HASLab
HIGH-ASSURANCE
SOFTWARE LABORATORY

14th ACM International Systems and Storage Conference (SYSTOR ’21)

S2Dedup: SGX-enabled Secure Deduplication

Mariana Miranda, Tânia Esteves, Bernardo Portela, João Paulo*
*INESC TEC & University of Minho, *NOVA LINCS & University of Porto*
Portugal

June 14–16, 2021

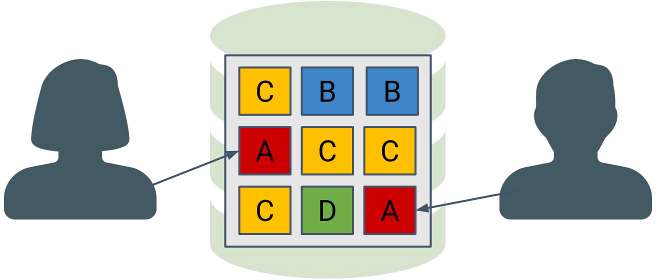
HASLab
HIGH-ASSURANCE
SOFTWARE LABORATORY

INESCTEC
TECNOLOGIA E CIÊNCIA
LABORATÓRIO ASSOCIADO

Universidade do Minho

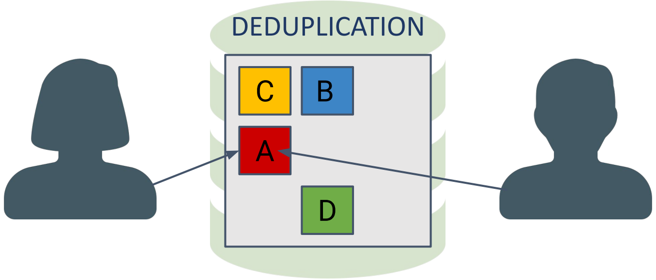
Context

- Exponential growth of digital information
- Identical data is being stored repeatedly
  - Deduplication analyzes stored data and eliminates redundant copies
- Security concerns when dealing with third-party storage services
  - Encrypt data before outsourcing it



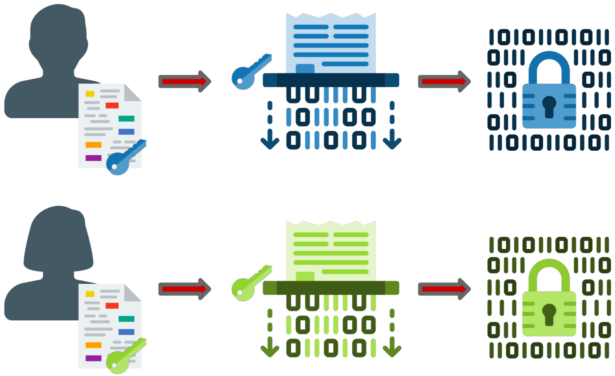
Context

- Exponential growth of digital information
- Identical data is being stored repeatedly
  - Deduplication analyzes stored data and eliminates redundant copies
- Security concerns when dealing with third-party storage services
  - Encrypt data before outsourcing it



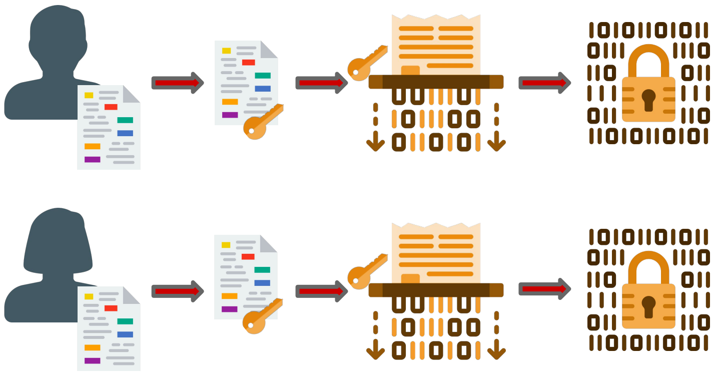
Problem

STANDARD ENCRYPTION



Same plaintexts → Different ciphertexts

CONVERGENT ENCRYPTION



Same plaintexts → Same ciphertexts

✗ NO DEDUPLICATION

✓ DEDUPLICATION

⚠ SECURITY ISSUES

Contributions

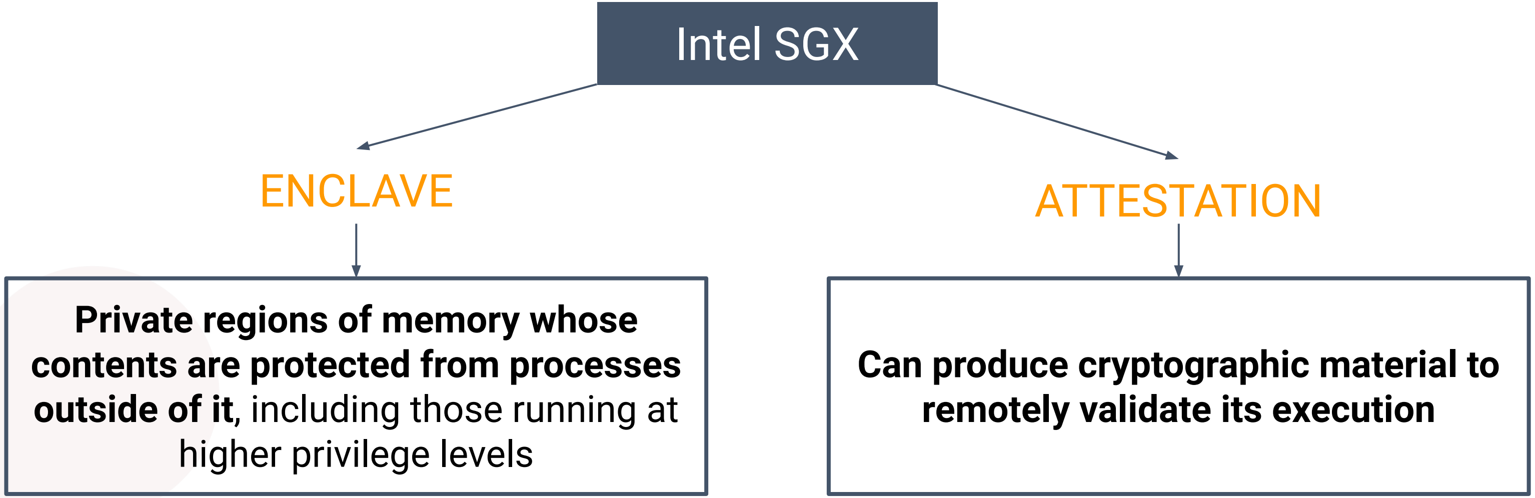
- Design and implementation of S2Dedup
  - An open-source[1] secure deduplication system that takes advantage of trusted hardware
- Integration of multiple secure deduplication schemes
- Epoch and exact frequency security scheme
  - Combines epochs and the idea of limiting the number of duplicates per chunk

[1] https://github.com/mmm97/S2Dedup

Background

Trusted Hardware

- Provides a secure trusted execution environment to perform critical operations in an untrusted software environment



Intel SGX

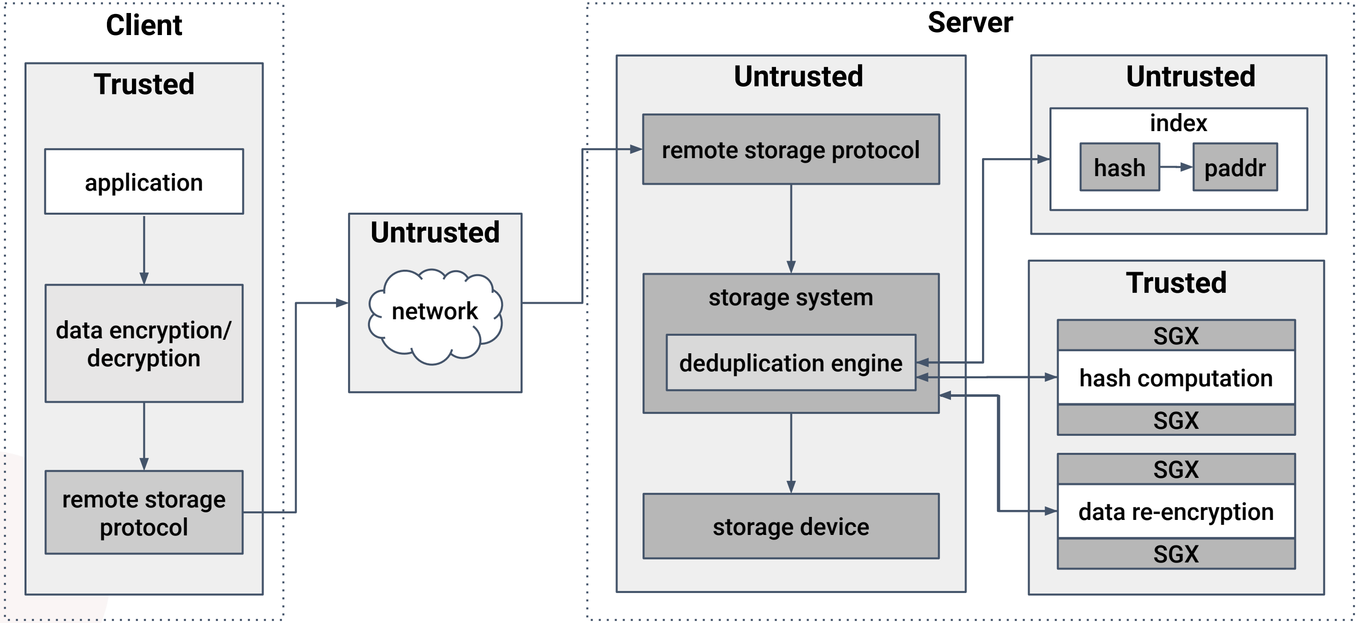
ENCLAVE

Private regions of memory whose contents are protected from processes outside of it, including those running at higher privilege levels

ATTESTATION

Can produce cryptographic material to remotely validate its execution

Architecture



Secure Solutions

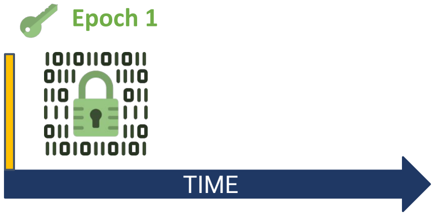
Plain Security

- Basic secure deduplication scheme
- Relies on the enclave to securely compute a block’s cryptographic hash and re-encrypt data
- Vulnerable to information leakage
    - For example, an attacker can still infer if deduplication occurred by checking if there are any changes in the deduplication metadata

Secure Solutions

Epoch Based[1]

- Performs deduplication in epochs
- Enables deduplication only for copies from the same epoch
- Epochs establish a temporal barrier
- Trade-off between security and space savings
- Still susceptible to in-epoch leakage

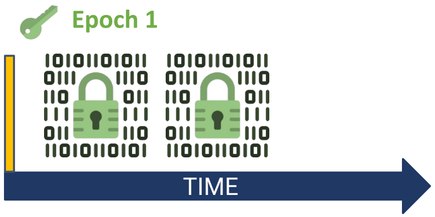


[1] Based on Hung Dang, Tien Tuan Anh Dinh, Ee-Chien Chang, and Beng Chin Ooi. Privacy preserving computation with trusted computing via scramble-then-compute. Proceedings on Privacy Enhancing Technologies, 2017(3):21–38, 2017.

Secure Solutions

Epoch Based[1]

- Performs deduplication in epochs
- Enables deduplication only for copies from the same epoch
- Epochs establish a temporal barrier
- Trade-off between security and space savings
- Still susceptible to in-epoch leakage

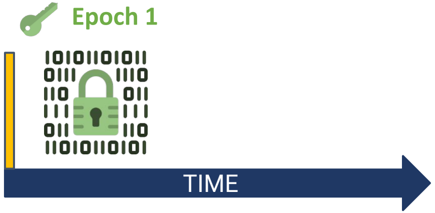


[1] Based on Hung Dang, Tien Tuan Anh Dinh, Ee-Chien Chang, and Beng Chin Ooi. Privacy preserving computation with trusted computing via scramble-then-compute. Proceedings on Privacy Enhancing Technologies, 2017(3):21–38, 2017.

Secure Solutions

Epoch Based[1]

- Performs deduplication in epochs
- Enables deduplication only for copies from the same epoch
- Epochs establish a temporal barrier
- Trade-off between security and space savings
- Still susceptible to in-epoch leakage

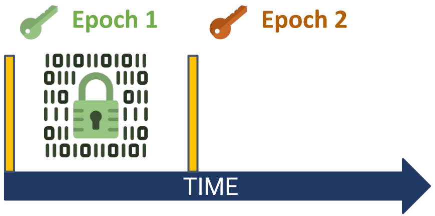


[1] Based on Hung Dang, Tien Tuan Anh Dinh, Ee-Chien Chang, and Beng Chin Ooi. Privacy preserving computation with trusted computing via scramble-then-compute. Proceedings on Privacy Enhancing Technologies, 2017(3):21–38, 2017.

Secure Solutions

Epoch Based[1]

- Performs deduplication in epochs
- Enables deduplication only for copies from the same epoch
- Epochs establish a temporal barrier
- Trade-off between security and space savings
- Still susceptible to in-epoch leakage

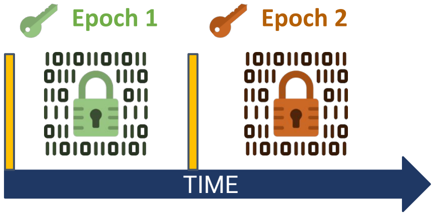


[1] Based on Hung Dang, Tien Tuan Anh Dinh, Ee-Chien Chang, and Beng Chin Ooi. Privacy preserving computation with trusted computing via scramble-then-compute. Proceedings on Privacy Enhancing Technologies, 2017(3):21–38, 2017.

Secure Solutions

Epoch Based[1]

- Performs deduplication in epochs
- Enables deduplication only for copies from the same epoch
- Epochs establish a temporal barrier
- Trade-off between security and space savings
- Still susceptible to in-epoch leakage

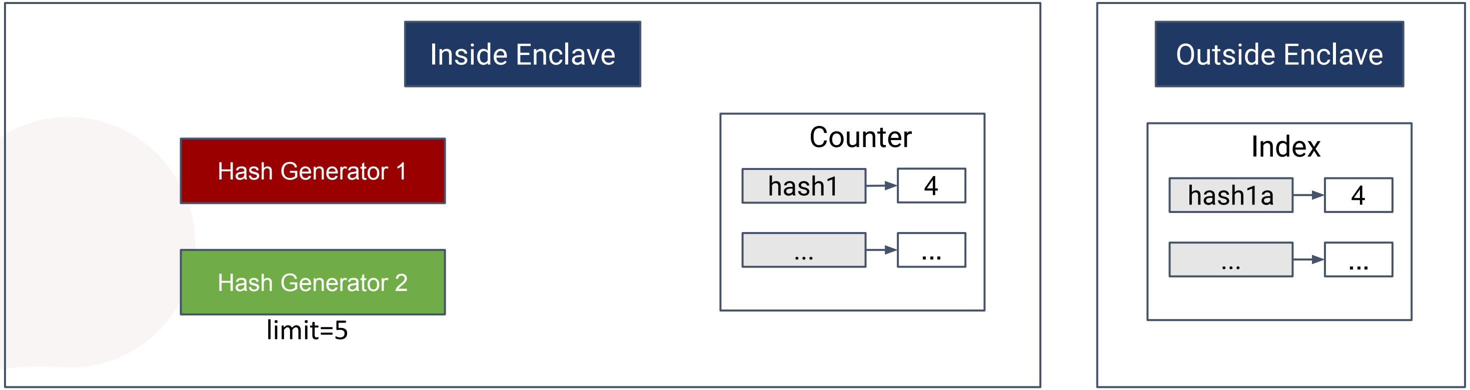


[1] Based on Hung Dang, Tien Tuan Anh Dinh, Ee-Chien Chang, and Beng Chin Ooi. Privacy preserving computation with trusted computing via scramble-then-compute. Proceedings on Privacy Enhancing Technologies, 2017(3):21–38, 2017.

Secure Solutions

Epoch and Exact Frequency Based

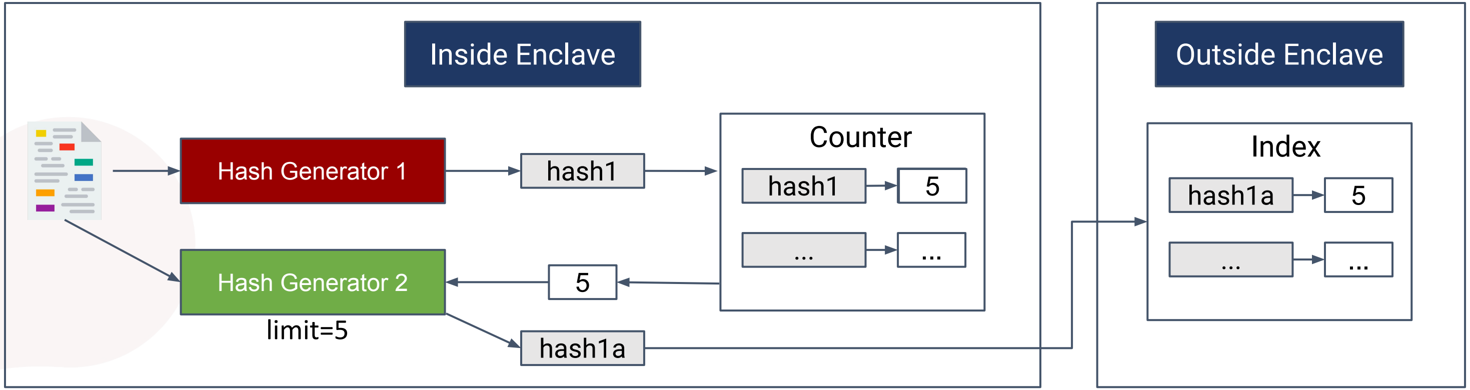
- Novel secure scheme
- Uses epochs and limits the number of duplicates per chunk
- Maintains an exact counter inside the enclave



Secure Solutions

Epoch and Exact Frequency Based

- Novel secure scheme
- Uses epochs and limits the number of duplicates per chunk
- Maintains an exact counter inside the enclave

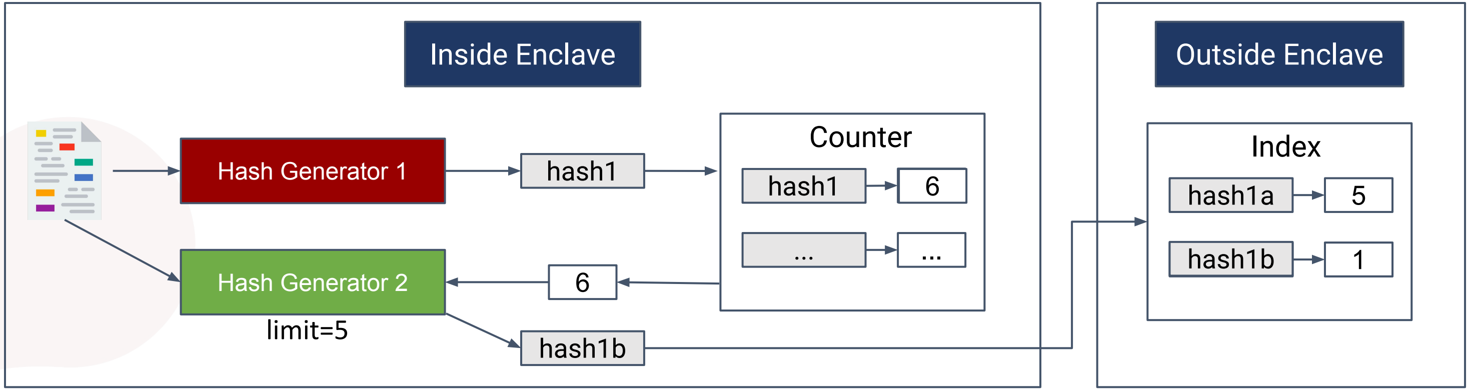


HASLab HIGH-ASSURANCE SOFTWARE LABORATORY

Secure Solutions

Epoch and Exact Frequency Based

- Novel secure scheme
- Uses epochs and limits the number of duplicates per chunk
- Maintains an exact counter inside the enclave



Secure Solutions

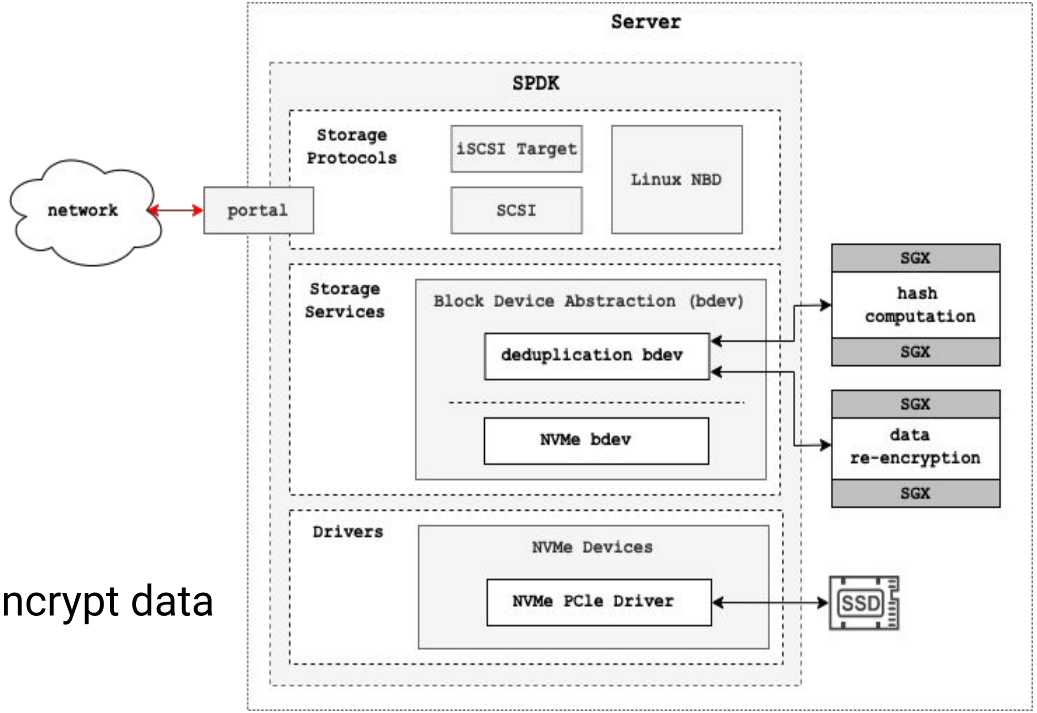
Estimated Frequency Based[1]

- Limits the maximum number of duplicates per block
- Relies on an estimated counter - *Count-Min Sketch*
- Can lead to deduplication loss
- Still discloses blocks with a lower number of duplicates

[1] Adapted from TED, Jingwei Li, Zuoru Yang, Yanjing Ren, Patrick PC Lee, and Xiaosong Zhang. Balancing storage efficiency and data confidentiality with tunable encrypted deduplication. In Proceedings of the Fifteenth European Conference on Computer Systems, pages *1–15, 2020*.

Prototype

- SGX
  - Trusted hardware platform
- SPDK
  - Server and client framework
- AES-XTS
  - Block cipher mode used to encrypt data



Evaluation

- How do different levels of security affect the system?

- 300 hours of experiments
  - Synthetic experiments[1]
    - 2 distributions: *dist_highperf and dist_kernels*
  - Realistic experiments[2]
    - 3 datasets: *mail, homes and web-vm*

[1] DEDISbench benchmarking tool

[2] Collected for three weeks at the Florida International University (FIU) Computer Science department

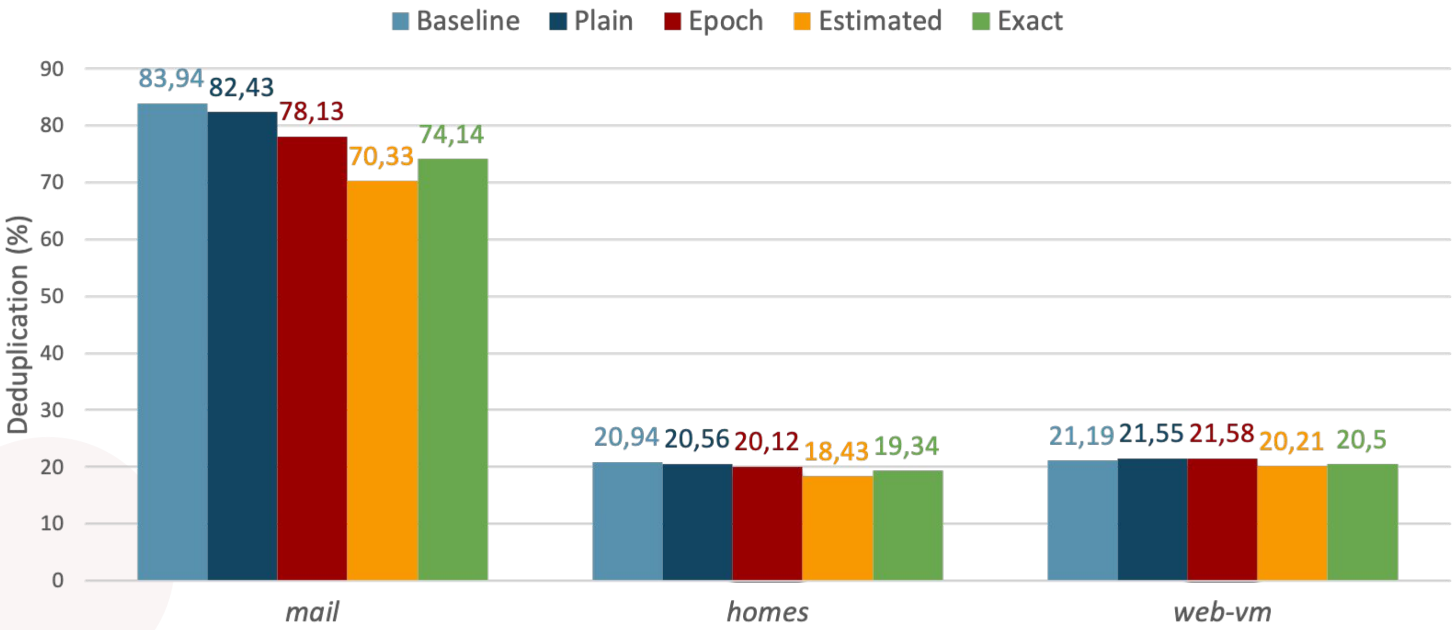
Evaluation

Realistic persistent writes - Throughput

		Throughput (MiB/s)				
		Baseline	Plain	Epoch	Estimated	Exact
mail	*S1*	1.2	1.2	1.2	1.2	1.2
	S200	180.56	166.44	168.61	148.05	149.05
	S400	234.93	168.17	169.77	151.51	150.34

Evaluation

Realistic persistent writes - Space Savings



Conclusions

- An open-source[1] secure deduplication system based on trusted hardware
- Support for multiple secure deduplication schemes
- Novel secure deduplication scheme
    - More robust security guarantees
    - Identical/improved deduplication effectiveness to the Estimated Frequency scheme
    - Identical performance to Estimated Frequency scheme

[1] https://github.com/mmm97/S2Dedup

Future work

- Further prevent from disclosing duplicate blocks by applying ORAM
- Expand S2Dedup for other types of deduplication
    - Offline deduplication
- Compare S2Dedup's schemes with other state-of-the-art solutions

HASLab
HIGH-ASSURANCE
SOFTWARE LABORATORY

14th ACM International Systems and Storage Conference (SYSTOR ’21)

S2Dedup: SGX-enabled Secure Deduplication

Mariana Miranda, Tânia Esteves, Bernardo Portela, João Paulo*
*INESC TEC & University of Minho, *NOVA LINCS & University of Porto*
Portugal

June 14–16, 2021

HASLab
HIGH-ASSURANCE
SOFTWARE LABORATORY

INESCTEC
TECNOLOGIA E CIÊNCIA
LABORATÓRIO ASSOCIADO

Universidade do Minho