

AUGUST 3-8, 2019
MANDALAY BAY / LAS VEGAS



AUGUST 3-8, 2019 MANDALAY BAY / LAS VEGAS

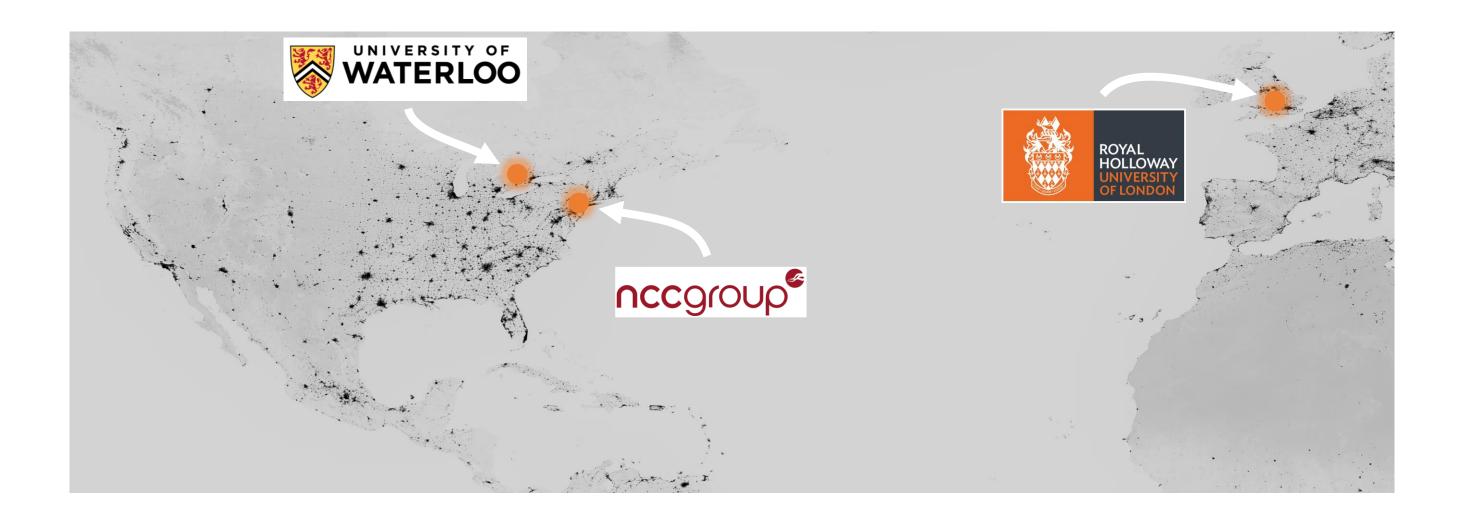
DATABASES: GENERIC ATTACKS ON RANGE **QUERIES**

Marie-Sarah Lacharité

Royal Holloway, University of London and NCC Group



About Me





Motivation: Data Breaches

Entity	Туре	Number of Records
First American	Finance	885,000,000
facebook	Social network apps	540,000,000
truecaller	Telephone directory	300,000,000
Capital One®	Finance	106,000,000
Quest Diagnostics™	Clinical laboratory	11,900,000
Desjardins	Finance	2,900,000





Motivation: Data Breaches

Healthcare IT News

Walgreens company announces data breach

According to the letter, an unknown person or persons broke into Crescent's billing center and stole the hardware, which may have contained patient names, addresses, phone numbers, Social Security numbers, health insurance data, dates of birth and clinical diagnoses. The group notified authorities three days later.

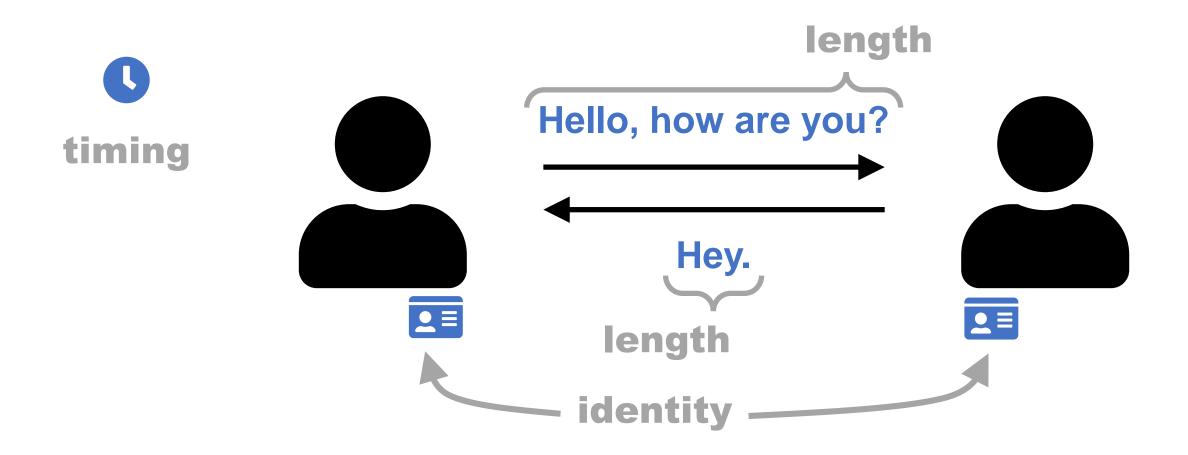
WSJ CYBERSECURITY

Capital One Breach Highlights Shortfalls of Encryption

Capital One said in a statement this week that it uses encryption "as a standard," but the method used by the hacker "enabled the decrypting of data." The bank didn't respond to questions about its encryption practices.

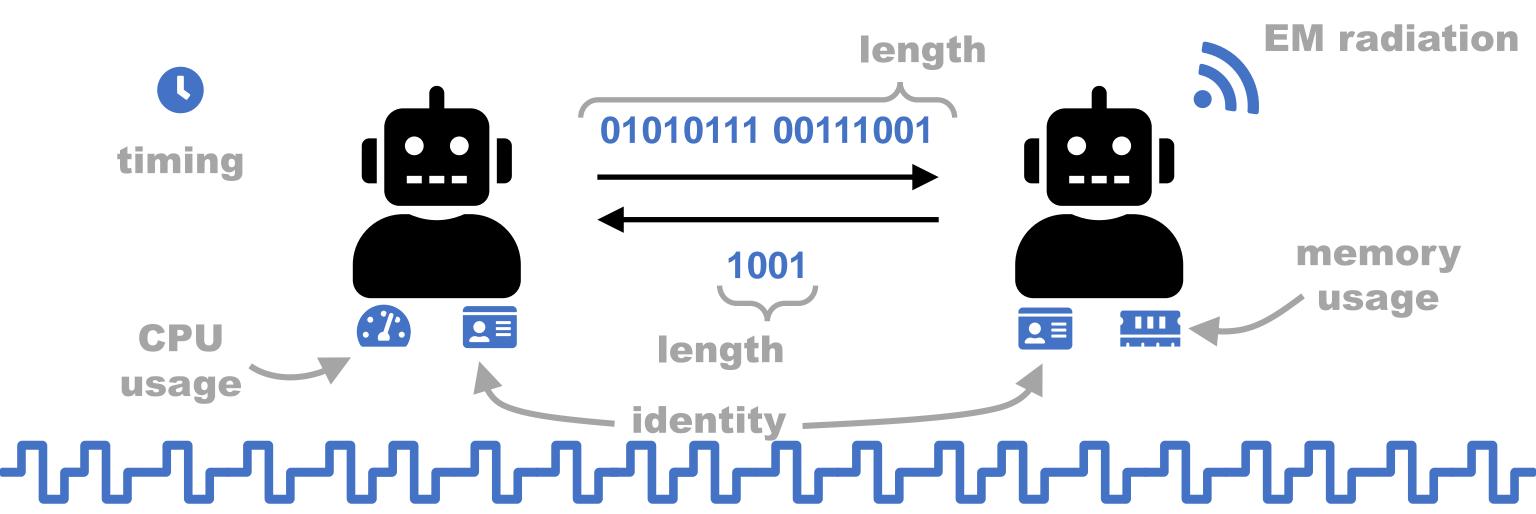


Side Channel Attacks





Side Channel Attacks

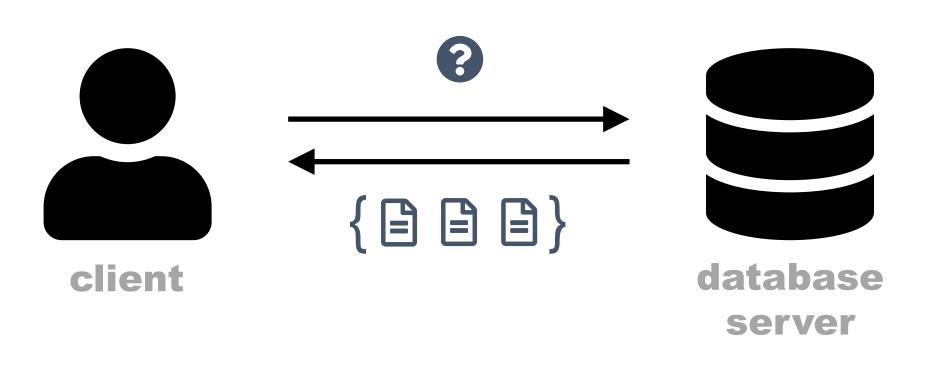




Side Channel Attacks

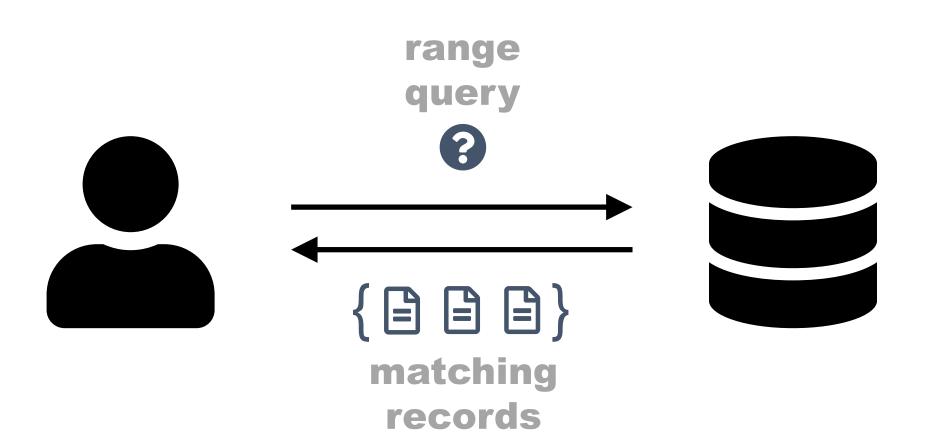
- SSH keystroke recovery from timing information
 [Song et al., USENIX 2001]
- Video stream identification from traffic burst analysis
 [Schuster et al., USENIX 2017]
- Message decryption from padding validity checks
 [Bleichenbacher, CRYPTO 1998]





ID	Value
1	3
2	1
3	15
4	41
5	1

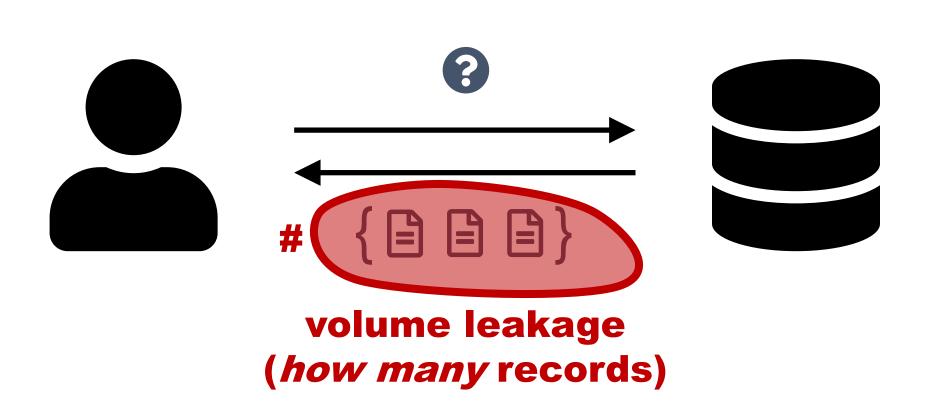




ID	Value
1	3
2	1
3	15
4	41
5	1



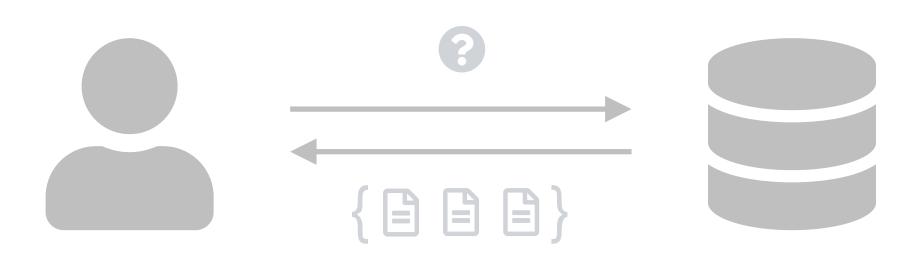
access pattern leakage (which records)



	ID	Value
EII)	1	3
	2	1
	3	15
	4	41
	5	1







ID	Value
1	3
2	1
3	15
4	41
5	1



Outline

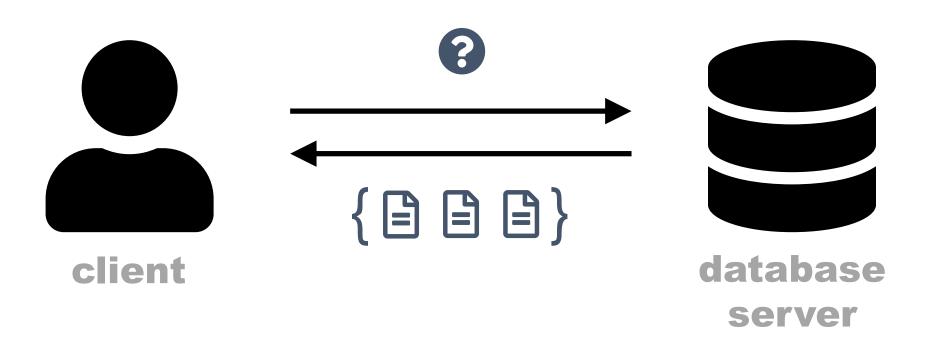
- 1. Existing approaches to securing a database
 - Securing data in transit, at rest, and in use
- 2. How to exploit leakage to break database encryption
 - Exploiting access pattern leakage and volume leakage
- 3. Security recommendations
 - Types of leakage, leaky operations, trade-offs



Outline

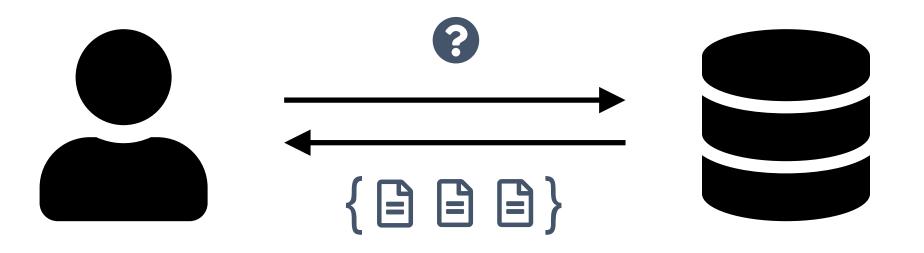
- 1. Existing approaches to securing a database
 - Securing data in transit, at rest, and in use
- 2. How to exploit leakage to break database encryption
 - Exploiting access pattern leakage and volume leakage
- 3. Security recommendations
 - Types of leakage, leaky operations, trade-offs





ID	Value
1	3
2	1
3	15
4	41
5	1





ID	Value
1	3
2	1
3	15
4	41
5	1
 	•••

database table

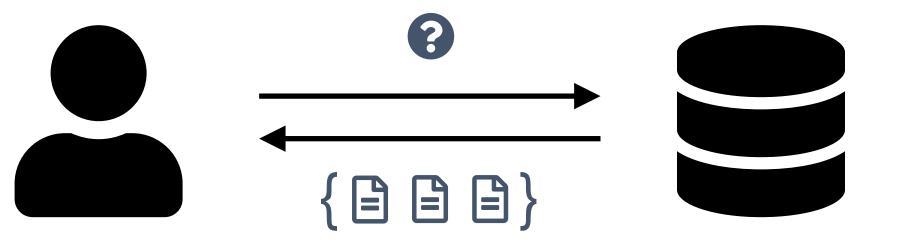


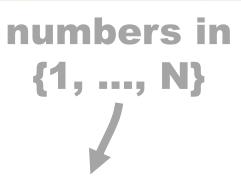




ID	Value
1	3
2	1
3	15
4	41
5	1

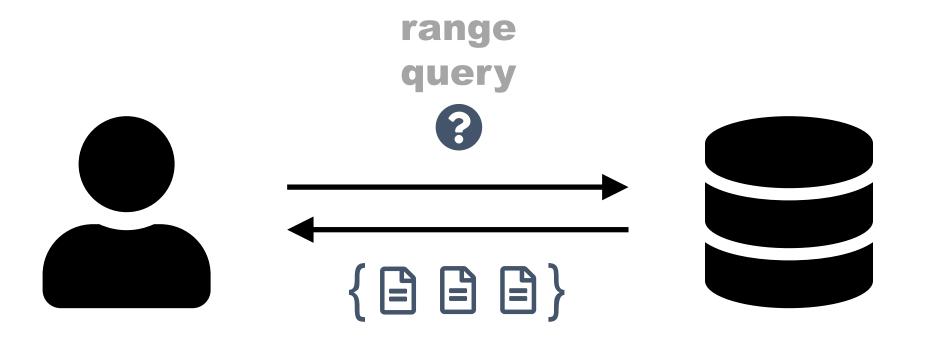






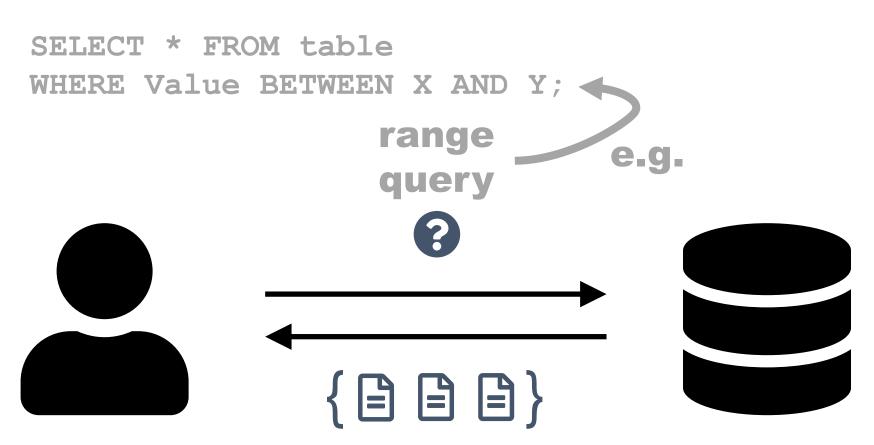
ID	Value
1	3
2	1
3	15
4	41
5	1





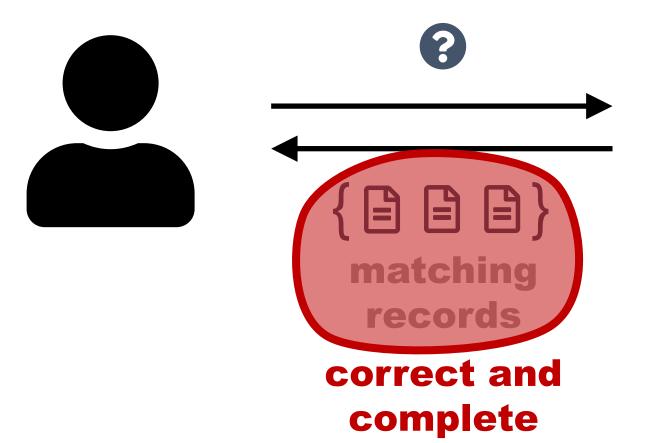
ID	Value
1	3
2	1
3	15
4	41
5	1
 	• • •

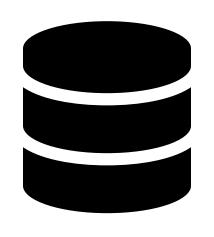


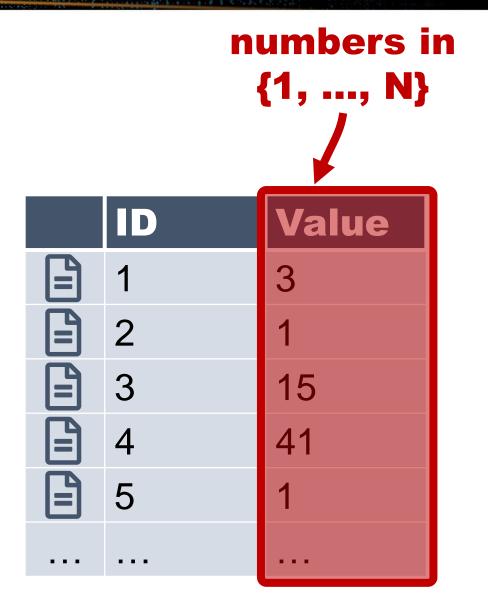


ID	Value
1	3
2	1
3	15
4	41
5	1



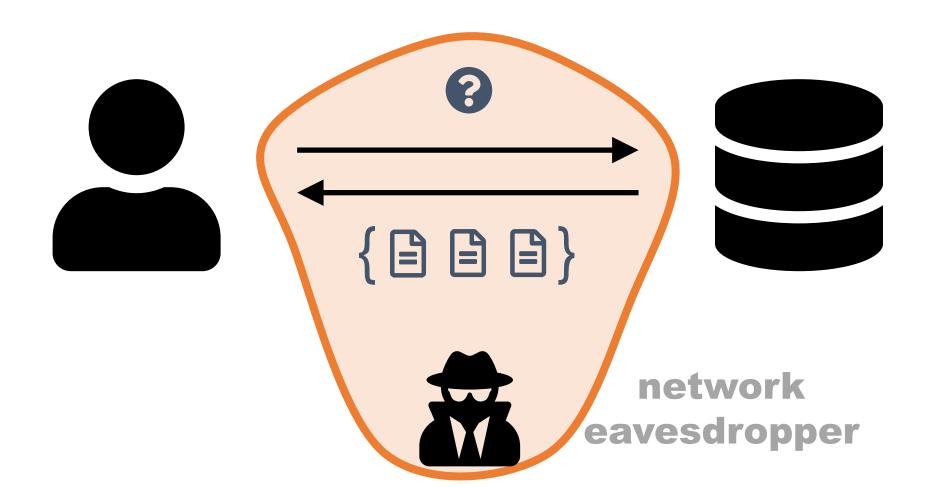








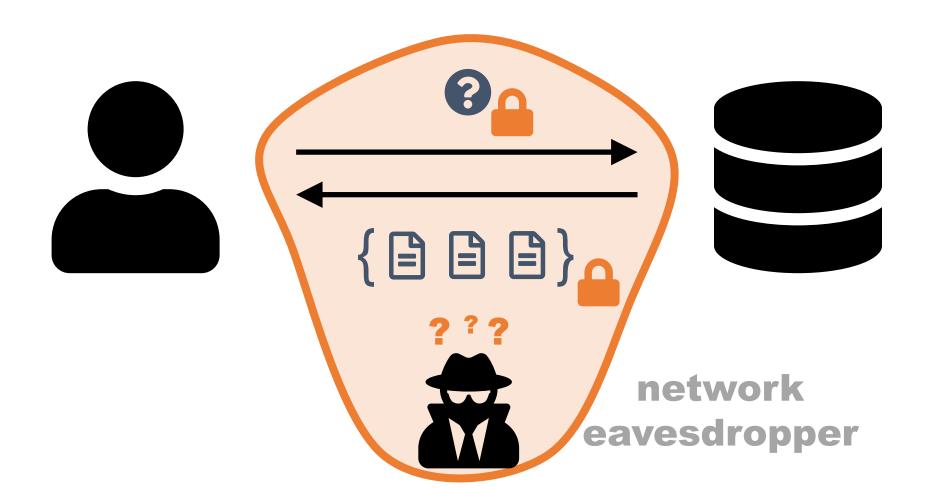
Encrypting Data in Transit



ID	Value
1	3
2	1
3	15
4	41
5	1

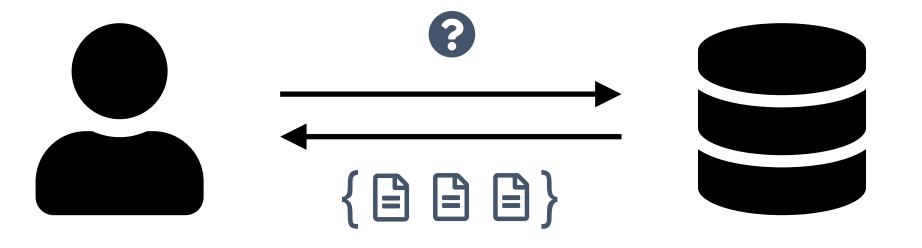


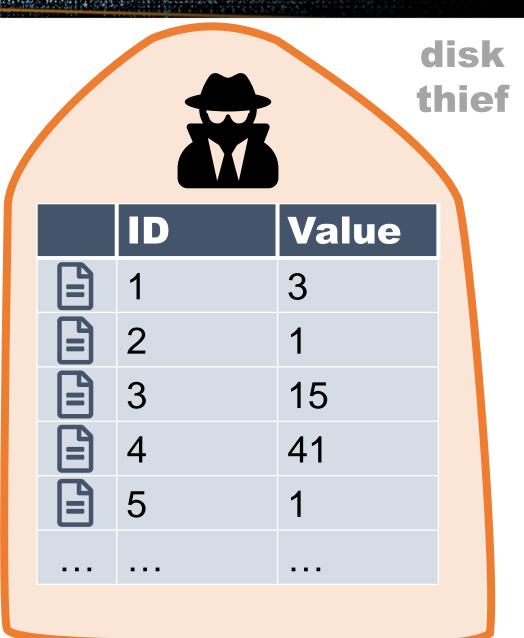
Encrypting Data in Transit



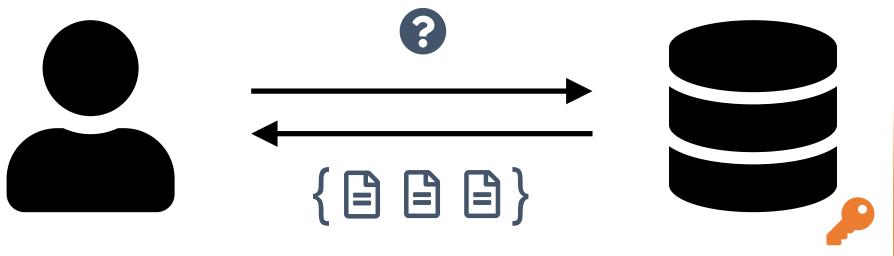
ID	Value
1	3
2	1
3	15
4	41
5	1
 	•••

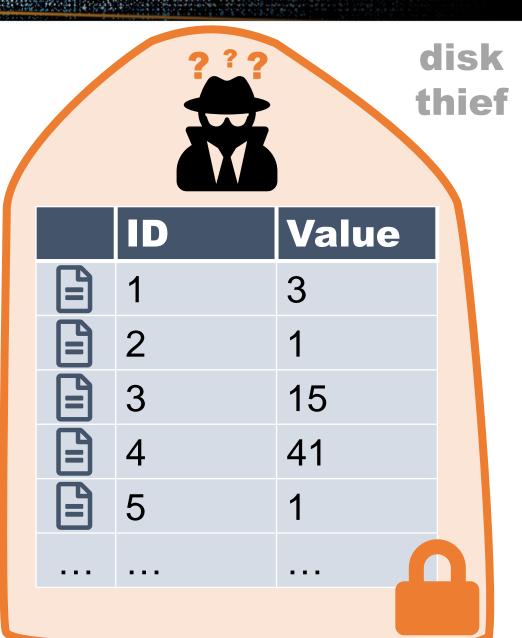




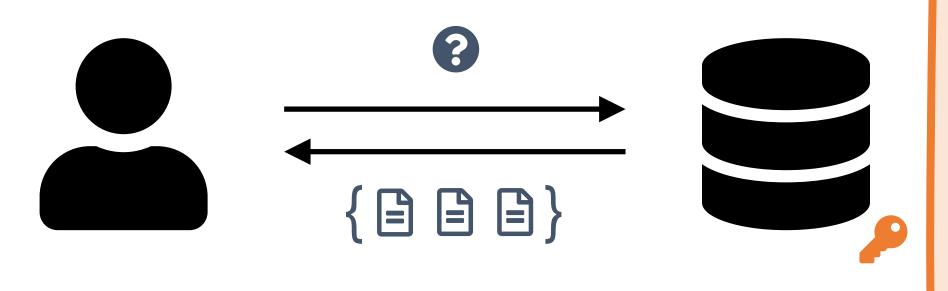


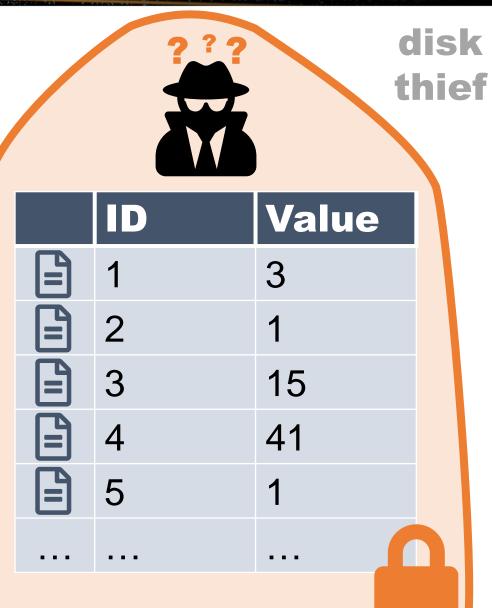














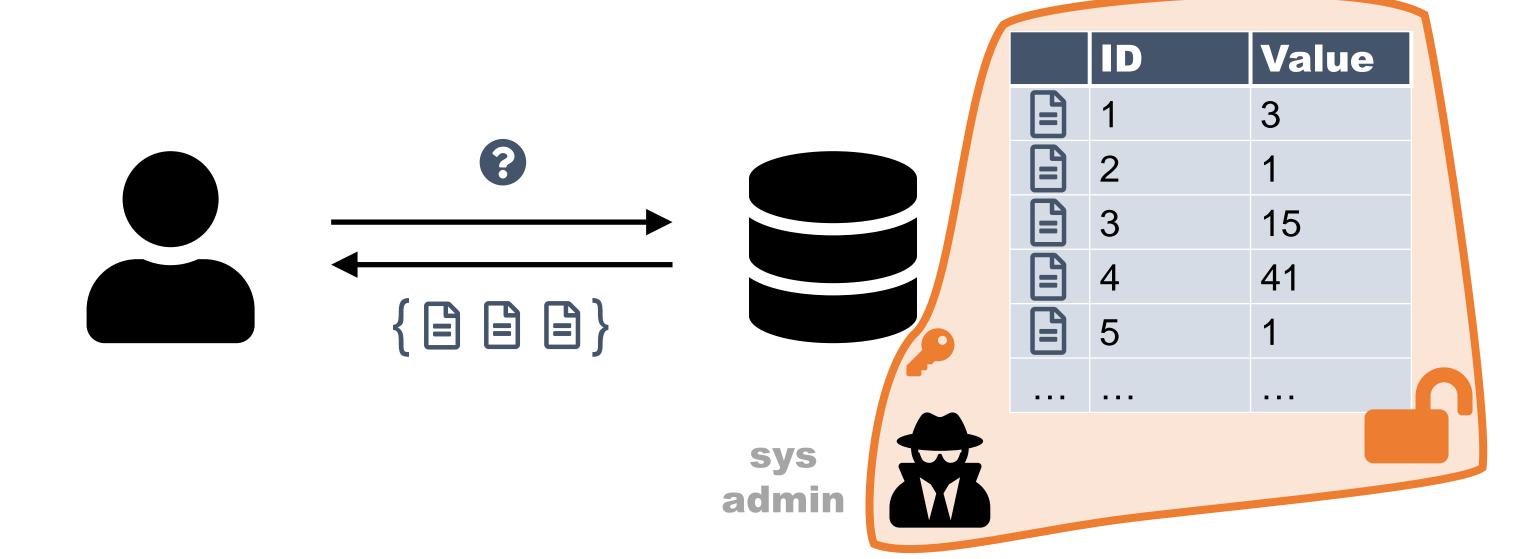




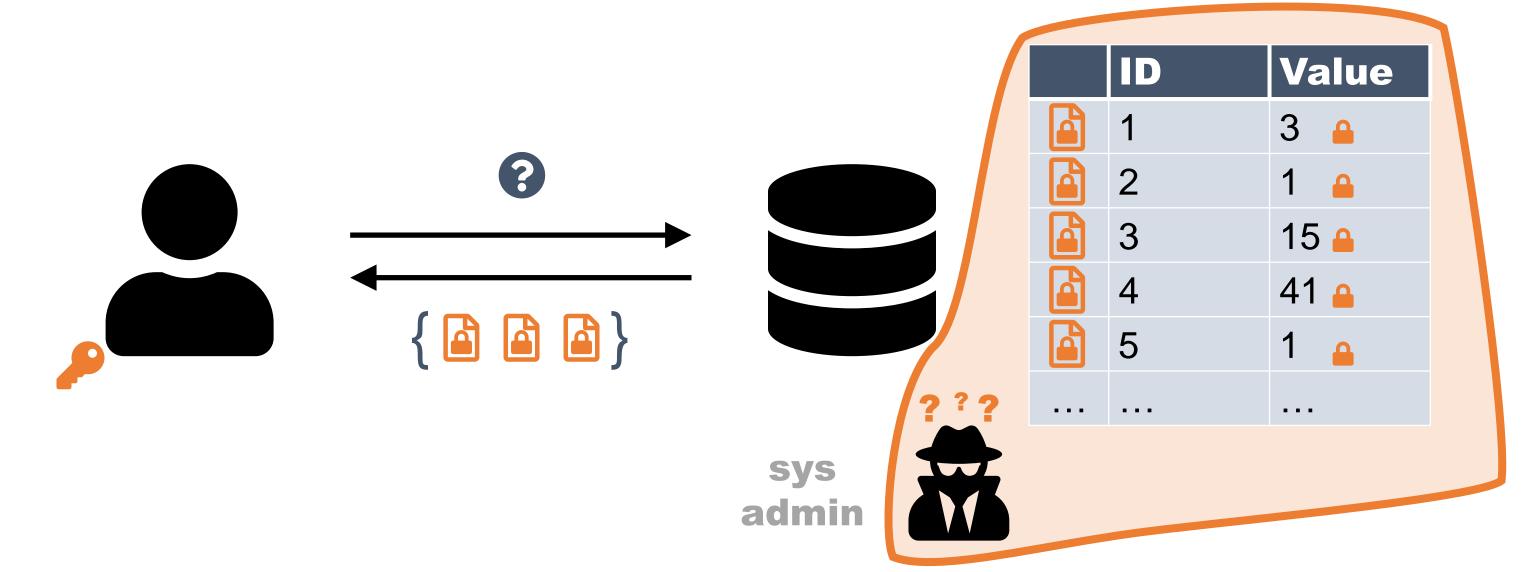




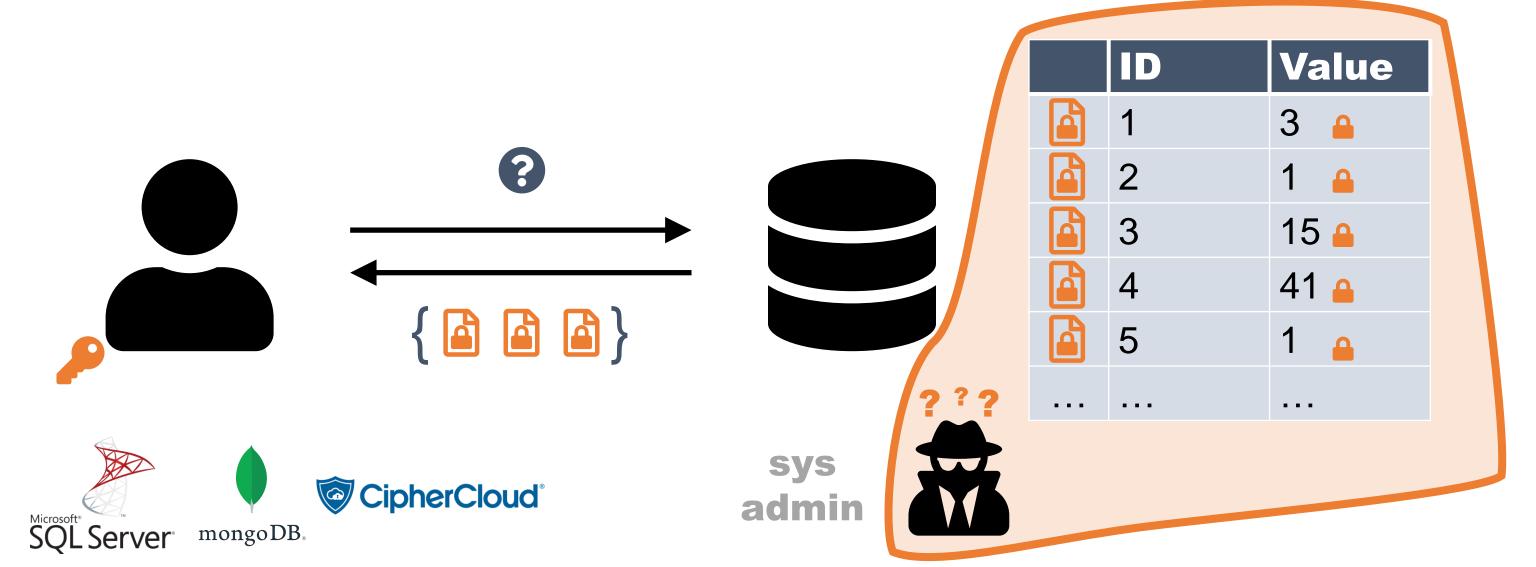














PLAINTEXT

ID	Value
1	3
2	1
3	15
4	41
5	1

DETERMINISTIC ENCRYPTION

ID	Value
1	0x18fa83
2	0x5449a1
3	0x8b7630
4	0x10cae8
5	0x5449a1

RANDOMIZED ENCRYPTION

ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb



PLAINTEXT

ID	Value
1	3
2	1
3	15
4	41
5	1

DETERMINISTIC ENCRYPTION

ID	Value
1	0x18fa83
2	0x5449a1
3	0x8b7630
4	0x10cae8
5	0x5449a1

RANDOMIZED ENCRYPTION

ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb



DETERMINISTIC ENCRYPTION

ID	Value
1	0x18fa83
2	0x5449a1
3	0x8b7630
4	0x10cae8
5	0x5449a1
•••	

Range queries are possible:

```
... WHERE Value BETWEEN 1 AND 3 becomes
```

```
... WHERE Value IN

(Enc(1), Enc(2), Enc(3))
```

Revealing repeated values is dangerous

[Naveed et al., CCS 2015]



PLAINTEXT

ID	Value
1	3
2	1
3	15
4	41
5	1
	•••

DETERMINISTIC ENCRYPTION

ID	Value
1	0x18fa83
2	0x5449a1
3	0x8b7630
4	0x10cae8
5	0x5449a1

RANDOMIZED ENCRYPTION

ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb

range queries impossible





PLAINTEXT

ID	Value	
1	3	
2	1	1 < 15
3	15	
4	41	
5	1	

ORDER-PRESERVING ENCRYPTION

ID	Value	
1	182	
2	84	84 < 2307
3	2307	04 \ 2301
4	8932	
5	84	

[Agrawal et al., SIGMOD 2004], [Boldyreva et al., EUROCRYPT 2009]



- "Ideal" OPE leaks approximate value and distance [Boldyreva et al., CRYPTO 2011]
- Revealing repeated values is even more dangerous [Naveed et al., CCS 2015] [Grubbs et al., S&P 2017]

ORDER-PRESERVING ENCRYPTION

ID	Value
1	182
2	84
3	2307
4	8932
5	84



Order-Revealing Encryption (ORE)

[Chenette et al., FSE 2016], [Lewi and Wu, CCS 2016]

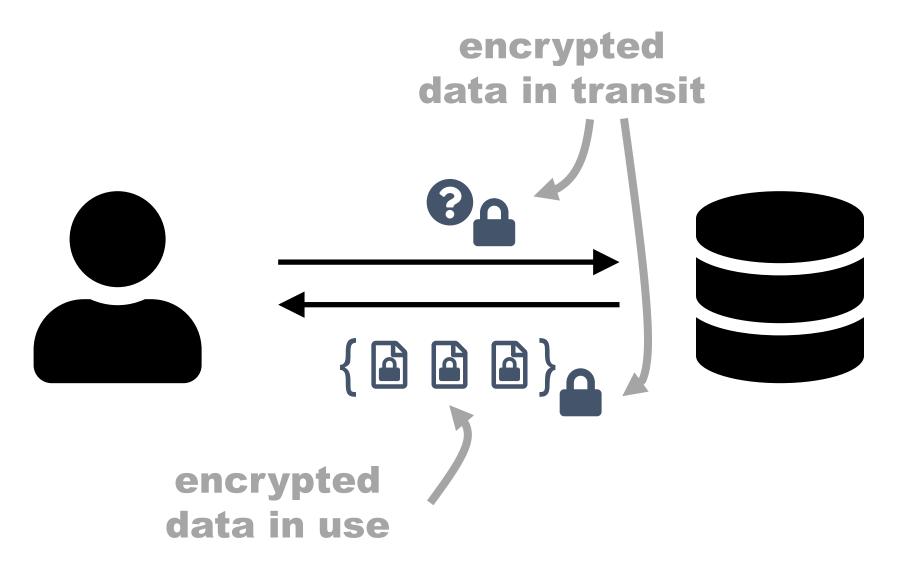
Partial Order-Preserving Encoding

[Roche et al., CCS 2016]

Custom search indices

[Boelter et al., eprint 2016/568]





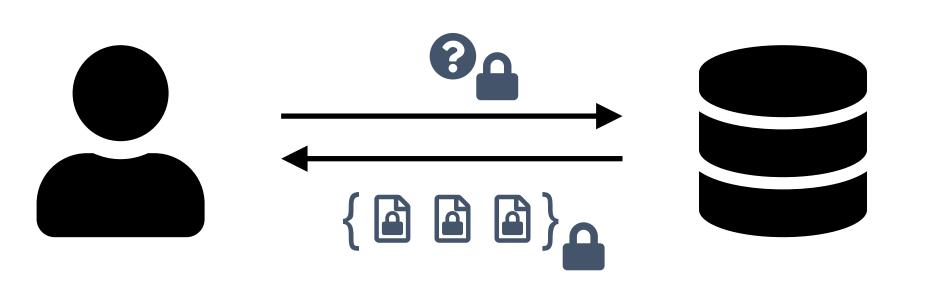
ID	Value
1	3 🛕
2	1 🖴
3	15 🖴
4	41 🖴
5	1 🛕

encrypted

data at rest



access pattern leakage (which records)



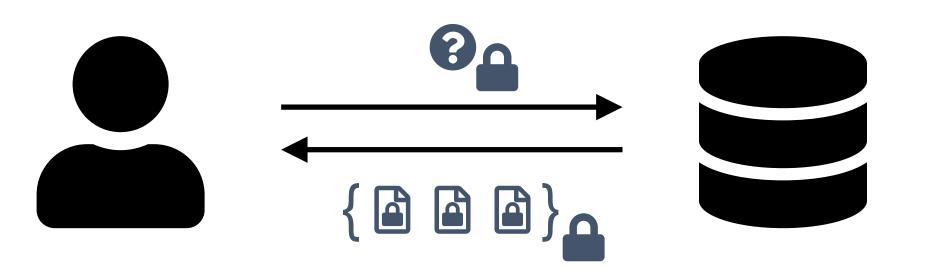
	ID	Value
4	1	3 🔒
4	2	1 🖴
	3	15 🖴
	4	41 🖴
4	5	1 🛕





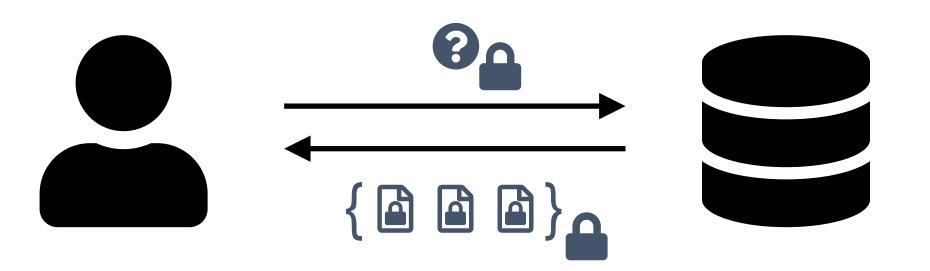
ID	Value
1	3 🛕
2	1 🛕
3	15 🖴
4	41 🖴
5	1 🛕





ID	Value
1	3 🛕
2	1 🖴
3	15 🖴
4	41 🖴
5	1 🛕





ID	Value
1	3 🛕
2	1 🖴
3	15 🖴
4	41 🖴
5	1 🛕

[Grubbs et al., HotOS 2017]



Existing Approaches

- No fixed definition of "encrypted database"
- Encryption mitigates threats (network, theft, server compromise)
- Despite these solutions, access pattern and volume can leak



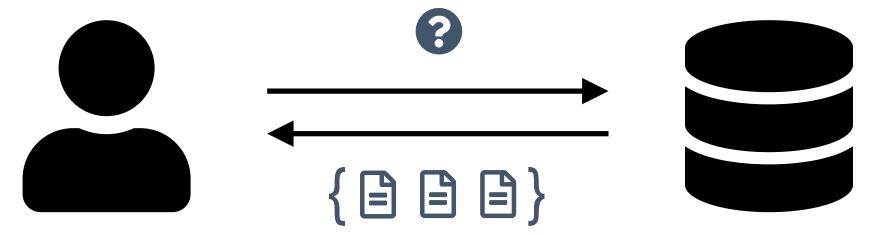
Outline

- 1. Existing approaches to securing a database
 - Securing data in transit, at rest, and in use
- 2. How to exploit leakage to break database encryption
 - Exploiting access pattern leakage and volume leakage
- 3. Security recommendations
 - Types of leakage, leaky operations, trade-offs



Exploiting Access Pattern Leakage



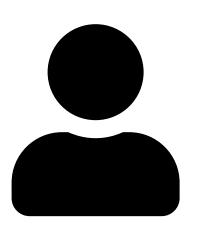


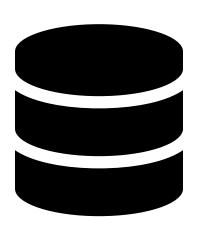
ID	Value
1	3
2	1
3	15
4	41
5	1





Example

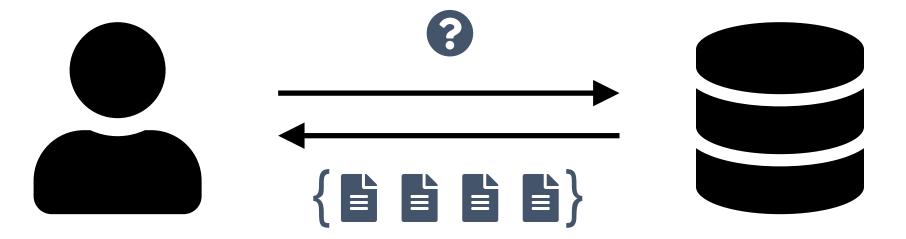




ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb
6	0x4073d2
7	0x2765be
8	0x74090f
9	0x5bae94
10	0xae60da



SELECT * FROM table
WHERE Value BETWEEN ? AND ?;

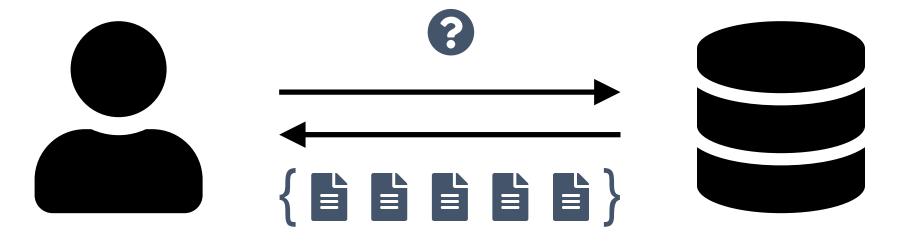


{2, 3,	5, 10}
--------	--------

ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb
6	0x4073d2
7	0x2765be
8	0x74090f
9	0x5bae94
10	0xae60da



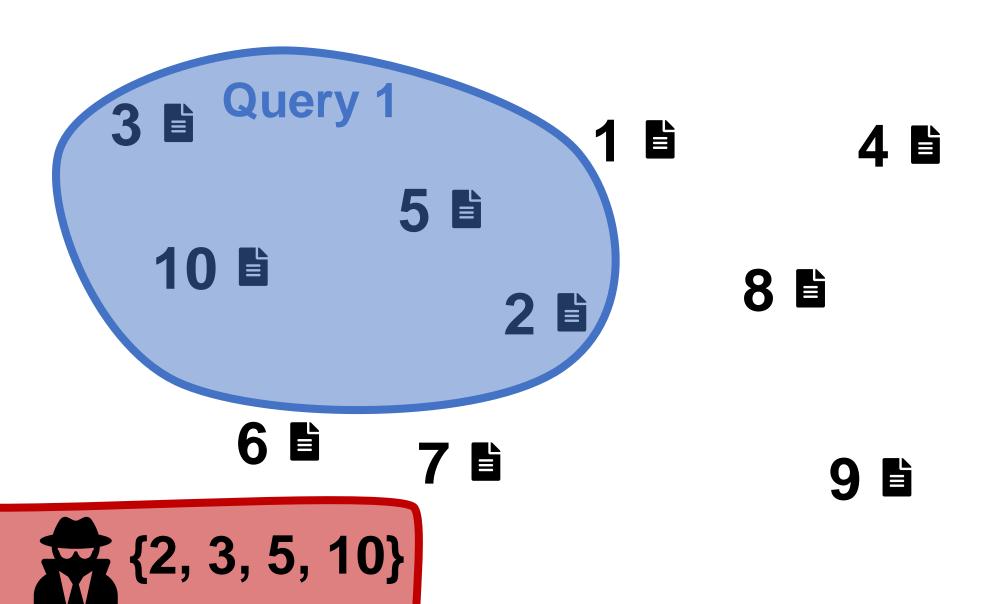
SELECT * FROM table
WHERE Value BETWEEN ? AND ?;



{2, 3, 5, 10}, {1, 2,	4, 5, 8}
-----------------------	----------

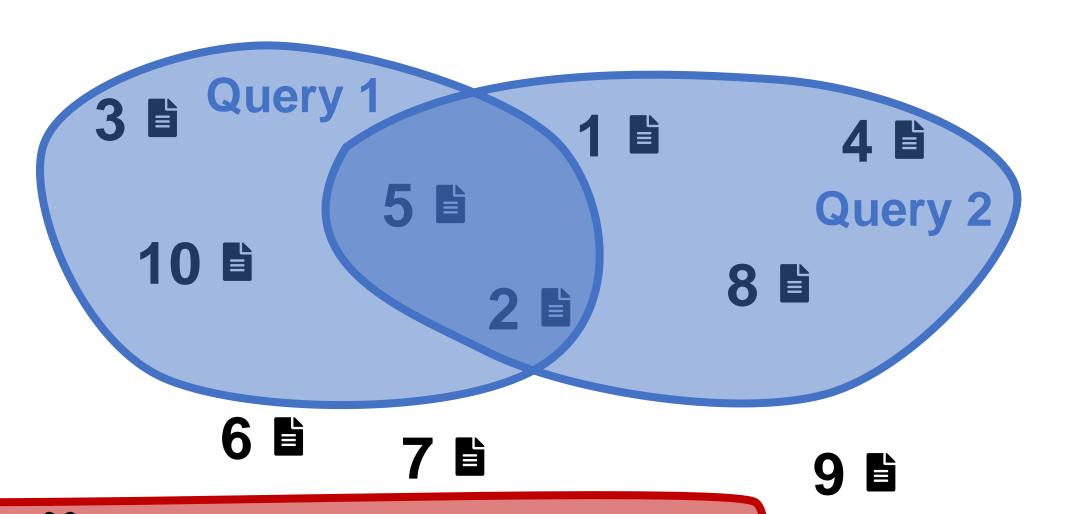
ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb
6	0x4073d2
7	0x2765be
8	0x74090f
9	0x5bae94
10	0xae60da





ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb
6	0x4073d2
7	0x2765be
8	0x74090f
9	0x5bae94
10	0xae60da





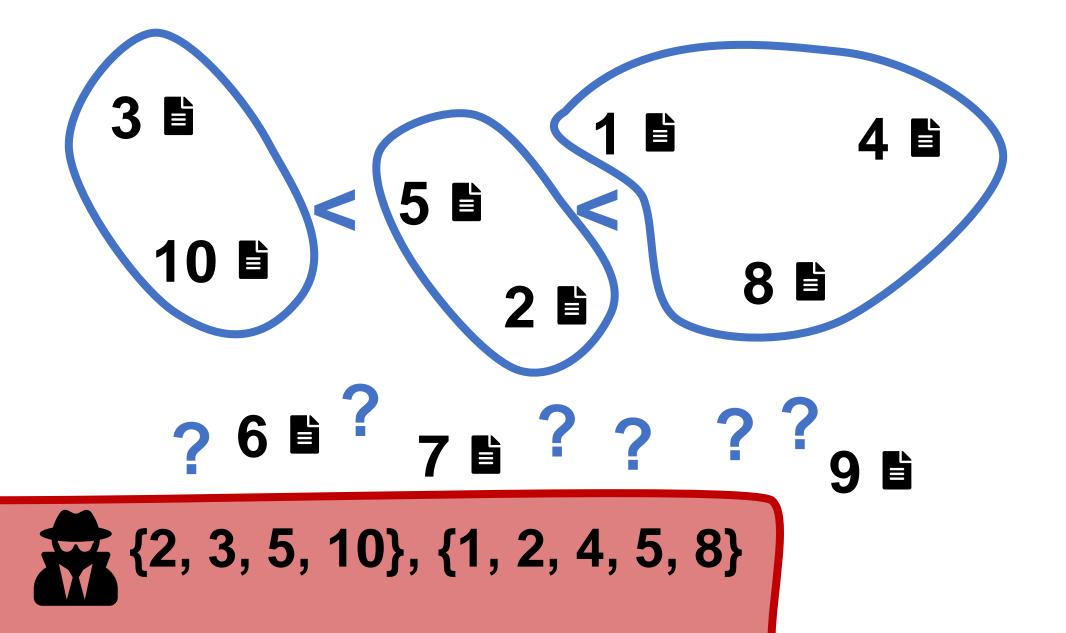
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb
6	0x4073d2
7	0x2765be
8	0x74090f
9	0x5bae94
10	0xae60da

Value



{2, 3, 5, 10}, {1, 2, 4, 5, 8}

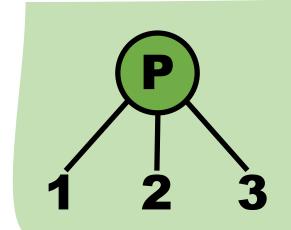




ID	Value
1	0x5239fb
2	0x8e9d98
3	0x5a9f2e
4	0x4ff8e1
5	0xe89cfb
6	0x4073d2
7	0x2765be
8	0x74090f
9	0x5bae94
10	0xae60da



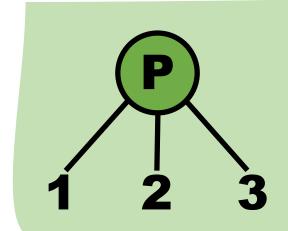
PQ Trees: Sets of Orderings



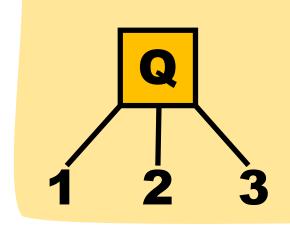
(1, 2, 3) or (1, 3, 2) or (2, 1, 3) or (2, 3, 1) or (3, 1, 2) or (3, 2, 1)



PQ Trees: Sets of Orderings



(1, 2, 3) or (1, 3, 2) or (2, 1, 3) or (2, 3, 1) or (3, 1, 2) or (3, 2, 1)

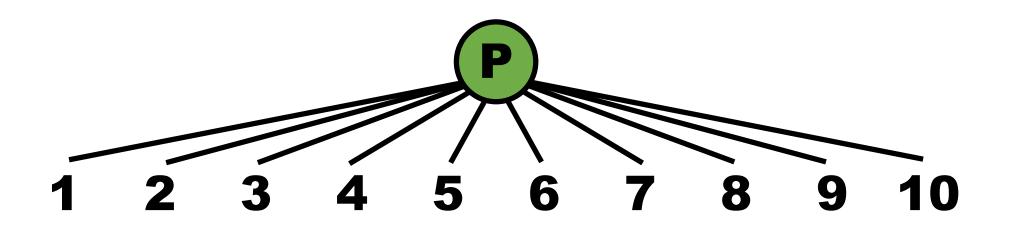


(1, 2, 3) or (3, 2, 1)

[Dautrich and Ravishankar, EDBT 2013] [Booth and Lueker, J. Comput. Sys. Sci., 1976]

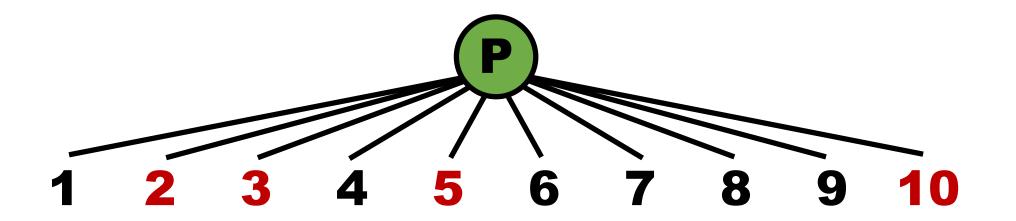


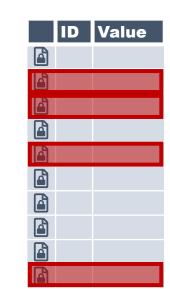
Example: Access Pattern Leakage



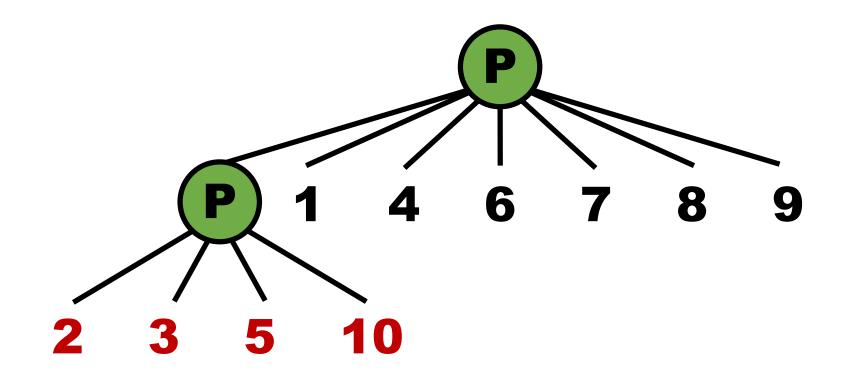
ID	Value

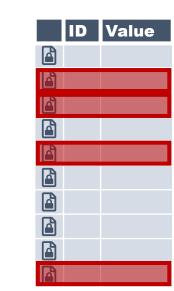




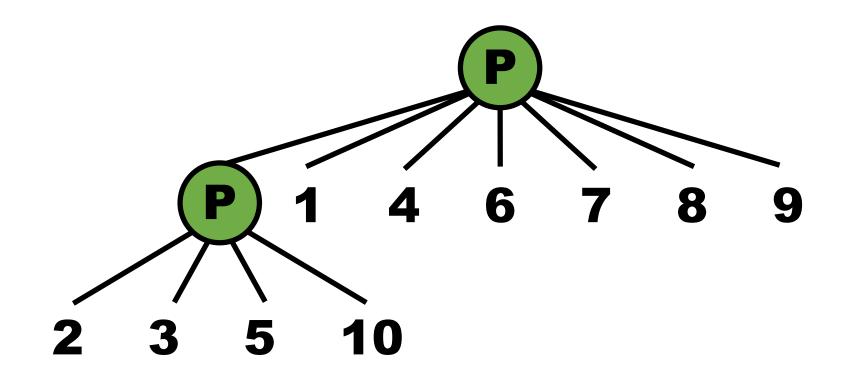






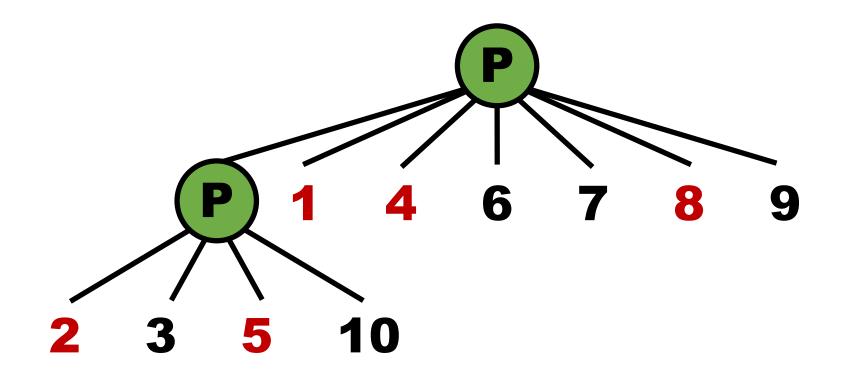


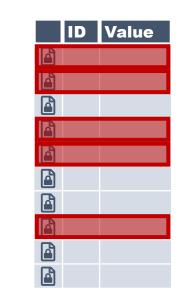




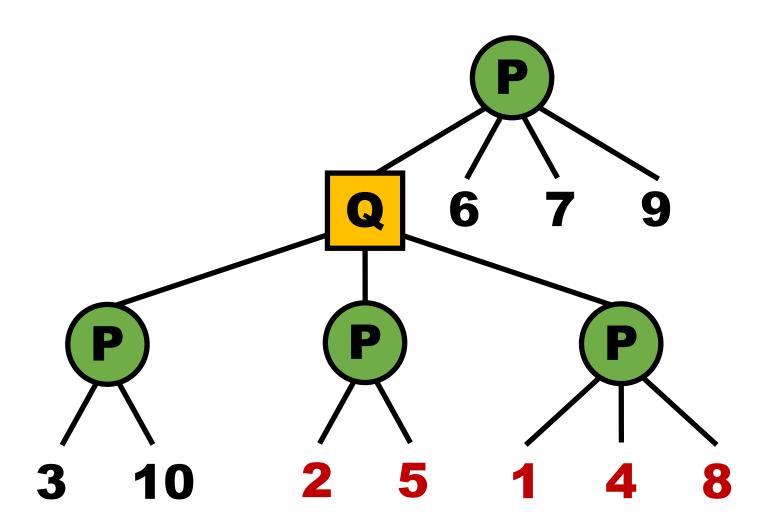
	ID	Value
屈		

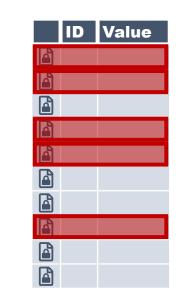




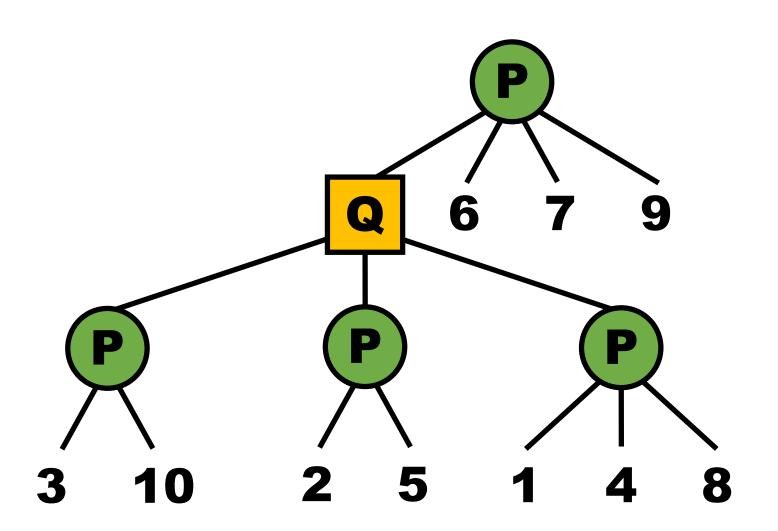






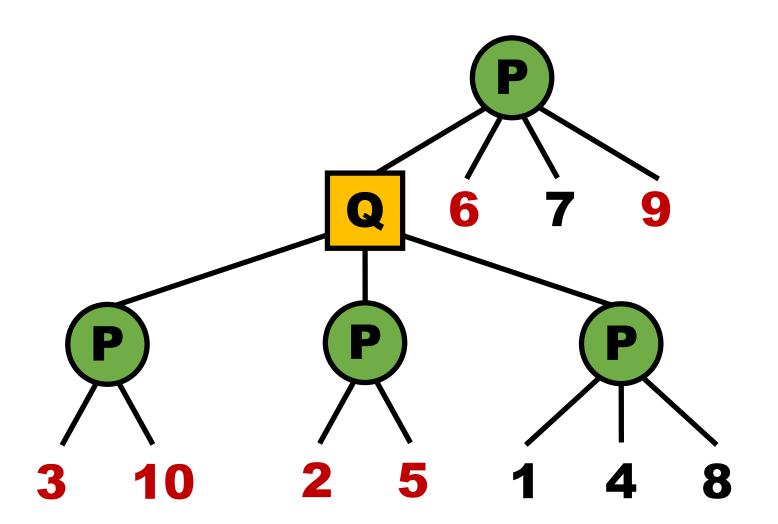


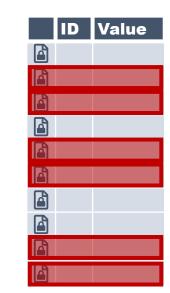




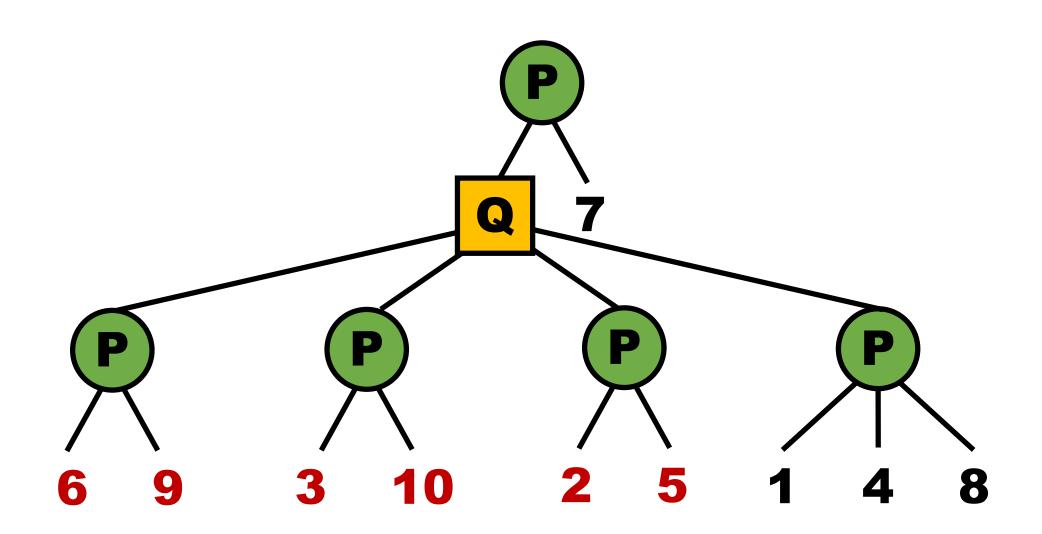
ID	Value

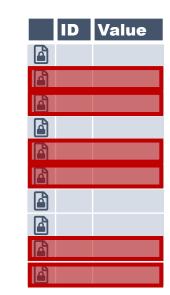




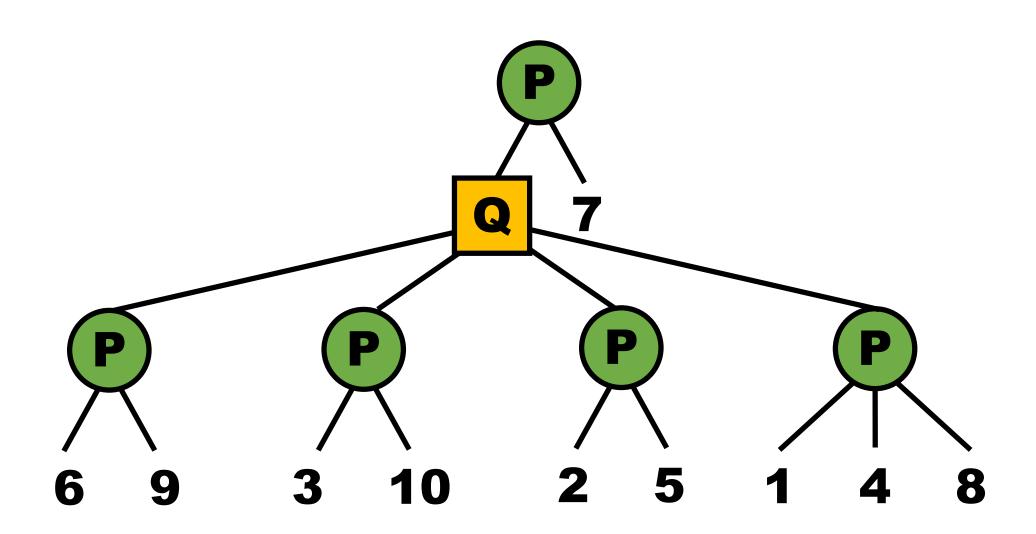






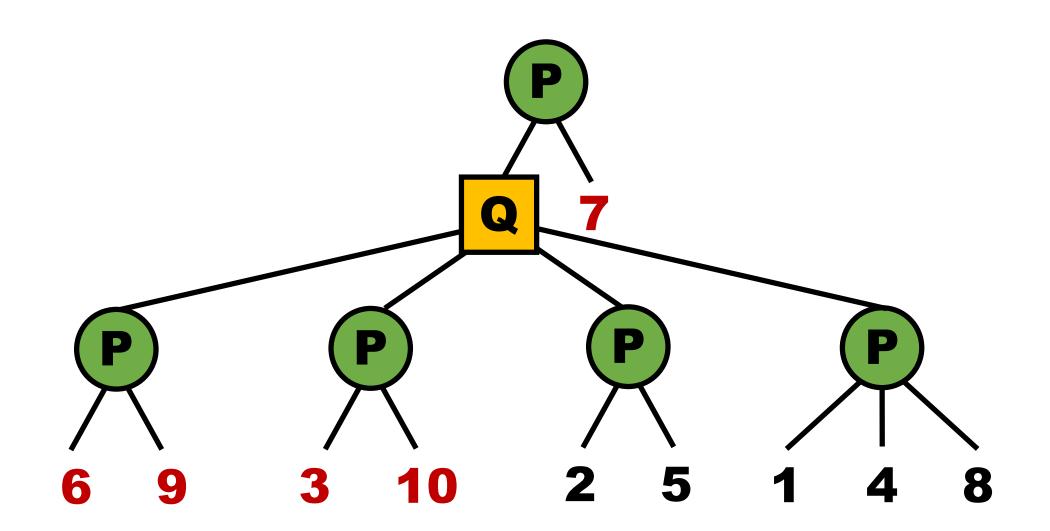


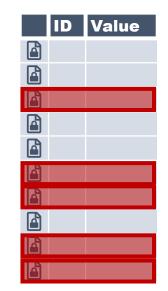




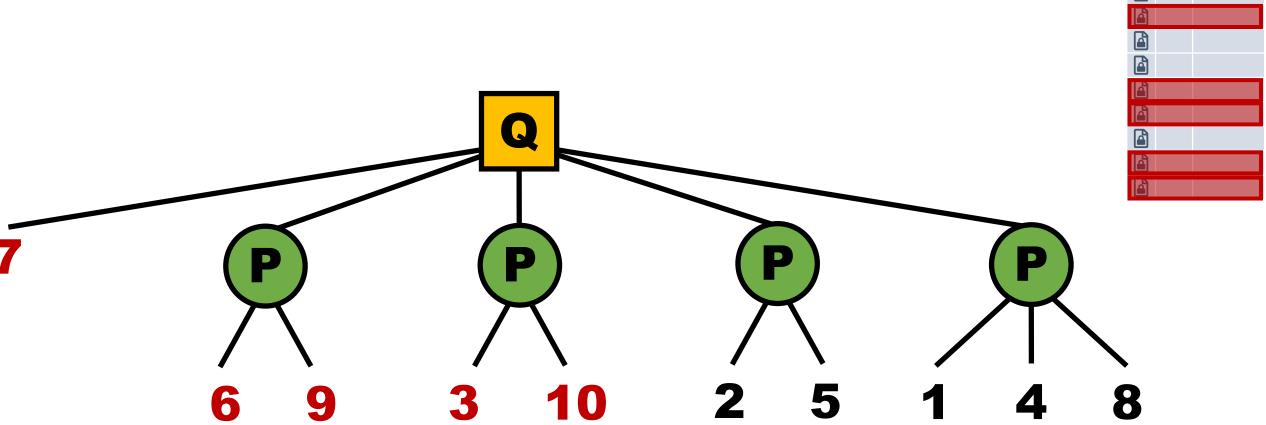
ID	Value



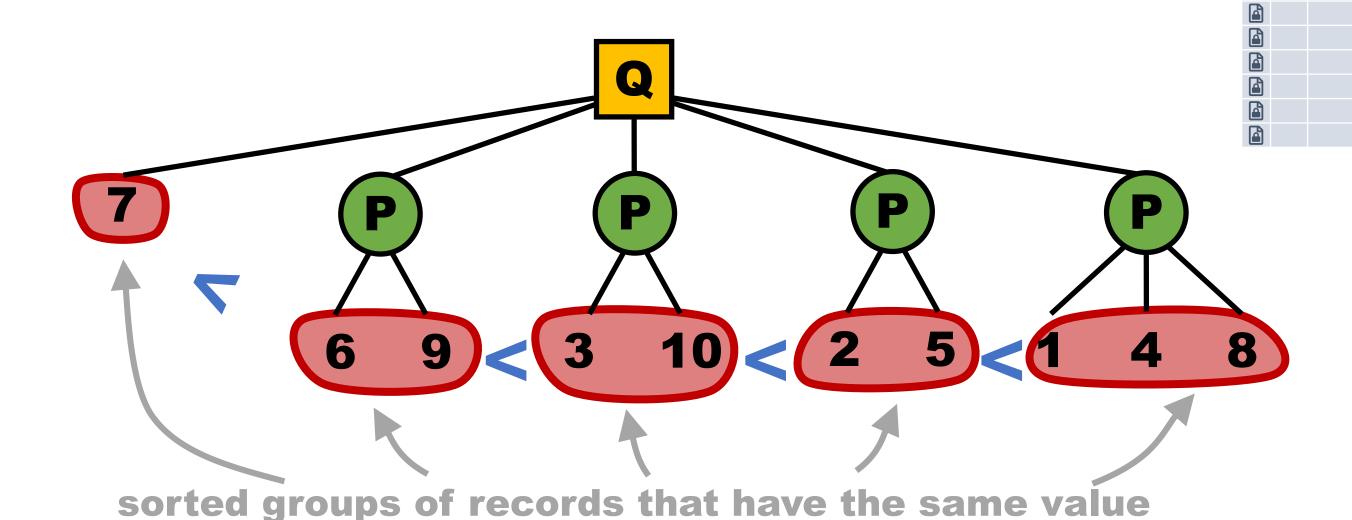




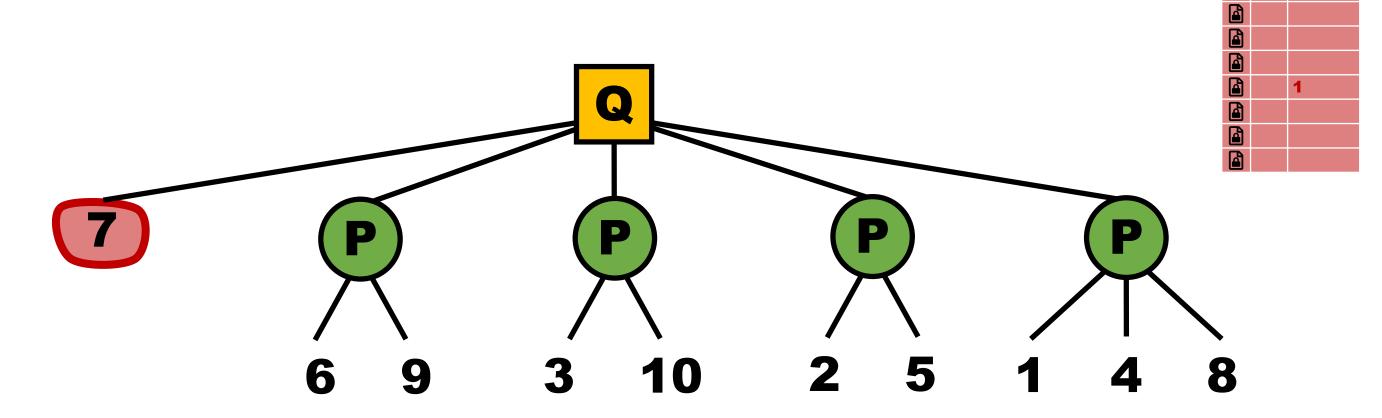




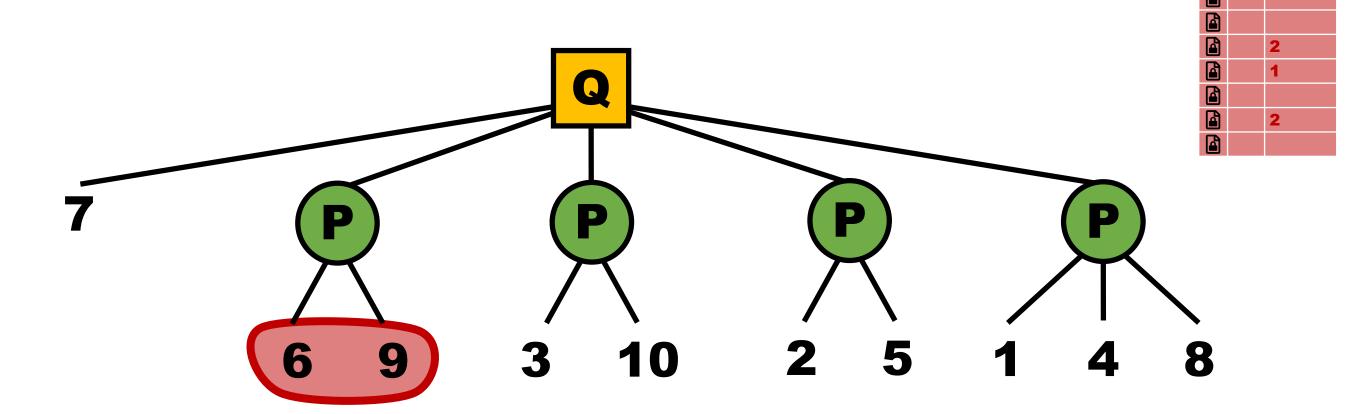




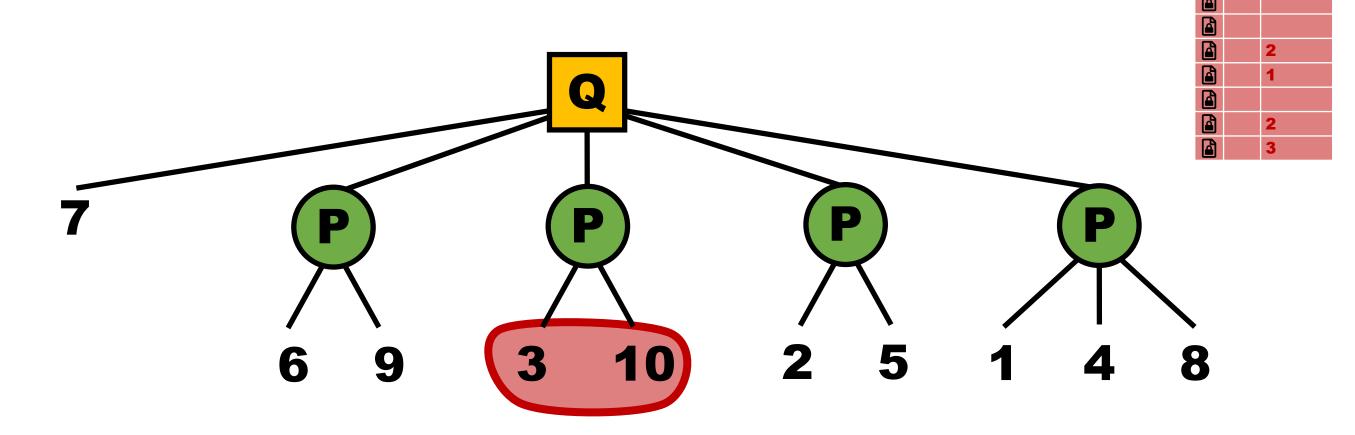




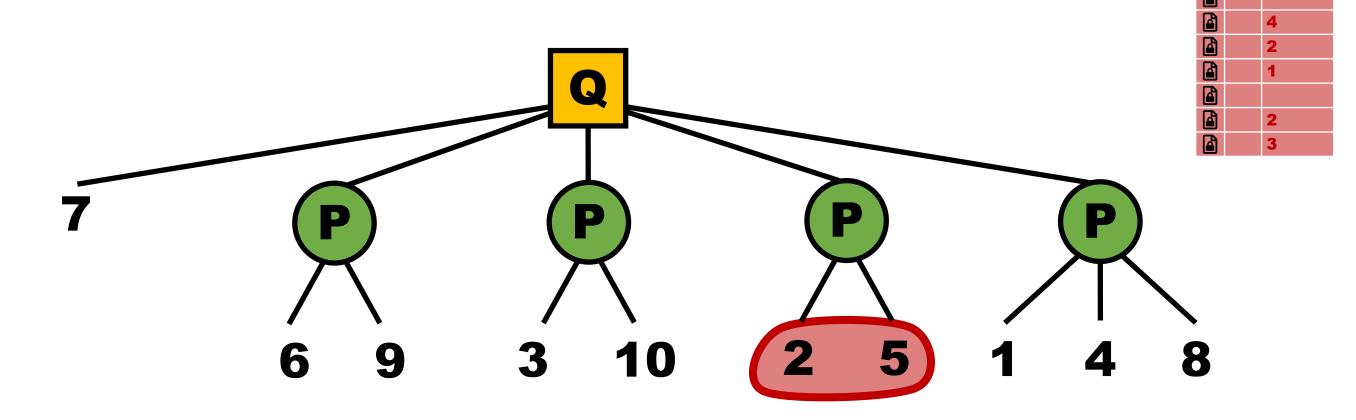




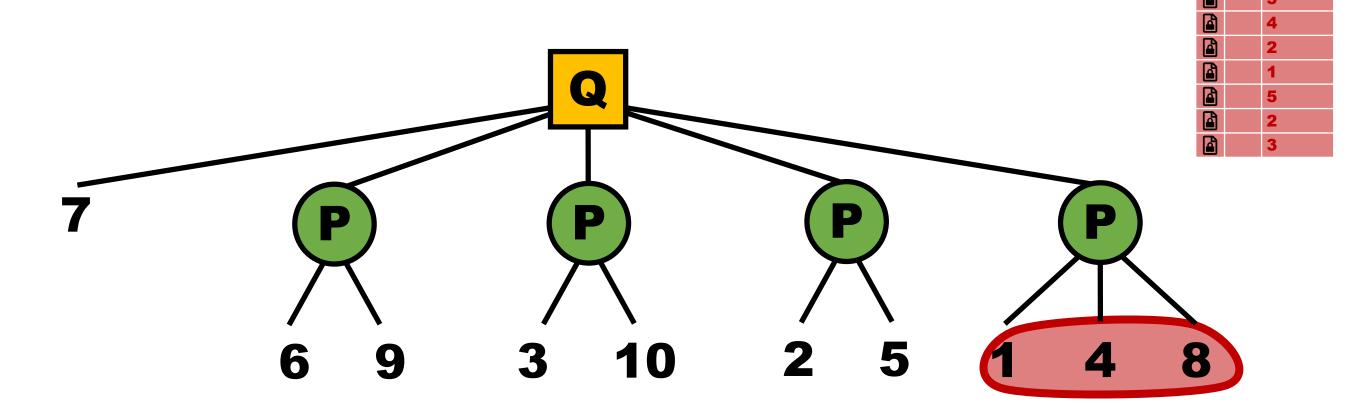
















ID	Value
1	5
2	4
3	3
4	5
5	4
6	2
7	1
8	5
9	2
10	3



How Many Queries Are Needed?

Suppose values are in {1, ..., N}.

EXACT RECONSTRUCTION



How Many Queries Are Needed?

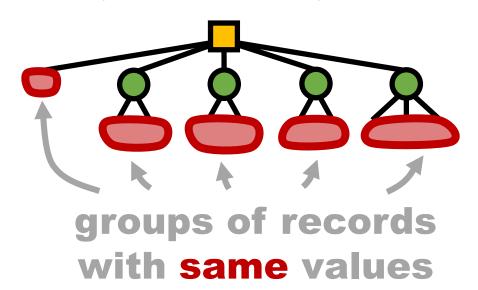
Suppose values are in {1, ..., N}.

EXACT RECONSTRUCTION

N log N queries

e.g.,

N=10: **23**, N=100: **461**, N=1000: **6908**





How Many Queries Are Needed?

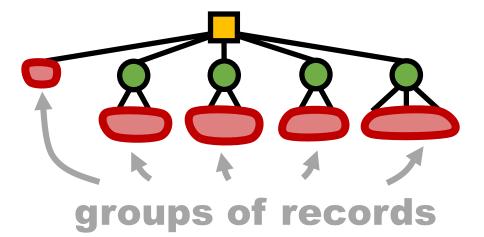
Suppose values are in {1, ..., N}.

EXACT RECONSTRUCTION

N log N queries

e.g.,

N=10: **23**, N=100: **461**, N=1000: **6908**



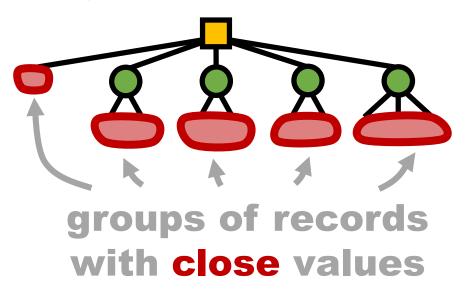
with same values

APPROXIMATE RECONSTRUCTION

Depends only on precision

(values within some % of N)

e.g., 10%: **23**, 5%: **60**, 2%: **460**



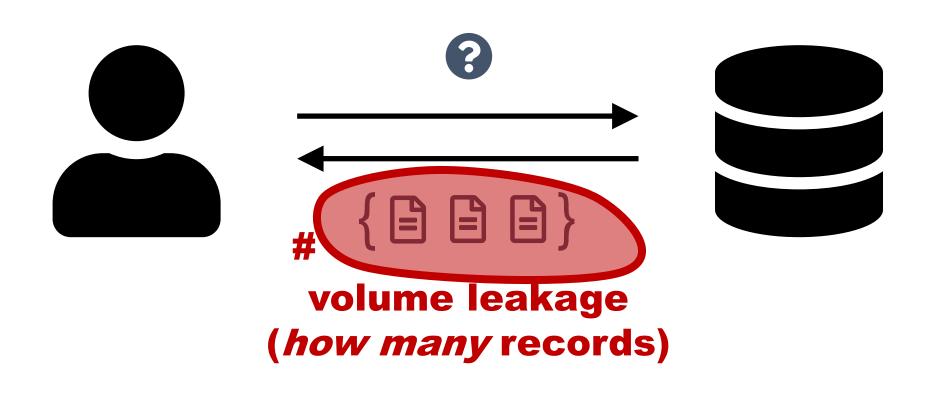


Exploiting Access Pattern Leakage

- Leaking which rows match a query can break encryption.
- PQ trees help organize the leakage along the way.
- Recovering approximate values takes even fewer queries.

Details: [Grubbs, Lacharité, Minaud, and Paterson, S&P 2019]





ID	Value
1	3
2	1
3	15
4	41
5	1





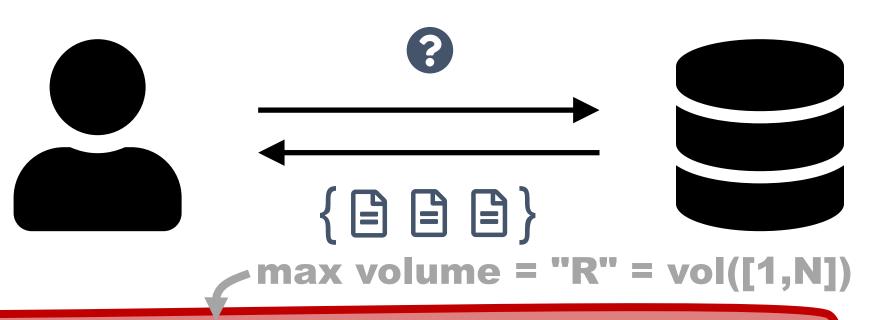
ID	Value
1	3
2	1
3	15
4	41
5	1



3, 16, 20, 5, 8, 11, 12, 1, 17, 19

all possible range volumes





ID	Value
1	3
2	1
3	15
4	41
5	1



3, 16, 20, 5, 8, 11, 12, 1, 17, 19



The Idea: Identify Elementary Volumes

ELEMENTARY RANGES	ELEMENTARY VOLUMES
[1,1]	# rows matching [1,1],
[1,2]	# rows matching [1,2],
[1,N-1]	# rows matching [1,N-1]
[1,N]	# rows matching [1,N]



The Idea: Identify Elementary Volumes

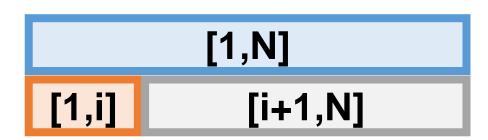
```
ELEMENTARY RANGES
                                  ELEMENTARY VOLUMES
                         [1,1]
                                  # rows matching [1,1],
                         [1,2]
                                  # rows matching [1,2],
                       [1,N-1]
                                  # rows matching [1,N-1]
                        [1,N]
                                  # rows matching [1,N]
vol([1,N]) - vol([1,N-1]) = vol([N,N])
```



Elementary Properties

1. Such volumes are **R**-complemented:

$$vol([1,i]) + vol([i+1,N]) = R$$



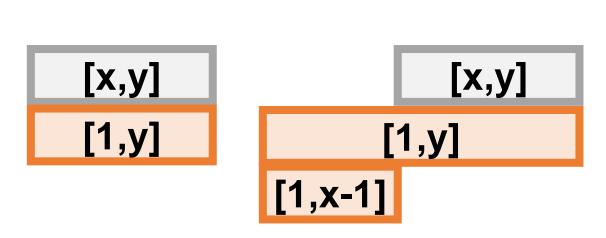


Elementary Properties

1. Such volumes are R-complemented:
vol([1,i]) + vol([i+1,N]) = R

[1,N]
[1,i] [i+1,N]

2. Every range [x,y] has the form [1,y] or [1,y] \ [1,x-1]





Elementary Properties

1. Such volumes are **R**-complemented: vol([1,i]) + vol([i+1,N]) = R

[1,N] [1,i] [i+1,N]

2. Every range [x,y] has the form [1,y] or $[1,y] \setminus [1,x-1]$ [x,y][1,y]

[x,y][1,y] [1,x-1]

Difference of any two such ranges is a range:

 $[1,y] \setminus [1,x] = [x+1,y]$

[1,y] [1,x][x+1,y]



Let's Build a Graph

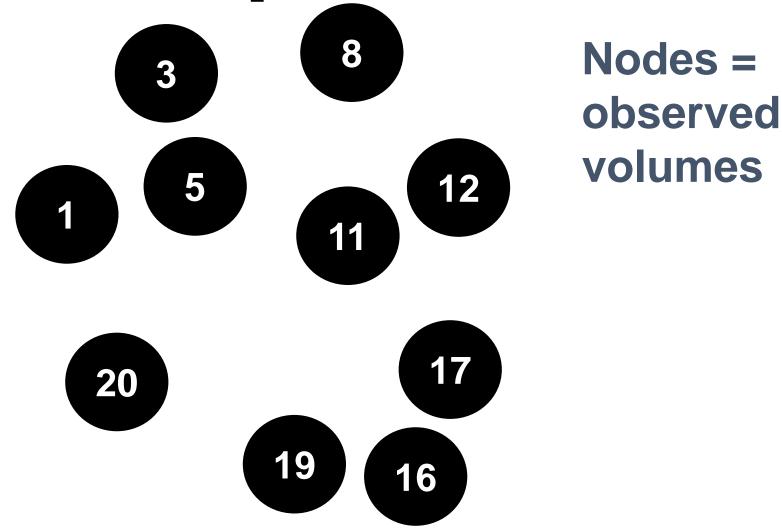


Let's Build a Graph

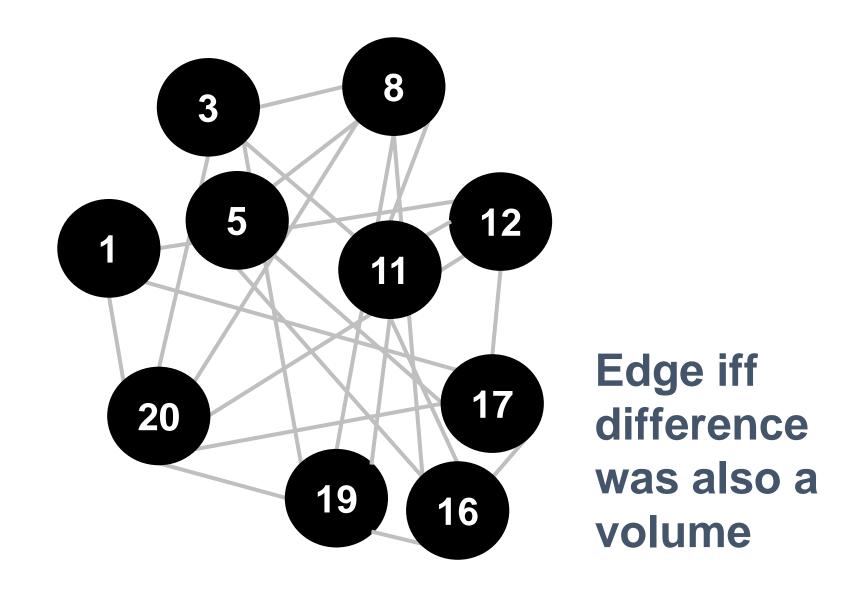
Nodes = observed volumes



Let's Build a Graph

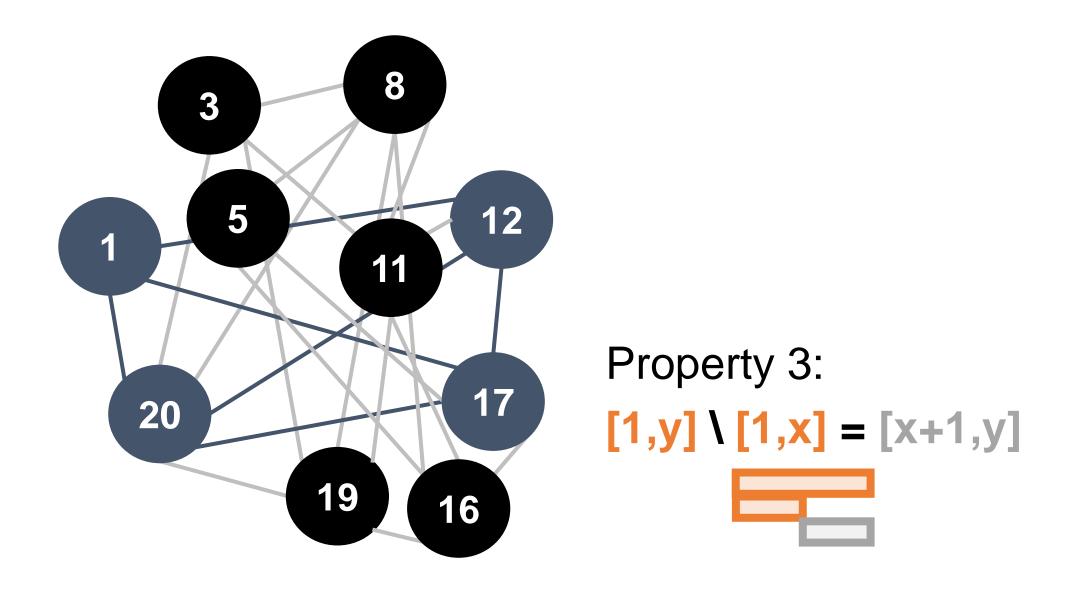






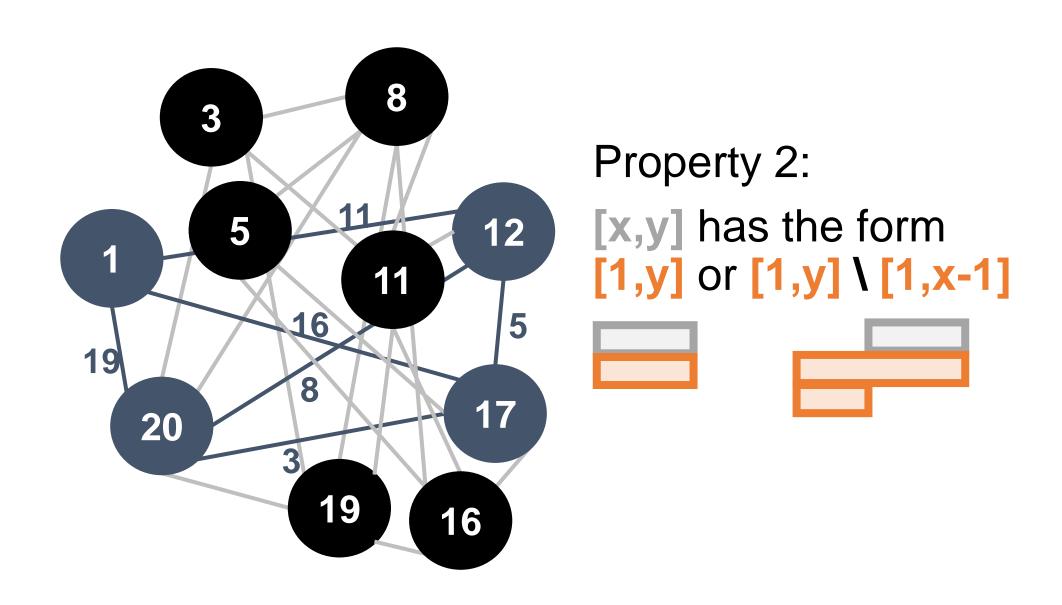


Elementary volumes form a clique





Elementary volumes form a clique... that generates all volumes





Goal: identify set of elementary volumes

Idea: build a graph and find a clique in it



Goal: identify set of elementary volumes

Idea: build a graph and find a clique in it

Phases:

- 1. Pre-processing
- 2. "Traditional" clique-finding

usually not necessary, see my paper for details



Phase 1: Pre-Processing

NECESSARY VOLUMES

REAL ELEMENTARY

ELEMENTARY

ELEMENTARY **VOLUMES**

CANDIDATE VOLUMES



Phase 1: Pre-Processing

NECESSARY VOLUMES

REAL ELEMENTARY

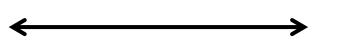
ELEMENTARY

ELEMENTARY

ELEMENTARY **VOLUMES**

CANDIDATE VOLUMES

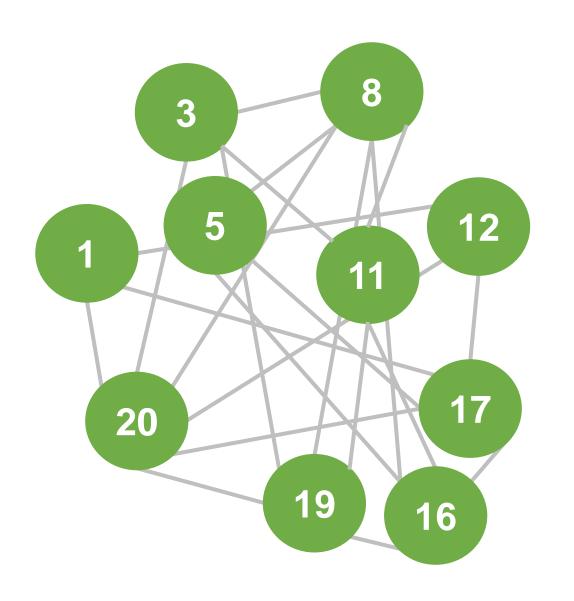
AUGMENT NECESSARY VOLUMES



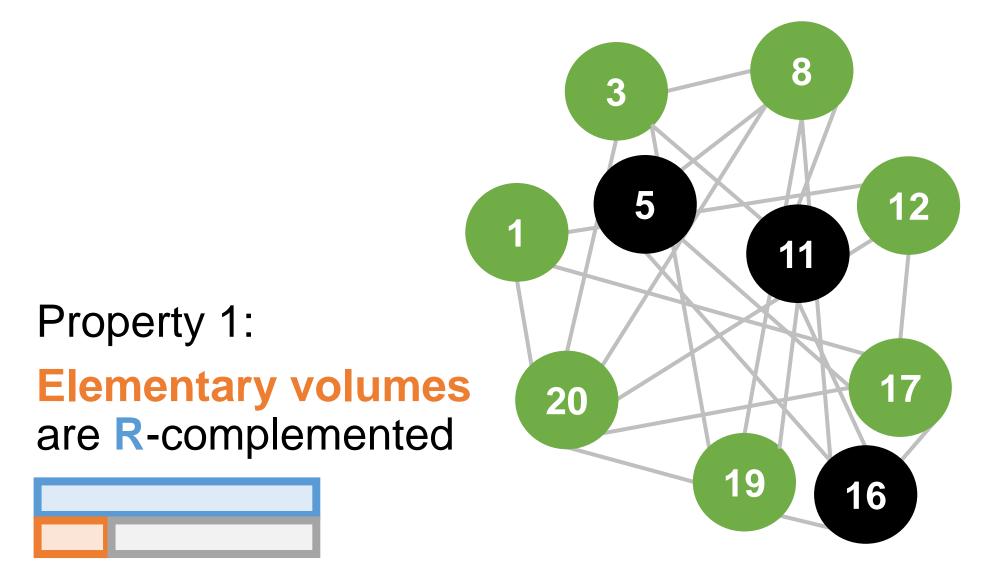
REDUCE CANDIDATE VOLUMES



Example



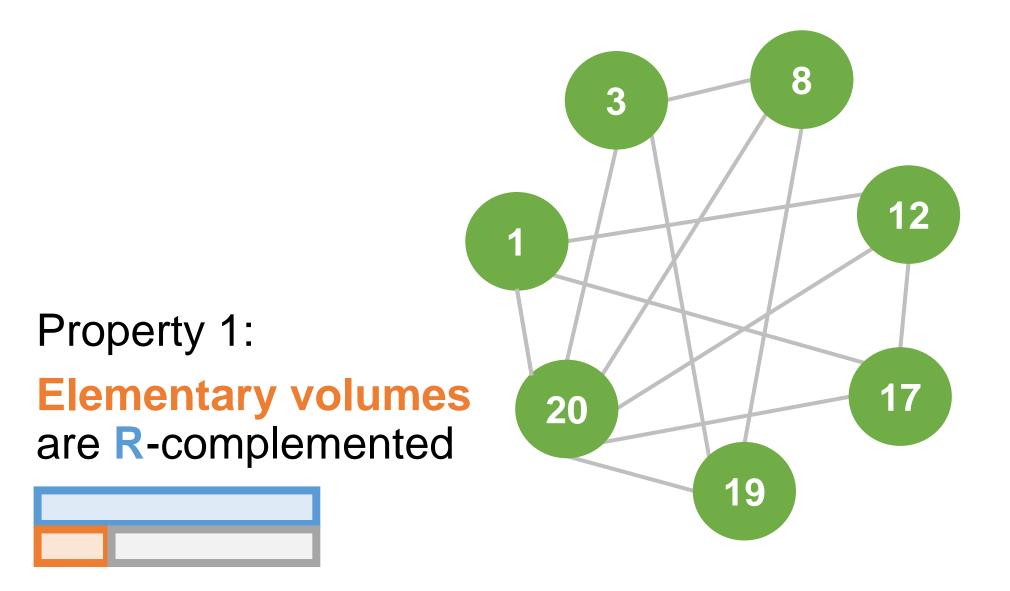




REDUCE

Remove nodes without *R*-complements

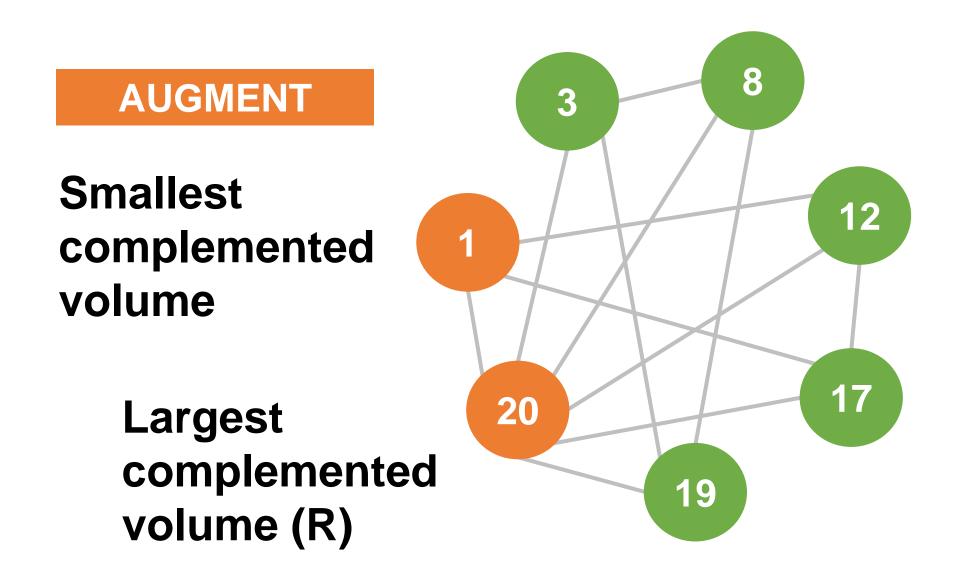




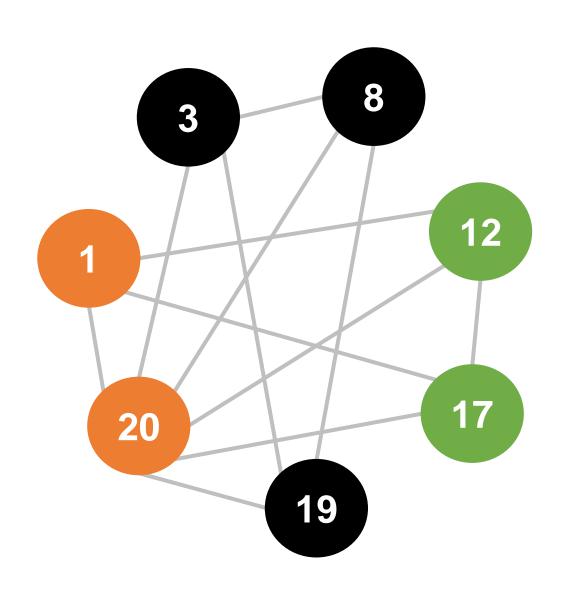
REDUCE

Remove nodes without *R*-complements







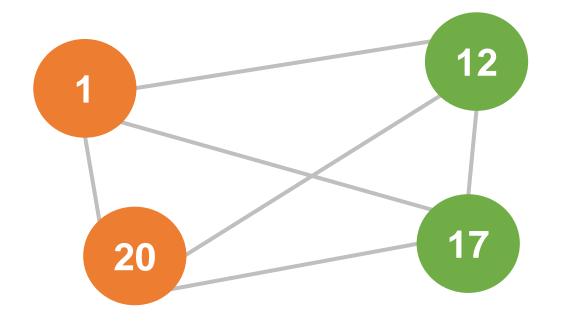


REDUCE

Remove nodes not adjacent to all necessary volumes



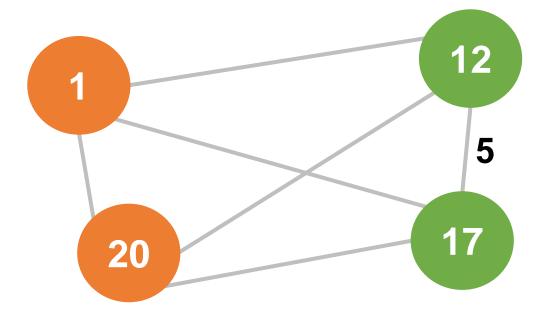
REDUCE



Remove nodes not adjacent to all necessary volumes



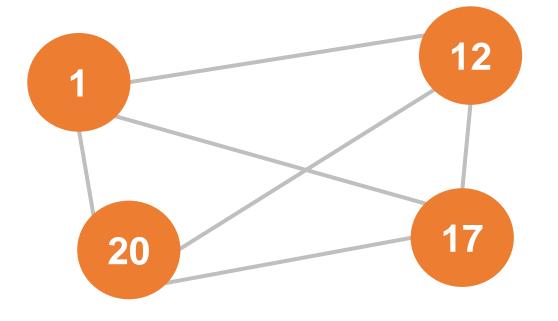
AUGMENT



Add endpoints of volumes that occur only once, as an edge

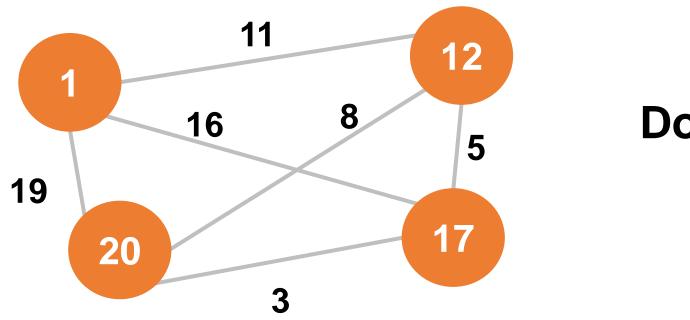


AUGMENT



Add endpoints of volumes that occur only once, as an edge

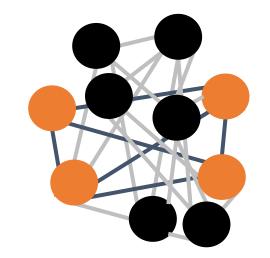




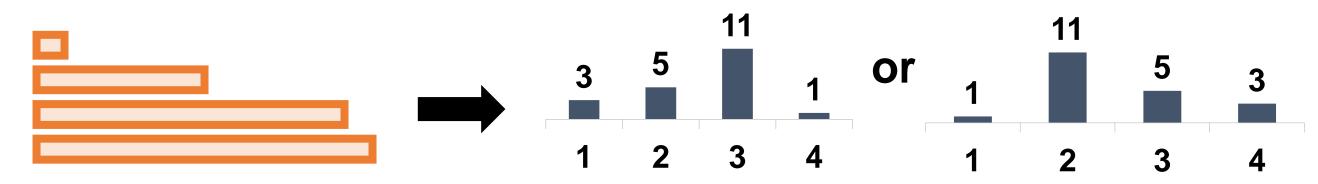
Done!



- Build a graph using all observed volumes
- Use properties of range queries to identify elementary volumes, which form a clique



 Use elementary volumes to directly reconstruct all counts in the database





How Many Queries Are Needed?

Suppose values in {1, ..., N}, queries drawn equally at random.



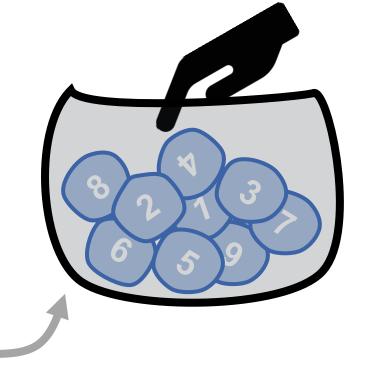
How Many Queries Are Needed?

Suppose values in {1, ..., N}, queries drawn equally at random.

N² log N queries expected to see all volumes

Coupon collector bound

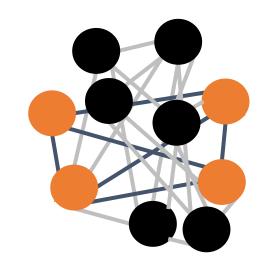
- Q: How many random draws expected till all coupons have been drawn at least once?
- **A**: ≈ N² log N



N(N+1)/2 "coupons" (one for every range)



- Build a graph using all observed volumes
- Use properties of range queries to identify elementary volumes, which form a clique



 Use elementary volumes to directly reconstruct all counts in the database

Details: [Grubbs, Lacharité, Minaud, and Paterson, CCS 2018]



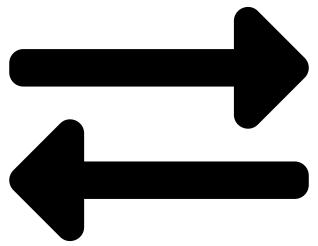
Outline

- 1. Existing approaches to securing a database
 - Securing data in transit, at rest, and in use
- 2. How to exploit leakage to break database encryption
 - Exploiting access pattern leakage and volume leakage
- 3. Security recommendations
 - Types of leakage, leaky operations, trade-offs



Approach

What can leak?



Where and when can it leak?



What Can Leak?

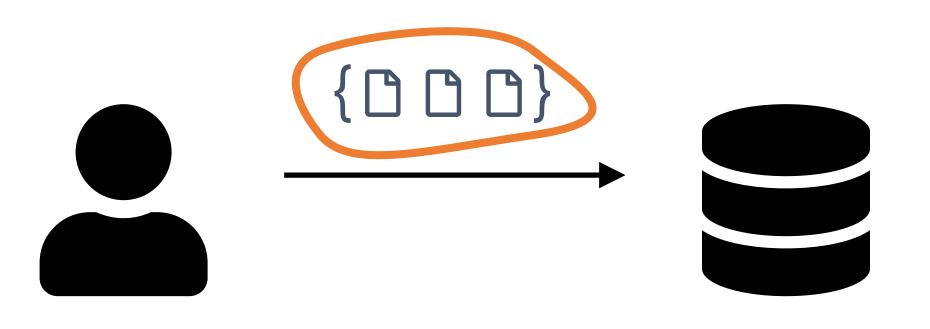
Properties related to...

- Values: order, distance, existence, number of distinct values, repetition, ...
- Queries: endpoints, repetition, width, inclusion, ...
- Responses: which rows matched, how many rows matched, repetition, ...

[Kamara et al., CRYPTO 2018]



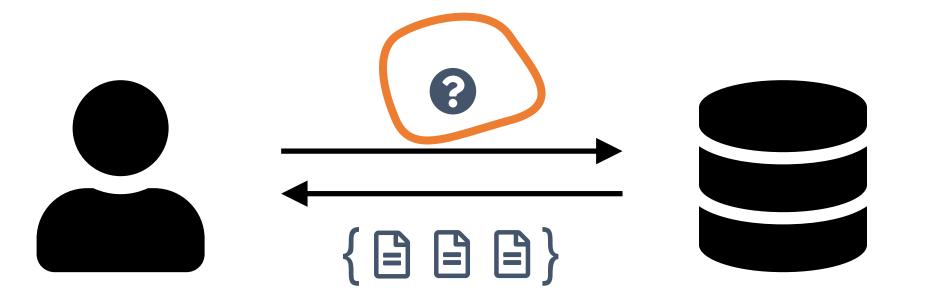
Where and When Can It Leak?



ID	Value
1	3
2	1
3	15
4	41
5	1



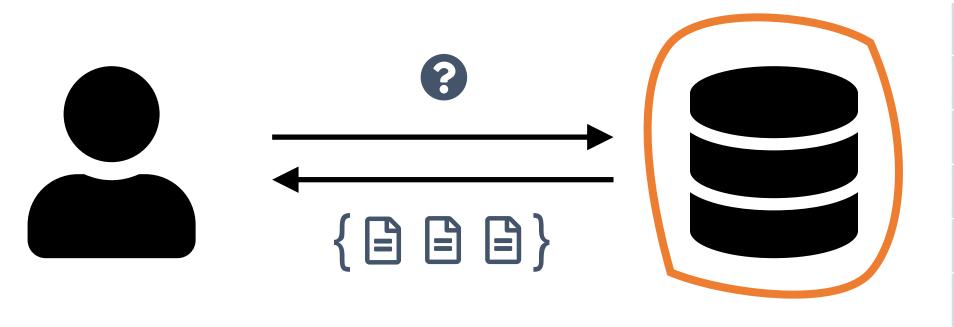
Where and When Can It Leak?



ID	Value
1	3
2	1
3	15
4	41
5	1
 	•••



Where and When Can It Leak?



ID	Value
1	3
2	1
3	15
4	41
5	1



Trade-offs



Trade-offs

MITIGATION TECHNIQUES

- Restricting query types
- Dummy records
- Dummy values
- Trusting hardware

•



Trade-offs

MITIGATION TECHNIQUES

- Restricting query types
- Dummy records
- Dummy values
- Trusting hardware

•

COSTS

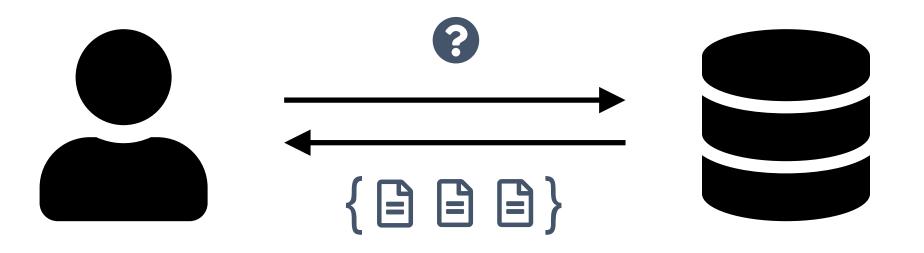
- Incomplete results
- Probabilistically correct results
- Efficiency
- Less compression/deduplication

•



Conclusion

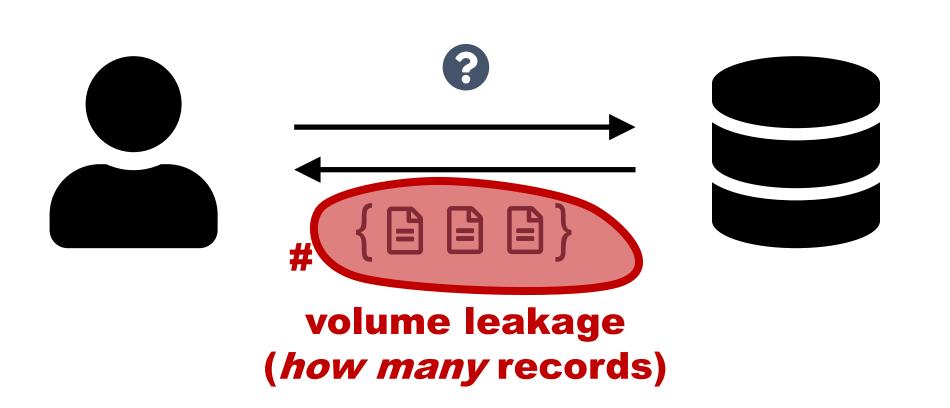


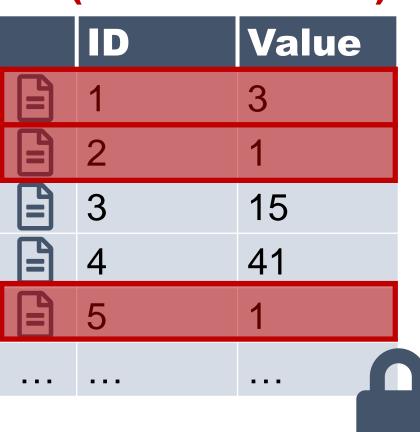


ID	Value
1	3
2	1
3	15
4	41
5	1

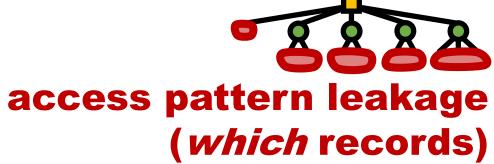


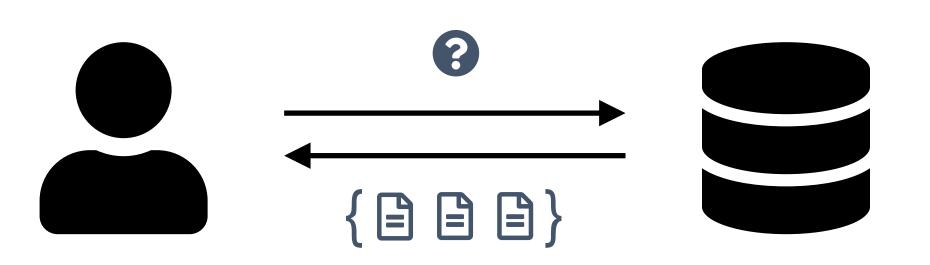
access pattern leakage (which records)

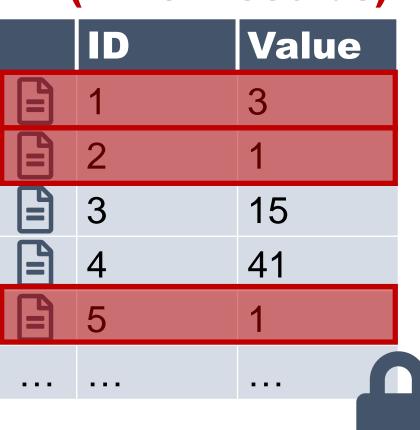




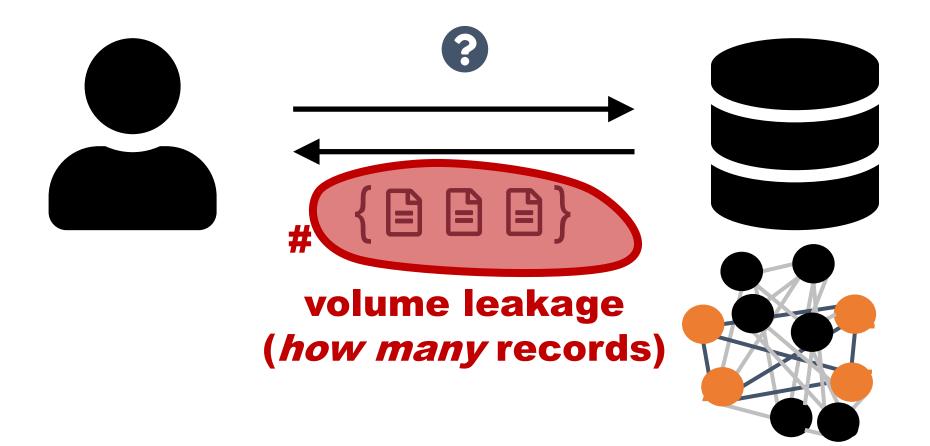






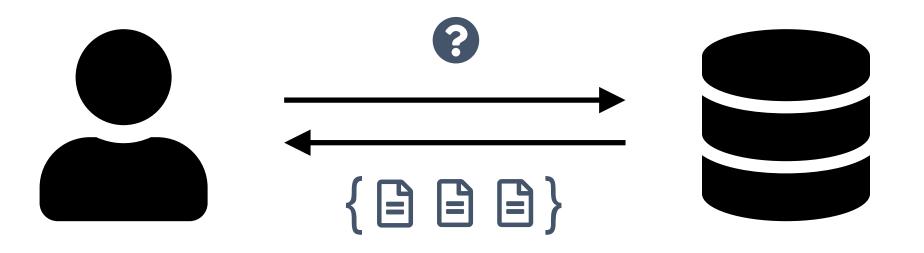






ID	Value
1	3
2	1
3	15
4	41
5	1

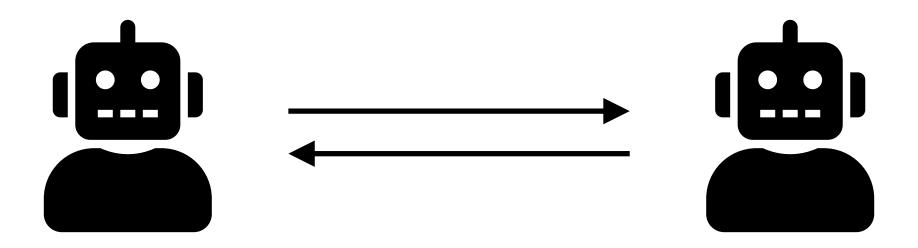




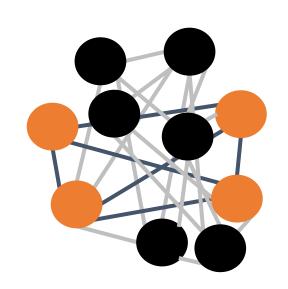
ID	Value
1	3
2	1
3	15
4	41
5	1



Side Channel Attacks

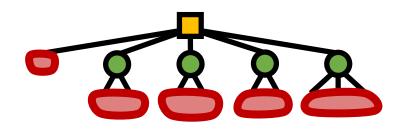






Thank you!

Questions?



mariesarah.lacharite@gmail.com @znevrfnenu



Black Hat Sound Bytes

- Databases have many unique side channels that leak information.
- Side channel attacks exploiting this leakage can break encryption.
- Understanding different kinds of leakage and during what operations they arise can help secure an encrypted database.