

Order-Revealing Encryption:

How to Search on Encrypted Data

Kevin Lewi and David J. Wu
Stanford University

Searching on Encrypted Data



The image shows a screenshot of the top portion of an Ars Technica article. At the top is a dark navigation bar with the 'ars TECHNICA' logo on the left, a search icon, a menu icon, and a 'SIGN IN' link. On the right side of the navigation bar is a small American flag icon with a dropdown arrow. Below the navigation bar, the article is categorized under 'RISK ASSESSMENT' in green text. The main headline reads 'Yahoo says half a billion accounts breached by nation-sponsored hackers'. Below the headline is a sub-headline: 'One of the biggest compromises ever exposes names, e-mail addresses, and much more.' At the bottom left of the article header, the author's name 'DAN GOODIN' and the date '9/22/2016, 1:21 PM' are displayed.

ars TECHNICA 🔍 ☰ SIGN IN ▾ 

RISK ASSESSMENT —

Yahoo says half a billion accounts breached by nation-sponsored hackers

One of the biggest compromises ever exposes names, e-mail addresses, and much more.

DAN GOODIN - 9/22/2016, 1:21 PM

Searching on Encrypted Data



The screenshot shows the top portion of a web browser displaying the New York Times website. The page is titled "BUSINESS DAY" and features a main headline: "Data Breach at Anthem May Forecast a Trend". The article is attributed to Reed Abelson and Julie Creswell, dated February 6, 2015. The page includes a navigation bar with "SECTIONS", "HOME", and "SEARCH" options, along with "SUBSCRIBE" and "LOG IN" buttons. Social media sharing icons for Facebook, Twitter, Email, and a share icon are visible below the byline.

SECTIONS HOME SEARCH The New York Times SUBSCRIBE LOG IN

BUSINESS DAY

Data Breach at Anthem May Forecast a Trend

By REED ABELSON and JULIE CRESWELL FEB. 6, 2015

f t e |

Searching on Encrypted Data



A screenshot of a web browser displaying a Reuters news article. The browser's address bar is blurred. The page features a dark navigation bar with the Reuters logo and various menu items. The article's title and date are prominently displayed.

EDITION: UNITED STATES ▾

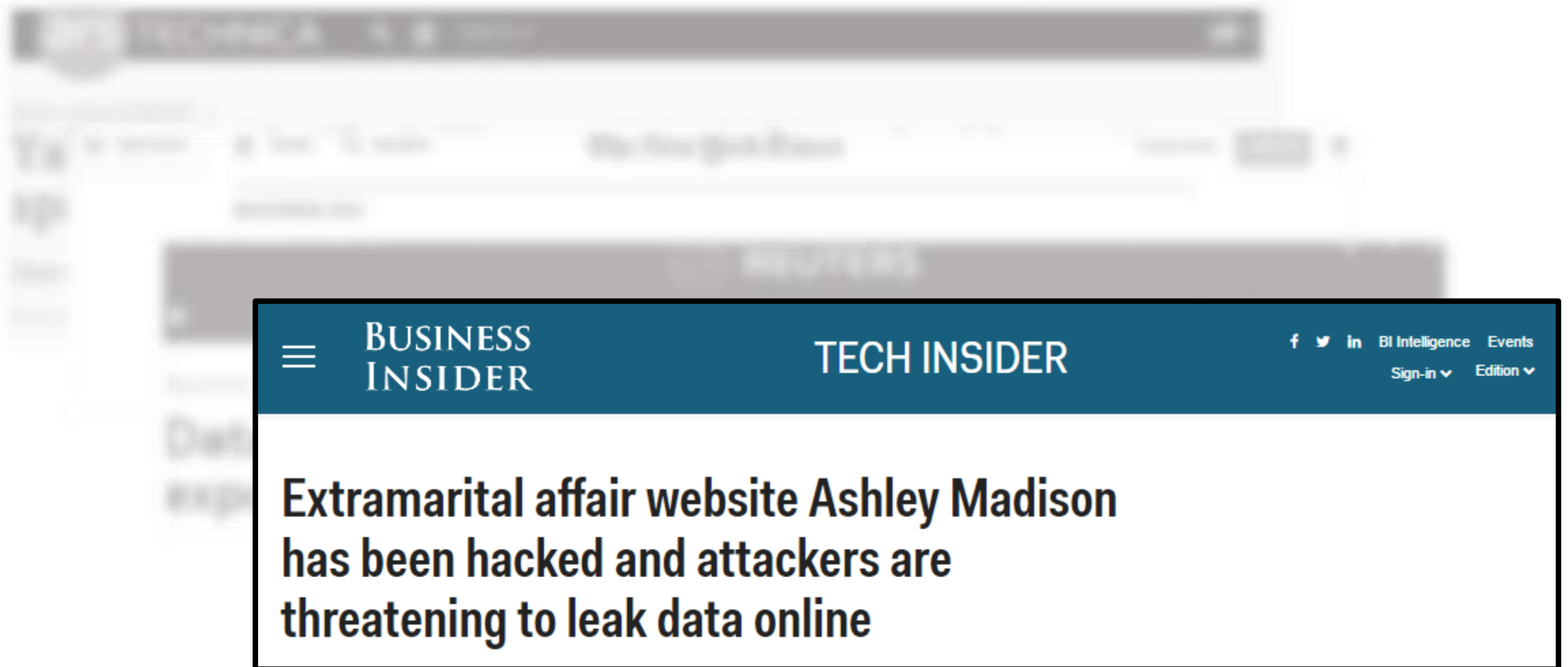
 **REUTERS**   

 [Business](#) [Markets](#) [World](#) [Politics](#) [Tech](#) [Commentary](#) [Breakingviews](#) [Money](#) [Life](#)   

POLITICS | Mon Dec 28, 2015 | 4:52pm EST

Database of 191 million U.S. voters exposed on Internet: researcher

Searching on Encrypted Data



≡ BUSINESS INSIDER

TECH INSIDER

f t in BI Intelligence Events
Sign-in ▾ Edition ▾

Extramarital affair website Ashley Madison has been hacked and attackers are threatening to leak data online

Searching on Encrypted Data



The image shows a screenshot of the TechWeek Europe website. The page features a dark blue header with the TechWeek Europe logo on the left, a search bar in the center, and navigation links for 'Follow us', 'NEWSLETTER', and 'CONNECT' on the right. Below the header is a dark blue navigation bar with links for 'Menu', 'Mobility', 'Networks', 'Cloud', 'Security', 'Workspace', 'Projects', 'Events', 'Tech Club', 'IT Life', 'Jobs', 'Mobility Focus', 'Whitepapers', and 'Partnerzone'. The main content area displays a news article with the headline 'eBay Asks 128 Million Customers To Change Their Passwords After Hack' and the byline 'Max Smolaks, May 21, 2014, 4:55 pm'. The article text is partially visible and blurred, showing the beginning of a paragraph: 'eBay has announced that its account password database has been hacked and attackers are threatening to leak data online'.

TechWeek
europe

Search ...

Follow us ▾ NEWSLETTER CONNECT

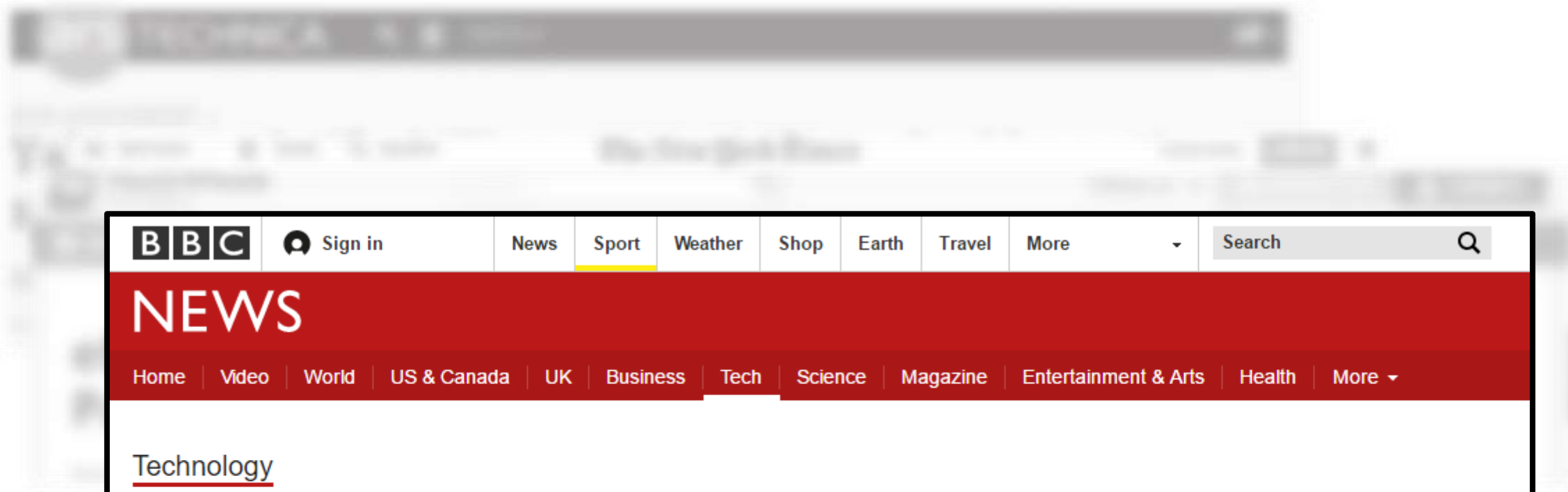
Menu Mobility Networks Cloud Security Workspace Projects Events Tech Club IT Life Jobs Mobility Focus Whitepapers Partnerzone

eBay Asks 128 Million Customers To Change Their Passwords After Hack

Max Smolaks, May 21, 2014, 4:55 pm

eBay has announced that its account password database has been hacked and attackers are threatening to leak data online

Searching on Encrypted Data



The image shows a screenshot of the BBC News website. The top navigation bar includes the BBC logo, a 'Sign in' button, and menu items for News, Sport, Weather, Shop, Earth, Travel, and More. A search bar is located on the right side of the navigation bar. Below the navigation bar, the word 'NEWS' is displayed in large white letters on a red background. Underneath, there is a secondary navigation bar with links for Home, Video, World, US & Canada, UK, Business, Tech, Science, Magazine, Entertainment & Arts, Health, and More. The main content area features a sub-section for 'Technology' with a red underline. The headline of the article is 'Millions of hacked LinkedIn IDs advertised 'for sale''. Below the headline, the date '18 May 2016' and the category 'Technology' are shown, along with a 'Share' button.

BBC Sign in News Sport Weather Shop Earth Travel More Search

NEWS

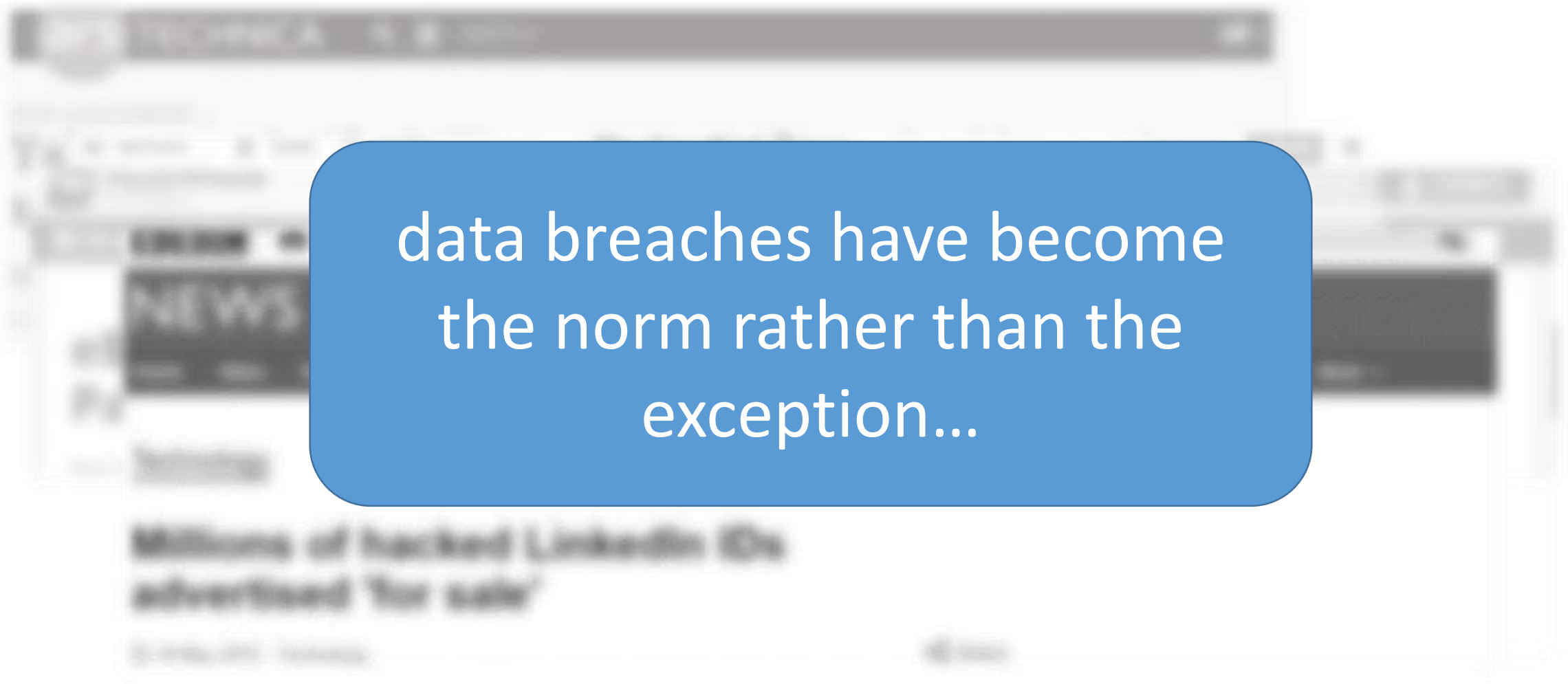
Home Video World US & Canada UK Business Tech Science Magazine Entertainment & Arts Health More

Technology

Millions of hacked LinkedIn IDs advertised 'for sale'

18 May 2016 | Technology [Share](#)

Searching on Encrypted Data

A blurred screenshot of a news article is shown in the background. A prominent blue rounded rectangle is overlaid on the center of the image, containing white text. The text in the blue box reads: "data breaches have become the norm rather than the exception...". The background text is mostly illegible due to blurring, but some words like "NEWS" and "Millions of hacked LinkedIn IDs" are visible.

data breaches have become
the norm rather than the
exception...

Why Not Encrypt?

“because it would have hurt Yahoo’s ability to index and search messages to provide new user services”
~Jeff Bonforte (Yahoo SVP)

Millions of hacked LinkedIn IDs
advertised for sale

Order-Revealing Encryption [BLRSZZ'15]

secret-key encryption
scheme

sk



client

$$ct_1 = \text{Enc}(sk, 123)$$

$$ct_2 = \text{Enc}(sk, 512)$$

$$ct_3 = \text{Enc}(sk, 273)$$



server

Which is greater:
the value encrypted
by ct_1 or the value
encrypted by ct_2 ?

(legacy-friendly)
range queries on
encrypted data

Order-Revealing Encryption [BLRSZZ'15]

given any two ciphertexts

$$ct_1 = \text{Enc}(sk, x)$$

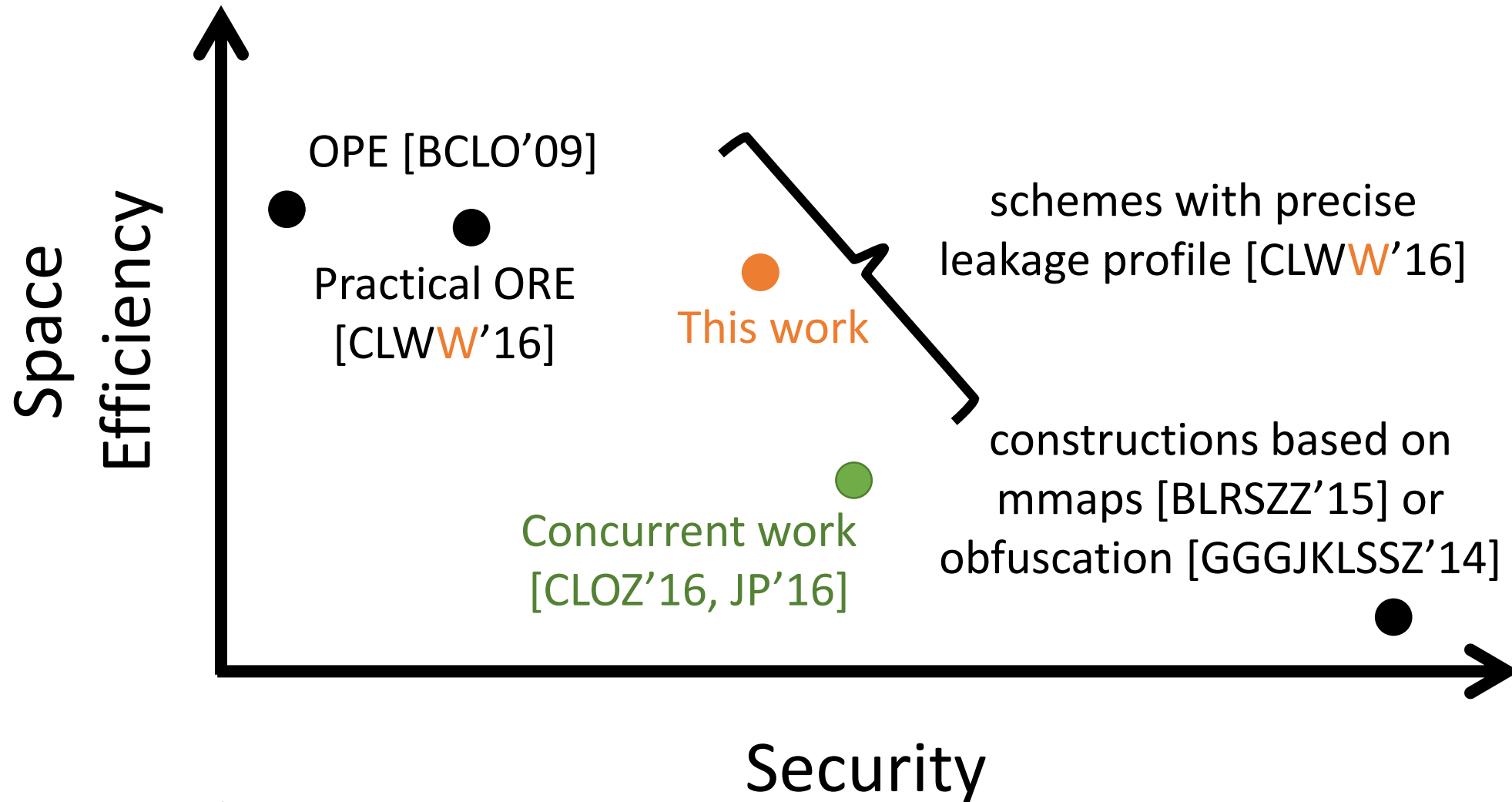
$$ct_2 = \text{Enc}(sk, y)$$

$$x > y$$

there is a public
function for performing
comparisons

OPE [BCLO'09]: comparison
function is numeric
comparison on ciphertexts

The Landscape of ORE



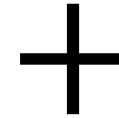
not drawn to scale

Inference Attacks [NKW'15, DDC'16, GSBNR'16]



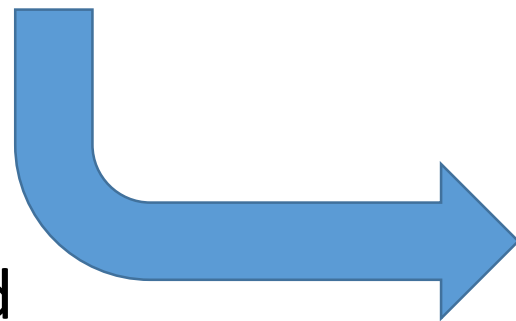
ID	Name	Age	Diagnosis
wpjOos	2wzXW8	SqX9l9	KqLUXE
XdXdg8	y9GFpS	gwilE3	MJ23b7
P6vKhW	EgN0Jn	S0pRJe	aTaeJk
orJRe6	KQWy9U	tPWF3M	4FBEO0

encrypted database



public information

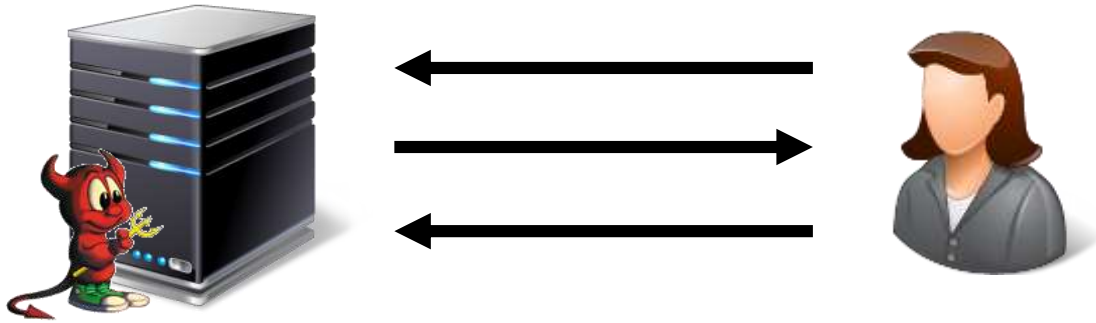
frequency and
statistical analysis



ID	Name	Age	Diagnosis
???	Alice	30-35	2
???	Bob	45-50	3
???	Charlie	40-45	2
???	???	40-45	4

plaintext
recovery

Online vs. Offline Security



adversary sees encrypted database +
queries and can interact with the database

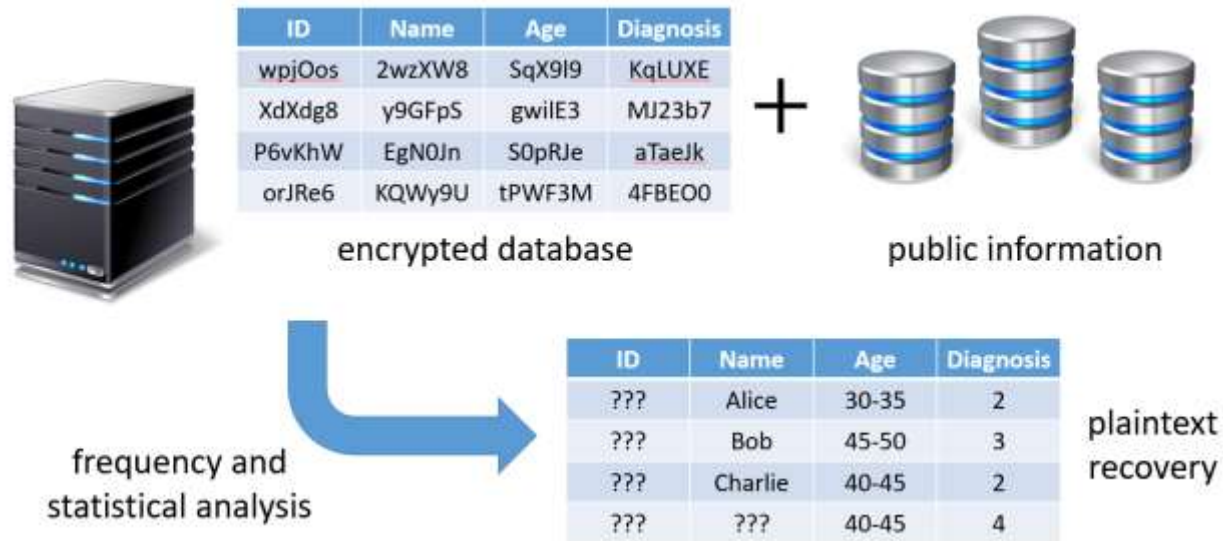
online attacks (e.g., active corruption)
offline attacks (e.g., passive snapshots)



adversary only sees contents
of encrypted database

typical database breach:
contents of database are stolen
and dumped onto the web

Inference Attacks [NKW'15, DDC'16, GSBNR'16]



PPE schemes always reveal certain properties (e.g., equality, order) on ciphertexts and thus, are vulnerable to offline inference attacks

Can we fully defend against offline inference attacks while remaining legacy-friendly?

This Work

Can we fully defend against offline inference attacks while remaining legacy-friendly?

Trivial solution: encrypt the entire database, and have client provide decryption key at query time

Desiderata: an ORE scheme that enables:

- perfect offline security
- limited leakage in the online setting

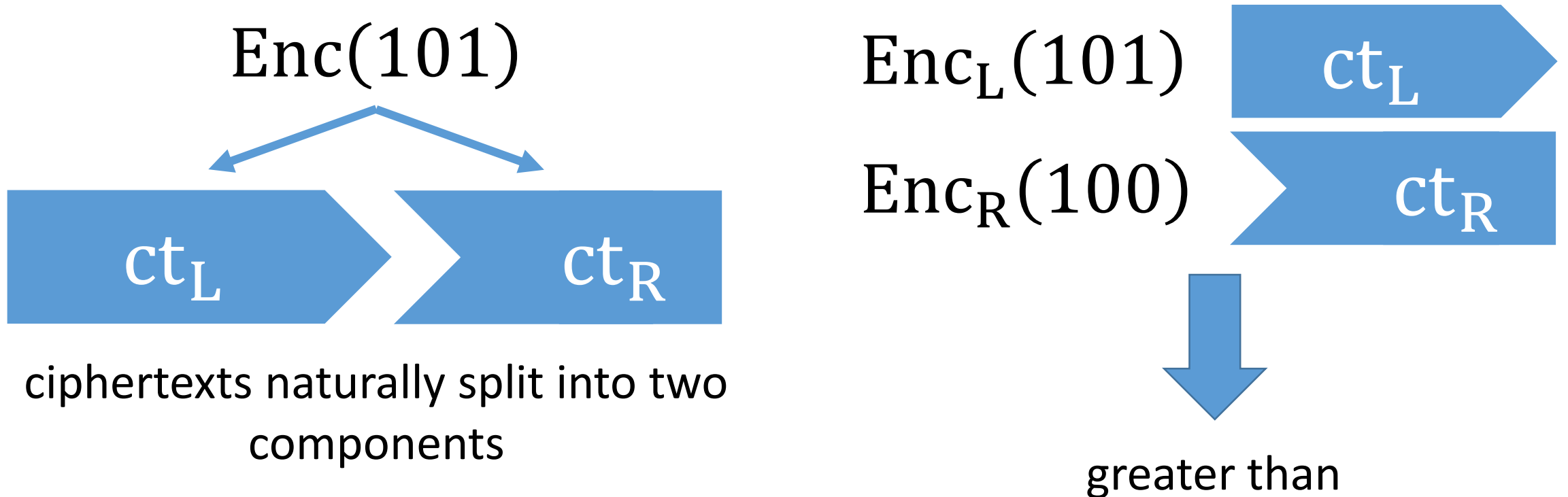


But no online security!

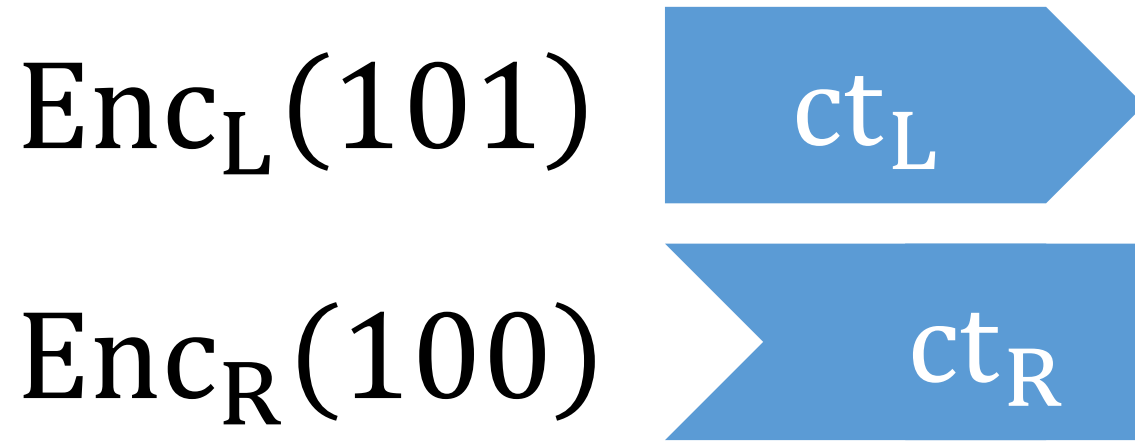
ORE with Additional Structure

Focus of this work: performing range queries on encrypted data

Key primitive: order-revealing encryption scheme where ciphertexts have a “decomposable” structure



ORE with Additional Structure



comparison can be performed
between left ciphertext and
right ciphertext

right ciphertexts provide
semantic security!



robustness against offline
inference attacks!

Encrypted Range Queries

ID	Name	Age	Diagnosis
0	Alice	31	2
1	Bob	47	3
2	Charlie	41	2
3	Inigo	45	4

build encrypted index

store right ciphertexts in sorted order

Age	ID
Enc _R (31)	Enc(0)
Enc _R (41)	Enc(2)
Enc _R (45)	Enc(3)
Enc _R (47)	Enc(1)

record IDs encrypted under independent key



Name	ID
Enc _R (Alice)	Enc(0)

Age	ID
Enc _R (31)	Enc(0)

Diagnosis	ID
Enc _R (2)	Enc(2)
Enc _R (2)	Enc(0)
Enc _R (3)	Enc(1)
Enc _R (4)	Enc(3)

separate index for each searchable column, and using independent ORE keys

Encrypted Range Queries

Encrypted database:

ID	Name	Age	Diagnosis
0	Alice	31	2
1	Bob	47	3
2	Charlie	41	2
3	Inigo	45	4



columns (other than ID) are encrypted using a semantically-secure encryption scheme

clients hold (secret) keys needed to decrypt and query database

Name	ID
Enc _R (Alice)	Enc(0)

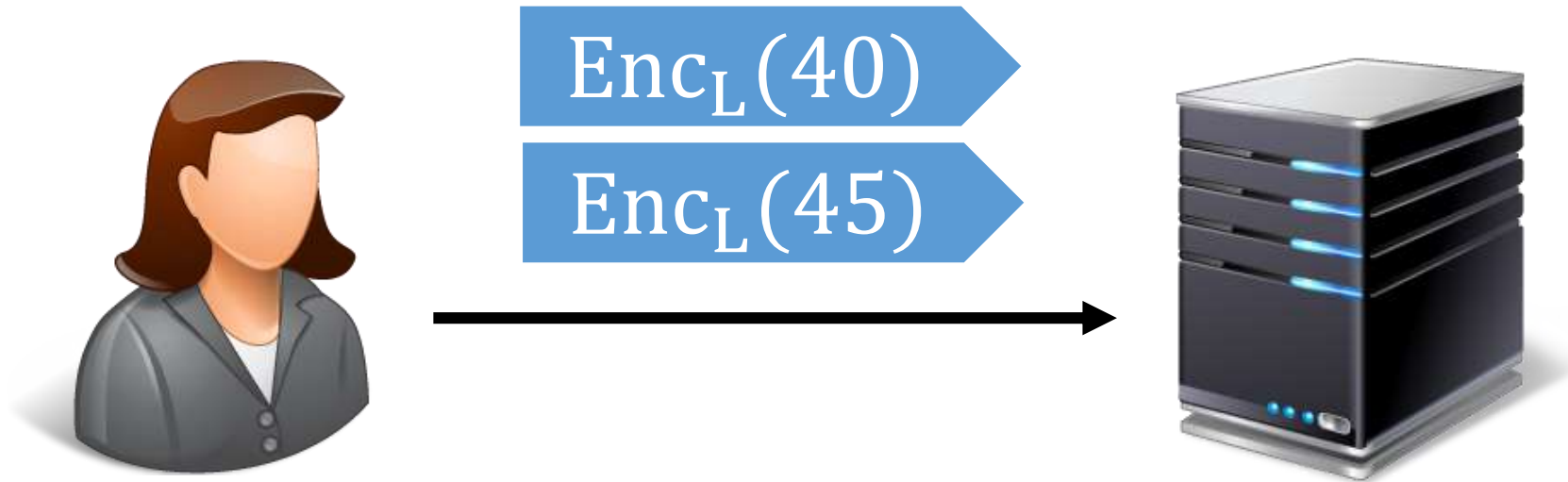
Age	ID
Enc _R (31)	Enc(0)

Diagnosis	ID
Enc _R (2)	Enc(2)
Enc _R (2)	Enc(0)
Enc _R (3)	Enc(1)
Enc _R (4)	Enc(3)

encrypted search indices

Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



$\text{Enc}_L(40)$

$\text{Enc}_L(45)$

Age	ID
$\text{Enc}_R(31)$	$\text{Enc}(0)$
$\text{Enc}_R(41)$	$\text{Enc}(2)$
$\text{Enc}_R(45)$	$\text{Enc}(3)$
$\text{Enc}_R(47)$	$\text{Enc}(1)$

Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



$\text{Enc}_L(40)$

$\text{Enc}_L(45)$

Age	ID
$\text{Enc}_R(31)$	$\text{Enc}(0)$
$\text{Enc}_R(41)$	$\text{Enc}(2)$
$\text{Enc}_R(45)$	$\text{Enc}(3)$
$\text{Enc}_R(47)$	$\text{Enc}(1)$

use binary search to determine endpoints (comparison via ORE)

Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



Age	ID
$\text{Enc}_R(31)$	$\text{Enc}(0)$
$\text{Enc}_R(41)$	$\text{Enc}(2)$
$\text{Enc}_R(45)$	$\text{Enc}(3)$
$\text{Enc}_R(47)$	$\text{Enc}(1)$

Diagram illustrating the query process. Two blue arrows point to the table: one labeled $\text{Enc}_L(45)$ pointing to the row with $\text{Enc}_R(45)$, and another labeled $\text{Enc}_L(40)$ pointing to the row with $\text{Enc}_R(41)$.

use binary search to determine endpoints (comparison via ORE)

Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



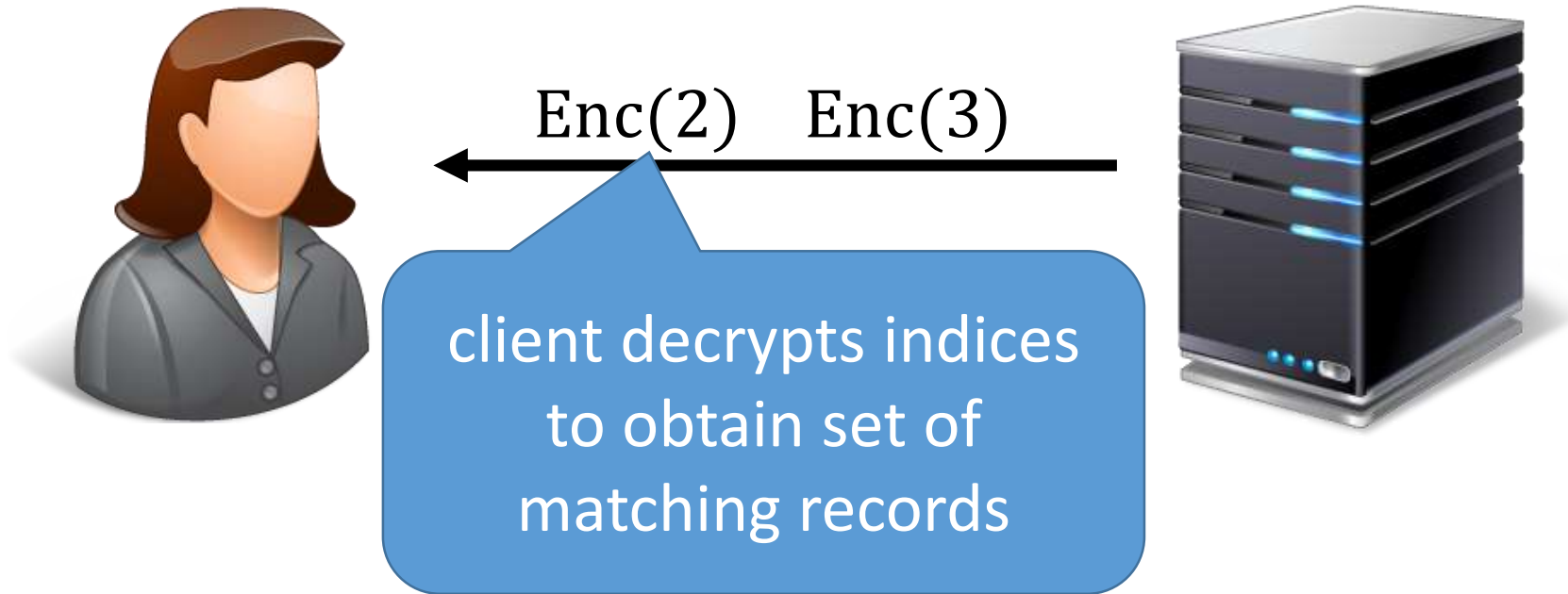
	Age	ID
$\text{Enc}_L(40)$	$\text{Enc}_R(31)$	$\text{Enc}(0)$
	$\text{Enc}_R(41)$	$\text{Enc}(2)$
$\text{Enc}_L(45)$	$\text{Enc}_R(45)$	$\text{Enc}(3)$
	$\text{Enc}_R(47)$	$\text{Enc}(1)$

return encrypted indices that match query

use binary search to determine endpoints (comparison via ORE)

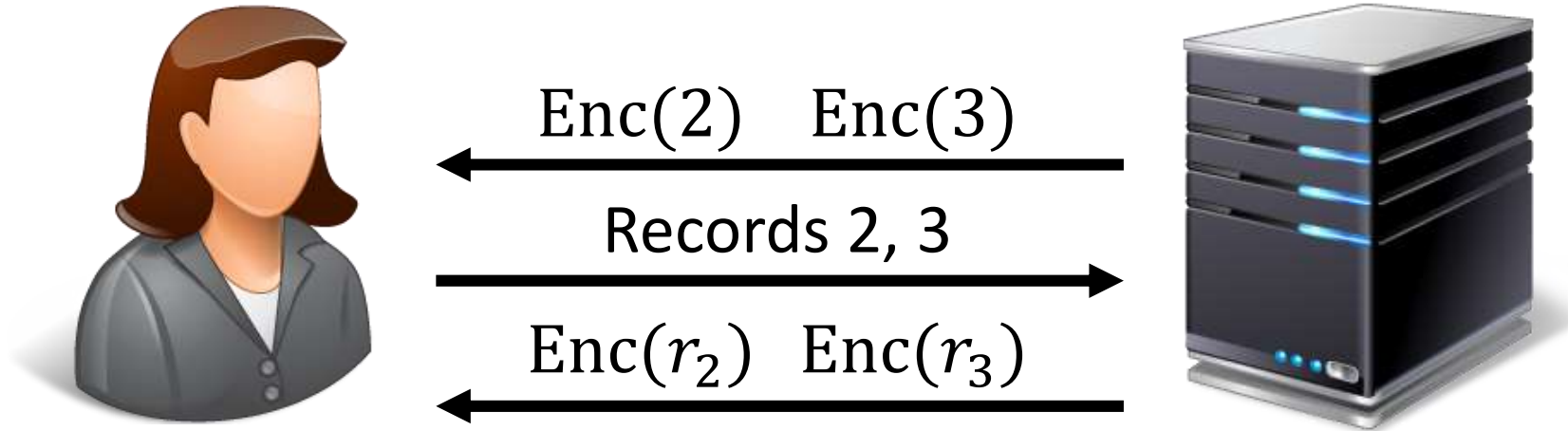
Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



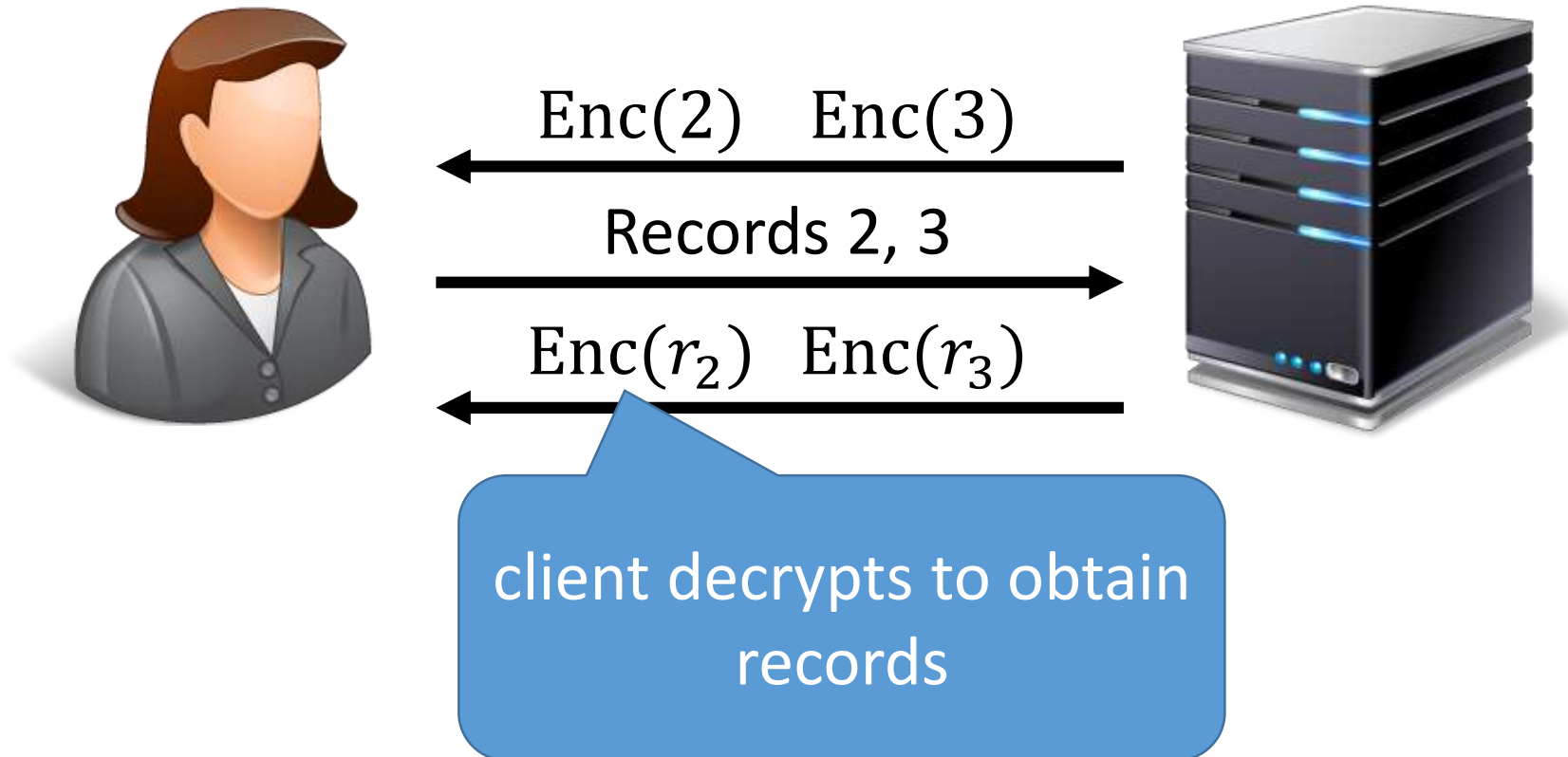
Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



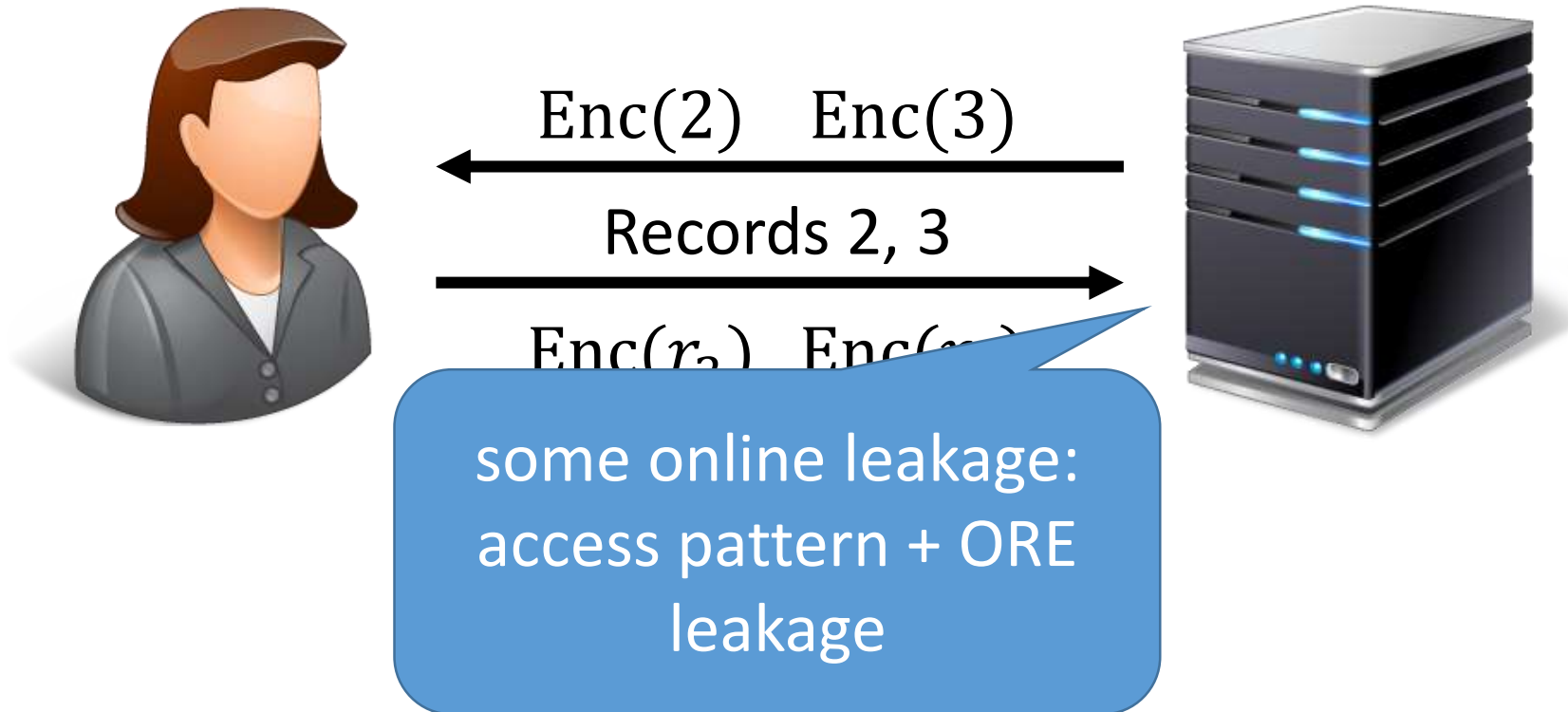
Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



Encrypted Range Queries

Query for all records where $40 \geq \text{age} \geq 45$:



Encrypted Range Queries

Encrypted database:

ID	Name	Age	Diagnosis
0	Alice	31	2
1	Bob	47	3
2	Charlie	41	2
3	Inigo	45	4



encrypted database is
semantically secure!
Perfect offline security

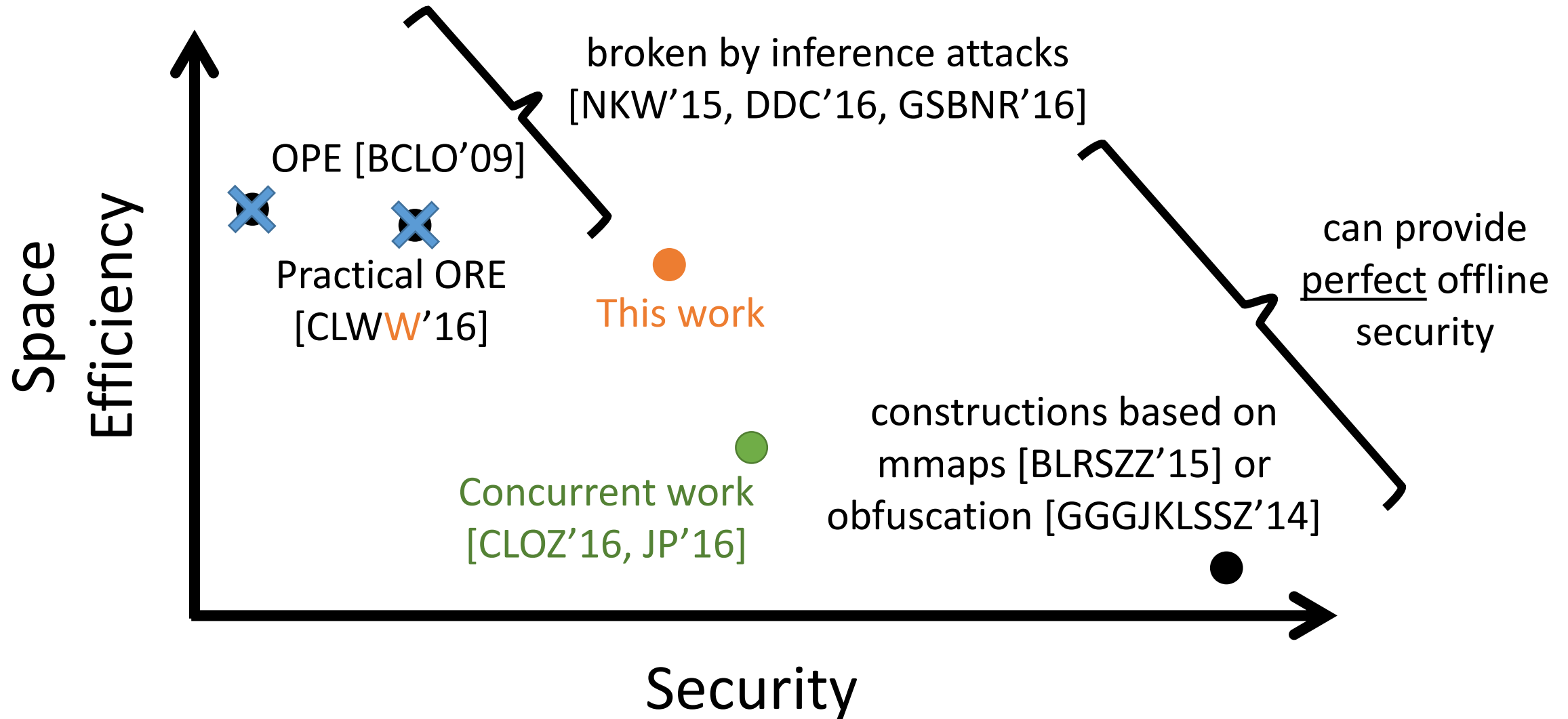
Name	ID
Enc _R (Alice)	Enc(0)

Age	ID
Enc _R (31)	Enc(0)

Diagnosis	ID
Enc _R (2)	Enc(2)
Enc _R (2)	Enc(0)
Enc _R (3)	Enc(1)
Enc _R (4)	Enc(3)

encrypted search indices

The Landscape of ORE



Not drawn to scale

Our New ORE Scheme

“small-domain” ORE with
best-possible security



domain extension
technique inspired by
CLW^W'16

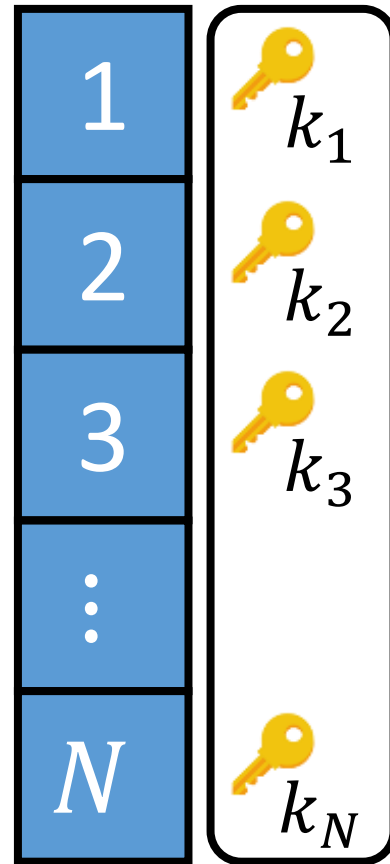


“large-domain” ORE
with some leakage

Small-Domain ORE with Best-Possible Security

Suppose plaintext space is small: $\{1, 2, \dots, N\}$

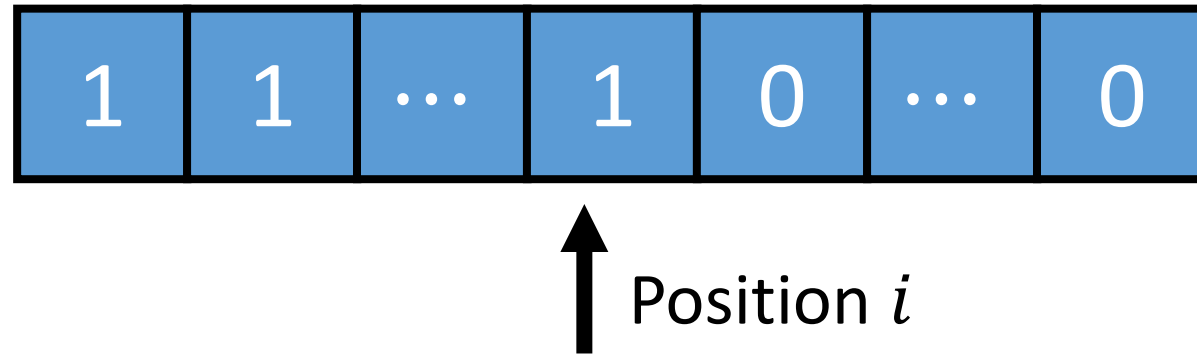
associate a key
with each value



(k_1, \dots, k_N) is the secret key
(can be derived from a PRF)

Small-Domain ORE with Best-Possible Security

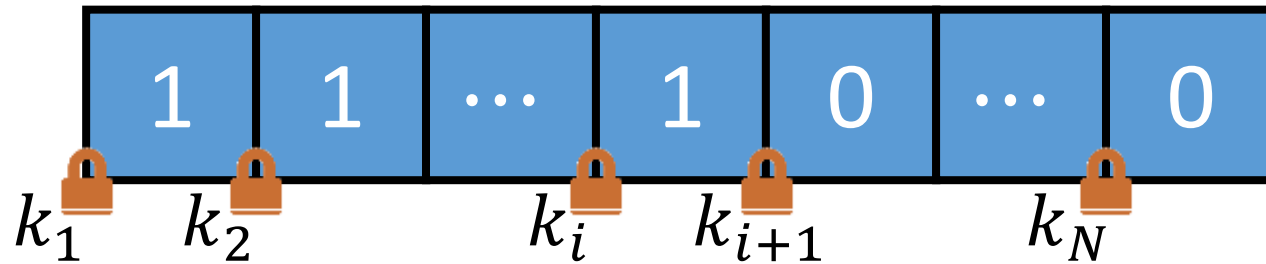
Encrypting a value i



Invariant: all positions $\leq i$ have value 1 while all positions $> i$ have value 0

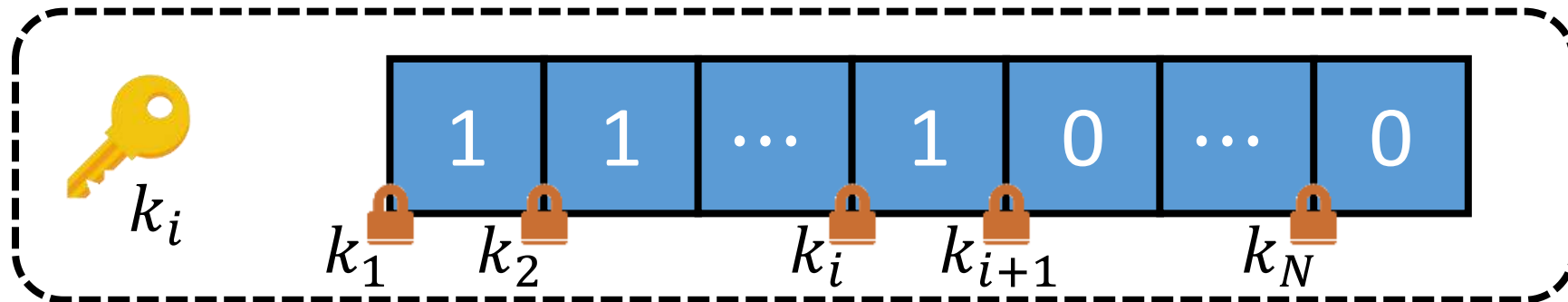
Small-Domain ORE with Best-Possible Security

Encrypting a value i



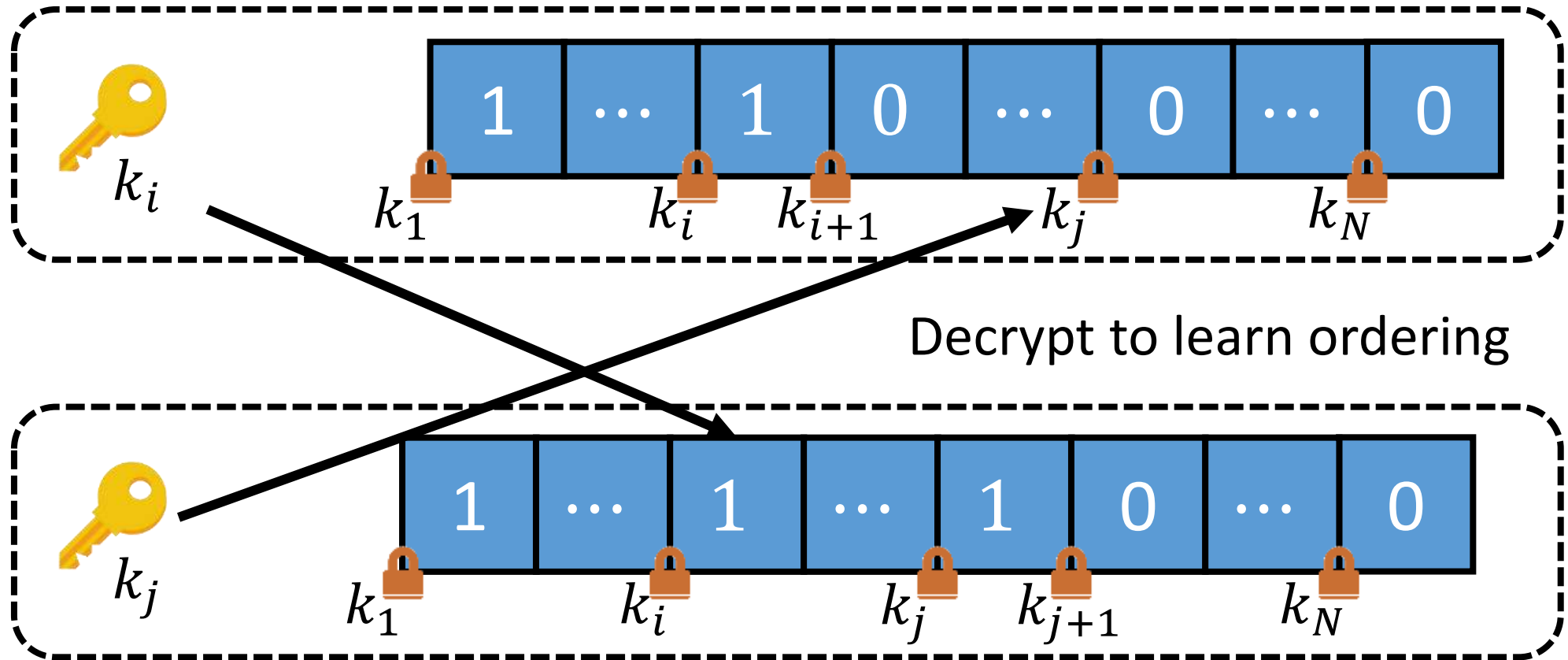
encrypt each slot
with key for that
slot

To allow comparisons, also give out key for slot i



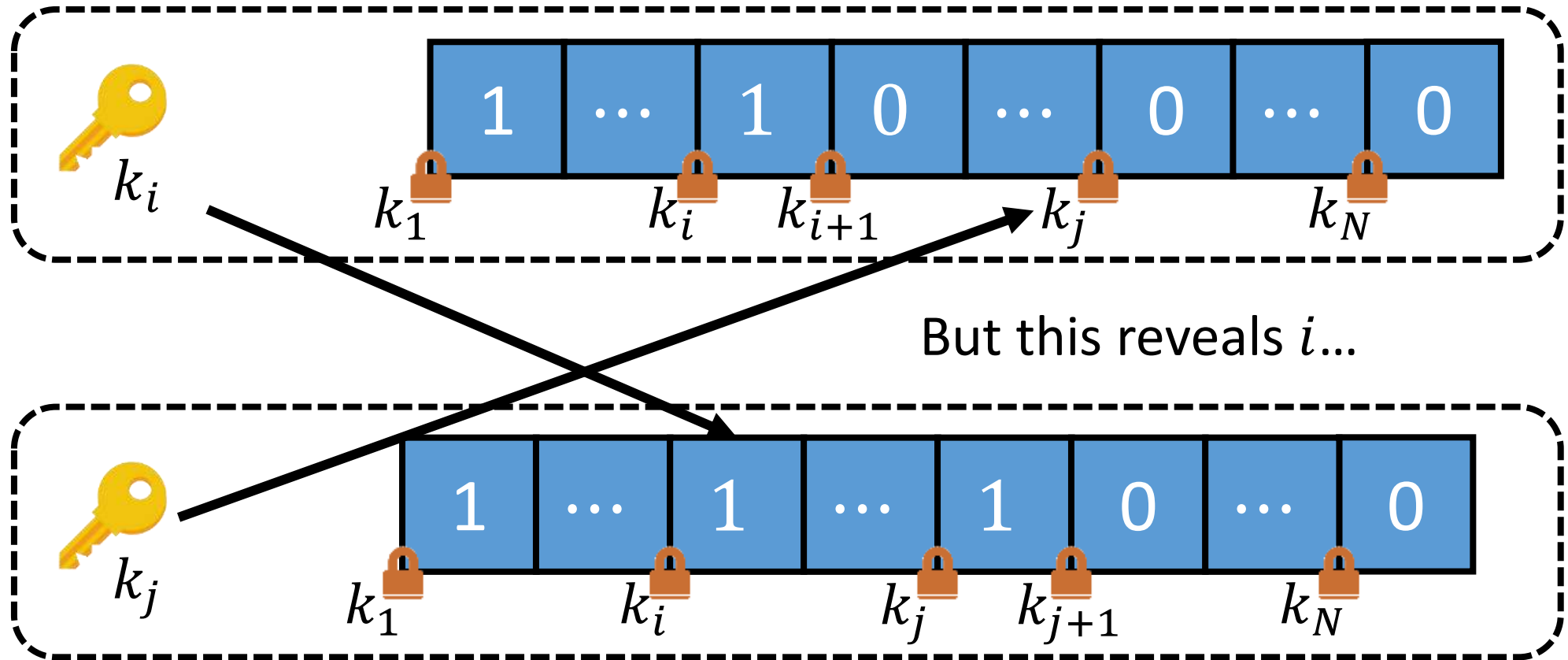
Small-Domain ORE with Best-Possible Security

Given two ciphertexts



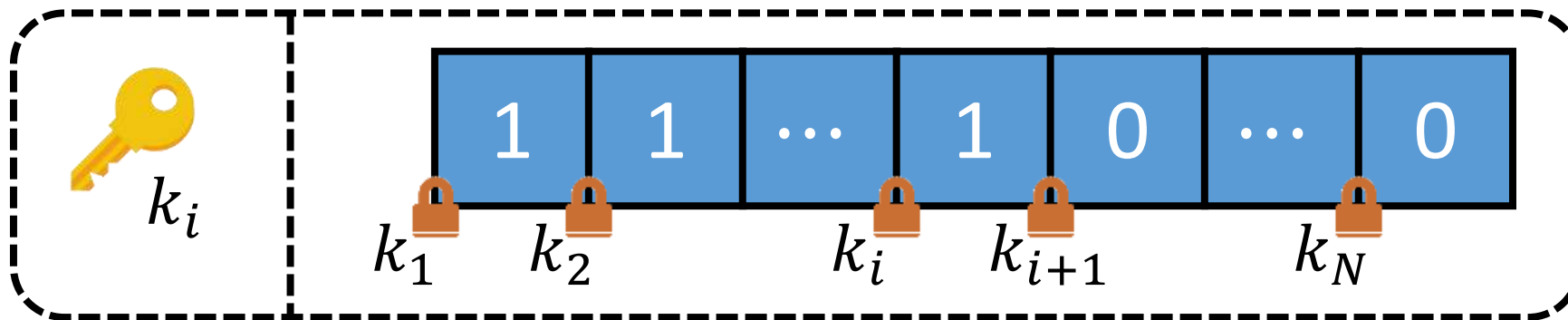
Small-Domain ORE with Best-Possible Security

Given two ciphertexts



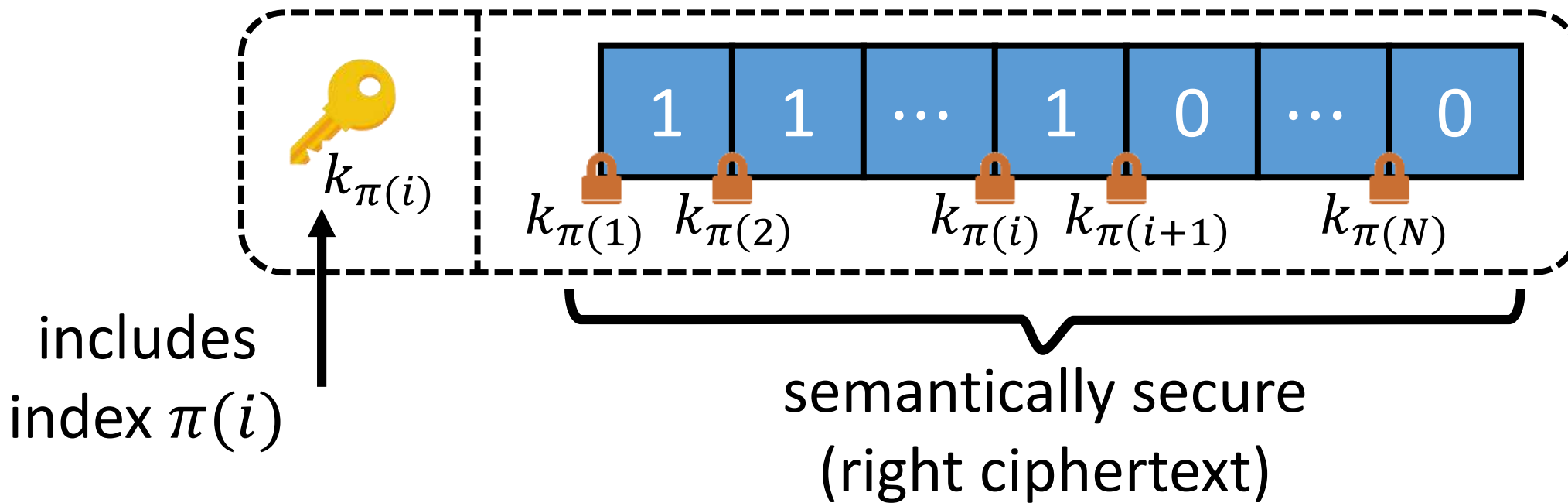
Small-Domain ORE with Best-Possible Security

Solution: apply random permutation π (part of the secret key) to the slots



Small-Domain ORE with Best-Possible Security

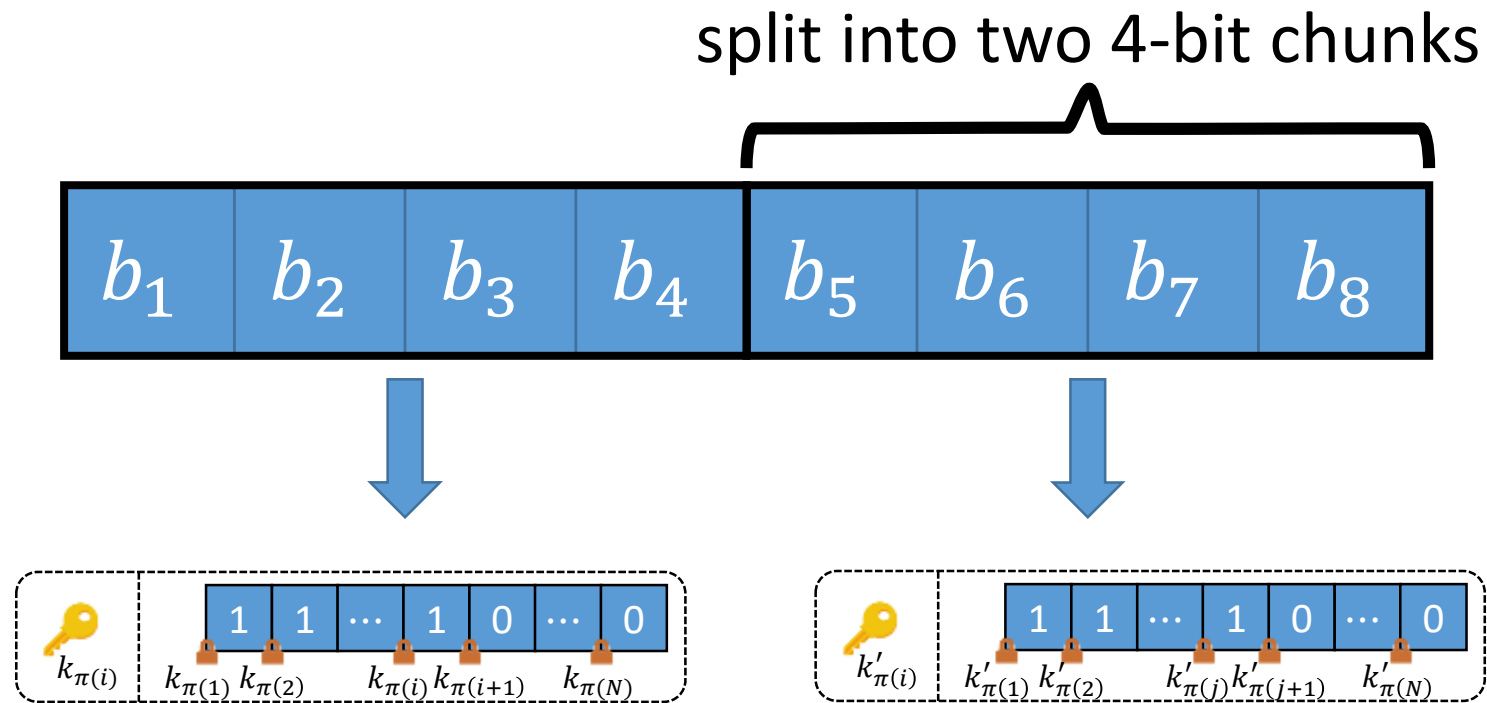
Solution: apply random permutation π (part of the secret key) to the slots



Achieves best-possible security, but ciphertexts are big

Domain Extension for ORE

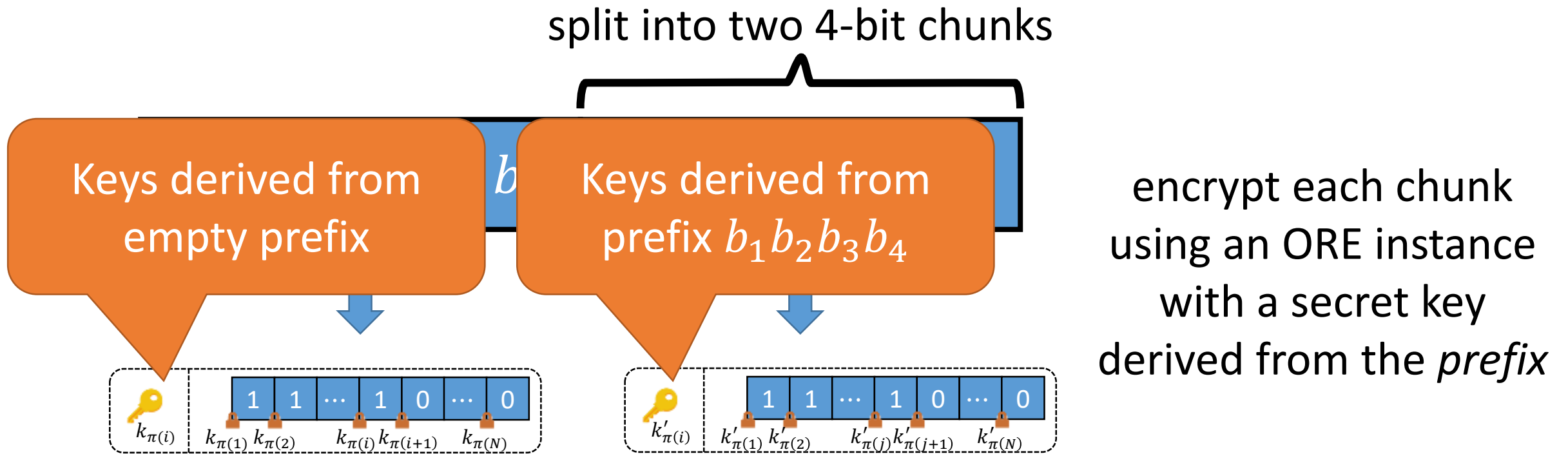
Key idea: decompose message into smaller blocks and apply small-domain ORE to each block



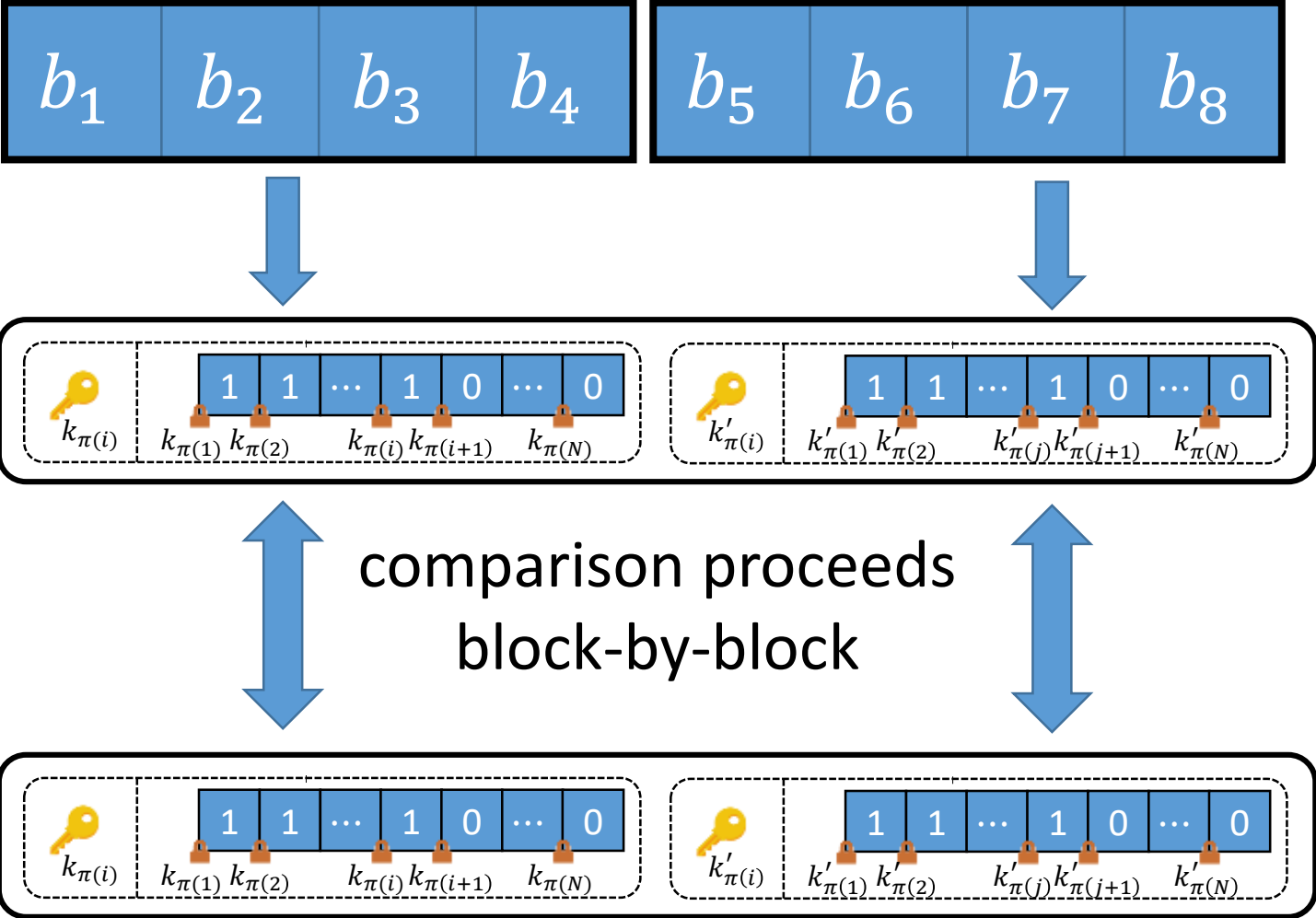
encrypt each chunk
using an ORE instance
with a secret key
derived from the *prefix*

Domain Extension for ORE

Key idea: decompose message into smaller blocks and apply small-domain ORE to each block



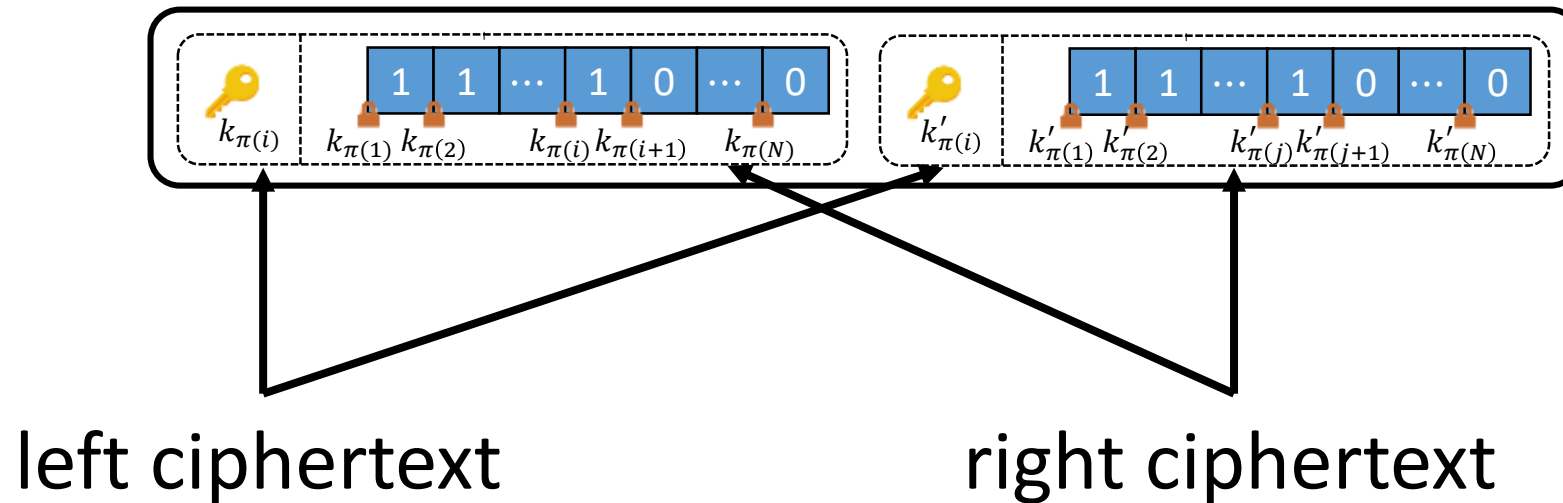
Domain Extension for ORE



Overall leakage: first **block** that differs

Domain Extension for ORE

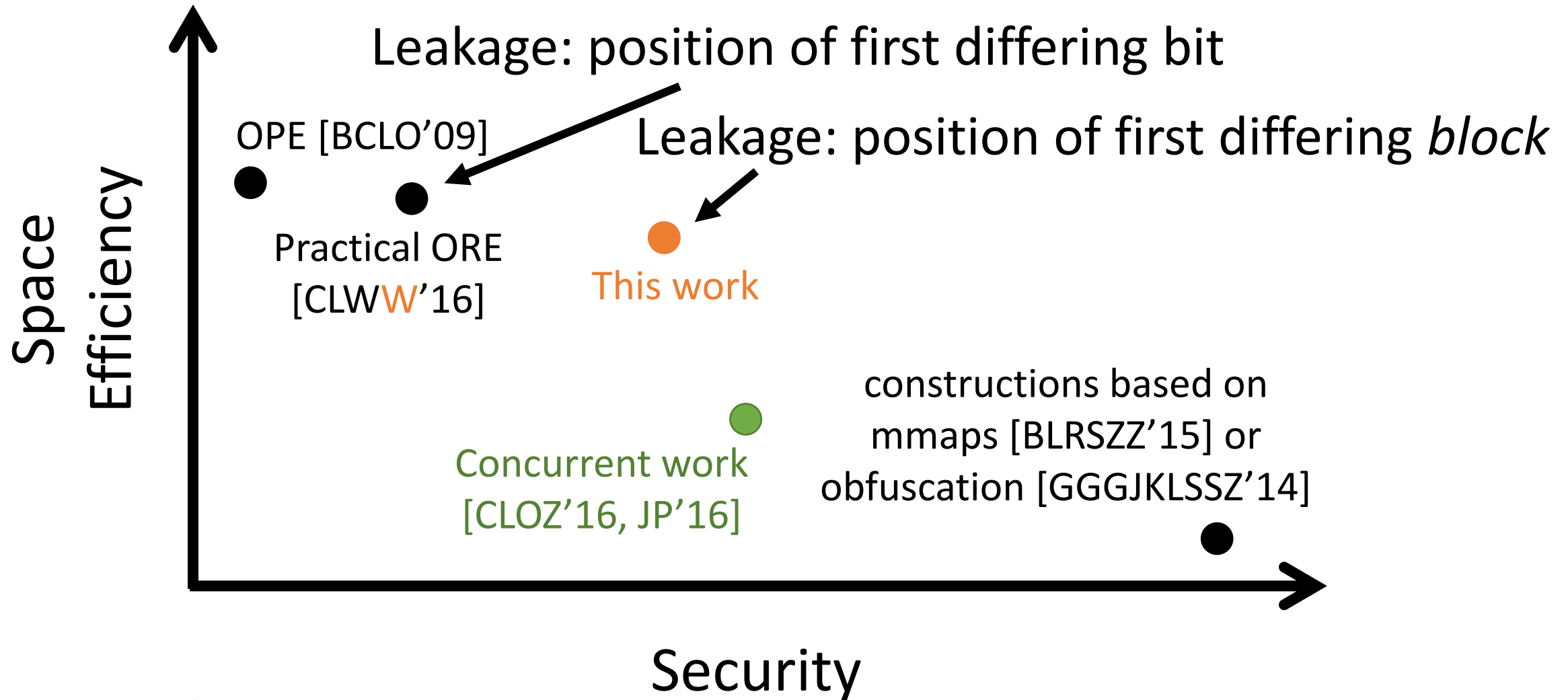
Same decomposition into left and right ciphertexts:



Right ciphertexts provide semantic security!

Note: optimizations are possible if we apply this technique in a non-black-box way to the small-domain ORE. See paper for details.

The Landscape of ORE



not drawn to scale

Performance Evaluation

Scheme	Encrypt (μs)	Compare (μs)	ct (bytes)
OPE [BCLO'09]	3601.82	0.36	8
Practical ORE [CLW ^W '16]	2.06	0.48	8
This work (4-bit blocks)	16.50	0.31	192
This work (8-bit blocks)	54.87	0.63	224
This work (12-bit blocks)	721.37	2.61	1612

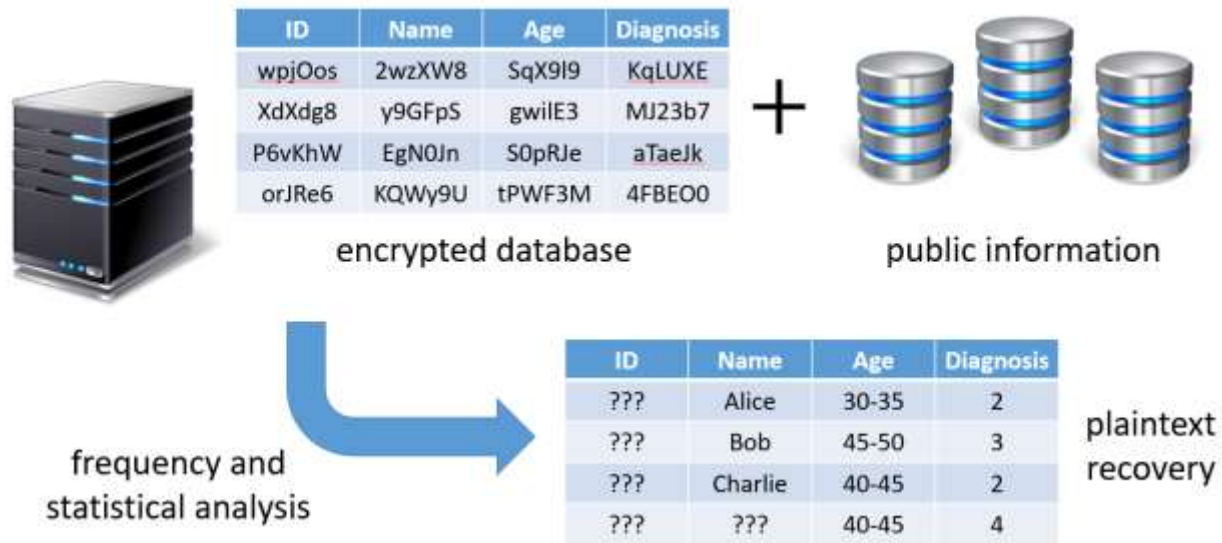
Benchmarks taken for C implementation of different schemes (with AES-NI). Measurements for encrypting 32-bit integers.

Performance Evaluation

Scheme	Encrypt (μs)	Compare (μs)	ct (bytes)
OPE [BCLO'09]	3601.82	0.36	8
Practical ORE [CLWW'16]	2.06	0.48	8
This work (4-bit blocks)	16.50	0.31	192
This work (8-bit blocks)	54.87	0.63	224
This work (12-bit blocks)	721.37	2.61	1612

Encrypting byte-size blocks is 65x faster than OPE, but ciphertexts are 30x longer. Security is substantially better.

Conclusions



- Inference attacks render most conventional PPE-based constructions insecure
- However, ORE is still a useful building block for encrypted databases

- Introduced new paradigm for constructing ORE that enables range queries in a way that is mostly legacy-compatible and provides offline semantic security
- New ORE construction that is concretely efficient with strong security
- In paper: new impossibility results for security achievable using OPE

Open Problems

- What kind of inference attacks are possible in the online setting?
 - Indices encrypted separately, so multi-column correlations harder to infer
 - More limited leakage profile (between left and right ciphertexts)
- Can we construct small-domain OREs (with best-possible security) and *sublinear* (in the size of the domain) ciphertext size from PRFs?
- Can we construct left/right ORE (from PRFs) where both left and right ciphertexts are *semantically secure*?



Questions?

Paper: <https://eprint.iacr.org/2016/612>
Website: <https://crypto.stanford.edu/ore/>
Code: <https://github.com/kevinlewi/fastore>