DDR-SSE: Duplicated Retrieval of Documents for System-wide Secure Searchable Symmetric Encryption

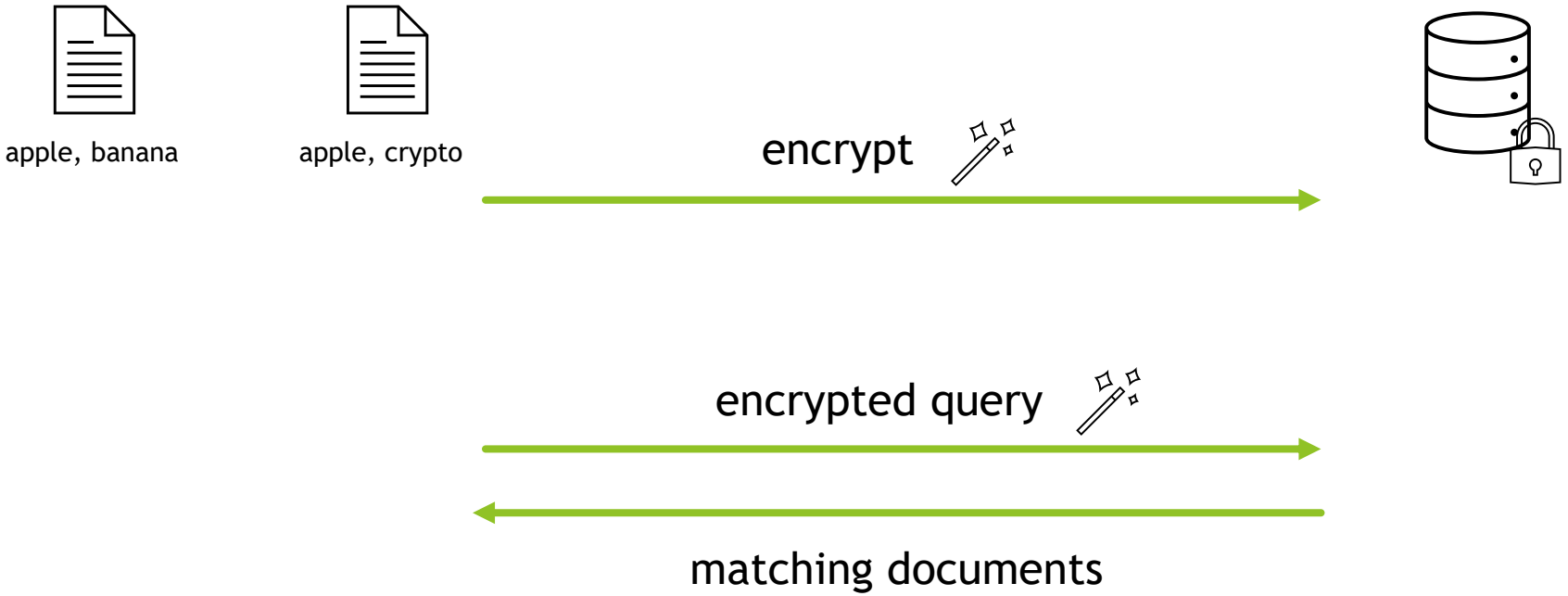
Zichen Gui
University of Georgia

Simon-Philipp Merz
ETH Zurich

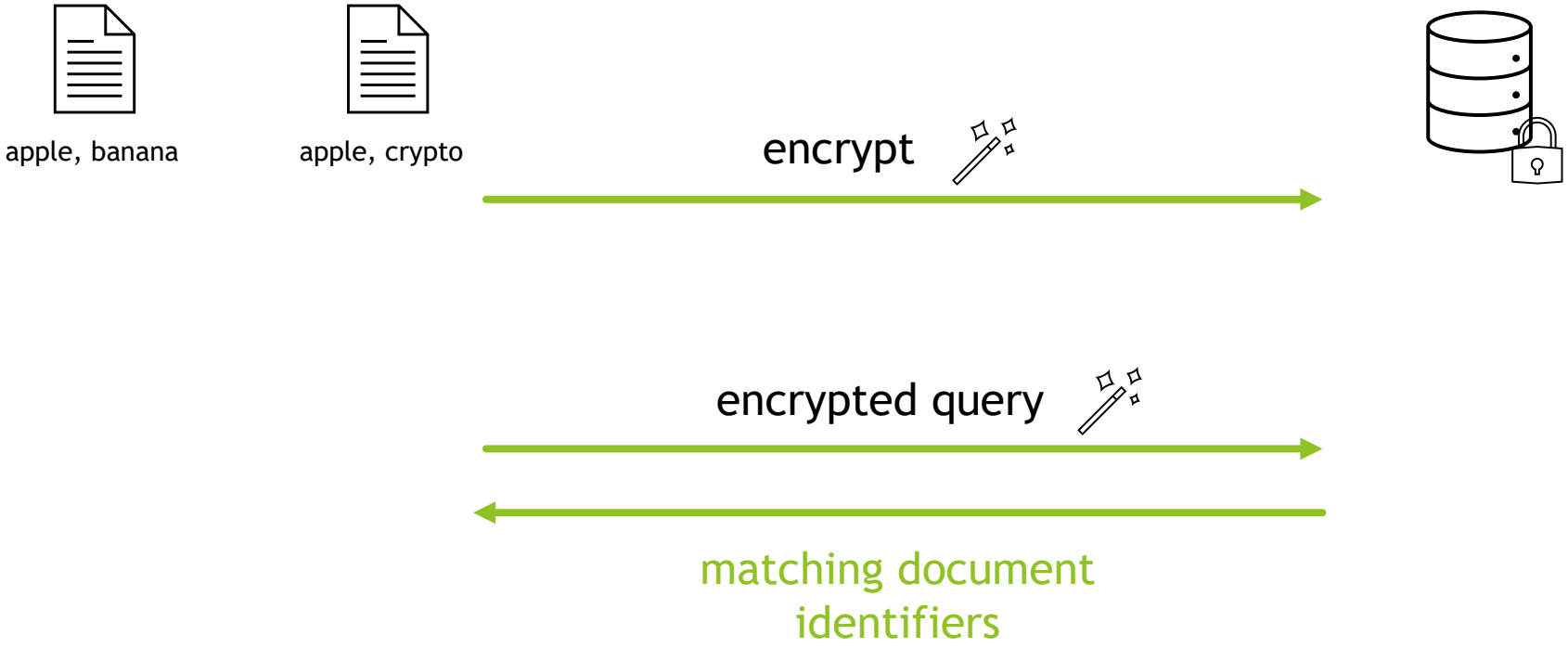
Kenneth G. Paterson
ETH Zurich

Sikhar Patranabis
IBM Research India

Searchable Encryption



Searchable Encryption



Most existing solutions only scale on document identifiers or short data

Duplication



apple, banana





apple, crypto



2 raw documents

Keyword	Documents
apple	
banana	
crypto	



4 encrypted documents so that distinct queries do not touch the same encrypted document

DDR-SSE

- The naïve duplication technique introduces two orders of magnitude storage blowup or more on a typical document collection.
- DDR-SSE asks the question: Can we get away with less duplication?
- There are already secure and efficient schemes for retrieving matching document identifiers. So, DDR-SSE focuses on the document retrieval part.

DDR-SSE

- Provides a system-wide solution that lets the client securely retrieves the actual documents in the end.
- Modular design by using an existing SSE scheme for index retrieval.
- Each (encrypted) document only has to be stored twice.
- Comes with a careful leakage cryptanalysis with existing attacks and our own attack.

DDR-SSE: Document Storage



Freshly encrypt each document twice

DDR-SSE: Document Retrieval



- $b \leftarrow \$\{0,1\}, r \leftarrow \$\{0,1,2\}$.
- If $b = 0$, retrieve the first r documents from the first copy and the remaining ones from the second copy.
- If $b = 1$, retrieve the first r documents from the second copy and the remaining ones from the first copy.

DDR-SSE: Document Retrieval



- $b = 0, r = 1.$
- Actual documents touched by the query are boxed in green.

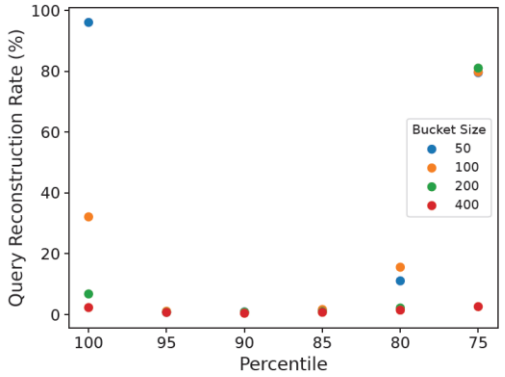
DDR-SSE: Main Result



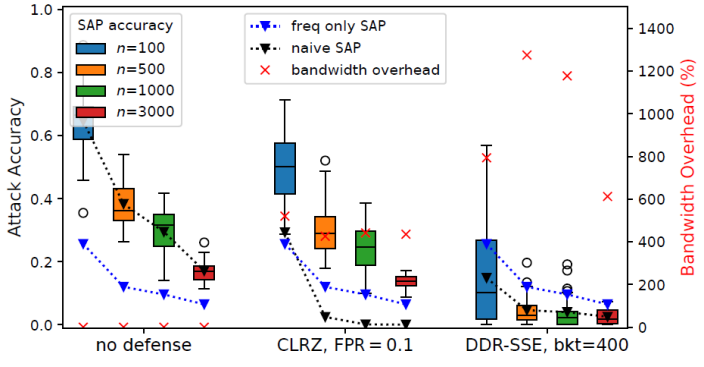


- $b = 0, r = 1.$
- Actual documents touched by the query are boxed in green.
- If two queries match n documents in common, the number of documents the queries touches simultaneously is almost uniformly distributed between 0 and n.
- See the paper for the precise distribution.

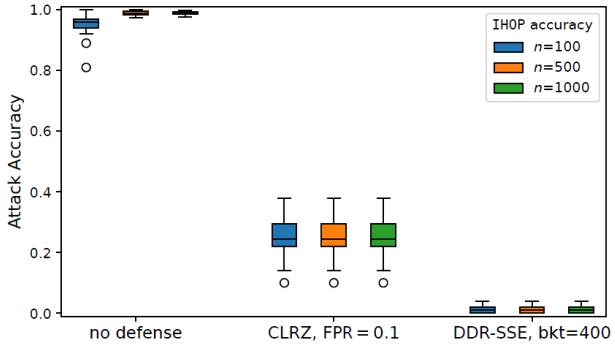
DDR-SSE: Cryptanalysis



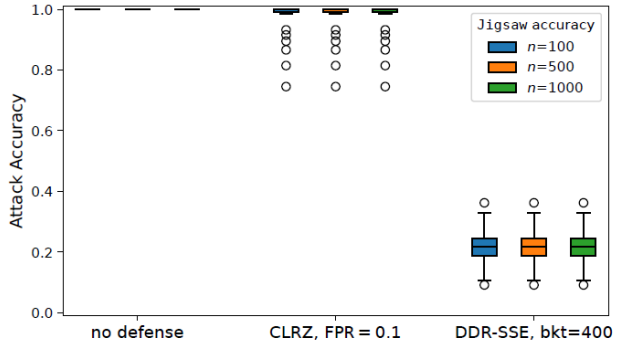
Our attack



SAP



IHOP



Jigsaw

DDR-SSE: Performance

- Raw Enron email corpus (400K documents): 1.3 GB
- DDR-SSE (8KB per document): 7.5 GB
- Search throughput: 35K documents/sec
- For comparison (document retrieval only):
  - EnigMap (ORAM-based): <8K documents/sec
  - Spiral Batch PIR: 0.61 documents/sec

Thank you!


Paper: 2026/753


Code


ARTIFACT
EVALUATED
USENIX
AVAILABLE


ARTIFACT
EVALUATED
USENIX
FUNCTIONAL


ARTIFACT
EVALUATED
USENIX
REPRODUCED