Towards a Threat Model and Security Analysis for Data Cooperatives

Abiola Salau, Ram Dantu, Kirill Morozov, Kritagya Upadhyay and Syed Badruddoja
Department of Computer Science and Engineering, University of North Texas, Denton, TX, 76207, U.S.A.

Keywords: Data Cooperatives, Threat Model, Secure Data Management, Cybersecurity, Security and Privacy.

Abstract: Data cooperative (called "data coop" for short) is an emerging approach in the area of secure data management. It promises its users a better protection and control of their data, as compared to the traditional way of their handling by the data collectors (such as governments, big data companies, and others). However, for the success of data coops, existing challenges with respect to data management systems need to be adequately addressed. Especially, they concern terms of security and privacy, as well as the power imbalance between providers/owners and collectors of data. Designing a security and privacy model for a data coop requires a systematic threat modeling approach that identifies the security landscape, attack vectors, threats, and vulnerabilities, as well as the respective mitigation strategies. In this paper, we analyze the security of data cooperatives, identify potential security risks and threats, and suggest adequate countermeasures. We also discuss existing challenges that hinder the widespread adoption of data coops.

1 INTRODUCTION

A data cooperative as defined in (Pentland et al., 2019) "refers to the voluntary collaborative pooling by individuals of their personal data for the benefit of the membership of the group or community". It is currently gaining a lot of attention from researchers as an approach that can address the power imbalance and declining trust among data providers and organizations that make profit from the data. Data coops serve as a fiduciary for the data providers to mediate between them and the data consuming companies/organizations in order to help them to negotiate the control and use of the subjects' personal data.

Although this shared data is regarded as a bedrock for the new knowledge economy, as it is central to government and organizational strategic decision making, advertisements, sales, research breakthroughs such as vaccine development (Radanliev et al., 2020), innovations (Chika and Adekunle, 2017), effective contact tracing during a pandemic etc. We may observe that it is being concentrated in the hands of only a few companies/organizations. This way, the data is locked up in disparate, unlinked silos being inaccessible to smaller companies or individual researchers who drive innovations (Blasimme et al., 2018). We illustrate this in Fig. 1. The result is a declining of trust between the involved parties. Therefore, this situation motivated and stimulated a search for better approaches to managing personal data, especially those giving a better control to the data owners.

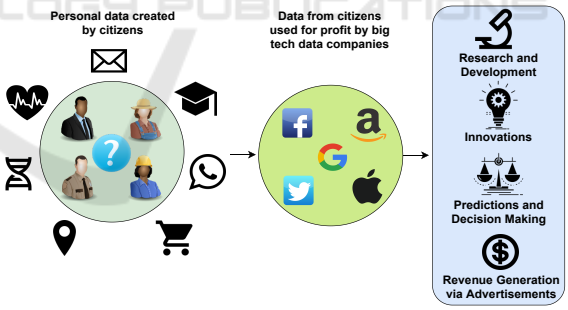

Figure 1: The big tech data companies collect data from the users of their services. These companies use the data for various purposes, e.g., targeted advertisement to the users and analytics for gaining business insights, while failing to adequately compensate the data providers.

Despite the numerous benefits as well as a transformative potential, that data cooperatives would offer, some challenges may hinder their implementation and/or success. *One of the main such challenges is the security and privacy of personal data—that is our focal point in this paper*. In a data coop, the data is an important asset, which, if compromised, can result in a great financial loss that may potentially threaten the

Salau, A., Dantu, R., Morozov, K., Upadhyay, K. and Badruddoja, S.
Towards a Threat Model and Security Analysis for Data Cooperatives.
DOI: 10.5220/0011328700003283
In *Proceedings of the 19th International Conference on Security and Cryptography (SECRYPT 2022)*, pages 707-713
ISBN: 978-989-758-590-6; ISSN: 2184-7711

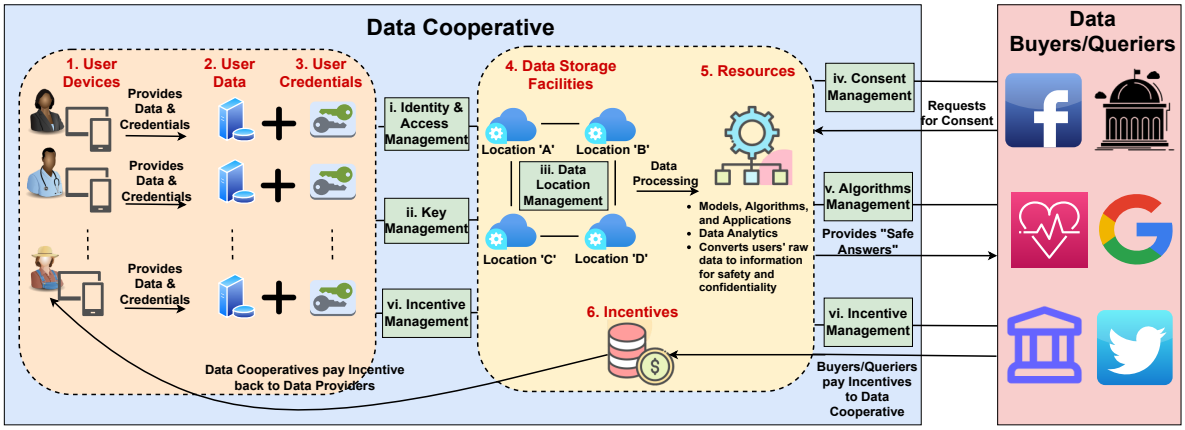

Figure 2: The Data Cooperative Ecosystem.

very existence of coops as a data management mechanism. For instance, in 2013, Target—one of the major retail stores in the US—was involved in a data breach that resulted in a loss of about $39 million for resolving claims by the affected financial organizations. To avoid such a situation and other dire consequences of security failures, it is important to identify and properly address the related security challenges.

To be specific, we will list the following major challenges, which, in our opinion, are hindering the widespread adoption of data cooperatives:

1. Security and privacy guarantees for the pooled data. This includes a deployment of advanced privacy-preserving technologies for data processing, such as privacy-preserving machine learning, federated learning, homomorphic encryption, differential privacy, and others.
2. A proper balance between strict control of personal data for the providers and economic value for the data coop.
3. Inclusivity of governance in a data coop. In particular, ensuring that all members can participate in the decision-making process.
4. Establishing and maintaining trust among participating members and the data coop fiduciaries.
5. Compliance with laws and regulations in the region(s) covered by the data coop with respect to the use, storing and deletion of personal data, portability, privacy, and data interoperability, e.g., GDPR[1] (*The General Data Protection Regulation*), HIPAA[2] (*The Health Insurance Portability and Accountability Act*), and others.
6. Auditing of the activities of the data coop. There should be a process that evaluates the fiduciary duties of the data coop.
7. Lack of awareness on the subject, in particular, that of potential benefits of data coops.

Our Contributions. To the best of our knowledge, this paper is the first to discuss data coops with an emphasis on security requirements. We, therefore, summarize the contributions of our paper as follows:

- We propose a systematic threat analysis with typical attack scenarios for data cooperatives.
- We present a detailed classification of threats and potential vulnerabilities in a data coop.
- Finally, we suggest defense mechanisms against the identified threats.

2 BACKGROUND AND RELATED WORKS

In this section, we briefly describe data cooperatives, and discuss previous results on threat modeling.

2.1 Data Cooperatives

The concept of data cooperatives is based on a presumption that there is a stronger impact on the collective data than the individually owned data (Gong, 2022). According to the data coop approach, individuals voluntarily pool their personal data in an institution and the institution is responsible for effective use and protection of the aggregated data. The collective data can be used for various purposes, for instance, it can be made available to companies with the permission of the data subjects for monetary reward. In another use case, the members themselves may use the pooled data to extract better analytical insights as

[1]https://eur-lex.europa.eu/eli/reg/2016/679/oj
[2]https://www.hipaa.com/

compared to when their individual data is used. According to (Pentland and Hardjono, 2020), a data coop ecosystem as shown in Fig. 2 comprises the data cooperative as a legal entity, members of the coop, and the external entities (or "queriers") who interact with the coop. Any interested individual who contributes their personal data (e.g., health data) to the coop may become its member.

Data coops can exist in a variety of forms depending on the kind of data that is pooled, and who are the potential members and queriers. Let us discuss some specific examples. Authors in (Blasimme et al., 2018), proposed a health data coop that makes health data more available for research as compared to data processing in individual organizational silos. An existing coop, which follows this approach, is MIDATA[3]. Data coops also have the potential to empower data providers by granting better control in the use of their data (Pentland et al., 2019). One example is HAT[4], a platform that grants individuals the right to the ownership of their personal data. The work (Hardjono and Pentland, 2019) discussed a data coop which is focused on decentralized data management. The Enigma platform (Zyskind et al., 2015) adds privacy to the data coop solution, thus enabling multiparty computation on data while protecting them from unauthorized access.

2.2 Previous Results on Threat Modeling

Threat modeling is a systematic approach to focusing on the vulnerabilities of a system through a process of analyzing and validating the respective threats and their mitigation methods. There exist a lot of approaches to threat modeling, which can be applied depending on the architecture/system to be protected, such as STRIDE (Microsoft, 2022), PASTA (UcedaVelez, 2012), and LINDDUN (Wuyts and Joosen, 2015), to mention just a few.

Although there are variations in the threat modeling approaches listed above, we observe that they generally share the following steps: *(i) identifying the items of value in the system (the assets) and the security landscape; (ii) identifying the attack entry points; (iii) defining potential attack taxonomy; (iv) identifying the threats and vulnerabilities; and (v) proposing countermeasures to the respective security gaps*. We emphasize that none of the above-mentioned approaches have been applied specifically to our use case of a data cooperative. Therefore, the main goal of our work is to close this gap by constructing the respective threat model, and then proposing mitigation strategies for the identified vulnerabilities. In addition, we discuss the existing challenges for this topic.

3 THREAT MODELING FOR DATA COOPERATIVES

3.1 Assets in a Data Cooperative

An asset of a data cooperative can be defined as any component of value within the system that is worth securing. Depending on the business goal of a data cooperative, the set of assets may differ. In Table 1, we list the assets which are common to any typical data coop.

Table 1: Assets in a Data Cooperative.

Asset	Description
Data	Members' personal data, pooled together; the most important asset of a data coop.
Resources	Assets such as algorithms, models, websites, applications, and software used to process the data. Others include identity, consent management, and access control tools, reputation management systems, etc.
Hardware and Storage Facilities	Data center facilities for storing and processing the data: hardware components and their peripherals, including computing devices and cloud servers.
User Credentials	Login details with which users are authenticated and granted access to resources in the data coop.
User Devices	Client devices to be used for communication with the data coop, e.g., personal computers, smartphones, etc.
Incentives	Form of motivation to the data subjects for their participation in the data pooling, such as monetary rewards, improved reputation, and information gain.

3.2 Points of Entry

An attack entry point is an avenue through which an attacker can exploit the system. For a data coop as a complex system, the entry points may be difficult to predict, as in principle, an attacker can break in through any component to which they can gain access. We list some of such entry points in Table 2.

[3]https://www.midata.coop/en/home/

[4]https://www.hubofallthings.com/

Table 2: Attack Entry Points.

Entry Point	Description
Members	Both current and former members of a data cooperative can be an attack entry point. Business process flaws, negligence of security audits, lack of risk analysis, poor security awareness of the members, missed security patches, improper waste disposal, and insufficient security/IT capacity are a few of the issues that make the participating individuals an attack entry point.
Technology	The various technologies used by members of the coop can be a source of security vulnerability, e.g., data sharing or pooling system, storing data on mobile devices, legacy systems without adequate security maintenance, internet browsers used to access the system, etc.
Network	There exist a variety of attack entry points provided by the network used by the data coop. For instance, unprotected network communication channels, open physical connections, IP addresses and ports, insecure network architecture, unused user IDs, etc.
Hardware	Hardware such as data storage devices, workstations, and servers are susceptible to flaws that may arise due to environmental conditions (e.g., dust, heat, humidity, etc.), design fault/defects, and flawed configurations, as well as the hardware being out of date (Salau et al., 2018).
Software	The software used by coop and/or its members can be a point of entry for an attack. Some of the related vulnerabilities are connected to insufficient testing, bugs, and design faults, unchecked user input, inadequate audit trail, use of unnecessary software such as bloatware, software as a service (SaaS) that relinquish control of data.

3.3 Attack Taxonomy

In this subsection, we briefly discuss various types of attacks against a data cooperative. Specifically, we classify the attacks into three main categories: access level, attack strategy, and motive.

3.3.1 Attacker's Access Level

An attacker can either be ***internal*** or ***external*** to the system based on their access level. Internal attackers are primarily insiders who have legitimate access to the use of the system. In this context, the individuals that make up the membership, the leaders, and the external entities who are given legal access to the data coop are regarded as internal attackers. These external entities (or data collectors) can be research institutions, data companies, etc. The external attackers are generally all other kinds of attackers who do not have legitimate access to the system.

3.3.2 Attacker's Strategy

The strategy of an adversary attacking data cooperative can be sub-categorized as follows. ***Technical vs. non-technical***: In a technical attack, the adversary uses sophisticated software, system knowledge, and expertise, while in a non-technical attack, the adversary adopts deception to have insiders reveal sensitive information or perform actions that could compromise security of the data coop, e.g., phishing and other types of social engineering. ***Active vs. passive***: An active attack disrupts the normal operations and functionality of the coop such as a denial of service (DoS) attack, while a passive attack monitors and possibly analyzes network traffic for vulnerabilities with the aim of gathering information about the coop system. ***Physical vs. logical***: A physical attack is performed directly against the data coop systems, hence resulting in physical damage to a device(s) or changes to the system configuration/properties. A logical attack does not directly cause physical damage to devices yet results in them not functioning properly.

3.3.3 Attacker's Motive

Due to the lack of space, we will only list the followings major types of attackers' motivation: profit, espionage, rivalry, and FIG (Fun, Ideology, Grudge). They will be discussed in detail in the full version of this paper.

4 SECURITY ANALYSIS FOR DATA COOPERATIVES

A *threat* is a potential danger concerning harm or loss to an asset of a system, while *vulnerability* is a potential attack point of entry to the system (Dahbur et al., 2011). As identified in Sec. 2, there are many threat modeling strategies available in the literature. In this paper, we use the STRIDE (Microsoft, 2022) model due to its categorization of threats in a manner that helps security professionals to provide answers about their system on the core aspects of information security—confidentiality, integrity, availability, and others (Stallings et al., 2012) (see Table 3).

4.1 Threats and Vulnerabilities

Next, we will discuss some potential threats and vulnerabilities in a data cooperative, which are categorized according to the STRIDE model.

Table 3: STRIDE Threat Categories and Security Violations.

Type of Threat	What Was Violated
Spoofing	Authentication
Tampering	Integrity
Repudiation	Non-repudiation
Information Disclosure	Confidentiality
Denial of Service (DoS)	Availability
Elevation of privilege	Authorization

Spoofing. *Identity theft.* Here, an attacker can steal the identity of a legitimate user to gain unauthorized access to the data cooperative system (various malicious tactics may be used). *Content Spoofing.* The adversary maliciously modifies data, which are sent between two members of the cooperatives. *Device Spoofing.* An adversary via eavesdropping intercepts a device or IP information of legitimate users in order to carry out a replay attack. *Session Spoofing.* Here, the adversary steals login credentials from legitimate users for future use. Typically, the adversary deployed the Man-in-the-Middle (MITM) or passive eavesdropping attack.

Tampering. *Data tampering.* In a data coop, the pooled data is a primary target for an attack. Likewise, the algorithms and software code—that it uses to manipulate the data—may be altered by an adversary. *Timestamp tampering.* The time at which an event occurred may be a crucial component of a data coop, such as, for instance, in a neighborhood-watch data coop (Salau et al., 2021). An attacker can manipulate timestamps, e.g., in order to make it appear that an event occurred at a different time. *Log files tampering.* Log files help to keep track of events happening behind the scene (Forte, 2009). An adversary can manipulate log file to cover up some other malicious activities it carried out. *Storage tampering.* Here, an attacker targets the data store intending to modify the data stored in it.

Repudiation. *Content repudiation.* In this form of attack, a malicious member in a data cooperative can deny sending, receiving, or manipulating some data. This can be backed up with log file tampering. *Activity hiding.* This is a situation where an adversary, after gaining access to the target system and carrying out an attack, also covers their track by carrying out passive attacks, e.g., in a hope to overfill the log files and hence to hide their malicious activities.

Information Disclosure *User Information and Data disclosure/breaches.* This covers an unauthorized access to sensitive personal data, such as users' health information, credentials, credit card details and so on which requires protection by laws such as GDPR. *Application Error Display.* When applications encounter an error and display an error message to the user, such the message may reveal sensitive information to an attacker. *Device Information Disclosure.* An intruder may intercept a communication between legitimate users and may discover device information such as its type, IP address, and/or location, which may used to coordinate future attacks.

Denial of Service *DoS.* Here, an adversary makes the system inaccessible or unusable to legitimate users. An example is making the pooled data unavailable either through packet flooding or by exploiting vulnerabilities that can lead the system to crash. *DDoS.* This can be seen as multiple simultaneous DoS attacks on a system. For example, multiple sources can launch such the attack on the data processing algorithms, for instance, overloading the system with "computationally expensive" requests.

Elevation of Privilege. *Unauthorized Privilege to Restricted Data.* An internal adversary, for instance, may be able to access data above its access level through malicious tactics such as phishing, brute force, or identity theft. *Abuse of privileges.* A legitimate user with admin privileges may attack the system by granting elevated privileges to unauthorized users.

4.2 Mitigation Strategies

In this section, we discuss various mitigation strategies and countermeasures to the threats and vulnerabilities identified in the previous section. Again, we use the STRIDE modeling approach in order to match the mitigation strategies to the vulnerabilities mentioned in Sec. 4.1.

Spoofing. Let us first discuss the countermeasures against spoofing. *Network Monitoring.* The network and communication channels must be well monitored for atypical activities using specialized network monitoring security tools. *Authentication.* Authentication systems must be robust. Connections between devices should be authenticated using secure systems such as IPSec, domain authentication, and others. Moreover, the members of the coop must adhere strictly to strong password policies. *Packet Filtering.* The coop should deploy packet filtering with deep packet inspection (DPI) techniques in order to detect anomalies such as outgoing traffic with IP address not consistent with that of the coop's network. *Encryption.* Data should be encrypted both at rest and in transit. Secure network protocols, such as Transport Layer Security (TLS), IPSec, and SSH, help to prevent spoofing attacks.

Tampering. *Cryptographic data integrity mechanisms.* Hash-based data integrity schemes such as HMAC can be used to ensure authenticity of the data

when transmitted. The blockchain is another technology that has been recently applied for data management, e.g., for keeping immutable logs of activities, especially for medical data. *Authenticated encryption.* Cryptographic primitives that combine encryption and data integrity into a single interface.

Repudiation. *Digital Signatures.* Using a digital signature is an effective approach to preventing repudiation(Piper, 2019). This would help to prevent either a sender or receiver from denying the actions they carried out. *Audit Trails.* Having a secure and tamper-resistant log of events as well as an audit log will help to enforce accountability, i.e., to identify who carried out certain actions.

Information Disclosure. The US government in the year 2021 gave an executive order on requirements to improve the nation's cybersecurity(Biden, 2021) to prevent information and data leakage. Accordingly, a data cooperative must prevent data leakage that may lead to information disclosure. Some techniques that can help achieve this are listed below. *Permissions and Access Control Monitoring.* A proper and regular evaluation of permissions and access rights would help to quickly spot elevated privileges before they are abused or exploited. *Avoid Self-Signed Certificates.* Certificates used on all communication channels by the members should be signed by a trusted certificate authority (CA). *Proper Software Testing.* Warning and error messages that are thrown by software applications should be properly vetted for sensitive information during software development and/or testing.

Denial of Service. *Network Monitoring.* Similarly to the countermeasures against spoofing, the network traffic should be monitored and analyzed for abnormal activities. The use of firewalls and/or intrusion detection systems will help to detect and block unwanted traffic, thus helping to counter (D)DoS attacks. *Good Response Plan and Redundancy.* Although this may appear as a reactive measure to a DDoS, members of a data cooperative should be aware of the security landscape and responsibilities in the event of a (D)DoS attack. Preparing a backup to servers and other network devices would also reduce the system downtime in case of a (D)DoS attack against the data coop.

Elevation of Privilege. *Principle of Least Privilege.* A user should only be given the privilege needed to execute a task. This can be achieved with the implementation of appropriate access rights and access lists (Gegick and Barnum, 2013). *Authorization and Authentication.* Only legitimate, authorized, and authenticated users should be able to execute actions on a resource.

5 CONCLUSION

Systematic threat and security analysis are essential to the development of a system such as a data coop. In this paper, we presented security and privacy as major challenges which data coops will be facing in the near future. Then, we presented some typical threats and vulnerabilities for such systems. In order to address them, we presented the respective countermeasures. Our model will be helpful for developers and security professionals of data management systems in identifying potential security risks for a data coop and building effective defense mechanisms.

ACKNOWLEDGMENTS

We thank National Security Agency for the partial support through grants H98230-20-1-0329, H98230-20-1-0403, H98230-20-1-0414, and H98230-21-1-0262.

REFERENCES

Biden, J. (2021). Executive order on improving the nation's cybersecurity. *Retrieved*, 5(22):2021.

Blasimme, A., Vayena, E., and Hafen, E. (2018). Democratizing health research through data cooperatives. *Philosophy & Technology*, 31(3):473–479.

Chika, Y.-B. and Adekunle, A. S. (2017). Smart fabrics-wearable technology. *Int. J. Eng. Technol. Manag. Res*, 4(10):78–98.

Dahbur, K., Mohammad, B., and Tarakji, A. B. (2011). A survey of risks, threats and vulnerabilities in cloud computing. In *Proceedings of the 2011 International conference on intelligent semantic Web-services and applications*, pages 1–6.

Forte, D. (2009). The importance of log files in security incident prevention. *Network Security*, 2009(7):18–20.

Gegick, M. and Barnum, S. (2013). Least privilege. *US-SERT, available at: www.us-cert.gov/bsi/articles/knowledge/principles/least-privilege (accessed 10 April 2022)*.

Gong, D. (2022). The data cooperative model: Combating the monopolization of data.

Hardjono, T. and Pentland, A. (2019). Data cooperatives: Towards a foundation for decentralized personal data management. *arXiv preprint arXiv:1905.08819*.

Microsoft (2022). Microsoft threat modeling tool threats.

Pentland, A. and Hardjono, T. (2020). 2. data cooperatives. In *Building the New Economy*. PubPub.

Pentland, A., Hardjono, T., Penn, J., Colclough, C., Ducharmee, B., and Mandel, L. (2019). Data cooperatives: Digital empowerment of citizens and workers.

Piper, F. (2019). Digital signatures for non-repudiation. In *Practical Data Security*, pages 93–103. Routledge.

Radanliev, P., De Roure, D., and Walton, R. (2020). Data mining and analysis of scientific research data records on covid-19 mortality, immunity, and vaccine development-in the first wave of the covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5):1121–1132.

Salau, A., Dantu, R., and Upadhyay, K. (2021). Data cooperatives for neighborhood watch. In *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 1–9. IEEE.

Salau, A., Yinka-Banjo, C., Misra, S., et al. (2018). Design and implementation of a fault management system. In *International Conference on Innovations in Bio-Inspired Computing and Applications*, pages 495–505. Springer.

Stallings, W., Brown, L., Bauer, M. D., and Howard, M. (2012). *Computer security: principles and practice*, volume 2. Pearson Upper Saddle River.

UcedaVelez, T. (2012). Real world threat modeling using the pasta methodology. *OWASP App Sec EU*.

Wuyts, K. and Joosen, W. (2015). Linddun privacy threat modeling: a tutorial. *CW Reports*.

Zyskind, G., Nathan, O., and Pentland, A. (2015). Enigma: Decentralized computation platform with guaranteed privacy. *arXiv preprint arXiv:1506.03471*.