32ND USENIX SECURITY SYMPOSIUM

香 港 大 學
THE UNIVERSITY OF HONG KONG

Efficient Unbalanced Private Set Intersection Cardinality and User-friendly Privacy-preserving Contact Tracing

Mingli Wu, Tsz Hon Yuen

The University of Hong Kong

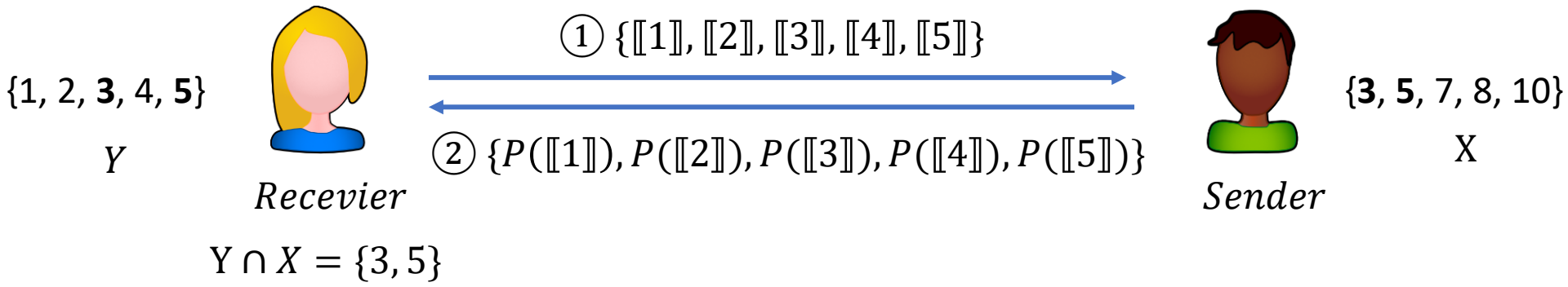
8/9/2023

Outline

- What is uPSI(-CA)?
- Related works
- Existing issues (deception attack and long item issue)
- Our solutions (VBF and PoL)
- Results
- Conclusion

What is uPSI(-CA)?

- Private set intersection (PSI) allows a $recevier$ (with a set Y) and a $sender$ (with a set X) to identify $\mathrm{Y} \cap X$ without revealing any information beyond it.



- Private set intersection Cardinality (PSI-CA): the $recevier$ secretly knows $|\mathrm{Y} \cap X|$.

- *Unbalanced* PSI-CA (uPSI-CA): $|Y| \ll |X|$. → Lowering the **communication costs**.

Related works (uPSI)

- FHE-based ones: CLR'17[1], **CMGDILR'21[2]**

1. CLR'17: item bit length $\delta \leq 32$
2. CMGDILR'21: slicing to support arbitrary δ

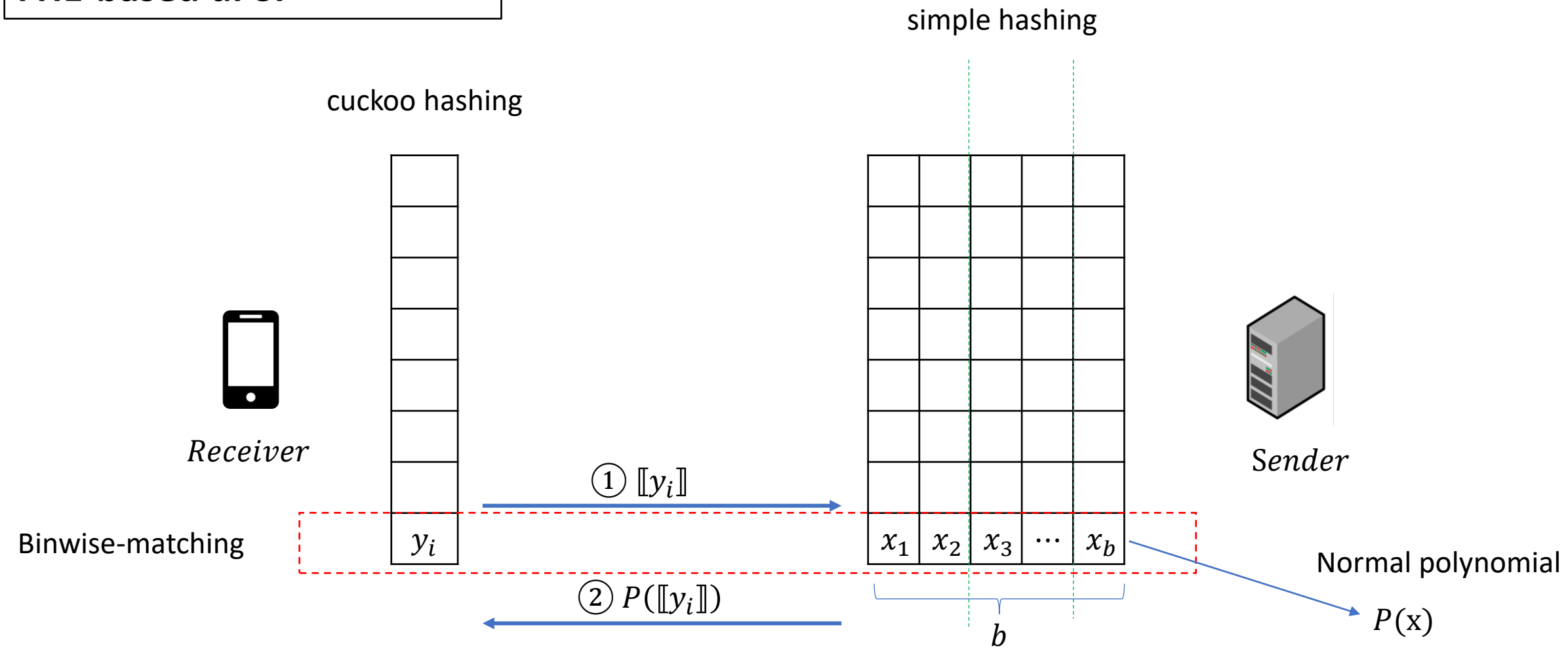
Strength — Communication cost $O(n_y \log(n_x))$

Weaknesses:
1. Prone to **deception attack**
2. Performance not good enough

Long item issue

1. Hao Chen, et.al. Fast private set intersection from homomorphic encryption. In CCS 2017, pages 1243–1255. ACM, 2017.
2. Kelong Cong, et.al. Labeled PSI from homomorphic encryption with reduced computation and communication. In CCS 2021, pages 1135–1150. ACM, 2021.

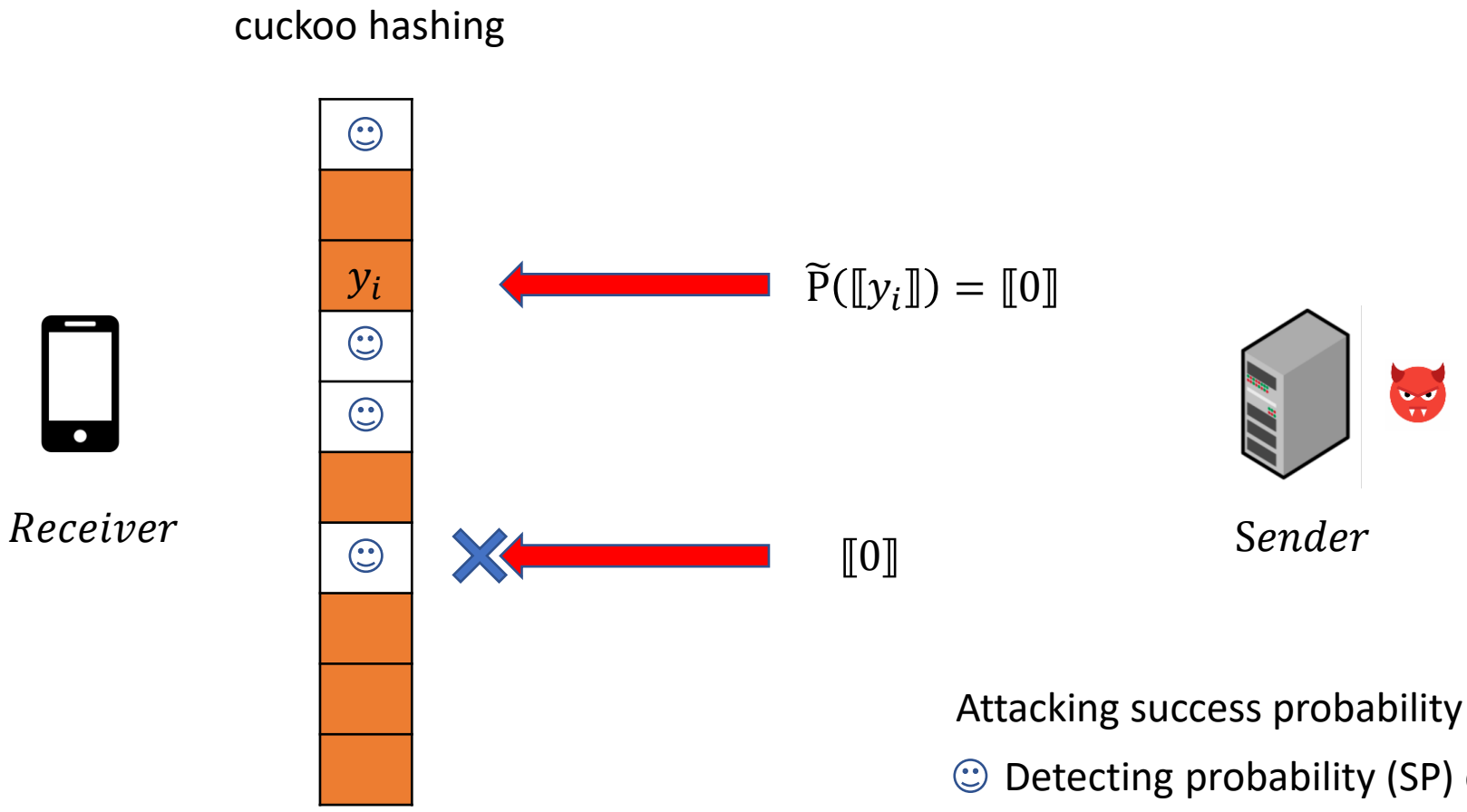
FHE-based uPSI



$$P(x) = \prod_{i=1}^{b}(y - x_i) = a_0 + a_1 y + a_2 y^2 + \cdots a_b y^b$$

Deception attack

In *private contact discovery*, to attract users, a service provider simply tells the user that he/she has many friends who are using the same application.



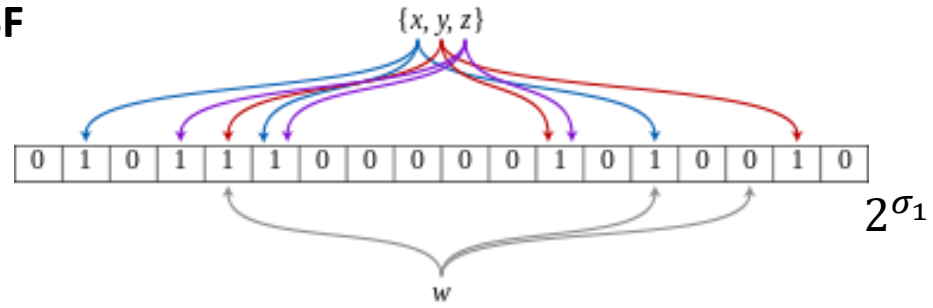
Attacking success probability (SP) as high as 0.67

☺ Detecting probability (SP) only 0.34

How we resolve the long item issue?

Scheme 1: Virtual Bloom Filter (VBF)

BF



$$x := \{h_1(x), h_2(x), \cdots, h_{k_v}(x)\}$$

Turn an arbitrary long item (e.g., 128 bit) into k_v short VBF sub-items with bit length σ_1.

k_v	2	3	4	5	6	7
σ_1	46	39	37	35	34	33

Example:

$$2 \times \sigma_1 = 2 \times 46 = 92\ bits$$

$n_y = 5535$ ↓ Permutation-based hashing

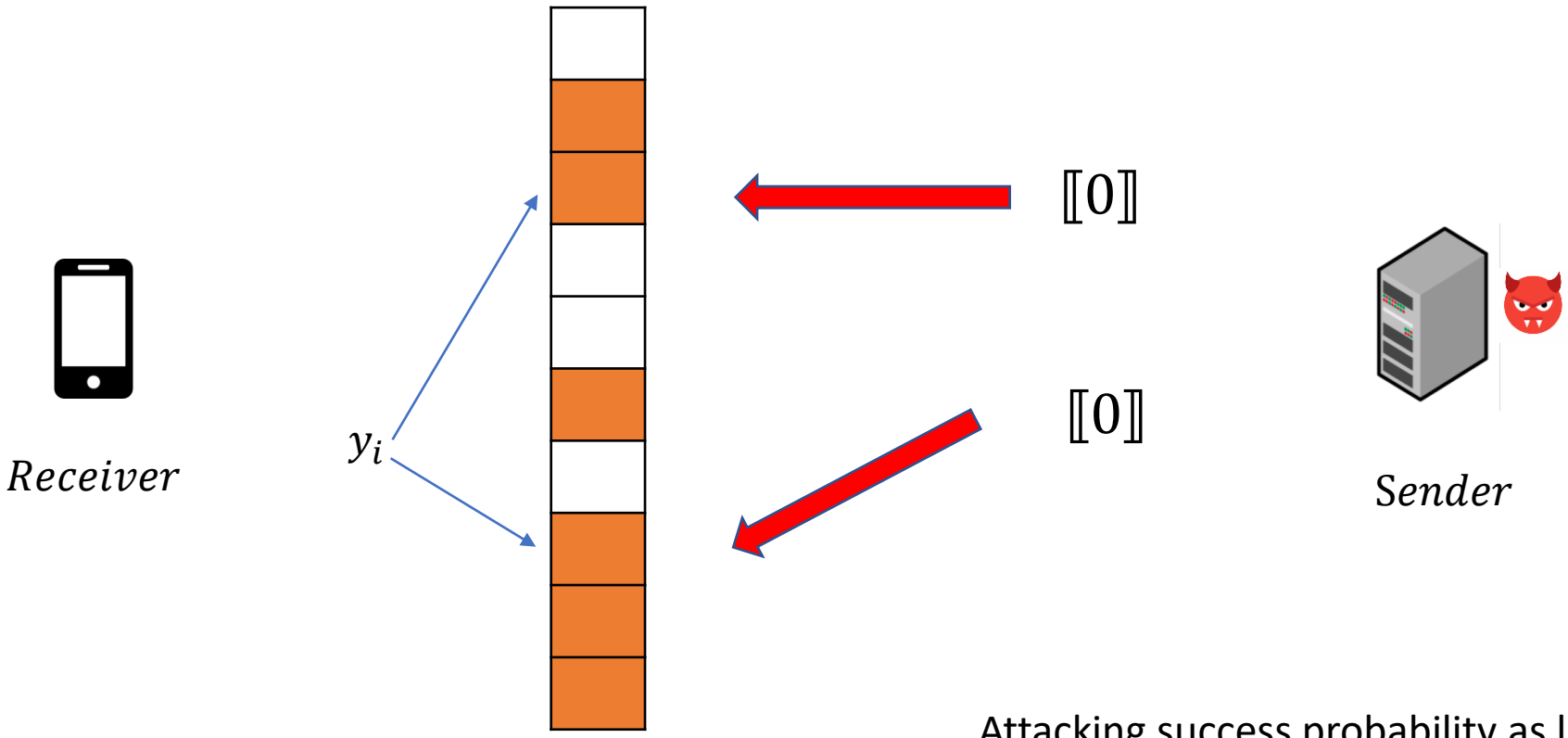
$$2 \times \sigma'_1 = 2 \times (46 - 13) = 66\ bits$$

↓

$$2 \times 33$$

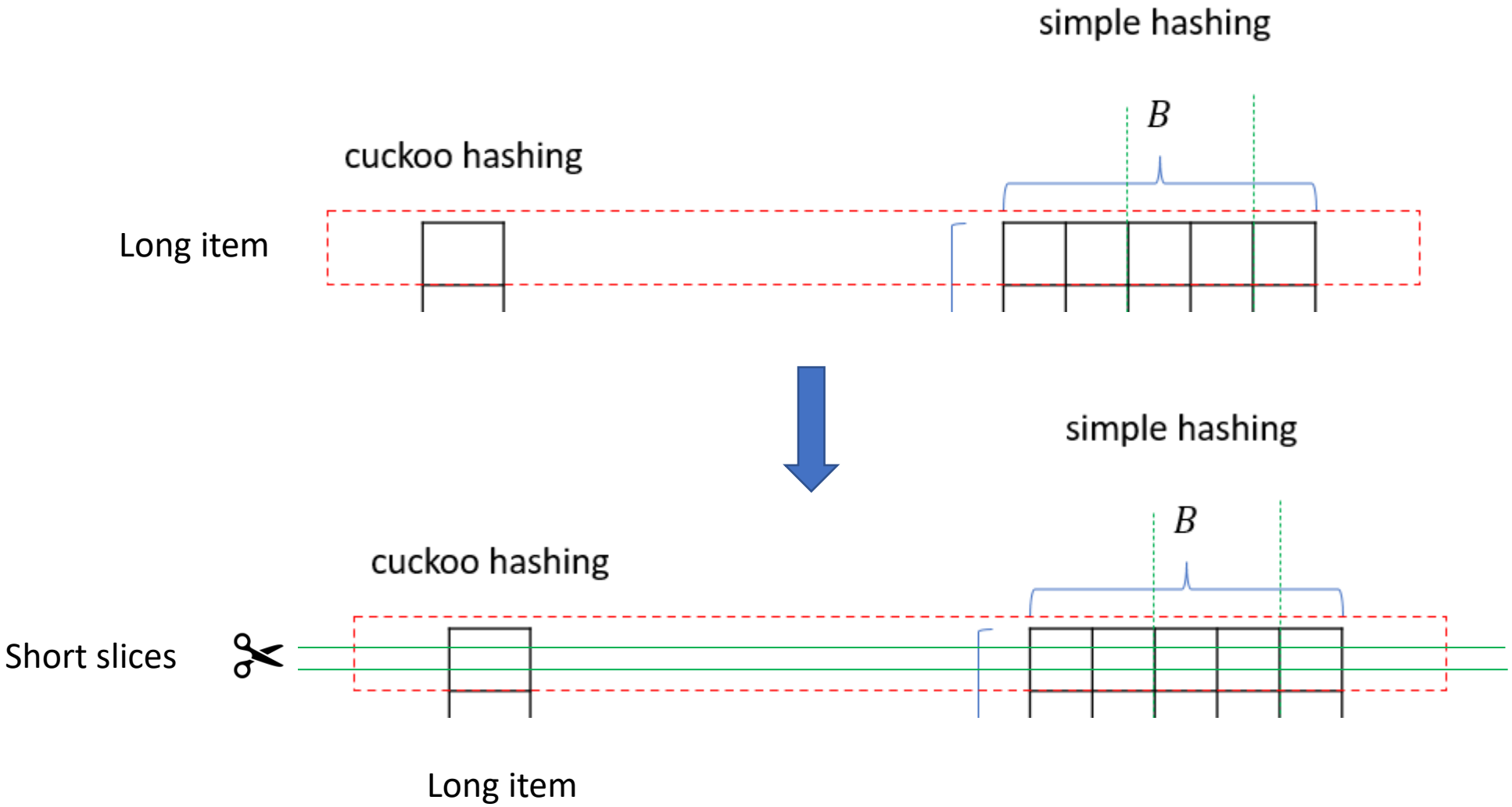
Deception attack

cuckoo hashing

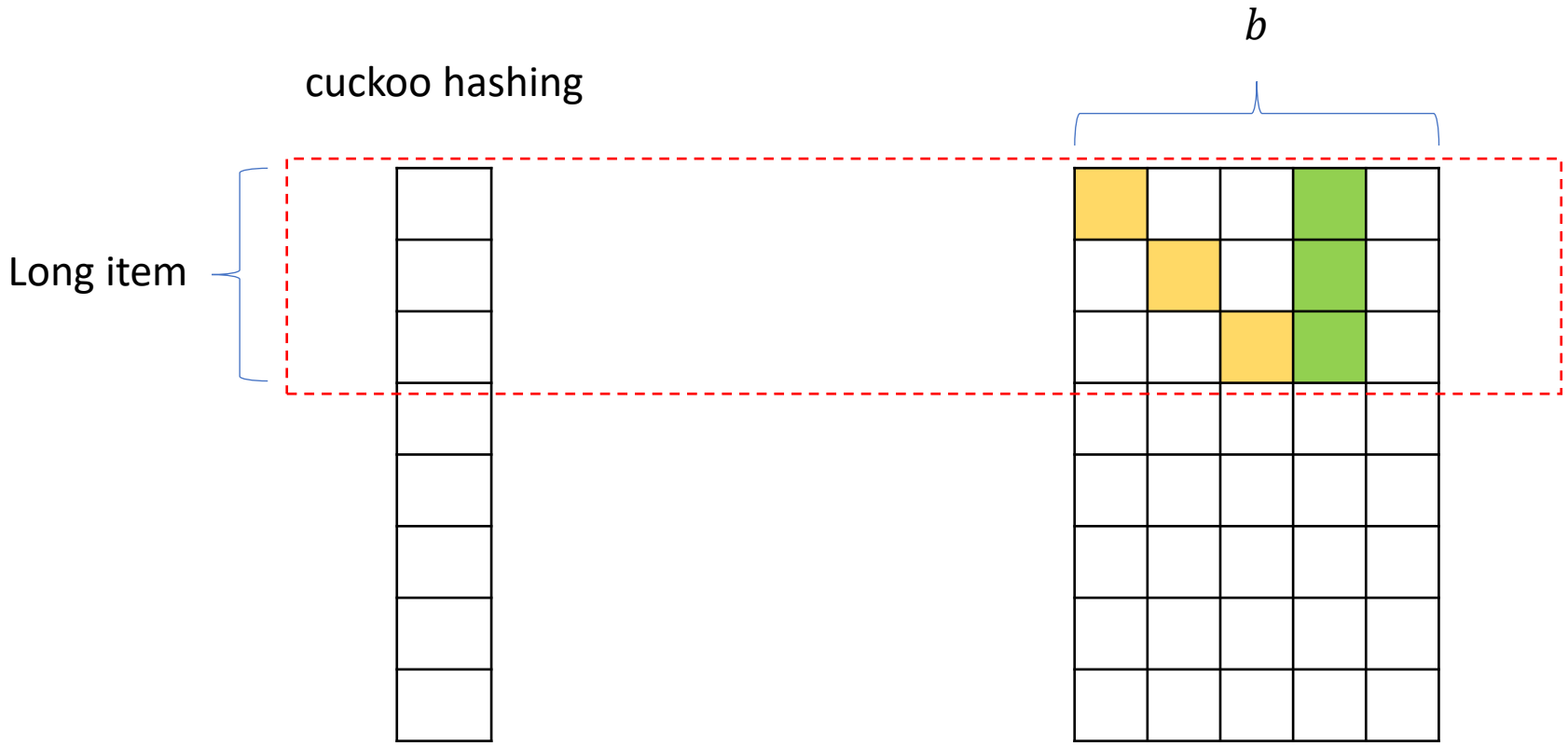


Attacking success probability as low as $0.67 \times \frac{1}{m}$

CMGDILR'21: slicing



Drawback: false positives

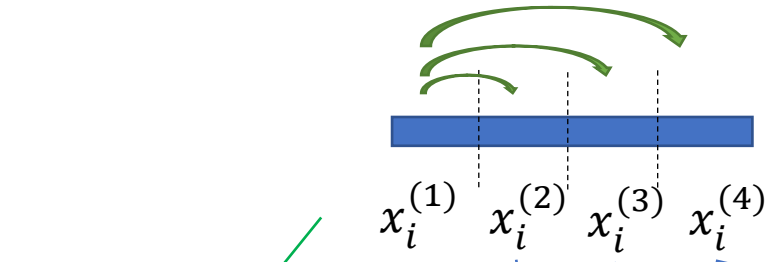


The total item bits are in (80, 110).

Scheme2: Polynomial links (PoL)

$$y = y^{(1)} \left| y^{(2)} \right| y^{(3)} | \cdots | y^{(k_s)}$$

Interpolation polynomials



$$x_i = x_i^{(1)} \left| x_i^{(2)} \right| x_i^{(3)} | \cdots | x_i^{(k_s)}$$

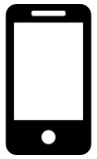
$x_i^{(1)}$ $x_i^{(2)}$ $x_i^{(3)}$ $x_i^{(4)}$

$P_1([\![y^{(1)}]\!]) = 0$

$P_2([\![y^{(1)}]\!]) = y^{(2)}$

$P_3([\![y^{(1)}]\!]) = y^{(3)}$

$P_4([\![y^{(1)}]\!]) = y^{(4)}$



Receiver

1. Only need to send $[\![y^{(1)}]\!]$, saving $(1 - k_s)/k_s$ communication cost.
2. Only encrypt $[\![y^{(1)}]\!]$, saving about $(1 - k_s)/k_s$ computation costs.

No false positives, so the total number of bits can be at most 80.
Saving more!

$3 \leq k_s \leq 11$

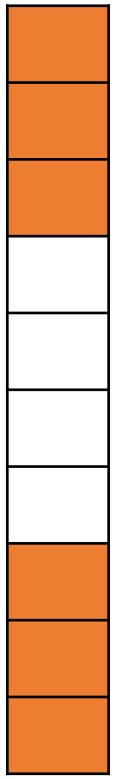
Sender

High computation cost.

Deception attack

cuckoo hashing

$Receiver$



$[\![0]\!]$

$Sender$

? Randomly guess $\frac{1}{2^{\sigma_2}}$

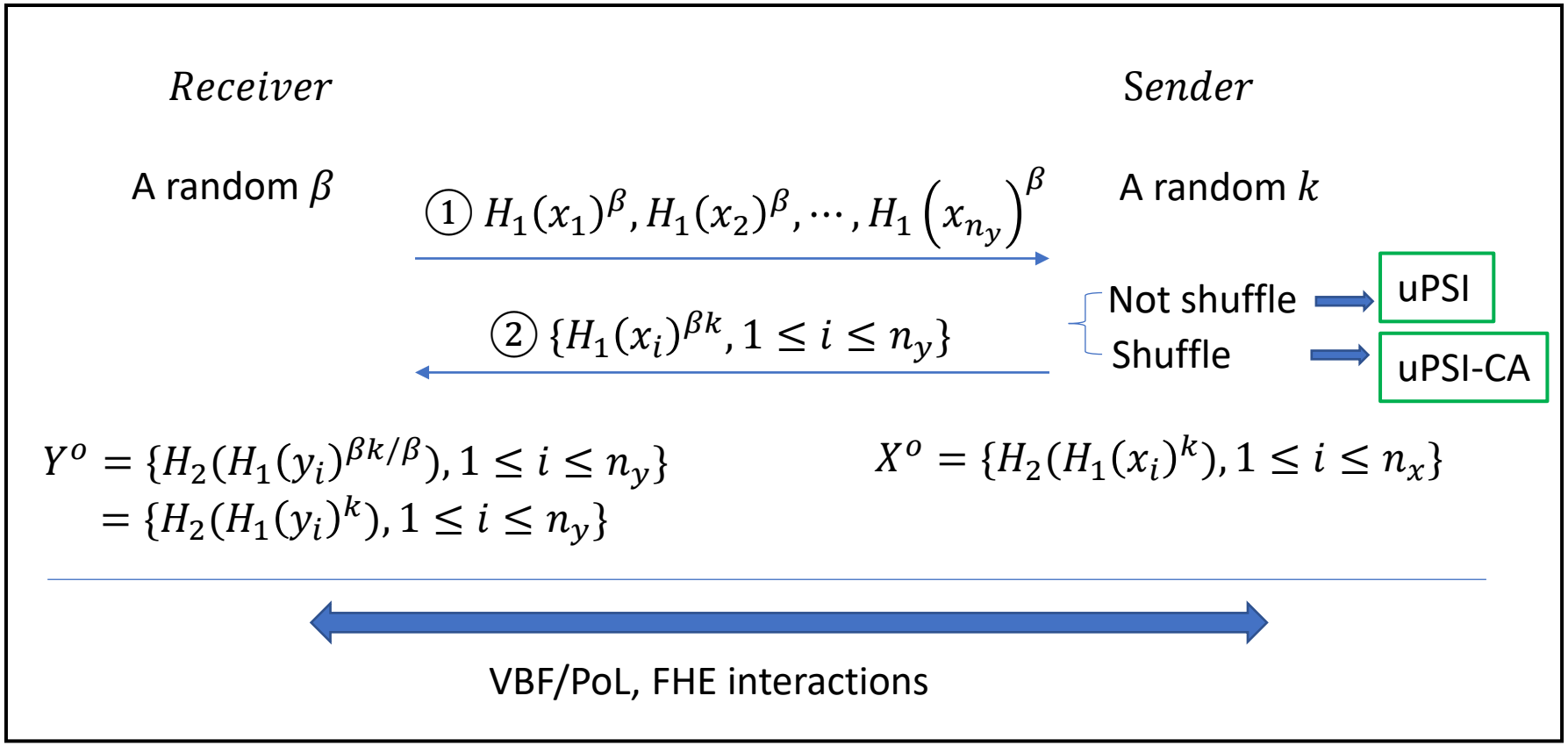
? Randomly guess $\frac{1}{2^{\sigma_2}}$

Attacking success probability as low as $0.67 \times \frac{1}{2^{(k_S-1)\sigma_2}}$

uPSI ⟶ uPSI-CA

Shuffled OPRF preprocessing

CANS'12 [1]



[1] Emiliano De Cristofaro et.al. Fast and private computation of cardinality of set intersection and union. In CANS 2012, pages 218–231. Springer, 2012.

Results

Deception attack

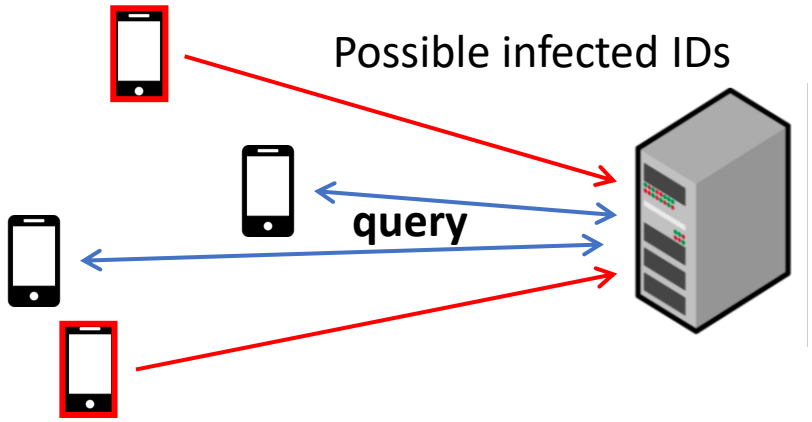
SP/DP	Prior FHE ones	VBF
$n_y = 1024$	$0.5/0.5$	$2.4 \times 10^{-4}/0.75$
$n_y = 2048$	$0.5/0.5$	$1.2 \times 10^{-4}/0.75$
$n_y = 5535$	$0.676/0.324$	$2.1 \times 10^{-5}/0.543$
$n_y = 11041$	$0.674/0.326$	$1.0 \times 10^{-5}/0.546$

Our uPSI (PoL) vs CMGDILR'21[1] :

Communication costs	42.04% ~ 58.85% cheaper
Online time (10Gbps)	1.81% ~ 63.00% faster
Online time (1Mbps)	7.65% ~ 247.69% faster
Sender offline	38.35% ~ 85.70% **slower**

1. Kelong Cong, et.al. Labeled PSI from homomorphic encryption with reduced computation and communication. In CCS 2021, pages 1135–1150. ACM, 2021.

uPSI-CA application (contact tracing)



1. Two-party designs

Protocols	Linkage attack	Query time(s)	User computation(s)	User communication(MB)
Google&Apple [2]	Yes	**6.640**	24.322	7.00
DP-3T [53]	Yes	387.736	0.384	448.00
Epione [52]	No	268.5/140.14	2.088/2.217	226.13/65.65
Ours (PoL)	No	60.524	**0.366**	**5.98**

2. Delegated design (third parties)

Ours (PoL)	No	**60.524**	0.366	5.98
Catalic [20]	No	97.79	**0.002**	**0.094**

Expensive communication costs for the backend server (e.g., >1GB) per query.

Conclusion

- 1. VBF and PoL to resolve the long item issue.
- 2. Handle the deception attack.
- 3. Secure and user-friendly contact tracing.

Q&A

mingliwu@hku.hk