USENIX Security ‘24

d-DSE: Distinct Dynamic Searchable Encryption Resisting Volume Leakage in Encrypted Databases

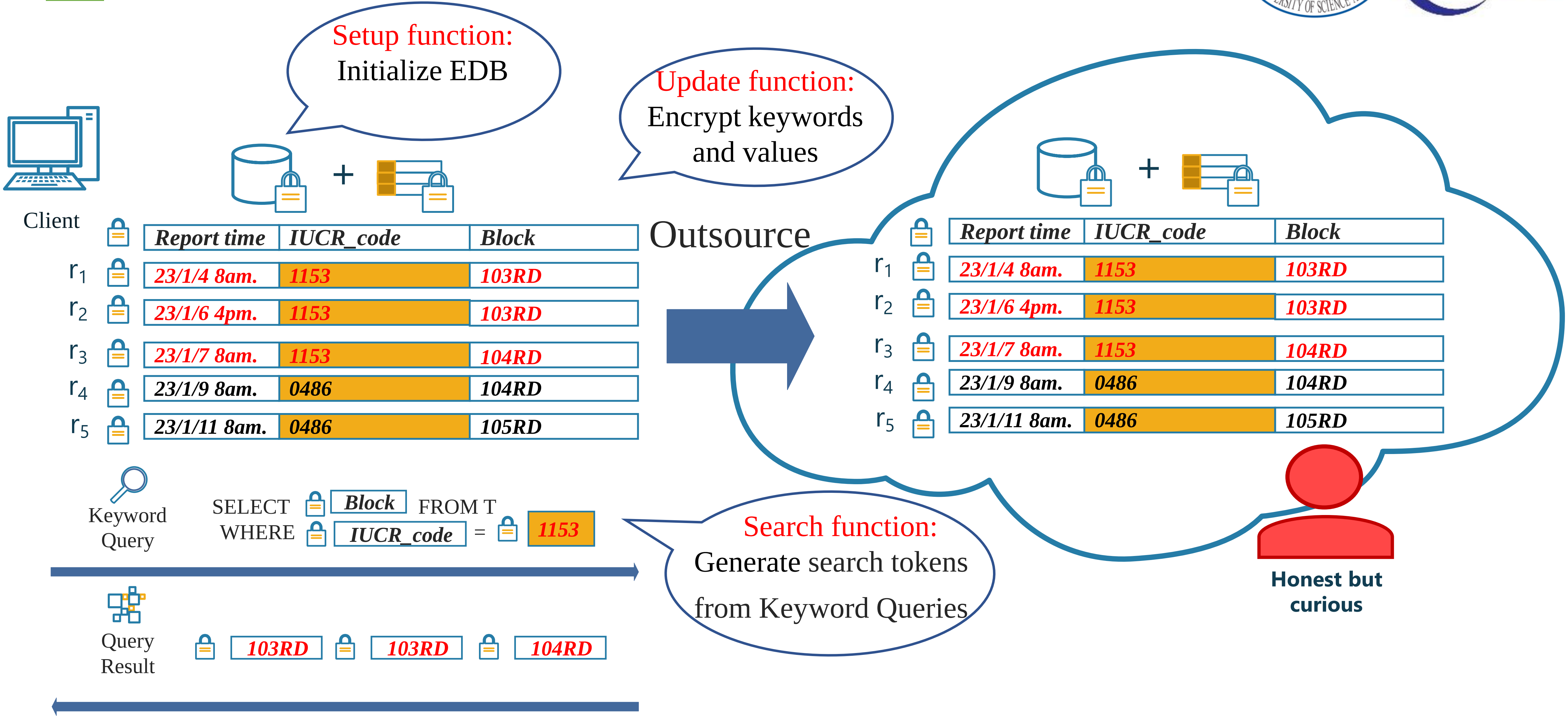
Dongli Liu, Wei Wang, Peng Xu, Laurence T. Yang, Bo Luo[+], Kaitai Liang*

Huazhong University of Science and Technology

[+]*The University of Kansas*

**Delft University of Technology*

What is DSE applied in EDB?



Setup function:
Initialize EDB

Update function:
Encrypt keywords and values

Client

	Report time	IUCR_code	Block
r_1	23/1/4 8am.	1153	103RD
r_2	23/1/6 4pm.	1153	103RD
r_3	23/1/7 8am.	1153	104RD
r_4	23/1/9 8am.	0486	104RD
r_5	23/1/11 8am.	0486	105RD

Outsource

	Report time	IUCR_code	Block
r_1	23/1/4 8am.	1153	103RD
r_2	23/1/6 4pm.	1153	103RD
r_3	23/1/7 8am.	1153	104RD
r_4	23/1/9 8am.	0486	104RD
r_5	23/1/11 8am.	0486	105RD

Keyword Query

SELECT Block FROM T
WHERE IUCR_code = 1153

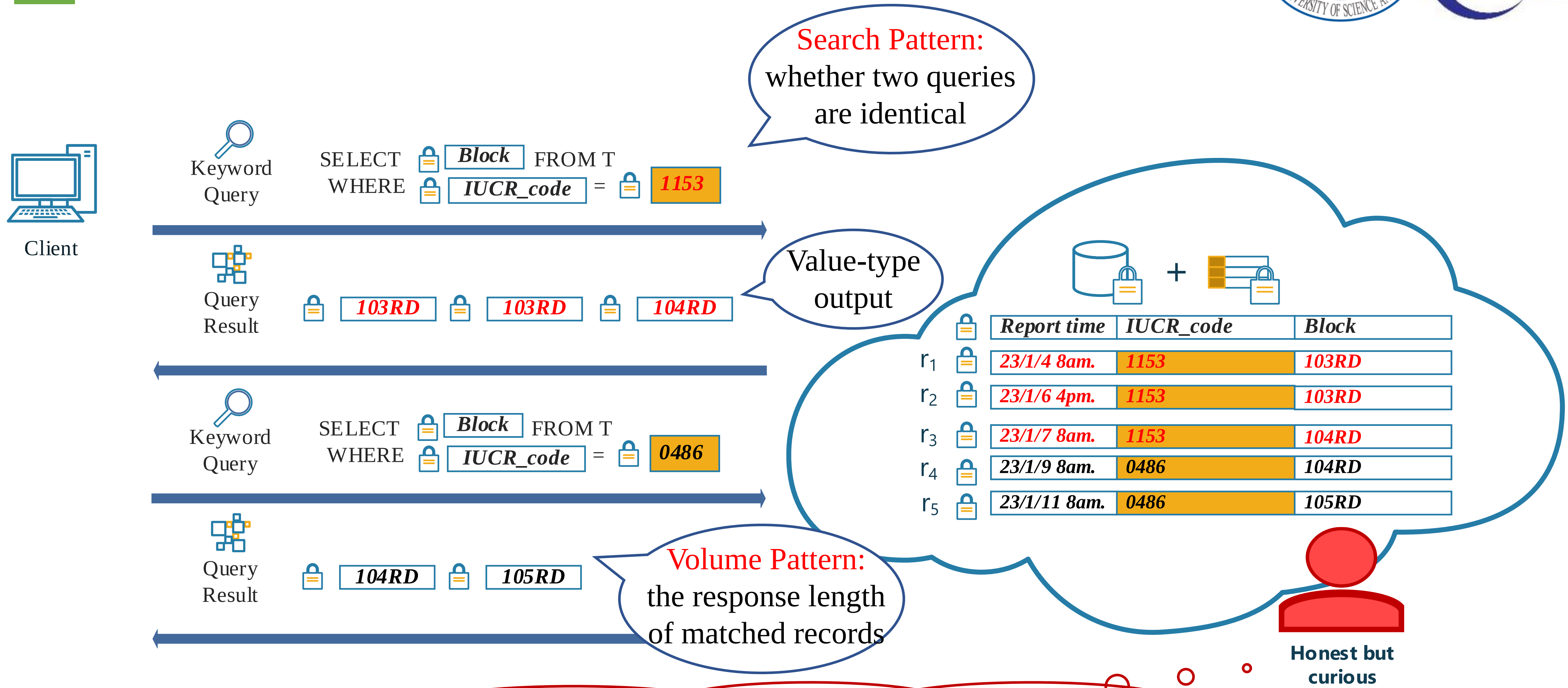
Search function:
Generate search tokens from Keyword Queries

Honest but curious

Query Result

103RD 103RD 104RD

Pattern & Volume leakage Attack



Client

Keyword Query

SELECT 🔒 *Block* FROM T
WHERE 🔒 *IUCR_code* = 🔒 *1153*

Query Result

🔒 *103RD* 🔒 *103RD* 🔒 *104RD*

Keyword Query

SELECT 🔒 *Block* FROM T
WHERE 🔒 *IUCR_code* = 🔒 *0486*

Query Result

🔒 *104RD* 🔒 *105RD*

Search Pattern: whether two queries are identical

Value-type output

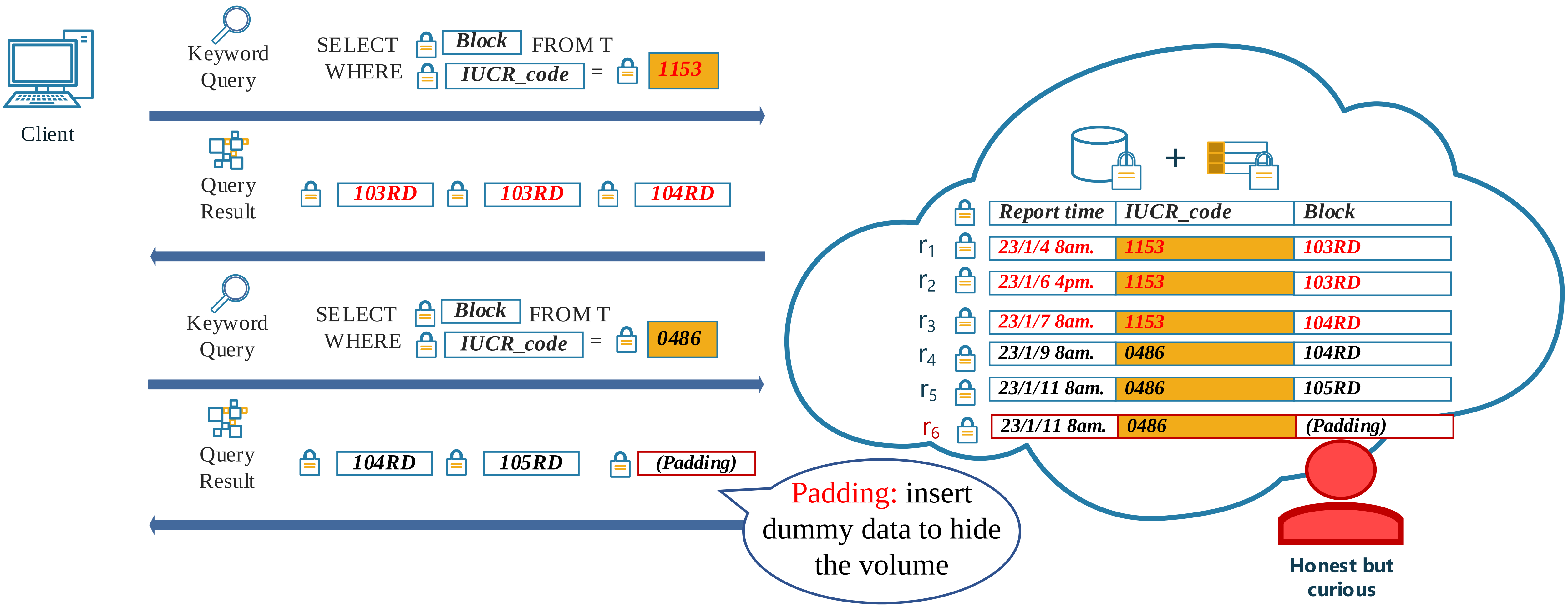
Volume Pattern: the response length of matched records

	Report time	*IUCR_code*	*Block*
🔒			
r_1 🔒	*23/1/4 8am.*	*1153*	*103RD*
r_2 🔒	*23/1/6 4pm.*	*1153*	*103RD*
r_3 🔒	*23/1/7 8am.*	*1153*	*104RD*
r_4 🔒	*23/1/9 8am.*	*0486*	*104RD*
r_5 🔒	*23/1/11 8am.*	*0486*	*105RD*

Honest but curious

Volume leakage attack: If I pre-known the ‘1153’ and ‘0486’ return 3 and 2-length result, I will infer the keyword from client queries.

Padding Strategy



SELECT Block FROM T
WHERE IUCR_code = 1153

103RD 103RD 104RD

SELECT Block FROM T
WHERE IUCR_code = 0486

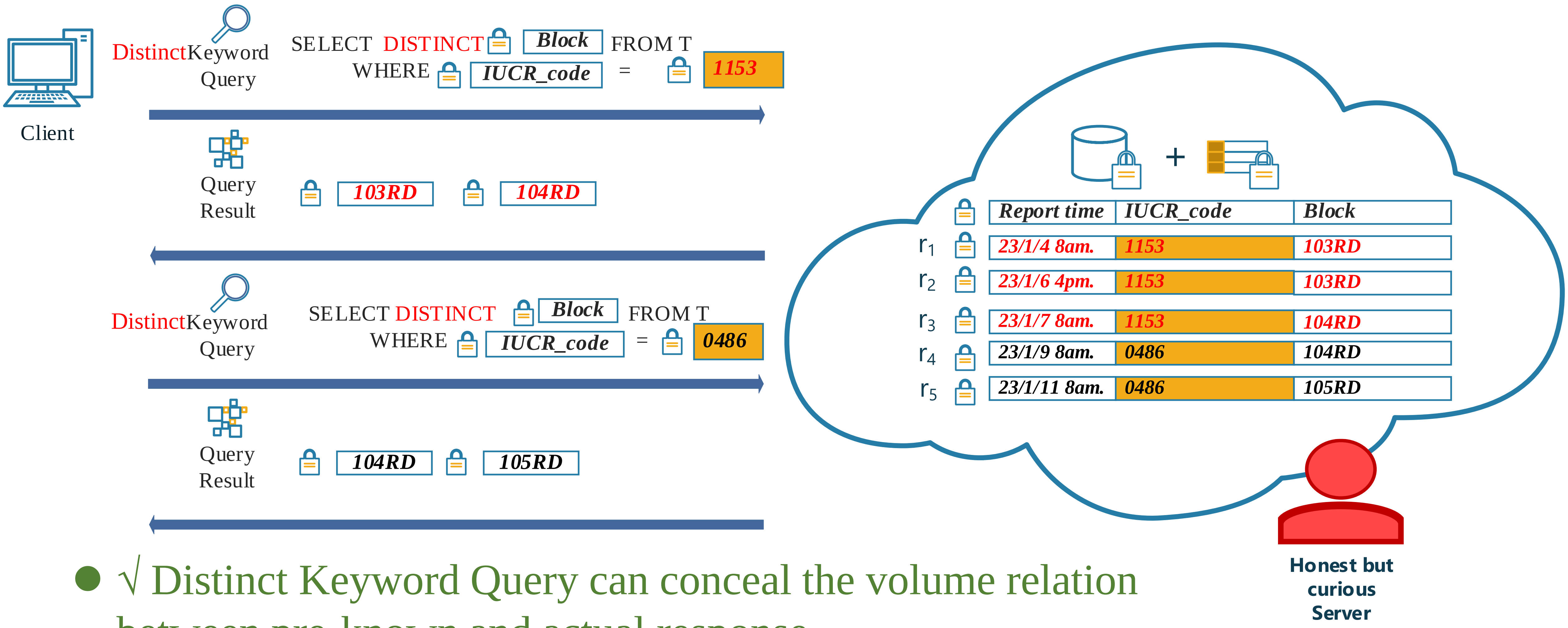
104RD 105RD (Padding)

Padding: insert dummy data to hide the volume

	Report time	IUCR_code	Block
r_1	23/1/4 8am.	1153	103RD
r_2	23/1/6 4pm.	1153	103RD
r_3	23/1/7 8am.	1153	104RD
r_4	23/1/9 8am.	0486	104RD
r_5	23/1/11 8am.	0486	105RD
r_6	23/1/11 8am.	0486	(Padding)

- √ fill false positive data to ensure the volume is consistent enough
- ×estimate the padding scale, hard to delete false positive data in EDB

Distinct Keyword Queries



DistinctKeyword Query

SELECT DISTINCT 🔒 *Block* FROM T
WHERE 🔒 *IUCR_code* = 🔒 *1153*

Query Result

🔒 *103RD* 🔒 *104RD*

DistinctKeyword Query

SELECT DISTINCT 🔒 *Block* FROM T
WHERE 🔒 *IUCR_code* = 🔒 *0486*

Query Result

🔒 *104RD* 🔒 *105RD*

		Report time	*IUCR_code*	*Block*
	🔒	*Report time*	*IUCR_code*	*Block*
r_1	🔒	*23/1/4 8am.*	*1153*	*103RD*
r_2	🔒	*23/1/6 4pm.*	*1153*	*103RD*
r_3	🔒	*23/1/7 8am.*	*1153*	*104RD*
r_4	🔒	*23/1/9 8am.*	*0486*	*104RD*
r_5	🔒	*23/1/11 8am.*	*0486*	*105RD*

Honest but curious Server

- √ Distinct Keyword Query can conceal the volume relation between pre-known and actual response
- ? How to ensure the **Security** of Distinct Keyword Query
- ? How to **feed back** queries in EDB

Contributions

- d-DSE: **d**istinct **D**ynamic **S**earchable Symmetric **E**ncryption
  - Search protocol => Distinct Search protocol
  - Forward and backward privacy (hides leakage from update operations)
  - Distinct with Volume Hiding Security (hides volume leakage)
  - Construction based on DSE
- Constructions for Queries in EDB
  - Update Queries
  - Keyword Queries from Distinct Keyword Queries
  - Join queries from Keyword Queries
- Experimental comparison between padding strategy and distinct search

Ⅰ. d-DSE

FP&BP In d-DSE

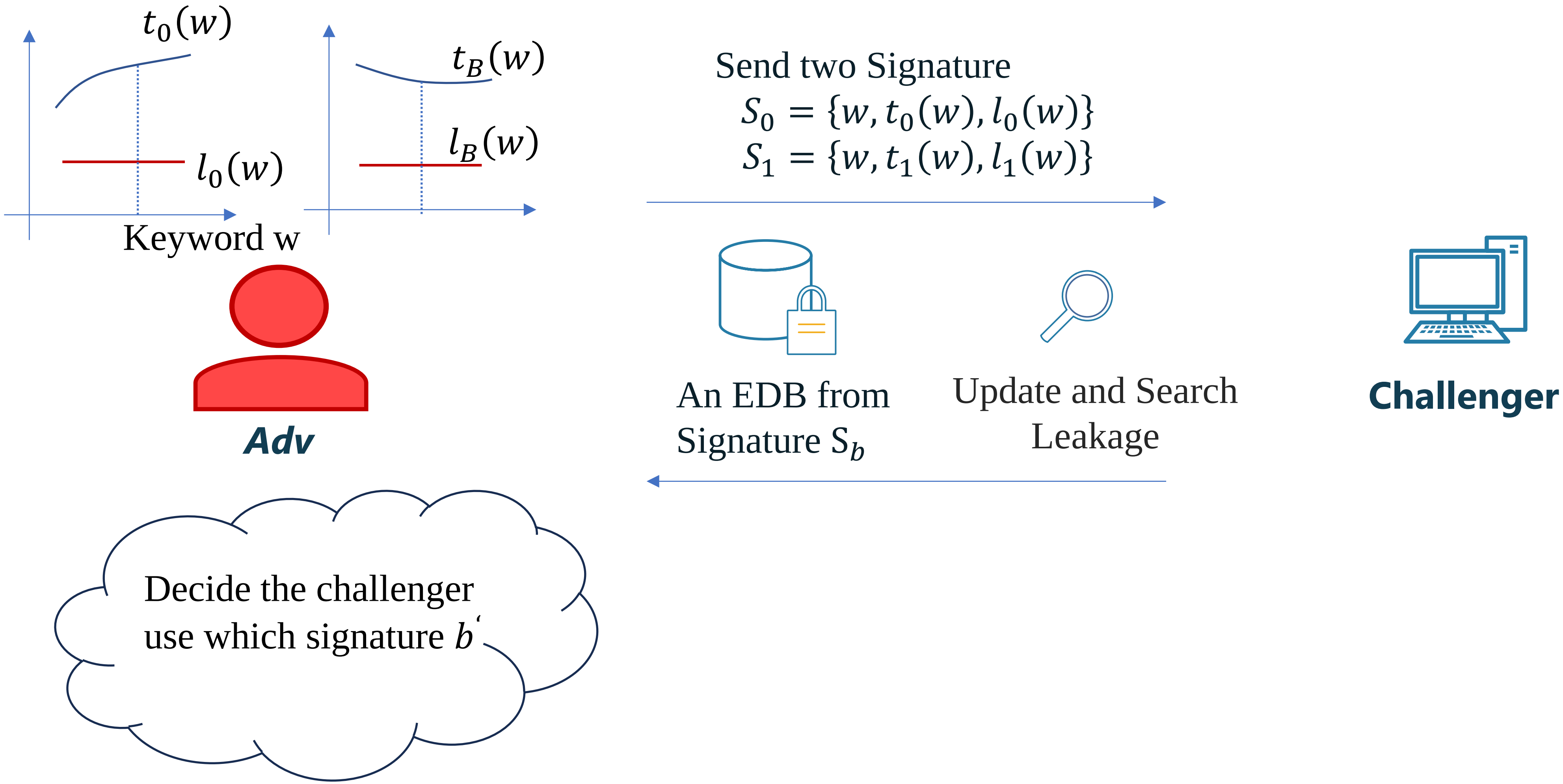
- **Observation 1**: The FP&BP in DSE [Bost et al., CCS’17] considers keyword/file-identifier input, but not keyword/value inputs

- **Observation 2**: If a deletion corresponds to multiple addition, the deletion will reveal which addition contains the same keyword/value

- Refine the FP&BP in d-DSE
  - FP should additionally hide the repetitive value information
  - BP should reach type-2 to hide the relation between addition and deletion

- *FP&BP =》 Volume-hiding：Hide from content, next to volume*

DISTINCT with Volume-Hiding (DwVH) Security



Send two Signature
$S_0 = \{w, t_0(w), l_0(w)\}$
$S_1 = \{w, t_1(w), l_1(w)\}$

An EDB from Signature S_b

Update and Search Leakage

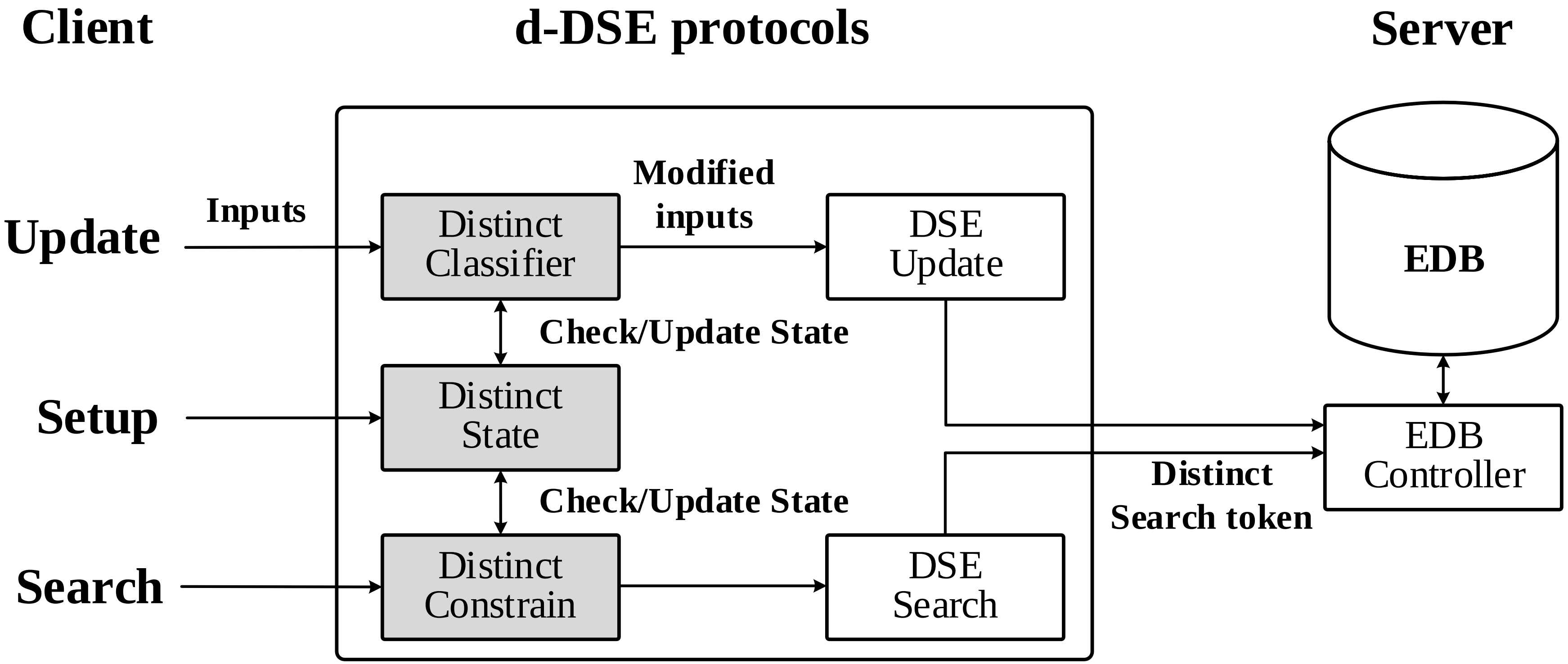
Challenger

Adv

Decide the challenger use which signature $b^{'}$

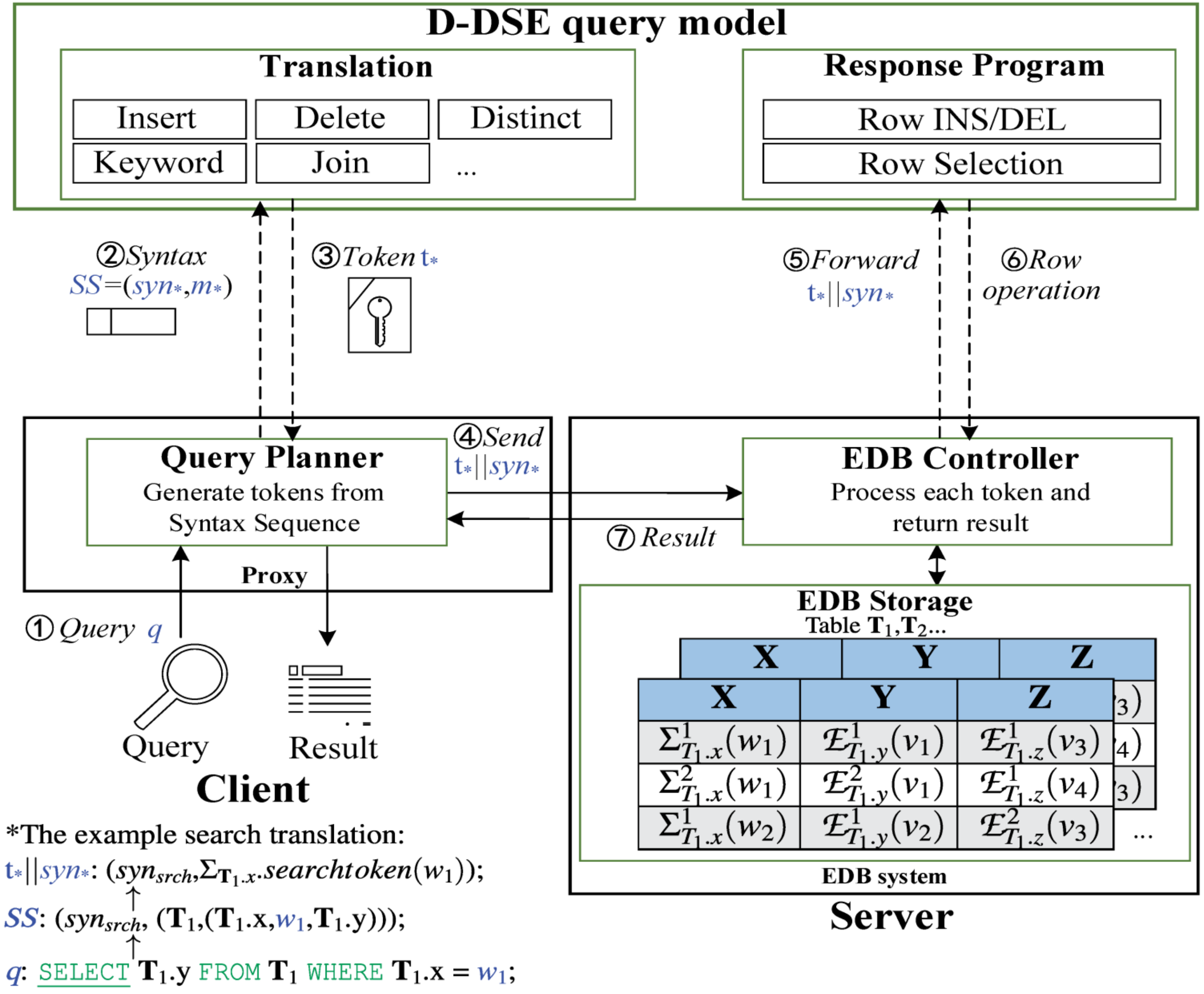
Construction based on DSE

- Design



Ⅱ. Constructions for Queries

d-DSE designed EDB



D-DSE query model

Translation

Insert | Delete | Distinct

Keyword | Join | ...

Response Program

Row INS/DEL

Row Selection

②*Syntax* $SS=(syn_*,m_*)$

③*Token* t_*

⑤*Forward* $t_*||syn_*$

⑥*Row operation*

Query Planner
Generate tokens from Syntax Sequence

Proxy

④*Send* $t_*||syn_*$

⑦ *Result*

EDB Controller
Process each token and return result

EDB Storage
Table $\mathbf{T}_1,\mathbf{T}_2$...

X	Y	Z
$\Sigma^1_{T_1.x}(w_1)$	$\mathcal{E}^1_{T_1.y}(v_1)$	$\mathcal{E}^1_{T_1.z}(v_3)$
$\Sigma^2_{T_1.x}(w_1)$	$\mathcal{E}^2_{T_1.y}(v_1)$	$\mathcal{E}^1_{T_1.z}(v_4)$
$\Sigma^1_{T_1.x}(w_2)$	$\mathcal{E}^1_{T_1.y}(v_2)$	$\mathcal{E}^2_{T_1.z}(v_3)$

...

EDB system

Server

① *Query* q

Query

Result

Client

*The example search translation:

$t_*||syn_*$: $(syn_{srch},\Sigma_{\mathbf{T}_1.x}.searchtoken(w_1))$;

$\uparrow$

SS: $(syn_{srch}, (\mathbf{T}_1,(\mathbf{T}_1.\mathrm{x},w_1,\mathbf{T}_1.\mathrm{y})))$;

$\uparrow$

q: SELECT $\mathbf{T}_1$.y FROM $\mathbf{T}_1$ WHERE $\mathbf{T}_1.\mathrm{x} = w_1$;

Query Construction

- **Step 1**: Apply the Update protocol for Update queries
- **Step 2**: Apply the Search protocol for distinct Keyword queries
- **Step 3**: Construct Keyword Queries Based on distinct Keyword Queries
  - Let client allocate a hash table to map the keyword w with the vector **d** constructed from (w,v,op) input.
  - The dimensions of **d** record the number of value in the value’s numerical (or lexicographical for string) order.
  - Update the number after Update queries
- **Step 4**: Construct Join Queries Based on Keyword Queries
  - Create a Keyword Query to find all values as *FOREIGN KEY* in another table
  - Use the *FOREIGN KEY* as keyword to perform Keyword Queries on the other table

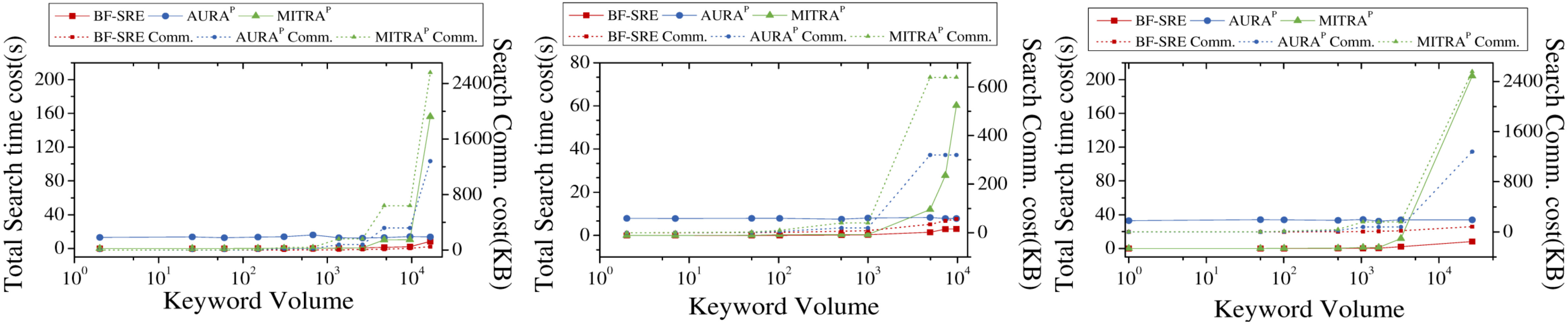
Ⅲ. Padding vs. Distinct Search

Comp. Configuration

- 3 real dataset
  - Crime incidents: 7,989,987 records of Street name/IUCR pairs
  - Enron mail: 5,190,199 records of email name/word pairs
  - Wikipedia: 4,565,948 records of document name/word pairs
- Compare d-DSE instance BF-SRE with DSE schemes (Mitra [CCS'18] and Aura [NDSS'21]) under the padding strategy proposed by Seal [USENIX security'20]
- Compare with Seal and ShieldDB [Vo et al., TKDE23] about the query performance

Comp. with Aura^P and Mitra^P

- **Main result after test:** Without adding dummy data, achieve around 15.27x and 30.54x communication advantage over Aura^P and Mitra^P



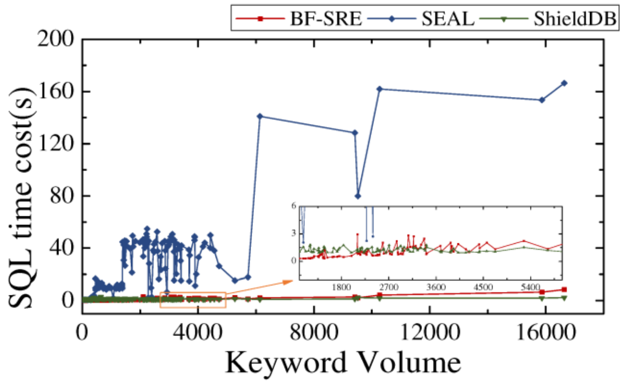
(a) Crime

(b) Wikipedia

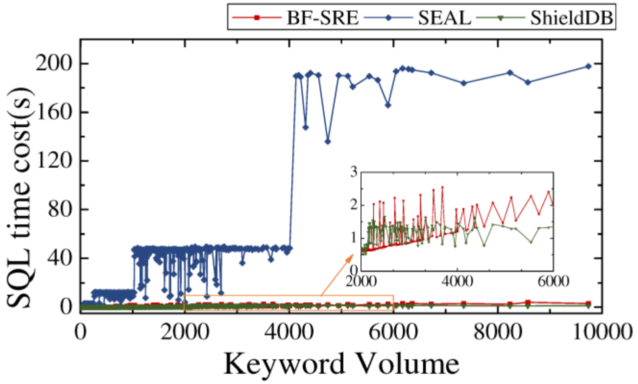
(c) Enron

Comp. with Seal and ShieldDB

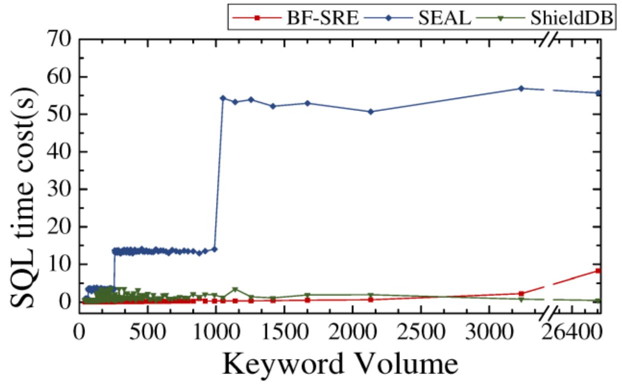
- **Main result after test:** Without adding dummy data, reduce around 6.36-53.14x communication advantage than Seal and ShieldDB



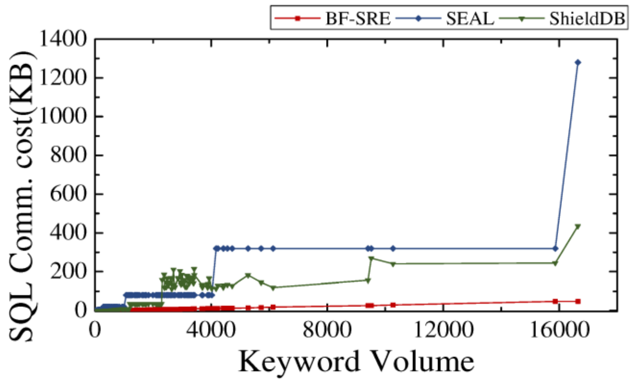
(a) Crime



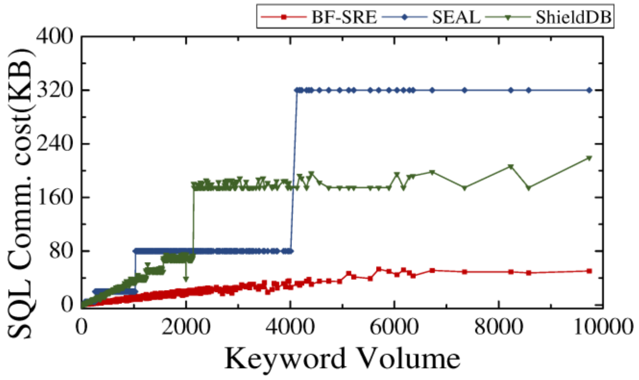
(b) Wikipedia



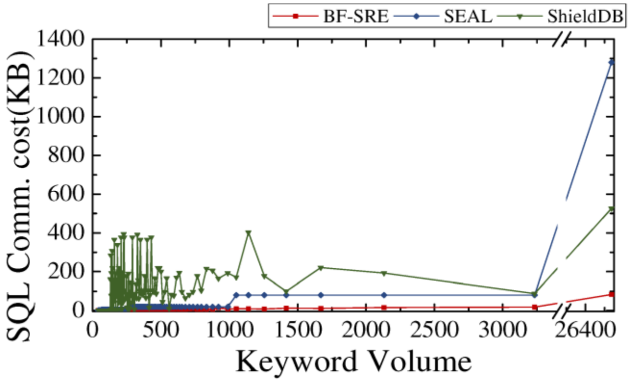
(c) Enron



(a) Crime



(b) Wikipedia



(c) Enron

Thank you for listening!

Code available: https://github.com/jd89j12dsa/ddse/tree/AEversion

ARTIFACT EVALUATED usenix ASSOCIATION AVAILABLE

ARTIFACT EVALUATED usenix ASSOCIATION FUNCTIONAL

ARTIFACT EVALUATED usenix ASSOCIATION REPRODUCED

Contact information:

→ nsffldl@hust.edu.com

→ viviawangwei@hust.edu.cn

→ xupeng@hust.edu.cn