基于可信执行环境的高性能加密重复数据删除研究

硕士学位论文答辩

匿名

电子科技大学计算机科学与工程学院（网络空间安全学院）

2022 年 5 月 13 日

外包数据存储

- Outsourcing data management to cloud is common in practice
  - 22% business data are stored in the cloud[*]
- Outsourcing storage should fulfill security and storage efficiency
  - Security: protect outsourced data against unauthorized access
  - Storage efficiency: reduce storage footprints

如何使用块

块的名称

- A
- B

如何使用定义、定理、引理、证明

定义 1 (定义名称)

定义内容

引理 1 (引理名称)

引理内容

定理 1 (定理名称)

定理内容（这里的定义、引理、定理分章节自动标号）

证明.

证明内容 □

```
\begin{table}[htbp]
  \caption{编号与含义}
  \label{tab:number}
  \centering
  \begin{tabular}{cl}
    \toprule
    编号 & 含义 \\
    \midrule
    1 & 4.0 \\
    2 & 3.7 \\
    \bottomrule
  \end{tabular}
\end{table}
公式~(\ref{eq:vsphere}) 的
编号与含义请参见
表~\ref{tab:number}。
```

表 1: 编号与含义

编号	含义
1	4.0
2	3.7

公式 (??) 的编号与含义请参见表 1。

测试环境与数据集

测试数据集

表 2: 真实世界数据集的特征

数据集	快照总数	去重前总数据量	重复数据删除系数
FSL	795	56.2 TiB	140.4
MS	143	14.4 TiB	6.0
Linux	84	44.9 GiB	1.3
CouchDB	83	22.9 GiB	1.5

测试平台

- 本地集群 (LAN)。万兆局域网内多台 Intel SGX 设备，每台设备均采用 Intel Core i7-10700 CPU，4 TB SATA 机械硬盘，32 GB DDR4 内存
- 云环境 (Cloud)。在两个不同区域的阿里云部署了多台规格为 *ecs.g7t.3xlarge* 的虚拟机 (VM) 来分别运行云服务端、密钥服务器和多个客户端。

密钥生成效率

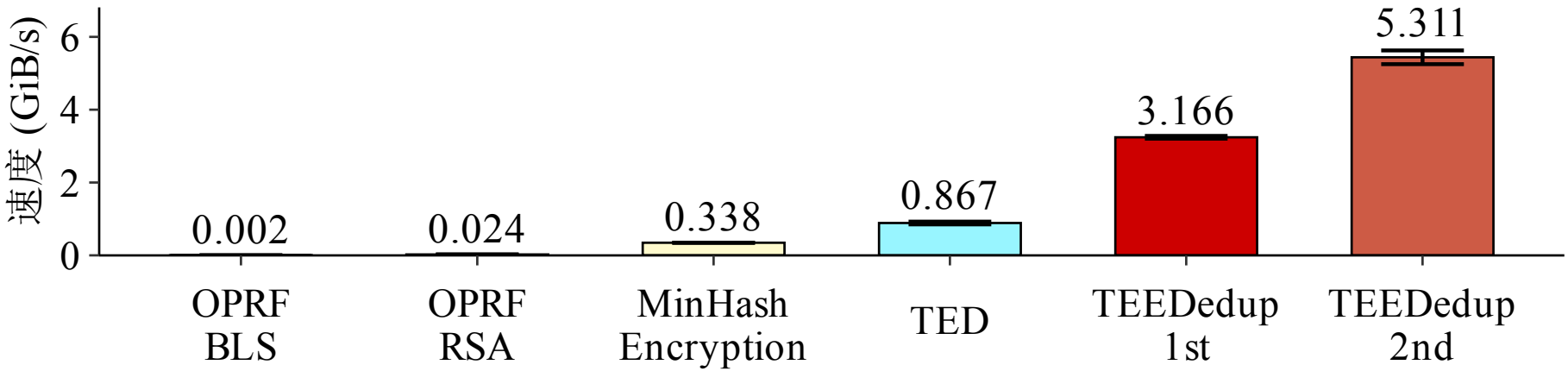


图 1: 单客户端密钥生成性能对比

- TEEDedup 提出的密钥生成方案相比 OPRF-BLS 和 OPRF-RSA 在第一轮（不含推测性加密）中实现了 1,583 倍和 131.9 倍的性能提升，而第二轮基于推测性加密，TEEDedup 将第一轮的密钥生成速度再次提高了 67.8%

数据块所有权证明效率

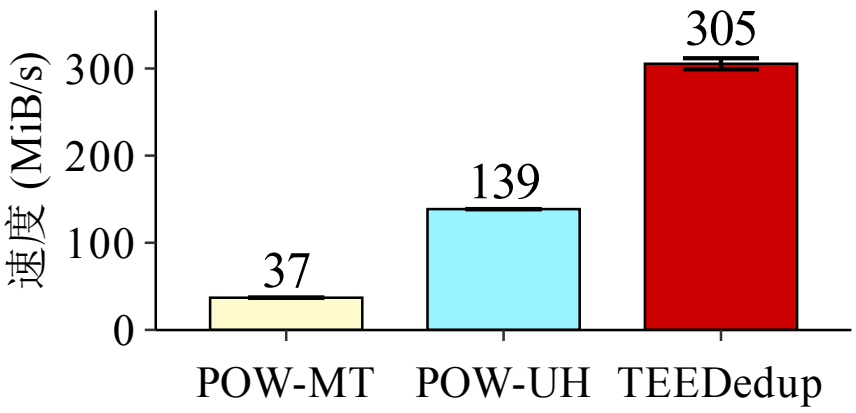


图 2: 数据所有权证明的计算性能 (不含网络开销)

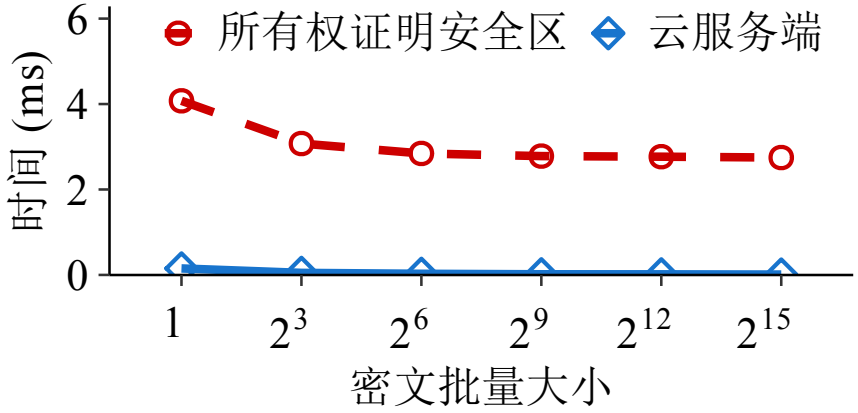


图 3: 所有权证明的计算开销 vs. 批量大小 (单位:ms/MiB)

- 由于 TEEDedup 避免了客户端中的纠删码编码和 Merkle 树构造，实现了相较于 PoW-MT 8.2 倍的性能提升。相较于安全性较弱的 PoW-UH 实现了 2.2 倍的性能提升。
- 云服务端的计算时间开销很低 (低于 0.05 ms)，而所有权证明安全区的计算时间随着批量大小增大从 4.1 ms 减少到 2.7 ms。

TEEDedup 原型系统性能

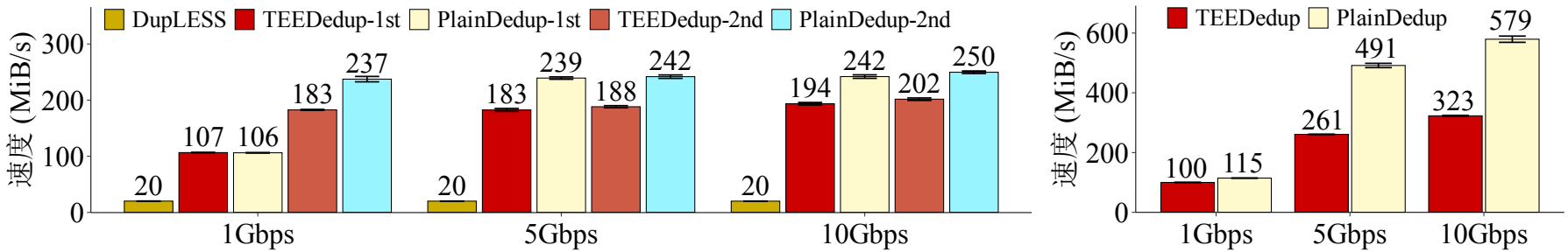


图 4: 单客户端在不同网络速度下的上传/下载性能

- TEEDedup 在第一轮和第二轮上传中分别相对 DupLESS 实现了 8.1 倍和 9.6 倍的性能提升。
- 与不安全的 PlainDedup 相比，TEEDedup 仅导致两轮上传速度分别下降约 17.5% 和 21.4%。

密文数据块的相似性检测

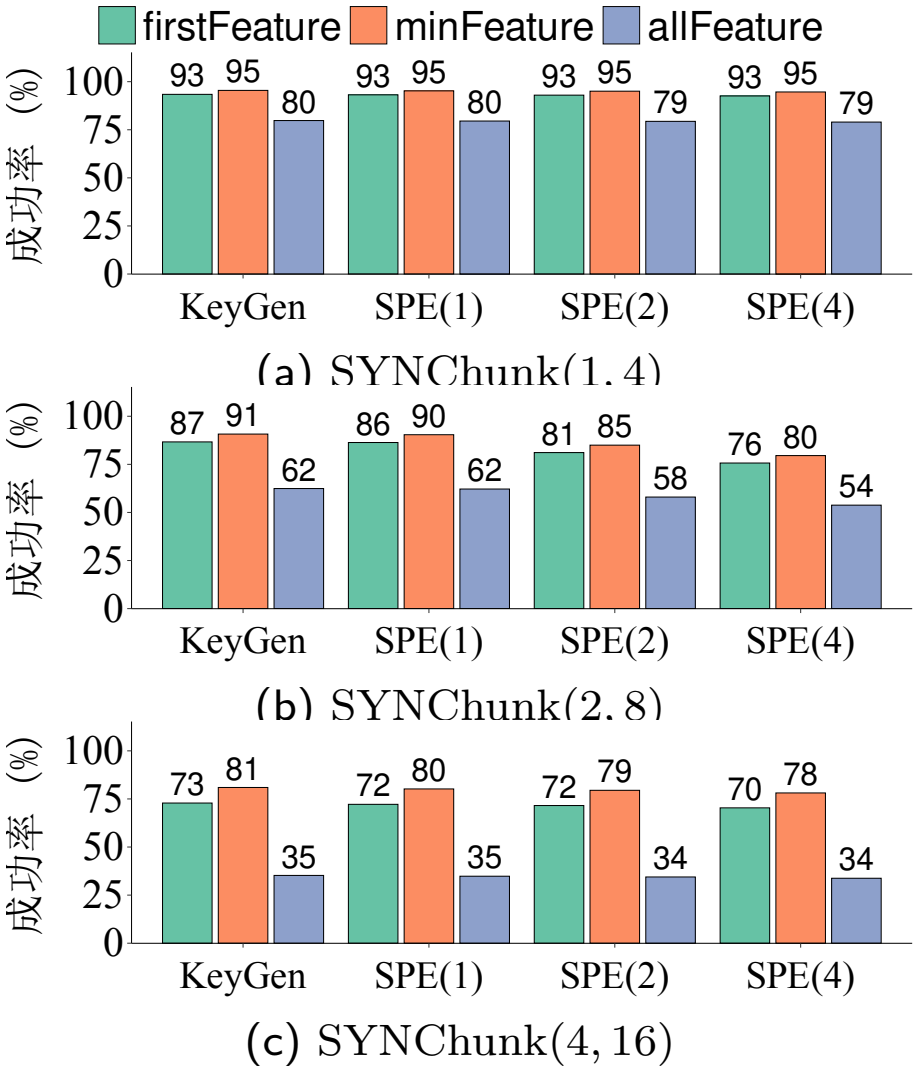


图 5: 密文数据块的相似性检测效果

- minFeature 方案对数据块内容的随机变化更不敏感。
- 相似性保留加密在加密后保留了较高的相似性: 通过检查密文数据块中的相似性指标，在 SYNChunk$(1, 4)$、SYNChunk$(2, 8)$、SYNChunk$(4, 16)$ 数据集中分别检测到至多 95.2%、90.4%、80.2% 的相似数据块。

推测内容攻击检测效果案例分析

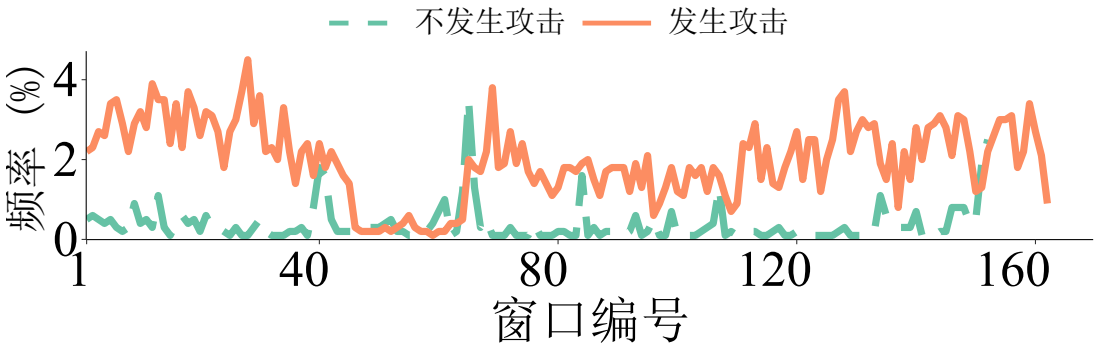
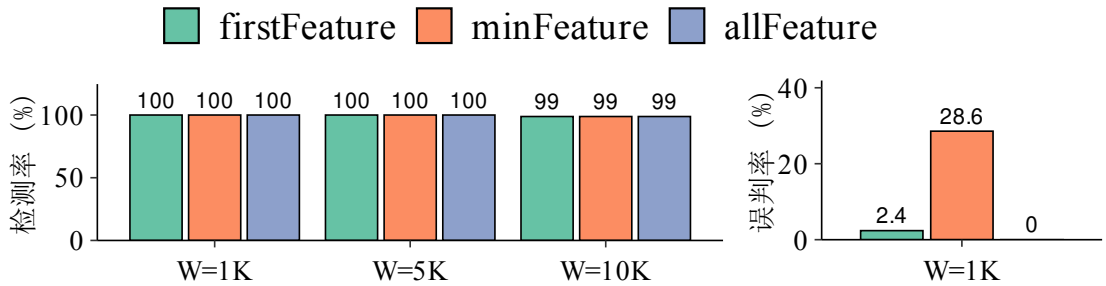


图 6: Linux 最新快照在包含/不包含攻击行为时每个窗口 (W=1 K) 的相似性指标的频率特征。

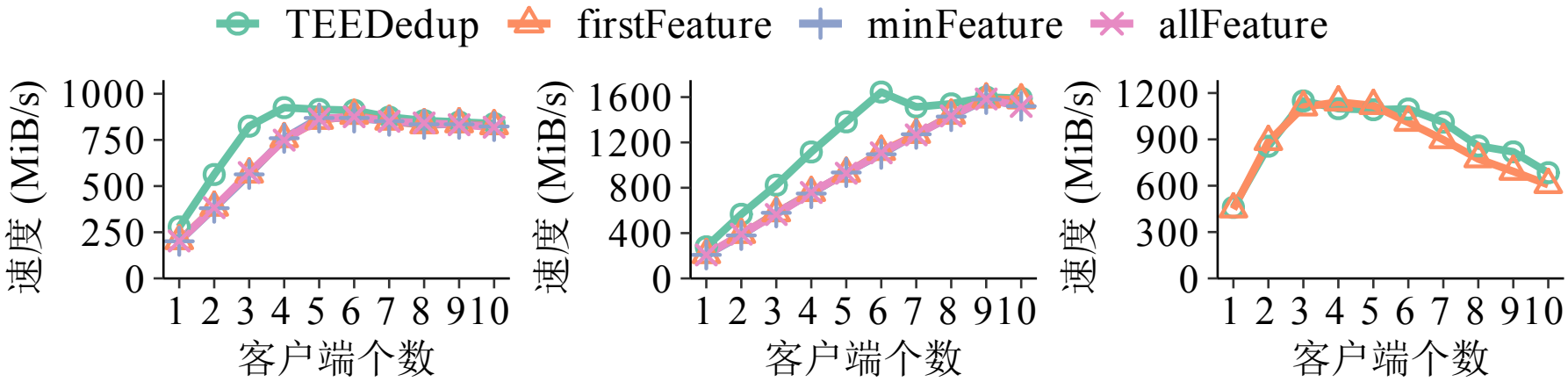


(a) 检测率 (b) 误判率

图 7: Linux 数据集中总体检测率和误判率

- 在对所有窗口大小使用相同检测阈值 T 时，随窗口大小增大误报率降低（W=5,10 K 时为 0）。
- `minFeature` 方案检测到更多相似数据块，相较于其他两种方案产生更高误报率。

TEEDedup+ 原型系统性能



(a) 第一轮上传 (b) 第二轮上传 (c) 下载

图 8: 多客户端上传/下载性能。所有 TEEDedup+ 实例的下载速度均相同的，本文将它们 (橙色) 与 TEEDedup (绿色) 进行比较。

- TEEDedup (4 客户端聚合上传速度 924.9 MiB/s) 比 TEEDedup+ 更早达到峰值性能 (6 客户端聚合上传速度 882.2 MiB/s)

攻读硕士学位期间取得的成果 I

[1] First author. Accelerating Encrypted Deduplication via SGX[C]. Proc.of USENIX ATC (CCF/A), 2021: 957-971.

[2] Third author. Balancing storage efficiency and data confidentiality with tunable encrypted deduplication[C]. Proc. of EuroSys (CCF/B), 2020: 1-15.

[3] Third author. Metadedup: Deduplicating metadata in encrypted deduplication via indirection[C]. Proc. of MSST (CCF/B), 2019: 269-281.

[4] Third author. Enabling Secure and Space-Efficient Metadata Management in Encrypted Deduplication[J]. IEEE Transactions on Computers (CCF/A), 2021.

[5] Third author. Tunable Encrypted Deduplication with Attack-Resilient Key Management[J]. ACM Transactions on Storage (CCF/A), 2022.

[6] Fifth author. Revisiting Frequency Analysis against Encrypted Deduplication via Statistical Distribution[C]. Proc. of IEEE INFOCOM (CCF/A), 2021.

攻读硕士学位期间取得的成果 II

[7] 第三发明人. 发明专利(授权), 一种加密重复数据删除系统中的高效元数据管理方法 [P]: CN110109617B, 2020 年 05 月 12 日.

[8] 第三发明人. 发明专利(授权), 一种可调节加密重复数据删除方法 [P]: CN111338572B, 2021 年 09 月 14 日.

[9] 第二发明人. 发明专利 (公开), 一种基于硬件安全区的高效加密重复数据删除方法 [P]: CN112947855A, 2021 年 06 月 11 日.

[10] 第三发明人. 发明专利 (公开), 基于加密数据去重的分布式密文共享密钥管理方法及系统 [P]: CN112152798A, 2020 年 12 月 29 日.

衷心感谢老师倾听
请各位老师指正！