified to issue commands or command parameters that originally could not be generated.

3. *Bots* – programs that act as players can be introduced to the game. The programs can exercise number of cheats including Collusion.
4. *Supp. update, TS, FD* – A cheater can suppress sending the state update, send the updates at a slower rate (FD) to gain advantage or incorrectly timestamp messages to gain advantage.
5. *Replay, Spoofing* – a player can obtain advantage by replicating messages by means of local software instead of using the tools provided by the game (example: sending impossibly fast series of missiles).
6. *Undo* – a player succeeds to undo a previously sent message after already receiving the opponents message and realizing that the original message was not optimal.
7. *BO* – in distributed schemes that use a Central Arbiter (CA), such as PP-CA [20], cheater may purposely withhold updates to his peers (but not to the CA), effectively covering own actions.
8. *DDoS* – a (cheating) player may use DDoS [17] attack to temporally disable the opponent to send messages and thus get advantage.
9. *Collusion* – unfavorable situation may occur whereby certain clients cooperate with one another in order to gain unfair advantage over others. Collusion via the use of external communication is difficult to eliminate due to the use of non-monitored means of communication [12].

10. *Robustness* – The robustness metric shows how much is the system or scheme or protocol fault tolerant (how much it is tolerant to node failure, at the worst to server node failure). The basic Client-Server architecture is the least robust and totally decentralized system is the most robust.
11. *User Data Privacy* – user profiles with scoring and possibly in-game funds and purchases are stored for future use. Client Server architectures use the server as a means for reliable storage, distributed systems have to deal with security risks as the data is spread on the network or stored locally at clients an thus easily available for tempering.
12. *Lack of Devices Situation* – all schemes that rely on an outer referee (an external entity that is not part of the game) rely on the availability of the referee. In decentralised architectures where refereeing is done by other players, the scheme relies on availability of adequate number of players.
13. *Exploiting AI* – Exploiting artificial intelligence (AI) cannot be handled within the protocol or the architecture. The idea that a player can use an AI to improve its decision making in a game is not possible to detect on the protocol level.
14. *Social Engineering** – social engineering is a very broad term. Generally, it involves using social information about a player to trick the person into revealing sensitive information pertaining to a game, i.e. passwords. Our protocol uses ECDSA public cryptography, and does not require passwords. The authentication is not necessary as messages exchanged between players are all signed and verified.

4 DECENTRALIZED ARCHITECTURE FOR THE EDGE

Previously proposed P2P architectures rely on some level of centralization. We propose a completely decentralized architecture for edge devices that would inherently circumvent the single point of failure (SPOF). In contrast to client-server (C/S) architectures, where the server is authoritative, P2P networks are arguably more exposed to cheats and vulnerabilities. To address the issues Web et al. propose RACS [28], a referee node that takes the authoritative role in case of conflicting or inconsistent states between players.

However, RACS does not address issues of node selection. In completely decentralized networks, deriving secure randomness is an open question. From a security point of view, a decentralized random generator must not be known in advance to avoid attacks and vulnerabilities based on information exposure (IE). At the time of writing, most networks rely on oracle networks secured with game theoretical incentive schemes [21, 10]. However, the security models for such systems require strong incentives, which are mostly based on staking mechanisms that introduce penalties for bad actors and rewards for good actors [13]. A recent paper proposed a mathematical construction for verifiable random functions (VDF) [4], an extension of time lock puzzles [22] that produce verifiable proofs of computation. More specifically,

VDFs are similar to time lock puzzles but require a trusted setup where the verifier prepares each puzzle using its private key. Additionally, a difficulty parameter can be adjusted to increase the amount of sequential work, thereby increasing the delay. The proof can be used as an entropy pool for a seeded random to derive randomness within a decentralized system while preventing attacks based in IE. We solve the requirement for a trusted set up (private key of the VDF) by using a blockchain structure. Blocks have a configurable block time parameter, which is used to adjust the difficulty parameter of the VDF, to target the block time. The consensus algorithm is a novel lottery draw scheme, where nodes draw lottery tickets in order to be voted as block producers.

Suppose the current block is H at height (canonical id) h. The block hash of block H is used to compute the VDF and obtaining proof H_p. Each node $n \in N$ then draws its own lottery ticket H_t, which is defined as the distance between the node's public key, and H_p. Since all nodes share the same H_p, and all nodes (asymptotically) computed H_p at the "same" time, the lottery draw is not predictable. A node is elected to be part of the validator set if H_t is within the v closest tickets, where v is a configurable parameter usually set to $\frac{P}{PPI}$, where P is the total number of players, and PPI is the number of players per instance. Nodes that belong to the validator set are considered referees, and block producers for block $H + 1$. The structure of the block is shown in Figure 1.

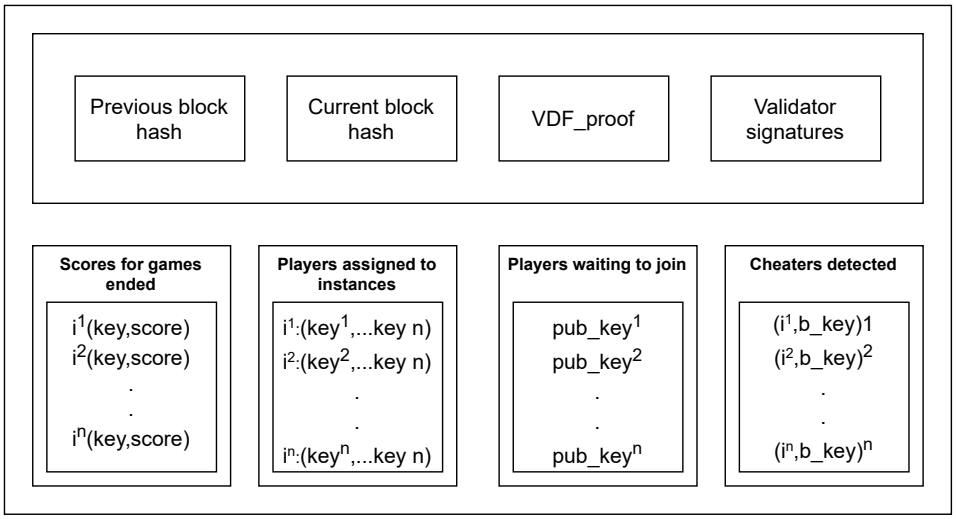

Figure 1. The Header of a block contains the previous block hash, current block hash, VDF proof, and signatures of all validators that signed the block. The body of the block contains a list of game instances that completed (combination of players public addresses), and their scores, a list of players waiting to join the next round (block), and a list of instances with assigned players that were in the waiting queue of the previous block.

Players are uniquely identified by their public key. Each player generates a public-private key pair before connecting to the P2P network. Upon joining the network,

players synchronize the last block to find the validator set. Querying the Distributed Hash Table – DHT [25], a node connects to the one or more validators to broadcast their intent to join a game. Once consensus is reached amongst the validators, a new block H will be forged that will include the player's public key in the players awaiting list. Players will receive block H, learn about their inclusion to the awaiting list, and wait for block $H + 1$. Note that the target block time is configurable with the VDF difficulty and should be set by the game operator. Upon receiving block $H + 1$, the players' public key will be assigned to an instance. Each instance has a unique ID, which is obtained by concatenating the public keys of players assigned to the instance. Each player parses the instance ID to obtain the public keys of opponents and connects to them by querying the DHT (with the public key) for their address. The last address of the instance ID is the assigned validator that will assume the role of the referee. Once the instance is resolved (game finished), the referee (also a validator) will inform all other validators and propose the inclusion of the decision/score in the next block. Validators will vote on the proposed block by signing it with their public key. Clients can verify their signature using their public key. In case a referee of an instance detected a cheater, a proof can be sent to the set of validators for confirmation. The details of how this is achieved are explained in detail in Section 5.

A candidate block $H+1$ is then transmitted (using gossip protocol) [15] through the P2P network. Each client accepts the block if and only if the block references the local block hash at height h, the provided proof H_p is valid, and the candidate block contains signatures of all validators whose public keys can be computed by each node using H_p. The nodes that are part of the validator set execute a matchmaking algorithm that must be deterministic (but can rely on randomness derived from H_p) and can use the previous block as input (list of players wanting to join). The matchmaking algorithm assigns player to game instances, and referees from the validator are set to act as authoritative nodes. The deterministic nature of the matchmaking algorithm is used to reach consensus amongst the validator nodes. The consensus is reached as all honest nodes will construct the same candidate block and will sign all candidate blocks equal to theirs. The result will be a candidate block signed by the majority of validators.

The construction assumes validators, referees, and players to be players. However, a set T of trusted nodes is required and assumed to be maintained by the game maintainers. These nodes are called full-nodes and are necessary to guarantee liveliness of the system even in extreme cases where there are no players in the network. Full-nodes are also responsible for permanently storing the blockchain and maintaining a DHT-based structure other nodes can query to discover other peers. Players are assumed to be lite clients that do not need to store the entire blockchain history in order to participate in the consensus [27]. Additionally, referees are assumed to be players as well. The matchmaking algorithm should avoid assigning players to be referees to their own game instance.

Each game can have one referee, which arguably decreases the robustness. All decisions about conflicts proposed by a referee must be presented to the validator

set in order to reach consensus and gather enough signatures to make the block valid. However, the referee can unexpectedly disconnect or even worse, be attacked by a player during the game. To circumvent this issue, any number of validators can be assigned as backups in case the referee is unresponsive.

Referees in DAMAGE are running the same protocol as RACS. However, in case a referee detects a cheating player, the proof (usually a set of states that allow validators to recreate/simulate the game) must be presented to the set of validators (also RACS referees). Consensus is reached if and only if $\frac{2}{3}$ validators agree [5]. Decisions about the proposed cheat detected is done by voting for the block. Each block contains a list of (*public key*, *instance key*) pairs and the type of cheat detected. Assuming the validator is honest, and the referee proposing the detected cheat is as well, both validators will reach the same conclusion and thus sign the block. In any other case, the block will only be signed by the malicious validator. Proposals that do not reach consensus are considered invalid blocks and will be rejected by the client protocol. A subset of nodes (without the majority vote) running modified clients may choose to accept the invalid block, thereby forking the chain [1]. In such cases, the next block would either resolve the fork if it is accidental or disconnect (network level) the subset of nodes with the modified protocol due to an invalid VDF proof on the forked chain.

5 SECURITY MODEL

All communication between peers provides the same level of security to that of C/S architectures by using public key (ECDSA) cryptography. DAMAGE provides a secure and completely decentralized protocol for selecting referees, and match players to game instances. However, the player and referee protocols are based on RACS [28] and therefore DAMAGE inherits cheat detection properties of RACS, and extends them with efficient and secure peer selection, peer synchronization, robustness, and some aspects of collusion.

5.1 Referee Selection

We address the issue of Referee selection by using VDF to derive randomness with which a lottery-based consensus is reached. The sequential nature of VDFs prevents IE attacks where a player would compute the VDF and, using the proof, obtain information about which nodes are part of the validator set and which node is assigned as the main referee to each game. Game operators should set the difficulty parameter according to their desired performance/security ratio. A more difficult VDF will result in players waiting to be matched to an instance longer (i.e., a few seconds), while a lower difficulty will potentially allow malicious players to discover the nodes that will be within the set of validators before others. We argue that knowing the set of validators and, consequently, the referee node for a game does not give the player a competitive advantage. This is further explained in the case of collusion.

5.2 Referee Trust

In RACS scheme referee nodes are assumed trustworthy (the authors acknowledge this to be an open issue). We solve this issue with the validator set. Even if a RACS-based referee is compromised, any player can dispute the referee and seek a decision by consensus within the set of validators. The player would then have to compromise $\frac{2}{3}$ nodes in the validator set, which is not known in advance and changed every block.

Instead of using only one referee per game, we propose to establish a Referee set (validator set of referees) that. Additionally, the cardinality of the validator set is a configurable parameter analogous to the trust level required by the game (higher trust requires bigger cardinality).

5.3 Synchronization

On the data layer, nodes synchronize through the blockchain. Blocks store the current state of the system on lite clients and the entire history on full-nodes maintained by the game operators. Blocks are gossiped across the network efficiently by maintaining a DHT that maps nodes (public keys) to their network addresses. Additionally, referees in the validator set must synchronize and reach consensus about the detected cheats and results of the games played. Due to the deterministic nature of the cheat detection algorithms, honest nodes will reach the same decision as the referee that reported the cheat. Consensus is reached if the majority of the validators sign the proposed block (which includes the decision about reported cheats).

5.4 Robustness

DAMAGE uses redundancy to increase fault tolerance. There are two main types of faults that can occur. A peer can fail (disconnect or violate protocol) before, after or during playing the game, and a peer acting as a referee (and also as a player in a different game) fails at the same time.

Player faults after entering matchmaking: A peer that faults after it announced inclusion to the validator set will cause the validators to match its public address to an instance. Other peers attempting to connect will fail and/or result in protocol violation. It is up to the client protocol to decide if the game instance can continue to run without the faulty peer or not. In case the instance must be destroyed, this can be trivially solved by extending the referee's protocol to label this as a "cheat". The referee will announce the instance destruction to the validator set.

Player faults during the game: If possible, the game instance should keep running. If not possible, the referee should notify the validator set about the destruction.

Player faults after the game: No effect on the system.

Faulty validators (referees) are arguably a bigger security issue. Even without the ability for players to know which referee will be assigned to their instance there is still a possibility of DDoS attacks on referees during the game. To combat this issue, validators form a randomly shuffled priority queue using the VDF proof. The priority queue is a backup queue of referees that will take over an instance in case the referee assigned faults. The fault tolerance can be increased on demand by increasing the size of the validator set. However, detecting a faulty referee must be done by peers playing in the instance. If messages from peer to referee are either latent or connection is dropped, client protocol will take the following steps:

1. Set up a seeded random with the latest block (local) of the VDF proof.
2. Compute the lottery draw results to find the public keys of the validator set.
3. Shuffle the validator set list with the same seed.
4. Contact the next validator (backup referee for its game).

5.5 User Profile Management

Previous research relied on a central authority for authenticating users and managing their profiles such as avatars, variables, and metadata. Our architecture can be extended with a completely decentralized storage and authentication service, a centralized authoritative server as well as inter-operability between both. A blockchain-based authentication service can be built in by extending the block structure [18]. Additionally, blocks can be used for persistent immutable storage. However, storing data in blocks raises scalability issues [31, 24] as the blockchain becomes hard to maintain even for full nodes. Hybrid approaches have been proposed where data is stored centrally whereas the signatures are stored on-chain [31]. This creates a tamper-proof system where data can be verified and trusted as any attempt to tamper with the data would invalidate the signature (hash) [24].

5.6 Lack of Devices Situation

The refereeing process relies on a set of validators that are randomly chosen for each block time-cycle. The randomness of selection ensures that a player cannot know who is refereeing the next game. Player's devices are used to act as validators. The pool of validators cannot be constructed if there are not enough players. It is developers' or game operator's task to supply enough (a fixed number that does not grow with player-base) resident secure services (servers) that act as starting validators. These actors also maintain the blockchain (full nodes).

5.7 Collusion

Assuming the game runs multiple instances, we argue collusion between players is not possible. We assume colluding players know and, hence, trust each other.

Figure 5 shows a graph of a simulation of 200 players as nodes. The edges represent the number of games (weight) a player was assigned another player as the referee for the instance the player was matched to. Simulating 1 000 blocks, the average degree was 200, and the graph density was 1. We observe that the assignment of a referee and opponents derived using the VDF proof are thereby random. Hence, players cannot know in advance which instance the set of validators will assign them to nor the referee that will observe the game. Section 6.2 and Figure 5 present an empirical evaluation of the "fairness" of the selection method based on VDF. Suppose the colluding players are able to compute the VDF proof faster then other players, and, hence, learn about the game instance assignment in advance. However, since the seed for the next VDF is the block hash, the colluding players can see at most one block time into the future. Every player must announce the desire to be matched to a game in the current block (players awaiting list). Matching awaiting players will be executed in the next block. Despite the ability to see one block in the future, colluding players seek assignment to the same instance since they must announce their willingness to play before they learn about the instance assignment even in the worst case scenario.

6 EVALUATION

The paper presents a scheme to eliminate all known cheats in a fully distributed game setting. The scheme eliminates the SPOF problem in previously presented hybrid P2P – Referee settings for solving game cheats. We base our solution on already presented solutions, mostly RACS [28]. We present simulation results leading to the following conclusions:

- The VDF based selection of referees and players is fair.
- The block propagation scales well.
- Block propagation times are acceptable for fast match making, and conflict resolution.
- Dynamic block size does not impact performance of the system.
- Players do not need to maintain a large number of outgoing connections.
- Latency and bandwidth do not substantially slow down information propagation through the network.

The scalability of the solution is addressed in two ways. The first is the scalability of the consensus mechanism and the ability to propagate state and state transition depending on the block size. In this test we show that the solution can scale to hundreds of thousands of nodes and achieve consensus.

The second scalability test is performed by introducing variance in latency and bandwidth to mimic the instability of home internet connections under standard TCP/IP parameters.

Since every player is also a node, and potentially a referee, we argue that scalability in terms of number of players can be derived by the aforementioned tests. Additionally, due to the nature of the P2P architecture, once players are matched into a game instance, the entire communication is done solely between them, and the referee of that instance, which is completely independent of the rest of the network, and hence does not impact the scalability.

In order to evaluate the solution a simulation environment was developed. We simulate a P2P network where each peer (player) has the following constraints:

- Local bandwidth constraints. Bandwidth constraints are assigned to peers joining the network based on the distribution obtained from the European report on network bandwidth [9].
- Maximal number of outgoing connections (out edge degree) is assigned to nodes (MAE), we ran tests with different values of this parameter, they are color-coded in Figure 3.
- Each node's connection is single-directional taking into account upload and download bandwidth constraints of sender and receiver.
- Each new connection is assigned a round trip time (RTT) to represent variance in latency. RTT values are assigned randomly fitting a Gaussian distribution on an interval $[30, 250]$ ms, the values were taken from the European report on network bandwidth [9].
- Actual throughput of each connection is estimated using the Mathis metric [16] with following parameters that were taken from real-life situations: maximum segment size (MSS) of 1460 bytes (most used in today's communications as shown in papers such as [23]), the connection's RTT, and a TCP packet loss probability of $p = 1.0 * 10^{-5}$ [19].

Nodes (players) join the network by connecting to one of the trusted nodes. Trusted nodes are those operated by the game maintainer and serve only as the entry point for new peers to discover other peers or if needed to persistent storage for player accounts. A node proceeds to run the peer discovery protocol building the DHT. When new nodes are discovered, the peer attempts to make new connections until the MAE limit is reached and the node is considered to be well connected. Examples of different architectures (presented by connected directed graph) for 20 nodes are presented in Figure 2.

Once a new block is forged the origin nodes propagates the block using a basic flooding algorithm simulating the bandwidth, and TCP constraints. In each simulation, multiple directed graphs are constructed following the above protocol. Simulations were carried out with different number of nodes to observe the scalability of the solution. We measure propagation time as the total time it takes for all nodes to receive a newly forged block.

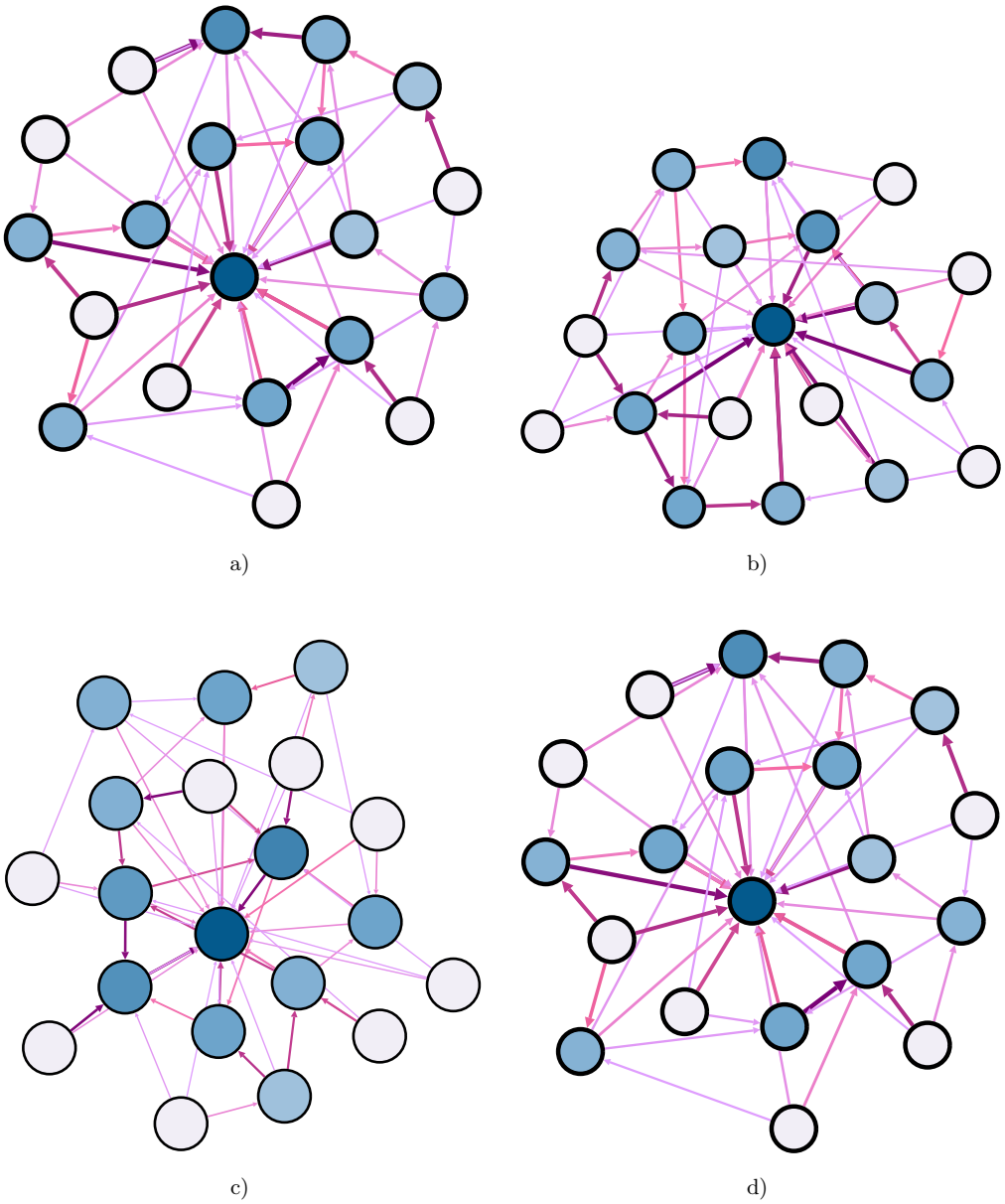

Figure 2. Examples of different architectures obtained by simulation. Number of nodes in all examples is $n = 20$.

6.1 Block Propagation Times

Blocks hold the state of the match making and games being played. Lowering the block time (VDF difficulty) would result into a more responsive experience. However, lower block times reduce security, and can cause network congestion. To avoid possible client synchronization issues the network must be able to reliably propagate blocks before new ones are forged. Additionally, the propagation times vary depending on the network topology, block sizes, and average node degree. We evaluate the scalability of propagating blocks in order to estimate viable block times, and show the scalability of the solution. From Figure 3 we observe that propagation times scale logarithmic as we increase the number of clients. Additionally, increasing the number of outgoing connection a node maintains reduces the average propagation times as it reduces the risks of unfavorable graph typologies.

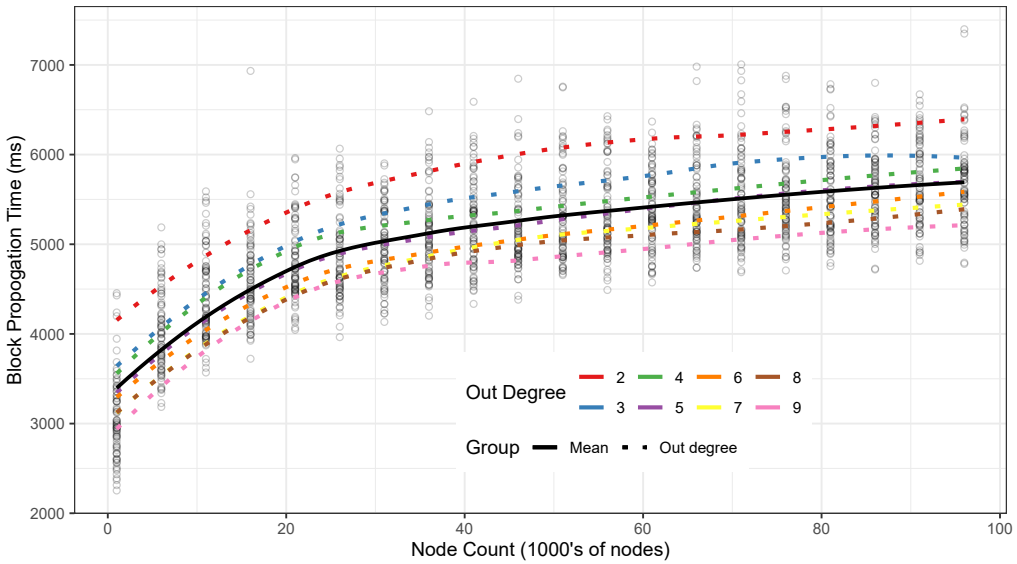

Figure 3. Simulation of block propagation on different graph configurations. Configurations are obtained by increasing the node count (number of players), and out degree using a block size of 1 MB.

Block size scales lineally with the number of clients. Every block has a constant size for the header, which is 64 bytes for previous and current block hash, 100 bytes for the VDF proof, and at least one validator signature of 64 bytes. As more players join the network, more games need to matched, assigned to instances, and scores saved. Figure 4 shows how the network scales with different block sizes. We observe that latency and bandwidth speeds of some nodes can cause considerable propagation slowdowns indicated by some outliers. However, this can be mitigated by having nodes maintain a dynamic number of outgoing connections increasing the

limit as blocks become larger (more players), and lowering the limit as blocks are smaller.

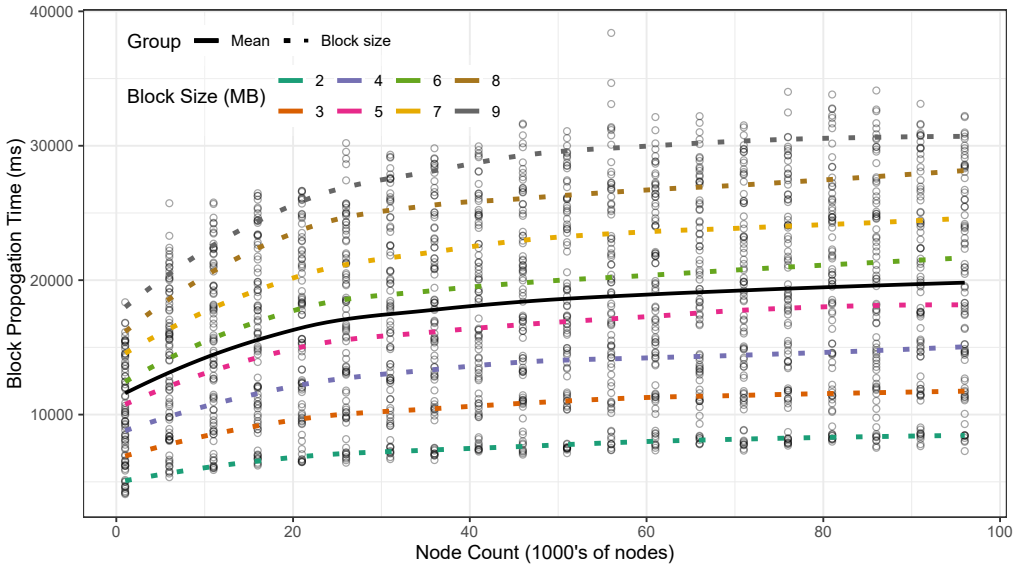

Figure 4. Simulation of block propagation on different graph configurations. Configurations are obtained by increasing the block size. Nodes were limited to 3 outgoing connections.

6.2 The VDF Based Selection of Referees and Players is Fair

Figure 5 shows a graph of a simulation of 200 players as nodes. The edges represent the number of games (weight) a player was assigned another player as the referee for the instance the player was matched to. Simulating 1 000 blocks, the average degree was 200, and the graph density was 1. We observe that the assignment of a referee and opponents derived using the VDF proof are thereby random. Hence, players cannot know in advance which instance the set of validators will assign them to nor the referee that will observe the game.

6.3 The VDF Method Scales Well

The setting presented in Section 6.2 and Figure 5 shows that the VDF based selection method is fair, the graph shows 200 players and 1 000 blocks, as this was the maximum feasible number combination that was still manageable to visualize. The setting was further evaluated with different parameters for the number of players, number of players per game and number of blocks. The number n of players per game: means a game where n players participate, Number of blocks: how much time the matchmaking process was observed. We evaluated the setting with 200, 1 000

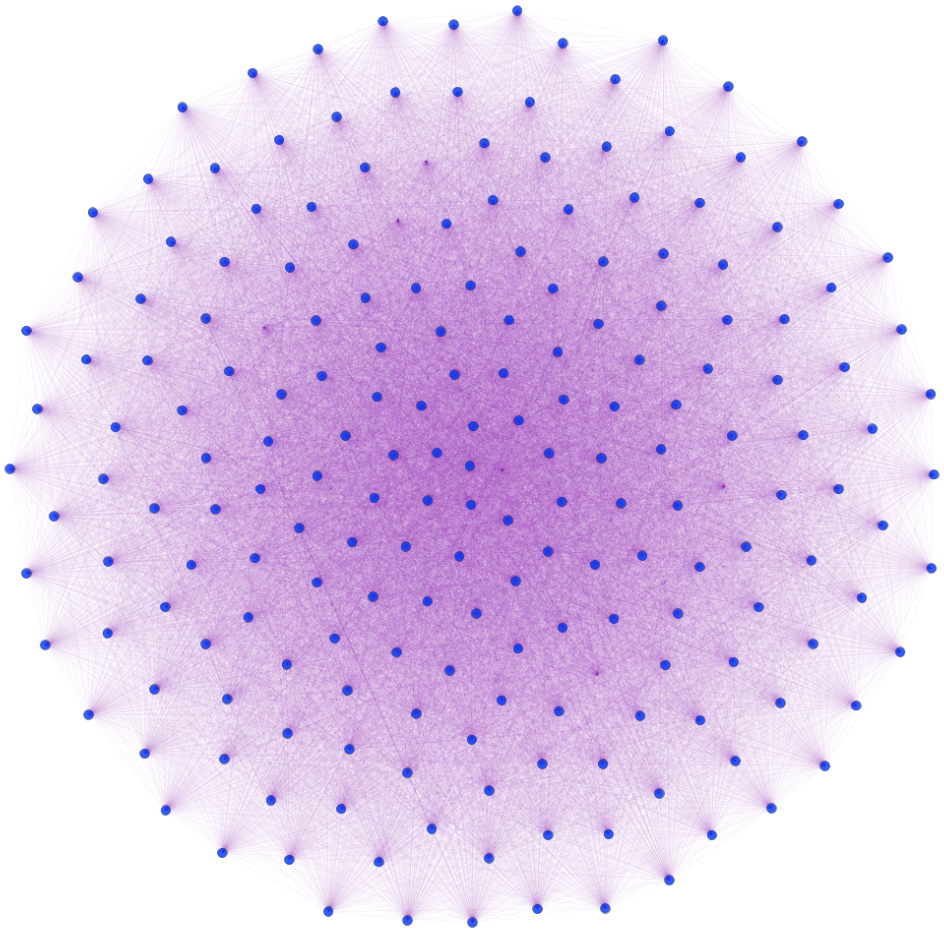

Figure 5. Simulation of 200 players in a 2-player game ($PPI = 2$) that played a total of 1 000 games. Nodes are players and edges represent instances where the destination node was referee for the instance the player was matched to.

and 10 000 players, 1 000, 2 000 and 10 000 blocks, we also changed the number of players per game. All results were consistent with the first test, the degree of all players was near the number of players and the density of the graph was near 1.

7 CONCLUSION AND FURTHER WORK

The paper proposes a blockchain-based, completely decentralized architecture for edge devices with no single point of failure that successfully addresses cheat problems. The presented solution is based on two hybrid approaches to P2P network games anti-cheat schemes that were based of server acting as referees. We propose

a completely decentralised approach while still retaining the same cheat resistance, actually in the case of Collusion we were able to partially address the issue. The proposed solution has not been fully implemented, we implemented the newly proposed building stones and executed empirical testing on a pilot setting. As the solution addresses the cheating problem in all aspects, a fully functional implementation is possible. DAMAGE is applicable to most game types. However, it is most suitable for turn based games, where potential latency does not impact user experience dramatically. Additionally, it reduces the complexity of the referee implementation due to the simple ordering of actions in the discrete time. Our results show, that the architecture scales automatically with the number of players thereby drastically reducing operation costs of running a multiplayer game. Every player added to the system also becomes a node, sharing its resources and contributing to verification as a potential referee.

Acknowledgment

The authors gratefully acknowledge the European Commission for funding the InnoRenew CoE project (Grant Agreement No. 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund).

REFERENCES

[1] Baliga, A.: Understanding Blockchain Consensus Models. Technical Report, Persistent Systems Ltd., 2017, pp. 1–14.

[2] Baughman, N. E.—Levine, B. N.: Cheat-Proof Playout for Centralized and Distributed Online Games. Proceedings of the Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (INFOCOM 2001), Vol. 1, 2001, pp. 104–113, doi: 10.1109/INFCOM.2001.916692.

[3] Baughman, N. E.—Liberatore, M.—Levine, B. N.: Cheat-Proof Playout for Centralized and Peer-to-Peer Gaming. IEEE/ACM Transactions on Networking (ToN), Vol. 15, 2007, No. 1, pp. 1–13, doi: 10.1109/TNET.2006.886289.

[4] Boneh, D.—Bonneau, J.—Bünz, B.—Fisch, B.: Verifiable Delay Functions. In: Shacham, H., Boldyreva, A. (Eds.): Advances in Cryptology – CRYPTO 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 10991, 2018, pp. 757–788, doi: 10.1007/978-3-319-96884-1_25.

[5] Castro, M.—Liskov, B.: Practical Byzantine Fault Tolerance. Proceedings of the Third Symposium on Operating Systems Design and Implementation, New Orleans, USA, 1999, pp. 173–186.

[6] Corman, A. B.—Douglas, S.—Schachte, P.—Teague, V.: A Secure Event Agreement (SEA) Protocol for Peer-to-Peer Games. First International Confer-

ence on Availability, Reliability and Security (ARES '06), 2006, 8 pp., doi: 10.1109/ARES.2006.15.

[7] Dobrilova, T.: How Much is the Gaming Industry Worth? Techjury, 2019.

[8] Dooley, K.: Designing Large Scale LANs: Help for Network Designers. O'Reilly Media, Inc., 2001, 404 pp.

[9] European Court of Auditors: Broadband in the EU Member States: Despite Progress, Not All the Europe 2020 Targets Will Be Met. Technical Report, European Court of Auditors, 2018.

[10] Gatteschi, V.—Lamberti, F.—Demartini, C.—Pranteda, C.—Santamaría, V.: Blockchain and Smart Contracts for Insurance: Is the Technology Mature Enough? Future Internet, Vol. 10, 2018, No. 2, Art. No. 20, 16 pp., doi: 10.3390/fi10020020.

[11] GauthierDickey, C.—Zappala, D.—Lo, V.—Marr, J.: Low Latency and Cheat-Proof Event Ordering for Peer-to-Peer Games. Proceedings of the 14th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '04), 2004, pp. 134–139, doi: 10.1145/1005847.1005877.

[12] Goodman, J.: A Hybrid Design for Cheat Detection in Massively Multiplayer Online Games. M.Sc. Thesis, McGill University, Montréal, 2008.

[13] Heiss, J.—Eberhardt, J.—Tai, S.: From Oracles to Trustworthy Data On-Chaining Systems. Proceedings of IEEE International Conference on Blockchain (Blockchain 2019), 2019, pp. 496–503, doi: 10.1109/Blockchain.2019.00075.

[14] Jamin, S.—Cronin, E.—Filstrup, B.: Cheat-Proofing Dead Reckoned Multiplayer Games. Proceedings of 2nd International Conference on Application and Development of Computer Games, Hong Kong, 2003, pp. 1–7.

[15] Kwiatkowska, M.—Norman, G.—Parker, D.: Analysis of a Gossip Protocol in PRISM. ACM SIGMETRICS Performance Evaluation Review, Vol. 36, 2008, No. 3, pp. 17–22, doi: 10.1145/1481506.1481511.

[16] Mathis, M.—Semke, J.—Mahdavi, J.—Ott, T.: The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm. ACM SIGCOMM Computer Communication Review, Vol. 27, 1997, No. 3, pp. 67–82, doi: 10.1145/263932.264023.

[17] Mirkovic, J.—Reiher, P.: A Taxonomy of DDoS Attack and DDoS Defense Mechanisms. ACM SIGCOMM Computer Communication Review, Vol. 34, 2004, No. 2, pp. 39–53, doi: 10.1145/997150.997156.

[18] Moinet, A.—Darties, B.—Baril, J.-L.: Blockchain Based Trust and Authentication for Decentralized Sensor Networks. 2017, pp. 1–6, arXiv: 1706.01730, doi: 10.1109/wimob.2017.8115791.

[19] Moltchanov, D.: A Study of TCP Performance in Wireless Environment Using Fixed-Point Approximation. Computer Networks, Vol. 56, 2012, No. 4, pp. 1263–1285, doi: 10.1016/j.comnet.2011.11.012.

[20] Pellegrino, J. D.—Dovrolis, C.: Bandwidth Requirement and State Consistency in Three Multiplayer Game Architectures. Proceedings of the 2nd Workshop on Network and System Support for Games, 2003, pp. 52–59, doi: 10.1145/963900.963905.

[21] Peterson, J.—Krug, J.—Zoltu, M.—Williams, A. K.—Alexander, S.: Augur: a Decentralized Oracle and Prediction Market Platform. 2015, pp. 1–16, arXiv: 1501.01042, doi: 10.13140/2.1.1431.4563.

[22] Rivest, R. L.—Shamir, A.—Wagner, D. A.: Time-Lock Puzzles and Timed-Release Crypto. Technical Report, Massachusetts Institute of Technology, 1996, pp. 1–9.

[23] Deering, S. R. H.: Internet Protocol, Version 6 (IPv6) Specification. RFC 2460, RFC Editor, 1998.

[24] Shafagh, H.—Burkhalter, L.—Hithnawi, A.—Duquennoy, S.: Towards Blockchain-Based Auditable Storage and Sharing of IoT Data. Proceedings of the 2017 on Cloud Computing Security Workshop (CCSW '17), Dallas, Texas, USA, 2017, pp. 45–50, doi: 10.1145/3140649.3140656.

[25] Stoica, I.—Morris, R.—Karger, D.—Kaashoek, M. F.—Balakrishnan, H.: Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications. ACM SIGCOMM Computer Communication Review, Vol. 31, 2001, No. 4, pp. 149–160, doi: 10.1145/964723.383071.

[26] Superdata: Market Brief – 2018 Digital Games and Interactive Entertainment Industry Year in Review, 2019.

[27] Vorick, D.—Champine, L.: Sia: Simple Decentralized Storage. Nebulous, 2014, pp. 1–8.

[28] Webb, S.—Soh, S.—Lau, W.: RACS: A Referee Anti-Cheat Scheme for P2P Gaming. Proceedings of the 17th International Workshop on Network and Operating Systems Support for Digital Audio and Video, 2007, pp. 34–42, doi: 10.1145/1542245.1542251.

[29] Yahyavi, A.—Kemme, B.: Peer-to-Peer Architectures for Massively Multiplayer Online Games: A Survey. ACM Computing Surveys (CSUR), Vol. 46, 2013, No. 1, Art. No. 9, 51 pp., doi: 10.1145/2522968.2522977.

[30] Yan, J.—Randell, B.: A Systematic Classification of Cheating in Online Games. Proceedings of 4th ACM SIGCOMM Workshop on Network and System Support for Games (NetGames '05), 2005, pp. 1–9, doi: 10.1145/1103599.1103606.

[31] Zyskind, G.—Nathan, O.—Pentland, A. S.: Decentralizing Privacy: Using Blockchain to Protect Personal Data. 2015 IEEE Security and Privacy Workshops, 2015, pp. 180–184, doi: 10.1109/SPW.2015.27.

Aleksandar Tošić is Teaching Assistant at the University of Primorska, and Research Assitant at InnoRenew CoE. His main research interests are distributed systems and distributed ledger technologies.

Jernej Vičič is Associate Professor at the University of Primorska, Primorska Institute for Natural Sciences and Technology in Koper, Slovenia. His main research interests are distributed systems and natural language processing.

2.5 Paper 5

Title: A WSN Framework for Privacy Aware Indoor Location
Authors: Aleksandar Tošić, Niki Hrovatin, Jernej Vičič
Year: 2022
Journal: Applied Sciences
DOI: 10.3390/app12063204
Link: https://www.mdpi.com/2076-3417/12/6/3204

applied sciences

MDPI

Article

A WSN Framework for Privacy Aware Indoor Location

Aleksandar Tošić [1,2,*,†,‡], Niki Hrovatin [1,2] and Jernej Vičič [1,‡]

[1] Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, 6000 Koper, Slovenia; niki.hrovatin@innorenew.eu (N.H.); jernej.vicic@upr.si (J.V.)
[2] InnoRenew CoE, Livade 6, 6310 Izola, Slovenia
* Correspondence: aleksandar.tosic@upr.si
† Current address: Glagoljaška 8, 6000 Koper, Slovenia.
‡ These authors contributed equally to this work.

Abstract: In the past two decades, technological advancements in smart devices, IoT, and smart sensors have paved the way towards numerous implementations of indoor location systems. Indoor location has many important applications in numerous fields, including structural engineering, behavioral studies, health monitoring, etc. However, with the recent COVID-19 pandemic, indoor location systems have gained considerable attention for detecting violations in physical distancing requirements and monitoring restrictions on occupant capacity. However, existing systems that rely on wearable devices, cameras, or sound signal analysis are intrusive and often violate privacy. In this research, we propose a new framework for indoor location. We present an innovative, non-intrusive implementation of indoor location based on wireless sensor networks. Further, we introduce a new protocol for querying and performing computations in wireless sensor networks (WSNs) that preserves sensor network anonymity and obfuscates computation by using onion routing. We also consider the single point of failure (SPOF) of sink nodes in WSNs and substitute them with a blockchain-based application through smart contracts. Our set of smart contracts is able to build the onion data structure and store the results of computation. Finally, a role-based access control contract is used to secure access to the system.

Keywords: WSN; indoor location; privacy; blockchain; COVID-19

check for updates

Citation: Tošić, A.; Hrovatin, N.; Vičič, J. A WSN Framework for Privacy Aware Indoor Location. *Appl. Sci.* **2022**, *12*, 3204. https://doi.org/10.3390/app12063204

Academic Editors: Asadullah Shaikh, Uffe Kock Wiil, Yousef Asiri and Agostino Forestiero

Received: 20 December 2021
Accepted: 11 March 2022
Published: 21 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

We have recently witnessed the coronavirus disease 2019 (COVID-19) outbreak caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2).At the time this manuscript was written, SARS-CoV-2 was still spreading and affecting billions of lives globally [1]. It is now well established from a variety of studies that the SARS-CoV-2 primary infection vectors are the respiratory droplets of infected people produced by coughing, sneezing, or talking [2–4]. Therefore, the rapid spread is driven by the social aspects of everyday life, which in recent days have been altered by the guidelines for preventing infection spread, such as the mandatory use of masks, cleaning and disinfection, and the introduction of social distancing. Respecting the mentioned guidelines is of particular concern in public buildings where multiple people share the same space, and the infection spread could endanger not only individuals but also halt the operations of organizations. Moreover, in confined spaces, the probability of infection is higher than outdoors since infection transmission is dependent on ventilation [4].

The role of IoT (Internet of Things) to prevent the spreading of the COVID-19 disease has already been discussed in [5–8], which conceptualize frameworks for monitoring the spread of the COVID-19 disease through heterogeneous sensor technology and apply data-driven inferences to forecast new outbreaks and predict virus mutations. However, the mentioned literature barely discusses privacy concerns and only recent studies [9] are deliberating over the privacy aspect of integrating such monitoring solutions in everyday

life. Even though encryption effectively provides data privacy, monitoring indoor activities by relying on wireless IoT devices could disclose contextual information on data transmission [10,11], not only posing risks to the privacy of individuals, but also compromising building security.

This has motivated us to extend the research on our privacy-aware IoT and blockchain-based indoor location system to counter the spread of the COVID-19 disease. The presented indoor location system is particularly suitable for application in medical facilities, public buildings, and residential homes as a framework for privacy-aware indoor location monitoring. The proposed solution could be applied for structural health monitoring, studying behavioral patterns of a building's occupants and health-related issues such as locating lost patients with memory and orientation disorders, fall detection, and also identifying violations of social distancing, counting the number of persons in a room, and determining when and which room needs surface disinfection due to over-utilization, etc.

The key contributions of the privacy-preserving framework are:

- A novel privacy-preserving indoor location system with querying capabilities: The network of sensors is embedded in the floor and senses the local force applied over it. It is non-intrusive and does not require active user interaction. Moreover, the raw sensory data collected by sensors describe the force applied to the floor and can only lead to unique user identification via walking gait analysis. However, the walking gait analysis [12] requires large amounts of data from individual users, and in our privacy-aware framework, the raw data do not leave the source sensor, therefore inhibiting similar attempts.
- A secure WSN with anonymous source location and sensor network identity: We propose a new querying protocol for WSN, which uses multi-layer encryption to conceal the network identity of sensor nodes, obfuscating the computation described in [13]. The protocol relies on particular messages similar to those used in the onion routing [14] to convey edge data processing information to sensor nodes and privately retrieve data.
- A blockchain-based fault tolerant indoor location system with no single point of failure(SPOF): We address the fault tolerance shortcomings of sink nodes [15] in traditional WSNs by substituting it with a smart contract, which handles the processing of queries, and storing the results. A decentralized role-based access control (RBAC) contract provides user access authorization to monitor individual building spaces defining privacy boundaries and further improves the security over traditional centralized approaches.

The remainder of the paper is structured as follows: In Section 2, we present the relevant literature. Section 3 highlights the core features of the proposed solution. In Sections 4 and 4.1, we present our onion route protocol and filtering. In Section 5, we detail how blockchain smart contracts can replace sink nodes. In Section 6, we provide the validation of the proposed framework, and finally give final remarks in Section 7.

2. Literature Review

Indoor real-time locating systems (RTLS) have been gaining relevance due to the widespread advances of devices and technologies and the necessity of location-based services. The interest of the mobile industry to accelerate the adoption of indoor position solutions turned into the foundation of the InLocation Alliance (ILA (InLocation Alliance): inlocationalliance.org, accessed on 19 December 2021). The goal of this alliance is to facilitate a rapid market adoption so that new business streams are opened up with context-aware applications in indoor environments. The ILA chose Wi-Fi and Bluetooth as their preferred technologies. Both proposed technologies require specialized apps on the mobile devices in order to produce satisfactory results [16].

A thorough and contemporary survey of the Indoor Positioning Systems (IPS) for IoT is presented in [17]; it presents indoor positioning concepts and a list of already used criteria that define IPS for IoT. Brena et al. [16] provide a classification of Indoor Positioning Systems

(IPS), basing the classification on a set of papers comparing different IPS approaches. This is a list of identified technologies: Infrared mobile reader, Infrared (IR), laser (passive), ultrasound passive, audible sound, magnetic, RFID mobile tag, RFID mobile reader, Wi-Fi, Bluetooth, ZigBee, UWB, tomographic technology (water resonance), camera infrastructure, cameras (portable), floor tiles, air pressure, inertial, ambient light, artificial light, indoor AGPS, cellular technology, TV, and FM. All the technologies that need any intervention from the user are out of the scope of this experiment, so all technologies based on wearables, which demand the installation of software on mobile devices or the users to act in a certain way, are out of the scope of the paper. All technologies based on audible and visible changes in the environment (such as the usage of fluorescent lighting) pose a distraction. Additionally, the use of video cameras and microphones presents a huge privacy concern and were thus eliminated from this study. Most IR systems require line-of-sight (LOS) clearance from the emitter to the sensor; in the context of IR IPS systems, the requirement of LOS clearance is a great disadvantage, as it suffers from no-detection areas, and the system performance is also affected by sunlight [18].

A metaheuristic for anomaly detection in IoT is proposed in [19], which is an extension of the work presented in [20]. The method is based on an activity footprints-based method to detect anomalies in IoT, but with small changes it can be used to track indoor activity.

The technology that "survived" the criteria posed by the presented study was "intelligent tiles", usually using pressure sensors. There has been some research in the area of employing pressure sensors to track the users' indoor behaviors, ranging from person tracking and indoor localization to fall prediction. The Smart Floor project at Georgia Institute of Technology [21] and ORL Active Floor at The Olivetti and Oracle Research Laboratory [21] provide location and identification without encumbering the users, but their highest levels of precision will not be reached until the user steps on the exact centers of the floor tiles, which for a reliable measurement would require conscious attention. Chan et al. [22] present a smart-sensored floor setting that draws energy to power the motion sensors from the integrated generators that are powered by normal floor activity such as walking or sport activity. Kaddoura et al. [23] present a cost-effective intelligent floor setting using pressure-sensing sensors that functionally competes with higher-cost systems. Shen and Shin [24] report on the development of a distributed sensing floor using an optical fiber sensor. However, all presented intelligent floor systems fail to properly address the privacy and data-sensitivity issues.

Privacy preservation in location systems has already been addressed in some works, although in different domains, such as [25], which proposes a location privacy method based on k-anonymity, and [26], which uses blockchain to achieve the desired behavior.

Cumulative pressure sensors [27] for large areas have been proposed to present a rough estimate of the number of persons present in a designated area (effectively measuring/counting the occupancy of a room). This technology is only useful for counting the number of occupants in the observed are; it lacks all the other IPS properties.

Google and Apple have jointly developed an exposure notification system (https://www.google.com/covid19/exposurenotifications/, accessed on 19 December 2021) based on a shared sense of responsibility to help the global community fight the pandemic by keeping track of contact. In the background, users' phones and surrounding phones share randomly generated privacy IDs via Bluetooth. Routinely, the application checks if some of the IDs that the phone has been exposed to have a "compromised" ID, the IDs of owners who have anonymously proclaimed to be infected. The exposure notification system does not monitor users' locations; Google, Apple and other users cannot see users' identities; and the data are only available to the public health authorities. This system does not address the same issues as the system proposed in this paper as the proposed system cannot be utilized as a substitute of the Google/Apples solution for the lack of a backward loop (the information of the infection case cannot be linked to the pseudo-anonymous identities used in our system).

Tošić et al. [28] present a non-intrusive fall detection solution based on a smart floor, which this paper extends to an indoor location system. The system enables a non-intrusive (with no need for special applications based on wearable devices, smartphones or any other devices) indoor location system with additional privacy preserving properties such as anonymity and sensor location/network anonymity. We achieved this by using the smart floor, coupled with onion routing for source location anonymity and blockchain for the final sink personal pseudo-anonymity.

2.1. Secure Data Processing in Network of Sensors

The data sourcing from a network of sensors is usually processed in a system external to the network; the processing system is often a cloud service. Solutions such as Transport Layer Security (TLS) are applied to provide a secure data transfer from sensor nodes to the data processing system. However, even though TLS solutions ensure data confidentiality, a number of studies [29–31] show that it is possible to associate TLS traffic patterns with activities monitored by the network of sensors.

The technique of Compressive Sampling (CS) found application in WSNs to severely reduce the sending data size by representing the data using a smaller number of samples than dictated by the Nyquist theorem [32,33]. Furthermore, the CS was not applied only to reduce the communication overhead but also to provide data confidentiality by changing the CS coefficients at each transmission by relying on a secure seed at sensor nodes [34]. In [35], the authors propose a CS data-gathering scheme that provides data confidentiality and protection against traffic analysis via the use of public-key Homomorphic Encryption [36] to secure the transmitted data [35]. However, in CS techniques, the data recipient can reconstruct and identify the data from individual nodes, and therefore, it is an appealing target for attackers since, if compromised, it could disclose the private data of several nodes. Moreover, CS requires that the data recipient node solves a linear programming equation to recover the original data; therefore, the computation load is introduced and does not take advantage of the processing power of nodes forming the sensor network.

Numerous studies [37,38] have focused on preserving sensor network privacy by aggregating data as they flow through the network. The technique is dubbed as in-network data aggregation and relies on aggregator nodes that aggregate the data from multiple sensor network nodes; however, it does this without the possibility for the aggregator node to disclose the private data of individual nodes. The survey [37] provides a classification of privacy-preserving data aggregation techniques, categorizing and describing them.

Even though privacy-preserving data aggregation could preserve the data privacy of individual nodes, the current solution only allows computing aggregates such as SUM, MAX, AVG, variance, etc. The mentioned aggregates could provide an overview of the monitored environment; however, they are not sufficiently descriptive for indoor location requirements. In this study, we propose a data acquisition layer based on the General Purpose Data and Query Privacy Preserving Protocol described in [13]. This technique allows the retrieval of arbitrary aggregated data without disclosing which nodes contribute to the data retrieval. The generated network traffic is uniform due to randomized paths and the sojourn time, therefore preventing traffic analysis attacks. The computing power of sensor nodes is utilized for data processing in situ. Moreover, in the present contribution, we present a technique coupled to a blockchain solution to secure query creation, ensure that only the message origin knows nodes contributing to the data retrieval, and eliminate the aggregator/sink node SPOF.

2.2. Role Based Access Control—RBAC

Traditional IoT access control schemes are mainly built on top of the well-known access control models, including the role-based access control model (RBAC) [39,40], the attribute-based access control model (ABAC) [41], and the capability-based access control model (CapBAC) [42]. In the RBAC-based schemes, the access control is based on the roles (e.g., administer and guest) of the subject. RBAC oversees the user role assignment and permission

assignment. Three implementations currently exist in the form of smart contracts for the Ethereum network [43]: RBAC-SC [44], Smart policies and OpenZepplin contracts (OpenZepplin contracts: https://github.com/OpenZeppelin/openzeppelin-contracts, accessed on 19 December 2021). The blockchain and RBAC service were used as an off-the-shelf service providing the necessary functionality and the scope of the paper does not support any analysis on the comparable properties of the presented solutions.

3. Architecture

Our framework makes use of three main innovations to implement unique properties, which we rely upon to address the limitations of existing indoor location systems. The architecture encompasses these as modules such that it allows interoperability between them to achieve an additive effect of their unique properties. In our implementation, we design a unique cost-effective passive indoor location system that relies on off-the-shelf sensors embedded in an additional layer between the tiling described in more detail in Section 3.2. At the local level, the sensors form a WSN which reduces the complexity of the large-scale implementations. The security and network anonymity [45] concerns are addressed by a specially designed computational model that relies on onion-routing [46] messages for network anonymity, and a general-purpose obfuscated computing model. By using multi-layer encryption and onion routing, nodes are able to collaborate in federated and distributed computations without ever revealing what the global computation is, nor the origin of the computation; further details can be found in Section 4. In the third module, we further improve the security and reliability of the solution by decentralizing the system to introduce much-needed fault tolerance, and secure the entire solution against a single point of failure (SPOF). By using blockchain, we are able to replace sink nodes with smart contracts. We implement an access control module that protects the underlying WSN against unauthorized queries, further detailed in Section 5.

In our vision, different deployments of indoor location systems have different requirements, ranging from personal home deployments (smart home) to health providers (hospitals, homes for older adults, clinics, nursing homes, etc.), and public buildings (municipalities, government buildings, etc.) illustrated in Figure 1. Using a global blockchain network, which stakeholders can participate in, we can inherit the same security level on all of the underlying sensor deployments.

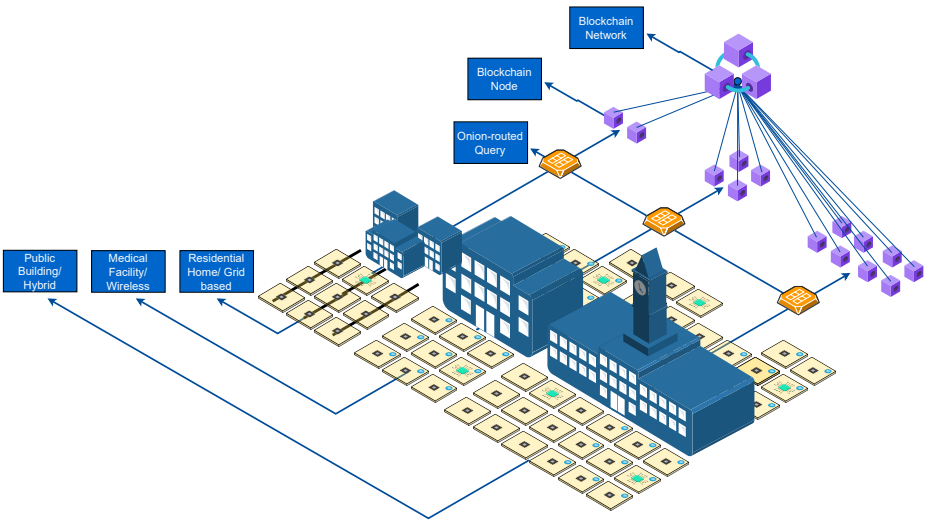

Figure 1. High level view of the presented architecture.

3.1. Cost Aspects of the Proposed Solution

The proposed solution was implemented with cost effectiveness as one of the most important factors. The retail value of the ICT hardware employed in the solution should not exceed USD 100 per square meter (10.76 square feet). One square meter would occupy nine tiles, around USD 20 for the controller and less than USD 80 for the nine pressure sensors. The solution scales linearly with no additional cost.

3.2. Non-Intrusive, Privacy-Preserving Indoor Location

Indoor location has many applications for structural health monitoring, studying the behavioral patterns of a building's occupants and health-related issues such as locating lost patients with memory and orientation disorders, healthy activities, etc. Most existing solutions for indoor location rely on wearable devices (i.e., location-aware bracelets), which require frequent charging and can generate invalid data in case the device is forgotten. Our approach is a passive system that does not require any maintenance or wearable device. We used off-the-shelf force resistors (FSR model 406), which are emended and centered inside a 30 $\times$ 30 cm tile of foam. Once force is applied, the foam and FSR deform, which can be measured as a voltage drop by the controller.

In a wired setting, each tile of foam is shaped like a puzzle piece, which ensures easy assembly. Each tile has two connectors on each face of the square to seamlessly connect to a neighbouring tile. It also includes a small chip for converting the analog signal to a digital that finally allows the collection of sensor readings over a one-wire type protocol. Every three-by-three grid of tiles contains one compute unit, which serves as a controller for the underlying sensors, and a WSN/blockchain node, as depicted in Figures 2–4. Figure 5 illustrates an assembled module in grid mode.

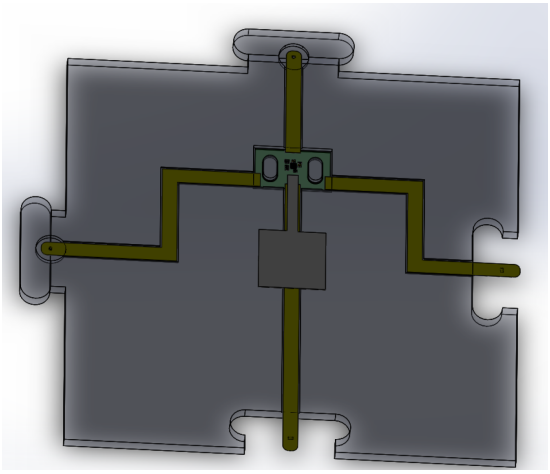

Figure 2. Bottom side of the foam tile.

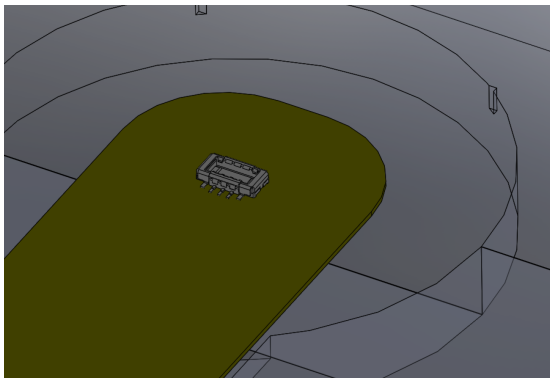

Figure 3. A detailed view of the male connector.

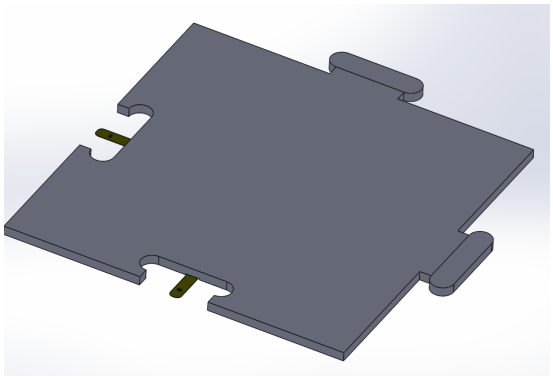

Figure 4. Upper side of the foam tile.

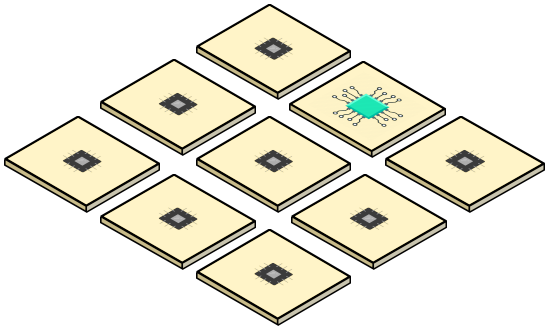

Figure 5. Grid-based connection of individual force sensors.

If physical connections are not suitable, a completely wireless configuration is possible but less cost effective. Figure 6 illustrates a module in full WSN mode.

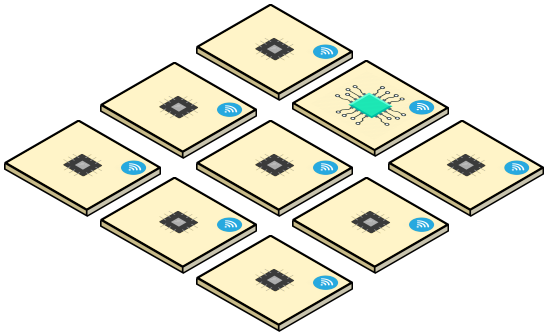

Figure 6. Wireless-enabled force sensors.

4. Secure, and Private Data Filtering, and Aggregation

In this section, we present the data acquisition layer, consisting of the General Purpose Data and Query Privacy Preserving Protocol described in [13]. The communication protocol presented in [13] is characterized by messages containing a layered object made of several encryption layers similar to the one employed in the onion routing [46]. Each layer of the layered object contains the IP address of the next receiver of the message, and since layers are produced using public key cryptography [47], the message must travel through the exact sequence of nodes defined at message construction. The technique of encoding the message path in the message is commonly known as source routing [48]. Path information is carried in encryption layers to restrict the knowledge obtained by nodes processing the message, which only learn about the sender and the next receiver of the message. Therefore, the whole message path is not revealed to any node receiving the message.

In addition to the layered object, messages specified by the communication protocol in [13] include a payload. The payload consists of computer code specified in a general-purpose programming language and a binary string that stores an aggregate. Therefore, it includes instructions specifying the data to retrieve and the aggregated data of nodes in the message path. In the following, we will refer to the onion message (OM) as a structure consisting of the layered object and the aforementioned payload. The OM payload is secured by symmetric key encryption to prevent malicious actors from tracking the OM and obtain values added by sensor nodes by comparing the aggregate pre and post OM processing. Moreover, encryption keys required to decipher the OM payload are delivered only to specific nodes in the OM path by enclosing symmetric encryption keys in the layered object. Nodes in the OM path are either: (a) processing the OM or (b) emulating OM processing.

(a) Nodes processing the OM obtain two symmetric encryption keys and the next-hop IP address from layer decryption of the layered object. The first symmetric encryption key is used to access the content of the OM payload. Next, the node executes the computer code and embeds results in the binary string. The OM payload is then encrypted using the second symmetric encryption key, and after a time-span affected by randomness, the OM is forwarded to the next-hop node.

(b) Nodes emulating OM processing only obtain the next-hop IP address from layer decryption of the layered object. These nodes retain the OM without accessing the payload for a time-span similar to nodes processing the OM, and then the message is forwarded to the next-hop node.

Therefore, external actors observing network communications are not able to identify nodes contributing to the aggregated result; consequently, they cannot associate activities occurring in the monitored environment with messages transiting network nodes.

4.1. Data Filtering and Aggregation

The framework for privacy-aware indoor location makes use of the privacy-preserving communication protocol described in [13] to securely convey to sensor nodes information related to data filtering and aggregation while maintaining the identity of interested nodes hidden from other entities except the message's origin.

The information related to data filtering and aggregation is delivered to sensor nodes in the form of computer code included in the payload of the previously described OM. Sensor nodes processing the OM execute the delivered computer code in a secure execution environment. The execution environment provides restricted access to the underlying sensor node system, allowing the executing computer code to access sensor readings recorded in the last h hours (h a fixed network parameter).

Since the described technique conveys general-purpose computer code to sensor nodes, it is possible to compute virtually any operation on the data of sensor nodes. Therefore, the presented technique can be used to count the number of persons in an environment, identify when and where the social distancing is violated, determine if a room was over-utilized and needs cleaning to prevent the spreading of the virus, etc.

In the following, we show how to verify if the social distancing is violated in a specific area of the monitored environment. The processing of a similar request begins as described in [13]. The sink node receives the request expressing the operation and the target location and starts constructing the OM to answer the request. First, the required operation is converted into a task specified in a general-purpose programming language. The task pseudo-code for addressing the verification of social distancing is shown in Algorithm 1. Then the set of nodes target of the request is selected and the sink node starts constructing OM. Since the communication protocol [13] relies on messages uniform size, the request will be resolved by issuing multiple OM.

An OM including the task given in Algorithm 1 being processed on a node of the smart floor sensor network described in Section 3.2 will perform the following: The data of the sensor network node is first filtered to a narrow time interval ($time_{start}$ and $time_{end}$); the narrower the time interval is, the more accurate the data acquired. All objects detected are filtered from the data by observing the data variance. Then, the data are filtered using the function FILTERSTATIONARY($data$,$time$) to remove all non-stationary activities, and the $time$ argument is used to determine when an activity is considered non-stationary. The observed phenomenon is considered non-stationary when it leaves the sensor in an amount of time lower than $time$ milliseconds. The threshold value must be of $time > (time_{end} - time_{start}) * \frac{1}{2}$; otherwise, repeat event detection may occur. The filtered data are discretized into a value array of underlying sensors, each value describing the number of observed events. The array of values is then stored in w, the data carrying binary string at the position determined by the two symmetric encryption keys and the linear probing technique. Since both symmetric encryption keys are known only to the current node and to the message's origin, other nodes processing the OM cannot identify which node contributed to which value in the data-carrying binary string. The OM is then reassembled and sent to the next-node IP address.

When the OM ends its path at the issuer sink node, the sink node uses the symmetric encryption key obtained from layered object decryption to decipher the OM payload and access the data carrying string. Moreover, the symmetric encryption key acts as the OM identifier. Thus, the sink node can uniquely identify the OM and use information about symmetric encryption keys and the OM path maintained from OM construction to associate the data in the data-carrying string to nodes in the OM path.

Therefore, the sink node gathers the results of all OM issued to resolve a request and uses the collected data to reconstruct the environment representation as shown in Figure 7 to detect where and when the social distancing was violated.

Algorithm 1: Data filtering to supervise social distancing violations

Input: D sensor node data
w binary string
S_1, S_2 symmetric encryption keys

Output: w' modified binary string

Function *Main (args)*:
t_{start},t_{end} ; // Time interval
$t_{contact}$; // Time in milliseconds

// Filter the interval of data between $time_{start}$ and $time_{end}$
D = FilterTimeInterval (D,t_{start},t_{end});
// Exclude objects from the data
D = FilterObjects (D);
// Remove all activities that are not stationary for $t_{contact}$ milliseconds
D = FilterStationary (D,$t_{contact}$);
// Discrertize the data into an value array of underlying sensors, each value describing the number of observed events
$tile_{status}$ = DiscretizeData (D);
// Use encryption keys to find the position in w where insert the data
$pos = (S_1 + S_2)\% \frac{size(w)}{size(tile_{status})}$;

// Use linear probing to insert data in w
while $w[pos * size(tile_{status})]! = null$ **do**
  $pos+ = 1$;
end
$w[pos * size(tile_{status})] = tile_{status}$;
return w ;
end

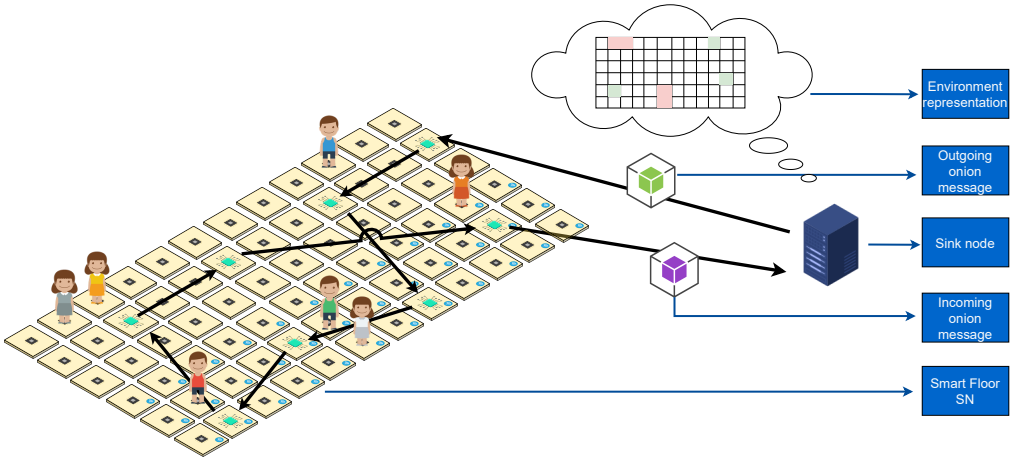

Figure 7. The figure displays the data acquisition layer relying on the privacy-preserving communication protocol in [13]. The environment representation highlights where the social distancing was violated (red-colored squares).

5. Blockchain for Secure Storage and Computation

Blockchain provides a secure, decentralized, transparent and immutable record that has gained a lot of attention. The unique set of properties it provides have directed researchers to seek other uses besides cryptocurrency. The first practical implementation

was Bitcoin, which uses Proof of Work to secure the blockchain coupled with the unspent transaction output (UTXO) transaction model. Even with the limited expressing power of Bitcoin's UTXO model, researchers have demonstrated that an access control can be built [40]. Microsoft implemented a decentralized identity solution running on the Bitcoin network [49], and Factom protocol, which uses the Bitcoin network as a decentralized notary service [50]. Smart contract platforms such as Ethereum use a state-based model in which state transitions are recorded in blocks. This paved the way for the development of smart contracts, Turing complete programs that are recorded on-chain. With smart contracts, more complex applications can be built. Our framework uses the OpenEthereum [51] private network as a smart contract platform that facilitates two main modules, namely Role-based access control (RBAC) and decentralized sink node for the underlying WSNs. In a permissioned setting, Ethereum is configured to run a proof of authority (PoA), in which only a selected group of nodes are configured as validators. In our use case, each building with an indoor location system operates at least one OpenEthereum node. However, preferably, most compute units that serve as sinks should run a light client.

5.1. WSN Sink

In order to perform queries and computation, WSNs are usually deployed with a sink node. Sensors in the network collect information from the environment and ultimately transfer the data to the sink node. In practical implementations, sink nodes usually reside in the cloud and seldom on-site. Whatever the case, sink nodes arguably present a single point of failure (SPOF) of the entire system [52]. Moreover, sink nodes are easier to identify as a target due to their fixed network identity and recognizable traffic patterns. In our solution, we achieve complete decentralization by replacing sink nodes with smart contracts. The sink contract keeps a record of public keys of all nodes in the network. The publicly exposed function *sendQuery()* enables users to initiate a query on a set of tiles and retrieve the result once submitted on-chain. The contract keeps a registry of all the computing nodes, their public keys, and references to which building/area they belong to. A query consists of a set of compute units and a function. To obtain the set of compute units, the sender can call the function *getComputingNodes()*, which checks the senders public key against RBAC and filters the set accordingly. The result is a subset of units, the sender has access to. Upon calling *sendQuery()* the contract creates an onion. The subset of computing nodes should be randomized to avoid using on-chain randomness when creating the onion.

Computing nodes in the WSN run a light client of OpenEthereum and are able to synchronize blocks with reasonable storage and resource requirements. Upon receiving a new block, each node checks the list of added onions to determine the starting node on the route. This is made possible by keeping encryption integrity checks on the first layer. The node whose key passes the integrity check is able to decrypt the first layer and initiate the query. Note that even if the onion is publicly available, no third party can decrypt it or determine the route the query will take. From a network point of view, every query has a sink node, which is pseudo-randomly selected amongst the set of nodes in the underlying WSN, as detailed in Section 4.

The route ends at the starting node, which submits a transaction to the contract storing the result of the computation encrypted with the public key of the original sender. This protects the results on the public ledger so that only the owner of the corresponding private key can view them.

5.2. Role-Based Access Control

RBAC is a smart contract deployed on the blockchain that allows the creation, removal, revocation, and transfer of roles to actors that interact with the sink node contract and underlying WSNs that are queried. Upon adding a new building, the transaction signer is automatically given the role of admin. We divide assets into buildings, areas, and sensors. Initially, each sensor must be registered using the public key and assigned to a area within a building. After the configuration, new roles can be assigned to each of the resources by

protecting their getter methods. This enables administrators to limit access to queries on individual area; i.e., an open space in a public building can be queried by anyone to learn how crowded it is. However, the offices of the public building can only be queried by the manager and occupants. Each of the public functions exposed by the sink contract is first filtered by the RBAC to determine if access is granted.

6. Validation

To validate the proposed privacy-aware framework, we designed an experiment to assess the average response time. We define the response time as the elapsed time between the execution of the sink smart contract and the subsequent transaction storing the result of the data filtering and aggregation. To conduct this investigation, we individually considered the latency introduced by blockchain operations (sink contract execution and subsequent result transaction) and the technique presented in Section 4 for data filtering and aggregation. Specifically, we validated the privacy-aware framework for the wireless configuration of the floor location system. We considered only the wireless configuration since the latency introduced by messages moving in the wireless multi-hop network is inherently higher than in wired settings.

6.1. Data Filtering and Aggregation

To evaluate the data filtering and aggregation duration, we used the simulator PPWSim [53]. PPWSim is based on the NS3 discrete-event simulation environment for Internet systems [54] and is designed to simulate the General Purpose Data and Query Privacy Preserving Protocol described in [13] and estimate network delays. We refer to the network delay as the latency for an OM (onion message) to travel from one node to the node at the next-hop address. To obtain valuable results to validate the proposed framework, we further extended PPWSim to estimate the delay of OM processing.

Experimental Setup

Since the detailed simulation description can be found in [53], in the following, we will outline the simulator parameters selected to obtain network delay results.

The simulator was set up to construct an ad hoc wireless network of 200 nodes. Nodes were deployed according to a grid structure; each node was equidistant from the closest nodes in cardinal directions. The simulated wireless communication conforms to the IEEE 802.11n standard operating at 2.4 GHz at the data rate of 13 Mbps (Modulation Coding Scheme index 1), abd the wireless communication range was set up to allow direct communication only between neighbouring nodes. The maximum transmission unit and maximum segment size were set to the ns-3 default value, 2296 bytes and 536 bytes, respectively.

OMs were transmitted over the TCP protocol, and the routing of packets in the multi-hop network was handled using the Optimized Link State Routing Protocol (OLSR) [55].

As described in [53], the simulator operates by issuing OMs from a node in the center of the network. OMs are issued sequentially; after an OM returns back to the issuer node, the following OM is issued. The central node was set up to issue 30 OM for each value of $n = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, the OM path length. OMs are constructed by randomly selecting n nodes to include in the OM path. The OM path is encoded in the layered object. Encryption layers of the layered object are produced using an ECC-based [56] public-key cipher of 256 b key length implemented in the Libsodium library [57]. Each encryption layer includes a shared secret, the next-hop IP address, two 32b symmetric encryption keys, and the inner encryption layer. To replicate the transfer of computer code, OMs are including a payload consisting of padding p = 2.5 k bytes. The OM size at n path length is given in Table 1.

To assess the OM processing delay using PPWSim, we had to first estimate Δ_{om} the maximum execution time of an OM. As described in [13], the Δ_{om} is a fixed network parameter depending on implementation specifics. The Δ_{om} is used to bound the OM

sojourn time on nodes to only a specific amount in order to achieve privacy preservation, as discussed in Section 4. To estimate the Δ_{om} specific to the privacy-aware framework, we measured delays introduced at each step of the OM execution on a node of the floor location system described in Section 3.2. Sixteen FSR sensors characterize each node of the floor location system, and one compute unit, in our implementation the ESP32-DevKitC V4. Table 2 presents the OM execution broken in individual operations, and the delays of operations are reported. We emphasize the fact that in the privacy-aware framework, the nodes of the floor system are executing only operations presented in Table 2. The OM construction involving the computation of many public-key encryption layers is achieved by the smart contract; therefore, this was executed on validator nodes of the blockchain as described in Section 5 and discussed in Section 6.2.

Table 1. Size of the layered object at the selected OM path lengths n. The row total gives the OM total size, including the payload of 2.5 kB.

n	10	20	30	40	50	60	70	80	90	100
Layered object (bytes)	840	1680	2520	3360	4200	5040	5880	6720	7560	8400
Total (bytes)	3340	4180	5020	5860	6700	7540	8380	9220	10,060	10,900

Table 2. OM execution broken in individual operations; the operation execution time was measured on the ESP32-DevKitC V4. Cryptographic operations were carried out using the Libsodium library [57]. The public-key cipher is ECC based using Curve25519 [58] and the symmetric key cipher is ChaCha20.

Operation	ECC Decryption	ECC Decryption	ChaCha20 Encryption	ChaCha20 Decryption	Data Processing
Data	1 B	1 kB	2.5 kB	2.5 kB	15 kB
Execution time	18.4 ms	18.9 ms	1.2 ms	1.1 ms	9.8 ms

Based on the data in Table 2, we estimated that the Δ_{om} appropriate to our system specifics is 35 ms. This value was derived for the OM size at the path length $n = 100$. As reported in Table 2, the ECC decryption is computation intensive only in deriving the shared secret. The ECC decryption of the layered object of 8400 bytes requires 22.2 ms, payload decryption and encryption require 2.3 ms, and payload content execution requires 9.8 ms. Therefore, we obtained a rough estimate of $\Delta_{om} = 35$ ms.

The Δ_{om} estimate was included in the PPWSim following the guidelines defined in [13] specifying that the technique ensures privacy preservation if the OM sojourn time on WSN nodes corresponds to $\Delta_{om} \times r$. r is a randomly chosen float bounded by $1 \leq r \leq 5$.

Therefore, in the extended version of PPWSim, nodes receiving the OM decipher the outer encryption layer of the layered object to reveal the next-hop IP address and the inner encryption layer. The layered object size is uniform by adding the padding of the same number of bytes as the removed layer. The payload is of uniform size, and after the sojourn time of $\Delta_{om} \times r$, the OM is forwarded to the next-hop node. Measurements are taken separately for network delays and OM processing, and the results are presented, respectively, in Figure 8 and Table 3.

Table 3. Delay introduced by OM processing at n nodes. Average and standard deviation are computed for 30 OMs at each value of n.

n	10	20	30	40	50	60	70	80	90	100
mean (seconds)	1.12	2.28	3.23	4.24	5.35	6.38	7.39	8.60	9.55	10.61
std	0.018	0.022	0.046	0.078	0.080	0.092	0.121	0.153	0.097	0.110

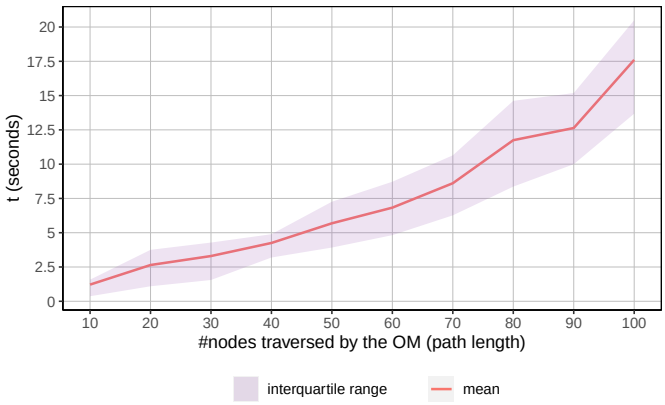

Figure 8. Required time for an OM to travel the selected path length. Measurements do not include the OM processing delay. Statistics are computed for 30 OMs at each OM path length.

6.2. Blockchain

As described in Section 5, the privacy-preserving framework relies on a PoA Ethereum blockchain maintained by a selected group of validator nodes operating in server farm-like settings. Therefore, the smart contracts responsible for RBAC and OM creation are executed on high-performance machines. Several studies [44,59] provide evidence that RBAC could function in similar settings, and reported results show that RBAC operations require low resource consumption. On the other hand, OM creation requires several public-key cryptography operations. We measured the time to create an OM of 100 encryption layers using Curve25519 [58] on a standard laptop (CPU: Intel i5, RAM: 16 GB). The OM construction took 19 ms of CPU time.

However, the time required to execute the mentioned contracts is negligible in the assessment of the framework response time since the blockchain state is propagated only at new block creation. Therefore, the nodes of the floor system running the light client can detect a new OM only after a new block is added to the blockchain. The PoA Ethereum block period is usually in the range from 2 to 15 s [60].

6.3. Discussion

We provided the validation of the WSN framework for privacy-aware indoor location by assessing its response time. The reported results show that applying the PoA Ethereum on the floor system does not introduce significant latency in response times. Nonetheless, it binds the detection of new OMs and the result transaction to a discrete basis imposed by the block period.

Moreover, the results show the applicability of the General Purpose Data and Query Privacy Preserving Protocol [13] to the indoor location floor system. The data in Figure 8 and Table 3 show that in the extreme scenario of OM path length $n = 100$, the OM Round-Trip-Time is generally less than 30 s. However, in practical implementations, the system will rather rely on multiple smaller OMs executed in parallel than one large OM. Therefore, the framework response time is reduced to approximately 10 s + two block periods if parallel OM execution is applied at $n = 50$.

7. Conclusions and Future Work

In this paper, we present a system for privacy-preserving, non-intrusive, and secure indoor location monitoring. We specifically design the system to not allow identification through data filtering. We present an innovative way of passively approximating location by measuring the force applied to the floor. We are able to distinguish objects from persons by observing the activity at the local level. The sensitized floor forms a WSN that is secure

from both external and internal adversaries. By designing a unique onion routing-based protocol, we were able to conceal the network identity of nodes in the WSN. Moreover, our onion-based approach allows a general-purpose computing model for distributed algorithms, and to the best of our knowledge, no comparable solution exists. To address the issue of SPOF on sink nodes, we used blockchain-based smart contracts that replace the onion creation and storage of query results. The blockchain operates in permissioned mode in which sink nodes are registered, and their public keys stored on the blockchain. We also show how using a blockchain-based RBAC is possible to further protect the query and data access. We validate our solution on our use case of tracking violations of indoor physical distancing restrictions to avoid the spread of COVID-19.

The presented solution aims at an implementation of a self-managing system to control the compliance to a set of predefined rules, such as the COVID-19 pandemic rules issued by local governments. The set of rules can be arbitrarily defined and modified without requiring updates of sensor nodes.

A typical use-case for the presented system would be the installation in a nursing home. The occupants are automatically pseudo-identified by the system in bedrooms and later tracked along the corridors of the building, ensuring an overview of the number of occupants in specific areas, triggering temporary blocks and sanitizing actions.

The system cannot be used as a critical contact signalling system (such as the exposure notification system by Google and Apple) as it is lacking a backward loop that would enable the information about an infection or critical contact to be attributed to a specific person.

Future work should explore more sophisticated algorithms for the detection and tracking of users. Data should be analyzed to advance our understanding of behavior in an effort to improve future building designs. Implementations that aim to identify occupants (i.e., elderly homes) should explore a key management scheme and extend the smart contracts to include the ability for users to grant the system permissions to use their data.

Author Contributions: Conceptualization, A.T.; methodology: A.T., N.H. and J.V.; software, A.T. and N.H.; validation, A.T., J.V. and N.H.; formal analysis, A.T., J.V. and N.H.; investigation, A.T., J.V. and N.H.; funding acquisition and resources, J.V.; data curation, N.H.; writing—original draft preparation, A.T., N.H.; writing—review and editing, J.V.; visualization, A.T. and N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by H2020 grant number 739574 and 857188 by the Slovenian Research Agency (ARRS) grant number J2-2504.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge the European Commission for funding the InnoRenew CoE project (H2020 Grant Agreement #739574) as well as the Slovenian Research Agency (ARRS) for supporting project number J2-2504.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Riffe, T.; Acosta, E. Data Resource Profile: COVerAGE-DB: A global demographic database of COVID-19 cases and deaths. *Int. J. Epidemiol.* **2021**, *50*, 390–390f. [CrossRef]
2. Bale, R.; Li, C.G.; Yamakawa, M.; Iida, A.; Kurose, R.; Tsubokura, M. Simulation of droplet dispersion in COVID-19 type pandemics on Fugaku. In Proceedings of the Platform for Advanced Scientific Computing Conference, Geneva, Switzerland, 5–9 July 2021; pp. 1–11.
3. Ooi, C.C.; Suwardi, A.; Ouyang, Z.L.; Xu, G.; Tan, C.K.I.; Daniel, D.; Li, H.; Ge, Z.; Leong, F.Y.; Marimuthu, K.; et al. Risk assessment of airborne COVID-19 exposure in social settings. *Phys. Fluids* **2021**, *33*, 087118. [CrossRef]
4. Sun, C.; Zhai, Z. The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. *Sustain. Cities Soc.* **2020**, *62*, 102390. [CrossRef] [PubMed]

5. Singh, R.P.; Javaid, M.; Haleem, A.; Suman, R. Internet of things (IoT) applications to fight against COVID-19 pandemic. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2020**, *14*, 521–524. [CrossRef] [PubMed]
6. Singh, P.K.; Nandi, S.; Ghafoor, K.Z.; Ghosh, U.; Rawat, D.B. Preventing covid-19 spread using information and communication technology. *IEEE Consum. Electron. Mag.* **2020**, *10*, 18–27. [CrossRef]
7. Kumar, K.; Kumar, N.; Shah, R. Role of IoT to avoid spreading of COVID-19. *Int. J. Intell. Netw.* **2020**, *1*, 32–35. [CrossRef]
8. Dong, Y.; Yao, Y.D. IoT platform for COVID-19 prevention and control: A survey. *IEEE Access* **2021**, *9*, 49929–49941. [CrossRef]
9. Garg, L.; Chukwu, E.; Nasser, N.; Chakraborty, C.; Garg, G. Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model. *IEEE Access* **2020**, *8*, 159402–159414. [CrossRef]
10. Chan, H.; Perrig, A. Security and privacy in sensor networks. *Computer* **2003**, *36*, 103–105. [CrossRef]
11. Gao, Y.; Ao, H.; Feng, Z.; Zhou, W.; Hu, S.; Tang, W. Mobile network security and privacy in WSN. *Procedia Comput. Sci.* **2018**, *129*, 324–330. [CrossRef]
12. Shi, Q.; Zhang, Z.; He, T.; Sun, Z.; Wang, B.; Feng, Y.; Shan, X.; Salam, B.; Lee, C. Deep learning enabled smart mats as a scalable floor monitoring system. *Nat. Commun.* **2020**, *11*, 4609. [CrossRef] [PubMed]
13. Hrovatin, N.; Tošić, A.; Mrissa, M.; Vičič, J. A General Purpose Data and Query Privacy Preserving Protocol for Wireless Sensor Networks. *arXiv* **2021**, arXiv:2111.14994.
14. Goldschlag, D.M.; Reed, M.G.; Syverson, P.F. Hiding routing information. In *International Workshop on Information Hiding*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 137–150.
15. Thulasiraman, P.; Haakensen, T.; Callanan, A. Countering passive cyber attacks against sink nodes in tactical sensor networks using reactive route obfuscation. *J. Netw. Comput. Appl.* **2019**, *132*, 10–21. [CrossRef]
16. Brena, R.F.; García-Vázquez, J.P.; Galván-Tejada, C.E.; Muñoz-Rodriguez, D.; Vargas-Rosales, C.; Fangmeyer, J. Evolution of indoor positioning technologies: A survey. *J. Sens.* **2017**, *2017*, 2630413. [CrossRef]
17. Farahsari, P.S.; Farahzadi, A.; Rezazadeh, J.; Bagheri, A. A Survey on Indoor Positioning Systems for IoT-based Applications. *IEEE Internet Things J.* **2022**, *early access*. [CrossRef]
18. Want, R.; Hopper, A.; Falcao, V.; Gibbons, J. The active badge location system. *ACM Trans. Inf. Syst. (TOIS)* **1992**, *10*, 91–102. [CrossRef]
19. Forestiero, A. Metaheuristic algorithm for anomaly detection in Internet of Things leveraging on a neural-driven multiagent system. *Knowl.-Based Syst.* **2021**, *228*, 107241. [CrossRef]
20. Forestiero, A. Self-organizing anomaly detection in data streams. *Inf. Sci.* **2016**, *373*, 321–336. [CrossRef]
21. Orr, R.J.; Abowd, G.D. The smart floor: A mechanism for natural user identification and tracking. In Proceedings of the CHI'00 Extended Abstracts on Human Factors in Computing Systems, The Hague, The Netherlands, 1–6 April 2000; pp. 275–276.
22. He, C.; Zhu, W.; Chen, B.; Xu, L.; Jiang, T.; Han, C.B.; Gu, G.Q.; Li, D.; Wang, Z.L. Smart floor with integrated triboelectric nanogenerator as energy harvester and motion sensor. *ACS Appl. Mater. Interfaces* **2017**, *9*, 26126–26133. [CrossRef]
23. Kaddoura, Y.; King, J.; Helal, A. Cost-precision tradeoffs in unencumbered floor-based indoor location tracking. In Proceedings of the Third International Conference On Smart Homes and Health Telematic (ICOST), Sherbrooke, QC, Canada, 4–6 July 2005.
24. Shen, Y.L.; Shin, C.S. Distributed sensing floor for an intelligent environment. *IEEE Sens. J.* **2009**, *9*, 1673–1678. [CrossRef]
25. Yang, X.; Gao, L.; Zheng, J.; Wei, W. Location privacy preservation mechanism for location-based service with incomplete location data. *IEEE Access* **2020**, *8*, 95843–95854. [CrossRef]
26. Shen, H.; Zhou, J.; Cao, Z.; Dong, X.; Choo, K.K.R. Blockchain-based lightweight certificate authority for efficient privacy-preserving location-based service in vehicular social networks. *IEEE Internet Things J.* **2020**, *7*, 6610–6622. [CrossRef]
27. Selamneni, V.; Dave, A.; Mondal, S.; Mihailovic, P.; Sahatiya, P. Large Area Pressure Sensor for Smart Floor Sensor Applications—An Occupancy Limiting Technology to Combat Social Distancing. *IEEE Consum. Electron. Mag.* **2021**, *10*, 98–103. [CrossRef]
28. Tošić, A.; Hrovatin, N.; Vičič, J. Data about fall events and ordinary daily activities from a sensorized smart floor. *Data Brief* **2021**, *37*, 107253. [CrossRef] [PubMed]
29. Gu, T.; Fang, Z.; Abhishek, A.; Mohapatra, P. IoTSpy: Uncovering Human Privacy Leakage in IoT Networks via Mining Wireless Context. In Proceedings of the 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, 31 August–3 September 2020; pp. 1–7.
30. Zhang, F.; He, W.; Liu, X. Defending against traffic analysis in wireless networks through traffic reshaping. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, USA, 20–24 June 2011; pp. 593–602.
31. Saltaformaggio, B.; Choi, H.; Johnson, K.; Kwon, Y.; Zhang, Q.; Zhang, X.; Xu, D.; Qian, J. Eavesdropping on fine-grained user activities within smartphone apps over encrypted network traffic. In Proceedings of the 10th {USENIX} Workshop on Offensive Technologies ({WOOT} 16), Austin, TX, USA, 8–9 August 2016.
32. Middya, R.; Chakravarty, N.; Naskar, M.K. Compressive sensing in wireless sensor networks—A survey. *IETE Tech. Rev.* **2017**, *34*, 642–654. [CrossRef]
33. Zheng, H.; Yang, F.; Tian, X.; Gan, X.; Wang, X.; Xiao, S. Data gathering with compressive sensing in wireless sensor networks: A random walk based approach. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *26*, 35–44. [CrossRef]
34. Hu, P.; Xing, K.; Cheng, X.; Wei, H.; Zhu, H. Information leaks out: Attacks and countermeasures on compressive data gathering in wireless sensor networks. In Proceedings of the IEEE INFOCOM 2014-IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 1258–1266.

35. Xie, K.; Ning, X.; Wang, X.; He, S.; Ning, Z.; Liu, X.; Wen, J.; Qin, Z. An efficient privacy-preserving compressive data gathering scheme in WSNs. *Inf. Sci.* **2017**, *390*, 82–94. [CrossRef]
36. Paillier, P. Public-key cryptosystems based on composite degree residuosity classes. In *International Conference on the Theory and Applications of Cryptographic Techniques*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 223–238.
37. Xu, J.; Yang, G.; Chen, Z.; Wang, Q. A survey on the privacy-preserving data aggregation in wireless sensor networks. *China Commun.* **2015**, *12*, 162–180. [CrossRef]
38. Bista, R.; Chang, J.W. Privacy-preserving data aggregation protocols for wireless sensor networks: A survey. *Sensors* **2010**, *10*, 4577–4601. [CrossRef]
39. Sandhu, R.S. Role-based access control. In *Advances in Computers*; Elsevier: Amsterdam, The Netherlands, 1998; Volume 46, pp. 237–286.
40. Di Francesco Maesa, D.; Mori, P.; Ricci, L. Blockchain Based Access Control. In *IFIP International Conference on Distributed Applications and Interoperable Systems*; Chen, L.Y., Reiser, H.P., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 206–220.
41. Hu, V.C.; Kuhn, D.R.; Ferraiolo, D.F.; Voas, J. Attribute-based access control. *Computer* **2015**, *48*, 85–88. [CrossRef]
42. Sandhu, R.S.; Samarati, P. Access control: Principle and practice. *IEEE Commun. Mag.* **1994**, *32*, 40–48. [CrossRef]
43. Achour, I.; Ayed, S.; Idoudi, H. On the Implementation of Access Control in Ethereum Blockchain. In Proceedings of the 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Virtual, 29–30 September 2021; pp. 483–487.
44. Cruz, J.P.; Kaji, Y.; Yanai, N. RBAC-SC: Role-based access control using smart contract. *IEEE Access* **2018**, *6*, 12240–12251. [CrossRef]
45. Wadaa, A.; Olariu, S.; Wilson, L.; Eltoweissy, M.; Jones, K. On providing anonymity in wireless sensor networks. In Proceedings of the Tenth International Conference on Parallel and Distributed Systems, Istanbul, Turkey, 15–20 July 2004; pp. 411–418.
46. Syverson, P.F.; Goldschlag, D.M.; Reed, M.G. Anonymous connections and onion routing. In Proceedings of the 1997 IEEE Symposium on Security and Privacy (Cat. No. 97CB36097), Oakland, CA, USA, 4–7 May 1997; pp. 44–54.
47. Rivest, R.L.; Shamir, A.; Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM* **1978**, *21*, 120–126. [CrossRef]
48. Sunshine, C.A. Source routing in computer networks. *ACM SIGCOMM Comput. Commun. Rev.* **1977**, *7*, 29–33. [CrossRef]
49. Microsoft. Decentralized Identity. 2018. Available online: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE2DjfY (accessed on 19 December 2021).
50. Snow, P.; Deery, B.; Kirby, P.; Johnston, D. Factom Ledger by Consensus. 2015. Available online: https://cryptochainuni.com/wp-content/uploads/Factom-Ledger-by-Consensus.pdf (accessed on 19 December 2021).
51. Buterin, V. Ethereum white paper. *GitHub Repos.* **2013**, *1*, 22–23.
52. Kohno, E.; Ohta, T.; Kakuda, Y. Secure decentralized data transfer against node capture attacks for wireless sensor networks. In Proceedings of the 2009 International Symposium on Autonomous Decentralized Systems, Athens, Greece, 23–25 March 2009; pp. 1–6.
53. Hrovatin, N.; Tošić, A.; Vičič, J. Ppwsim: Privacy Preserving Wireless Sensor Network Simulator. *SSRN* **2021**. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3978796 (accessed on 19 December 2021).
54. Henderson, T.R.; Lacage, M.; Riley, G.F.; Dowell, C.; Kopena, J. Network simulations with the ns-3 simulator. *SIGCOMM Demonstr.* **2008**, *14*, 527.
55. Clausen, T.; Jacquet, P.; Adjih, C.; Laouiti, A.; Minet, P.; Muhlethaler, P.; Qayyum, A.; Viennot, L. *Optimized Link State Routing Protocol (OLSR)*. RFC; INRIA. 2003. Available online: https://hal.inria.fr/inria-00471712/ (accessed on 19 December 2021).
56. Miller, V.S. Use of elliptic curves in cryptography. In *Conference on the Theory and Application of Cryptographic Techniques*; Springer: Berlin/Heidelberg, Germany, 1985; pp. 417–426.
57. Libsodium. The Sodium Crypto Library. Available online: https://libsodium.gitbook.io/doc/ (accessed on 28 May 2021).
58. Bernstein, D.J. Curve25519: New Diffie-Hellman speed records. In *International Workshop on Public Key Cryptography*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 207–228.
59. Rahman, M.U.; Baiardi, F.; Guidi, B.; Ricci, L. Protecting personal data using smart contracts. In *International Conference on Internet and Distributed Computing Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 21–32.
60. Schäffer, M.; Angelo, M.D.; Salzer, G. Performance and scalability of private Ethereum blockchains. In *International Conference on Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 103–118.

Chapter 3

Conclusions

The studies in this thesis provide several insights into blockchain technology, from the analysis of existing cryptocurrency networks to the application of blockchain on other domains and finally the design of a more efficient protocol.

Existing public blockchain networks such as Bitcoin have left a negative connotation with discussions on energy inefficiency and illicit activity leaving regulators in a difficult position trying to protect consumers. This thesis makes no attempt at clarifying any of the aforementioned issues, rather it examines the space from a technical perspective whilst not shying away from its popular application. The study in Articles (2 and 3) examine the potential of using a very known method for anomaly detection an it's application to graph based networks such as transaction networks. In Article 3, we report that Benford's law can be used as a robust tool for first level screening of entire ledgers. We contribute towards understanding the differences in the structure of each ledger further strengthening the argument for a ledger agnostic analysis tool. We confirm that generally cryptocurrency transaction data conforms to Benford's law and report the following findings:

- Data-sets need to be filtered to exclude programmable transactions such mining pool payouts. These usually occur either periodically or once the miner has accumulated a specific amount of coins. Most mining pools have default presets, which results in a large mount of transactions with the same value (default payout amount). Filtering out transactions that originate from known mining pools is important in order to prevent skewing the distribution of digits.

- Transaction amounts must be converted to the market value at the time of acceptance in our case USD denominated. We assume this a result of general pricing in USD to avoid exposure due to price volatility of underlying cryptocurrency. More importantly, the nonconformity of transaction networks with native amounts increases the expectation of a low false positive rate.

- The highest conformity is achieved by aggregating all transaction on a daily basis.

We conduct empirical tests and show that Benford's law can be use in general graph based networks. Moreover, our findings reported in Article 2 provide evidence that time series graphs conform to Benford's law given sufficiently long intervals. The ability to perform time interval conformity tests we are able to narrow down the search space significantly. Future research should aim at investigating the lower bound on the time interval in which Benford's law can be reliably used.

On the other hand, for blockchain technology to truly have a transformative role, it's application must extend the current boundaries of public blockchains. Our findings report that while the unique properties of blockchain protocols can benefit other domains, there are intricate relationships between protocol design pertaining to the CAP theorem, and the benefits these protocols provide in other domains. Arguably, there are use cases where these trade-offs can be made. However, we argue that there currently is no "one size fits all" solution. Our findings report that protocols with domain specific design in block structure, consensus mechanism, and network topology are better suited in pinpointing the necessary trade-offs for an efficient application.

Specifically, we report that a decentralized orchestrator for a truly decentralized edge computing network can be implemented. The study in Article 1 introduces a unique protocol to derive decentralized and secure randomness that does not require significant computing power thereby making it suitable for edge devices. The proposed protocol uses VDFs as the entropy source for secure randomness. Nodes participating in consensus compute a funtion $p = vdf(bh, bd)$ where p is the proof, bh is the SHA256 hash of the current block, and bd is the difficulty of the current block. We show that p is a sufficiently secure source of entropy for generating randomness. Moreover, given delay imposed on the nodes computing the VDF prevents malicious nodes to peek into the future. Using the shared seed, nodes are able to self-elect into consensus roles for each slot without communication overhead. The theoretical assumptions made in on the protocol level are further strengthened with a full implementation and empirical study of the performance and behaviour of the network. The main conclusions made from observing the networks performance are as follows:

- The consensus protocol reduces the network footprint and is suitable for edge devices with limited resources
- VDF proofs are sufficient sources of entropy, and introduce desired security properties to the protocol
- A decentralized orchestrator can reliably perform migrations to balance the resource consumption across the nodes but requires deterministic execution for verifiability.

- Using CRIU, docker containers can be migrated at run-time with considerable improvement in payload size, and consequently transfer time.

Future work should focus on solving variability of computation pertaining to docker containers. In permissionless network, an attacker can successfully execute the migrated container to produce the migration proof and then stops execution. There is currently no reliable way to verify liveliness of the application after the migration. Existing solutions such as Intel SGX are not generic and inherently impose hardware restrictions on the protocol. Research should focus on utilizing economic incentives and game theory to overcome specific hardware implementations of trusted computation.

Our findings report that a block structure tailored to a specific application is a viable alternative to existing monolithic blockchains where an enforced block structure is modeled as a state machine and versatility is added with smart contracts. In Article 1 and 4 we report two examples of a modified block structure. We show that a completely trustless and server-less multiplayer game architecture can be implemented by encoding game states in blocks. Our simulations show that the protocol is sufficiently scalable, and can address a set of Sybil cheats previously suggested protocols could not. Future research should investigate the feasibility of using the protocol with a full implementation. Moreover, researchers should explore the potential of using economic incentives to stimulate the fruition of a sharing economy in an effort to distribute the currently unused resources evenly.

Finally, we report that functionality of existing monolithic chains such as Ethereum can be extended with the use of smart contracts for specific use-cases where scalability can be sacrificed in favor of security and decentralization. The study in Article 5 makes the case for a more secure, and privacy preserving protocol for sensor networks. Using smart contracts, the underlying sensor network can be abstracted away from users in a secure and privacy preserving way.

Perhaps the most important contribution of this thesis is in giving readers variety in methods and approaches in utilizing blockchain technology for specific use-cases. The overall findings of the studies reported in this thesis suggest that regardless of it's current limitations, blockchain technology can still have a transformative role in many sectors.

Bibliography

[1] C. C. Agbo, Q. H. Mahmoud, and J. M. Eklund. Blockchain technology in healthcare: a systematic review. In *Healthcare*, volume 7, page 56. MDPI, 2019.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, et al. A view of cloud computing. *Communications of the ACM*, 53(4):50–58, 2010.

[3] A. Basiri, E. S. Lohan, T. Moore, A. Winstanley, P. Peltola, C. Hill, P. Amirian, and P. F. e Silva. Indoor location based services challenges, requirements and usability of current solutions. *Computer Science Review*, 24:1–12, 2017.

[4] E. Bayrak, J. Conley, and S. Wilkie. The economics of cloud computing. 2011.

[5] F. Benford. The law of anomalous numbers. *Proceedings of the American philosophical society*, pages 551–572, 1938.

[6] D. Boneh, J. Bonneau, B. Bünz, and B. Fisch. Verifiable delay functions. In *Annual international cryptology conference*, pages 757–788. Springer, 2018.

[7] C. W. Cai. Disruption of financial intermediation by fintech: a review on crowdfunding and blockchain. *Accounting & Finance*, 58(4):965–992, 2018.

[8] S. De Angelis, L. Aniello, R. Baldoni, F. Lombardi, A. Margheri, and V. Sassone. Pbft vs proof-of-authority: Applying the cap theorem to permissioned blockchain. 2018.

[9] T. Dobrilova. How much is the gaming industry worth?, 2019.

[10] P. Dutta, T.-M. Choi, S. Somani, and R. Butala. Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation research part e: Logistics and transportation review*, 142:102067, 2020.

[11] Y. Gao, H. Ao, Z. Feng, W. Zhou, S. Hu, and W. Tang. Mobile network security and privacy in wsn. *Procedia Computer Science*, 129:324–330, 2018.

[12] A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag. Blockchain in healthcare and health sciences—a scoping review. *International Journal of Medical Informatics*, 134:104040, 2020.

[13] B. Hayes. Cloud computing, 2008.

[14] M. Mettler. Blockchain technology in healthcare: The revolution starts here. In *2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)*, pages 1–3. IEEE, 2016.

[15] S. Nakamoto. Bitcoin whitepaper. Technical report, Bitcoin.org, 2008.

[16] P. Nerurkar, D. Patel, Y. Busnel, R. Ludinard, S. Kumari, and M. K. Khan. Dissecting bitcoin blockchain: Empirical analysis of bitcoin network (2009–2020). *Journal of Network and Computer Applications*, 177:102940, 2021.

[17] C. T. Nguyen, D. T. Hoang, D. N. Nguyen, D. Niyato, H. T. Nguyen, and E. Dutkiewicz. Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities. *IEEE Access*, 7:85727–85745, 2019.

[18] S. Noether, A. Mackenzie, et al. Ring confidential transactions. *Ledger*, 1:1–18, 2016.

[19] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu. Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646, 2016.

[20] Superdata. Market brief — 2018 digital games & interactive entertainment industry year in review, 2019.

[21] A. Tošić, N. Hrovatin, and J. Vicic. A wsn framework for privacy aware indoor location. *Appl. Sci*, 12:3204, 2022.

[22] A. Tošić and J. Vičič. A decentralized authoritative multiplayer architecture for games on the edge. *Computing and Informatics*, 40(3):522–542, 2021.

[23] A. Tošić and J. Vičič. Use of benford's law on academic publishing networks. *Journal of Informetrics*, 15(3):101163, 2021.

[24] A. Tošić, J. Vičič, M. D. Burnard, and M. Mrissa. A blockchain-based edge computing architecture for the internet of things. 2021.

[25] J. Vicic and A. Tošić. Application of benford's law on cryptocurrencies. *J. Theor. Appl. Electron. Commer. Res*, 17:313–326, 2022.

[26] A. Yahyavi and B. Kemme. Peer-to-peer architectures for massively multiplayer online games: A survey. *ACM Computing Surveys (CSUR)*, 46(1):9, 2013.

Povzetek v slovenskem jeziku

Uvod

Mineva desetletje od izuma Bitcoin protokola, ki je trenutno največje in najbolj uporabljano omrežje P2P v zgodovini. Od takrat je bilo razvitih veliko novih protokolov, ki naj bi izboljšali protokol Bitcoin [15] ali uvedli nove koncepte, ki izkoriščajo edinstvene lastnosti tehnologije veriženja blokov. Protokoli veriženja blokov (Blockchain Protocols) se uporabljajo za gradnjo omrežij, v katerih se domneva, da vozlišča niso poštena. Vozlišča v omrežju hranijo popolnoma podvojeno lokalno kopijo globalne knjige (Ledger), shranjeno v obliki verige blokov (Blockchain). Veriga je povezana tako, da vsak blok vsebuje zgoščeno vrednost svojega predhodnika. Da bi dodali blok, vozlišča v omrežju oblikujejo soglasje o tem, kateri naj bo naslednji blok. Od začetnega dokaza o delu (Proof of Work) ali Nakamotovega soglasja [15], v katerem vozlišča tekmujejo v reševanju matematičnega problema in s tem dokazujejo porabljeno računsko moč, so bili predlagani tudi drugi mehanizmi soglasja, kot so dokaz o deležu (Proof of Stake – PoS), dokaz o avtoriteti (Proof of Authority – PoA) itd. Obstoječi mehanizmi soglasja omogočajo kompromise med hitrostjo, varnostjo in decentralizacijo. Na dobljeni sistem lahko gledamo kot na nespremenljivo knjigo, ki je javna/pregledna in preverljiva. Morda je največji dosežek tehnologije veriženja blokov ta, da protokol ne predvideva poštenosti sodelujočih vozlišč. Kot takega ga je mogoče obravnavati kot stroj zaupanja, ki ga potencialno nezaupljivi deležniki uporabljajo za odprte, preverljive in pregledne transakcije.

Uporaba tehnologij veriženja blokov v drugih sistemih in razvoj novih konceptov je ustvarila obilico inovacij in rast raziskovalnih dejavnosti na tem področju. Področja, kot so zdravstvo, upravljanje dobavne verige in finance, so bila deležna veliko pozornosti zaradi svojih očitnih potencialnih koristi. Naše raziskave širijo iskanje možnih koristi tehnologije veriženja blokov v sistemih, kjer je njihova korist manj očitna. Identificiramo štiri teme, iz področja tehnologije veriženja blokov, kjer lahko le-ta zagotovi uporabne učinke, izboljšave ali sistemske transformacije obstoječih rešitev. Čeprav so predstavljene raziskave na prvi pogled nekoliko nepovezane, jih povezujejo prispevki k uporabi in sprejemanju tehnologije veriženja blokov. Zaradi raznolikosti

raziskovalnih tem so razdelki edinstveno strukturirani tako, da obravnavajo vsako temo posebej. Vsak razdelek sledi splošni strukturi disertacije.

K decentraliziranemu robnemu računalništvu

V zadnjih letih je računalništvo v oblaku postalo pogosto uporabljena arhitektura za večino aplikacij. Geografski premik računalniških storitev (centralizacija k ponudnikom oblačnih storitev) so spodbudili številni dejavniki, od enostavnosti vzdrževanja programske opreme [4], zanesljive kakovosti storitev (Quality of Service – QoS) [13], prilagodljivosti strojne opreme in stroškov (Capital expenditures – CapEx do Operating Expenses – OpEx) [2], itd. Vendar s pričakovano rastjo generiranja in porabe podatkov ter shranjevanja in zagotavljanja storitev v okoljih računalništva v oblaku arhitektura potiska zahteve glede pasovne širine omrežja do zgornjih meja [19]. Robno računalništvo v svoji najpreprostejši obliki je mogoče opredeliti kot arhitekturo, v kateri je računanje premaknjeno na rob omrežja, da se izkoristi geografska bližina za zmanjšanje zakasnitve in izboljšanje pasovne širine. Ta nedavna sprememba paradigme poskuša obravnavati preveč geografsko centralizirano arhitekturo oblaka. Vendar pa distribucija storitev do roba prinaša nove izzive, kot so dodeljevanje virov, selitev storitev in aplikacij, zaupanje, itd.

K robustni analizi transakcijskih omrežij kriptovalut

Od začetka Bitcoina je bilo razvitih veliko alternativnih protokolov. Nekateri še vedno temeljijo na verigi blokov, kjer se transakcije shranjujejo in posledično časovno žigosajo v blokih, da se s soglasjem ustvari kanonična veriga. Drugi uporabljajo podatkovne strukture, ki temeljijo na usmerjenih acikličnih grafih, kjer ni ene same kanonične verige. Namesto tega se transakcije sklicujejo in potrjujejo prejšnje transakcije, da bi povečali prepustnost sistema z žrtvovanjem decentralizacije in varnosti. Poleg tega je mogoče spremeniti strukturo transakcije, da se doseže zasebnost, tj. z uporabo obročnih podpisov v Monero [18]. Forum [14] napoveduje, da bo do leta 2025 10% svetovnega domačega proizvoda shranjenega v javnih knjigah, ki temeljijo na verigi blokov. Naraščajoče zanimanje je navdihnilo številne razvijalce, raziskovalce in inovatorje, da prizadevajo iskanju izboljšav in premikanju omejitev obstoječih sistemov . Učinke je mogoče opazovati v velikem številu kriptovalut in omrežij, ki trenutno obstajajo. Hitrost s katero nastajajo nova omrežja dodatno povečuje tveganje za regulatorje pri zaščiti potrošnikov in stabilnosti finančnega sistema. Ob predpostavki, da goljufije rastejo vzporedno s hitrostjo in skupno vrednostjo, zaklenjeno v teh omrežjih, je metoda za hitro in učinkovito odkrivanje nepravilnosti izjemnega pomena. Vendar pa morajo z rastjo inovacij v tem prostoru uporabljene tehnike iskati generično rešitev,

ki nima veliko predpostavk o strukturi blokov in transakcij.

Zagotavljanje tehnološko agnostičnega orodja za analizo odprtih knjig za sprožanje alarmov glede sumljivega vedenja, ki zahteva nadaljnjo, bolj natančno analizo, ima jasne prednosti. Čeprav je od prve transakcije prve kriptovalute - Bitcoin (BTC) [15] minilo več kot desetletje, je bilo šele zadnjih nekaj let shranjenih zadostno število transakcij v dovolj dolgem časovnem okviru, ki omogoča zanesljive statistične analize.

K decentralizirani arhitekturi večigralskih iger

Leta 2019 je bila industrija iger na srečo vredna skoraj 135 milijard z ocenjeno rastjo 10% na leto [9]. Nedavni trendi v smeri iger za več igralcev so bili zelo uspešni, saj so igre, kot je Fortnight, samo v letu 2018 zaslužile več kot 2,4 milijarde prihodkov [20]. Steam, največja platforma za distribucijo iger, poroča, da ima kar 18,5 milijona sočasnih uporabnikov. Ta obseg povpraševanja zahteva, da se strežniki, ki podpirajo računalništvo v oblaku, selijo v realnem času. Poleg tega se je omrežna zakasnitev zmanjšala zaradi lokalizacijskih pristopov, pri katerih se strežniš ustvarijo geografsko blizu odjemalcev, če je to mogoče. Vendar pa je vzdrževanje na tisoče ali celo milijone igralcev skupaj s strojno in programsko infrastrukturo zelo drago in kompleksno [26]. Nedavno idejo o "delitveni ekonomiji" je mogoče uporabiti skupaj s premikom paradigme na robno računalništvo. Natančneje, odjemalci na robu sistema lahko izkoristijo skupno rabo virov, kot sta pasovna širina in računalniška moč, s čimer razbremenijo centralizirane strežnike.

To je mogoče doseči z uporabo arhitekture enakovrednih (Peer to Peer – P2P). Arhitekture iger P2P so bile obsežno preučene, vendar niso bile široko sprejete [26]. Glavne težave so tesno povezane s pomanjkanjem avtoritete in zaupanja. Centralizirane arhitekture te težave rešujejo z centraliziranimi strežniki, ki služijo kot avtoriteta. Naloge strežnika so simulacija igranja, preverjanje in razreševanje konfliktov v simulaciji ter shranjevanje stanja igre. Arhitekture P2P za več igralcev so prej lahko obravnavale nekatere vektorje goljufanja, vendar so zahtevale določeno raven centralizacije.

K notranji lokaciji brez zaupanja in ohranjanju zasebnosti (WSN)

Notranji lokacijski sistem je lahko ena izmed mnogih aplikacij, ki se zavedajo lokacije, na področjih medicine, robotike, industrijske optimizacije, psihologije, varnosti itd. Večina trenutnih rešitev zahteva poznavanje položaja stanovalcev znotraj stavbe v danem trenutku. Obstoječi pristopi k zbiranju podatkov o lokaciji na kraju samem imajo tako težave z uporabnostjo kot tehnološke ovire. Tipične izvedbe vključujejo, vendar niso omejene, nosljive naprave (tj. lokacijsko zaznavne zapestnice), ki jih lahko

nevedni uporabniki zavržejo ali pa zahtevajo pogosto interakcijo (npr. polnjenje baterije), podporo na kraju samem in vzdrževanje. Senzorska omrežja, ki se ne zanašajo na nosljive naprave, običajno vključujejo kamere in mikrofone skupaj s programsko opremo za samodejno zaznavanje obrazov, ki imajo psihološki vpliv na potnike in povzročajo pomisleke glede zasebnosti.

Čeprav šifriranje učinkovito zagotavlja zasebnost podatkov, bi spremljanje dejavnosti v zaprtih prostorih z zanašanjem na brezžične naprave interneta stvari lahko razkrilo kontekstualne informacije o prenosu podatkov [11], kar ne bi predstavljalo le tveganja za zasebnost posameznikov, ampak tudi ogrozilo varnost zgradbe.

Raziskovalni nameni, cilji in hipoteze

Splošni cilj te študije je raziskati primernost omrežij blockchain za reševanje težav in izboljšanje delovanja sistema na več področjih. Preizkusili bomo lastnosti omrežij blockchain, identificirali možnosti za izboljšave in ustvarili napredek k trenutnim implementacijam, ki povečujejo zmogljivost in širijo uporabnost omrežij blockchain. Natančneje, obravnavamo naslednja raziskovalna vprašanja:

RC-1: Ali je mogoče implementirati decentralizirani orkestrator za selitev aplikacij v realnem času?
H-1: Skupna uporaba eksperimentalne kontrolne točke/obnovitve v uporabniškem prostoru (CRIU) in verige blokov z razširljivim konsenznim protokolom se lahko uporabi za izvajanje decentralizirane orkestracije brez ene same točke napake (SPOF).
RC-2: Ali je mogoče Benfordov zakon uporabiti za odkrivanje goljufij v transakcijskih omrežjih s kriptovalutami?
H-2: Transakcijska omrežja s kriptovalutami so v skladu z Benfordovim zakonom in jih je mogoče uporabiti za odkrivanje morebitnega nepravilnega vedenja.
RC-3: Ali je možno odpraviti odpornost na Sybil napade v obstoječih P2P arhitekturah za večigralske igre.
H-3: Protokol za veriženje blokov z varno decentralizirano naključnostjo je mogoče uporabiti za dodeljevanje igralcev v instance in shranjevanje stanj. Naključno izbrani kvorum sodnikov lahko zagotovi zadosten odpor proti kategoriji goljufijam Sybil.
RC-4: Ali je učinkovito računanje na podlagi lokacije v zaprtih prostorih možno doseči na decentraliziran način, ne da bi pri tem kršili zasebnost podatkov in računanja?
H-4: Senzorsko omrežje za notranjo lokacijo je mogoče poizvedovati tako, da računske naloge zavijete v večplastno šifriranje, kar odpravi potrebo po občutljivem prenosu podatkov. Za ohranitev zasebnosti senzorjev lahko pametne pogodbe, ki temeljijo na verigi blokov, izvajajo ovijanje računskih nalog in tako ločijo uporabnike od osnovnega

senzorskega omrežja.

Metode

Članek 1: Decentralizirana samouravnotežna arhitektura za splet stvari, ki temelji na verigi blokov

V raziskavi predlagamo nov omrežni protokol, ki temelji na tehnologiji veriženja blokov. Protokol uvaja spremembo običajne stukture blokov, ki namesto transakcij hranijo stanje sistema. Na ta način je vsak nov blok v verigi novo stanje, celotna veriga pa zgodovina prehodov med stanji celotnega sistema. Da bi dosegli zadostno hitrost in ohranili decent ralizacijo predlagamo tudi nov mehanizem soglasja, ki si močno izposoja ideje sodobnih PoS mehanizmov. Decentralizirano naključnost, ki je potrebna za tovrstne mehanizme dosežemo z vpeljavo preverljivih funkcij zakasnitve. Protokol je uporabljen za implementacijo decentraliziranega orkestratorja aplikacij, ki je primeren za naprave na robu. To trditev podkrepimo z popolno implementacijo protokola, obsežno analizo učinkovitosti sistema na testnih omrežjih velikih do 1000 vozlišč.

Članek 2/3: Uporaba Benfordovega zakona v transakcijskih omrežjih kriptovalut in Uporaba Benfordovega zakona na citatnih omrežjih

V študiji 3 smo preiskovali ali so v splošnem grafovska omrežja v skladu z Benfordovim zakonom. Za analizo smo pripravili podatkovne zbirke popularnih kriptovalut. Podatkovne zbirke vsebujejo vrednosti transakcij skozi čas izražene v ameriških dolarjih. V raziskavi poročamo, da so analizirana omrežja v splošnem v skladu z Benfordovim zakonom. Za omrežja, ki niso v skladu pa argumentiramo z dejstvi, da imajo neetična ozadja oziroma v nekaterih primerih tudi dokazana prevara. V raziskavi pokažemo kako lahko Benfordov test uporabljamo kot robustno orodje, ki ne predpostavlja nobene strukture transakcij za detekcijo anomalij in potencialne neetične uporabe. Uporabnost orodja nadaljnje okrepimo z študijo 3, kjer predstavimo uporabno vrednost orodja tudi na drugih grafovskih omrežjih. Prav tako je pomemben doprinos možnost uporabe orodja v primeru podatkov časovne vrste kot so transakcijsih grafi. Testiranje manjših časovnih rezin tako omogoča nadaljnje zožiti preiskovalni prostor, kjer se potencialne anomalije nahajajo.

Članek 4: Decentralizirana autoritativna arhitektura za večigralske igre na robu

V študiji smo preučevali:

1. ali lahko protokoli veriženja blokov zadostijo potrebam po nizkih latencah v večigralskih arhitekturah,
2. ali lahko takšna arhitektura prepreči znane goljufije kot ostali P2P protokoli,
3. ali lahko decentralizirana arhutektura prepreči tudi goljufije, ki na tajnem dogovarjanju.

Predstavljene simulacije so pokazale, da je predlagani protokol lahko primeren za večigralske igre tudi do 100.000 igralcev. Prav tako prikažemo kako lahko vse obstoječe metode za ugotavljanje goljufij uporabimo v okviru predlaganega protokola in dodatno pokažemo, da goljufije, ki temeljijo na tajnem dogovarjanju igralcev v praksi niso mogoče, saj protokol zaščiti decentralizirano naključnost tako, da izvor naključnosti ni mogoče predvideti v prihodnost.

Članek 5: Ogrodje WSN za notranjo lokacijo, ki upošteva zasebnost

Študija predlaga novo paradigmo računanja na podatkih senzorjev. Običajno arhitekture temeljijo na pošiljanju podatkov po omrežju na centralni strežnik, kjer se izvajajo analize. V raziskavi predlagamo novo arhitekturo, ki s pomočjo večnivojske enkripcije omogoči prenos izračunov namesto podatkov. V duhu porazdeljenega računanja je tako mogoče senzorje uporabiti za kolaborativno računanje, ki ne predvideva prenosa podatkov po omrežju. V raziskavi pokažemo, da je ta protokol bolj varen tako za podatke, kot za zaščito omrežne identite senzorjev pred zunanjimi nasprotniki. Predlagano ogrodje dodatno zaščitimo z uporabo blockchain protokola Ethereum. S pomočjo pametnih pogodb prikažemo kako s pomočjo javne avtentikacije senzorsko omrežje ločimo od uporabnikov. Omrežje Ethereum uporabimo tudi kot javno hrambo kriptiranih izračunanih rezultatov. Predlagano ogrodje podkrepimo z empirično raziskavo na primeru inovativnega sistema za notranjo lokacijo uporabnikov.

Zaključek

Raziskave v tej disertaciji preučujejo uporabo tehnologije veriženja blokov v domenah izven finačnega sektorja. S tem prispevajo k boljšemu razumevanju posameznih gradnikov in lastnosti tehnologije. Pregled literature vsebuje ključne prispevke znanosti

na posameznem preučevanem področju in izpostavi obstoječe pomanjkljivosti. Da bi tehnologija veriženja blokov dosegla transformativen učinek na izbranih področjih, je potreben razvoj domensko specifičnih protokolov, ki so energetsko varčni. Članek 1 predstavlja primer unikatnega protokola, ki je uporabljen za implementacijo decentraliziranega orkestratorja aplikacije, ki je primeren za računanje na robu. Omrežja, ki temeljijo na veriženju blokov so transparentna in zato ponujajo unikatno priložnost za podatkovno analizo. Transakcijski grafi zgrajeni z preiskovanjem blokov so veliki kar otežuje preiskovanje. Članka 2 in 3 predstavita robusten test, ki temelji na Benfordovem zakonu, ki omogoča preliminarno testiranje celotnih grafov za iskanje anomalij. Protokol (članek 1) s prilagojeno strukturo blokov smo uporabili za implementacijo ogrodja namenjenega večigralskim igram. V raziskavi izboljšamo stanje tehnike z detekcijo goljufij, ki temeljijo na tajnih dogovorih med igralci. Poleg novih protokolov pokažemo uporabo priljubljenega omrežja Ethereum (Članek 5) v domeni senzorskih omrežjih. Poleg inovativnega protokola za zaščito zasebnosti podatkov, izračunov in uporabnikov predstavimo tudi ogrodje, ki omogoča ločitev uporabnikov od senzorskega omrežja. Predlagano ogrodje podkrepimo z empirično raziskavo na primeru inovativnega sistema za notranjo lokacijo uporabnikov. Kljub kontroverzni zgodovini tehnologije veriženja blokov in njene trenutno najbolj razširjene uporabe, je njena uporaba v drugih domenah lahko izjemno koristna. Iz raziskav je razvidno, da obstoječa javna omrežja niso primerna za vsesplošno uporabo. Razvoj novih protokolov lahko prispeva k iskanju kompromisov med skalabinostjo, varnostjo in decentralizacijo, ki jih zahteva posamezna aplikacija. Kljub teoretičnim omejitvam porazdeljenih sistemov je smiselno raziskovati nove rešitve s primernimi kompromisi.