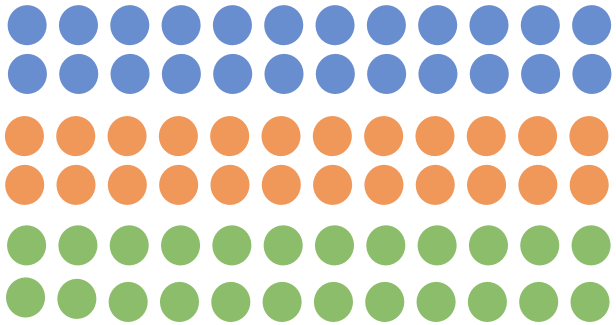
TimeCrypt: Encrypted Data Stream Processing at Scale with Cryptographic Access Control

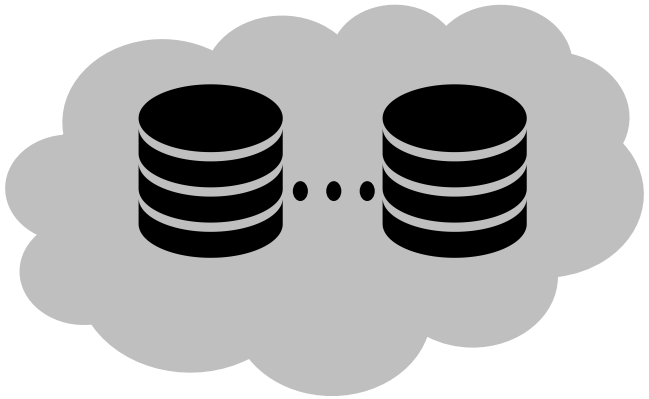
Lukas Burkhalter, Anwar Hithnawi, Alexander Viand,
Hossein Shafagh, Sylvia Ratnasamy

Time Series Data
is Emerging Everywhere

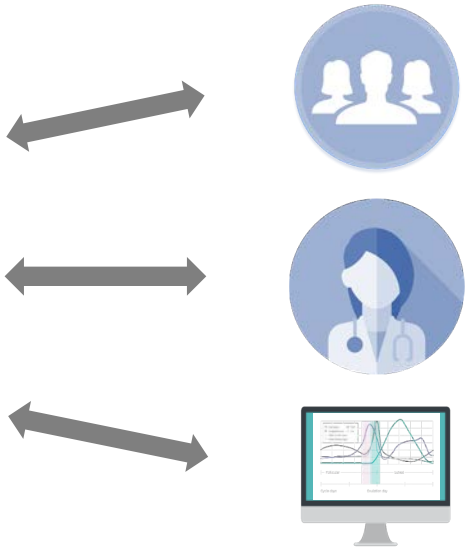
Monitoring, Telemetry, Internet of Things

Data Sources





Analytics



Applications/Services

Problem: High resolution sensitive data!
Server compromise results in privacy risks

Time

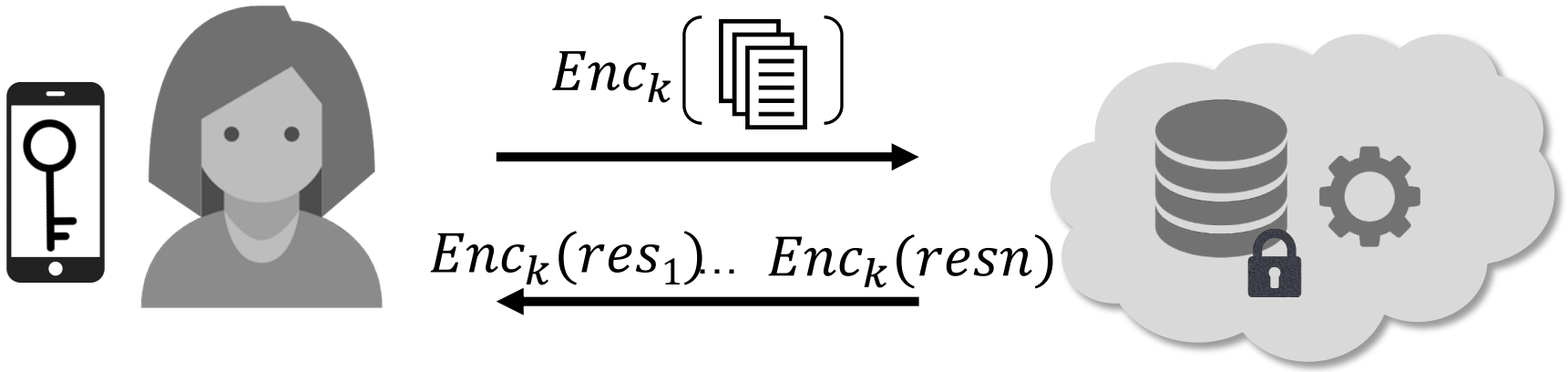
Data Breaches

> 45 billion records

*since 2004 https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

Encrypted Data Processing

- Keep data encrypted while in-use → preserve confidentiality and functionality
- Encrypted Databases → relational databases, graph databases, key-value stores
  - E.g., CryptDB [SOSP'11], BlindBox [SIGCOMM'15], Seabed [OSDI'16], Talos [SenSys'15]



Can we enable encrypted data processing for time series workloads?

Challenge I
Scalability and Interactivity

Time Series Databases

Time series workloads are different:

- ✓ Primarily INSERTS to recent time interval (append)
- ✓ Statistical queries over time ranges

Requirements:

- ✓ High throughput writes
- ✓ Large volumes of data
- ✓ Support for time-based queries

Time Series Databases

Time series workloads are different:

- ✓ Primarily INSERTS to recent time interval (append)
- ✓ Statistical queries

Requirements:

- ✓ High throughput w
- ✓ Large volumes of data
- ✓ Support for time-based queries

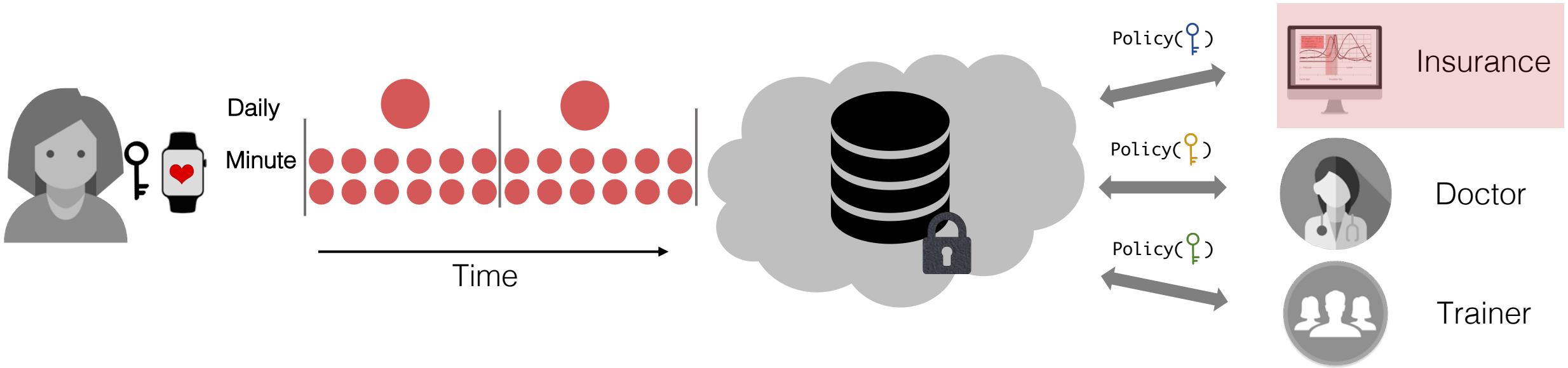
Scalability and Latency

- Memory Expansion (~100x)
- Enc/Dec Time (~milliseconds)
- Ciphertext Aggregation (~1000x)



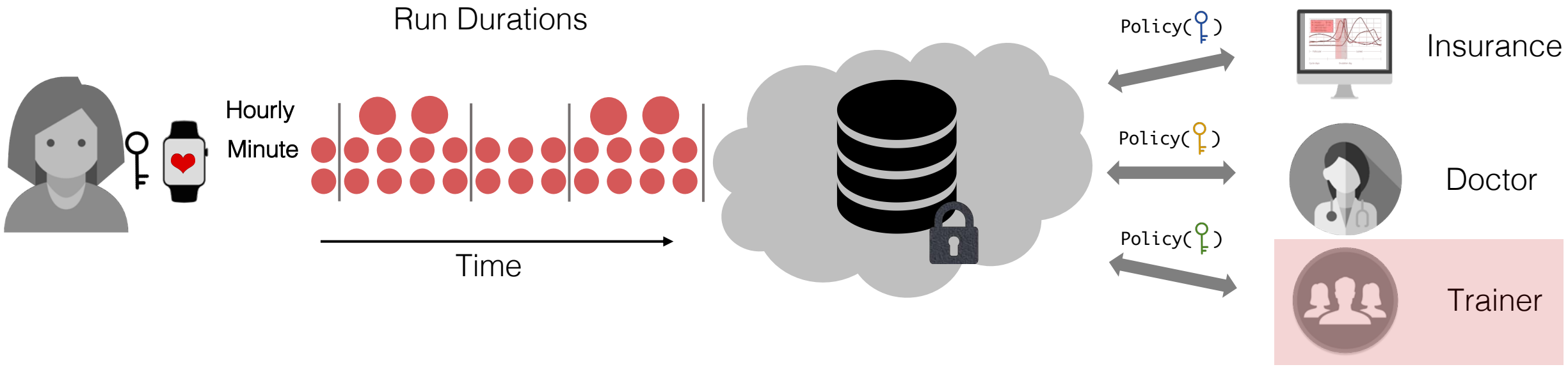
Challenge II
Secure Sharing

Selective Data Sharing



Selective Data Sharing

Selective Data Sharing



How to enable users to **selectively** share their encrypted data?
Enforce access control semantics **cryptographically**

TimeCrypt in a Nutshell

Data is encrypted end-to-end:

✓ Scalable computation over large volumes of encrypted data

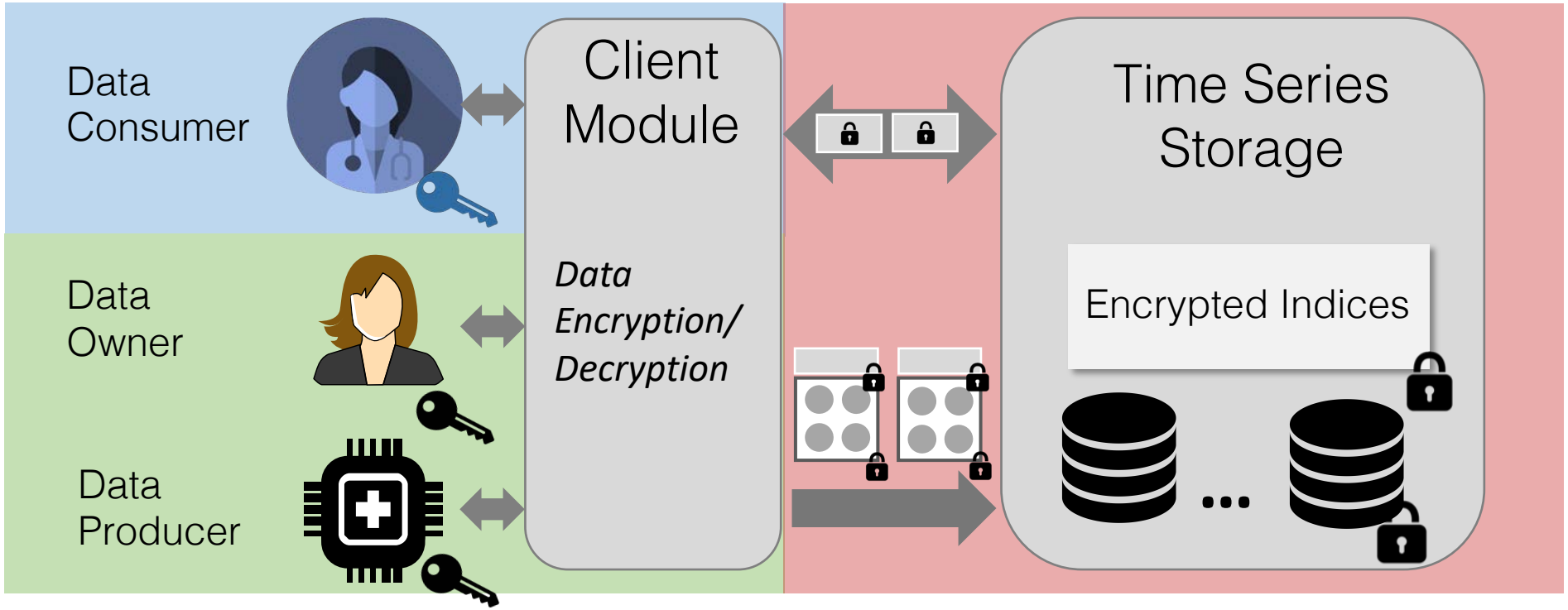
✓ Key time-series data functionalities, analytics, lifecycle operations

✓ Cryptographic access control → selective access to encrypted data

✓ Verifiable computation

Overview and Threat Model

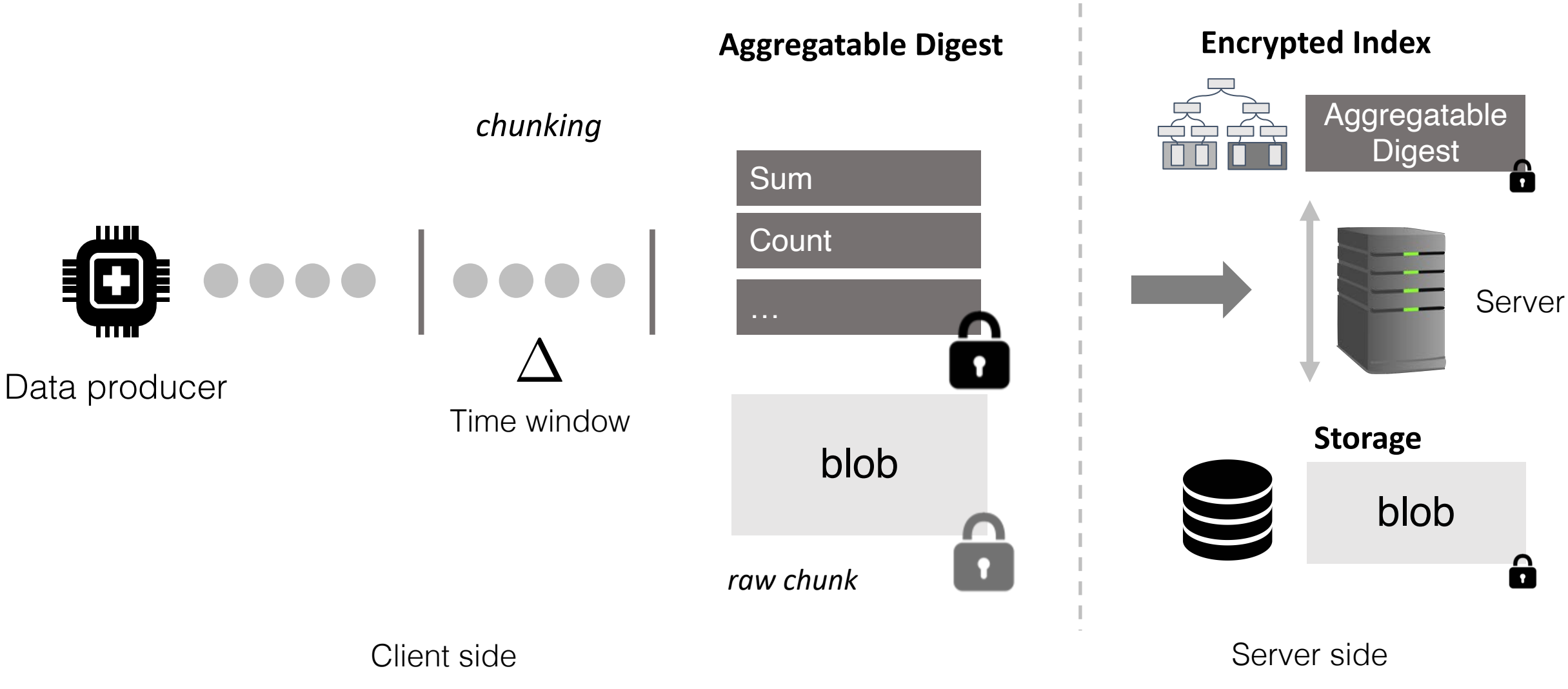
Semi-Trusted: ***Data access according to an access policy***



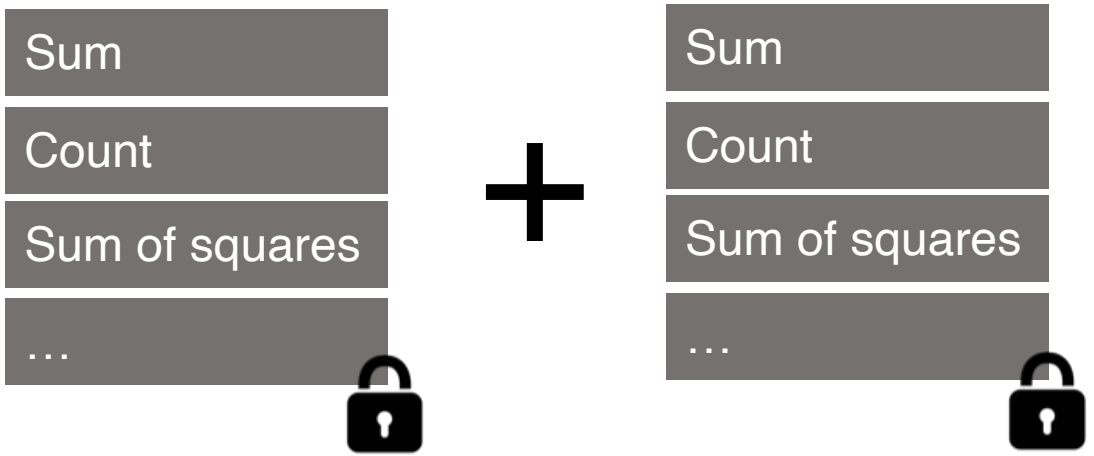
Trusted: ***Full data access***

Untrusted: ***Confidentiality + Integrity***

Writing Data Streams



Aggregatable Digest

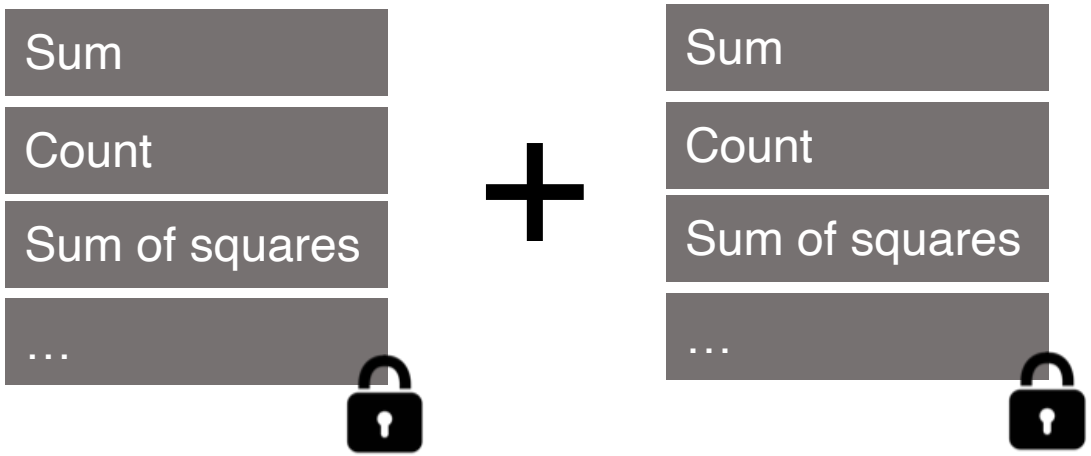


- **Additive homomorphic** encryption is the underlying construction

$$m_1 + m_2 = \mathrm{Dec}(\mathrm{Enc}(m_1) \oplus \mathrm{Enc}(m_2))$$

How to support statistics and analytics beyond addition?

Aggregatable Digest

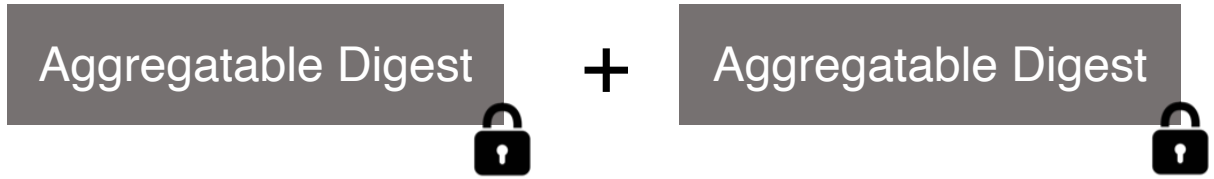


- **Additive homomorphic** encryption is the underlying construction

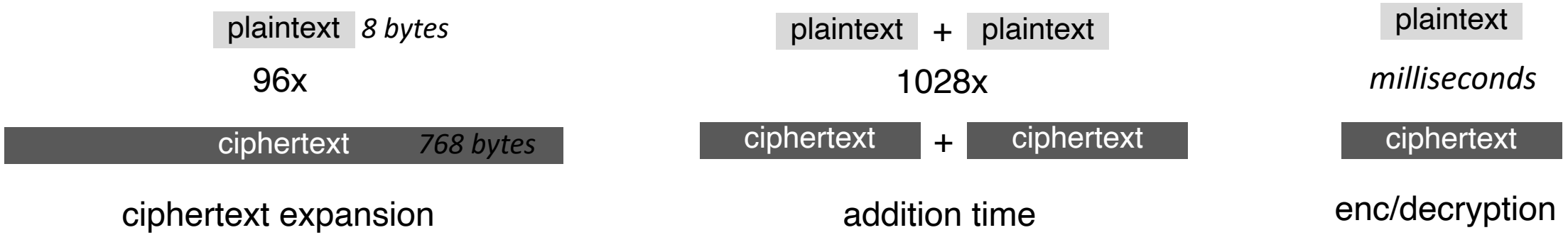
$$m_1 + m_2 = \text{Dec}(\text{Enc}(m_1) \oplus \text{Enc}(m_2))$$

- Leverage known encoding techniques → If we can compute sum privately, then we can compute f(·) privately
  - average, sum, count, variance, min/max (approx.), histograms (approx.), least-squares regression, …

Homomorphic Encryption



Problem: Homomorphic encryption based on asymmetric cryptography is expensive (e.g., Paillier, EC-ElGamal)

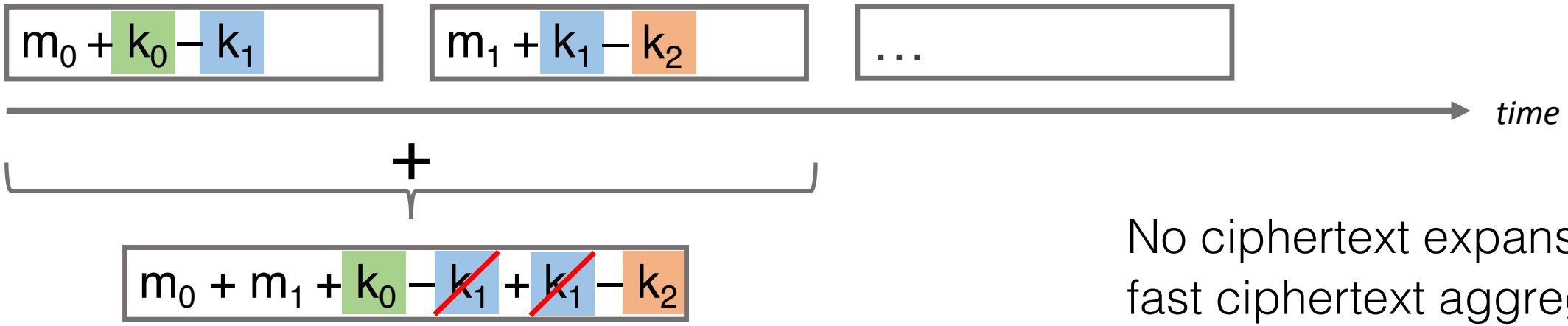


TimeCrypt Encryption

Given a key stream: $k_0, k_1, k_2, k_3, k_4, k_5, \ldots$

[Castelluccia et al. 05] *Symmetric homomorphic encryption*

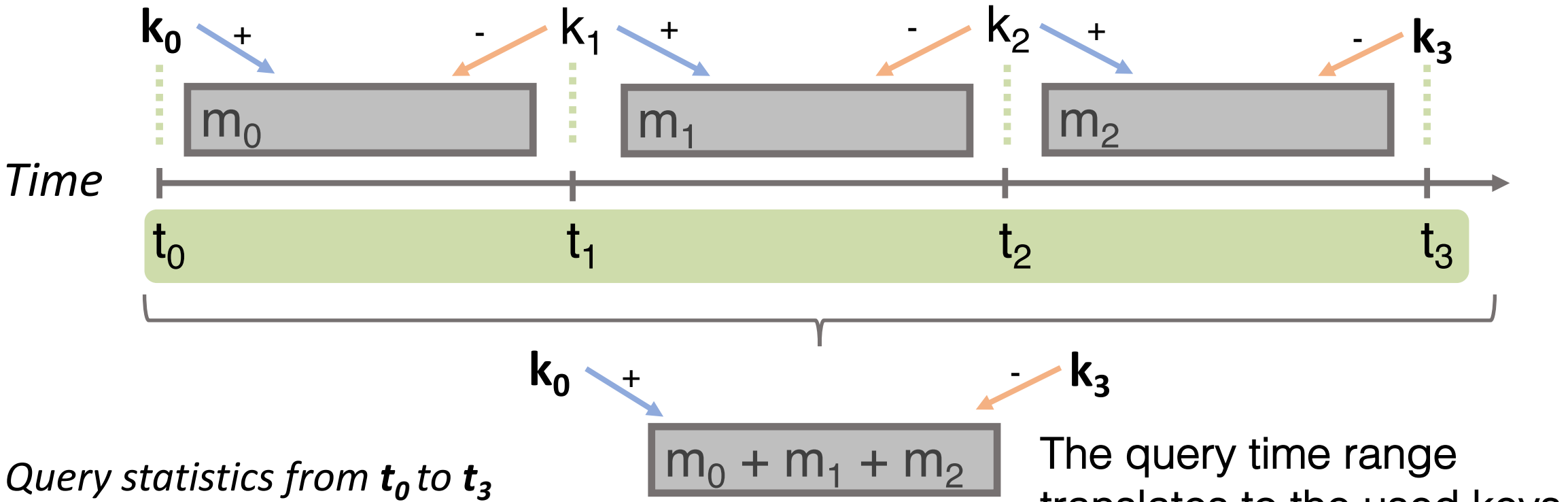
+/- is addition modulo M



No ciphertext expansion +
fast ciphertext aggregation

$$m_0 + \ldots + m_N + k_0 - k_1 + k_1 - \ldots + \ldots - k_{(N+1)}$$

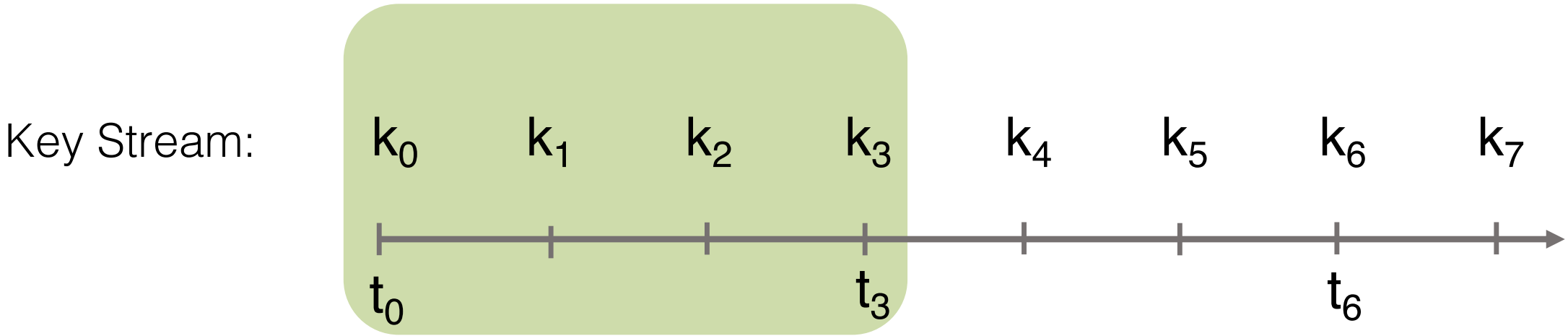
Key Stream to Time Encoding



Query statistics from t_0 *to* t_3

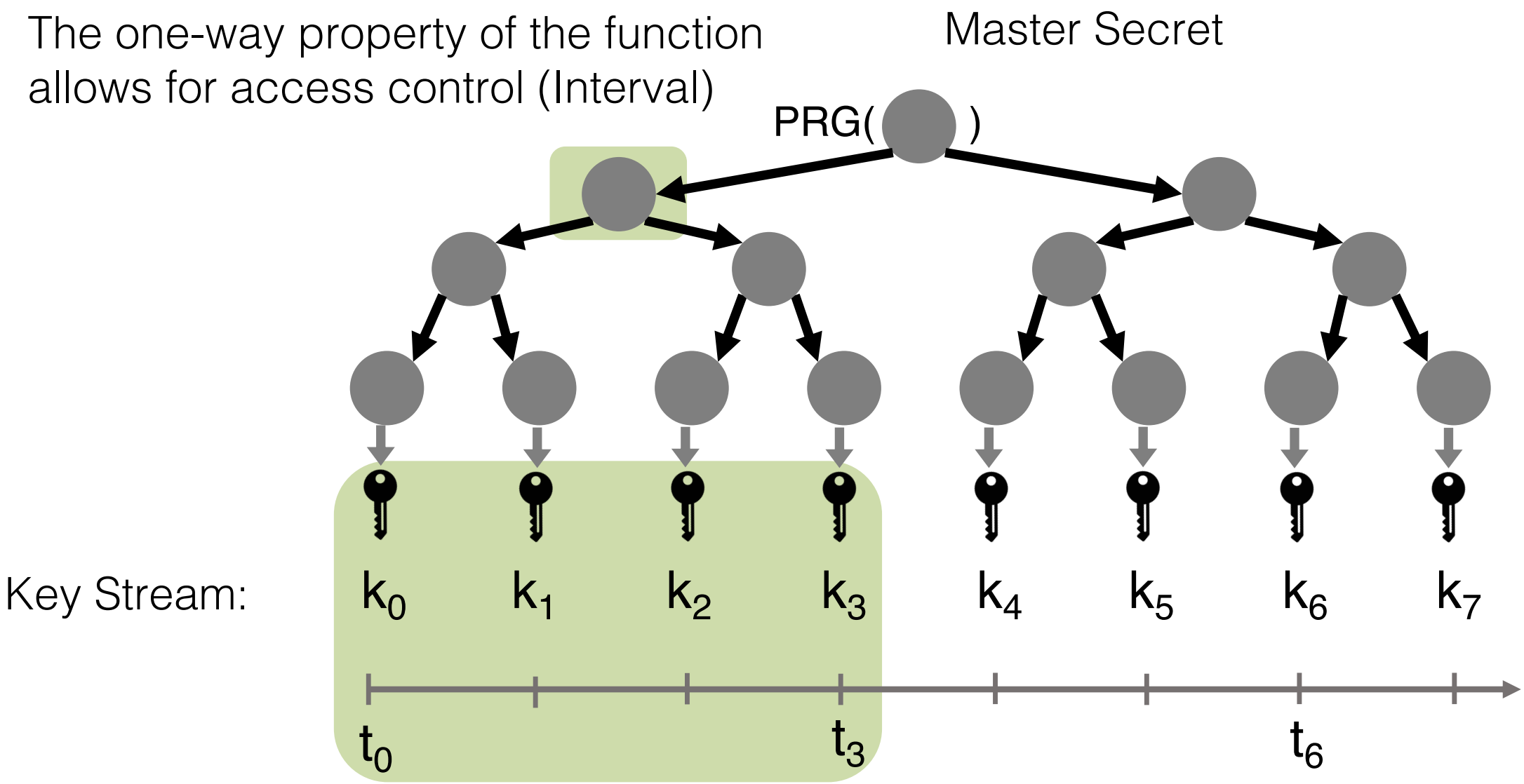
The query time range translates to the used keys

Time Interval Access Restriction



Tree-based Key Derivation

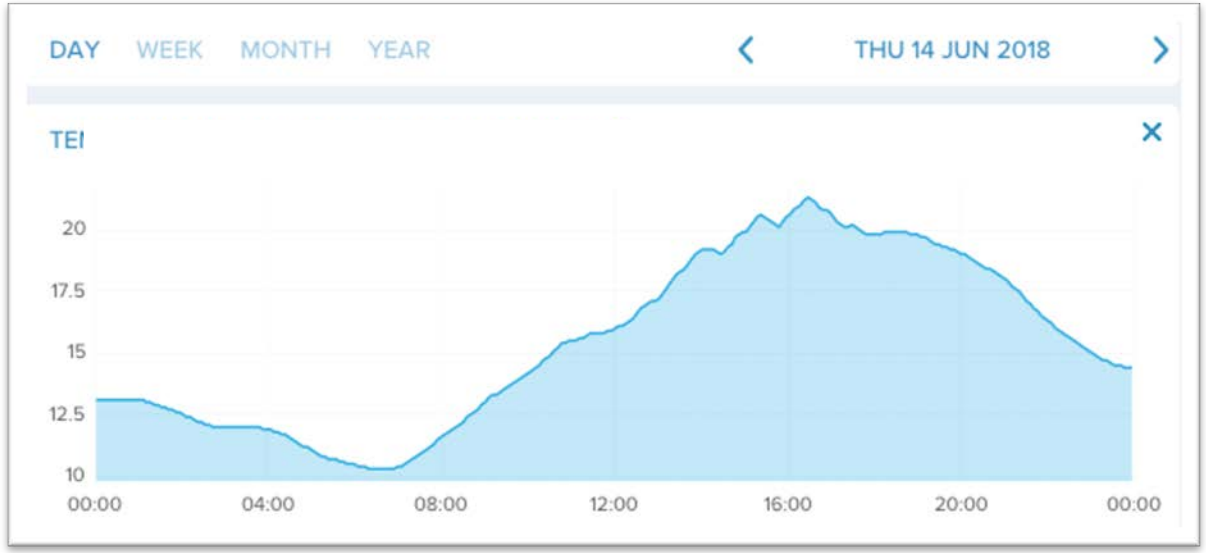
The one-way property of the function allows for access control (Interval)

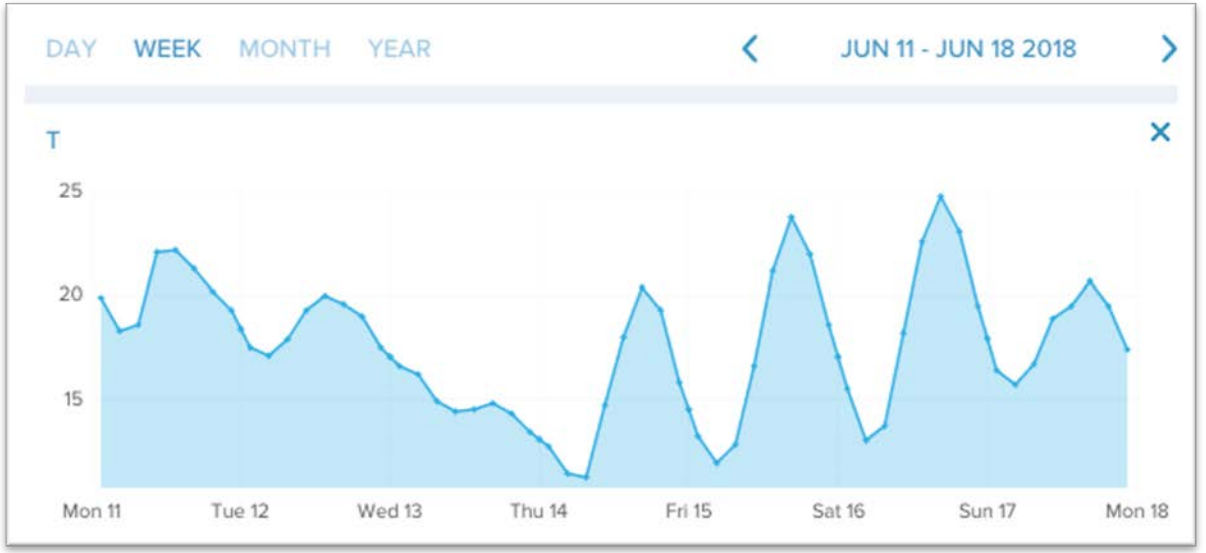


Access Restriction at Resolution Level

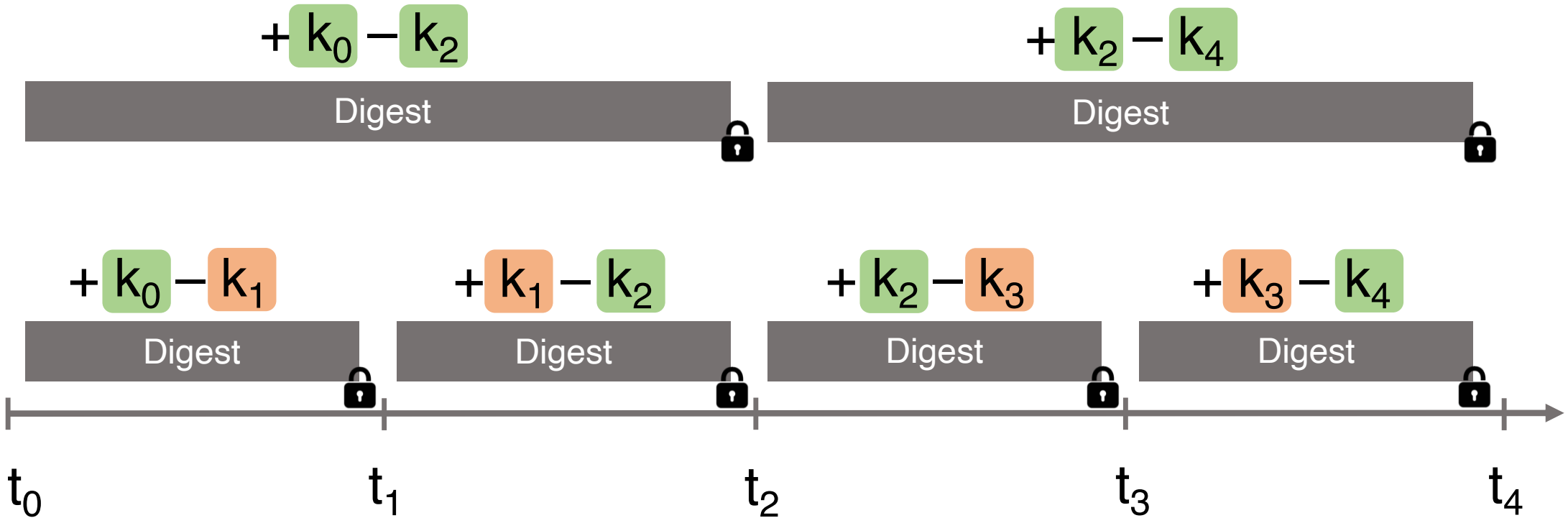
How to share aggregated information of a certain granularity?

Per hour aggregates vs Per day aggregates

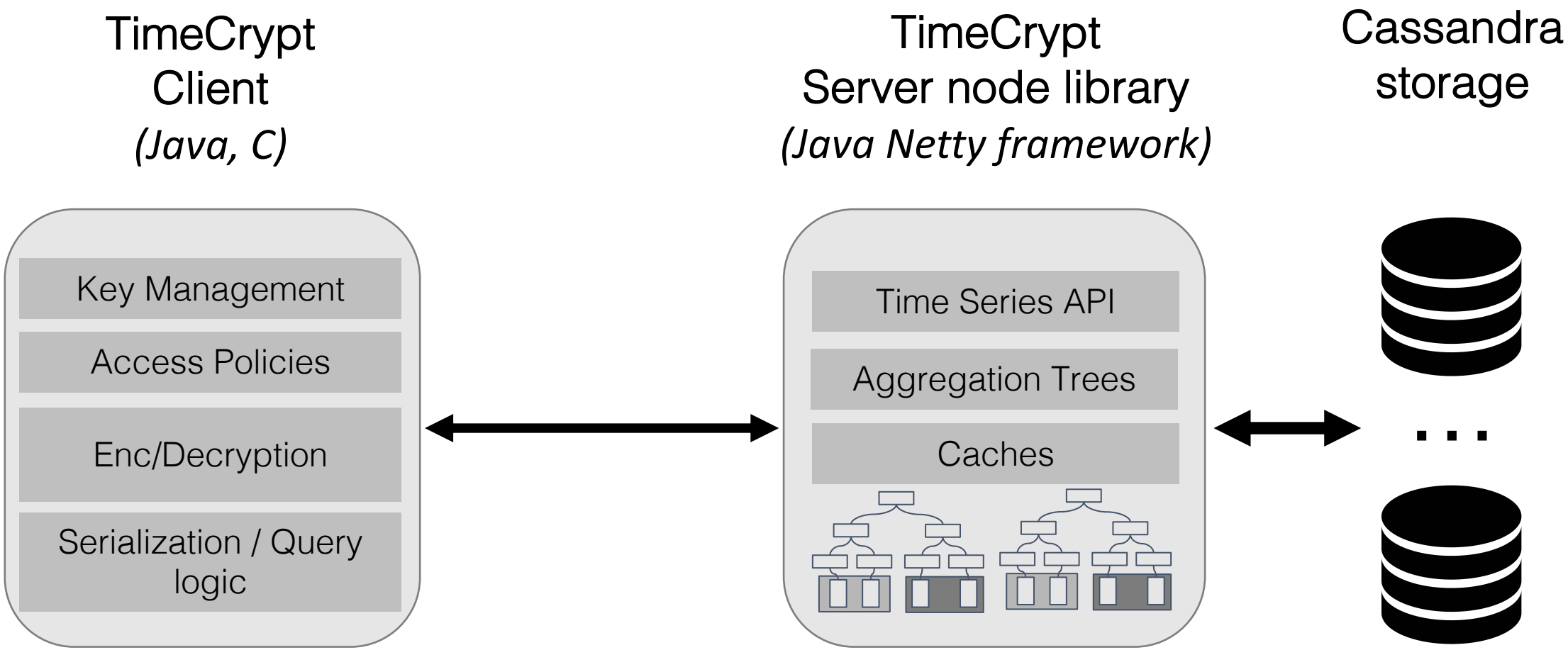




Only share the outer keys of the desired granularity $k_0, k_2, k_4, \ldots$



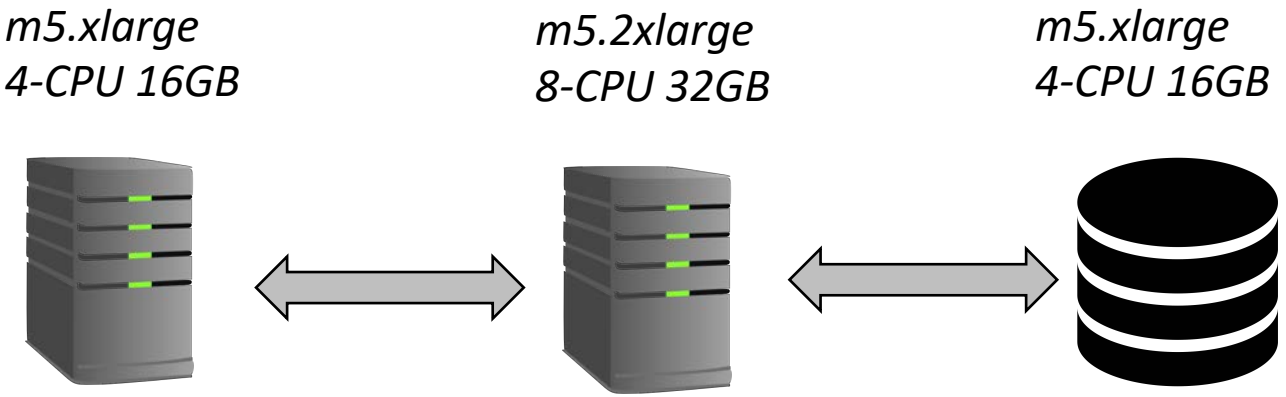
TimeCrypt Implementation



Evaluation

Health Dashboard Application

Medical Sensor Data

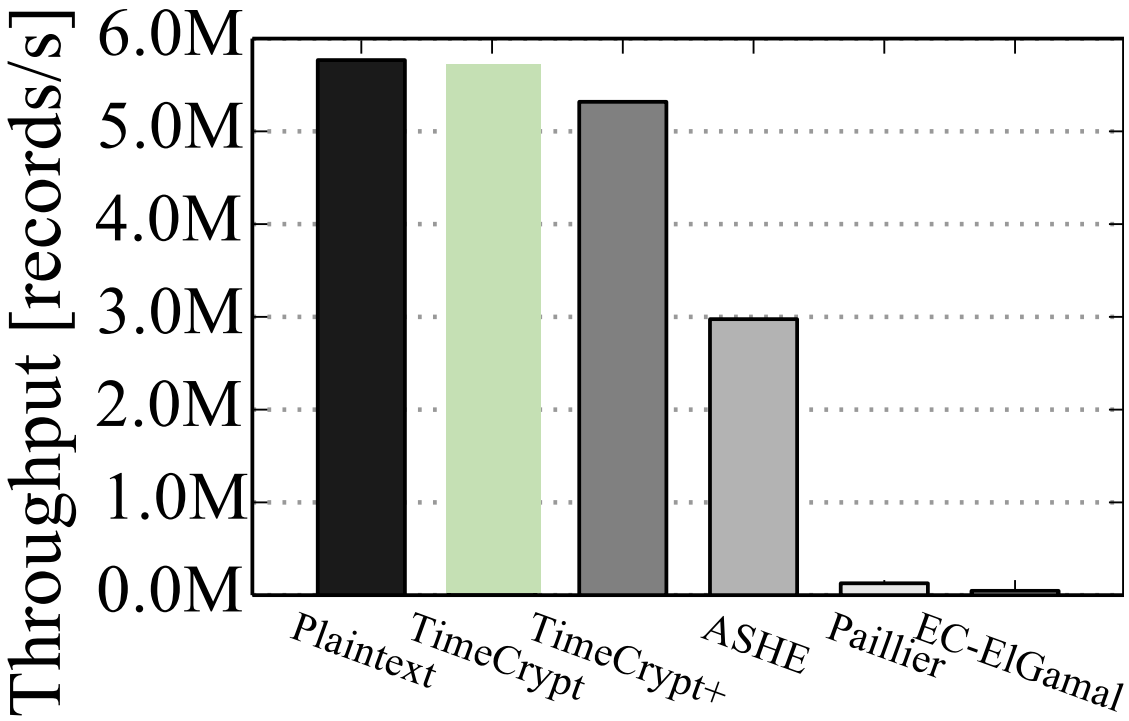


100 Clients

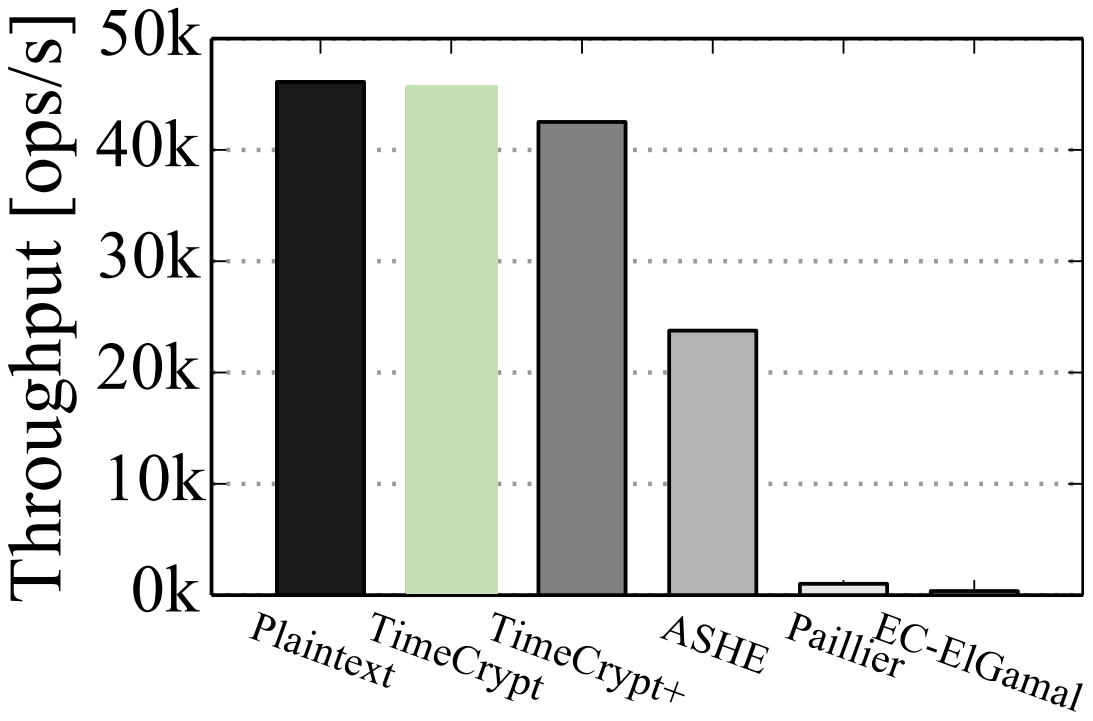
4 range queries per 1 chunk insert

50Hz data rate/stream, 10s chunks

System Performance



Ingest

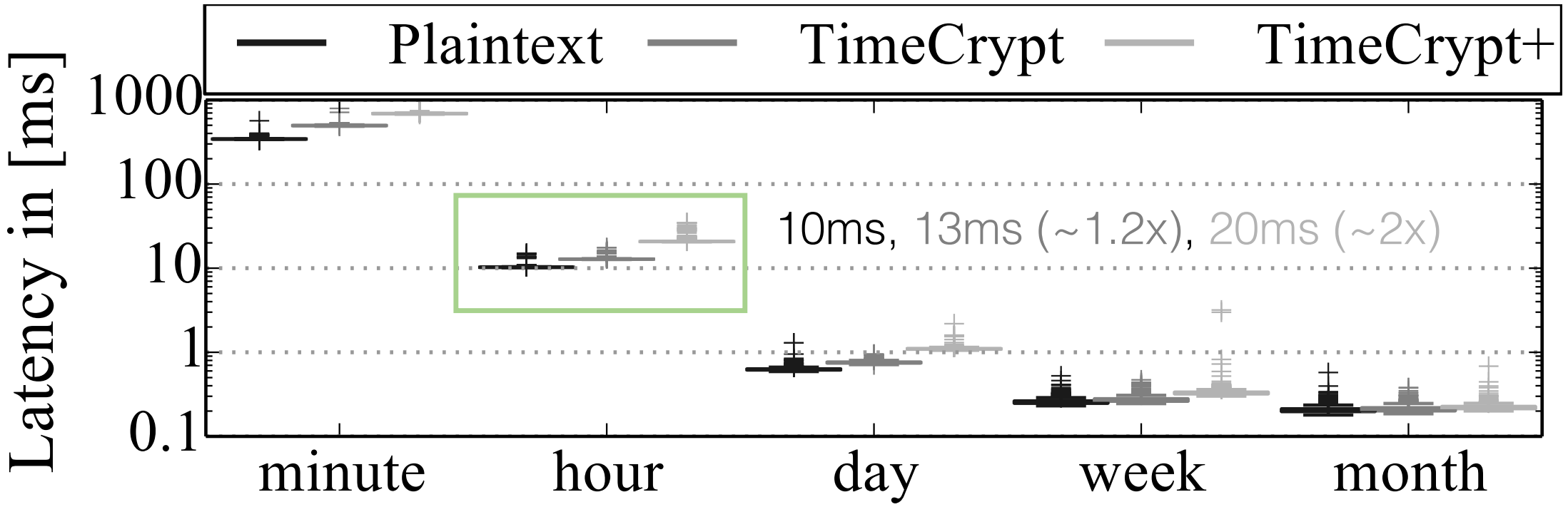


Statistical Range Queries

Throughput under heavy load of

4/1 read-write ratio, 49k streams

Health Dashboard Queries



Latency for statistical queries over one month, based on our health app

120M data records, 241920 chunks (1chunk/10s)

Summary

- TimeCrypt is an efficient system that augments time series datastores with encrypted data processing capabilities
  - Protects **confidentiality** of sensitive time series data
  - Supports computation integrity on encrypted data

- **TimeCrypt's Encryption:** Efficient construction that couples encrypted data processing with crypto-enforced access-control for time series streams

- Source code available at: https://timecrypt.io/