PrivacyGuard: Enforcing Private Data Usage with Blockchain and Attested Execution

Ning Zhang[1], Jin Li[2], Wenjing Lou[1], and Y. Thomas Hou[1]

[1] Virginia Polytechnic Institute and State University, USA
[2] Guangzhou University, China

Abstract. In the upcoming evolution of the Internet of Things (IoT), it is anticipated that billions of devices will be connected to the Internet. Many of these devices are capable of collecting information from individual users and their physical surroundings. They are also capable of taking smart actions, which are usually from a backend cloud server in the IoT system. While IoT promises a more connected and smarter world, this pervasive large-scale data collection, storage, sharing, and analysis raise many privacy concerns.

In the current IoT ecosystem, IoT service providers have full control of the collected user data. While the original intended use of such data is primarily for smart IoT system and device control, the data is often used for other purposes not explicitly consented to by the users. We propose a novel user privacy protection framework, PrivacyGuard, that aims to empower users with full privacy control of their data. PrivacyGuard framework seamlessly integrates two new technologies, blockchain and trusted execution environment (TEE). By encoding data access policy and usage as smart contracts, PrivacyGuard can allow data owners to control who can have what access to their data, and be able to maintain a trustworthy record of their data usage. Using remote attestation and TEE, PrivacyGuard ensures that data is only used for the intended purposes approved by the data owner. Our approach represents a significant departure from traditional privacy protections which often rely on cryptography and pure software-based secure computation techniques. Addressing the fundamental problem of data usage control, PrivacyGuard will become the cornerstone for free market of private information.

1 Introduction

The emergence of the Internet of Things (IoT) is the result of rapid advancement in technology in multiple fields. In the past two decades, we have witnessed an explosive deployment of *communications and networking technologies*, especially wireless technologies. At the same time, *mobile devices* have transformed from limited embedded systems to highly capable general purpose computing platforms. A variety of *mobile devices* with increased capability and intelligence are being introduced at a speed of approximately half a billion each year in recent years [1]. New life-changing *mobile apps* are being introduced every day.

IoT promises a more connected and smarter world. However, as a wide variety of things are increasingly embedded around us and more and more data about us are collected, shared, and analyzed, there is an increased concern on privacy. Individuals share personal information with people or organizations within a particular community for specific purposes. For example, individuals may share their medical status with healthcare professionals, product preferences with retailers, and real-time whereabouts with their loved ones. When information shared within one context is exposed in another outside of the intended context, people may feel a sense of privacy violation [2]. This *contextual nature of privacy* implies that privacy protection techniques need to address at least two aspects: 1) what kind of information can be exposed to whom, under what conditions; and 2) what is the "intended purpose" or "expected use" of this information.

Much research has been done to address the first privacy aspect. There has been a large body of research work on *data access control* that aims to ensure that only authorized data consumers can access private user data [3–11]. Another line of research is *data anonymization* that tries to ensure if sensitive data needs to be published, it is published anonymously, i.e. the personal identifiable information is removed from the data and the linkability between the published data and individual users is carefully eliminated [12–16]. Only recently, there have been a few works that attempted to address the second aspect of privacy, i.e., data used only for the intended purposes [17, 18]. In fact, with the current practice, once an authorized user gains access to the data, how this user would use the data, whether or not he/she would use the data for purposes not consented by the user, or simply pass the data to another party (i.e., data monetization) is up to this new ”data owner." Legal or regulatory measures may be taken to put some constraints on this, but technical approaches that allow users to specify and enforce the intended use of their data are lacking in general.

In this paper, we propose *PrivacyGuard*, a private data utilization framework, to address this very challenging privacy problem in IoT – how to empower a data owner in an IoT system to have full control over how his/her personal data is used. The data owner should not only be able to control who can have what access to his/her data, but also be ensured that the data is used only for the intended purposes. To realize the envisioned functionality of PrivacyGuard, there are three key requirements.

- User shall be able to define his/her own data access policy concerning to whom she will share the data at what time for what purpose and at what price. The framework shall also support rich encoding of different data utilization conditions.
- There shall be strict enforcement on the data policy set forth by the data owner. Each usage of the user data shall have a verified proof that it is compliant with the policy and data content is well protected during the utilization.
- Each data usage shall be recorded on a platform that offers non-repudiation and transparency.

In PrivacyGuard, users' privacy policies are embedded as smart contracts on a blockchain platform. In recent years, blockchain, the technology behind *Bitcoin* [19] and *Ethereum* [20], has emerged as a popular technology for distributed public repository of data. Bitcoin [19], exploiting the blockchain as a public ledger to store cryptocurrency exchanges (called transactions), is the first implementation of blockchain technology. Other emerging platforms also using the blockchain are quickly gaining popularity, such as Ethereum [20], HyperLedger [21], IOTA [22]. *Smart contract*, a program that runs on the blockchain and has its correct execution enforced by the consensus protocol, has seen fast adoption and increased use in the Ethereum platform. In PrivacyGuard, smart contracts are used to facilitate the transactions of private data utilization on the private data market, providing access control, tamper-resistant record of data utilization.

Smart contract provides a mechanism to ensure desired privacy protection at the protocol level. However, when the program is running on a third party computer (such as in the Cloud) which is not fully trusted by the data owner, the confidentiality of user data as well as the faithful execution of the protocols can no longer be guaranteed. Pure software-based approaches, such as homomorphic encryption and secure multi-party computation, for secure computation in the cloud have been investigated extensively in the past decade. However, the heavy overhead on generic constructions of secure computation makes practical adoption infeasible with the current computing power. We propose to take a different approach to support generic computation by developing a system level security mechanism exploiting trusted platform such as the Intel SGX enclave technology, which provides a hardware-enforced isolated secure execution environment. By processing data within the Intel SGX enclave, it is possible to ensure the data confidentiality and enforce the intended data usage.

2 PrivacyGuard Overview: A Framework Enabling User Control on IoT Data Usage

Things in IoT can take many different forms, from simple RFIDs attached to merchandises, smart thermostats installed in the classrooms, to wearable medical devices on patients and video cameras at home.

Some powerful devices, such as IP cameras and smart TVs, can connect directly through the Internet to the backend application server in the cloud. Some other IoT systems, such as Samsung smart things, make use of something like a smart hub to orchestrate communications between heterogeneous things. However, in most cases, the intelligence of the system is hosted at a cloud backend, therefore all the data generated from the system is stored within the vendor cloud. Data collected by IoT devices could be used directly by the vendor IoT applications. They could also be shared with other services, including various big data analytics tasks.

The huge amount of data collected by IoT and the desire of broad information sharing raise serious privacy concerns.

- **Confidentiality Protection on User Data** When data is generated under the current paradigm, it is stored in the vendor's cloud storage. The data is often stored in plaintext, and the access control on user data relies on the vendor system. Even when the data is stored in an encrypted storage, the user does not control the encryption key. When the data is less sensitive such as video from a driveway camera, plaintext storage might be acceptable. However, video from a camera in the bedroom can contain sensitive private information and would require an appropriate level of protection, and such data should not be exposed even to the service provider. Therefore, user-controlled, rather than service provider-controlled, encryption/decryption is fundamental in IoT data privacy.

- **Verifiable User-controlled Fine-grained Data Access** Under the current paradigm, once the data is uploaded to the vendor cloud, it belongs to the vendor under a service agreement. A user could grant access to his data to someone. But there is no way for the user to find out who actually accessed his private data, not to mention for what purpose. Lack of transparency and verifiability on data access often prompts users to choose the most restrictive data sharing agreement. This is evidential in a recent study on data sharing practice among windows error reporting users, where most people choose not to share data when they do not know how the data may be used. A public service that keeps track of user data usage and makes it auditable by data owners is therefore essential to not only protect user privacy but also promote data sharing in the community.
- **Provable Legal Binding on User Data Usage** Service level agreements and legal contracts are the only control over how data is stored, shared, and mined under the current IoT ecosystem. As more and more devices are connected to the network, we are witnessing an economic drive of intelligence collected from mining the IoT data. On one hand, the intelligence reaped from mining IoT data could help provide quality service, increase convenience, lower the cost of operation, etc. On the other hand, misuse of such information could lead to injustices, such as a patient being denied of health insurance due to a health condition inferred from his medical IoT system. To realize the grand vision of a more connected and smarter world, the capability to provide flexible and provable legal binding over the use of user personal data is the utmost capability our society needs in the era of Internet of Things.

2.1 PrivacyGuard Architecture

As shown in the left half of Figure 1, an IoT system can be divided into four layers based on the technical supports they provide. The lowest layer is the *Thing/Device layer,* which is made up of various smart objects integrated with sensors and actuators. This is the IoT system's interface to the physical world. The sensors and actuators will interact with their physical environment, allowing real-time information to be collected and processed (mostly signal processing). Layer 2 is the *Network layer* which provides interconnectivity of various wireless access technologies, and supports routing functions. The highest is the *Cloud layer*, which is where the backend services/applications reside. It is a data concentration point and where most of the data analytics happens.

Between the cloud layer and the network layer is what we call the *Service Support layer*. This layer has more computation and storage capability and is capable of carrying out some important information processing tasks, possible through data analytics. From a security and privacy point of view, security control and device management, process modeling and information flow control, such as data filtering, aggregation, can all happen at this layer. The placement of this layer depends on the network architecture of an IoT system. For a Cloud-based IoT system, the layer would be in the Cloud. For an IoT system that adopts a Device-Gateway-Cloud architecture by leveraging edge computing, this layer could be at the edge node.

Figure 1 shows the system architecture of the proposed *PrivacyGuard* framework. Although we have been using the term *users* to refer to individuals or organizations

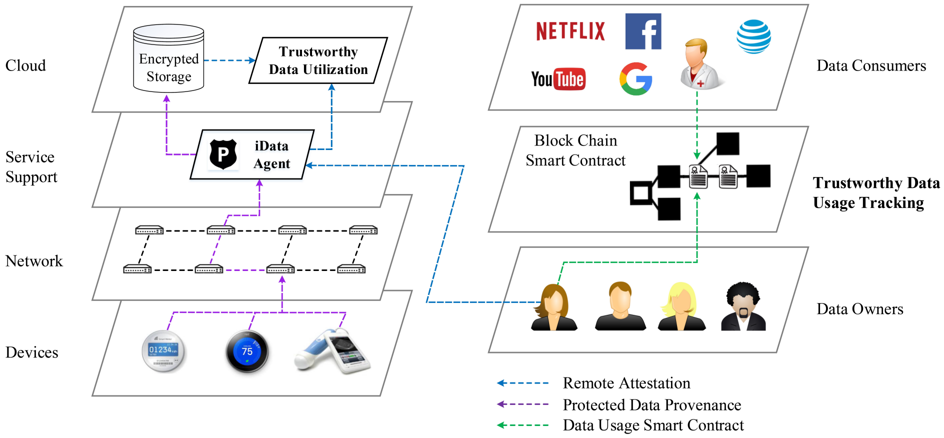
Cloud
Service Support
Network
Devices
Encrypted Storage
Trustworthy Data Utilization
iData Agent
Data Consumers
Block Chain Smart Contract
Trustworthy Data Usage Tracking
Data Owners
Remote Attestation
Protected Data Provenance
Data Usage Smart Contract

Fig. 1. IoT System Architecture and Proposed PrivacyGuard Framework

who are using an IoT system, in what follows, we differentiate two roles that an IoT user can take. We refer to the individual or organization that owns the IoT devices and produces IoT data as *data owner* and the entity that needs to access and use IoT data as *data consumer*.

Main Components There are three main components in the PrivacyGuard architecture.

- **Blockchain:** We employ an external blockchain (such as Ethereum) to enable an accountable distributed data repository for publishing access policy and facilitating data use recording. For data access control, a data owner can encode the terms and conditions regarding the access to his/her personal data as a smart contract. Data uses are recorded as transactions that interact with the smart contract. Here the blockchain serves as a public, auditable, and irreversible data repository, thus providing transparency of user policies as well as public verifiability of data usage.
- **iDataAgent (an Enclave):** iDataAgent is a trusted entity and is an instance of the iDataAgent program running in a TEE. iDataAgent acts as a broker for user data. Any data that goes in and out of the user data repository will go through iDataAgent. Private user data collected by the IoT devices will first be sent to the iDataAgent for processing. iDataAgent manages the keys for the data owner and the data encryption/decryption for that user. Sensitive data will be encrypted by the iDataAgent before pushed to the cloud for storage. iDataAgent is also responsible to remotely attest the function execution enclave in the data consumer before passing the data decryption key to it.
- **Encrypted Storage:** Private user data will always be encrypted when they are at rest in the cloud. This will ensure data confidentiality at rest against the cloud service provider.

2.2 Workflow

In what follows we outline the workflow of the proposed PrivacyGuard. We separate the workflow into three stages: data generation (encrypting user data), data access binding generation (contract negotiation), and data utilization (contract execution).

- **Data Generation and Key Management** In this stage, user data is collected and uploaded to the cloud storage. We propose to build a trusted entity, iDataAgent, at the service support level using Intel SGX secure enclave technology. The framework allows individual data owners to manage the keys used to encrypt/decrypt their data before uploading to the cloud storage through iDataAgent. A straightforward solution to initialize the master secret between a data owner and his iDataAgent enclave is to bootstrap it when a data owner first signs up for the service. Upon successful remote attestation of the iDataAgent enclave, the data owner can transmit his secret key to iDataAgent through the secure channel established along with the remote attestation. This key can then be used to derive data encryption/decryption and integrity check keys for this user. When the user data is generated, it is transmitted to iDataAgent instead of directly to the service provider such as Samsung smart home cloud. iDataAgent encrypts user data using the derived keys before pushing them to the cloud for storage. There are multiple ways to secure the communications between user IoT devices and iDataAgent. To minimize the changes necessary to the current IoT system implementation, we assume that the IoT devices can be reconfigured to connect to iDataAgent rather than Samsung Smart Home server, and rely on existing SSL/TLS implementations to establish the secure channel.
- **Policy Generation and Contract Negotiation** Our framework allows a data owner to define the access policy for the data he generated. The policy is encoded in a smart contract and committed to the blockchain. A smart contract involves at least the following information, *Policy = [data type, data range, operation, consumer, expiration, cost]*, where the "'intended use" of data of certain *type, range* is coded as *operation*, which can be arbitrary computer programs attestable by iDataAgent.
- **Data Utilization - Contract Execution** Smart contract, by design, can only embed some simple logics (functions) and the trustworthy execution of those functions is enforced by the consensus protocol. The "intended use" of the data can be arbitrary computer programs. Thus, it is impractical to embed them into a smart contract and have their trustworthy execution results enforced by the consensus protocol. In PrivacyGuard, we propose to use smart contract and blockchain for trustworthy bookkeeping of user access policy, consumer data usage record, and secure payment transfer. We use the trusted entity iDataAgent to ensure that only programs for the "approved use" can have access to the data and that the program will be executed in a remotely attested separate TEE for contract execution.

  When a data consumer app requests the use of the data, iDataAgent remotely attests the contract execution environment and the function to be executed on data. Only when both the environment is trustworthy and the function to be executed is as specified in the smart contract, will iDataAgent pass the data decryption key to the contract execution enclave. Encrypted data can be obtained by the execution enclave from the Cloud storage. Note that an additional layer of defense can be built on the Cloud storage to grant access only to encrypted data as specified in the data access contract. When the contracted operation is finished, the contract execution enclave will commit a transaction to the blockchain to certify that the contracted operation is finished, thus finalizing the final transaction and recording the instance of data usage. In addition, it will clean up all the key materials as well as data inside the enclave to prevent data reuse.

3 Conclusion

In this position paper, we propose *PrivacyGuard*, a novel user privacy protection framework that aims to empower data owners with full privacy control of their data. Two important aspects of data privacy shall be addressed: 1) how to allow data owners to control who can have what access to their data, and be able to maintain a trustworthy record of their data usage; and 2) how to ensure that data is only used for the intended purposes approved by the data owner. To accomplish the afore-mentioned privacy goals, the proposed PrivacyGuard framework seamlessly integrates two new technologies, *blockchain* and *trusted execution environment (TEE)*.

The proposed approaches are novel, representing a significant departure from traditional privacy protection researches that rely on cryptography and pure software-based secure computation techniques. Hardware-assisted approaches will provide a more powerful and more practical solution to the very challenging privacy problem. The unique combination of blockchain and TEE technologies will enable new privacy protection capabilities, i.e., verifiable data usage tracking and data use compliance enforcement. We believe PrivacyGuard framework is a foundational technology for user privacy control in the era of Internet-of-things and data intelligence.

Acknowledgment

This work was supported in part by US National Science Foundation under grants CNS-1446478 and CNS-1443889.

References

1. Cisco visual networking index: Global mobile data traffic forecast update, 2016-2021 white paper. `https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html`.
2. National privacy research strategy. `https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf`.
3. Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM Conference on Computer and Communications Security*, CCS '06, pages 89–98, New York, NY, USA, 2006. ACM.

4. Amit Sahai. Ciphertext-policy attribute-based encryption. In *In Proceedings of the IEEE Symposium on Security and Privacy*, pages 321–334, 2007.
5. Melissa Chase and Sherman S.M. Chow. Improving privacy and security in multi-authority attribute-based encryption. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, pages 121–130, New York, NY, USA, 2009. ACM.
6. Yvo Desmedt and Arash Shaghaghi. Function-based access control (fbac): From access control matrix to access control tensor. In *Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats*, MIST '16, pages 89–92, New York, NY, USA, 2016. ACM.
7. Adam Bates, Ben Mood, Masoud Valafar, and Kevin Butler. Towards secure provenance-based access control in cloud environments. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, CODASPY '13, pages 277–284, New York, NY, USA, 2013. ACM.
8. Enrico Bacis, Sabrina De Capitani di Vimercati, Sara Foresti, Stefano Paraboschi, Marco Rosa, and Pierangela Samarati. Mix&slice: Efficient access revocation in the cloud. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 217–228, New York, NY, USA, 2016. ACM.
9. Guojun Wang, Qin Liu, and Jie Wu. Hierarchical attribute-based encryption for fine-grained access control in cloud storage services. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*, CCS '10, pages 735–737, New York, NY, USA, 2010. ACM.
10. Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Achieving secure, scalable, and fine-grained data access control in cloud computing. In *IEEE INFOCOM 2010*, San Diego, CA, USA, March 2010.
11. Sabrina De Capitani Di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. Encryption policies for regulating access to outsourced data. *ACM Transactions on Database Systems (TODS)*, 35(2):12, 2010.
12. Cynthia Dwork. Differential privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II*, ICALP'06, pages 1–12, Berlin, Heidelberg, 2006. Springer-Verlag.
13. Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1), March 2007.
14. Latanya Sweeney. K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, October 2002.
15. Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115. IEEE, 2007.
16. Pierangela Samarati. Protecting respondents identities in microdata release. *IEEE transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.
17. Guy Zyskind, Oz Nathan, and Alex 'Sandy' Pentland. Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW)*. IEEE, 2015.
18. Guy Zyskind, Oz Nathan, and Alex Pentland. Enigma: Decentralized computation platform with guaranteed privacy. *arXiv preprint arXiv:1506.03471*, 2015.
19. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.
20. Ethereum: Blockchain app platform. `https://www.ethereum.org/`.
21. Christian Cachin. Architecture of the hyperledger blockchain fabric. In *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, 2016.
22. M Divya and Nagaveni B Biradar. Iota-next generation block chain. *International Journal Of Engineering And Computer Science*, 7(04):23823–23826, 2018.