Toward Enforceable Data Usage Control in Cloud-based IoT Systems

Wenjing Lou, Y. Thomas Hou, Virginia Tech
Ning Zhang, Washington University in St. Louis

Award #:
CNS-1916902, CNS-1916926

Data Privacy at a Crossroads

The abundance of rich varieties of data is enabling many transformative big-data applications in cloud-based IoT that have profound societal impacts. However, there are also increasing concerns regarding the improper use of individual users' private data, either by the cloud or third-parties. Many argue that the technology that customizes our experience in the cyber domain is threatening the fundamental civil right to privacy.

Three Recent Data Privacy Violations (Among Many Others):

Incident	Year	Cause and Records Exposed	Potential Improper Use
First American Data Leak	2019	885 million personal and financial records including SSNs, bank account numbers, mortgage and tax records remained unprotected on company website for 2 years.	Identity thefts and financial scams
Facebook-Cambridge Analytica Scandal	2018 (Started In 2015)	A Facebook API original designed to allow a third-party app to poll the profiles of participants, was misused by Cambridge Analytica to collect up to 87 million profiles through their social network without their consent.	Social engineering for political purposes
Equifax Data Breach	2017	An intrusion into Equifax's computer system followed by a cybercrime identity theft event that potentially impacted 143 million consumers.	Mass identity thefts and credit abuses

The Root Cause of Data Privacy Violation in Cloud:

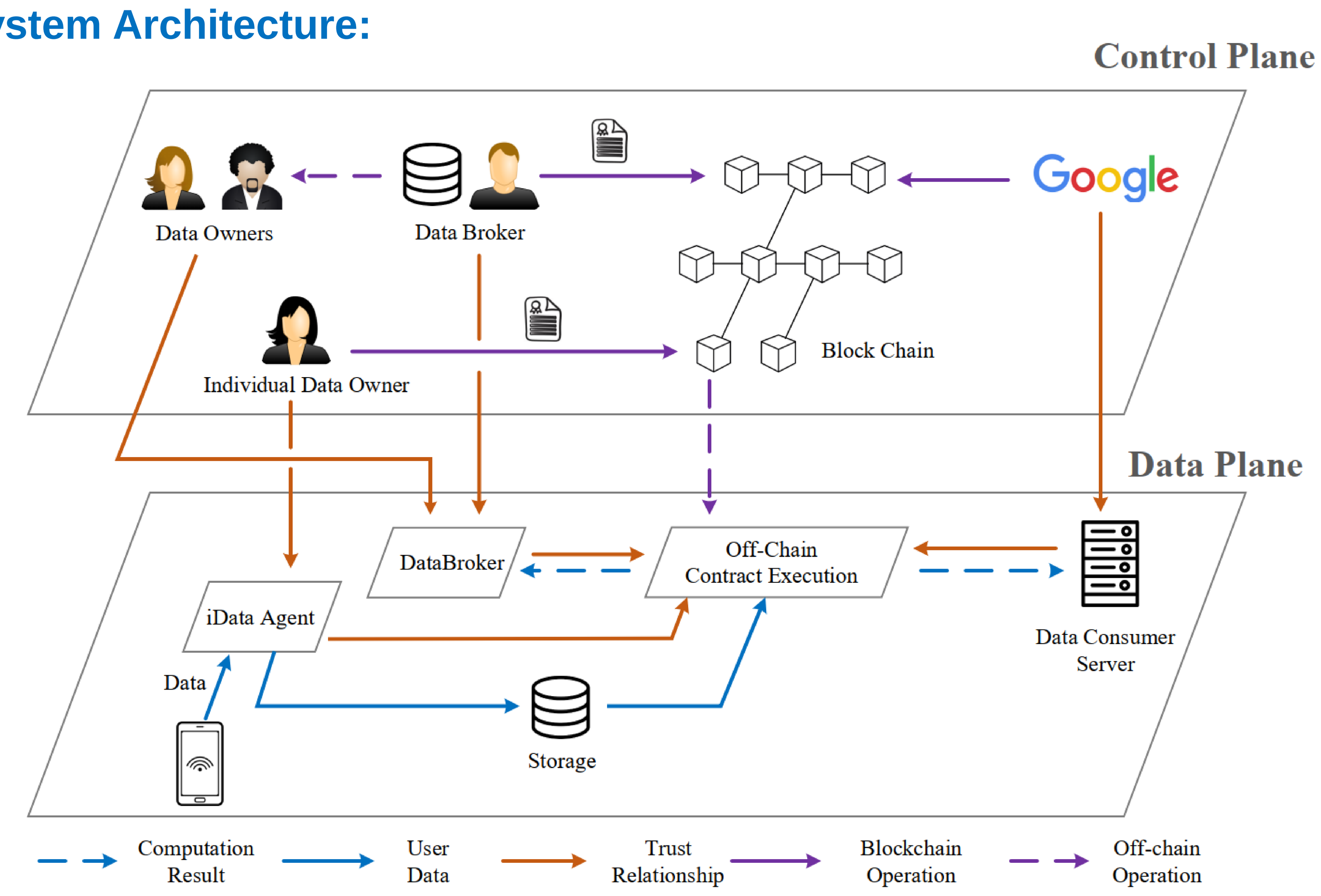
- ❑ Lack of effective security mechanisms to control data access & usage
- ❑ Centralized storage of unencrypted data. Users lose control of their plaintext data once they are provisioned to third parties.
- ❑ Lack of non-repudiable data usage auditing service and legal binding.

Research Objectives

- ✓ **Confidentiality Protection on User Data:** Data encryption/decryption are fully controlled by data owners. Untrusted parties (cloud and data consumers) can not obtain or possess data owners' plaintext data.
- ✓ **User-Controlled Fine-Grained Verifiable Data Access and Usage Recording:** Data owners are able to control who can access which data items under what conditions for what usage. The data usage records should be non-repudiable and auditable by data owners.
- ✓ **Enforceable Legal Binding on User Data Usage:** The security mechanism of a data sharing system should be able to capture user-defined privacy policies and then enforce the compliance of the policies during the execution of data access.
- ✓ **Enabling Data Market and Sharing Economy:** The three objectives above are essential to not only protecting data owner privacy but also promoting a vibrant data sharing economy, in which data owners can confidently sell the right to use their data to data consumers for profit without worrying about data leakage or misuse.

Our Solution — PrivacyGuard

System Architecture:



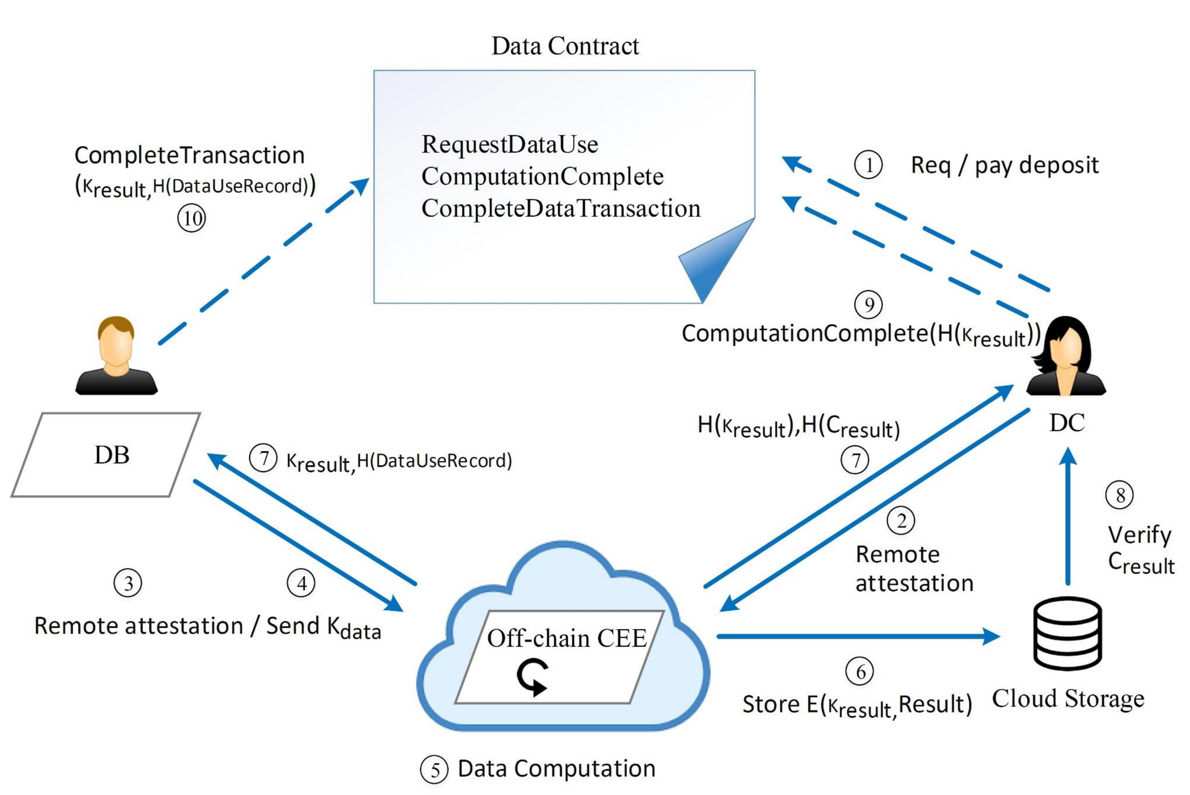
Key Technology and Methodology:

- ❖ **Blockchain smart contract** for data access, usage control, and usage record tracking. (eg. Ethereum)
- ❖ **Trusted execution environment (TEE)** for the execution of data analytics task without exposing plaintext data into untrusted domain. (eg. Intel SGX)

System Components:

- ❑ **Data Owner (DO):** An individual user that has data to share.
- ❑ **iDataAgent (iDA):** A TEE-enabled entity that manages the private data for an individual DO.
- ❑ **Data Broker (DB):** A TEE-enabled entity that manages the private data for a collection of DOs.
- ❑ **Data Consumer (DC):** An entity that wants to perform computation on DO's private data. Google and Facebook are good examples.
- ❑ **Off-Chain Contract Execution Environment (CEE):** A TEE-enabled entity that decrypts DO's encrypted private data and executes DC's computation on them inside its secure enclave. CEE's enclave will be attested by the DC and the iDA/DB before the computation.
- ❑ **Blockchain Smart Contract:** A blockchain application that enforces a DO's data access and usage rules and records DC's usage history. It is invoked by DC and concluded by iDA/DB.
- ❑ **Cloud Storage:** An entity that stores DOs' encrypted private data and CEE's encrypted computation result.

Commit Protocol for Off-chain Contract Execution (Still in Work)



Notes:

① allows DC to invoke contract via payment.

⑨ records DC's usage of DO data in blockchain.

⑩ publishes the result Key K_{result} and transfers DC's deposit to the involved DOs.

In the end, the objectives are met:

- ✓ DOs get money for allowing computation on their private data without exposing the plaintext data.
- ✓ DC obtains the computation result by following DOs' policies and without knowing DOs' data.
- ✓ Data usage is recorded in blockchain and auditable by the public.

Microbenchmark Performance

Remote attestation to AWS Cloud Intel SGX server: $\approx 33ms$

Remote attestation + K_{data} **provision** to local Intel SGX server: $\approx 70ms$

Calling smart contract function in Ethereum blockchain: $10 \sim 16s$

Project Agenda

- ❖ Use formal verification methods to prove the security of PrivacyGuard and the commit protocol.
- ❖ Implement all components of PrivacyGuard system. Design an automated evaluation scheme that takes performance measurements on remote attestations and blockchain smart contract function calls.
- ❖ Deploy PrivacyGuard on public cloud services (eg. AWS) and perform large-scale experiments on different smart contract platforms and machine learning algorithms to evaluate performance and cost.

Preliminary work: Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, "A survey of distributed consensus protocols for blockchain networks," arXiv preprint arXiv:1904.04098, Apr, 2019.