Unlinkable Data Sharing with Dynamic Access Control

Kevin Röbert
University of Hamburg
Hamburg, Germany
kevin.roebert@uni-hamburg.de

Dominik Kaaser
TU Hamburg
Hamburg, Germany
dominik.kaaser@tuhh.de

Mathias Fischer
University of Hamburg
Hamburg, Germany
mathias.fischer@uni-hamburg.de

Abstract

In an increasingly information-driven society, the volume of digital footprints left by individuals has surged significantly. Safeguarding the anonymity of data generated by computing devices is becoming more challenging as these offer deep insights into personal behaviors. We propose a user-centric and privacy-preserving data space for unlinkable data sharing based on a central intermediary. By integrating differential privacy techniques with fine-grained access control, our system allows data providers to store their data confidentially and unlinkable at the intermediary. Data consumers can then locate and request data via this intermediary, ensuring that data providers remain informed without revealing the origin of the data. Additionally, the intermediary facilitates continuous data sharing, requiring only a single data upload. Our approach is designed to protect data providers from both external and internal attackers, as well as from an honest-but-curious intermediary.

CCS Concepts

• **Security and privacy → Privacy-preserving protocols**; **Access control**; **Pseudonymity, anonymity and untraceability**.

Keywords

private data sharing, differential privacy, access control, data spaces

ACM Reference Format:
Kevin Röbert, Dominik Kaaser, and Mathias Fischer. 2025. Unlinkable Data Sharing with Dynamic Access Control. In *The 40th ACM/SIGAPP Symposium on Applied Computing (SAC '25), March 31-April 4, 2025, Catania, Italy*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3672608.3707708

1 Introduction

In today's world, the convergence of digital and physical domains has led to an unprecedented increase in data generation. The extensive collection of user and sensor data raises two significant issues: the lack of transparency in how the data is used and the diminishing control over it. Often, such data has a personal dimension, enabling deep insights into individuals' preferences [3] and behavioral patterns [6]. It has become increasingly challenging for individuals to understand which data is being collected about them and with whom it is being shared. This is especially relevant as data breaches increase in frequency and magnitude [7]. While in Europe, the General Data Protection Regulation (GDPR) has been implemented to enhance privacy – and often parts of the world might follow – it has unfortunately not met expectations [17]. The unchecked data collection practices also continue despite the GDPR. Thus, it is vital to ensure that users have sovereignty over their data, i.e., with whom it is shared, for how long, and for what purpose.

A fundamental challenge lies in managing user data, particularly regarding privacy and security. Data spaces have emerged as a promising solution, offering a secure and trustworthy environment [13]. These data spaces can be realized through a data intermediary [2, 10] that securely stores data and obtains explicit permission from data providers. Such an intermediary aligns the interests of data providers – who seek to maintain sovereignty over their data – with the needs of data consumers for research and innovation. As mediators, data intermediaries can bridge the gap between providers and consumers in the context of data spaces.

The Urban Data Trustee model highlights the critical issue of data monopolies. For example, in the Google Sidewalk Labs project in Toronto [2], a single company's exclusive control over urban data led to public distrust and low acceptance. In such cases, intermediaries can play a crucial role in preserving data sovereignty by ensuring that personal data is shared according to the preferences of individual data subjects. However, a centralized intermediary possessing unrestricted access to user data becomes an attractive target for malicious actors. To address this concern, we have identified four functional and two non-functional requirements for an intermediary-based system. The functional requirements are *data confidentiality* and *controlled access* – to prevent unauthorized data retrieval – and *anonymity* and *unlinkability* – to protect the privacy of data providers and ensure their communications cannot be associated with them. The non-functional requirements include *efficiency*, demanding optimal resource utilization, and *scalability*, requiring the system to scale linearly with increases in data volume, providers, and consumers. We also assume the presence of various types of attackers. An *external attacker* operates outside the system, aiming to exploit vulnerabilities, eavesdrop on traffic, or gain unauthorized access. An *internal attacker* acts within the system, leveraging privileged access to violate protocol specifications or access data without authorization. The intermediary is assumed to be honest-but-curious: it follows established protocols but may attempt to infer additional information. While it lacks the keys to decrypt stored data, it may analyze traffic patterns or metadata to deanonymize data providers or infer encrypted content. Crucially, its curiosity is limited to passive observation without active interference in the protocol flow.

In response to these challenges, it is vital to implement techniques that protect user data while ensuring that the intermediary cannot directly read the stored data. Achieving this goal calls for privacy-preserving methods that guarantee unlinkability for data providers, enabling them to exercise access control over their data.

SAC '25, March 31-April 4, 2025, Catania, Italy

ACM ISBN 979-8-4007-0629-5/25/03
https://doi.org/10.1145/3672608.3707708

The main contribution of this paper is a novel approach for secure, privacy-preserving, and unlinkable data sharing with dynamic access control. We propose integrating differential privacy techniques with fine-grained dynamic access control mechanisms. Our system, which relies on an intermediary, ensures both the confidentiality of the data held by the intermediary and its unlinkability.

The remainder of this paper is structured as follows: Section 2 summarizes related work. Section 3 presents our proposed system. Finally, Section 4 concludes the paper and summarizes future research directions.

2 Related Work

Privacy-preserving data sharing systems have been a focus of research for many years, encompassing both centralized [1, 4, 8, 14] and decentralized [9, 11, 15, 16] approaches. Traditional Private Information Retrieval (PIR) systems prioritize consumer privacy but often neglect provider protection and robust access control mechanisms. For instance, Chor et al.[8] introduced a PIR system without security or access restrictions. Persona[4], a centralized intermediary, uses attribute-based encryption for fine-grained sharing policies. However, it accumulates metadata, violating unlinkability, and lacks efficient open search and integrity guarantees. Similarly, SeDaSC [1] and Credential [14] enhance data sharing in cloud environments but compromise provider anonymity due to metadata exposure and limited support for many-to-many interactions. Decentralized systems such as Tor [11] and Freenet [9] focus on sender-receiver privacy but lack mechanisms for controlled access and efficient distribution of large datasets. Peer-to-peer systems like OneSwarm [15] and SQL-based retrieval frameworks like PrivApprox [5] improve anonymity through decentralized routing or proxy use. However, OneSwarm's protocol vulnerabilities undermine its anonymity guarantees, while PrivApprox lacks integrity assurances and fine-grained access control. Blockchain-based approaches, such as the system by Naz et al. [16], provide role-based access using IPFS and smart contracts but fail to ensure unlinkability.

In summary, existing systems address anonymity for data consumers and, in some cases, data providers, but they lack fine-grained access control mechanisms. Most solutions fail to meet unlinkability requirements and often limit data searches to predefined identifiers, lacking support for efficient open search. In contrast, our approach ensures that data providers retain full control over their data, while data consumers require explicit permission to access it.

3 Unlinkable Data Sharing with Dynamic Access Control

We propose an approach that enables data providers to share their data anonymously and unlinkable with data consumers via a central data intermediary that is honest but curious. With the help of the intermediary, consumers can submit requests for data approval to providers. The providers check the requests and can accept or reject them. The data is only forwarded to the consumer once the providers have given their consent.

Our unlinkable data sharing system with dynamic access control consists of three entities: data providers, data consumers, and a data intermediary and is shown in Figure 1. The system has four main functions: *Initialization*, *Storage* of data, *Localization* of data and data *Retrieval*. These functions are described in the following.

Initialization: During the initialization phase, providers and consumers sign up with the intermediary using their chosen credentials. After successful registration, providers and consumers get a unique ID for logging in.

Storage: The intermediary must facilitate the storage of encrypted data blocks without linking them to the provider, ensuring that metadata such as creation time or origin location remains exclusively with the provider. We assume a set of data providers $\mathcal{P}$ and a set of data consumers C who exchange data via a data intermediary $\mathcal{I}$. A provider $p \in \mathcal{P}$ encrypts a plaintext d_i into a ciphertext block b_i using the function $b_i \leftarrow enc(d_i, K_i)$, where K_i is a symmetric encryption key, $i \geq 1$, and all blocks have a fixed size l. Conversely, a consumer $c \in C$ decrypts a ciphertext block b_i back into plaintext d_i using the function $d_i \leftarrow dec(b_i, K_i)$. The intermediary $\mathcal{I}$ manages a database containing n tuples $(i_1, b_1), \ldots, (i_n, b_n)$, where i represents the address of an encrypted data block b_i. Plaintexts that span multiple blocks are divided into uniformly sized segments, each assigned a unique address i. A data provider, such as an application collecting data, can encrypt and store data for secure and private sharing with a third party. The provider encrypts plaintext d_i with key K_i, generating ciphertext block b_i. A uniformly random address i is assigned, and the resulting tuple (i, b_i) is transmitted anonymously (e.g., via Tor) to the intermediary $\mathcal{I}$. Thus, the intermediary has no access to the plaintext data and cannot determine which provider owns which data.

Localization: A key responsibility of the intermediary is to match incoming consumer requests with suitable providers while minimizing system overhead caused by unnecessary messaging. To achieve this, the intermediary maintains a mapping of topics to providers, where each topic represents a type of data (e.g., location or humidity) that a provider can supply. However, this mapping is sensitive, as it reveals information about providers and must therefore be protected. The objective is to balance minimizing system message load with safeguarding the privacy of the topic-provider relationships. For this purpose, we propose the use of differential privacy to construct so-called anonymity sets. Providers initially submit a list of topics for which they can offer data, enabling the intermediary to establish the mapping between topics and providers. To ensure unlinkability, providers augment or obfuscate their topic lists by introducing noise–either by adding unrelated topics or omitting actual ones. The resulting set of topics, $\mathcal{T}' \subseteq \mathbb{T}$, creates an anonymity set in which the provider's genuine topics are indistinguishable from the added "cover" topics. For instance, a provider with location data might include unrelated topics like humidity in their list. The intermediary then records the anonymized topic set $\mathcal{T}'$ alongside the provider's identifier as a tuple $(\mathcal{T}', \text{ID})$ in its topic lookup table. This approach complicates an attacker's efforts to discern the specific topics a provider can supply, enhancing privacy while maintaining the system's functionality. To construct these anonymized topic sets, we propose to use randomized response techniques, a variant of differential privacy. Randomized response works by asking individuals to randomly flip a coin in private, then answer the question truthfully if the coin lands on heads. Otherwise, the individual flips a second coin in private, answering "Yes" if the coin lands on heads or "No" if it lands on tails. The ability

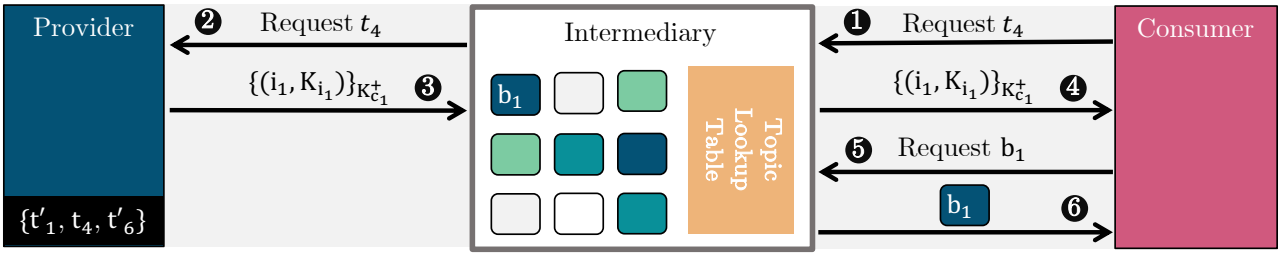

Figure 1: Model of our approach. On the left, a data provider (p_1) creates a set of topics ($\{t'_1, t_4, t'_6\}$), distinguishing real from cover topics (primed ones). On the right, a data consumer (c_1) requests data on topic t_4. The intermediary stores encrypted data blocks b_i with addresses and a lookup table.

to plausible deny the true response preserves the participants' privacy. Additionally, the algorithm is proven to be ε-differentially private [12], enabling the construction of anonymity sets locally.

In general, data localization works as follows. A data consumer, e.g., a research institution, expresses interest in obtaining data by sending a request to the intermediary. The data intermediary then attempts to locate data providers with potentially relevant data and forwards the request to them. Crucially, the intermediary does not know which data belongs to which specific provider. To achieve this, the intermediary utilizes a lookup table that lists the topics associated with the data each provider holds.

Retrieval: Data consumers can request data for a specific topic from the intermediary. The intermediary then looks up the topic in its table (which contains all topics reported by all providers) and forwards the request to all potential providers. To prevent the intermediary from distinguishing which data providers have data for a particular topic and which do not, the protocol requires all data providers that have reported the topic to respond to the request. This ensures that no information is leaked based on the responses. When a data provider holds actual data, it encrypts the data decryption key using the data recipient's public key (K^+). The provider then sends the encrypted key, along with the addresses of the relevant data blocks, to the data recipient via the intermediary in response to the request. The data consumer can use its private key (K^-) to decrypt the message and retrieve the tuple (i, K_i), where i is the address of the data block, and K_i the respective decryption key. Finally, the consumer proceeds to download the blocks and decrypt them using the shared key K_i. If a data provider does not have data for a requested (cover) topic, it responds with a dummy key and a fixed but arbitrary block address.

4 Conclusion

This paper presents an unlinkable data sharing system with dynamic access control, designed to enhance users' sovereignty in deciding with whom their data is shared.

In future work, we will test various differential privacy methods for our data localization and identify ways to allow individuals to balance privacy with message load. Additionally, we will evaluate these methods, providing a theoretical analysis of message complexity and proofs of their differential privacy, along with an ε-privacy value.

Acknowledgments

This work is funded by the European Union – NextGenerationEU. The views and opinions expressed are solely those of the author(s) and do not necessarily reflect the views of the European Union nor the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

[1] Mazhar Ali, Revathi Dhamotharan, Eraj Khan, Samee U Khan, Athanasios V Vasilakos, Keqin Li, and Albert Y Zomaya. 2015. SeDaSC: secure data sharing in clouds. *IEEE Systems Journal* 11, 2 (2015).
[2] Anna Artyushina. 2020. Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics* 55 (2020), 101456.
[3] Naveen Farag Awad and M. S. Krishnan. 2006. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quarterly* 30 (2006).
[4] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. 2009. Persona: an online social network with user-defined privacy. In *ACM SIGCOMM '09.*
[5] Martin Beck, Pramod Bhatotia, Ruichuan Chen, Christof Fetzer, Thorsten Strufe, et al. 2017. PrivApprox: Privacy-Preserving Stream Analytics. In *USENIX ATC'17.*
[6] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *MobileHCI'13.*
[7] Long Cheng, Fang Liu, and Danfeng Yao. 2017. Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7, 5 (2017), e1211.
[8] Benny Chor, Niv Gilboa, and Moni Naor. 1997. *Private information retrieval by keywords.* Technical Report. Dept. of Computer Science, Technion.
[9] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W Hong. 2000. Freenet: A distributed anonymous information storage and retrieval system. In *Designing privacy enhancing technologies: international workshop on design issues in anonymity and unobservability.* Springer.
[10] Sylvie Delacroix and Neil D Lawrence. 2019. Bottom-up data trusts: Disturbing the 'one size fits all'approach to data governance. *International data privacy law.*
[11] Roger Dingledine, Nick Mathewson, Paul F Syverson, et al. 2004. Tor: The second-generation onion router.. In *USENIX Security'04.*
[12] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* (2014).
[13] Alon Halevy, Michael Franklin, and David Maier. 2006. Principles of dataspace systems. In *ACM PODS '06.*
[14] Felix Hörandner, Stephan Krenn, Andrea Migliavacca, Florian Thiemer, and Bernd Zwattendorfer. 2016. CREDENTIAL: a framework for privacy-preserving cloud-based data sharing. In *11th ARES.*
[15] Tomas Isdal, Michael Piatek, Arvind Krishnamurthy, and Thomas Anderson. 2010. Privacy-preserving p2p data sharing with oneswarm. *ACM SIGCOMM '10* (2010).
[16] Muqaddas Naz, Fahad A Al-zahrani, Rabiya Khalid, Nadeem Javaid, Ali Mustafa Qamar, Muhammad Khalil Afzal, and Muhammad Shafiq. 2019. A secure data sharing platform using blockchain and interplanetary file system. *Sustainability* 11, 24 (2019).
[17] Mathew J Schwartz. 2019. GDPR: Europe Counts 65,000 Data Breach Notifications So Far. *Online: https://www. bankinfosecurity. com/gdpr-europe-counts-65000-data-breachnotifications-so-far-a-12489* (2019).