A Data Computing Platform with Privacy Evidence

Prashanth Mohan

Overview

- Increased regulatory pressure to ensure data privacy. Much utility for external access to data.
- However, once data is released there is no guarantee on how data is consumed.
- Evidence of data privacy and security should be required of all applications.
- Applications perform both aggregation-based mining as well as silo-based services.

Differential Privacy

\[
\Pr[A(D) \in S] \leq \exp(\varepsilon) \times \Pr[A(D') \in S]
\]

Maximize the accuracy of queries from statistical databases while minimizing the chances of identifying its records.

- Differentially private results are composable, i.e. \( f(g(x)) \) is differentially private for any \( f \), if \( g(x) \) is differentially private.
- Accuracy of answer improves with data set size.
- Sample and Aggregate - asymptotically normal queries are allowed.

Privacy Evidence

End user

- Privacy policy
- Privacy evidence

App developer

- API
- App

Sealed container

Privacy evidence

Privacy policy

API

App

Private data

Layer

Storage

Layer

App

Layer

Web Frontend

Data Analyst

1. Computation
2. Accuracy
Differentially Private Answer

Data Set Manager

Computation Manager

Data Set

Manager

Computation

Manager

Dataset

Blocks

T

T1

T2

T3

... T

Program

1. Data Set
2. Privacy
Budget (\( \varepsilon \))

End Users

T1

T2

T3

... T

Automation

Cloud

Trusted Boot Sequence

End Users

tboot

Boot Loader

Hardware

BIOS

Linux Kernel

Initial RAM FS

Secure Block Device

Root File System

Read/Write on demand

TPM (Remote Attestation)

Secure Block Device

TPM Chip (Remote Attestation)

IPTables

Controller

ACL Store

Friend Share

LXC Containers

Ether

Pad

K/V Proxy

FS Proxy

DeDup

Storage

Untrusted Computation

SEE

SEE

SEE

RPC Layer

Web Frontend

App

Layer