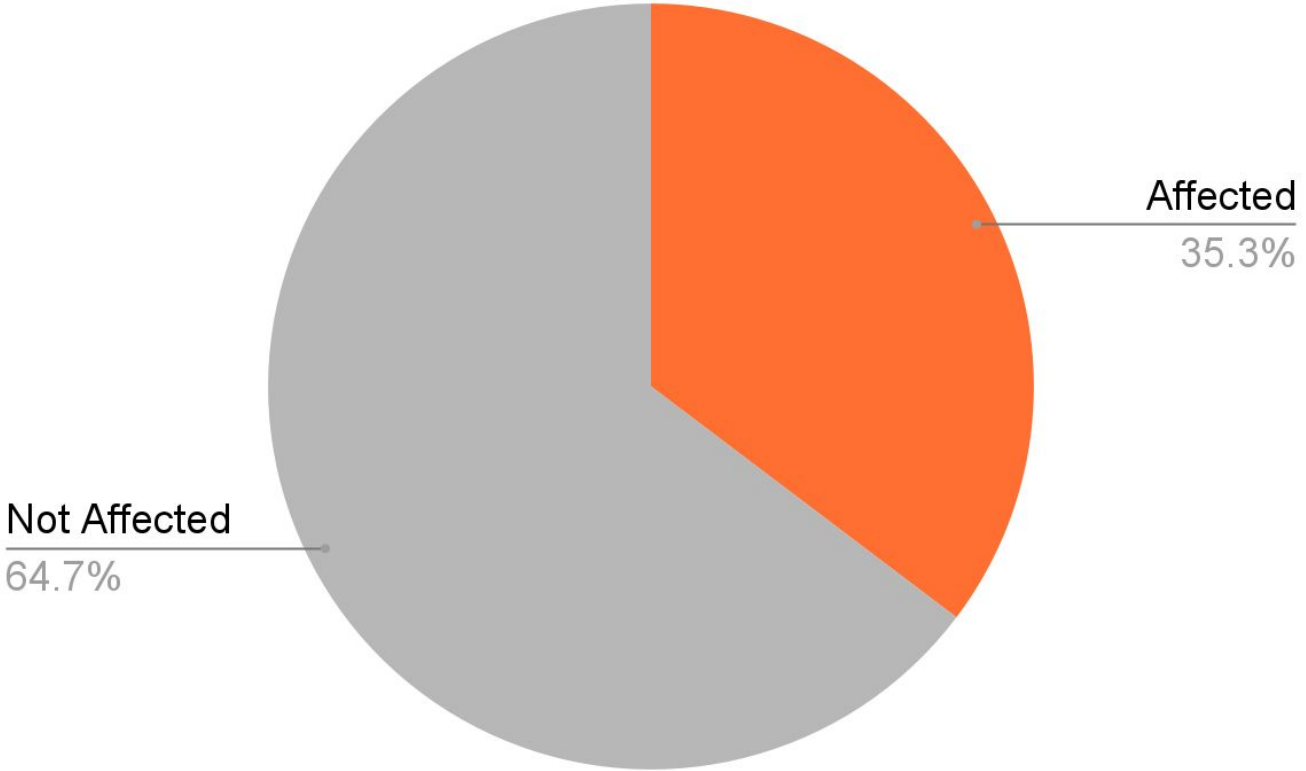
Share of U.S. residents affected by health data breaches in 2015



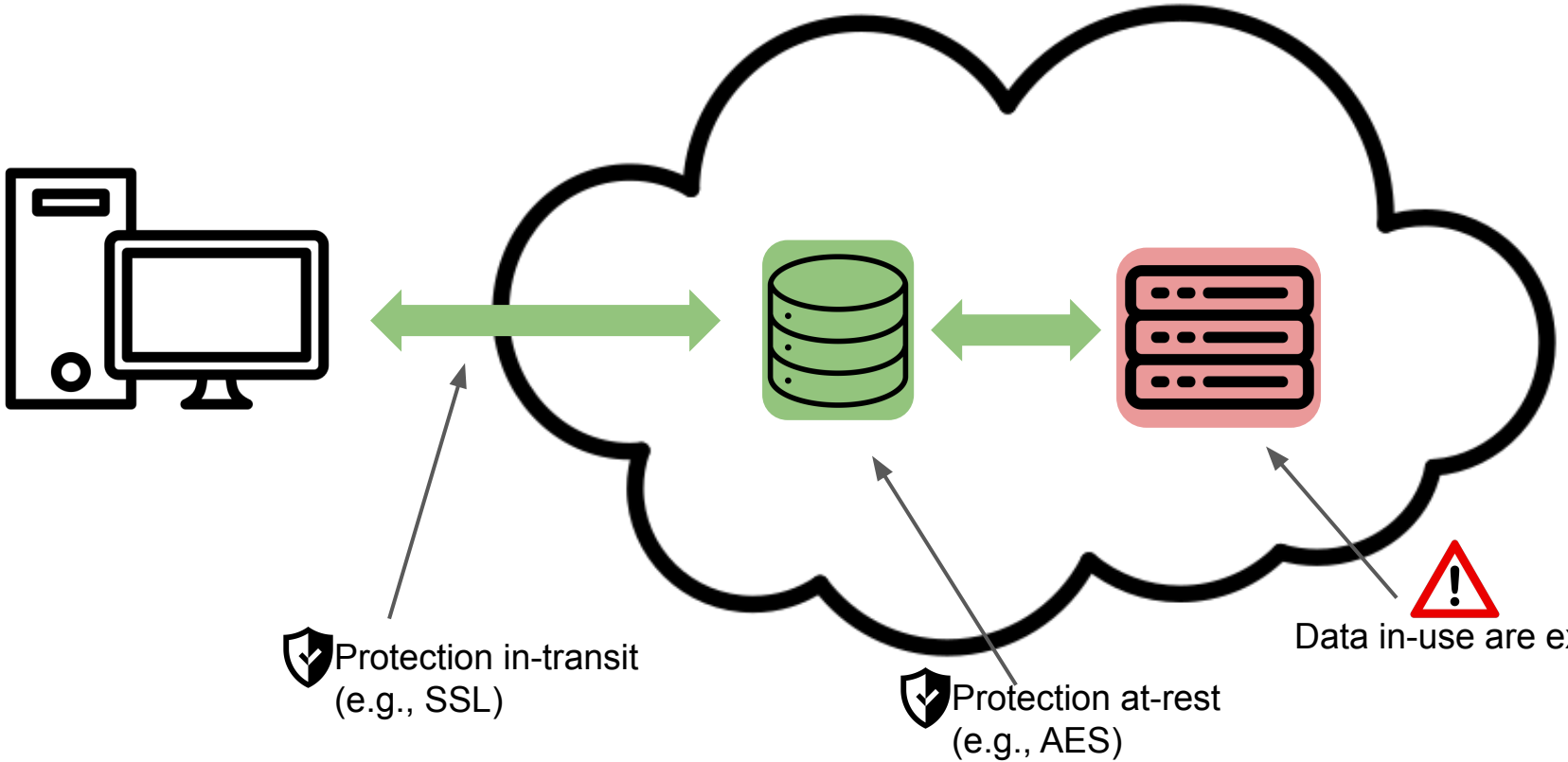
United States; US Department of Health and Human Services (Office for Civil Rights), Statista 2022

Technische Universität Berlin

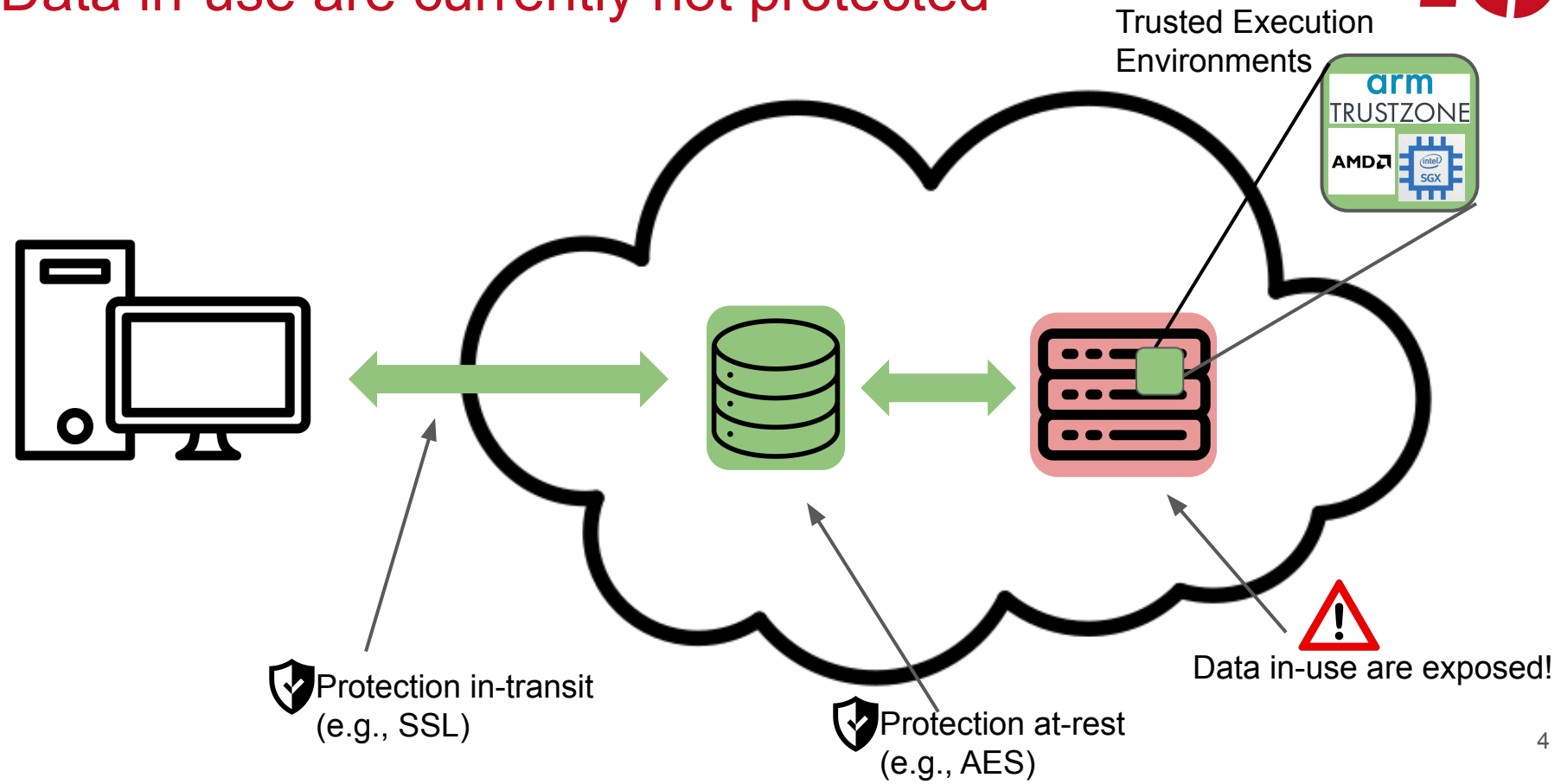
What Is the Price of Joining Securely? Benchmarking Equi-Joins in Trusted Execution Environments

Kajetan Maliszewski, Jorge-Arnulfo Quiané-Ruiz, Jonas Traub, Volker Markl

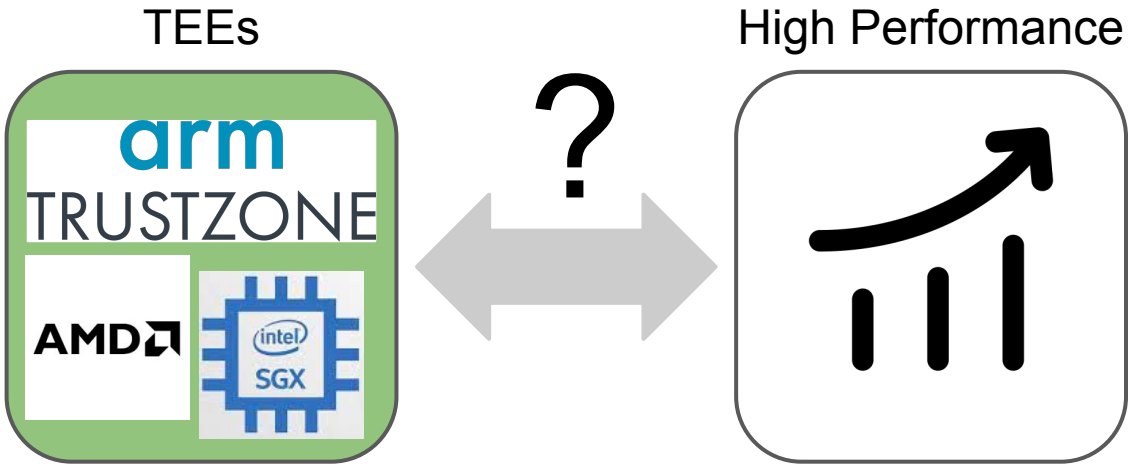
Data in-use are currently not protected



Data in-use are currently not protected

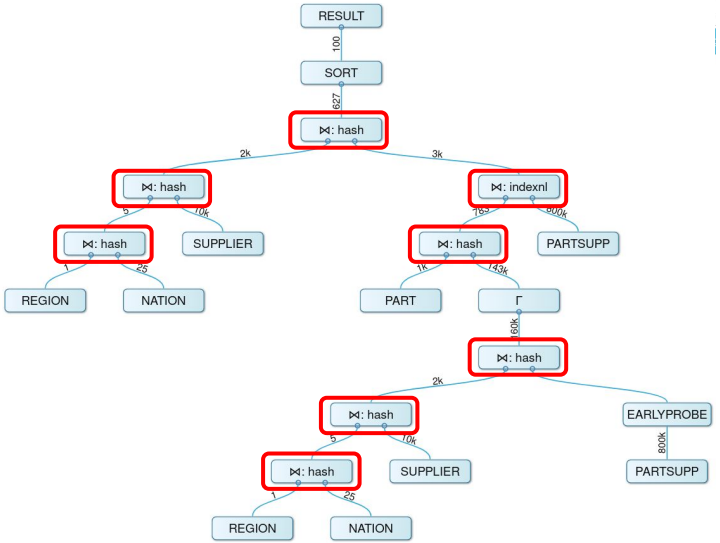


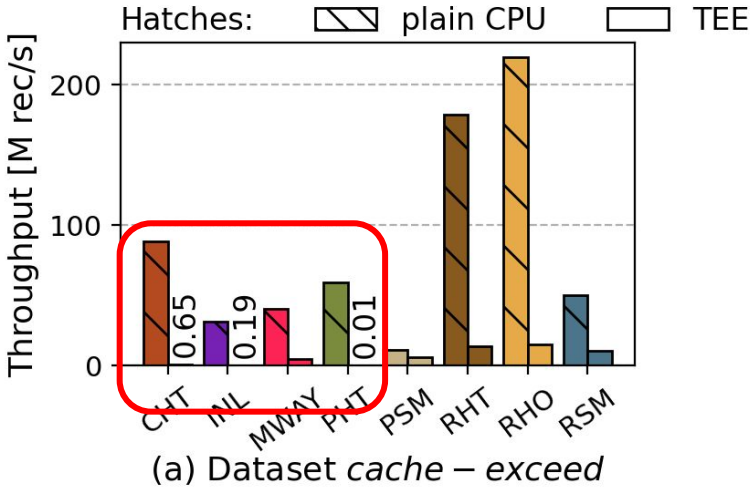
Performance of TEEs is an open challenge



Joins are ubiquitous and expensive

- Eight join operators in TPC-H Q2
- Two orders of magnitude worse in TEEs





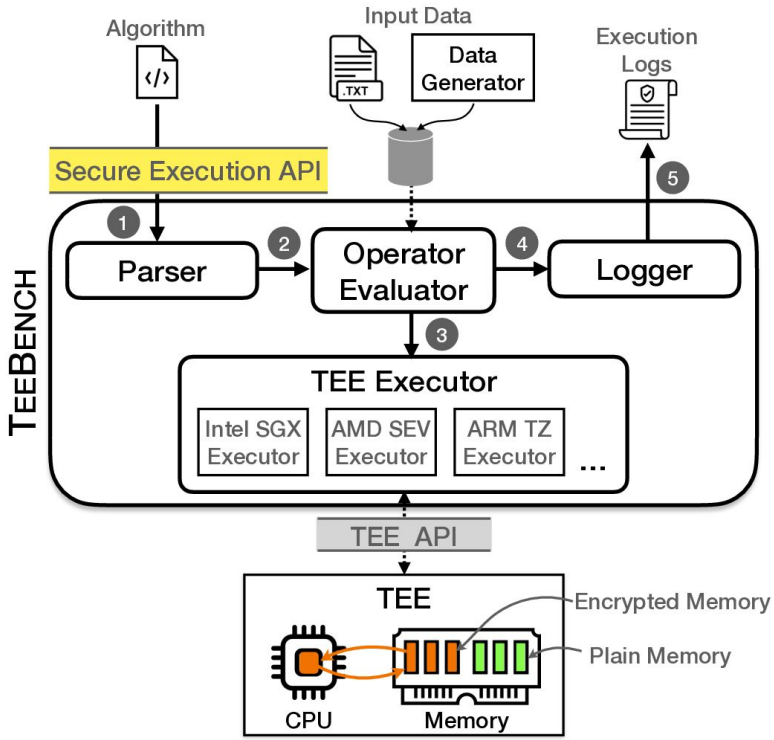
We benchmark all families of join algorithms

Family		Join Algorithm	
hash-based		CHT	Concise Hash Table [1]
		PHT	Parallel Hash Table [2]
sort-merge		PSM	Parallel Sort-Merge
		MWAY	Multi-Way Sort-Merge [3]
radix-based		RHT	Radix Hash Table [3]
		RHO	Radix Hash Optimized [4]
		RSM	Radix Sort-Merge
nested-based		INL	Index Nested Loop

[1] Barber et al., *Memory-efficient hash joins*, PVLDB 2014
[2] Blanas et al., *Design and evaluation of main memory hash join algorithms for multi-core CPUs*, SIGMOD 2011
[3] Balkesen et al., *SMulti-Core, Main-Memory Joins: Sort vs. Hash Revisited*, PVLDB 2014
[4] Balkesen et al., *Main-memory hash joins on multi-core CPUs: Tuning to the underlying hardware*, ICDE 2013

TeeBench is a fair referee for enclave benchmarks

- Plug&Play experience
- TeeBench can execute on any TEE

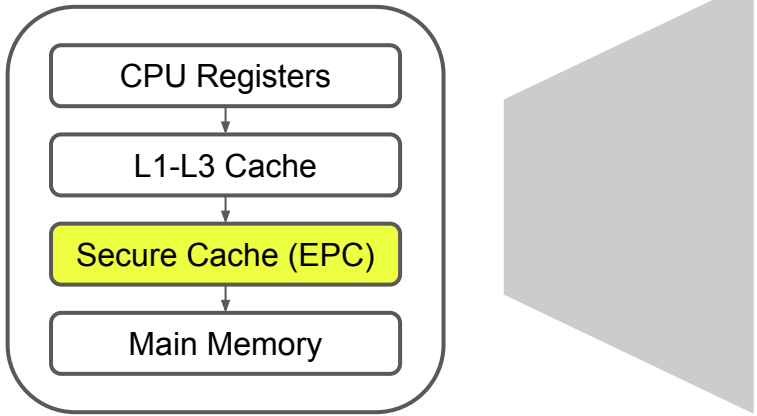


LESSONS LEARNED

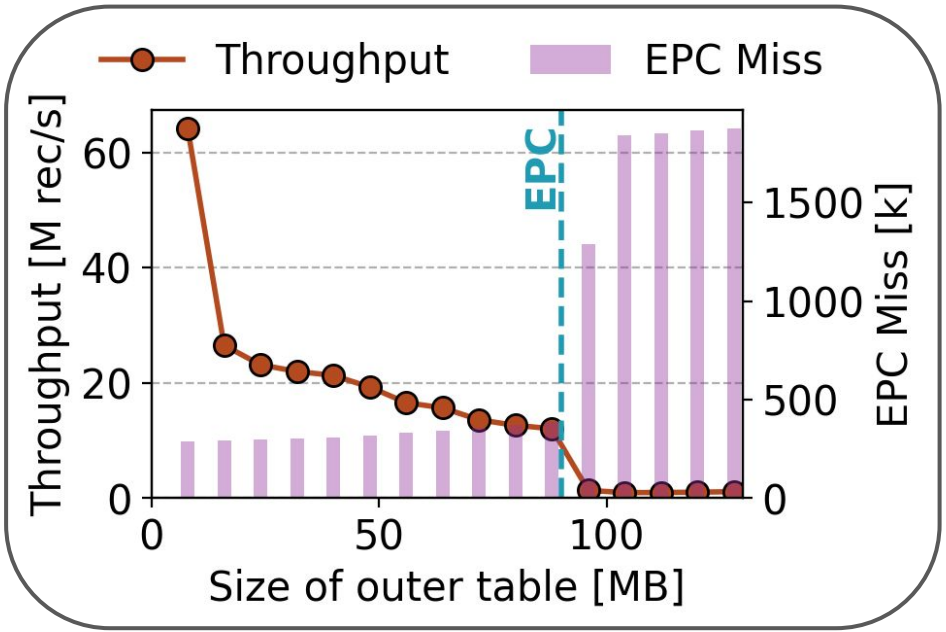
- Hardware Counters
- Multi-threading
- Data Encryption
- Hardware Spectrum

Lesson 1: Fit your data structures into EPC

SGX memory model

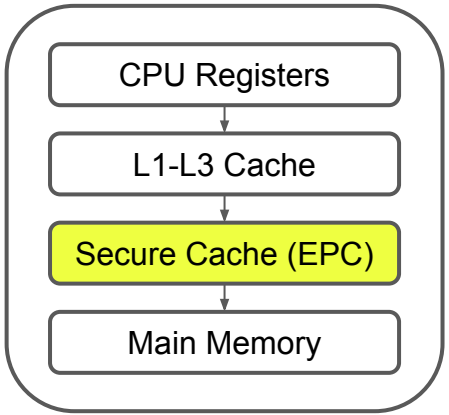


Performance implication

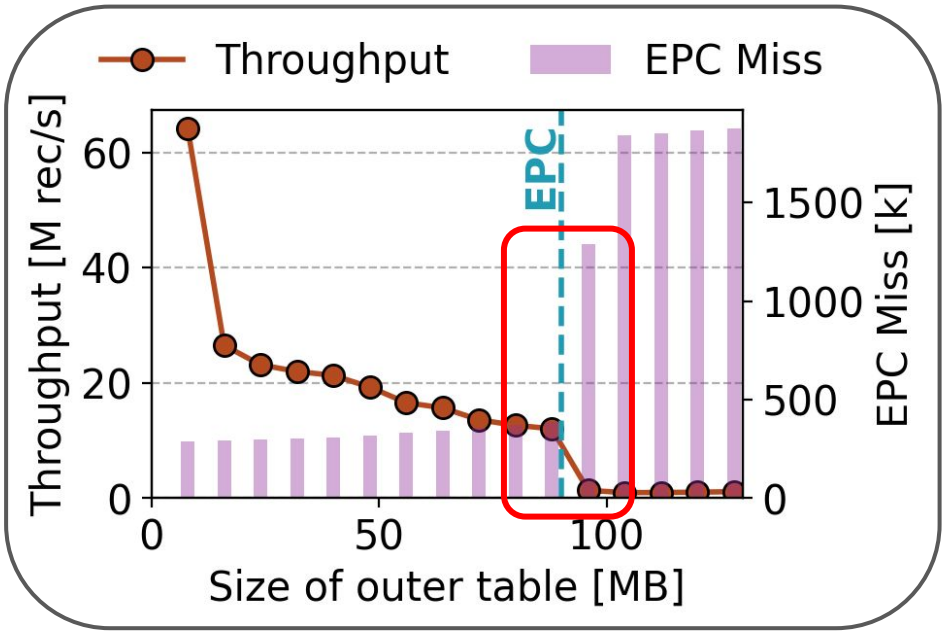


Lesson 1: Fit your data structures into EPC

SGX memory model

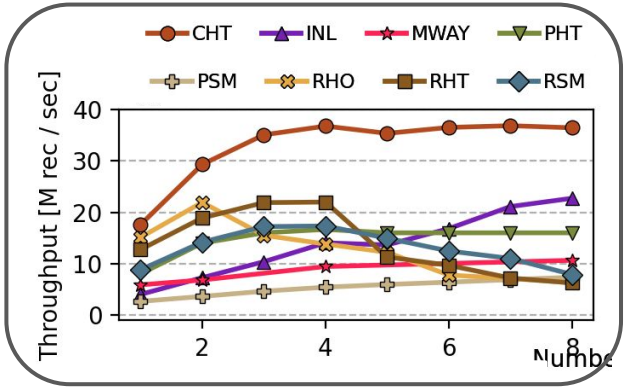


Performance implication

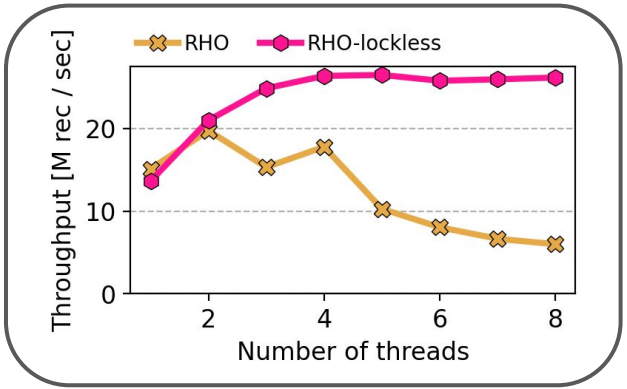


Lesson 2: Mutex is the new bottleneck

Now:
multi-threading can be a performance bottleneck



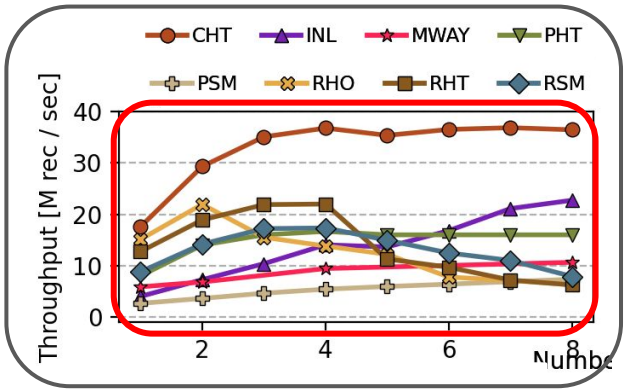
Simple mitigation:
avoid OS interaction

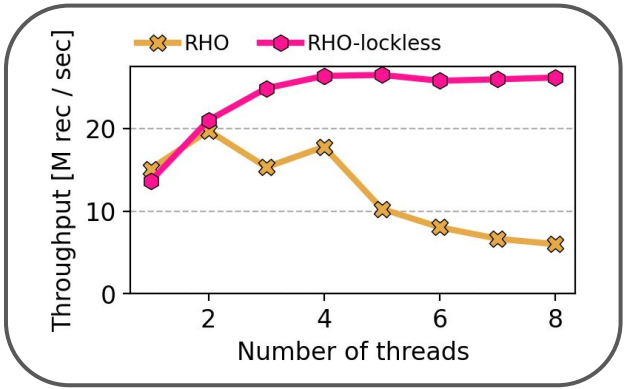


Lesson 2: Mutex is the new bottleneck

Now:
multi-threading can be a performance bottleneck

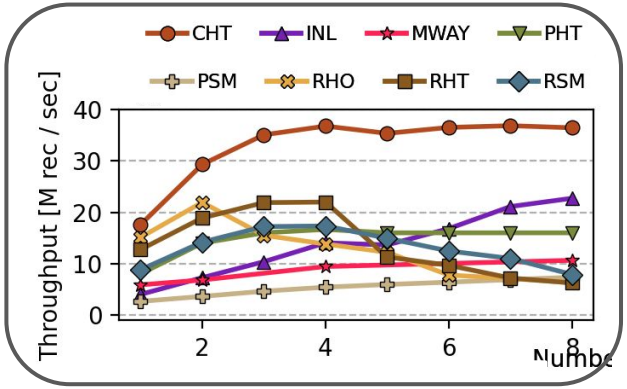
Simple mitigation:
avoid OS interaction



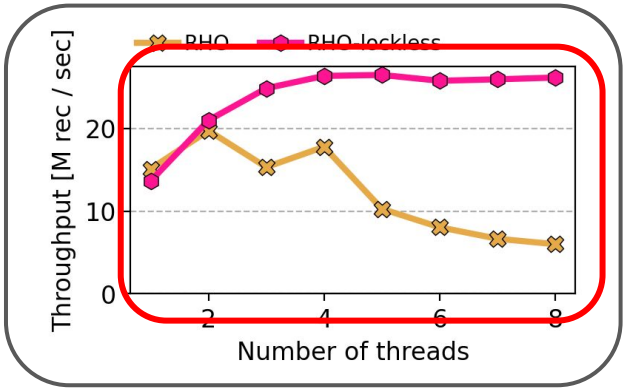


Lesson 2: Mutex is the new bottleneck

Now:
multi-threading can be a performance bottleneck

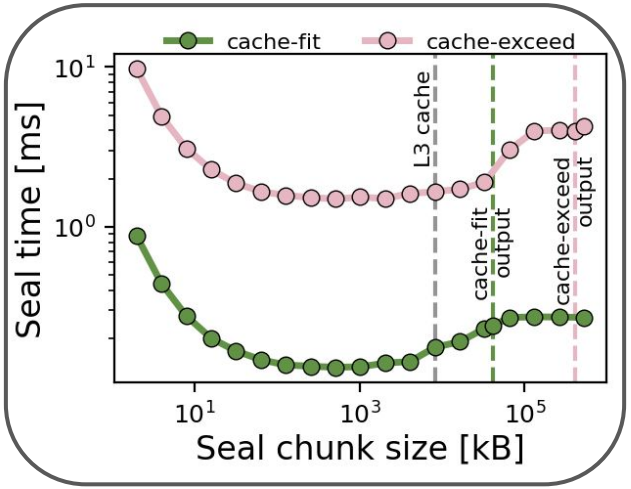


Simple mitigation:
avoid OS interaction

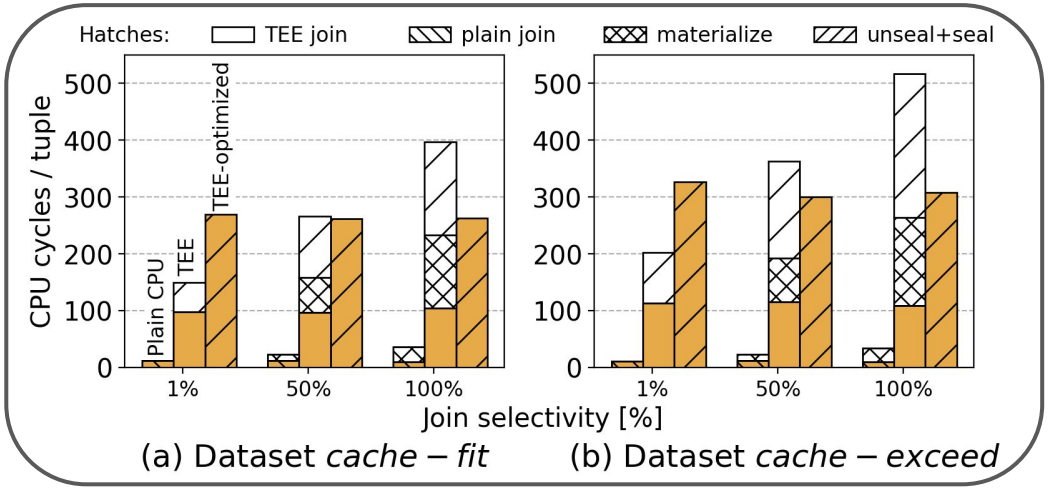


Lesson 3: Use less memory during data encryption

Data encryption in chunks

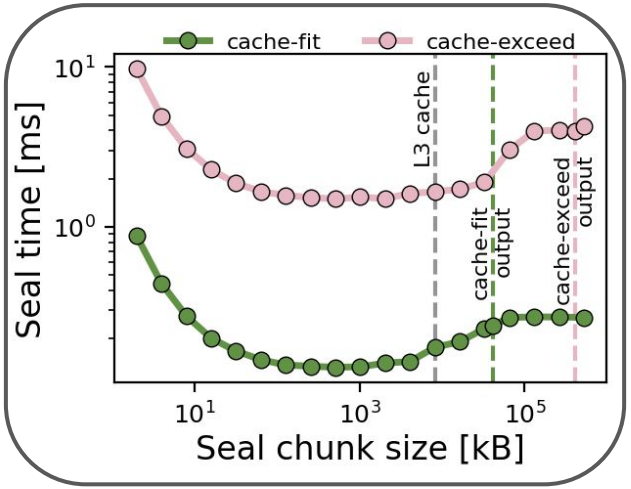


Performance of encryption and materialization

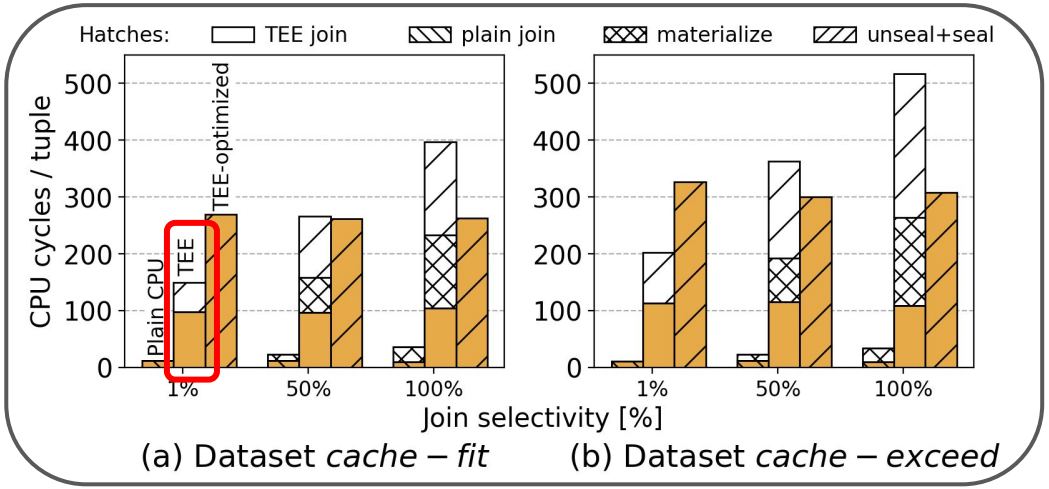


Lesson 3: Use less memory during data encryption

Data encryption in chunks

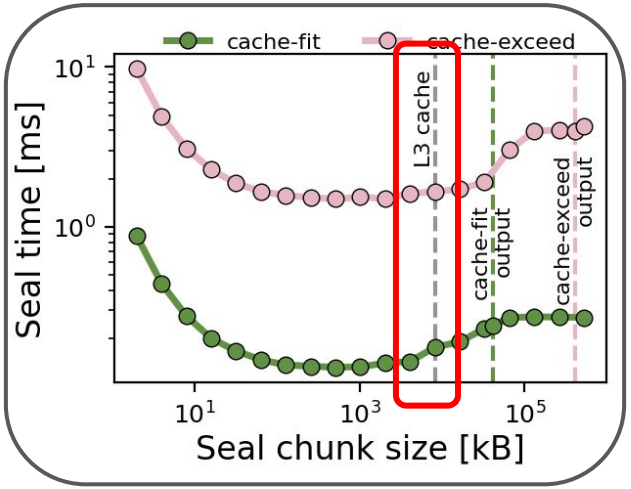


Performance of encryption and materialization

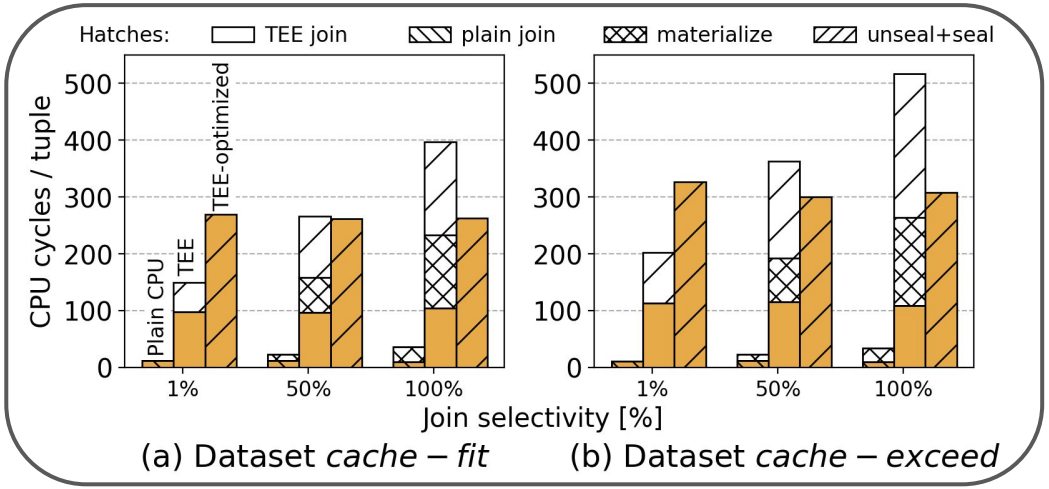


Lesson 3: Use less memory during data encryption

Data encryption in chunks

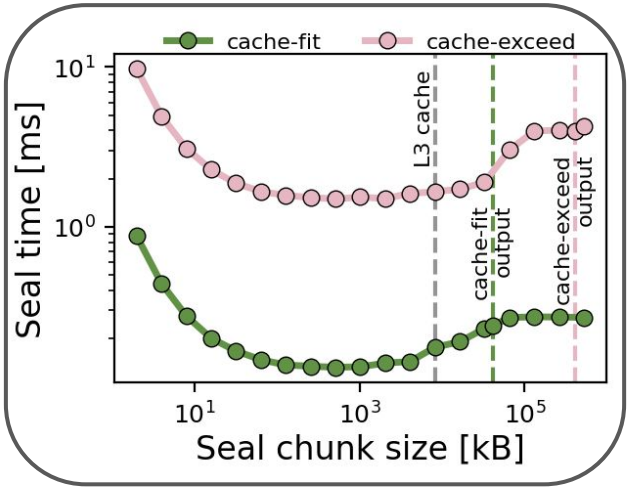


Performance of encryption and materialization

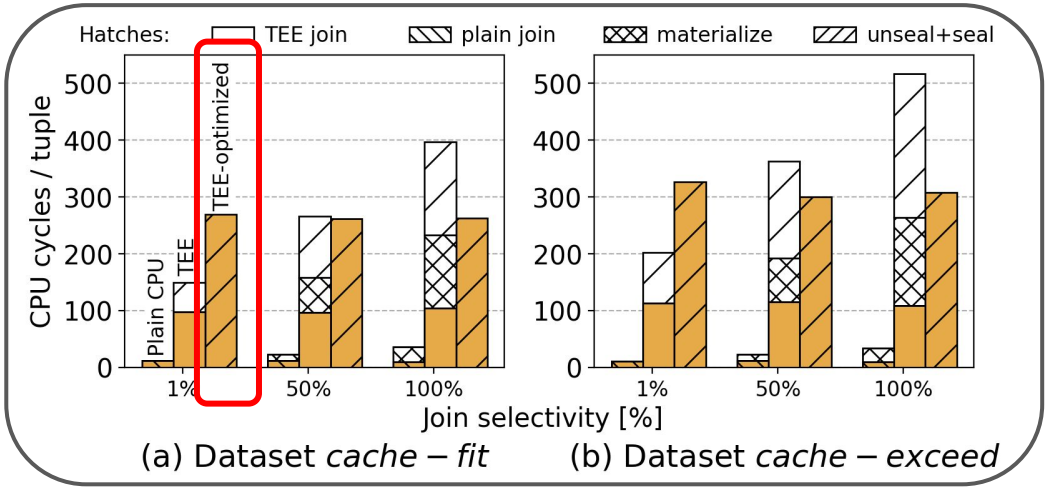


Lesson 3: Use less memory during data encryption

Data encryption in chunks



Performance of encryption and materialization

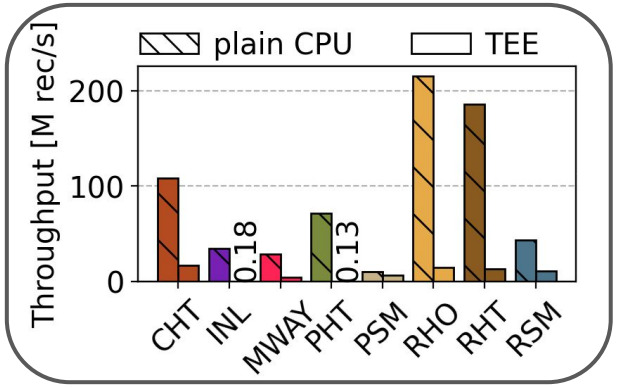


Lesson 4: More throughput is not always better

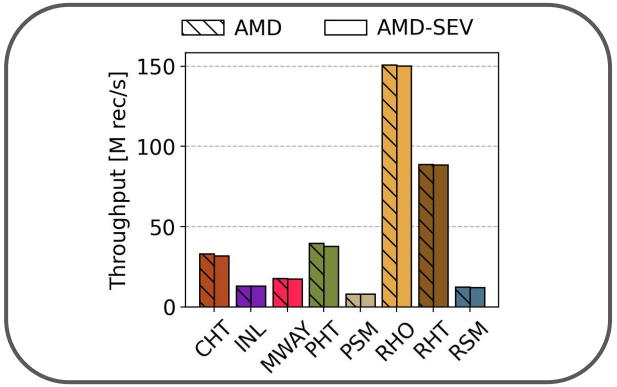
Intel SGXv1

AMD SEV

Performance

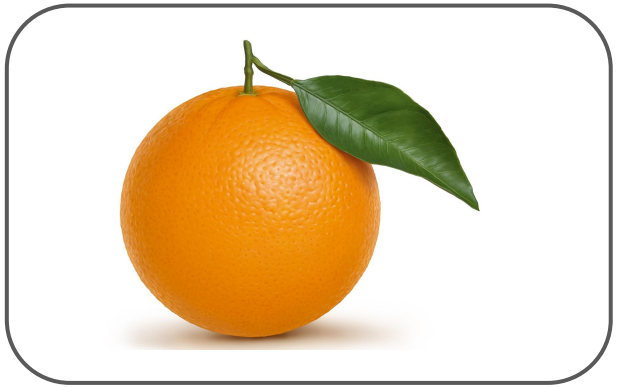


<

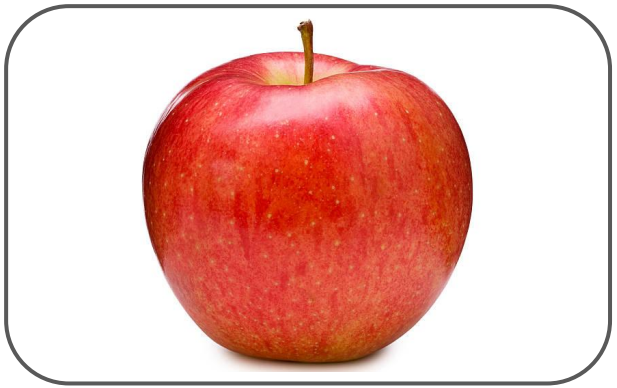


?

Security



≠≠

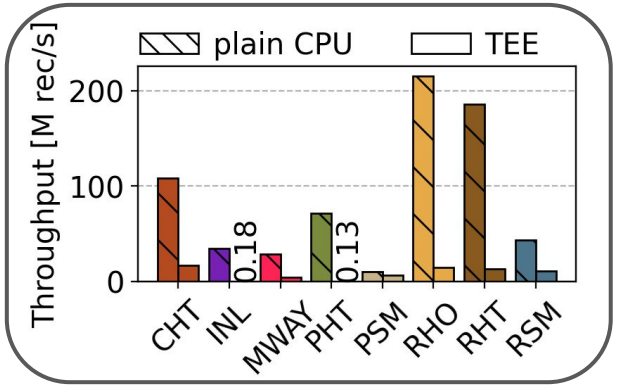


Lesson 4: More throughput is not always better

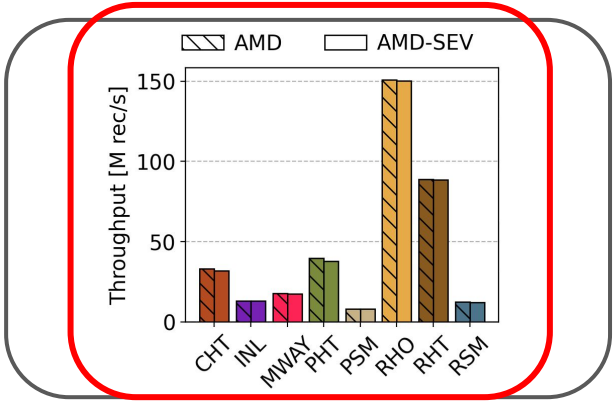
Intel SGXv1

AMD SEV

Performance

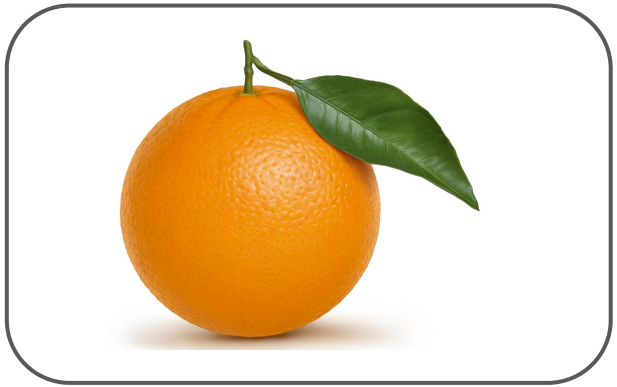


<

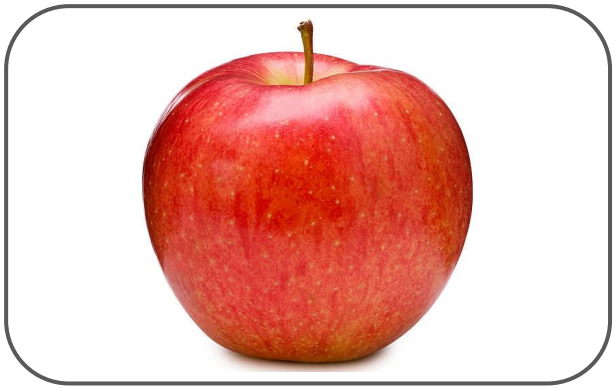


?

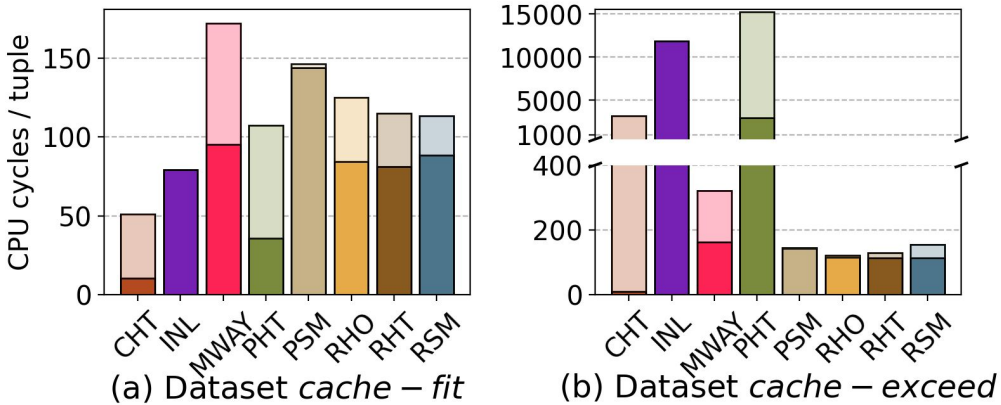
Security



≠≠



The results trail the blaze for future research

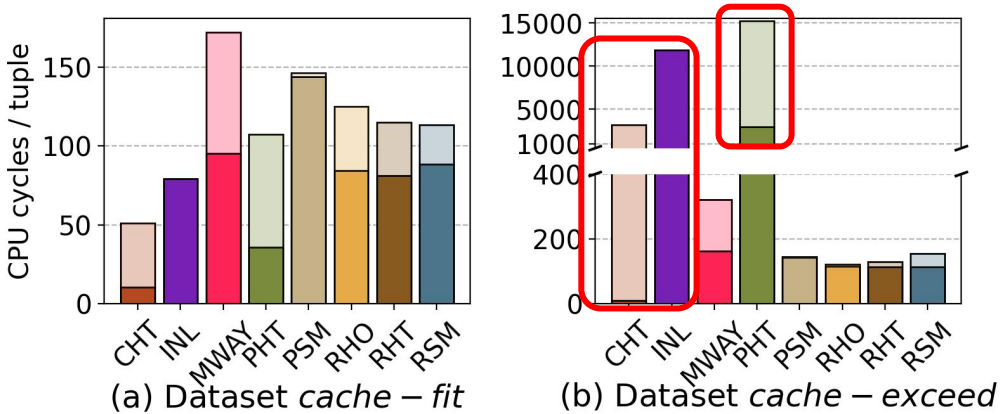


Hash-based joins for small data

Radix-based joins for large data

Memory consumption crucial to high performance

The results blaze a trail for future research



Hash-based joins for small data

Radix-based joins for large data

Memory consumption crucial to high performance

We call for TEE-native algorithms

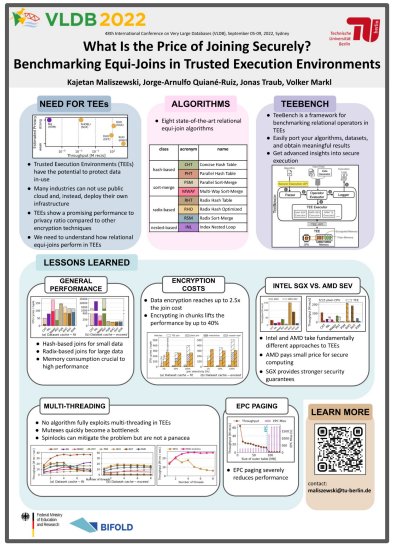
Takeaways

TEEs can protect data in-use in cloud DBs

Existing join algorithms do not fit TEEs

We need TEE-native solutions

Poster C3



The future is confidential!