Proofs of Data Residency

Checking whether Your Cloud Files Have Been Relocated

Hung Dang, Erick Purwanto, Ee-Chien Chang
School of Computing
National University of Singapore

Data Geolocation

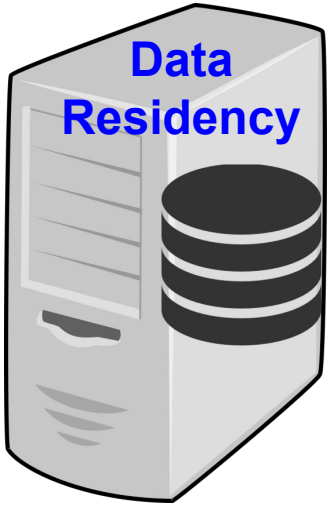
➢ A file ***F*** is ***stored at*** a particular location ***L***.

Is It Relevant?

- Various *legislations* and *directives* regulate possessing and storage of data *across national borders*.
  - Australian Privacy Act
  - EU Data Protection Directive
- *Fault tolerance* of storage system relies on *replicating* the data *across geographically* separated drives.

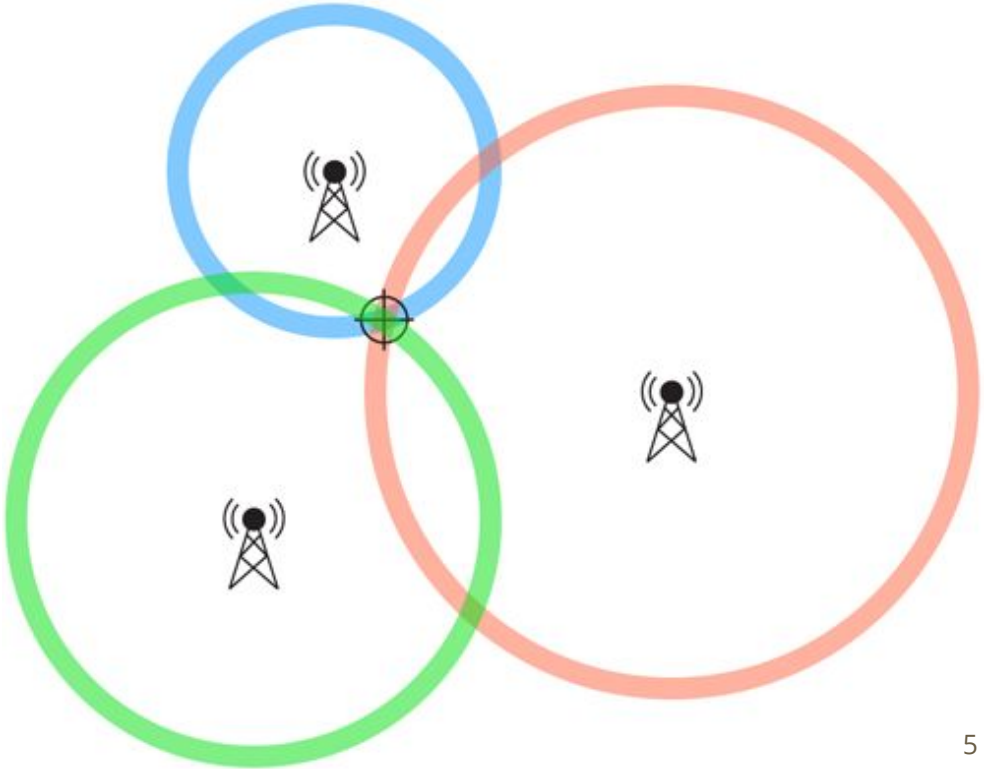
Data Geolocation - The Breakdown

- Check if the file *F* is stored at a particular location *L*.
- Check if the file *F* is **stored on** a server S which is located **at** a particular location *L*.



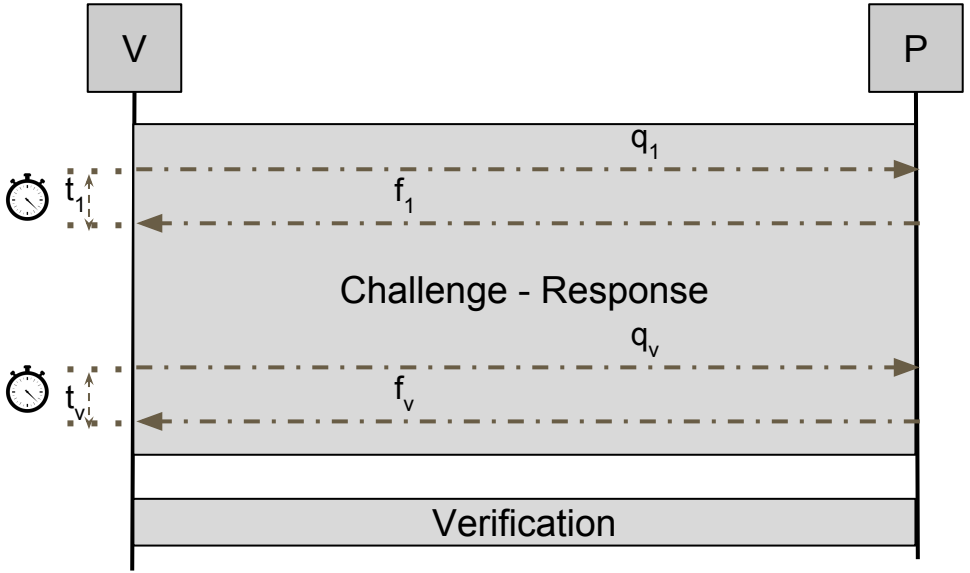
Host Geolocation

- Triangulation: Executing *distance bounding protocols* from various landmarks
- Key assumption: There exists a correlation between *distance* and *round-trip-time*



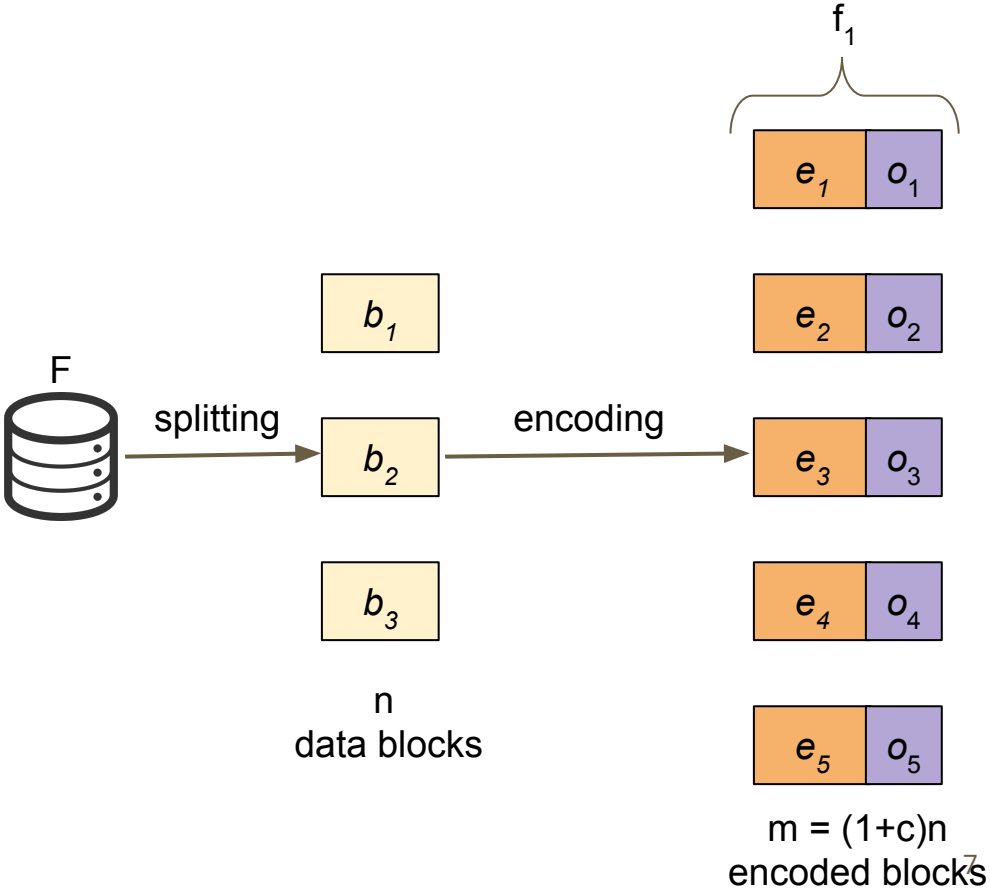
Data Residency

- *F* is stored on a server *S*.
- *F* is *retrievable from local drives* of the server *S*.
  - Retrievability attested with Proof of Retrievability.
  - Storage locality checked by timings the POR response latencies.



POR - A Brief Review

- F can be reconstructed from any *n* valid encoded blocks
  - Data lost if more than cn blocks *deleted* or *corrupted*
  - Each encoded block e_i is authenticated by a tag o_i
- Audit of size *v* detects data loss w.h.p $(1 - 1/(1+c)^v)$.



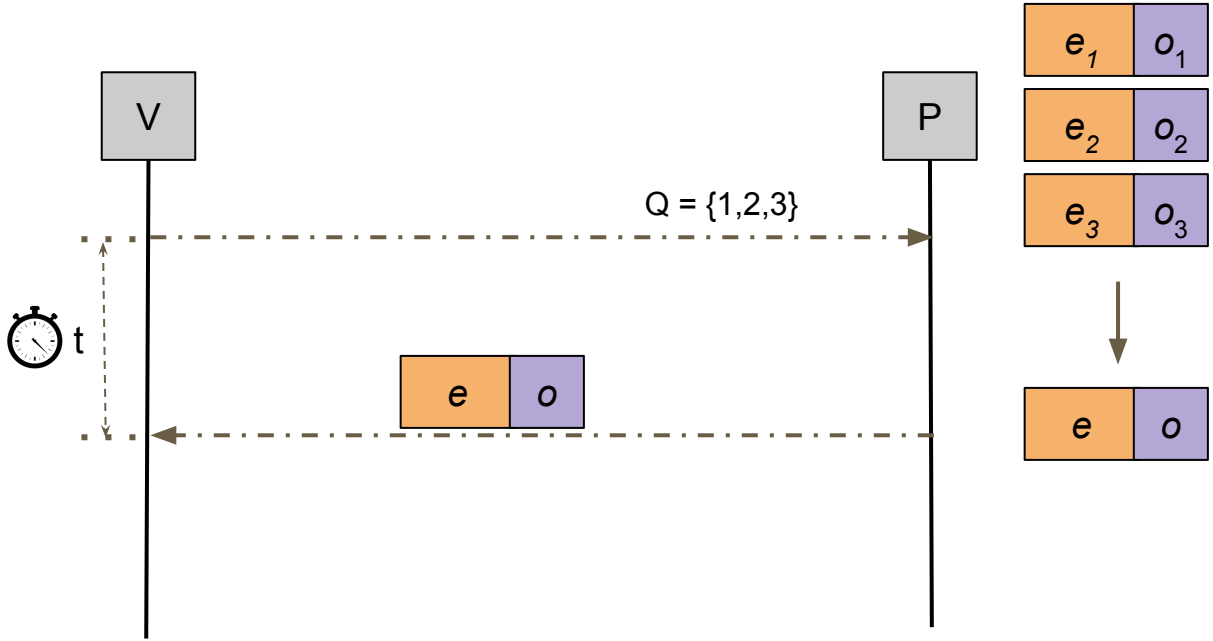
Why Timed POR Complicated?

The timing measurements consist of:

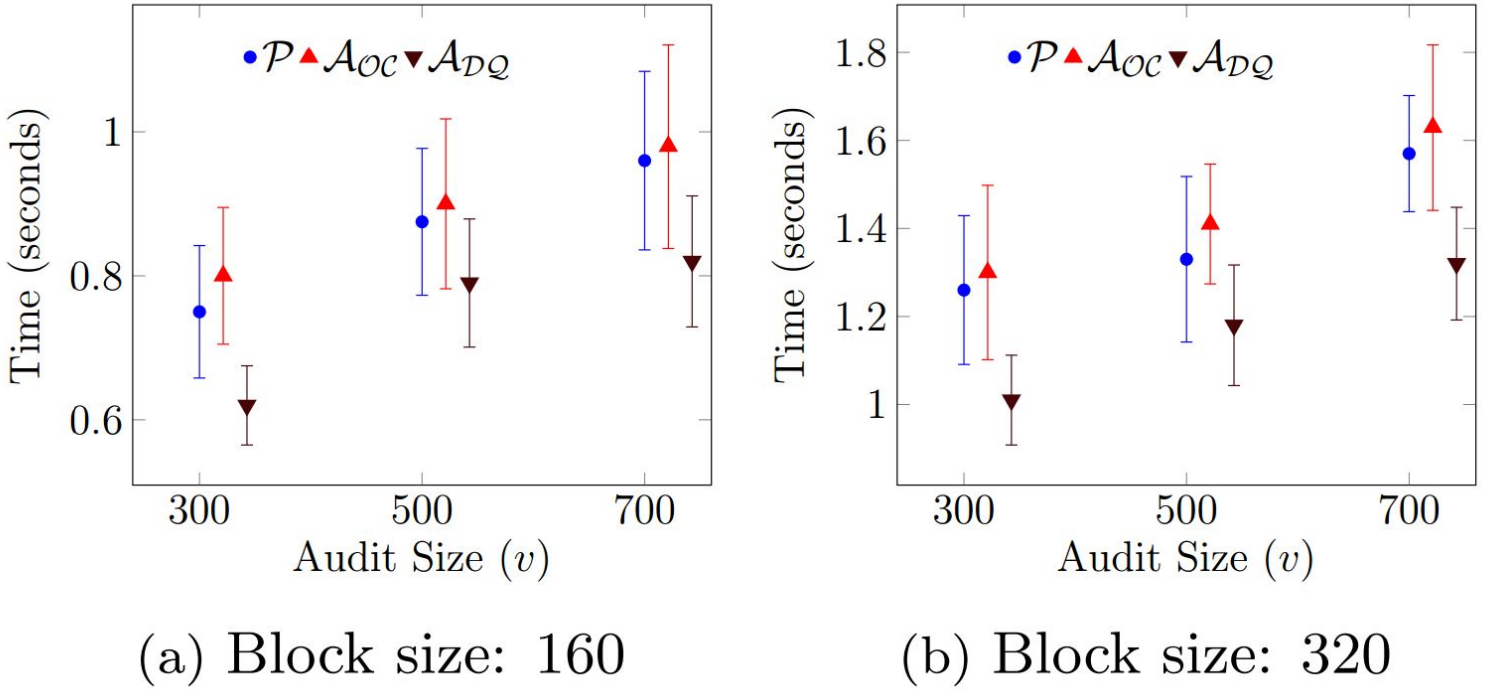
- Challenge-response transmission time — ***noisy***
- Fetching time — ***(a bit) noisy, malleable***
- Computation time — ***malleable***

Vulnerable Construction (timing SW-PoR)

- t consists of:
  - Transmission time
  - Fetching time
  - Computation time
- Accept if:
  - t < threshold
  - Response is valid

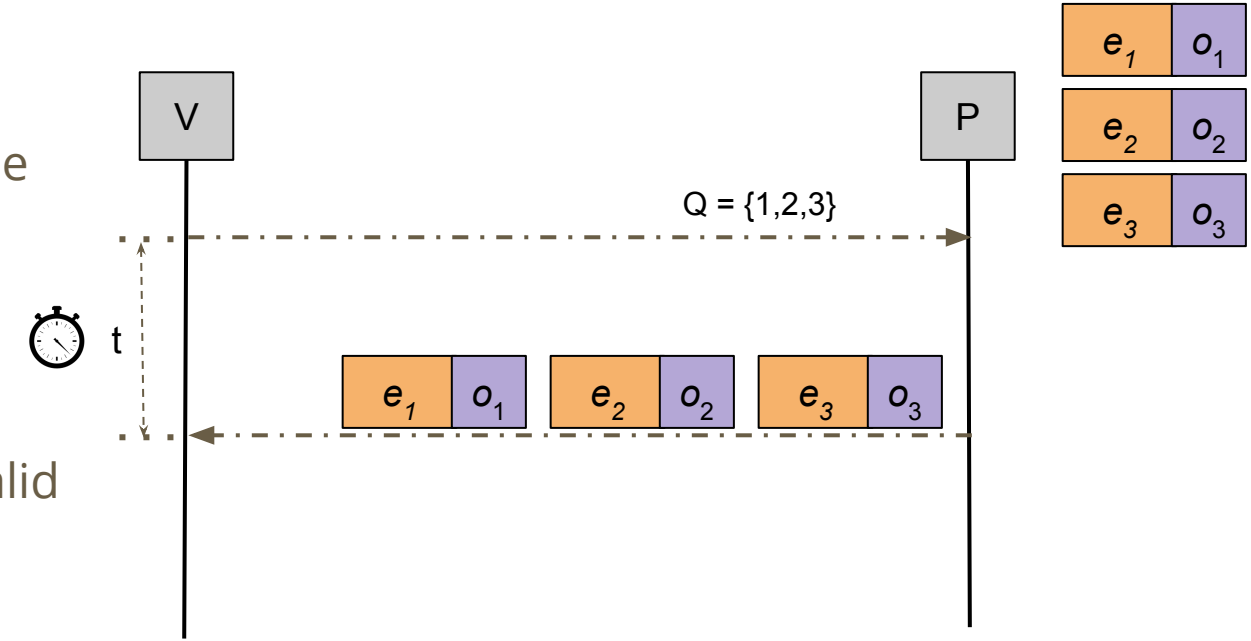


Vulnerable Construction (timing SW-PoR)

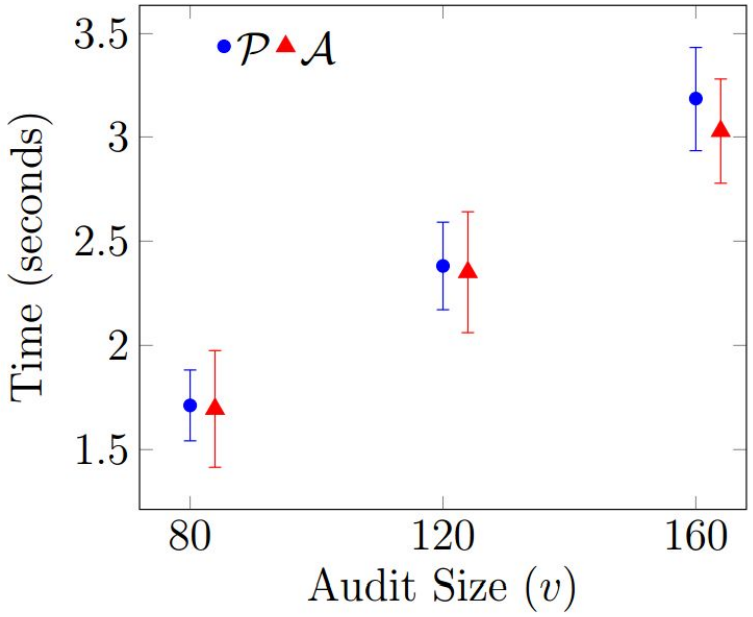


Vulnerable Construction (timing JK-PoR)

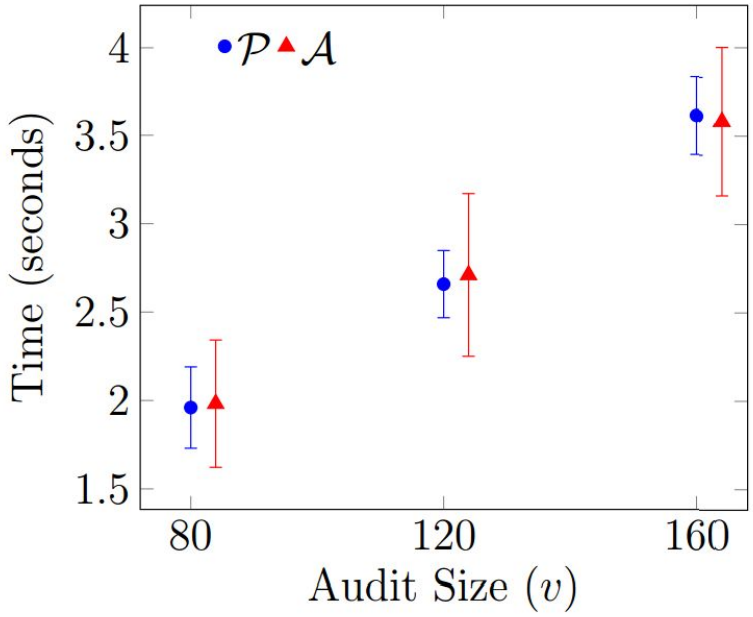
- t consists of:
  - Transmission time
  - Fetching time
- Accept if:
  - t < threshold
  - Responses are valid



Vulnerable Construction (timing JK-PoR)



(a) 512-byte block



(b) 1024-bytes block

And the attention goes to...

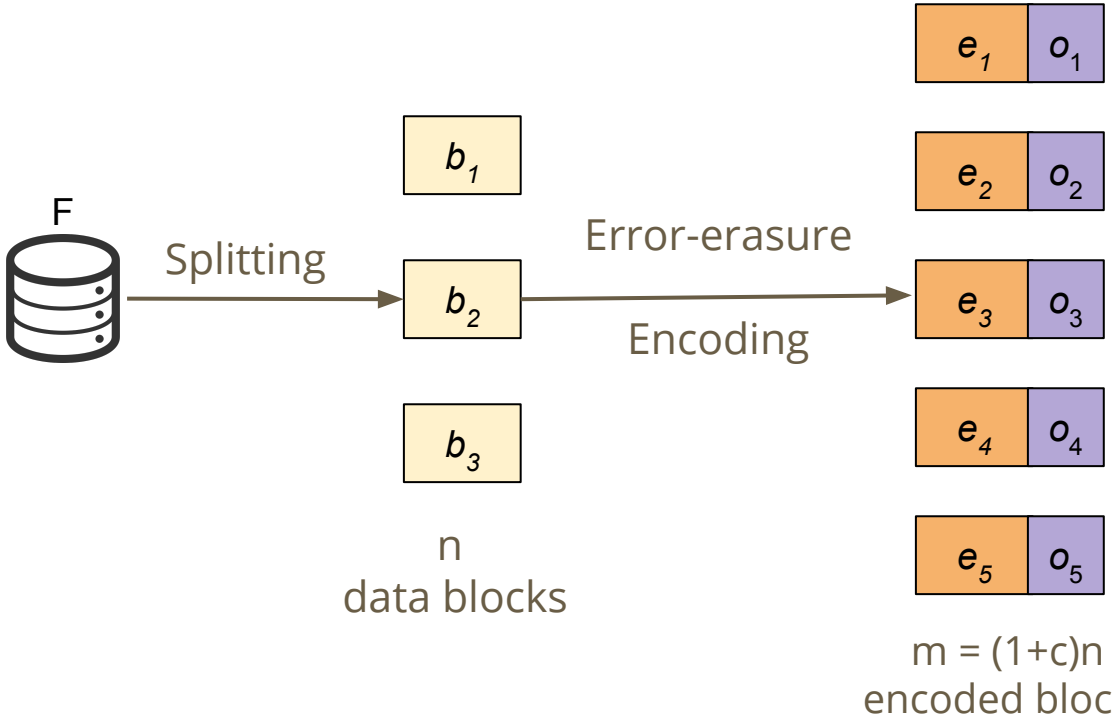
- Computation Time
  - *Eliminated*
- Fetching Time
  - *Minimal + Consistent*

Authenticator-based PoR

Atomic Operation

- Transmission Time
  - *Minimizing the noise*

The Construction - Setup Phase

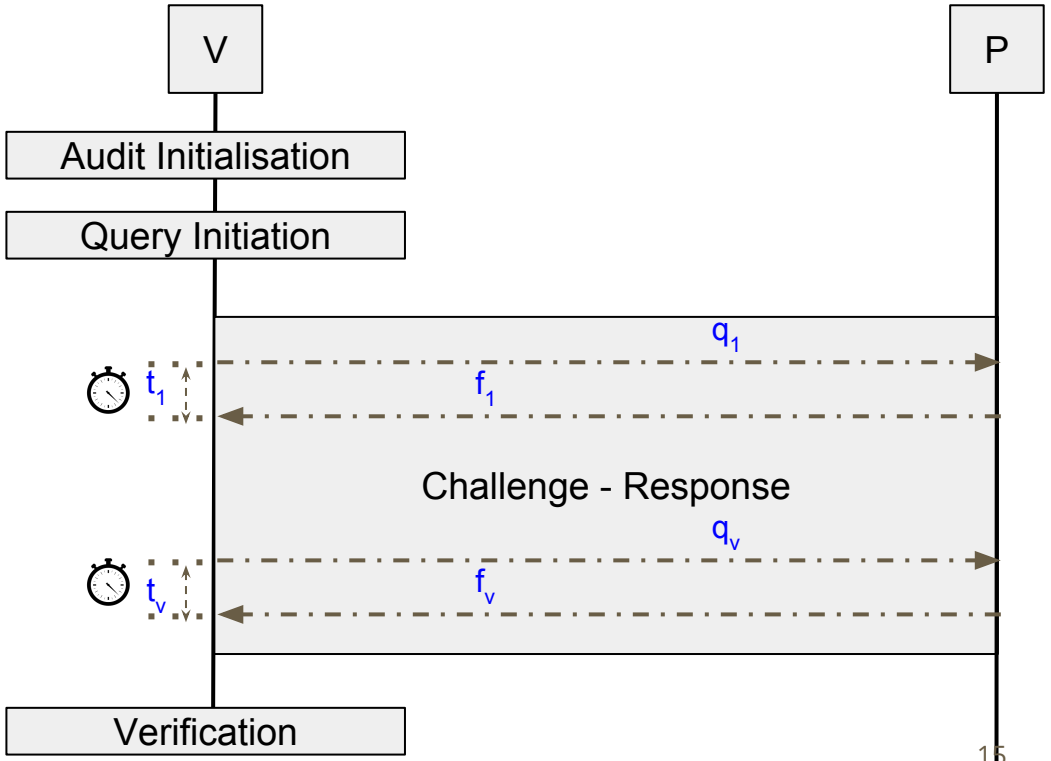


$o_i = HMAC(sk, F_{ID}||i||e_i)$

$f_i = o_i || e_i$

The Construction - Residency Checking

- Audit Initialisation
  - Audit size *v*
  - Latency threshold *d*
  - Late delivery threshold *l*
- Query Initiation
  - Pick *v* challenges at random
- Challenge-Response
  - Measures latency of *each* query and its response
- Verification*
  - Decision is made based on $\langle f_1,\ldots,f_v\rangle$, $\langle t_1,\ldots,t_v\rangle$, *d* and *l*

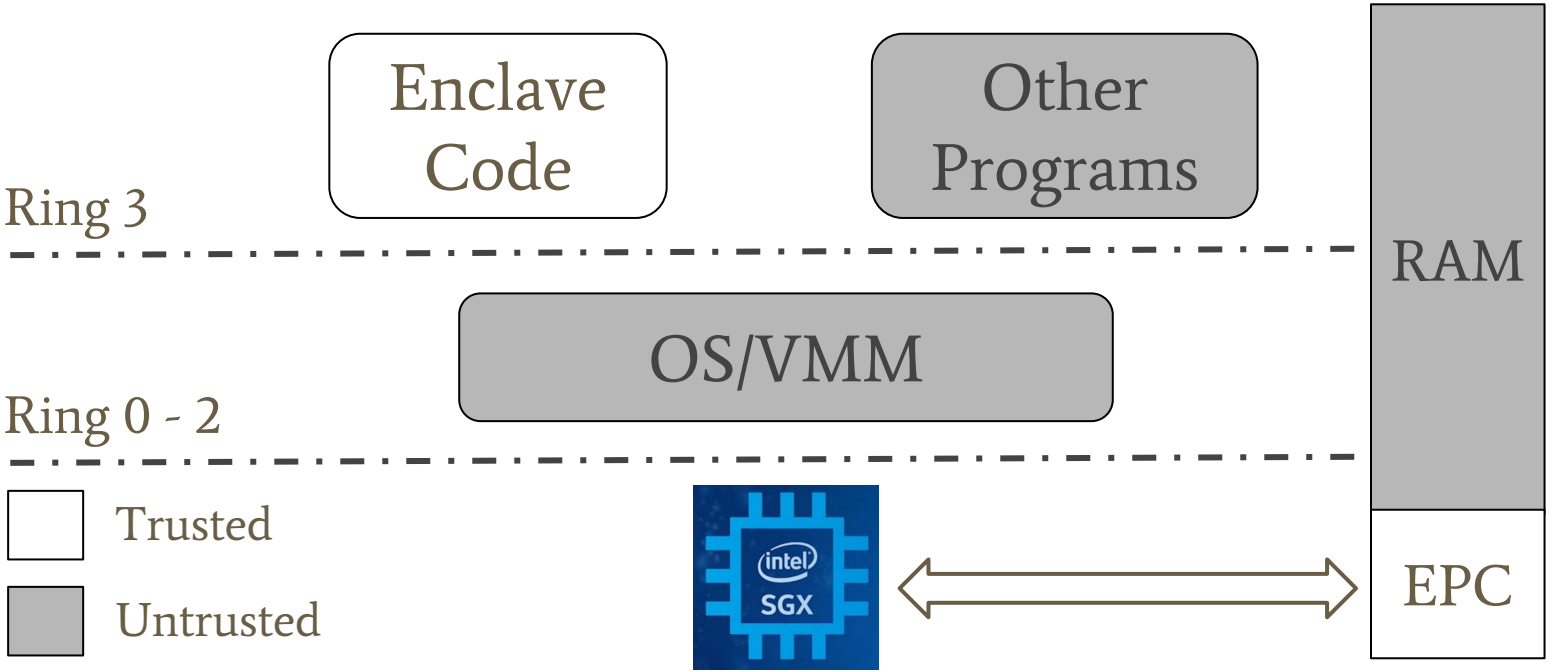


*Invalid response leads to immediate failure

Implementations

- N-ResCheck: Verifier and Prover communicate over the network
  - Employs TCP for transmission of challenges and responses
  - Subject to high level of noise
- E-ResCheck: Entrust a trusted unit on the storage server
  - Verifier and Prover co-locate on the same physical system, minimizing noisy factor (i.e., transmission time)
  - Implemented with SGX-enabled processor

SGX: Hardware-root of Trust

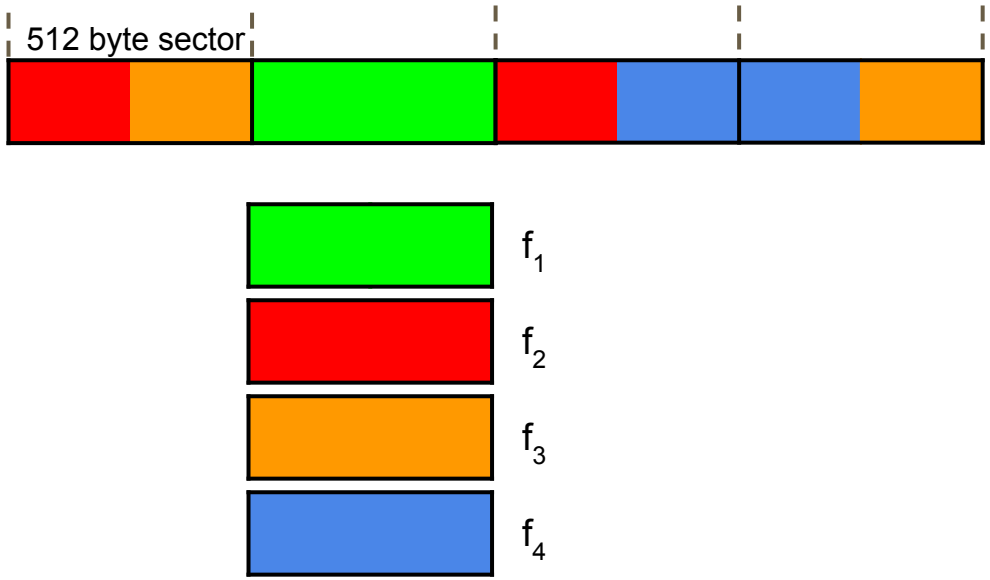


Atomic Operation - The Block Size

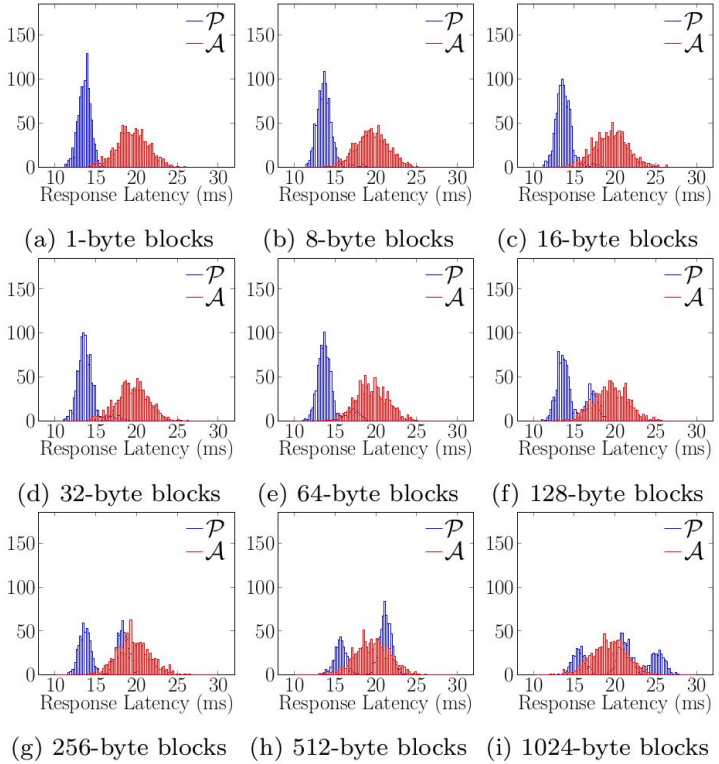
- A block may span across multiple non-contiguous sector

⇒ *high variance in fetching time*

- Small blocks *fitting entirely* in one disk sector (w.h.p) make timings more reliable.



Atomic Operation - The Block Size



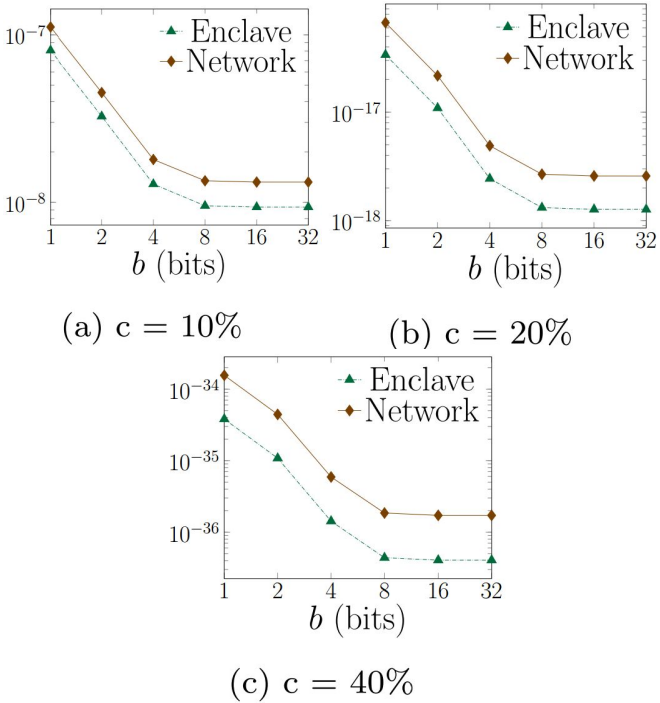
(a) 1-byte blocks (b) 8-byte blocks (c) 16-byte blocks

(d) 32-byte blocks (e) 64-byte blocks (f) 128-byte blocks

(g) 256-byte blocks (h) 512-byte blocks (i) 1024-byte blocks

Atomic Operation - The MAC Length

- Small blocks entail short authentication
- With limited access to verification oracle, short authentication tags do not compromise security
  - E.g., c = 40%, v = 300 and 16 bits MAC, probability of detection is [1 - 2^{-145}]



Related Work

Proofs of Retrievability

Timed Challenge-Response Protocols

Locality of Storage

Protected Execution Environment

Conclusion and Extension

- Residency of the data demands attention.
- PoDR provides a mechanism to establish data residency.
- Extensions
  - Finer granularity (e.g., different machines in the same data warehouse)
  - Dynamic PoDR - supporting data updates

Q & A

Hung Dang

hungdang@comp.nus.edu.sg

False Acceptance Rate

- Hit is the number of challenges certainly arrive “late”
- l is the late delivery threshold
- v is the audit size
- b is the authentication tag length

$$\Pr(Hit \leq l) + \sum_{x=1}^{v-l} \Pr(Hit = x + l) \cdot (2^{-bx} + \mu(\lambda))$$