MOZAIK: An End-to-End Secure Data Sharing Platform

Aysajan Abidin[*¶] Enzo Marquet[†¶] Jerico Moeyersons[‡¶]

Xhulio Limani[§¶] Erik Pohle[*¶] Michiel Van Kenhove[‡¶]

Johann M. Marquez-Barja[§] Nina Slamnik-Kriještorac[§] Bruno Volckaert[‡]

ABSTRACT

The proliferation of Internet of Things (IoT) devices has led to exponential data growth that can be harnessed for personalized services, cost savings, and environmental benefits. However, collecting and sharing this data comes with significant risks, including hacking attacks, breaches of sensitive data, and non-compliance with privacy regulations. This paper proposes a comprehensive, end-to-end secure system, MOZAIK, for privacy-preserving data collection, analysis, and sharing to address these challenges. We perform a requirements analysis from the perspectives of security, privacy, legal, and functionality, highlighting the various mechanisms employed to safeguard sensitive data throughout the entire data cycle. This includes the use of lightweight encryption, distributed computation, and anonymous communication mechanisms to reduce security and privacy risks and to protect against single points of failure. MOZAIK provides a trusted and secure platform for data sharing and processing that can enable the creation of a data market and data economy.

KEYWORDS

Secure Data Collection, Data Sharing, Private Data Analytics, MPC, Data Protection, Data Intermediaries.

ACM Reference Format:
Aysajan Abidin, Enzo Marquet, Jerico Moeyersons, Xhulio Limani, Erik Pohle, Michiel Van Kenhove, Johann M. Marquez-Barja, Nina Slamnik-Kriještorac, and Bruno Volckaert. 2023. MOZAIK: An End-to-End Secure Data Sharing Platform. In *Data Economy (DE '23), June 18, 2023, Seattle, WA.* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The pervasive use of internet-connected devices, from smart wearables to city-wide sensor deployments, has led to a wealth of information that can be used to personalize services in an efficient manner. However, the risks associated with data breaches, hacking, and misuse of personal data pose significant challenges to realizing the benefits of the IoT-enabled future. To address these challenges, this paper proposes MOZAIK, a secure and privacy-friendly distributed IoT-data collection and analytics system that reduces the risks associated with IoT-enabled devices by using lightweight encryption, distributed computation, and anonymous communication mechanisms to protect the privacy of data throughout the entire data lifecycle. MOZAIK provides a trusted and secure platform for distributed data sharing and processing. The proposed architecture can be implemented as part of a data marketplace set-up e.g., as part of the trading engine, thus enabling the creation of the data economy. To give an example for a use case of MOZAIK, we consider an e-health setting. The user's wearable device collects and streams heartbeat sensor data to the MOZAIK system where heartbeat anomaly detection algorithms periodically evaluate the latest data and may warn to the user, if irregularities are detected.

Next to presenting the MOZAIK architecture, we conduct a thorough requirements analysis from the perspectives of functionality, legal, security, and privacy. This analysis aims to identify the key features and capabilities that the system must have to meet the needs of its users and stakeholders, while ensuring that it is secure and protects the privacy of sensitive data. From a functionality perspective, we evaluate MOZAIK's ability to collect, store, and process data from IoT-enabled devices, as well as its ability to support analytics and decision-making. We also consider the system's scalability and performance. For security, we apply the threat modeling framework STRIDE [11] to identify and mitigate security risks in MOZAIK. For privacy, we apply the LINDDUN framework [12] to analyze the possible privacy threats to MOZAIK and discuss MOZAIK's ability to implement data protection by design, as well as its ability to provide users with control over their data and its subsequent use. Through this requirements analysis, MOZAIK is designed and developed to meet the highest standards of functionality, security, and privacy, and provide users and stakeholders with a trusted and secure platform for data sharing and processing.

The remainder of the paper is organized as follows: Section 2 presents the MOZAIK's system model and Section 3 provides a legal requirements analysis. Section 4 details the functionalities of MOZAIK focusing on secure data collection, secure data storage, and secure data processing and sharing. Section 5 presents the performance security and privacy threat analysis while we discuss related work in Section 6. Finally, Section 7 concludes the paper.

[*] Author affiliation: imec-COSIC, KU Leuven, Belgium
firstname.lastname@esat.kuleuven.be
[†] Author affiliation: CiTiP, KU Leuven, Belgium
enzo.marquet@kuleuven.be
[‡] Author affiliation: IDLab-imec, Ghent University, Belgium
firstname.lastname@ugent.be
[§] Author affiliation: University of Antwerp-imec, IDLab-Faculty of Applied Engineering, Belgium
firstname.lastname@uantwerpen.be
[¶] Authors have equal contribution. Author names alphabetical.

DE '23, June 18, 2023, Seattle, WA
© 2023 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

2 MOZAIK SYSTEM

In this section, we detail the system architecture for MOZAIK and interactions among the system entities.

2.1 System model

MOZAIK is an end-to-end secure and privacy-friendly distributed IoT-data collection and analytics system. As shown in Figure 1, the MOZAIK system involves multiple components, including IoT nodes (or end-devices), users (data and device owners), a gateway, a storage unit, computing entities, and a web server.

- IoT nodes are devices that generate data, such as sensors, smart home devices, and wearables. These are responsible for collecting and transmitting data to either the gateway or directly to the storage unit.
- Users, also referred to as data owners, are individuals or organizations that own the data generated by the IoT nodes.
- The gateway serves as a central point for data collection, where data from multiple IoT nodes, belonging to the same user, is aggregated and stored. The gateway also provides secure communication channels between IoT nodes and other components of the system, such as the storage unit and the computing entities.
- The storage unit (i.e., the Obelisk platform[1]) is responsible for storing the encrypted data collected from IoT nodes securely and efficiently. Obelisk allows distributed computation engines to securely fetch the data while keeping sensitive data private and protected against hacking attacks.
- Distributed computing entities are responsible for performing data analytics on the encrypted data stored in the storage unit upon such request from the users. These entities use efficient, decentralized, and secure Multi-Party Computation (MPC) protocols to analyze the data using statistical, combinatorial, or machine learning algorithms. Protection mechanisms are also employed to protect data against information leakage. The data is encrypted using state-of-the-art lightweight encryption.
- The web server acts as the interface for the users (data owners) to interact with the system. The web server is responsible for providing a user-friendly dashboard that enables data owners to manage their data and control its access. It serves as a gateway for the users to upload their data to the system, initiate data processing tasks, and monitor the progress of those tasks. The web server also handles authentication and authorization tasks, ensuring that only authorized users have access to the system.

3 LEGAL REQUIREMENTS

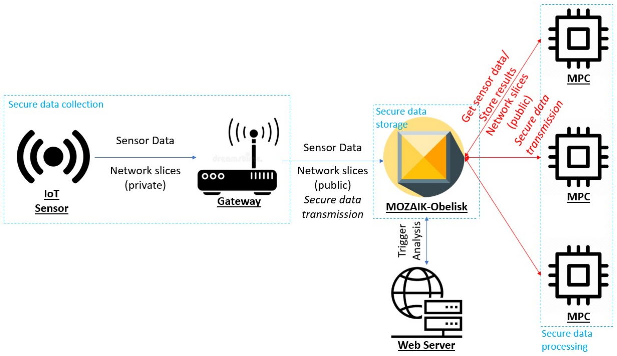

Figure 1: MOZAIK architecture overview

Building a data marketplace or a platform to enable data sharing, which includes personal data, comes with specific data protection requirements in the EU. Selling personal data as a commodity in line with current and future data protection legislation is being researched extensively. Moreover, under the General Data Protection Regulation (GDPR), data subjects retain certain rights on their personal data after collection or generation of the latter and over the whole course of the processing activities. Additionally, the GDPR strictly regulates the processing of personal data according to general principles such as proportionality and notably data minimization. Conforming with the GDPR remains challenging for data marketplaces. Privacy enhancing technologies (PETs) are to be welcomed as a valuable technical answer to the challenge of processing in line with the GDPR's principles. However, while PETs themselves have widely been investigated by computer scientists, their intrinsic value in light of data protection obligations has not yet been subject to significant interdisciplinary, technical and legal research. Additionally, the usage of PETs can greatly assist the incorporation of data intermediaries, as they rely on neutrality, transparency and control, from both user and company side.

The benefit of processing personal data in accordance with the GDPR raises users' trust and avoids high GDPR fines. In addition, ensuring adequate security has been enshrined in the EU legal framework concerning data[2]. Besides the standard processing requirements[3], it is critical for a data marketplace to assess the threats and risks data sharing is liable to and only then, it becomes possible to address them in appropriate fashion, e.g., through the use of tailored PETs. These threats, based on the LINDDUN framework [12], are linkability, identifiability, non-repudiation, detectability, disclosure of information, unawareness and non-compliance. When considering data marketplaces/sharing, the threats of linkability and identifiability (e.g., through inference) are a major factor because of the (possibly limitless) number of parties involved. A data marketplace should thus be able to adequately assess and tackle these threats. To comply with the GDPR[4], specific PETs can be used not only to tackle these issues, but also to be able to demonstrate compliance by following a privacy by design approach, another key element of the GDPR. MOZAIK utilizes a combination of PETs which lead to a better protection for data subjects. Below, in Section 4, the benefits of the MOZAIK architecture will be discussed and afterwards, in Section 5, the paper will outline how the architecture mitigates the most important threats.

4 FUNCTIONALITIES

The main functionalities of MOZAIK can be divided into three categories: data collection, data storage, and data processing and

[1] https://github.com/idlab-discover/obelisk

[2] e.g., Data Governance Act (DGA), the proposed Data Act.

[3] Lawfulness, minimisation, accuracy, storage limitation, integrity and confidentiality, and accountability

[4] Art. 5 and 24 GDPR

sharing. With MOZAIK we cover the entire value chain, from the data collection devices themselves (such as smart home or wearable devices) to the network providers that transport the data, cloud infrastructure and service providers, analytics, data marketplace providers, and end customers.

- Secure data collection: MOZAIK aims to provide a secure and privacy-friendly distributed IoT-data collection system that can collect data from various IoT devices. The data collection system uses state-of-the-art lightweight encryption and distributed computation to keep the sensitive data private while also protecting against single points of failure.
- Secure data storage: MOZAIK's secure data storage unit stores encrypted data and protects the data from unauthorized access and/or tampering. The storage unit should be robust against various attacks such as denial of service attacks and hacking attempts.
- Secure data processing and sharing: MOZAIK's data analytics feature aims to enable secure and efficient data processing, utilizing machine learning (ML) algorithms while preserving the privacy of the data. To achieve this, the data analysis should be executed using secure and decentralized protocols, such as MPC protocols, and privacy protecting mechanisms should be developed to minimize privacy risks. At the same time, MOZAIK enables creating a hybrid personal and non-personal data sharing marketplace that complies with prevailing data protection regulations, such as the GDPR and the DGA, while allowing data owners to retain control over their data and its subsequent use.

Below, we provide a detailed description of these functionalities. Afterwards, we describe how these functionalities confirm to the legal requirements as set out in Section 3.

4.1 Secure collection

Requirements. The IoT technology has become increasingly important to businesses and citizens due to the growing number of smart devices that can be remotely controlled, which enables real-time services aimed at improving safety and quality of life [2]. IoT technology is a broad field that involves a wide variety of different devices with diverse technology requirements, thus necessitating solutions to manage the connectivity of massive IoT devices in both private and public networks. For certain IoT services, such as e-health, automotive, transport, and logistics services, requirements on low latency, and ultra-reliability, are essential. In such cases, network operators must schedule data transfers with minimal latency to enable services such as remote healthcare, smart grids, and intelligent transport systems.

To support massive machine-type services, the network must be able to handle and differentiate traffic from millions of different devices efficiently. Furthermore, it is necessary to manage and control thousands of concurrent network devices to access the same network simultaneously. Some IoT services, such as e-health, require low-latency communication and high throughput because the data is directly related to a person's health condition considering that it is important to have efficient diagnosis and quick decisions. WiFi technology is widely used globally to connect smart devices in order to facilitate the everyday operations of industrial processes, healthcare providers, as well as individuals, thereby improving efficiency and increasing safety. However, there are several technological challenges associated with IoT communication, including interference between devices that communicate or connect to the network at the same time and the inefficient use of allocated bandwidth, which can lead to wasted resources in cases of insufficient activity by IoT devices. As the number of IoT devices per km^2 continues to increase and is expected to reach up to one million devices per km^2 in the coming years, it is essential to address these challenges to ensure efficient communication and connectivity.

MOZAIK solution. The data generated by an IoT device or sensor needs to be sent to MOZAIK Obelisk, the secure storage entity of MOZAIK. This needs to be done in a secure way as outlined in the requirements for secure data collection. As Figure 1 shows, the device sends its (encrypted) data to the gateway using, e.g., WiFi.

Network slicing concepts could be applied to improve the performance and reliability of sending data to the gateway, but is not always required (e.g., in a home environment). Once the data has arrived at the gateway, it will be sent in a secure way to MOZAIK Obelisk for storage, using network slices.

Network Slicing technology [17] has emerged as a promising solution to address the challenge of managing the connectivity of massive IoT devices in private or public networks. With IoT, slicing means that different services are assigned different resources based on their unique requirements for connectivity and security. Our research in this area aims to define service requirements and use them to create the required categories of network slices.

In Wi-Fi networks, network slicing enables the simultaneous connection of massive devices to different services with different requirements, leading to optimal network utilization. Achieving network slice isolation, where the network slices share the same resources without reducing performance and throughput, is critical.

Moreover, data collected from small devices must be secured and protected from cyber threats, making network slicing a crucial requirement for moving data from devices to the cloud. By creating virtualized networks, network slicing can tailor the network to specific needs, including performance and security requirements. This technology can provide enhanced performance and security measures, such as ultra-low latency, high reliability, encryption, and authentication, particularly for small devices that may be susceptible to cyber-attacks and other security breaches.

Network slicing is also vital for ensuring that data is only accessible to authorized users when moved to the cloud. By creating customized virtualized networks, sensitive data can be protected from unauthorized access and cyber threats. Overall, the concept of network slicing is a critical component in managing the connectivity of IoT massive devices and ensuring their security and privacy. To address this challenge, we introduce mechanisms that control the slice's isolation from data collection (i.e., sensors) to data storage and processing (i.e., the cloud). These mechanisms are based on protected orchestration processes and network management policies [14]. Additionally, we ensured that storage components within the slice, such as cache proxies, backpropagate security, and privacy policies are repurposed as part of the management policies. This will enable adaptability in current systems to changes in these policies and improve existing network solutions.

4.2 Secure data storage

Requirements. When storing data in the cloud, precautions need to be taken to make sure the data is secure and cannot be accessed by unauthorized entities. Especially when the application and its data is deployed on third-party multi-tenant cloud service providers (CSP), it must be guaranteed that the data is safe and cannot be accessed by the public, a malicious tenant or the CSP.

A first requirement to prevent unauthorized access to the data is to use a system with secure access control built-in. Access control provides adequate security to prevent theft or unauthorized modifications to the data at the application level.

A second requirement is to secure the underlying system: when the underlying system is attacked, and a malicious actor gains access to the operating system of the system where the application is running, they could still gain access to the data and modify it. When the underlying system is breached, any security measures taken by the application become irrelevant. Security of the underlying system can be achieved by using strict isolation between the components of the system, e.g., by using virtual machines (VM) to create a hardware-virtualized environment for each component. An attacker may gain access to, e.g., a publicly available access point, through a discovered vulnerability, but due to the isolated environment, cannot get access to the data stored in the other isolated components of the system.

Next to preventing unauthorized access, the system should also have fallback security measures to make the data practically useless when stolen. This can be done by applying data anonymization, encryption or obfuscation. The stored data on its own does not reveal anything about the data subject. This introduces new challenges, e.g., computing on the data becomes harder or less useful.

MOZAIK solution. To store data, MOZAIK utilizes Obelisk, a cloud-based storage platform that focuses on IoT-data with an attached web server for client-side user interfaces. Obelisk allows secured data ingestion, storage, streaming and retrieval through HTTPS REST (REpresentational State Transfer) APIs (Application Programming Interface). With regards to data transport, Obelisk ensures server authentication, data confidentiality and integrity towards its clients (IoT devices, applications, end users) through the use of the Transport Layer Security (TLS) protocol, a proven and widely adopted cryptographic protocol designed to secure communication over a computer network. Obelisk has been designed and implemented by means of an event-based, asynchronous microservices-based software architecture. [5]

To protect the data stored in Obelisk, a stateless (i.e., no data is stored) role-based access controlled (RBAC) API (MOZAIK Obelisk API) is introduced, which ensures only authorized users can access data. Data can only be accessed by using this API, which in turn is the only publicly available access point to Obelisk. This does not suffice to protect the data at the system level though. An unauthorized entity can still try to gain access to the data by attacking or accessing the underlying system where the application is running. A corrupt CSP could also access the data stored on their hardware.

To protect the data at a system level, a strict isolation between different components and the hardware is used. This can be accomplished by using VMs, as stated in the requirements, but this is not the ideal solution when working with a microservices-based architecture like Obelisk. A VM creates a hardware-virtualized environment that is isolated from other components of the system and the underlying hardware, but comes with the cost of a significant performance overhead (e.g., booting times and CPU- and RAM usage) [1]. In a modern microservices-based architecture, containers are used in combination with a container orchestration platform, like Kubernetes[5], to manage the large amount of running containers. Containers bundle all the necessary dependencies and software of an application in a standardized and portable format and offer relatively fast startup times with a smaller computational overhead compared to VMs, which allows large applications to scale horizontally in a seamless manner. A container runtime manages the container life cycle and provides software-based container isolation. Software-based container isolation exposes the risk of container runtime escaping vulnerabilities and system privilege escalation [15]. To mitigate these risks, a secure container runtime is used. A secure container runtime encapsulates each container in a lightweight virtual machine, called a microVM, which ensures a strict isolation between different components of the system and the underlying hardware. When an attacker gains access to one of the components (i.e., the MOZAIK Obelisk API, as it is the only publicly available access point to the data storage), it is stuck inside its microVM environment and cannot access the stored sensitive data in other components.

4.3 Processing and sharing

Requirements. The overarching requirement for data processing is confidentiality of the user's data while it is being processed. Since the processing takes place outside of the control of the user, e.g., on powerful cloud servers that are hosted by a second or third party, we require that the entity that processes the data cannot learn meaningful information about this data to prevent unauthorized data collection and misuse.

In addition, another goal is to ensure integrity of the user's data. This amounts to two requirements. First, the data that is the input to the processing as well as the data that is output from the data processing needs to be integrity protected. Otherwise, a malicious entity may change (parts of) the user data and thus alter the processing functionality. A wrong and misleading result would be returned to the user who cannot detect the forgery. For example, a malicious entity might alter, elevate or replace medical sensor data that a user sends to process in order to force a wrong diagnosis. Second, the data processing function must remain equivalent to the functionality the user agreed on. This means that the processing entity should not be able to deviate from the computation. Despite the data being confidential to the processing entity, the entity may learn information by computing a (slightly) different function and observing the user's behaviour, e.g., via a selective-failure attack.

The user data is generated by IoT devices, i.e., lightweight, embedded devices with weak hardware and requirements for low power consumption, memory and computation. Therefore, the selected approach and solution to provide end-to-end confidentiality and integrity of the IoT sensor data must be lightweight for the parts that are executed on the IoT device itself. Further, typical security

[5] https://kubernetes.io/

threats due to the nature of the IoT device, such as nonce misuse [4], must be explicitly addressed.

The data processing is a critical part of the system. In order to further reduce risks for the user when processing, we require a distributed trust model. Instead of entrusting one entity with data processing (following an approach that satisfies the aforementioned requirements on data confidentiality, integrity and lightweightness) we require a distributed approach that involves multiple independent entities that jointly process the data where, importantly, the corruption of some entities will not lead to a security issue. The benefits of distributing are increased resilience and security. In order to attack the confidentiality, multiple independent entities have to be corrupted, increasing the attack cost.

MOZAIK solution. Regarding privacy-preserving data processing, two PETs come in mind that allow confidentiality of the data while being processed. Fully homomorphic encryption (FHE) and secure multi-party computation (MPC). With FHE, the data processing is carried out by a single powerful cloud server that can largely remain untrusted. Confidentiality and privacy properties result from homomorphisms on the ciphertext of particular encryption schemes, so the cloud server cannot know the decryption key of the data. Consequently, to provide end-to-end confidentiality, the IoT device itself must create the FHE ciphertext. Barring the computational overhead, current state-of-the-art FHE schemes have large ciphertexts that make their use in IoT and embedded devices hardly feasible at the moment.

We settle for another approach. We employ lightweight symmetric encryption, i.e., dedicated low-power, low-area AEAD schemes from the NIST lightweight cryptography competition[6], and combine the processing phase with a distributed decryption phase using MPC. This relieves much of the performance pressure from the IoT devices and adds great flexibility and interoperability with existing soft- and hardware implementations. The workload is shifted towards computationally powerful and high-bandwidth cloud servers. In MPC, privacy of the input values is achieved by secret-sharing inputs and intermediate secret values such that reconstruction is infeasible even if multiple servers collude up to a threshold.

The lightweight AEAD scheme must offer a security level of 128 bit for confidentiality [3, 6] at at least 64 bit for authenticity [6]. Since longer nonces imply more infrequent key rotation, we require at least 48 bit nonces. Nonce misuse is a real problem in IoT devices, so we require well-defined nonce misuse resistance for the AEAD scheme as well.

4.4 Legal

The proposed architecture addresses several key legal issues a secure and privacy-friendly IoT-data collection and analytics systems encounter such as risk management, appropriate security against threats and flexibility. The introduction of carefully selected PETs such as MPC [8], Obelisk, RBAC, network slicing ensures that data subjects are sufficiently protected against linkability and identifiability through inference.

Specifically, MPC significantly reduces identifiability upon running computations on shared personal data without reducing utility of said data. Plus, as processing parties do not know the identity of the person, the linkability of personal data becomes neigh impossible. Additionally, MPC is a PET implemented to achieve data minimization as well as integrity and confidentiality.

Lastly, Obelisk offers the data subjects control over their data, thus preventing purpose creep among the processing while also promoting transparency. As with any data storage, the Obelisk setup implements state-of-the-art measures to provide the required integrity and confidentiality. Also, the flexibility of the data storage under Obelisk allows for additional PETs to be implemented by adapting the API. Implementing additional PETs with the purpose of, e.g., obfuscation or de-identification, allows for a greater level of data protection depending on the data's sensitivity.

The introduction of these PETs in the MOZAIK architecture provides adequate state-of-the-art technical and organisational measures to safeguard the (personal) data of data subjects and users, which means controllers or data holders can use this architecture as the basis for a marketplace to run computations on the shared data in line with the GDPR and DGA. This will greatly aid data intermediaries in proving their compliance with the GDPR.

5 THREAT ANALYSIS

This section first describes the threat model, and then analyzes the security and privacy of MOZAIK using both STRIDE [11] and LINDDUN [12] frameworks, respectively.

5.1 Adversarial model

We consider the following adversarial model.

- IoT nodes: These are considered honest and not compromised, are securely provisioned with secret key material, and are capable of encrypting collected data. While we leave it outside the scope of this paper, we do, however, note that they could be compromised by attackers who could gain access to the device's firmware or credentials, and compromise the data collected by the device.
- Users: Users could be dishonest and may try to gain unauthorized access to other users' data.
- The gateway: It could be compromised by attackers who could intercept data being transmitted from the IoT nodes to the storage unit or web server. We do not consider the cases where the gateway is target of denial of service (DoS) attacks to disrupt data transmission.
- The storage unit: We assume that the storage unit is semi-honest: while offering its services faithfully it may attempt to infer private information about the user, for example, by building a profile for users based on their access patterns.
- Computing entities: The adversarial model for these follow that of the employed MPC protocols for private computing tasks, e.g., semi-honest adversaries with honest majority.
- The web server: The web server is assumed to be hosted on a secure server and offers a secure interface between the user and the other entities in the system, as well as a secure user authentication and authorization mechanism.

Note that our threat model does not include network adversaries, such as, malicious network providers. Although there are anonymous communication solutions against such adversaries, it is beyond the scope of this paper.

[6] https://csrc.nist.gov/Projects/Lightweight-Cryptography/

5.2 Security analysis

Here we perform security analysis using the STRIDE [11] framework to identify the possible security threats to MOZAIK.

- Spoofing: An attacker may attempt to impersonate a legitimate entity in order to gain unauthorized access to sensitive data. To mitigate this threat, MOZAIK employs secure mutual authentication between the communicating entities.
- Tampering (with data): An attacker may attempt to tamper with the information stored in Obelisk and/or exchanged between the entities. MOZAIK mitigates this threat by using authenticated encryption of all data exchanged and stored.
- Information disclosure: An attacker may attempt to gain unauthorized access to sensitive data by, say, eavesdropping on messages exchanged between the MOZAIK entities. Authenticated encryption of data also mitigates this threat.
- Repudiation: Disputes may arise when an entity denies to have performed certain tasks or accessed services. MOZAIK employs digital signatures to ensure non-repudiation, as well as other appropriate dispute resolution mechanisms.
- Denial of Service (DoS): With this threat an attacker could disrupt the system's availability. This impacts IoT nodes, the gateway, and the Obelisk platform. In MOZAIK, only Obelisk has countermeasures against DoS for the moment.
- Elevation of Privilege: An attacker may attempt to gain elevated privileges to the system. To mitigate this threat, MOZAIK uses secure container environments in combination with secure access control and authorization mechanisms to ensure that only authorized users can access the system and its sensitive data.

5.3 Privacy analysis

Now we perform a privacy threat analysis using LINDDUN [12]. Note, the list excludes *disclosure of information* since it belongs to the security category and is already covered.

- *Linkability* refers to the ability to distinguish whether two items of interest (IOI) are linked, without necessarily knowing the actual identity of data subject. MOZAIK ensures unlinkability by keeping the data encrypted throughout its lifecycle – i.e., during collection, storage, and processing (performed over encrypted data) – and by employing privacy-preserving authentication (PPA) based on anonymous credentials.
- *Identifiability* refers to the ability to identify the data subjects from the data collected or from other activities. Identifying data subjects from the data collected is a none issue, since the data is encrypted, but Obelisk could build access patterns for each (anonymous) user identity in an attempt to identify the real user identity. This risk is mitigated by using anonymous access control as part of the PPA.
- *Non-repudiation* refers to the data subject being unable to deny a claim (e.g., having performed a certain action). While security demands non-repudiation, privacy requires plausible deniability. Hence, a right balance is needed.
- *Detectability* refers to the ability to distinguish whether an IOI about a data subject exists. By design, MOZAIK guarantees user undetectability since all data associated with a user is end-to-end encrypted and anonymous identities and credentials are used for accessing the services.
- *Unawareness* refers to data subjects being unaware of their personal data's collection, storage, or processing activities. In MOZAIK, the users fully control when to upload what data from their (IoT) devices. Also, since data is encrypted by default, the risk to the users' privacy is minimized.
- *Non-compliance* refers to not being compliant with data protection principles or regulations – in particular, the processing, storage, or handling of personal data is not compliant with legislation, regulation, and/or policy. The MOZAIK architecture provides intermediaries with a means to implement technical and organizational measures that helps mitigate the risks and, at the same time, provide accountability by proving that these measures have been implemented.

6 RELATED WORK

The MyHealthMyData project [13] focuses on distributed storage through ledgers and access control for medical data sharing. However, unlike MOZAIK, they do not consider secure data collection nor privacy-preserving use (computation) since the authorized party obtains the shared data in the clear. In the same vein, Veeningen et al. [16] implements MPC in three pilot use-cases to perform privacy-preserving computation in the medical domain on collected data. But Veeningen et al.'s analysis does also not include secure data collection. They do provide a way for confidential processing for the analysis, but with the disadvantage of having to trust an intermediary such as a hospital, a worker's union or a GP. The secure data collection of MOZAIK bypasses this need for trust entirely.

There exist two other relevant projects: Agora: A Privacy-Aware Data Marketplace [10] and H2020 KRAKEN (BroKeRage and MArKet platform for pErsoNal data)[7] [7, 9]. They both focus on data sharing, Agora with smartcontracts and KRAKEN with privacy-preserving analytics. The goal of MOZAIK is thus in line with these projects, but MOZAIK takes an umbrella approach towards secure and scalable IoT data collection, transmission, storage and processing. By doing so, MOZAIK covers the full data life cycle.

The unique value proposition of MOZAIK is the created synergy of the operations at each data cycle point.

7 CONCLUSION

In conclusion, the increased use of internet-connected devices has generated a wealth of data that can fuel various (personalized and more efficient) services. However, the associated risks to the security and privacy of data and even individual users are significant. To address these challenges, we propose an end-to-end secure and privacy-friendly data-sharing platform MOZAIK. The proposed MOZAIK architecture utilizes lightweight encryption, distributed computation, network slicing, and various cryptographic mechanisms to ensure the privacy of data throughout the data lifecycle. Additionally, the paper conducts a requirements analysis from various perspectives, including functionality, legal, security, and privacy, to ensure that MOZAIK meets the needs of its users and stakeholders while maintaining the highest standards of security and privacy. Overall, MOZAIK enables the creation of a trusted data market and data economy.

[7] https://krakenh2020.eu/

ACKNOWLEDGEMENTS

This work is supported by the Flemish Government through FWO SBO project MOZAIK S003321N.

REFERENCES

[1] Kavita Agarwal, Bhushan Jain, and Donald E. Porter. 2015. Containing the Hype. In *Proceedings of the 6th Asia-Pacific Workshop on Systems* (Tokyo, Japan) *(APSys '15)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/2797022.2797029

[2] Stephanie B. Baker, Wei Xiang, and Ian Atkinson. 2017. Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities. *IEEE Access* 5 (2017), 26521–26544. https://doi.org/10.1109/ACCESS.2017.2775180

[3] Elaine Barker. 2020. *Recommendation forKey Management: Part 1 – General.* NIST Special Publication 800-57 Part 1 Revision 5. https://doi.org/10.6028/NIST.SP.800-57pt1r5

[4] Hanno Böck, Aaron Zauner, Sean Devlin, Juraj Somorovsky, and Philipp Jovanovic. 2016. Nonce-Disrespecting Adversaries: Practical Forgery Attacks on GCM in TLS. In *10th USENIX Workshop on Offensive Technologies (WOOT 16)*. USENIX Association, Austin, TX. https://www.usenix.org/conference/woot16/workshop-program/presentation/bock

[5] Vincent Bracke, Merlijn Sebrechts, Bart Moons, Jeroen Hoebeke, Filip De Turck, and Bruno Volckaert. 2021. Design and evaluation of a scalable Internet of Things backend for smart ports. *Software: Practice and Experience* 51, 7 (2021), 1557–1579.

[6] Federal Office for Information Security. 2023. *BSI TR-02102-1: Cryptographic Mechanisms: Recommendations and Key Lengths, Version 2023-01.* Federal Office for Information Security.

[7] Silvia Gabrielli, Stephan Krenn, Donato Pellegrino, Juan Carlos Pérez Baún, Pilar Pérez Berganza, Sebastian Ramacher, and Wim Vandevelde. 2022. *KRAKEN: A Secure, Trusted, Regulatory-Compliant, and Privacy-Preserving Data Sharing Platform.* Springer International Publishing, Cham. 107–130 pages. https://doi.org/10.1007/978-3-030-98636-0_6

[8] Lukas Helminger and Christian Rechberger. 2022. Multi-Party Computation in the GDPR. In *Privacy Symposium 2022*, Stefan Schiffner, Sebastien Ziegler, and Adrian Quesada Rodriguez (Eds.). Springer International Publishing, Cham, 21–39. https://doi.org/10.1007/978-3-031-09901-4_2

[9] Karl Koch, Stephan Krenn, Donato Pellegrino, and Sebastian Ramacher. 2021. Privacy-Preserving Analytics for Data Markets Using MPC. In *Privacy and Identity Management*, Michael Friedewald, Stefan Schiffner, and Stephan Krenn (Eds.). Springer International Publishing, Cham, 226–246. https://doi.org/10.1007/978-3-030-72465-8_13

[10] Vlasis Koutsos, Dimitrios Papadopoulos, Dimitris Chatzopoulos, Sasu Tarkoma, and Pan Hui. 2020. Agora: A Privacy-aware Data Marketplace. In *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, Singapore, Singapore, 1211–1212. https://doi.org/10.1109/ICDCS47774.2020.00156

[11] Microsoft. 2009. *The STRIDE Threat Model.* Retrieved 2023-03-05 from https://msdn.microsoft.com/en-us/library/ee823878(v=cs.20).aspx

[12] Deng Mina, Wuyts Kim, Scandariator Riccardo, Preneel Bart, and Joosen Wouter. 2011. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineeering Journal* 16, 1 (2011), 3–32.

[13] Edwin Morley-Fletcher. 2017. MHMD: My Health, My Data. In *EDBT/ICDT Workshops*. 1 pages.

[14] Fatima Salahdine, Qiang Liu, and Tao Han. 2022. Towards Secure and Intelligent Network Slicing for 5G Networks. *IEEE Open Journal of the Computer Society* 3 (2022), 23–38. https://doi.org/10.1109/OJCS.2022.3161933

[15] Murugiah Souppaya, John Morello, and Karen Scarfone. 2017. NIST Special Publication 800-190 Application Container Security Guide. (2017). https://doi.org/10.6028/NIST.SP.800-190

[16] Meilof Veeningen, Supriyo Chatterjea, Anna Zsófia Horváth, Gerald Spindler, Eric Boersma, Peter van der Spek, Onno van der Galiën, Job Gutteling, Wessel Kraaij, and Thijs Veugen. 2018. Enabling Analytics on Sensitive Medical Data with Secure Multi-Party Computation. *Studies in health technology and informatics* 247 (Jan. 2018), 76–80.

[17] Shalitha Wijethilaka and Madhusanka Liyanage. 2021. Survey on Network Slicing for Internet of Things Realization in 5G Networks. *IEEE Communications Surveys & Tutorials* 23, 2 (2021), 957–994. https://doi.org/10.1109/COMST.2021.3067807