Searchable Encryption

Leyla TEKİN

Izmir Institute of Technology

24.12.2019

Outline

- Introduction
- Base Schemes
- Proposed Approaches
    - Extended Schemes
- Experimental Results
- Conclusion

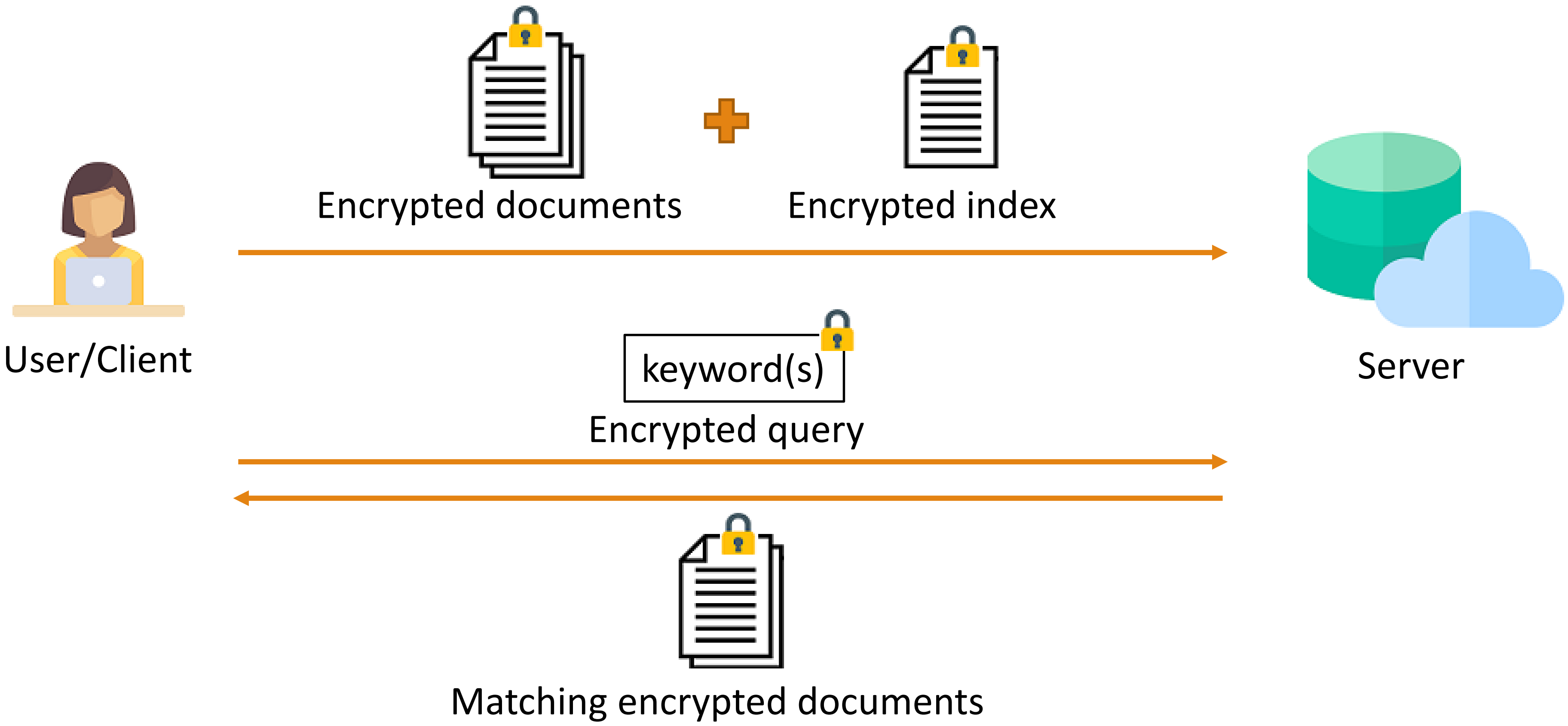
Problem

- Storing *sensitive* data on untrusted servers
- *Encryption* reduces security and privacy risks
- But, it removes search capabilities

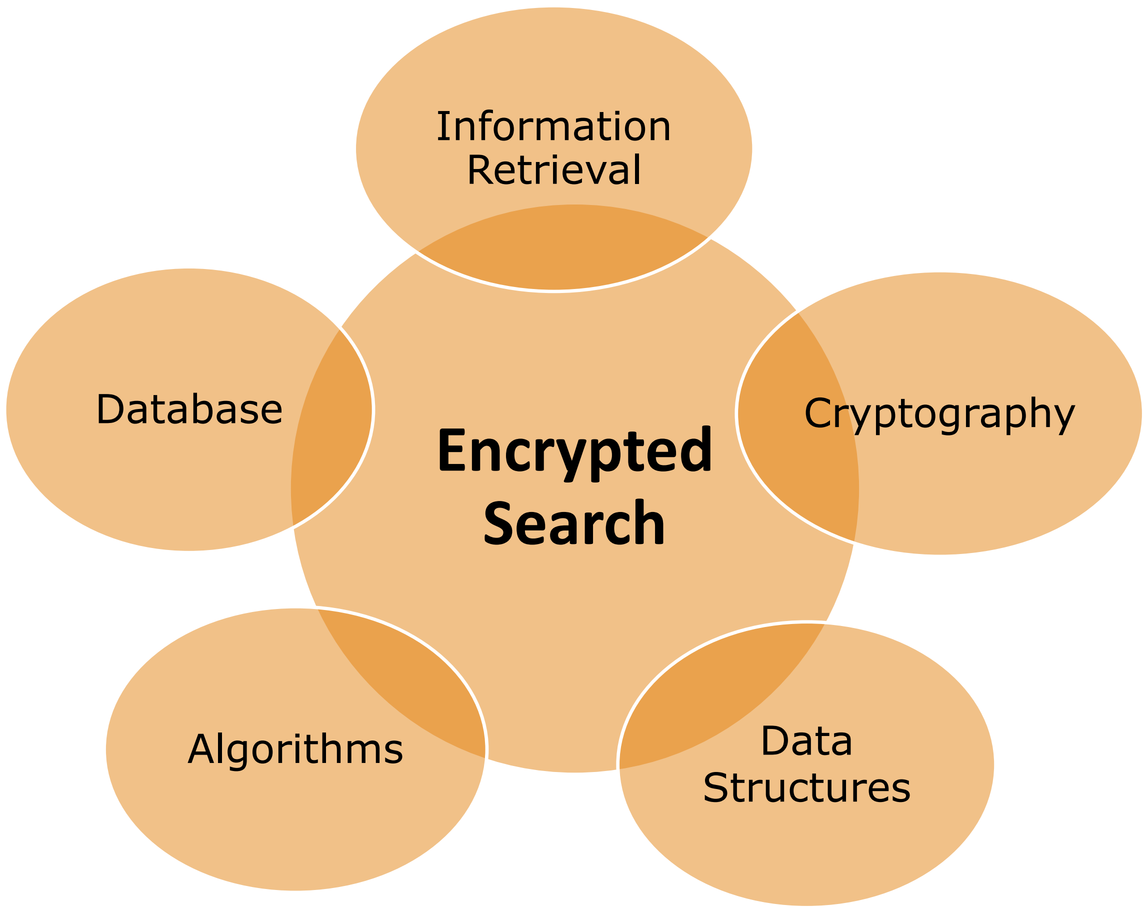
➢ **Solution**: Emergence of Searchable encryption(SE) schemes

Components of SSE Schemes

- Setup, Search, Update*



Related Areas



Motivation

- The static SSE schemes (Cash et al. (2014))
- The dynamic SSE schemes (Kamara and Moataz (2017))

<u>Contribution:</u>

- *Different approaches* for secure *single- and multi-keyword* ***ranked*** searches
- Extend the base schemes
- Modify the structures

Base Schemes

- Response-revealing (RR) or Response-hiding (RH)

- The base schemes:
  - RR2Lev
  - RH2Lev
  - DynRR
  - DynRH

Label	Value
w_1	D_2, D_{10}
w_2	D_4, D_5, D_{10}
w_5	$D_1, D_3, D_5, D_6, D_{10}$
w_7	$D_1, D_2, D_3, D_5, D_6, D_9, D_{10}$

Towards RR2Lev – Basic Scheme

- A key K is chosen.
- For each keyword:
  - Two keys $\boldsymbol{K_1}$ and $\boldsymbol{K_2}$
  - K_1 for PRF and K_2 for encryption
  - A label is associated by applying the PRF and the key K_1 to **a keyword-specific counter**
  - The identifier is encrypted with the key K_2
  - The label/encrypted identifier pair is added to the dictionary.

Label	Value
$F(K_1^1, 0)$	$E(K_2^1, D_2)$
$F(K_1^1, 1)$	$E(K_2^1, D_{10})$
$F(K_1^2, 0)$	$E(K_2^2, D_4)$
$F(K_1^2, 1)$	$E(K_2^2, D_5)$
$F(K_1^2, 2)$	$E(K_2^2, D_{10})$
$F(K_1^5, 0)$	$E(K_2^5, D_1)$
$F(K_1^5, 1)$	$E(K_2^5, D_3)$
$F(K_1^5, 2)$	$E(K_2^5, D_5)$
$F(K_1^5, 3)$	$E(K_2^5, D_6)$
$F(K_1^5, 4)$	$E(K_2^5, D_{10})$
$\ldots$	$\ldots$

RR2Lev

- The result sets of keywords as small, medium and large.
- If **small** ($|\mathrm{DB}(w)| \leq b$):
  - a block of b identifiers in the dictionary
- If **medium** ($b < |\mathrm{DB}(w)| \leq Bb$):
  - blocks of B identifiers in the array
  - a block of b pointers in the dictionary.
- If **large** ($Bb < |\mathrm{DB}(w)| \leq B^2b$):
  - a block of b pointers in the dictionary
  - blocks of B pointers in the array
  - blocks of B identifiers in the array

Label	Value
$F(K_1, 0)$	$E(K_2, D_2 \parallel D_{10})$
$F(K_1, 0)$	$E(K_2, 17 \parallel X)$
$F(K_1, 0)$	$E(K_2, 67 \parallel 20)$
$F(K_1, 0)$	$E(K_2, 6 \parallel X)$

	$\ldots$
6	$E(K_2, 90 \parallel 14 \parallel 85)$
	$\ldots$
14	$E(K_2, D_5 \parallel D_6 \parallel D_9)$
	$\ldots$
17	$E(K_2, D_4 \parallel D_5 \parallel D_{10})$
	$\ldots$
20	$E(K_2, D_6 \parallel D_{10} \parallel X)$
	$\ldots$
67	$E(K_2, D_1 \parallel D_3 \parallel D_5)$
	$\ldots$
85	$E(K_2, D_{10} \parallel X \parallel X)$
	$\ldots$
90	$E(K_2, D_1 \parallel D_2 \parallel D_3)$
	$\ldots$

RH2Lev

- Response hiding scheme
- Encrypted identifiers in the index → $\mathbf{K_3}$

Label	Value
$F(K_1, 0)$	$E(K_2, E(K_3, D_2) \parallel E(K_3, D_{10}))$
$F(K_1, 0)$	$E(K_2, 17 \parallel X)$
$F(K_1, 0)$	$E(K_2, 67 \parallel 20)$
$F(K_1, 0)$	$E(K_2, 6 \parallel X)$

	$\ldots$
6	$E(K_2, 90 \parallel 14 \parallel 85)$
	$\ldots$
14	$E(K_2, E(K_3, D_5) \parallel E(K_3, D_6) \parallel E(K_3, D_9))$
	$\ldots$
17	$E(K_2, E(K_3, D_4) \parallel E(K_3, D_5) \parallel E(K_3, D_{10}))$
	$\ldots$
20	$E(K_2, E(K_3, D_6) \parallel E(K_3, D_{10}) \parallel X)$
	$\ldots$
67	$E(K_2, E(K_3, D_1) \parallel E(K_3, D_3) \parallel E(K_3, D_5))$
	$\ldots$
85	$E(K_2, E(K_3, D_{10}) \parallel X \parallel X)$
	$\ldots$
90	$E(K_2, E(K_3, D_1) \parallel E(K_3, D_2) \parallel E(K_3, D_3))$
	$\ldots$

- An additional algorithm : **resolve**

DynRR

- Allocates two dictionaries D and D_{count} and a list P.
  - **$\mathbf{D_{count}}$** at the user *as state*
  - **D** and **P** at the server

$K_1, K_2, 4$ →

L = [$D_1, D_3, D_5, D_6, D_{10}$] ←

→ I = [1, 3]

Label	Value

Label	Value
$F(K_1, 0)$	$E(K_2, D_2)$
$F(K_1, 1)$	$E(K_2, D_{10})$
$F(K_1, 0)$	$E(K_2, D_4)$
$F(K_1, 1)$	$E(K_2, D_5)$
$F(K_1, 2)$	$E(K_2, D_{10})$
$F(K_1, 0)$	$E(K_2, D_1)$
$F(K_1, 2)$	$E(K_2, D_5)$
$F(K_1, 4)$	$E(K_2, D_{10})$
...	...
...	...

P = [0, 1, 2, 3, 4]

L = [$D_1, D_3, D_5, D_6, D_{10}$]

DynRH

- **Encrypted** identifiers in the index
- **Resolve** algorithm

Label	Value
$F(K_1, 0)$	$E(K_2, E(K_3, D_2))$
$F(K_1, 1)$	$E(K_2, E(K_3, D_{10}))$
$F(K_1, 0)$	$E(K_2, E(K_3, D_4))$
$F(K_1, 1)$	$E(K_2, E(K_3, D_5))$
$F(K_1, 2)$	$E(K_2, E(K_3, D_{10}))$
$\ldots$	$\ldots$

Proposed Approaches for Ranked SE Schemes

Scheme	Sorted	OPE	Paillier
SR-RR2Lev	✓	✓	✓
SR-RH2Lev	✓	✓	✓
SR-DynRR		✓	✓
SR-DynRH		✓	✓
MR-RR2Lev			✓
MR-RH2Lev			
MR-DynRR			✓
MR-DynRH			

- **static** & single-keyword:

$$\mathrm{score}(w, F_{\mathrm{id}}) = \frac{1}{|F_{id}|}\left(1 + \ln(f_{\mathrm{id}, w})\right).$$

- **static** & multi-keyword:

$$\mathrm{score}(w, F_{\mathrm{id}}) = \frac{1}{|F_{id}|}\left(1 + \ln(f_{\mathrm{id}, w})\right)(1 + \frac{N}{f_w}).$$

- **dynamic** & single-keyword:

$$\mathrm{score}(w, F_{\mathrm{id}}) = f_{id,w}.$$

- **multi-keyword:**

$$\mathrm{score}(Q, F_{\mathrm{id}}) = \sum_{w \in Q} score(w, Fid).$$

Sorted Ranked SE Schemes

- Only *single keyword* searches
- The identifiers in descending order of relevance scores
- *Static*

OPE-Based Ranked SE Schemes

- A relevance score for each keyword-document pair

Label	Value
$F(K_1, 0)$	$E(K_2, \langle D_2, E(K_{ope}, 130)\rangle \parallel \langle D_{10}, E(K_{ope}, 42)\rangle)$
$F(K_1, 0)$	$E(K_2, \langle 17, \underline{R}\rangle \parallel \langle X, \underline{R}\rangle)$
$F(K_1, 0)$	$E(K_2, \langle 67, \underline{R}\rangle \parallel \langle 20, \underline{R}\rangle)$
$F(K_1, 0)$	$E(K_2, \langle 6, \underline{R}\rangle \parallel \langle X, \underline{R}\rangle)$

- $x < y \rightarrow \mathrm{Enc}(x) < \mathrm{Enc}(y)$
- *Single keyword* searches

	$\ldots$
6	$E(K_2, \langle 90, \underline{R}\rangle \parallel \langle 14, \underline{R}\rangle \parallel \langle 85, \underline{R}\rangle)$
	$\ldots$
14	$E(K_2, \langle D_5, E(K_{ope}, 79)\rangle \parallel \langle D_6, E(K_{ope}, 174)\rangle \parallel \langle D_9, E(K_{ope}, 279)\rangle)$
	$\ldots$
17	$E(K_2, \langle D_4, E(K_{ope}, 209)\rangle \parallel \langle D_5, E(K_{ope}, 79)\rangle \parallel \langle D_{10}, E(K_{ope}, 76)\rangle)$
	$\ldots$
20	$E(K_2, \langle D_6, E(K_{ope}, 147)\rangle \parallel \langle D_{10}, E(K_{ope}, 87)\rangle \parallel \langle X, \underline{R}\rangle)$
	$\ldots$
67	$E(K_2, \langle D_1, E(K_{ope}, 130)\rangle \parallel \langle D_3, E(K_{ope}, 199)\rangle \parallel \langle D_5, E(K_{ope}, 69)\rangle)$
	$\ldots$
85	$E(K_2, \langle D_{10}, E(K_{ope}, 52)\rangle \parallel \langle X, \underline{R}\rangle \parallel \langle X, \underline{R}\rangle)$
	$\ldots$
90	$E(K_2, \langle D_1, E(K_{ope}, 84)\rangle \parallel \langle D_2, E(K_{ope}, 139)\rangle \parallel \langle D_3, E(K_{ope}, 119)\rangle)$
	$\ldots$

OPE-Based SR-DynRR: A dynamic example

- **No** *multi-keyword* scheme
  - *3 + 15 < 19*
  - Enc(3) + Enc(15) ≰ Enc(19)

$K_1, K_2, 4$

L = [D_3, D_6, D_1, D_{10} , D_5]

I = [D_3, D_6]

Label	Value
$F(K_1,0)$	$E(K_2, \langle D_2, E(K_{ope}, 130)\rangle)$
$F(K_1,1)$	$E(K_2, D_{10}, E(K_{ope}, 42)\rangle)$
$F(K_1,0)$	$E(K_2, D_4, E(K_{ope}, 209)\rangle)$
$F(K_1,1)$	$E(K_2, D_5, E(K_{ope}, 79)\rangle)$
$F(K_1,2)$	$E(K_2, D_{10}, E(K_{ope}, 76)\rangle)$
$F(K_1,0)$	$E(K_2, D_1, E(K_{ope}, 130)\rangle)$
$F(K_1,2)$	$E(K_2, D_5, E(K_{ope}, 69)\rangle)$
$F(K_1,4)$	$E(K_2, D_{10}, E(K_{ope}, 87)\rangle)$
…	…
…	…

Ᵽ = { }

D_1	0
D_3	1
D_5	2
D_6	3
D_{10}	4

L = [D_3, D_6, D_1, D_{10} , D_5]

Paillier-Based Ranked SE Schemes

- **Resolve**

Label	Value
$F(K_1, 0)$	$E(K_2, \langle D_2, E(K_{pai1}, 130)\rangle \parallel \langle D_{10}, E(K_{pai1}, 42)\rangle)$
$F(K_1, 0)$	$E(K_2, \langle 17, \underline{R}\rangle \parallel \langle X, \underline{R}\rangle)$
$F(K_1, 0)$	$E(K_2, \langle 67, \underline{R}\rangle \parallel \langle 20, \underline{R}\rangle)$
$F(K_1, 0)$	$E(K_2, \langle 6, \underline{R}\rangle \parallel \langle X, \underline{R}\rangle)$

- *Single keyword* searches
- *Multi-keyword* searches (only response-revealing)

	$\ldots$
6	$E(K_2, \langle 90, \underline{R}\rangle \parallel \langle 14, \underline{R}\rangle \parallel \langle 85, \underline{R}\rangle)$
	$\ldots$
14	$E(K_2, \langle D_5, E(K_{pai1}, 79)\rangle \parallel \langle D_6, E(K_{pai1}, 174)\rangle \parallel \langle D_9, E(K_{pai1}, 279)\rangle)$
	$\ldots$
17	$E(K_2, \langle D_4, E(K_{pai1}, 209)\rangle \parallel \langle D_5, E(K_{pai1}, 79)\rangle \parallel \langle D_{10}, E(K_{pai1}, 76)\rangle)$
	$\ldots$
20	$E(K_2, \langle D_6, E(K_{pai1}, 147)\rangle \parallel \langle D_{10}, E(K_{pai1}, 87)\rangle \parallel \langle X, \underline{R}\rangle)$
	$\ldots$
67	$E(K_2, \langle D_1, E(K_{pai1}, 130)\rangle \parallel \langle D_3, E(K_{pai1}, 199)\rangle \parallel \langle D_5, E(K_{pai1}, 69)\rangle)$
	$\ldots$
85	$E(K_2, \langle D_{10}, E(K_{pai1}, 52)\rangle \parallel \langle X, \underline{R}\rangle \parallel \langle X, \underline{R}\rangle)$
	$\ldots$
90	$E(K_2, \langle D_1, E(K_{pai1}, 84)\rangle \parallel \langle D_2, E(K_{pai1}, 139)\rangle \parallel \langle D_3, E(K_{pai1}, 119)\rangle)$
	$\ldots$

Paillier-Based Ranked SE Schemes: A multi-keyword example

Label	Value
$F(K_1,0)$	$E(K_2, \langle D_2, E(K_{pai1}, 130)\rangle \parallel \langle D_{10}, E(K_{pai1}, 42)\rangle)$
$F(K_1,0)$	$E(K_2, \langle 17, R\rangle \parallel \langle X, R\rangle)$
$F(K_1,0)$	$E(K_2, \langle 67, R\rangle \parallel \langle 20, R\rangle)$
$F(K_1,0)$	$E(K_2, \langle 6, R\rangle \parallel \langle X, R\rangle)$

	$\ldots$
6	$E(K_2, \langle 90, R\rangle \parallel \langle 14, R\rangle \parallel \langle 85, R\rangle)$
	$\ldots$
14	$E(K_2, \langle D_5, E(K_{pai1}, 79)\rangle \parallel \langle D_6, E(K_{pai1}, 174)\rangle \parallel \langle D_9, E(K_{pai1}, 279)\rangle)$
	$\ldots$
17	$E(K_2, \langle D_4, E(K_{pai1}, 209)\rangle \parallel \langle D_5, E(K_{pai1}, 79)\rangle \parallel \langle D_{10}, E(K_{pai1}, 76)\rangle)$
	$\ldots$
20	$E(K_2, \langle D_6, E(K_{pai1}, 147)\rangle \parallel \langle D_{10}, E(K_{pai1}, 87)\rangle \parallel \langle X, R\rangle)$
	$\ldots$
67	$E(K_2, \langle D_1, E(K_{pai1}, 130)\rangle \parallel \langle D_3, E(K_{pai1}, 199)\rangle \parallel \langle D_5, E(K_{pai1}, 69)\rangle)$
	$\ldots$
85	$E(K_2, \langle D_{10}, E(K_{pai1}, 52)\rangle \parallel \langle X, R\rangle \parallel \langle X, R\rangle)$
	$\ldots$
90	$E(K_2, \langle D_1, E(K_{pai1}, 84)\rangle \parallel \langle D_2, E(K_{pai1}, 139)\rangle \parallel \langle D_3, E(K_{pai1}, 119)\rangle)$
	$\ldots$

$K^1{}_1, K^1{}_2, K^2{}_1, K^2{}_2$

D_2	$E(K_{pai1}, 130)$
D_4	$E(K_{pai1}, 209)$
D_5	$E(K_{pai1}, 79)$
D_{10}	$E(K_{pai1}, 118)$

D_2	$E(K_{pai1}, 130)$
D_4	$E(K_{pai1}, 209)$
D_5	$E(K_{pai1}, 79)$
D_{10}	$E(K_{pai1}, 118)$
D_{10}	$E(K_{pai1}, 76)$

$$E(42) * E(76) = E(42 + 76)$$

- No response-hiding schemes for multi-keyword searches

Experimental Results

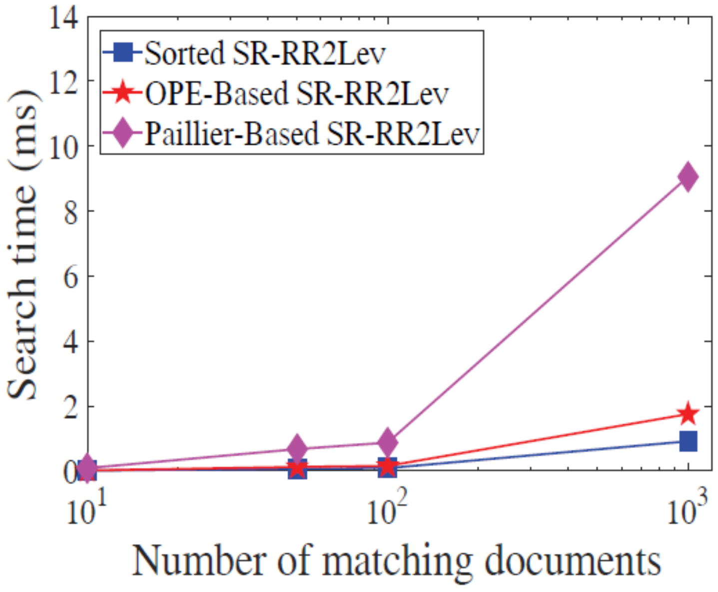
- Intel Core i7 8750 2.2GHz CPU with 16GB RAM running Windows 10

- RFC dataset (text files)
- AES in CTR mode with 256-bit key
- HMAC-SHA512 for key-based hash function

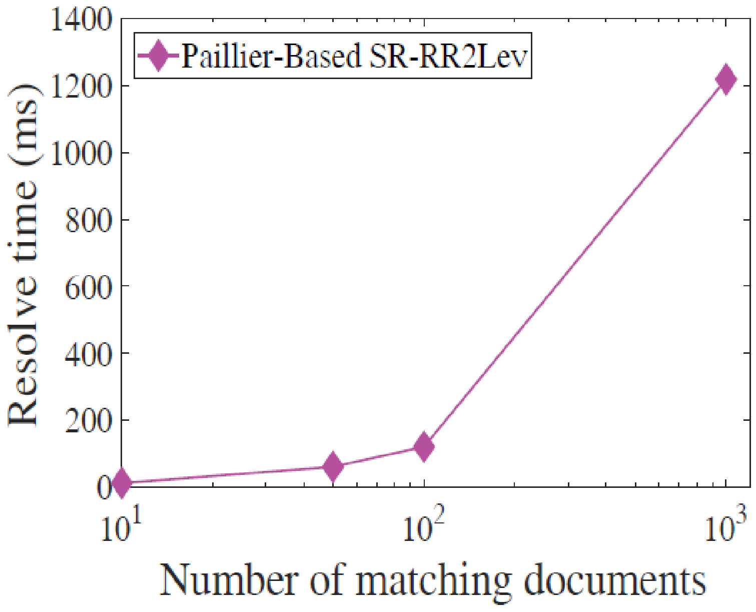
- <u>Ciphertexts:</u>
  - OPE → 64-bit
  - Paillier → 1024-bit*

* NIST recommendation, https://www.keylength.com/en/4/

Benchmark results of schemes based on RR2Lev for a single keyword query

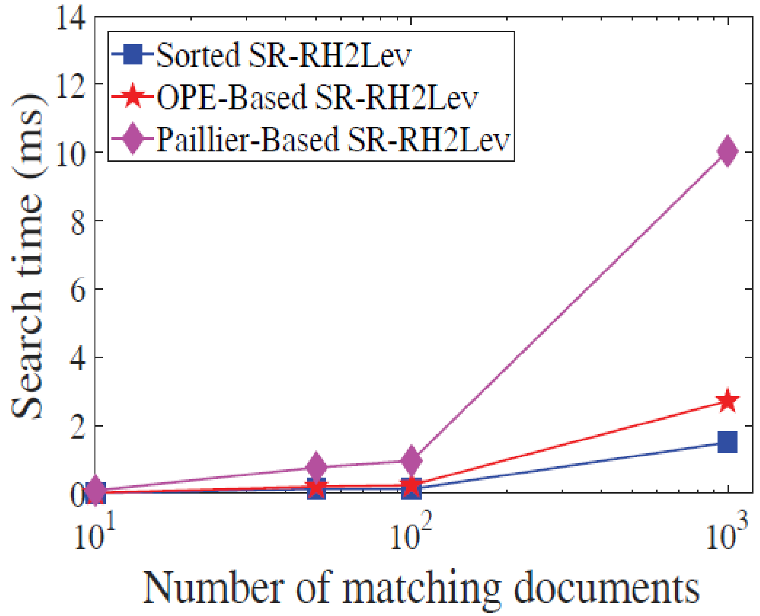


(a)

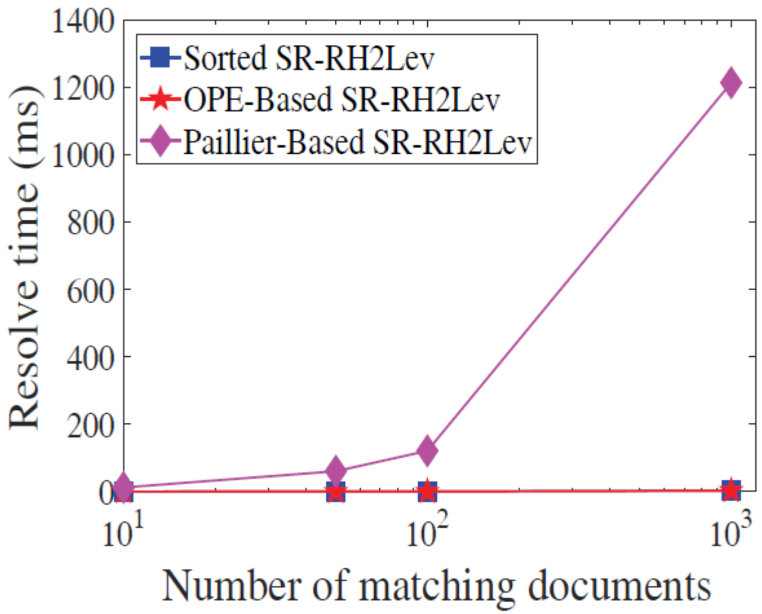


(b)

Benchmark results of schemes based on RH2Lev for a single keyword query

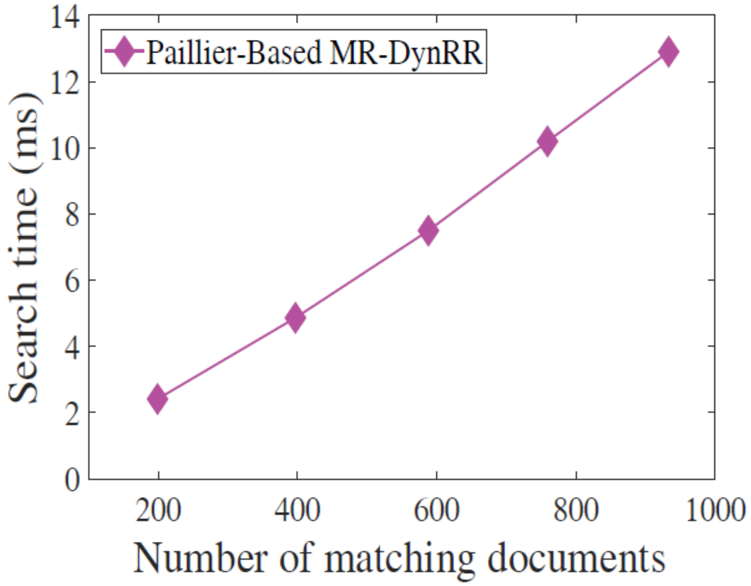


(a)

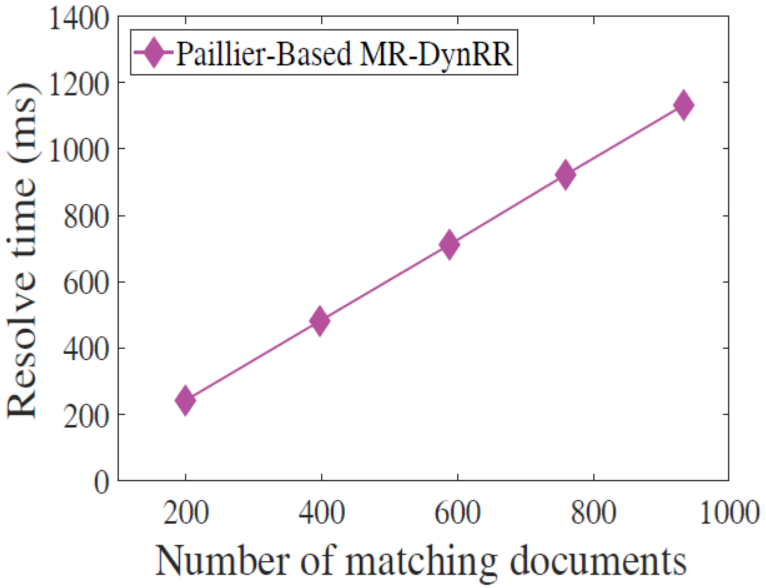


(b)

Benchmark results of scheme based on DynRR for a multi-keyword query



(a)



(b)

Conclusion

- Approaches for single- and multi-keyword ranked searches
- The extended schemes → the properties also differ.
- Advantages and disadvantages of each scheme

Thanks for your attention! Any questions?