The 1st International Conference on Ubiquitous Security (UbiSec 2021)

The 6th International Workshop on Trusted Computing (IWTC 2021)

Secure Search in Cloud Computing

Presenter: Qin Liu

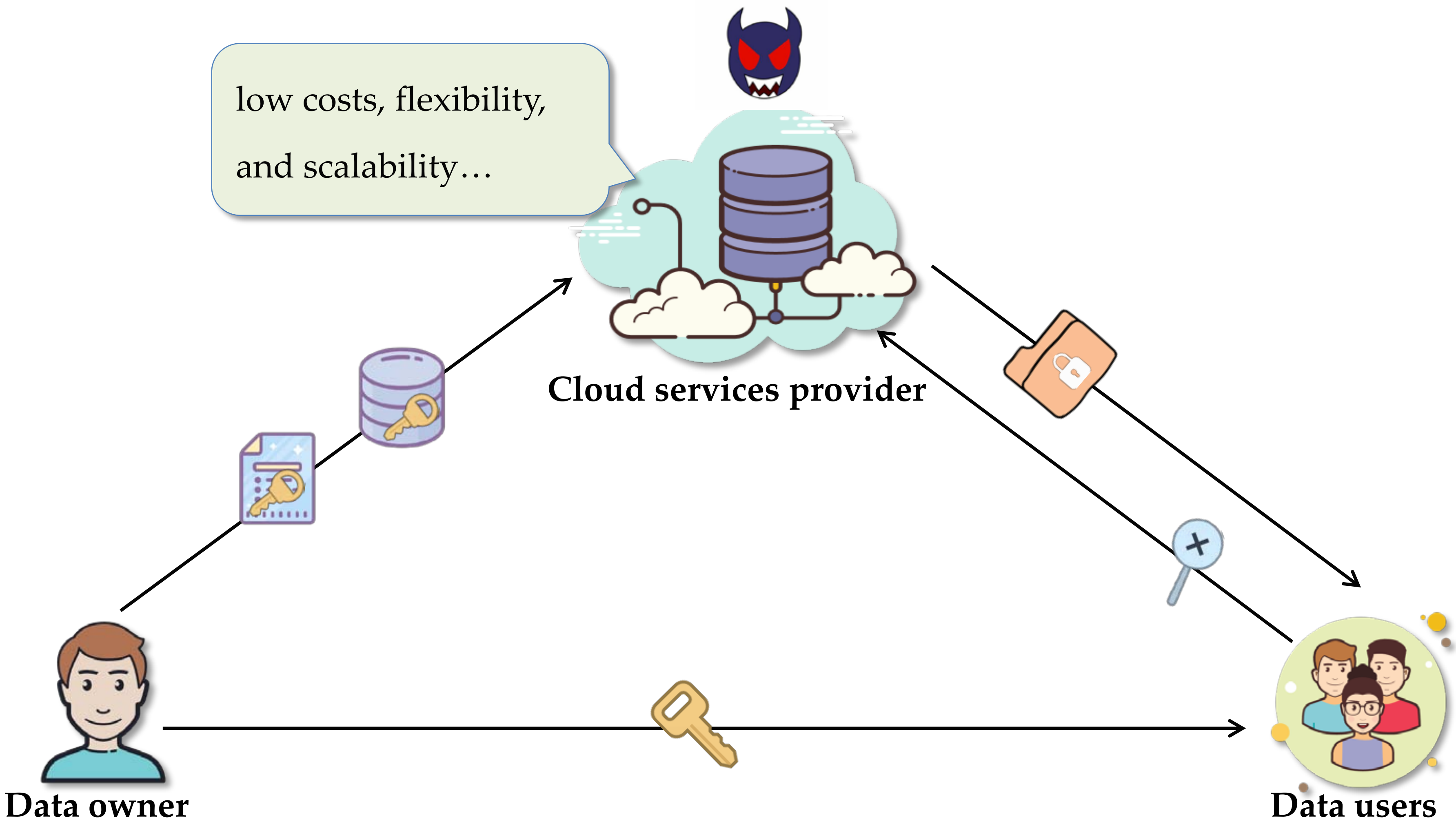
Hunan University

Dec. 30, 2021

Outline

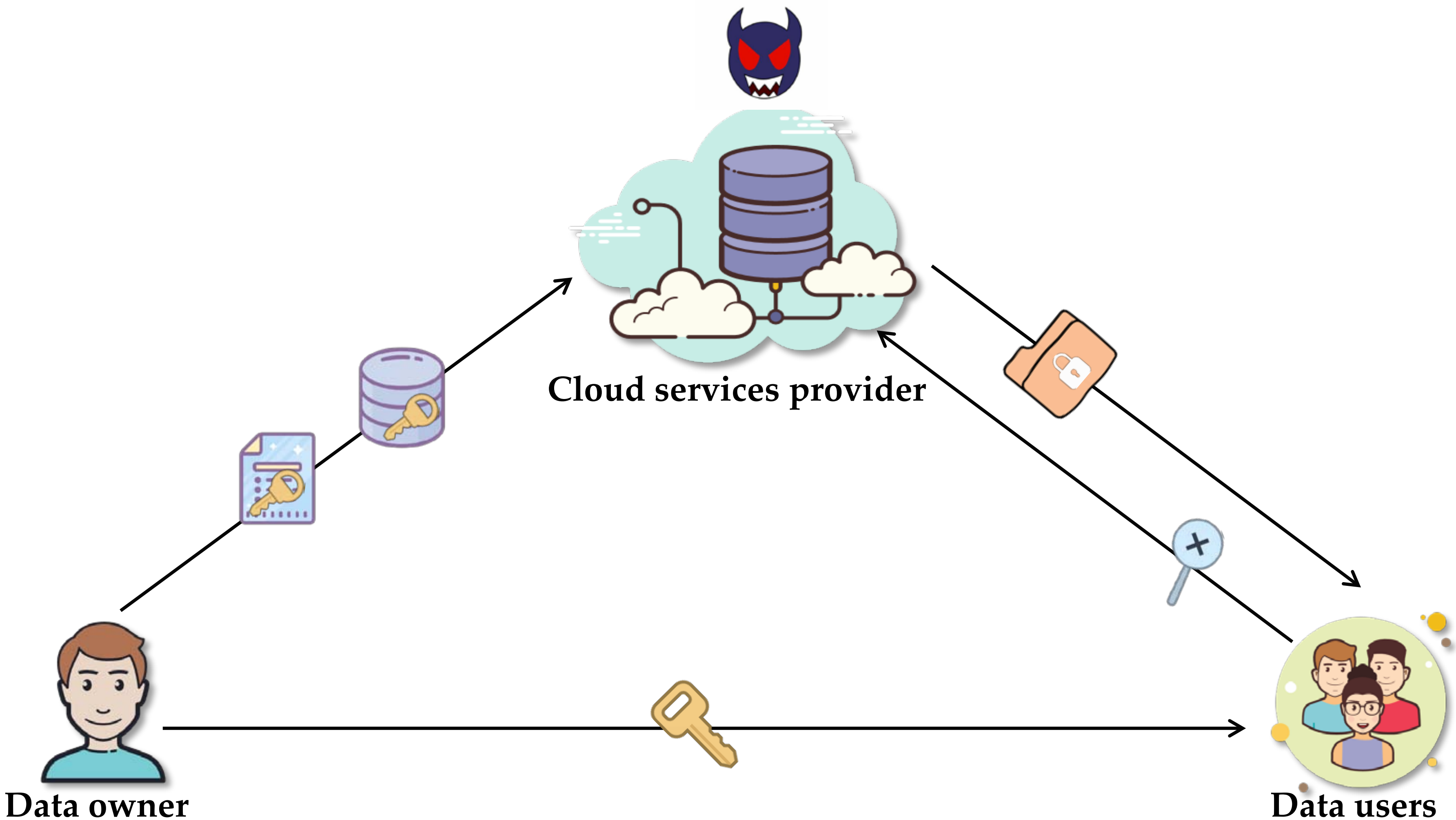
1. Introduction
2. Previous Work
3. Ongoing Work
4. Future Work

Introduction



- **Searchable encryption (SE)** is a tool that allows the cloud server to perform secure searches over encrypted data

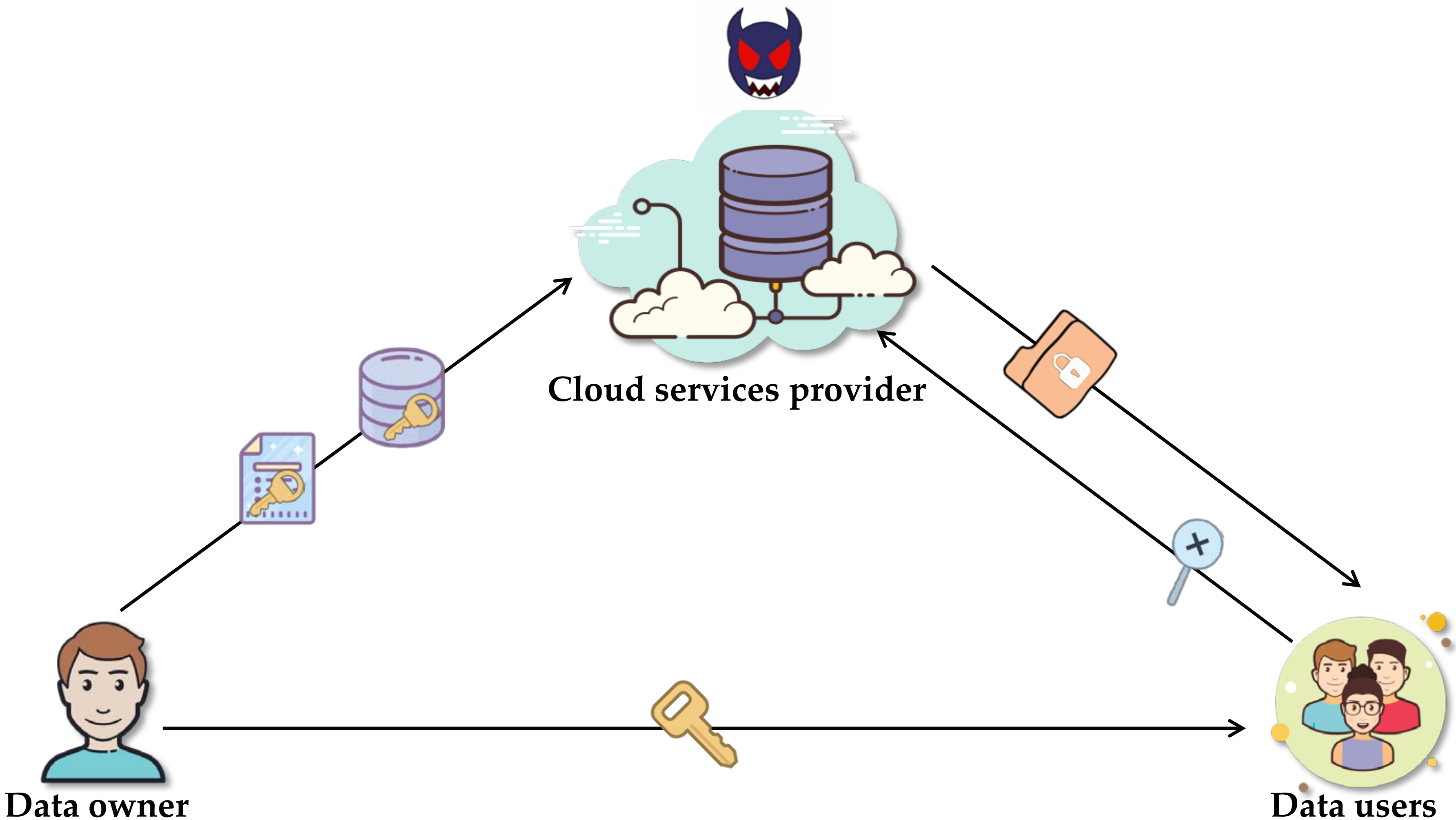
Introduction



Searchable Encryption

Symmetric key setting: the keys encrypting the index and the token are the same
Asymmetric key setting: the keys encrypting the index and the token are different

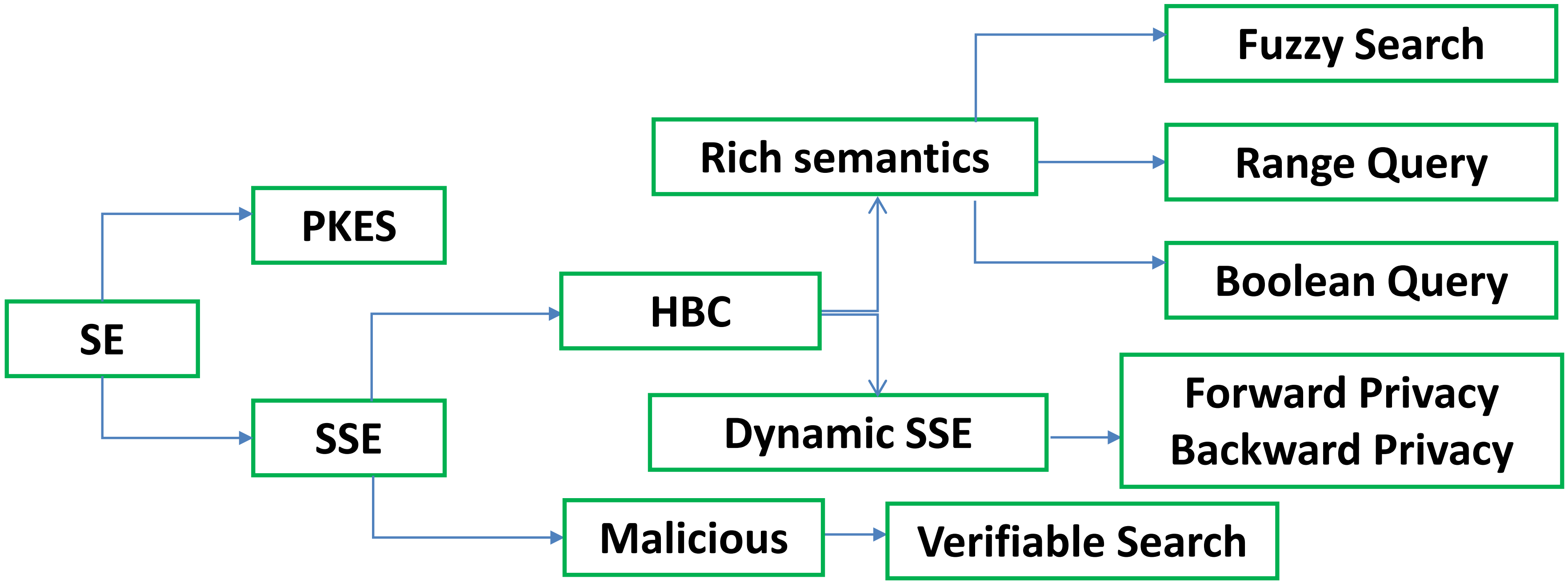
Introduction



Common Leakage in SSE

Access pattern: which files have been returned
Search pattern: whether two searches were performed for the same keyword

Overview of Our Work



Q. Liu, , Y. Peng, J. Wu, T. Wang, and G. Wang, “Secure Multi-Keyword Fuzzy Searches with Enhanced Service Quality in Cloud Computing” IEEE Transactions on Network and Service Management (TNSM), 2020

Q. Liu, Y. Tian, J. Wu, T. Peng, and G. Wang, "Enabling Verifiable and Dynamic Ranked Search Over Outsourced Data," IEEE Transactions on Services Computing(TSC), 2019.

Q. Liu, X. Nie, X. Liu, T. Peng, and J. Wu, "Verifiable Ranked Search over Dynamic Encrypted Data in Cloud Computing," Proc. of IWQoS 2017.

based on Comparable Inner Product Encoding," Proc. of CNS 2018.

L. Du, K. Li*, Q. Liu*, Z. Wua, S. Zhang, "Dynamic Multi-Client Searchable Symmetric Encryption with Support for Boolean Queries, Information Sciences," 2019.

B. Hu, **Q. Liu**, X. Liu, T. Peng, G. Wang, J. Wu, "DABKS: Dynamic Attribute-based Keyword Search in Cloud Computing," Proc. of ICC 2017.

Outline

1. Introduction
2. Previous Work
3. Ongoing Work
4. Future Work

IEEE Transactions on Services Computing 2020

Prime Inner Product Encoding for Effective Wildcard-based Multi-Keyword Fuzzy Search

Qin Liu[a], Yu Peng[a], Shuyu Pei[a], Jie Wu[b], Tao Peng[c] and Guojun Wang[c]

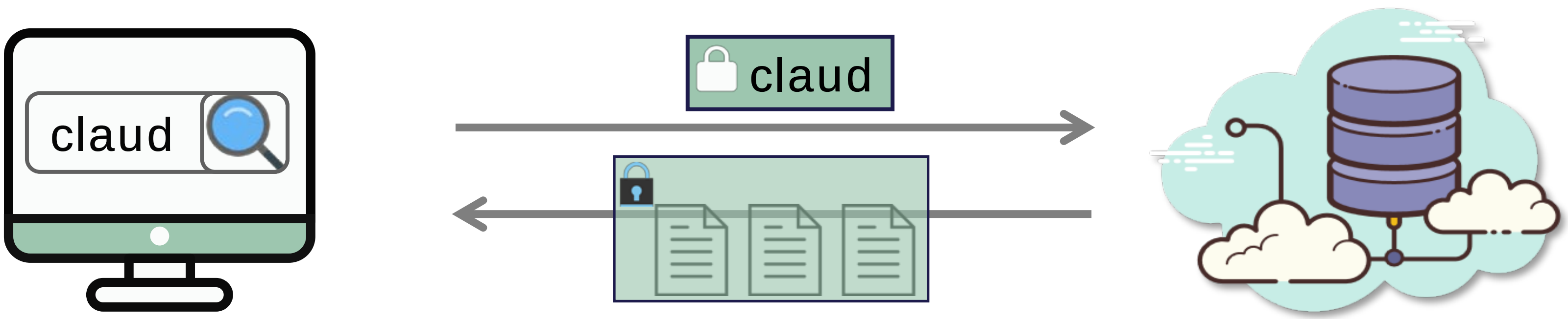
[a] Hunan university

[b] Temple university

[c] Guangzhou university

Introduction to Secure Fuzzy search

Alice wants to retrieve files containing keyword “cloud” from cloud servers.



The misspelling of a query keyword will cause an error result to be returned.

- **Fuzzy search.** The tolerance of misspelling o
- **Flexibility.** The user can specify different sea
- **Efficiency.** An efficient index structure to facilitate parallel searches.

- **Inverted index**

O(r)

w_1	ind1	ind2	ind3
w_2	ind2	ind3	
w_3	ind1	ind2	

Related work on Secure Fuzzy Search

INFOCOM2010 [1]

single keyword, predefined set

S_{CASTLE}

$= \{CASTLE, * CASTLE,$

$* ASTLE, C * ASTLE, C$

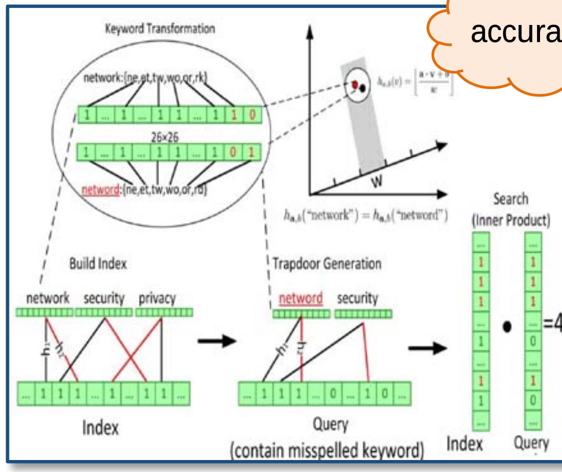
$* STLE, \cdots, CASTL * E, CASTL$

$*, CASTLE *\}$

Pre-set edit distance = 1

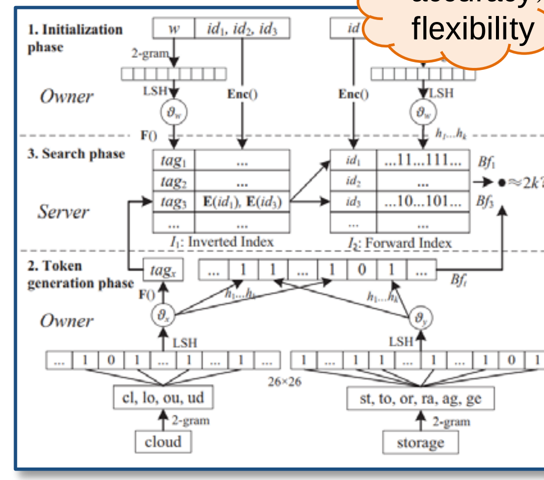
INFOCOM2014 [2]

accuracy



TSC2017 [4]

accuracy, flexibility



Schemes	Multi-keyword fuzzy Search	Flexibility	Indexes	Building blocks
INFOCOM2010 [1]	✗	-	-	Predefined set
INFOCOM2014 [2]	√	✗	Forward index	LSH、bloom filter
TIFS2016 [3]	√	✗	Forward index	LSH、bloom filter
TSC2017 [4]	√	✗	Forward + inverted index	LSH、bloom filter
TDSC2019 [5]	√	✗	tree	-
JNCA2020 [6]	√	✗	tree	LSH、bloom filter

Contributions of Our Work

The Prime Inner Product Encoding (PIPE) Scheme

Main idea

- Encoding a query keyword or an index keyword into a vector filled with primes or reciprocals of primes, such that the result of vectors' inner product is an integer only when two keywords are similar.

Compared with Previous Fuzzy Search Schemes

- Greater flexibility. Vectors are organized into prime-related matrices to support multi-semantic queries.
- Higher efficiency. A keyword balanced binary (KBB) tree is built to support parallelizable and dynamic search.
- Enhanced robustness. A query matrix is extended by random noises to resist linear analysis attacks.

Basic scheme: Prime Inner Product Encoding ($PIPE_0$)

Files	keywords
D_1	{“hello”, “key”}

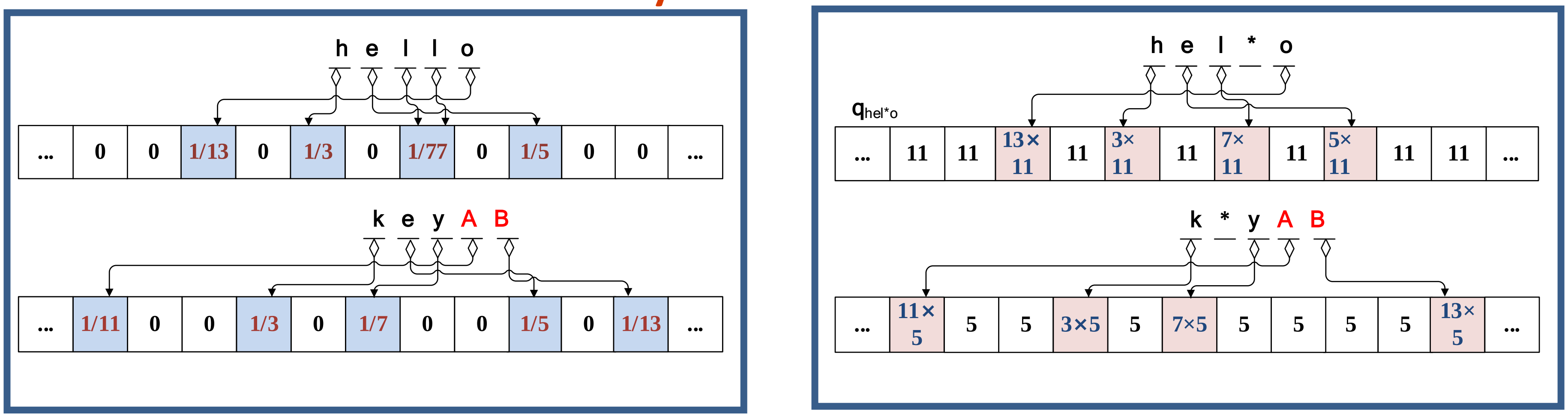
$\mathcal{P}$ $3, 5, 7, 11, 13$

A, B, C, D, E

Dummy characters

Max length of universal keywords $L = 5$

hel*o, k*y



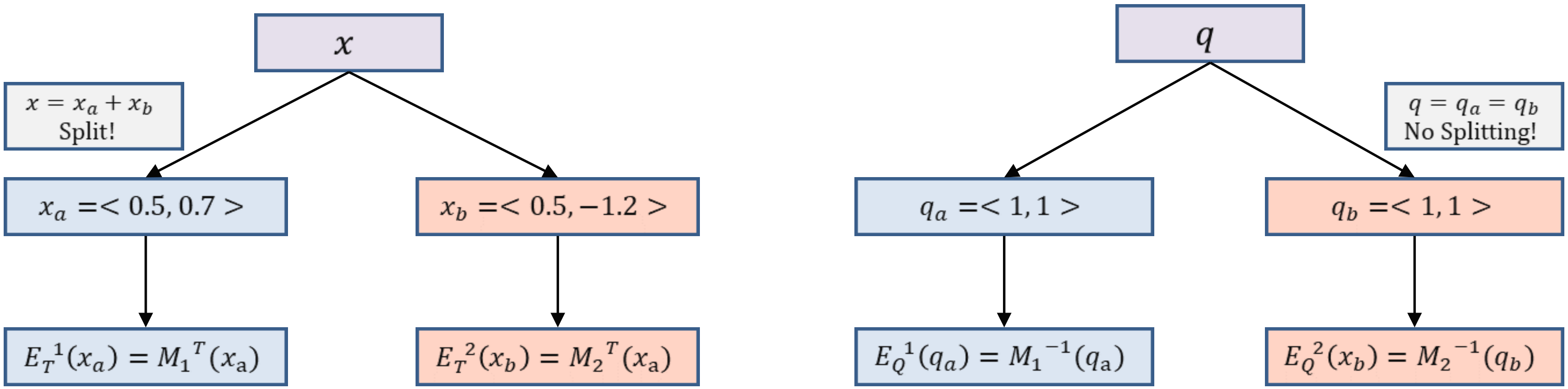
- The inner products between the index vectors and the query vectors

$$R = \begin{bmatrix} \mathbf{p}_{hello} \cdot \mathbf{q}_{hel*o} = 34, & \mathbf{p}_{hello} \cdot \mathbf{q}_{k*y} \approx 3.11 \\ \mathbf{p}_{hello} \cdot \mathbf{q}_{k*y} = 3.12, & \mathbf{p}_{key} \cdot \mathbf{q}_{k*y} \approx 21 \end{bmatrix}$$

- For AND queries, if each column of R has at least one integer, the query q matches the file D
- For OR queries, if at least one element in R is an integer, the query q matches the file D

Secure KNN

$x =< 1, -0.5 >, q =< 1,1 >$; ensure $x \cdot q = 0.5$ can be recovered on E(DB)



$$E_T^1(x_a) \cdot E_Q^1(q_a) + E_T^2(x_b) \cdot E_Q^2(q_b)$$
$$= x_a \cdot q_a + x_b \cdot q_b = (0.5 + 0.7) + (0.5 - 1.2) = 0.5$$
$$= x \cdot q$$

secret key:

(1)M_1; (2)M_2;

(3)the random splitting configuration

$M_1 * M_1^{-1} = I$

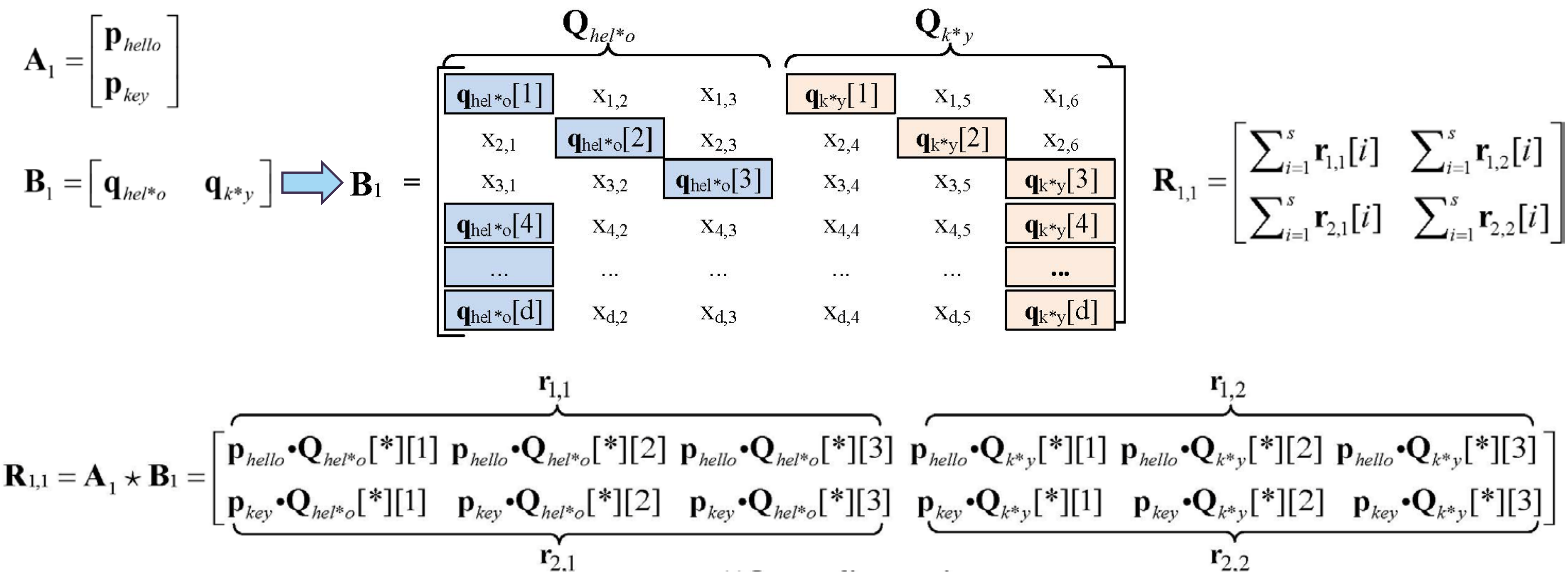
$M_2 * M_2^{-1} = I$

- The multiplication of the plaintext matrices to be calculated based on their encrypted forms.

W. K. Wong, D. W.-l. Cheung, B. Kao, et al, “Secure knn computation on encrypted databases,” in Proc. of SIGMOD, 2009.

Advanced scheme: PIPE_S

- Secure KNN failed to resist linear analyses. **ICDE2013 [6]**



> Each column of matrix Q contains at least one element of vector q.

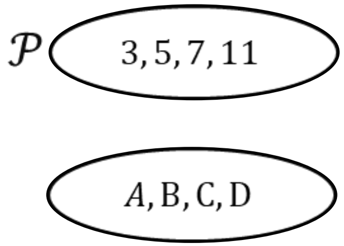
> The sum of the random numbers at the l-th row, denoted as δ_l, is equal to $t_l q[l]$ where $t_l = 0$ or $(t_l + 1)$ is a prime outside primes set $\mathcal{P}$.

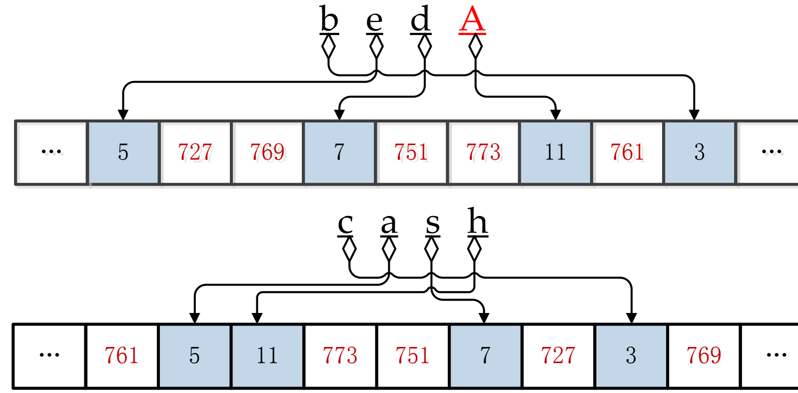
$\alpha = p \cdot q + X$, where $X \in R$ is a random number that has no linear relationship with the result of $p \cdot q$. Therefore, it is impossible for the adversary to decompose $p \cdot q$ from α.

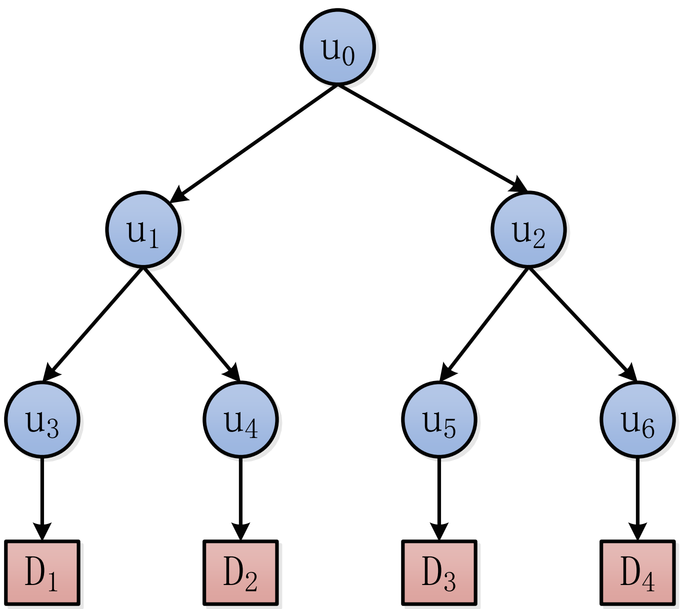
Tree-based Index

Files	keywords
D_1	{“bed”, “cash”}
D_2	{“cash”}
D_3	{“cat”, {“pen”}
D_4	{“love”}

Max length of universal keywords $L = 4$







U=<nid, data, fid, lchild, rchild>

u3. data

…	5	727	769	7	751	773	11	761	3	…
…	761	5	11	773	751	7	727	3	769	…

u4. data

…	761	5	11	773	751	7	727	3	769	…
…	1	1	1	1	1	1	1	1	1	…

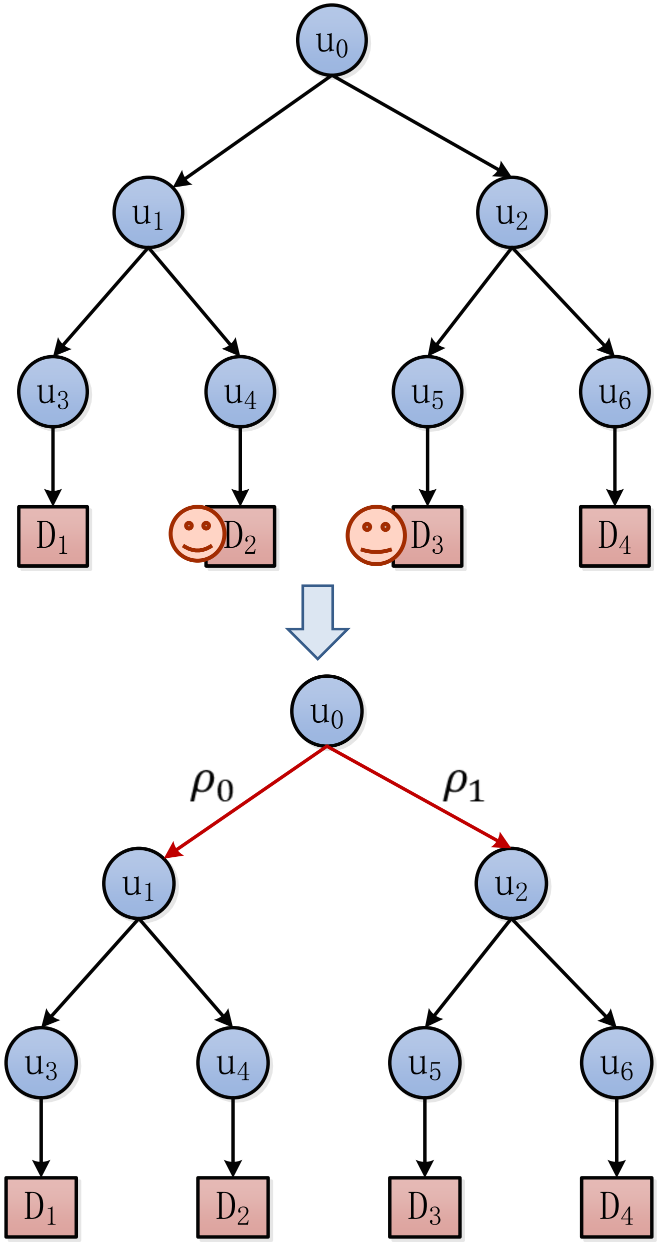
u1. data

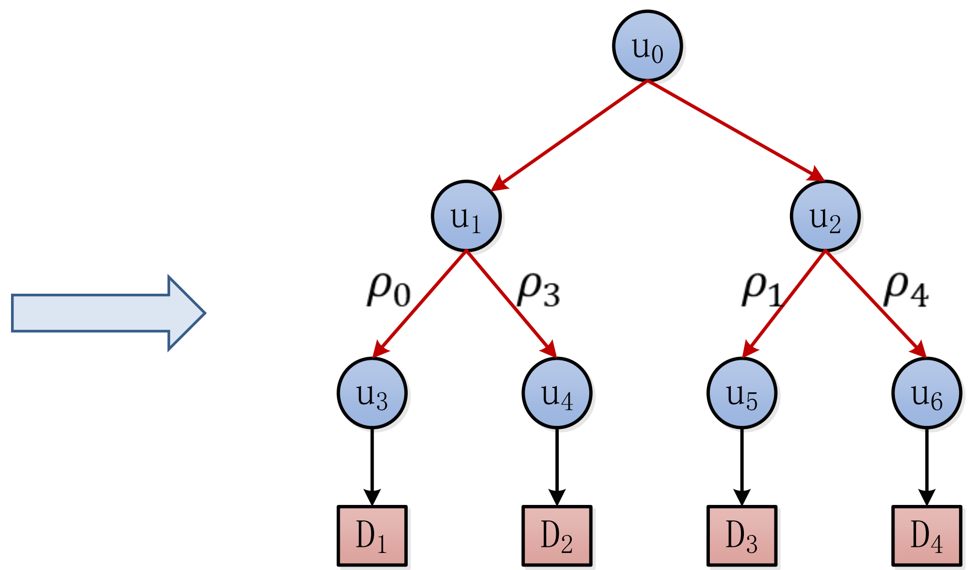
…	5× 761	727 ×5	769 ×11	7× 773	751	773 ×7	11× 727	761 ×3	3× 769	…
…	761	5	11	773	751	7	727	3	769	…

Parallel search

Let $P = \{\rho_0, \rho_1, \rho_2, \rho_3\}$ be a set of 4 available processors in the system

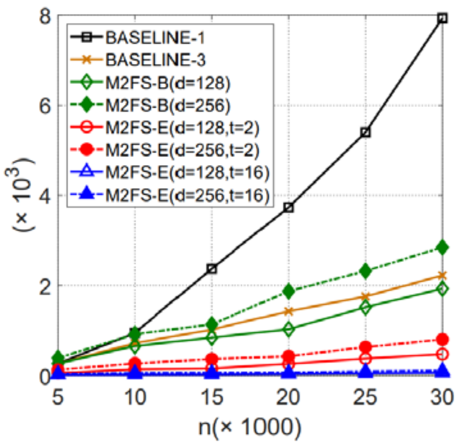
$$O\left(\frac{r \log n}{t}\right)$$



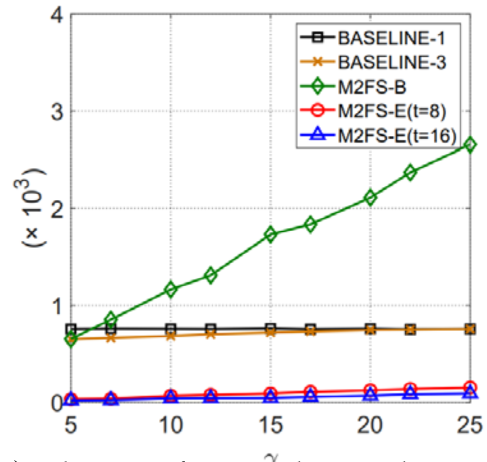


Evaluation

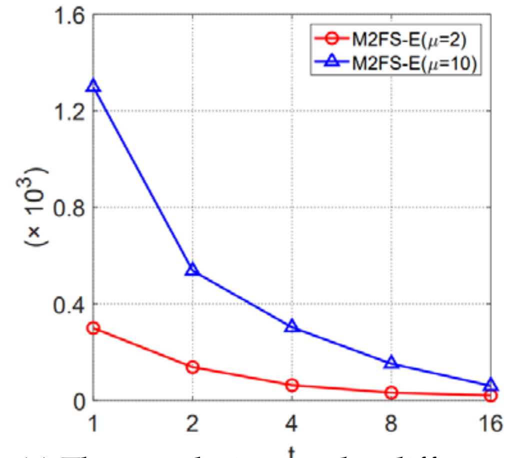
- Comparison of the execution time (*ms*) for AND queries



（a）The time for searching *n* files

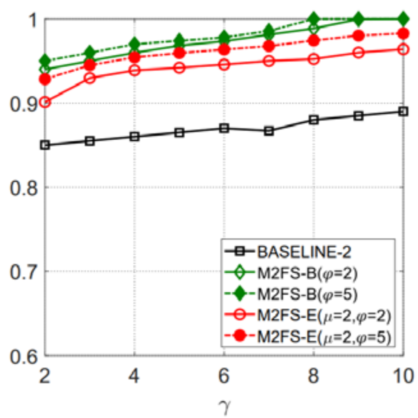


（b）The time for searching γ keywords

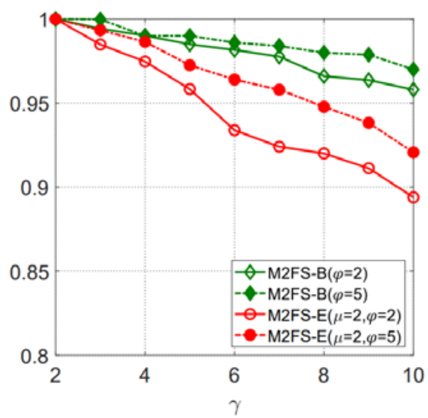


(c) The search time under different t

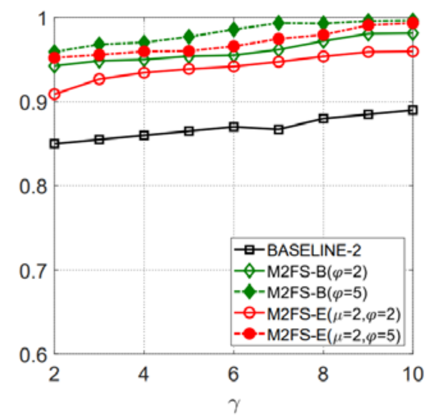
- Comparison of the execution time (*ms*) for AND queries



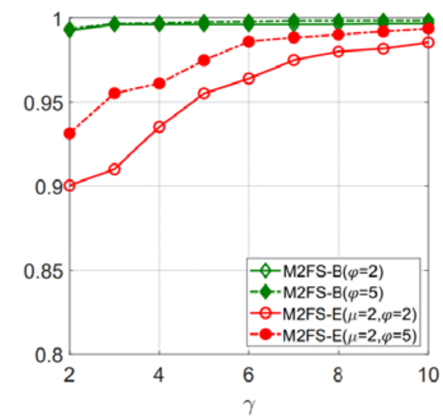
(a) Precision of AND queries



(b) Recall of AND queries



(c) Precision of OR queries



(d) Recall of OR queries

References

[1] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in Proc. of INFOCOM, 2010. . **[INFOCOM2010]**

[2] B. Wang, S. Yu, W. Lou, and Y. T. Hou, “Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud,” in Proc. of INFOCOM, 2014. **[INFOCOM2014]**

[3] Z. Fu, X. Wu, C. Guan, X. Sun, and K. Ren, “Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement,” IEEE Transactions on Information Forensics and Security, 2016. **[TIFS2016]**

[4] J. Chen, K. He, L. Deng, Q. Yuan, R. Du, Y. Xiang, and J. Wu, “EliMFS: achieving efficient, leakage-resilient, and multi-keyword fuzzy search on encrypted cloud data,” IEEE Transactions on Services Computing, 2017. **[TSC2017]**

[5] Q. Liu, Y. Peng, S. Pei, J. Wu, T. Peng and G. Wang, "Prime Inner Product Encoding for Effective Wildcard-based Multi-Keyword Fuzzy Search," IEEE Transactions on Services Computing, 2020. **[TSC2020]**

[6] B. Yao, F. Li, and X. Xiao, “Secure nearest neighbor revisited,” in Proc. of ICDE, 2013. **[ICDE2013]**

IEEE CNS 2018

Secure and Efficient Multi-Attribute Range Queries based on Comparable Inner Product Encoding

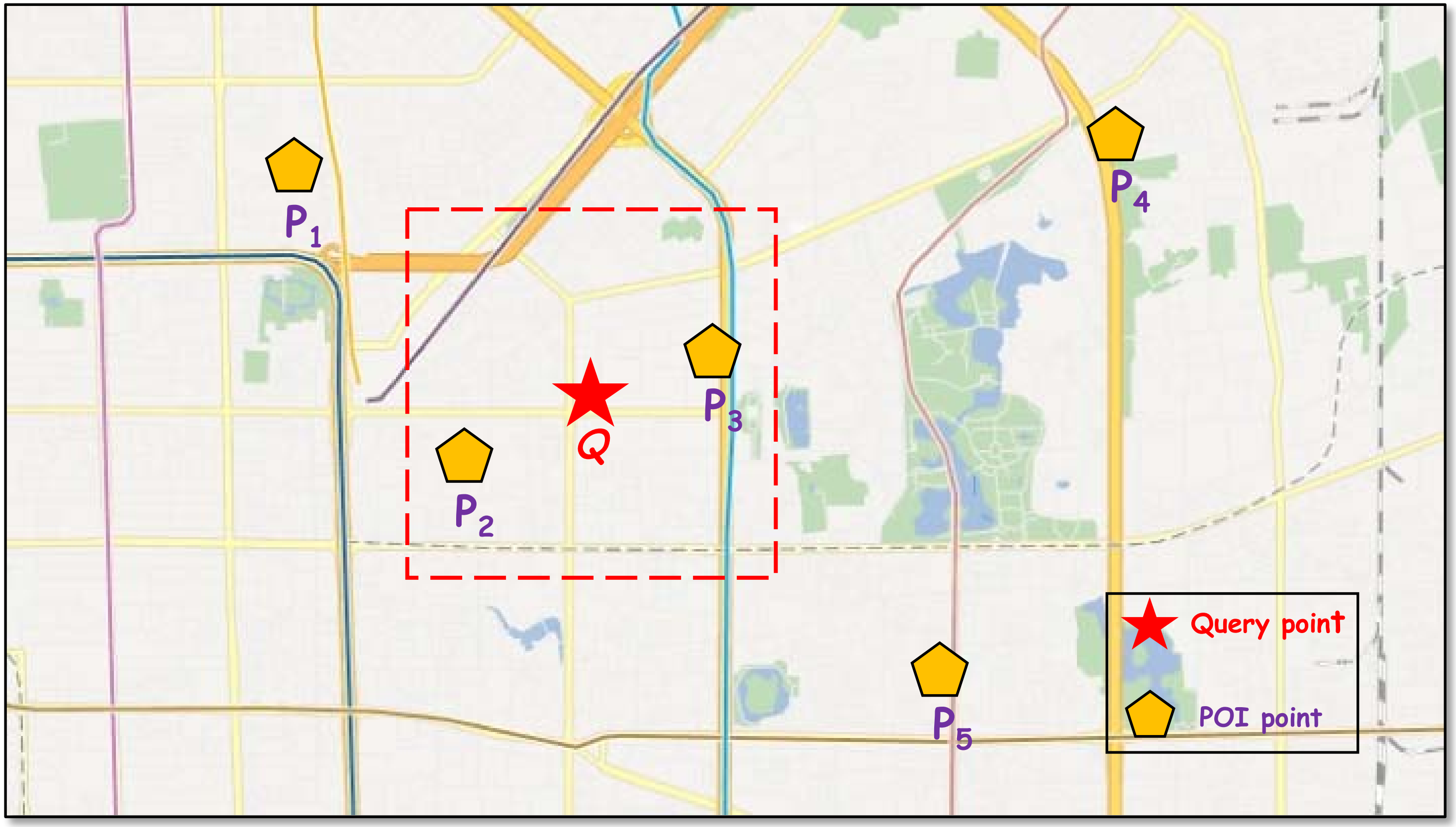
Qin Liu[a], SiXia Wu[a], Shuyu Pei[a], Jie Wu[b], Tao Peng[c] and Guojun Wang[c]

[a] Hunan university

[b] Temple university

[c] Guangzhou university

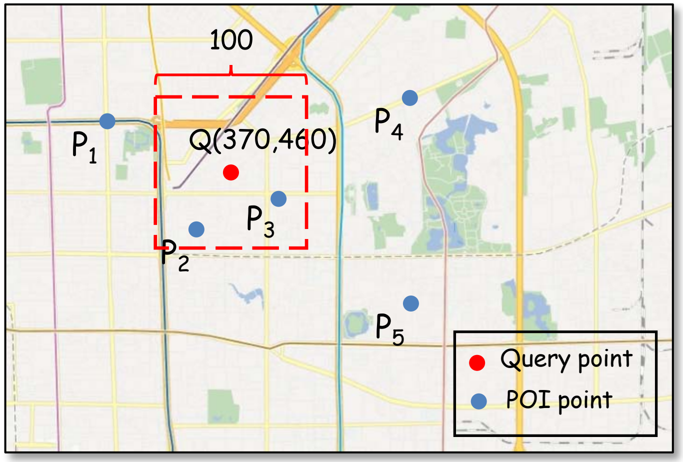
Introduction to Secure Range Query



Location based services(LBS)

- LBS uses of location technology to obtain the current location of the device and provides query services through the mobile Internet.
- E.g. **range query** or ***k*****NN query.**

Introduction to Secure Range Query



Point	x-coordinate	y-coordinate
P_1	300	480
P_2	350	420
P_3	400	440
P_4	450	520
P_5	450	300

- 2-dimensional range query is used in Location Based Services(LBS).
- E.g. Q =(370, 460) and edge length = 100, the result of range query is $\{P_2, P_3\}$.
- Besides, multi-dimensional range query has wide application prospect.
  - (Age in [20,40] AND Blood Pressure in [100, 130] AND Weight in [60, 80])

Challenge in secure range query: Comparisons need to be performed based on ciphertextes!

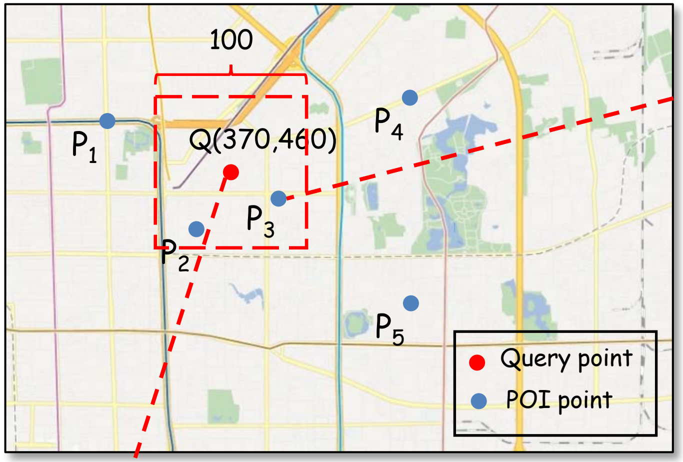
Related work on Secure Range Query

Schemes	Efficiency	Scalability	Security	Privacy
Most of OPE	✔	✔	✗	✔
Ideal OPE	✗	✔	✔	✔
ORE	✔	✔	✗	✔
Homomorphic	✗	✔	✔	✔
CIPE scheme	✔	✔	✔	✔

Contributions

- **Enhanced security.** It can resist inference attacks that existing OPE schemes are vulnerable to.
- **Higher efficiency.** It needs only around **1.4s** on average while performing two-attribute range queries on **1 million** encrypted data records.

Basic Scheme: $CIPE_0$



$P_3(400,440)$

$\mathbf{p}_{3x}$	400	400	1	1
$\mathbf{p}_{3y}$	440	440	1	1

Index vector constructions

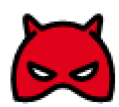
$$\mathbf{p}_{3x} \cdot \mathbf{q}_{xl} = 2 \times (-400 + 320) < 0$$
$$\mathbf{p}_{3x} \cdot \mathbf{q}_{xu} = 2 \times (-400 + 420) > 0$$
$$\mathbf{p}_{3y} \cdot \mathbf{q}_{yl} = 2 \times (-440 + 310) < 0$$
$$\mathbf{p}_{3y} \cdot \mathbf{q}_{yu} = 2 \times (-440 + 510) \geq 0$$

√

Query([320,420], [410,510])

$\mathbf{q}_{xl}$	-1	-1	320	320
$\mathbf{q}_{xu}$	-1	-1	420	420
$\mathbf{q}_{yl}$	-1	-1	410	410
$\mathbf{q}_{yu}$	-1	-1	510	510

Query vector constructions



- The distance between attribute values
- The equality of attribute values

Noise addition!

$$v_{k,l} = p_{i_k} \bullet q_{j_{k,l}} = (\sum C + \sum \overline{C})(b_{j_{k,l}} - a_{i_k})$$

> 0

Advanced scheme: $CIPE_s$

- Secure KNN has been proved unable to resist chosen plaintext attacks(CPA)

$$p \cdot q = p'_{|\alpha|} \cdot q'_{|\alpha|} + p'_{|\beta|} \cdot q'_{|\beta|}$$

Extend Query matrices!

$\overline{C}$ =(0.5, 0) and $\overline{C}$ =(0, 0.01)

-0.5	0	165	0	330	330	-1	-1	$\mathbf{q}_{1,l}$
0	-0.01	0	4.3	430	430	-1	-1	$\mathbf{q}_{1,u}$

-0.15	1.105	-1
-2	0	2
-2	-32.65	49.5
0	-3	3
99	-63.3	-6
99	-300	230.7
-0.3	0.21	0
-0.3	0	0.21

(a) $\mathbf{Q}_{1,l}$ **(d=8, s=3).**

Set r = 0.3 and t = - 0.21

0	1	0	-1	2	4	-3	-3
-2	-0.02	0.006	-1	1	-4	5	1
3.1	1	0	-2.58	1.9	-2	-3	-1
-5	-3	-2	8.6	0	10	0	0
2	4	-3	-258	860	-3	-1	1
2.5	3.5	-4	-1	-258	860	1.5	-2.5
3	-6	3	-0.5	-1.5	0.6	-2	2
3	-3	0	1	-4	3	0.6	-2

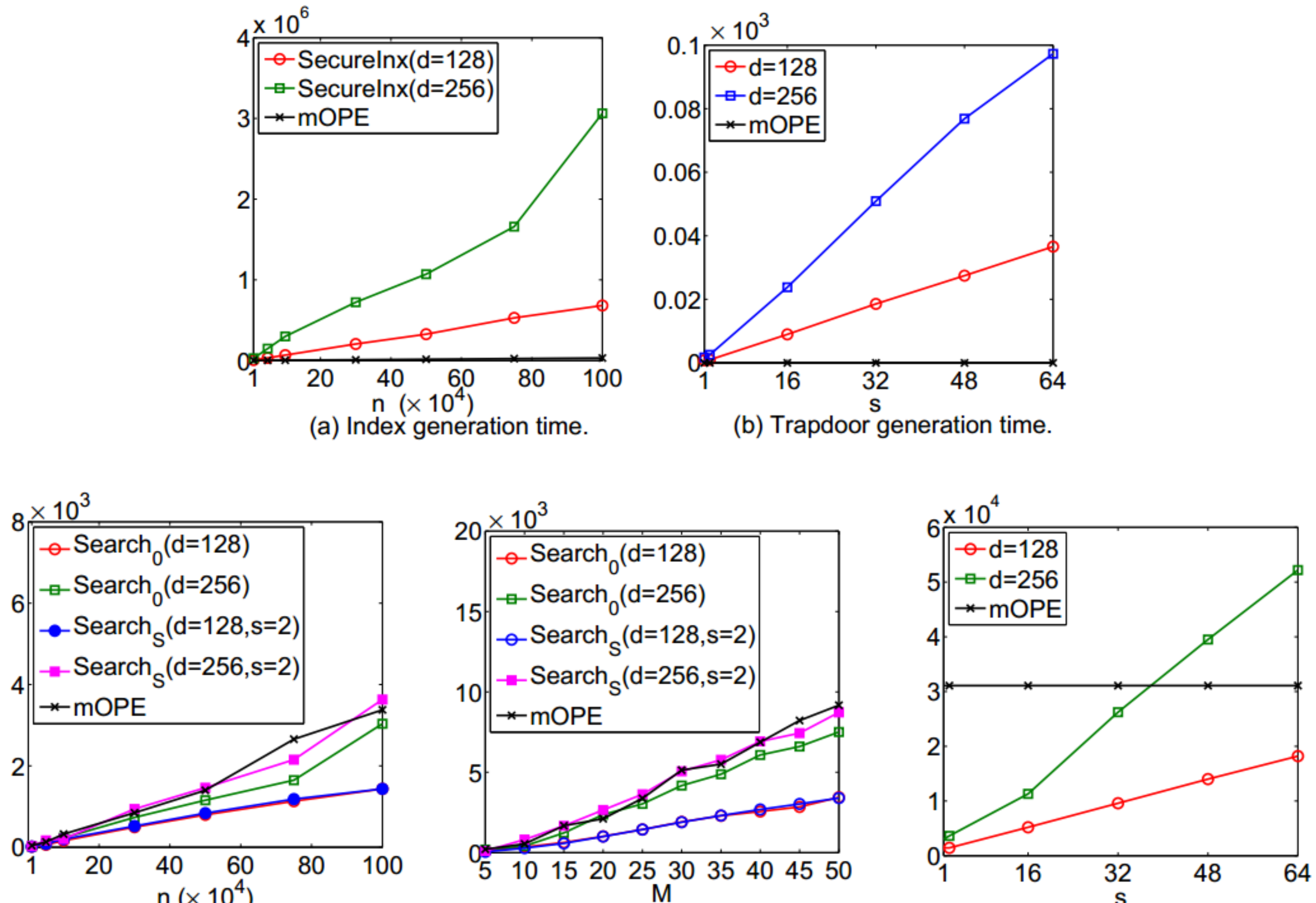
(b) $\mathbf{Q}_{1,u}$ **(d=8, s=8).**

Set r = 2 and t = -0.6

$$p \cdot q \neq p'_{|\alpha|} \cdot q'_{|\alpha|} + p'_{|\beta|} \cdot q'_{|\beta|}$$

Evaluation

- Comparison of the execution time (ms) between CIPE and mOPE



(a) Index generation time.
(b) Trapdoor generation time.
(c) Search time.
(d) Search time.
(e) Search time.

R. A. Popa, F. H. Li, and N. Zeldovich, “An ideal-security protocol for order-preserving encoding,” in Proc. of S&P, 2013

Outline

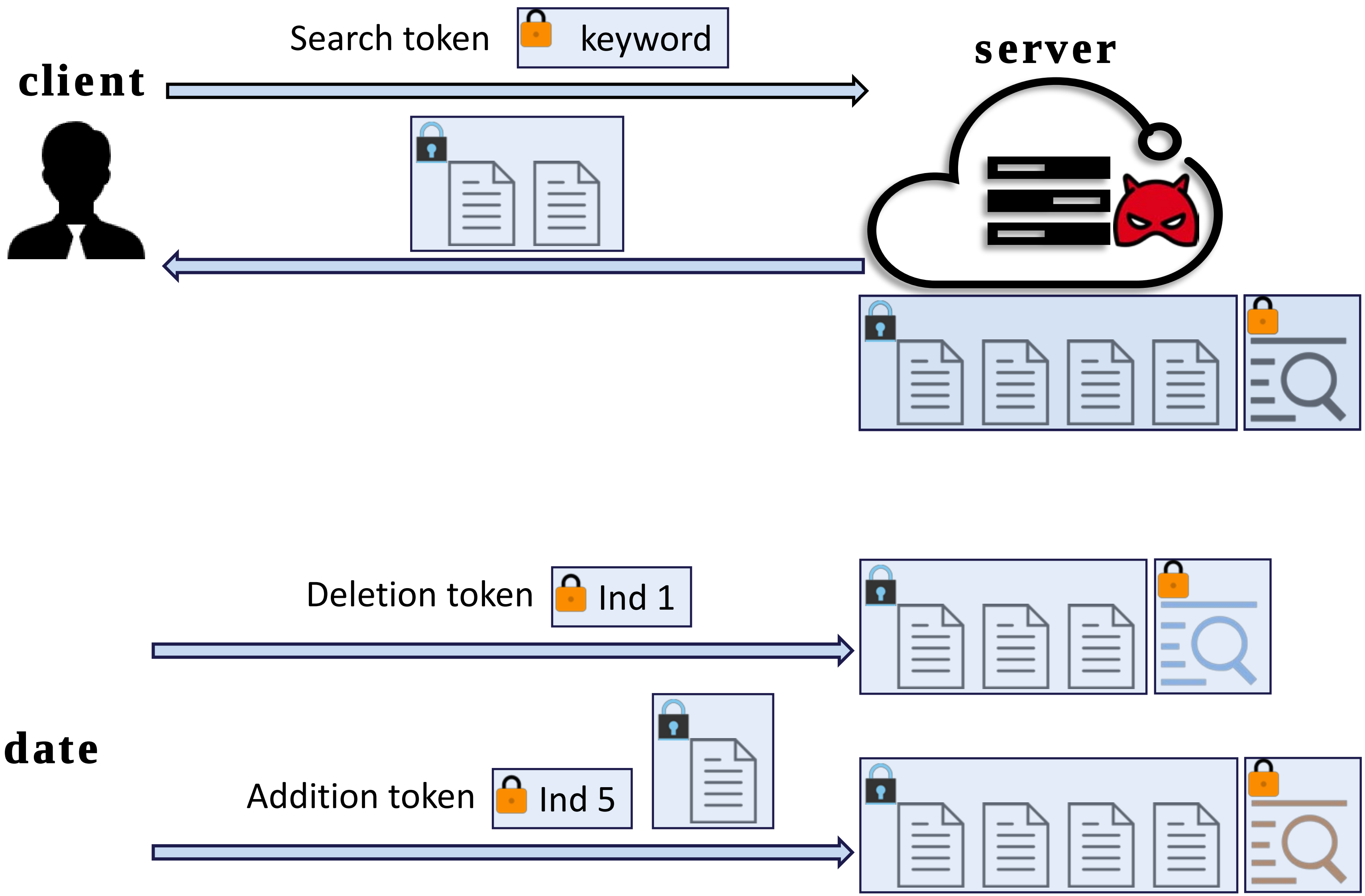
1. Introduction
2. Previous Work
3. Ongoing Work
4. Future Work

Ongoing Work

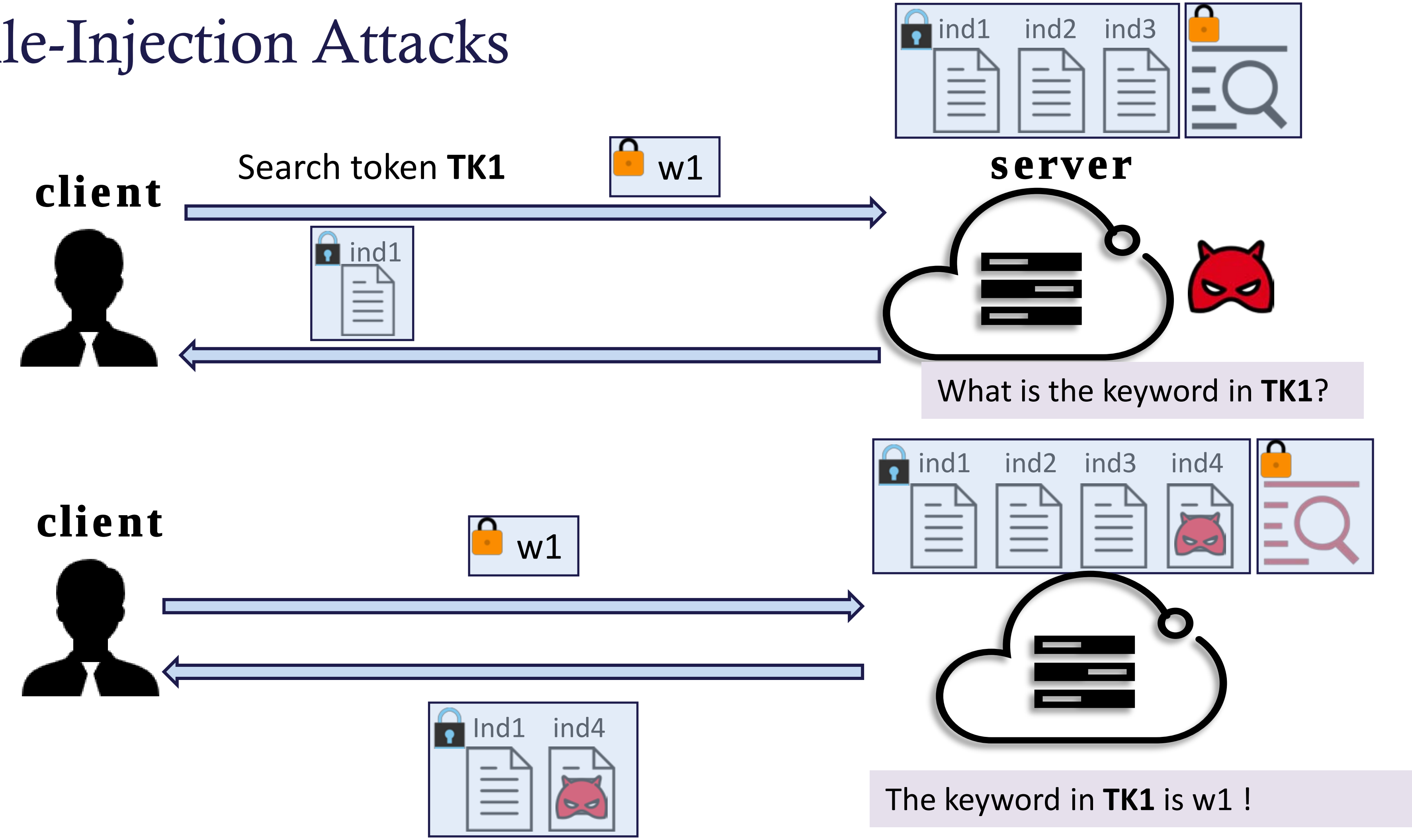
Forward and Backward Privacy of DSSE

Qin Liu[a], Yu Peng[a], Hongbo Jiang[a], Guojun Wang[b] , Tian Wang[c], and Jie Wu[d]

Dynamic searchable symmetric encryption (DSSE)



File-Injection Attacks



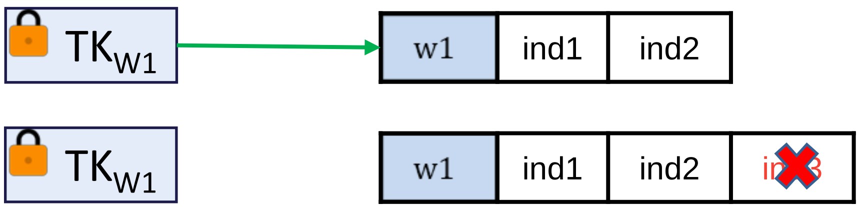
Forward Privacy is required in DSSE to resist file injection attack !

- **Why query privacy is important?**
  - Keywords are part of the files. File content can be recovered.
  - Keywords can be used to classify files and help other attacks.

Y. Zhang et al, " All your queries are belong to us: the power of file-injection attacks on searchable encryption," in Proc. of USENIX, 2016.

Forward Privacy (FP)

- Forward privacy (FP) requires that the newly added files cannot be linked to previous search tokens.



2005 — Chang et al.[1]
- The first FP construction
- High communication cost due to the ever-growing query size

2014 — Stefanov et al.[2]
- The first FP scheme with sublinear search time
- Based on oblivious RAM
- High communication cost

2016 — Bost et al.[3]
- Formally defined forward privacy
- Low communication cost
- High computation cost of asymmetric cryptography

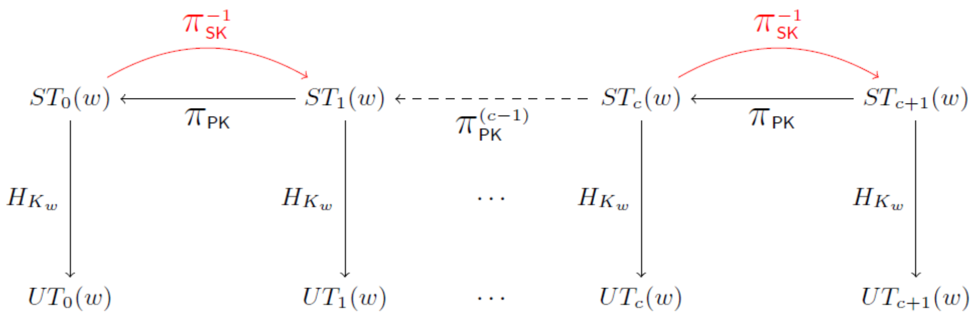
[1] Chang et al."Privacy preserving keyword searches on remote encrypted data. " in Proc. ACNS, 2005.

[2] Stefanov et al."Practical Dynamic Searchable Encryption with Small Leakage. " in Proc. NDSS, 2014.

[3] Bost et al."Σοφος: Forward secure searchable encryption." in Proc. CCS, 2016.

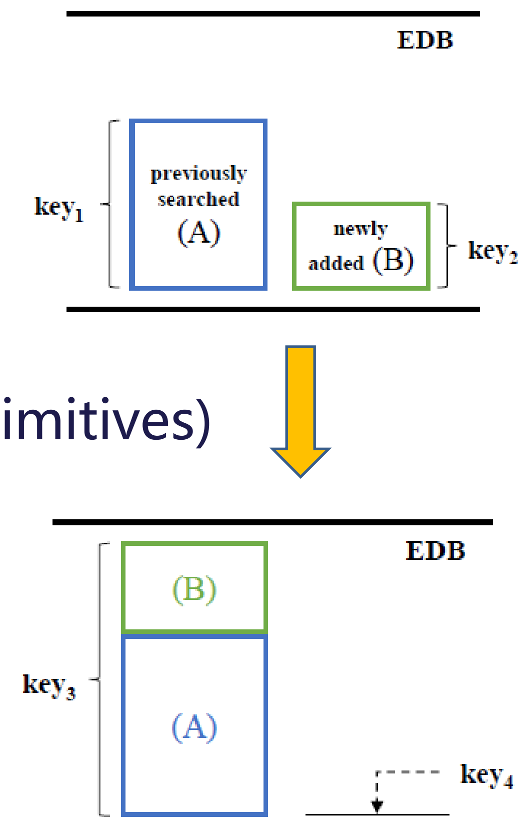
State-of-the-art FP Schemes

- Sophos: Trapdoor permutation (TDP)



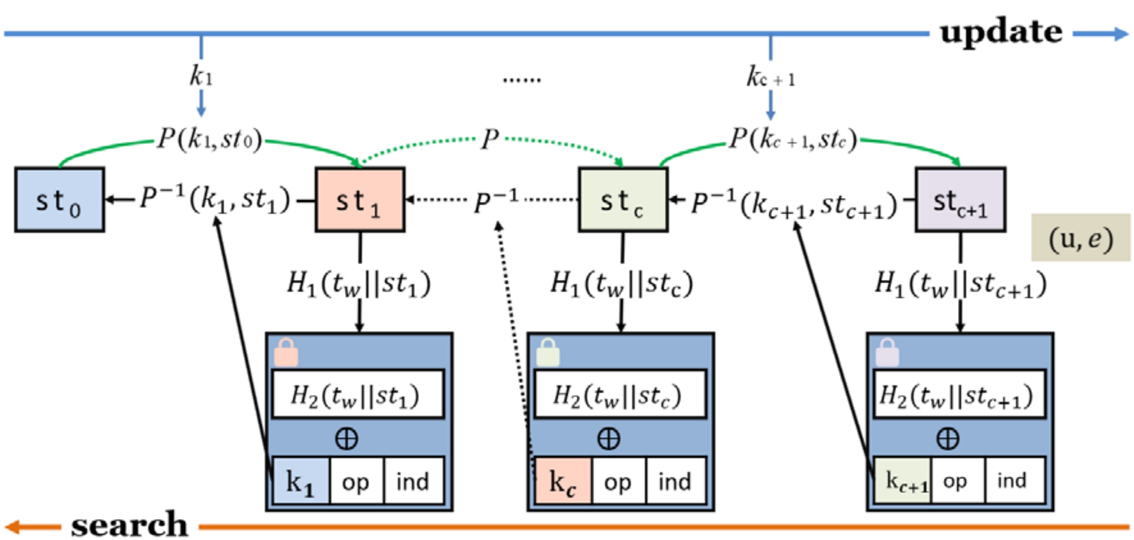
Relations among tokens. Operations in red can only be done by the client, using the secret key SK

- Dual dictionary



Re-Encryption
Storage cost

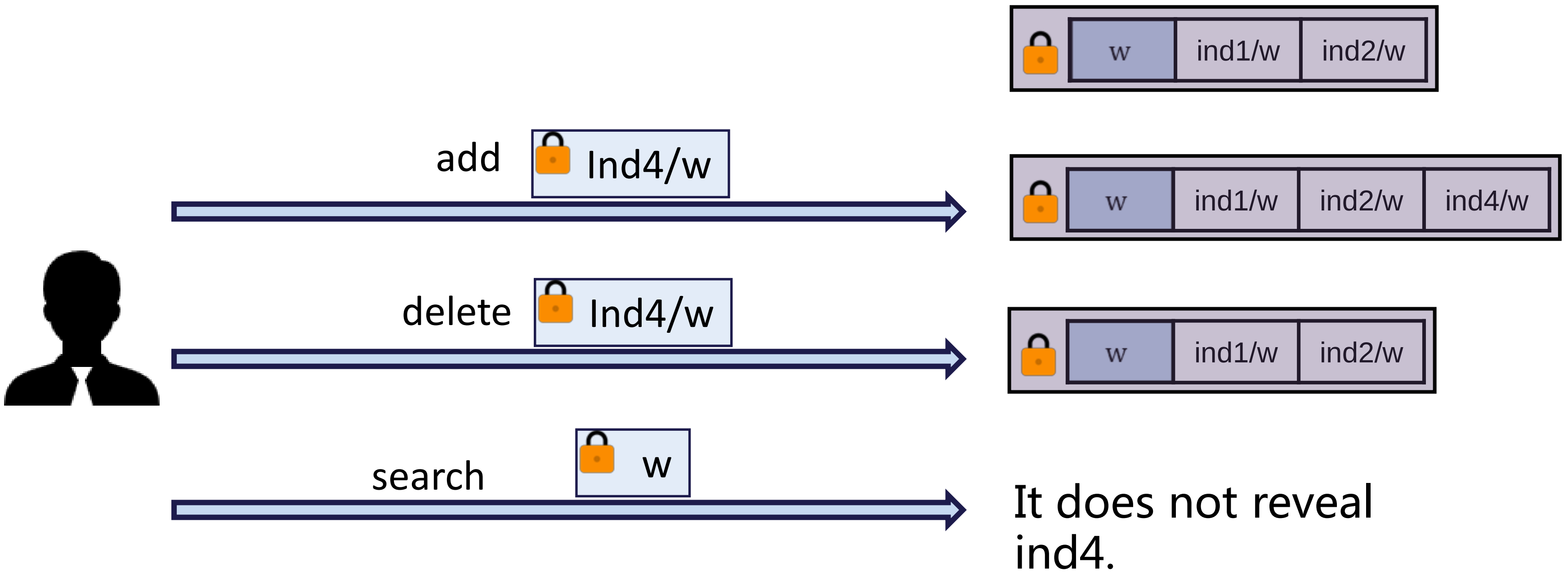
- Fast: Pseudorandom permutation (symmetric primitives)



Lack of actual deletion

Backward Privacy (BP)

- **Backward privacy (BP):** the deleted files cannot be searched any more.

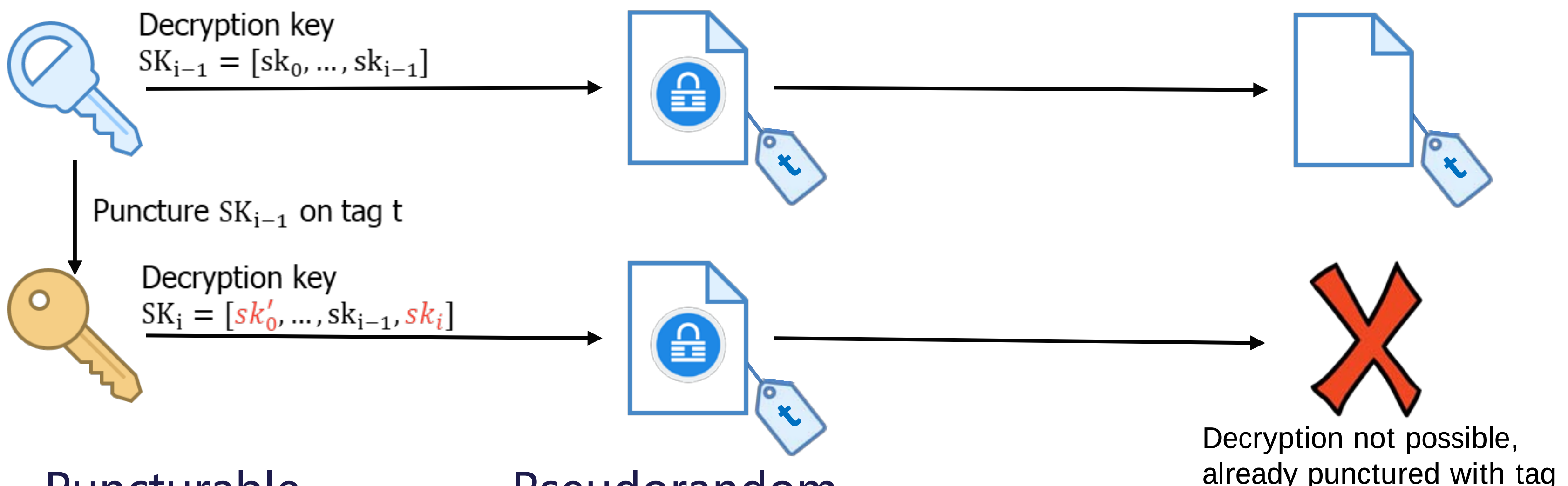


- Backward privacy: the deleted files cannot be searched.

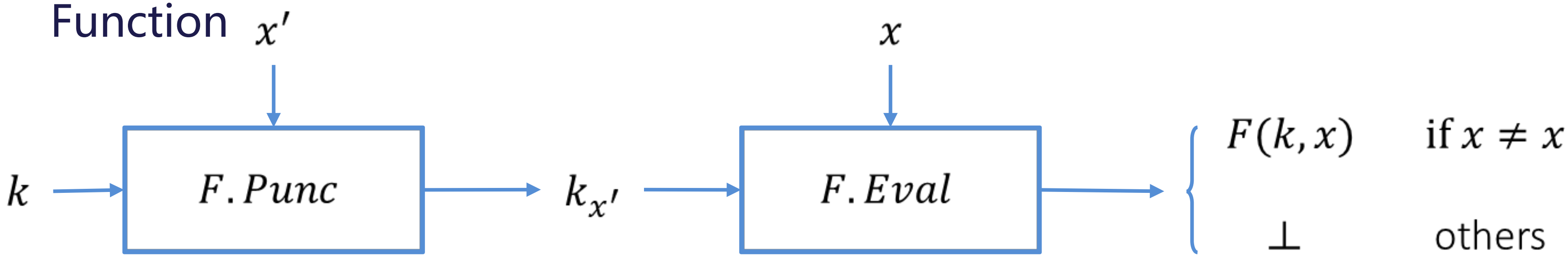
[1] Bost, Raphaël, Brice Minaud, and Olga Ohrimenko. "Forward and backward private searchable encryption from constrained cryptographic primitives," in Proc. of CCS, 2017.

State-of-the-art- BP Schemes

- Constrained pseudorandom function (CPRF)
- Public-key Puncturable Encryption



- Puncturable Pseudorandom Function



- Symmetric Puncturable Encryption (SPE)
- FP + 2 round trip

Related Work on Forward & Backward Privacy

Schemes	Forward privacy	Backward privacy	Search round trip	Building blocks
Sophos [1]	√	✗	-	Trapdoor permutation
FAST [2]	√	✗	-	Pseudorandom function
Dual [3]	√	✗	-	Dual dictionary
$\text{Diana}_{\text{del}}$ [4]	√	BP-3	2	Constrained pseudorandom functions
Janus [4]	√	BP-3	1	Puncturable encryption
Janus++ [6]	√	BP-3	1	Symmetric puncturable encryption
Fish-bone [7]	√	BP-3	2	Symmetric key encryption
Fides [4]	√	BP-2	2	From Sophos
Mitra [8]	√	BP-2	2	-
Moneta [4]	√	BP-1	3	obvious RAM
Orion [8]	√	BP-1	O(log N)	obvious RAM

Our Scheme with FP & BP

- **Basic scheme (FP)**
  - **A hybrid index structure** that incorporates the merits of both inverted indexes and forward indexes, but is much more simple and efficient.

- **Advanced scheme (FP+BP)**
  - Hybrid index + Symmetric Puncturable Encryption (SPE)
  - File-based BP

ind	$head^{ind}$
ind1	
ind2	
…	

Our Future Work

Secure Search in Emerging Computing

Thanks for your attentions

Email: gracelq628@hnu.edu.cn