Unbalanced PSI

Applications, Constructions, and Combinations with PIR

Dr Christian Weinert – Presented at WPEC 2024, September 24

NIST Workshop on Privacy-Enhancing Cryptography

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Royal Holloway, University of London (RHUL)

ROYAL HOLLOWAY UNIVERSITY OF LONDON

The Cryptography Group

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- RHUL has a long history of cryptographic research, dating back to at least **1987**
- Currently:
  - 11 members of staff
  - ~15 PhD students

- Find out more: https://cryptography.isg.rhul.ac.uk/people/

Definition

ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Private Set Intersection (PSI)

Alice

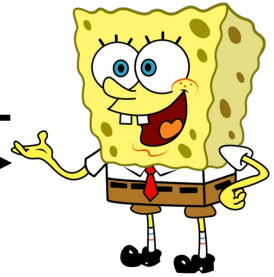
Set X

Z

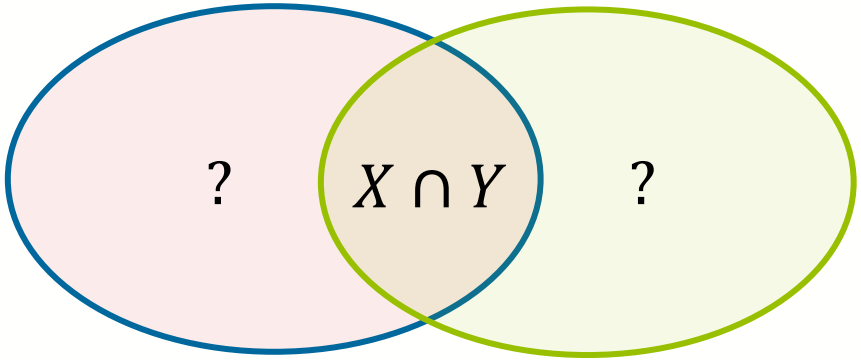
$Z = f(X, Y) = X \cap Y$

Set Y

Z



Bob



Private Set Intersection (PSI)

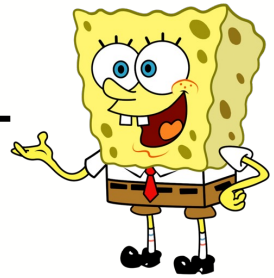
PSI Receiver

Set X

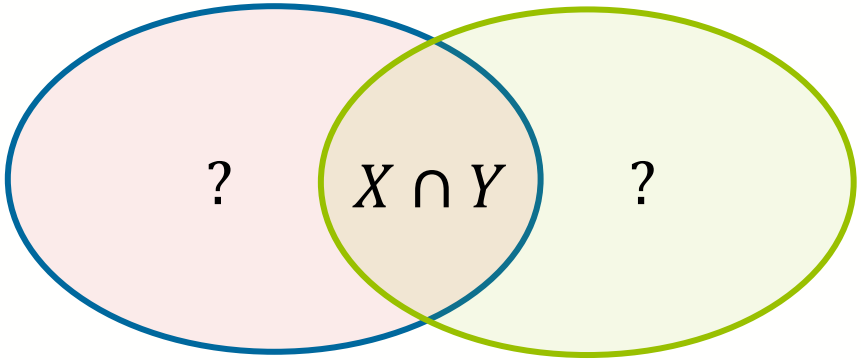
Z

$Z = f(X, Y) = X \cap Y$

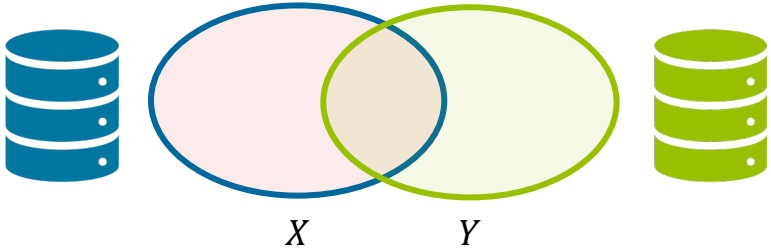
Set Y



PSI Sender

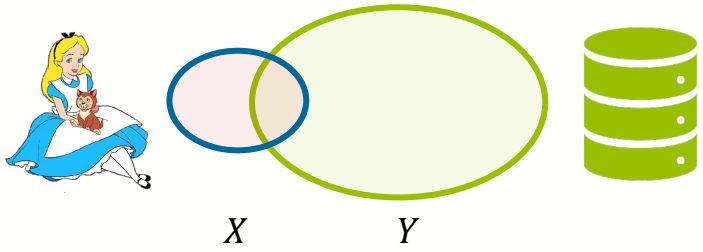


Balanced vs Unbalanced PSI (uPSI)



Balanced PSI

$|X| \approx |Y|$



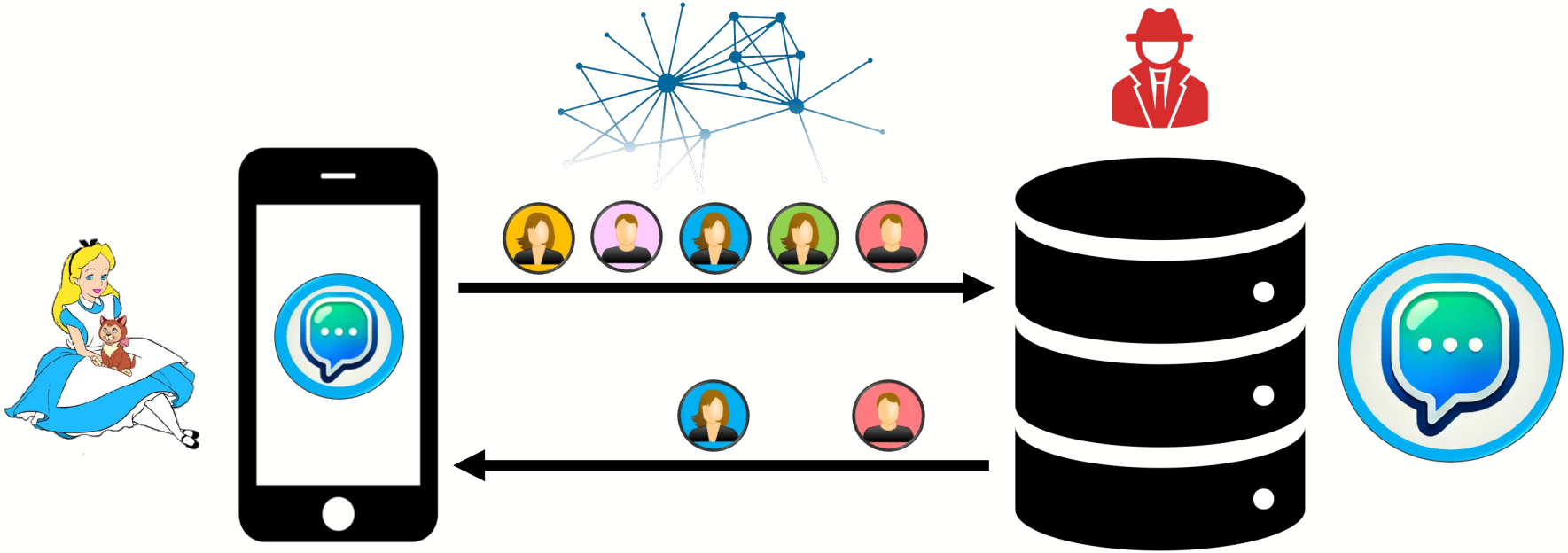
Unbalanced PSI (uPSI)

$|X| \ll |Y|$

Applications and Deployments

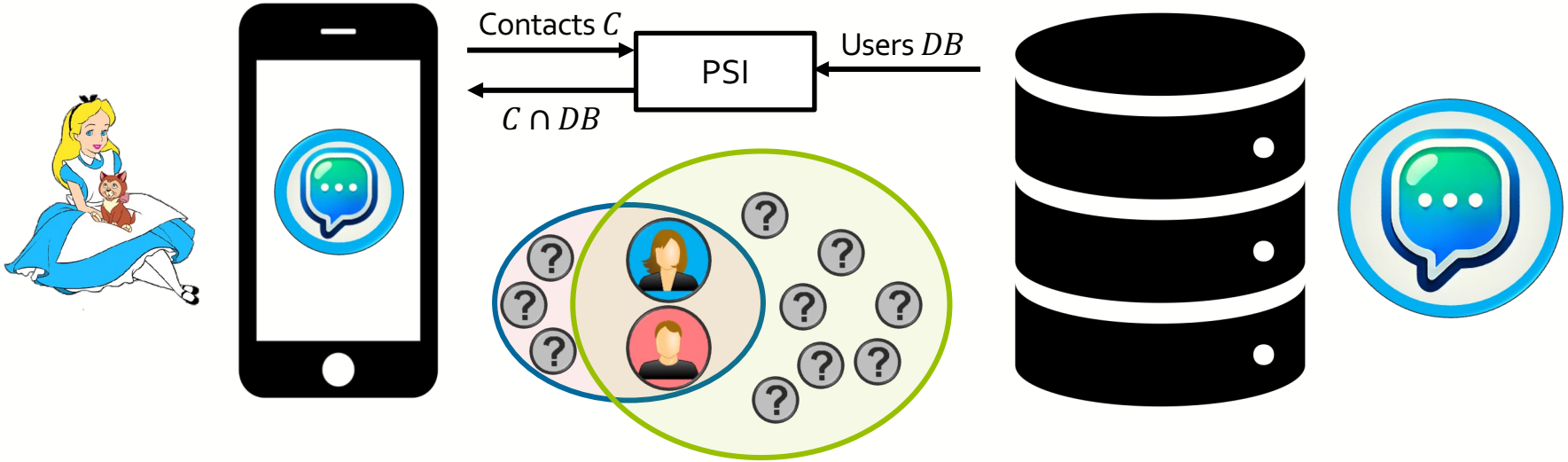
ROYAL HOLLOWAY UNIVERSITY OF LONDON

Mobile Contact Discovery



https://contact-discovery.github.io/

Mobile Contact Discovery



https://contact-discovery.github.io/

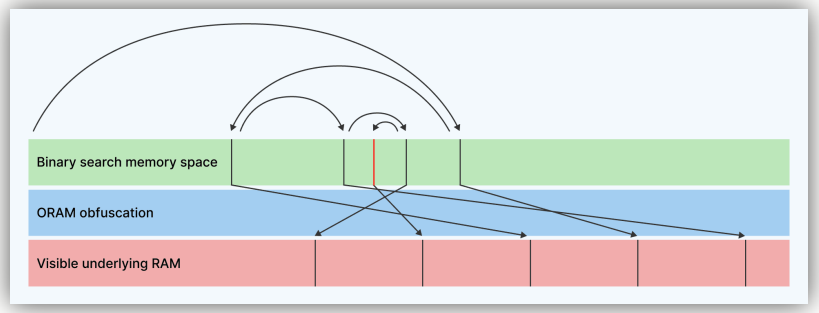
Intel SGX-based Contact Discovery Service in Signal

Performance requirements: 1B+ users, 10k contacts, <2s latency, <10MiB communication

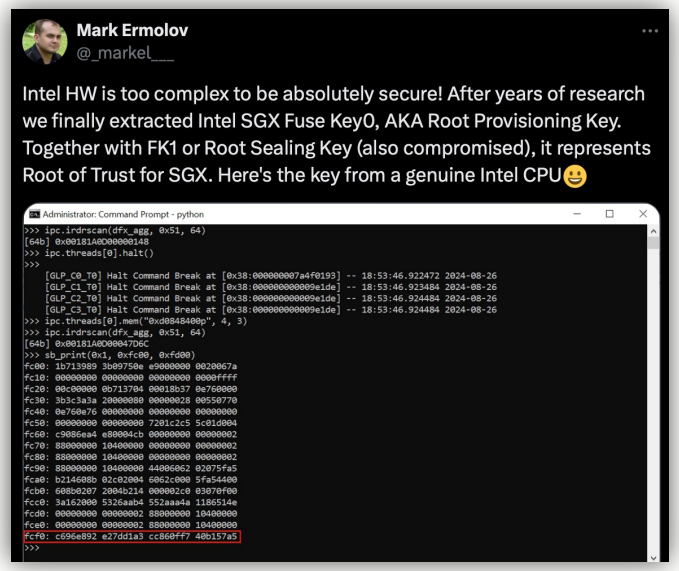
CDSI SGX Enclave

The Contact Discovery Service (CDSI) SGX Enclave provides endpoints to expose a table of Signal user records indexed by phone number that allows users to securely discover which of their contacts are also Signal users.

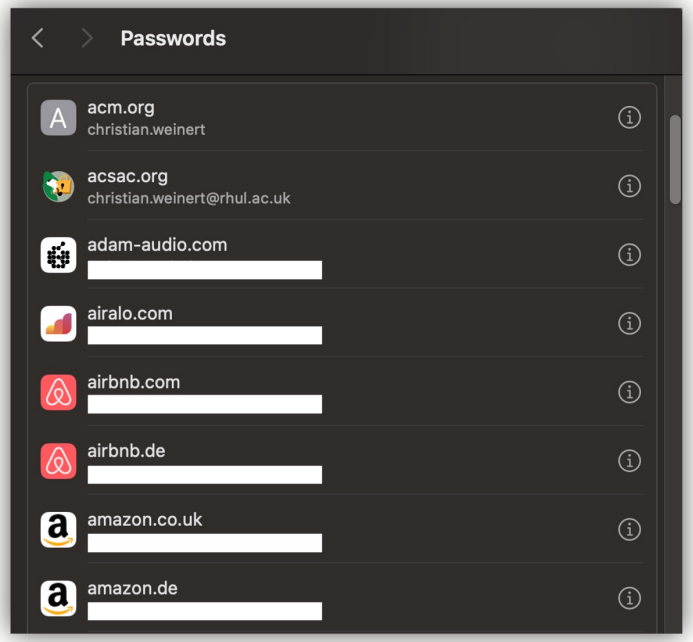
While SGX provides memory encryption and attestation that are essential to the security of this enclave, it does not guard against memory access pattern or timing side-channels. To close memory-access pattern side channels at the architectural level, the table or user records is stored in Oblivious RAM (ORAM). To close timing and access-pattern side-channels at the local level, the code is written using branchless idioms and oblivious algorithms. Please see Side-channel Resistant Programming Idioms before contributing.

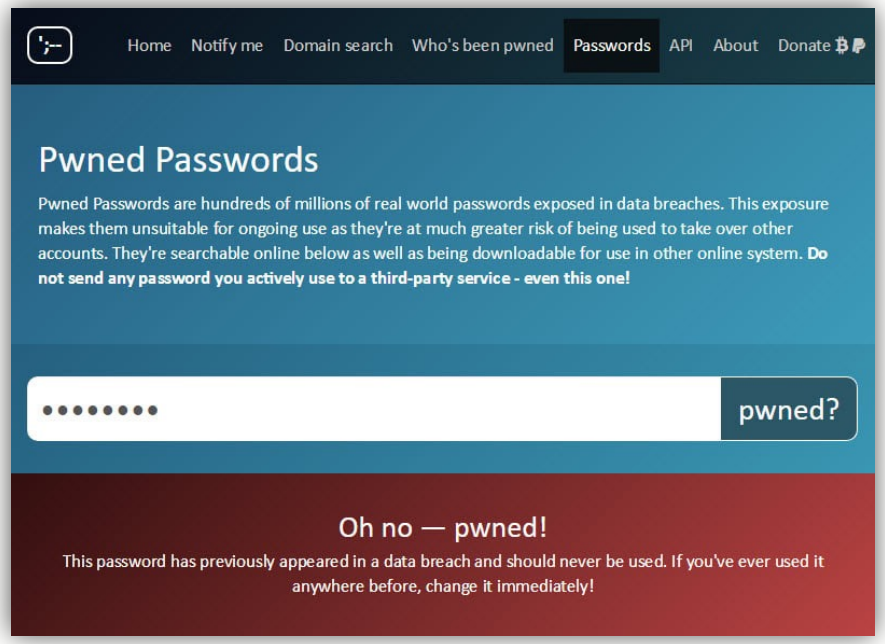


Path ORAM [Sig22]

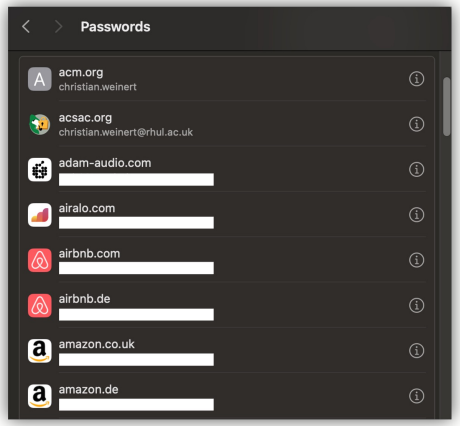


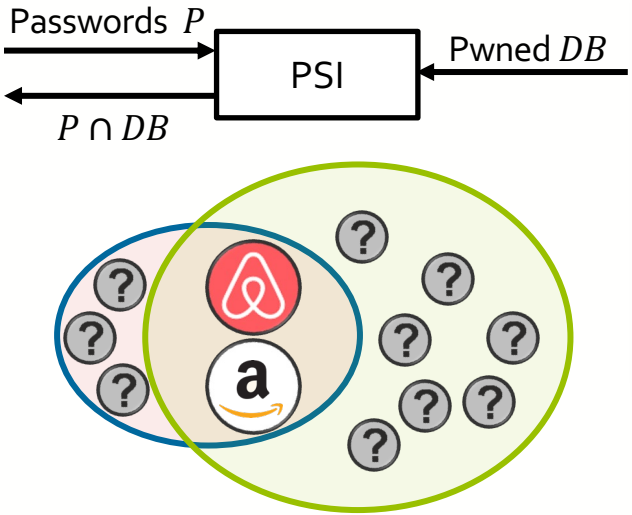
Discovery of Leaked/Compromised Passwords

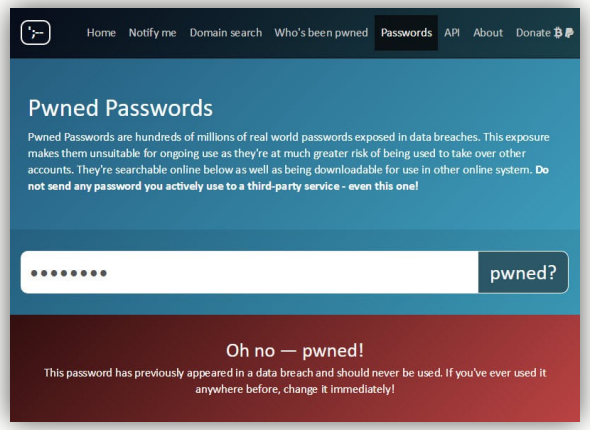




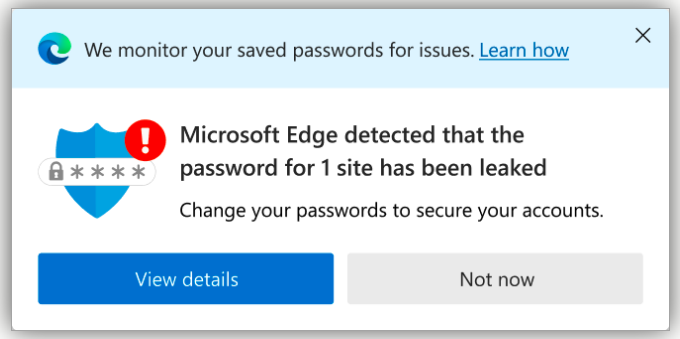
Leaked/Compromised Passwords



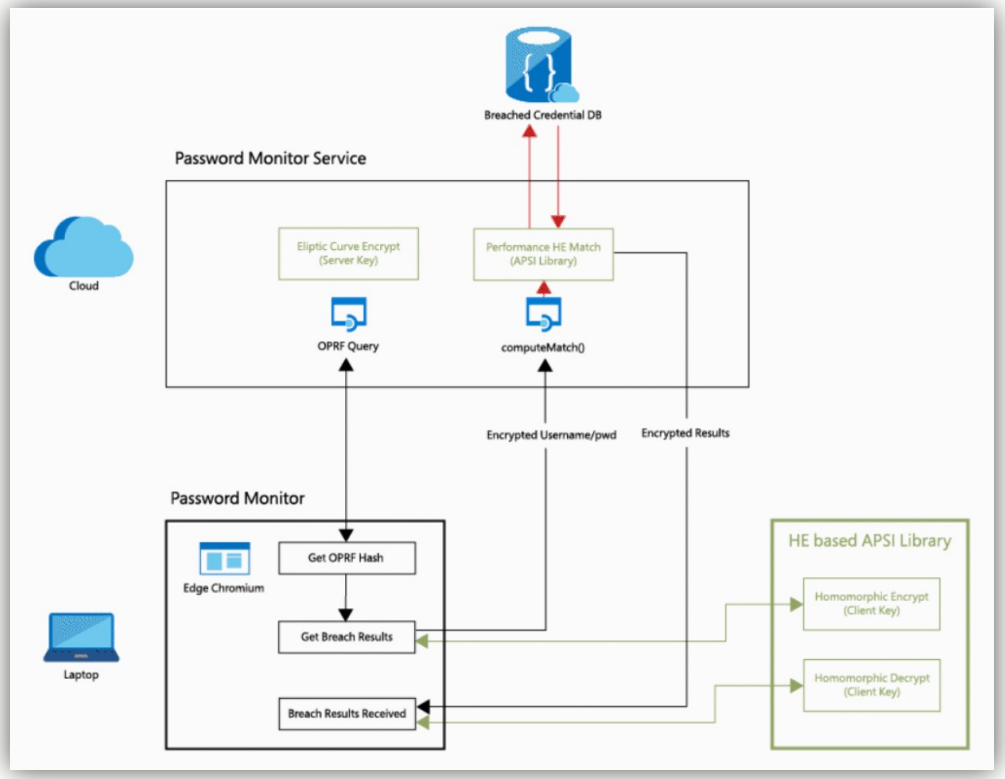




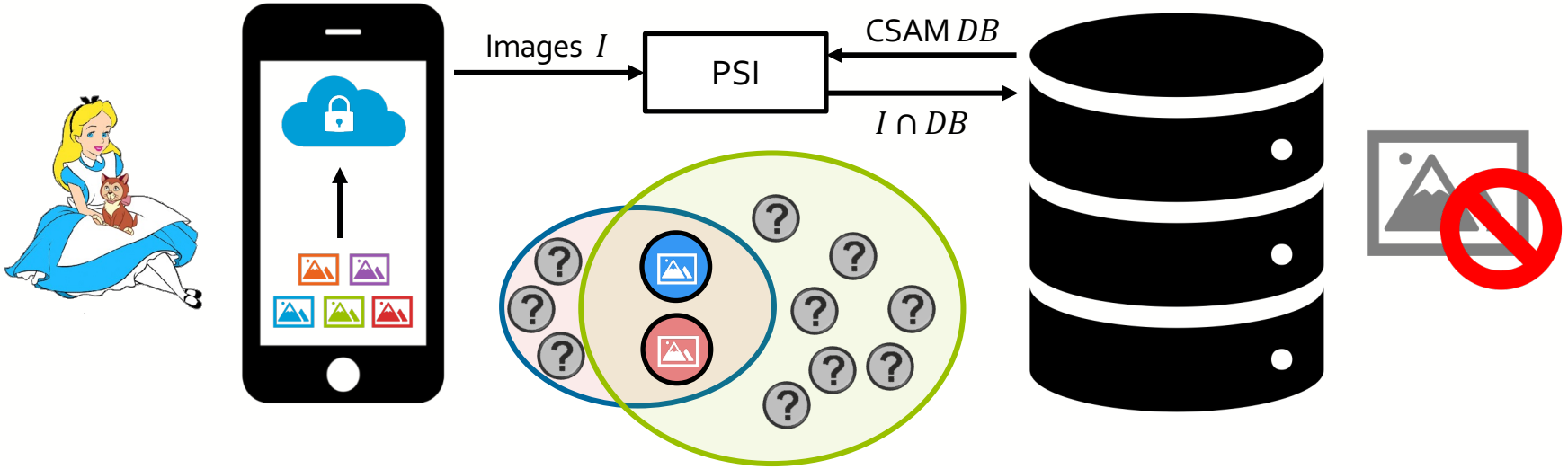
Password Monitor in Microsoft Edge



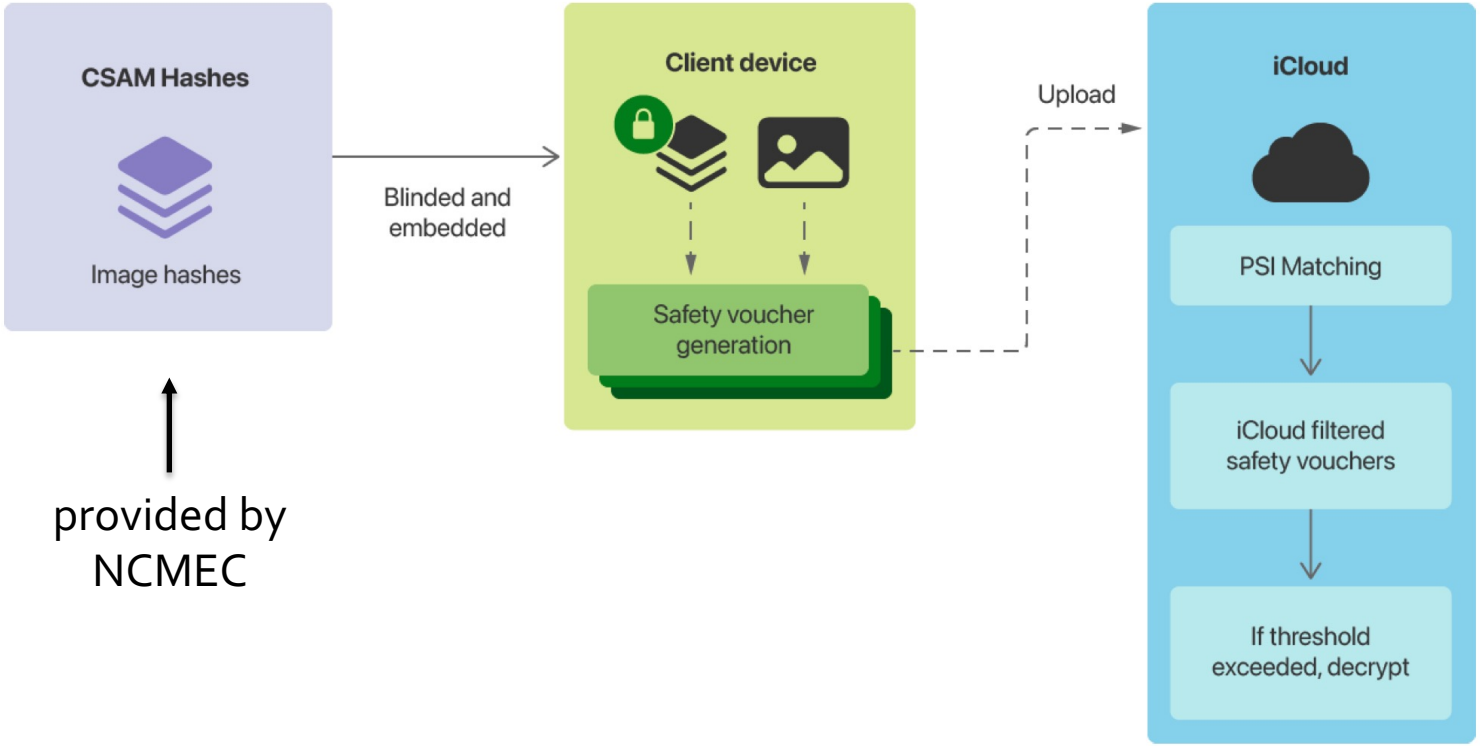
Framework based on FHE-based unbalanced PSI [Mic21]



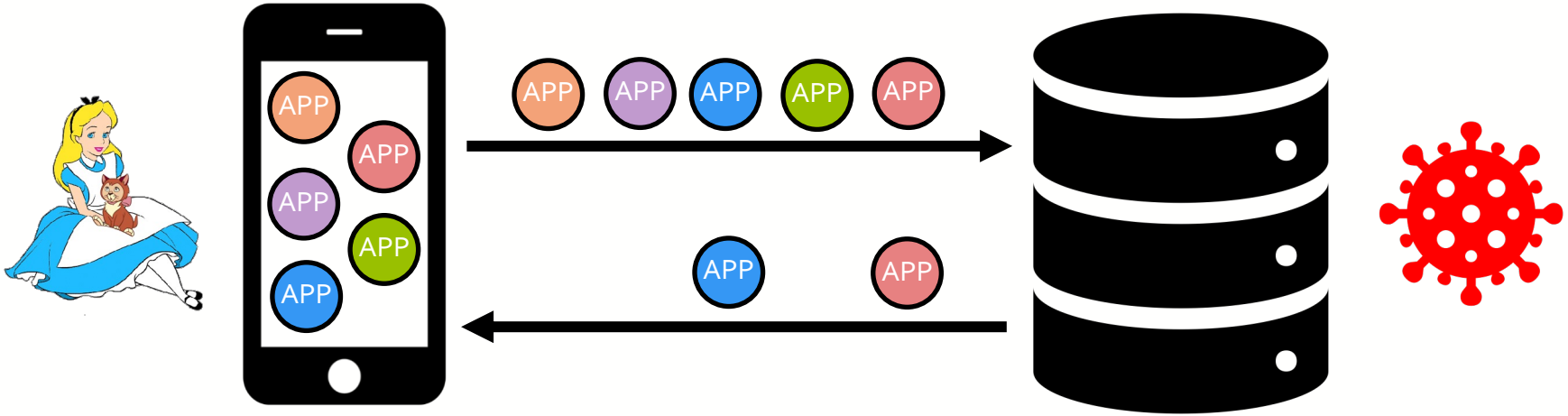
Scanning E2E-encrypted Content



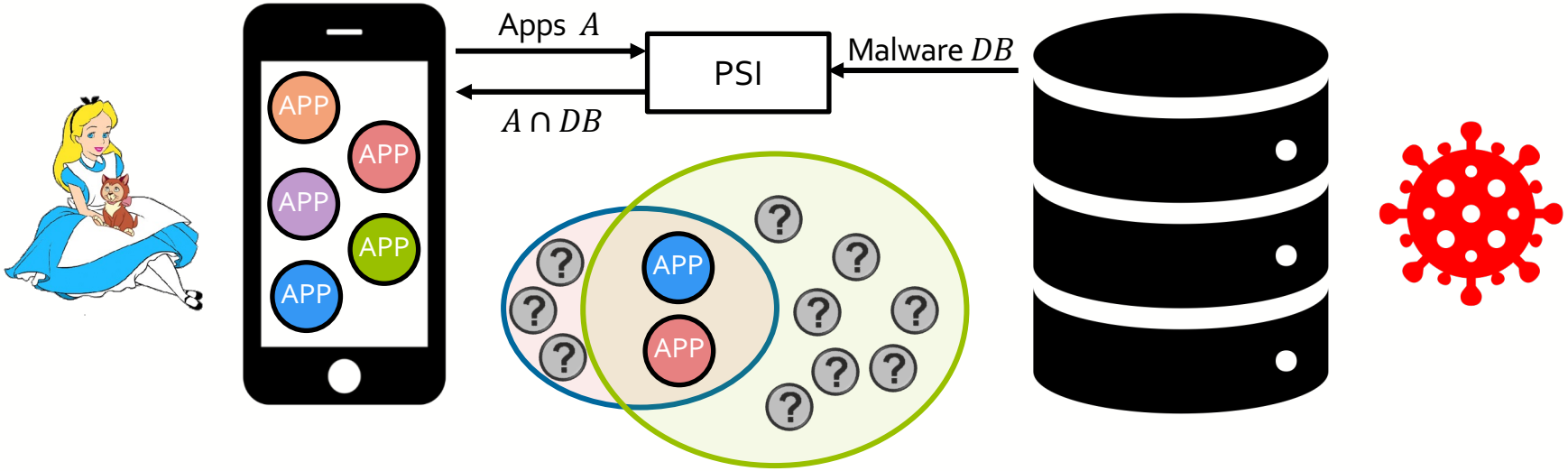
Apple's iCloud CSAM Detection [App21]



Mobile Malware Detection Service [KLS+17]



Mobile Malware Detection Service [KLS+17]



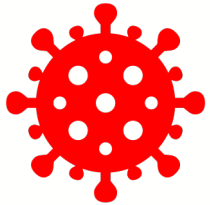
Contact Tracing (via uPSI Variants)

Log:

EU7X8KW
ksF2wXN
Lz9EhD7
MT84MhB
kgn27p2

NkZ6H63

Report EU7X8KW



EU7X8KW

Log:

NkZ6H63
8hbR2d8
uKk7srf
SNmfHQD
VRJLMsx

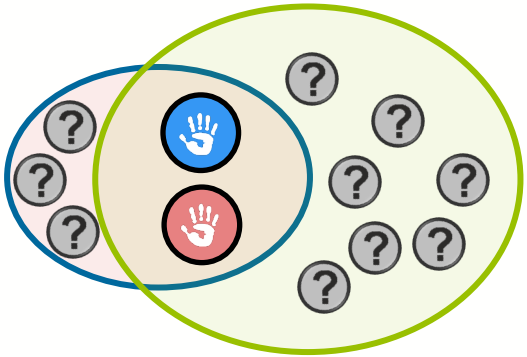
Contact Tracing (via uPSI Variants)

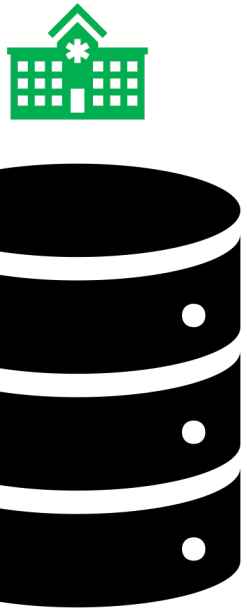
Logged IDs

$|IDs \cap DB|$

PSI

Infected DB





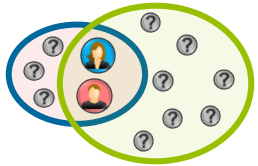
Relevant works: [DPT20 , TSS+20 , WY23]

Constructions

ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

OPRF-based uPSI – Framework

Client inputs $x_{i\in\{1,\dots,n\}}$



Precomputation Form suggested by [KLS+17]

1. Setup Phase

Client-independent Precomputation Phase, $O(|\mathrm{Y}|)$

Apply PRF to all server inputs with key k and insert them in probabilistic data structure PD for membership testing

PD ←

2. Base Phase

Client-specific Setup Phase, $O(|\mathrm{X}|)$

(e.g., OT precomputation, GC transfer, etc.) ↔

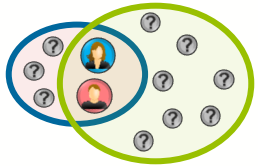
3. Online Phase $(O(|\mathrm{X}|))$

x_i →
e_i ←
OPRF
k →

Check if e_i is in PD

OPRF-based uPSI – Instantiations

Client inputs $x_{i\in\{1,\dots,n\}}$



1. Setup Phase

Client-independent Precomputation Phase, $O(|\mathrm{Y}|)$

Bloom Filter [KLS+17],
Cuckoo Filter [RA18, KRS+19]

PD

Apply PRF to all server inputs with key k and insert them in probabilistic data structure PD for membership testing

2. Base Phase

Client-specific Setup Phase, $O(|\mathrm{X}|)$

(e.g., OT precomputation, GC transfer, etc.)

OPRF Options
[KLS+17, KRS+19, …]:

- Naor-Reingold
- AES/LowMC GC
- RSA
- …

3. Online Phase $(O(|\mathrm{X}|))$

x_i

e_i

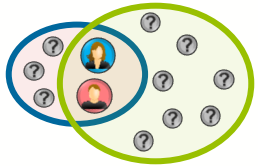
OPRF

k

Check if e_i is in PD

OPRF-based uPSI – Performance for $|X| = 2^{10}, |Y| = 2^{28}$

Client inputs $x_{i\in\{1,\ldots,n\}}$



1. Setup Phase

Client-independent Precomputation Phase, $O(|Y|)$

[KRS+19]: ≈ 1GB / 15s

Apply PRF to all server inputs with key k and insert them in probabilistic data structure PD for membership testing

PD

2. Base Phase

Client-specific Setup Phase, $O(|X|)$

(e.g., OT precomputation, GC transfer, etc.)

[KRS+19]: ≈ 2MB / 0.1s

3. Online Phase $(O(|X|))$

[KRS+19]: ≈ 2MB / 0.6s

x_i

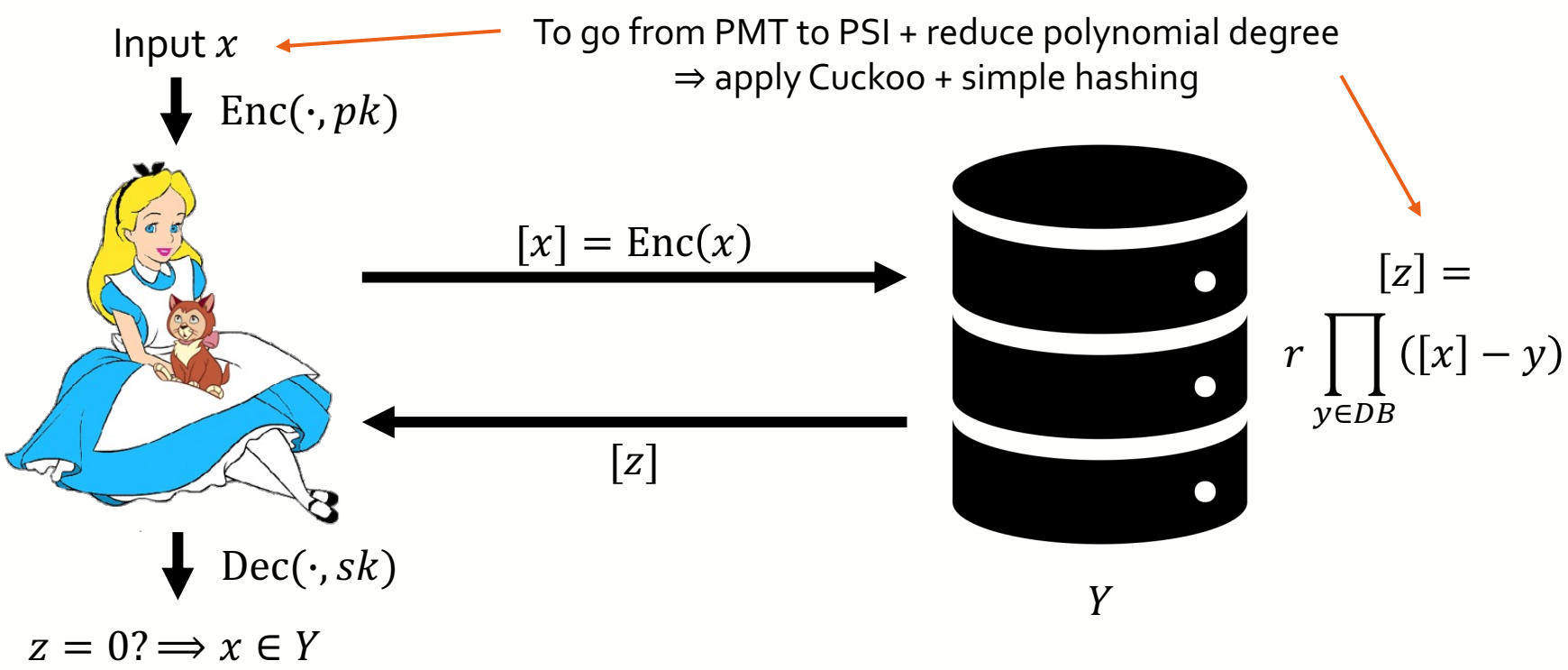
e_i

OPRF

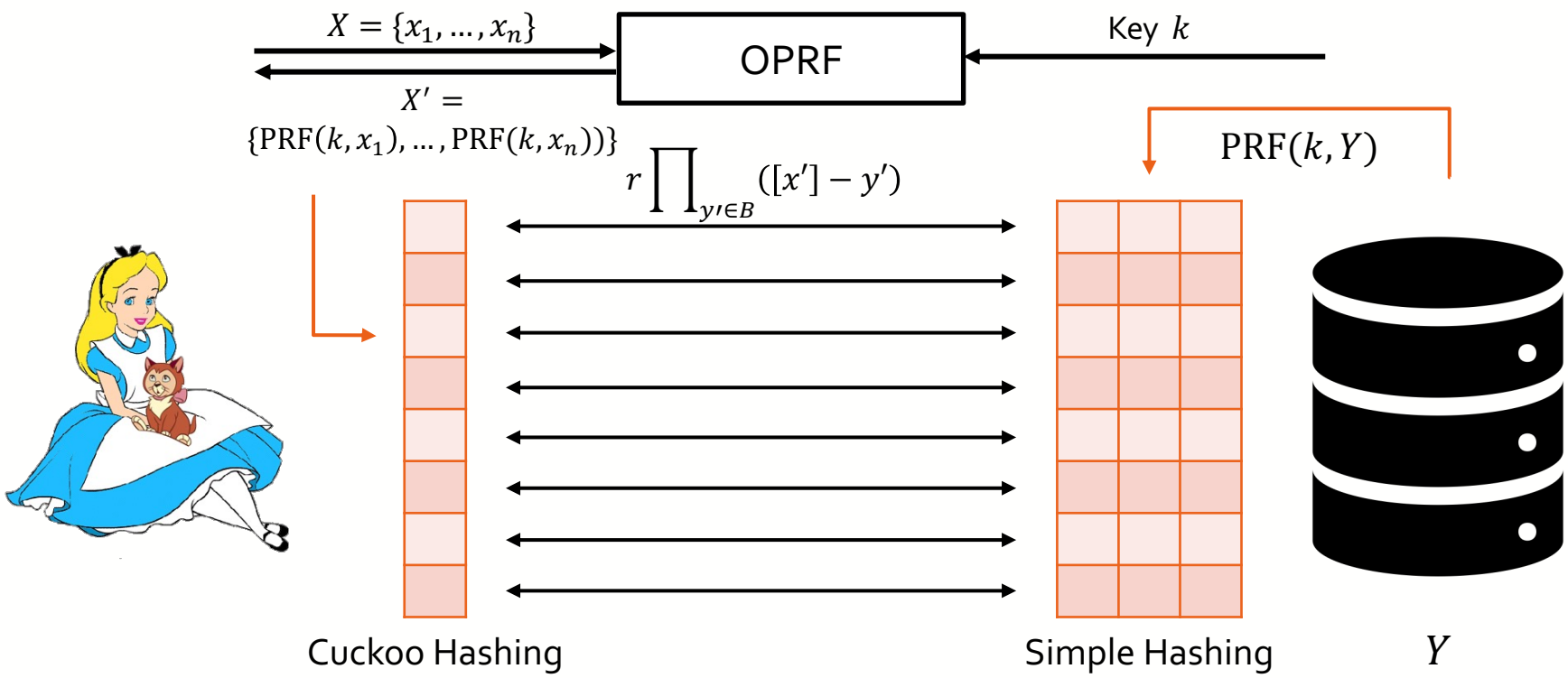
k

Check if e_i is in PD

FHE-based uPSI – General Idea



FHE-based uPSI – Framework



FHE-based uPSI – Improvements

[CLR17]
- Hashing
- Batching
- Windowing
- Partitioning
- Modulus switching

[CHLR18]
- OPRF pre-processing
- Malicious client security
- Support for arbitrary length inputs + labels

[CMG+21]
- Extremal postage-stamp bases
- SIMD packing with prime fields
- Paterson-Stockmeyer algorithm

[WY23]
- Virtual Bloom filter encoding
- Slicing + polynomial links
- Prevents "deception attacks"

FHE-based uPSI – Performance for $|X| = 2^{10}, |Y| = 2^{28}$

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Protocol	Time [s]		Comm. [MB]
	Server Setup	Online	
[CHLR18]	4,628	12.1	18.57
[CMG+21]	2,033	8.21	12.56
[WY23]	3,298	8.00	5.22

All results are in LAN setting with 32 threads on the server side

Note: "PEPSI" [MLE+24] for circuit-based uPSI reduces computation complexity from $O(|Y| \log |Y|)$ to $O(|Y|)$ but is concretely less efficient for (labeled) uPSI (>100s online time)

Combinations with PIR

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Private Information Retrieval (PIR)

Alice

Index i

$DB[i] + \text{leakage}(DB)$

PIR

Database DB

Bob

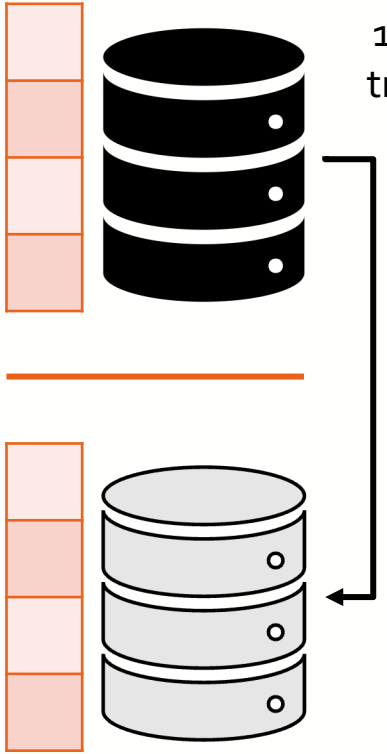
non-trivial PIR: communication overhead $< O(|DB|)$

single-server PIR

2/multi-server PIR

"PIR-PSI" [DRRT18] – General Idea

Issue: no server input privacy

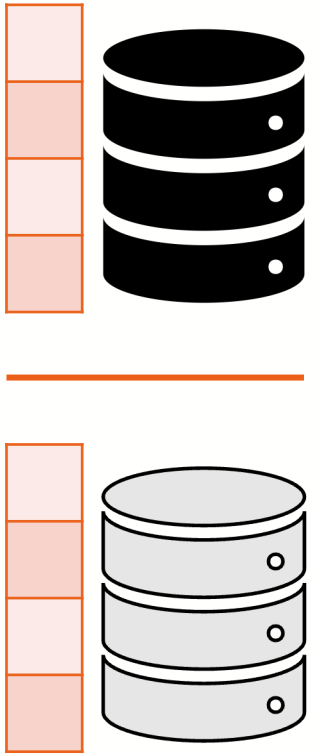


4. Compute $Y[H(x)] = v_1 \oplus v_2$

"PIR-PSI" [DRRT18] – PIR Step

2. Send random mask r

3. Multi-Server PIR Query for $H(x)$ via [BGI16]



1. One-time transfer of Y

4. Send $v_1 \oplus r$

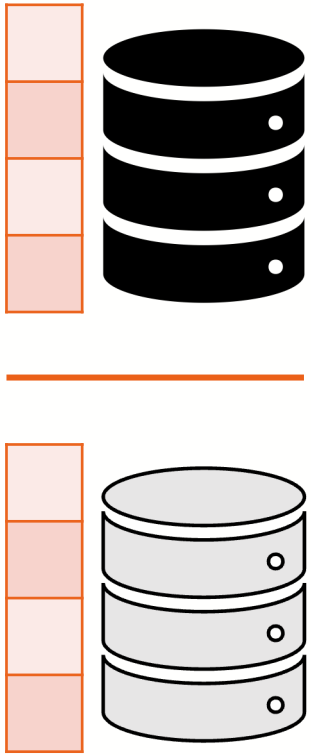
"PIR-PSI" [DRRT18] – PSI Step

$x \oplus r$

$x \in Y$

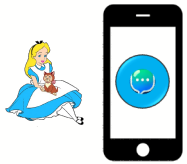
PSI
[KKRT16]

$Y[H(x)]$
$\oplus r$

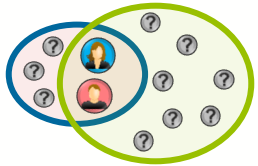


Server 2 knows masked PIR result $v_1 \oplus v_2 \oplus r = Y[H(x)] \oplus r$

OPRF-based uPSI w/ PIR Lookup [HSW23]



Client inputs $x_{i\in\{1,\dots,n\}}$



Instantiated in [HSW23] with "Offline/Online"-PIR [KC21]

1.1 Server Setup Phase

Client-independent Precomputation Phase to prepare PD

1.2 Per-Client Setup Phase

Client-specific Precomputation Phase, $O(\sqrt{|\mathrm{Y}|})$

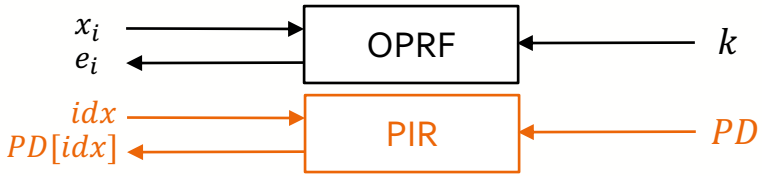
2. Base Phase

Client-specific OPRF Setup Phase, $O(|\mathrm{X}|)$

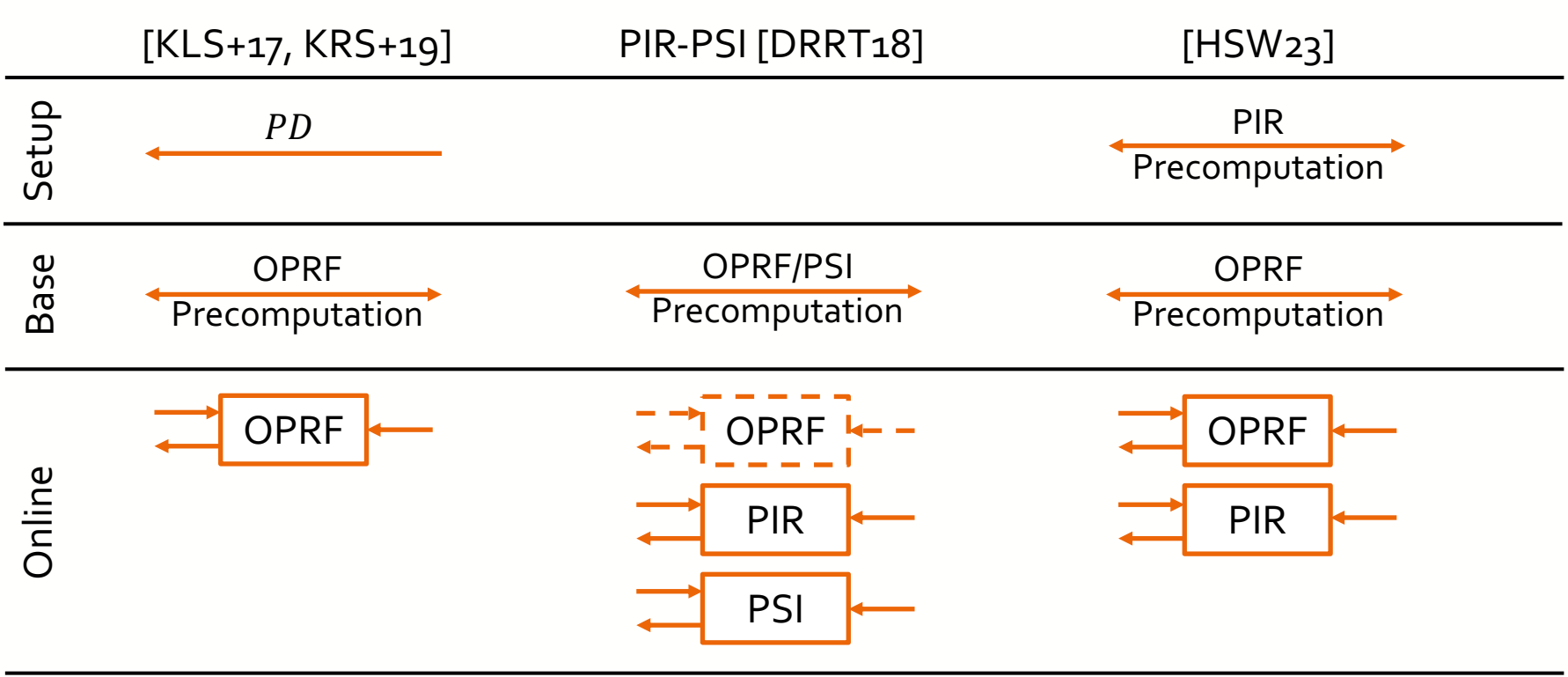
3. Online Phase $(O(|\mathrm{X}| + \log(|\mathrm{Y}|))$

Compute idx from e_i

Check if e_i is in PD



Comparison of OPRF-based uPSI – Conceptual



Comparison of OPRF-based uPSI – Performance

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Protocol	$\|Y\|$	Time [s]			Comm. [MiB]	
		Server Setup (per client)	Client Setup	Online	Setup	Online
[KRS+19]	2^{28}	-	15.17	0.63	1072.14	2.06
[DRRT18]		-	-	13.22	-	5.05
[HSW23]		63.71	35.26	1.08	66.00	4.02
	2^{31}	525.09	286.78	1.37	264.00	4.72

Disclaimer: Some results are "cherry picked"; for client sets with size $|X| = 2^{10}$; gentle multi-threading optimizations (4 client + 8 server threads) are considered for [DRRT18, HSW23]

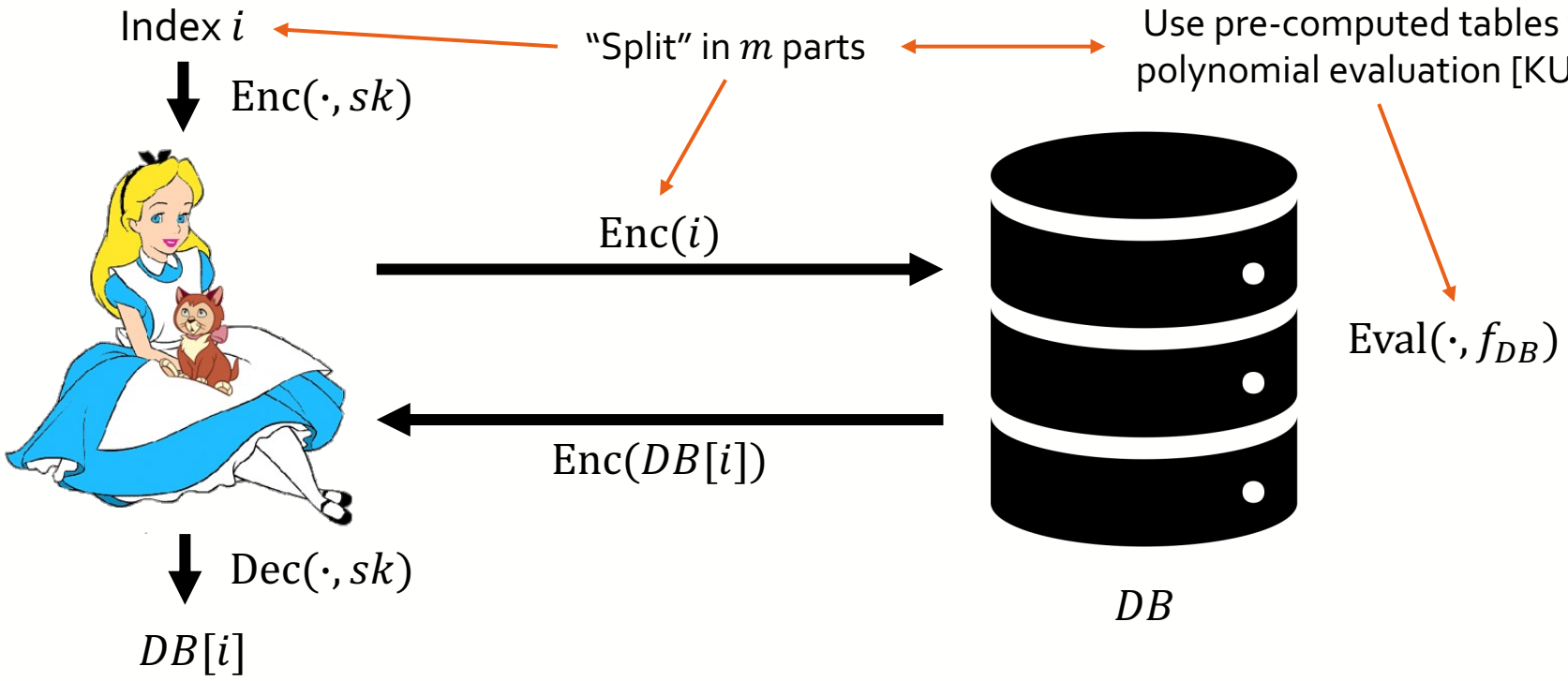
OPRF-based uPSI w/ Single-Server PIR?

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Type	Offline (client-ind.)		Offline (client-dependent)			Online		SOTA Example
	Comp	Storage	Comp	Comm	Hint size	Comp	Comm	
Stateless PIR	$\tilde{O}(N)$	$\tilde{O}(N)$	–			$\tilde{O}(N)$	$\tilde{O}(1)$	[MCR21]
Stateful PIR	$\tilde{O}(N)$	$\tilde{O}(N)$	$\tilde{O}(N)$	$\tilde{O}(\sqrt{N})$	$\tilde{O}(\sqrt{N})$	$\tilde{O}(\sqrt{N})$	$\tilde{O}(1)$	[ZLTS23]
DEPIR	$\tilde{O}(N)$	$\tilde{O}(N)$	–			$\tilde{O}(1)$	$\tilde{O}(1)$	[LMW23]

[LLMT24]: DH-style OPRF [JL10] + SimplePIR [HHC+23]

Doubly-Efficient PIR (DEPIR) [LMW23]



Doubly-Efficient PIR (DEPIR) – Concrete Performance

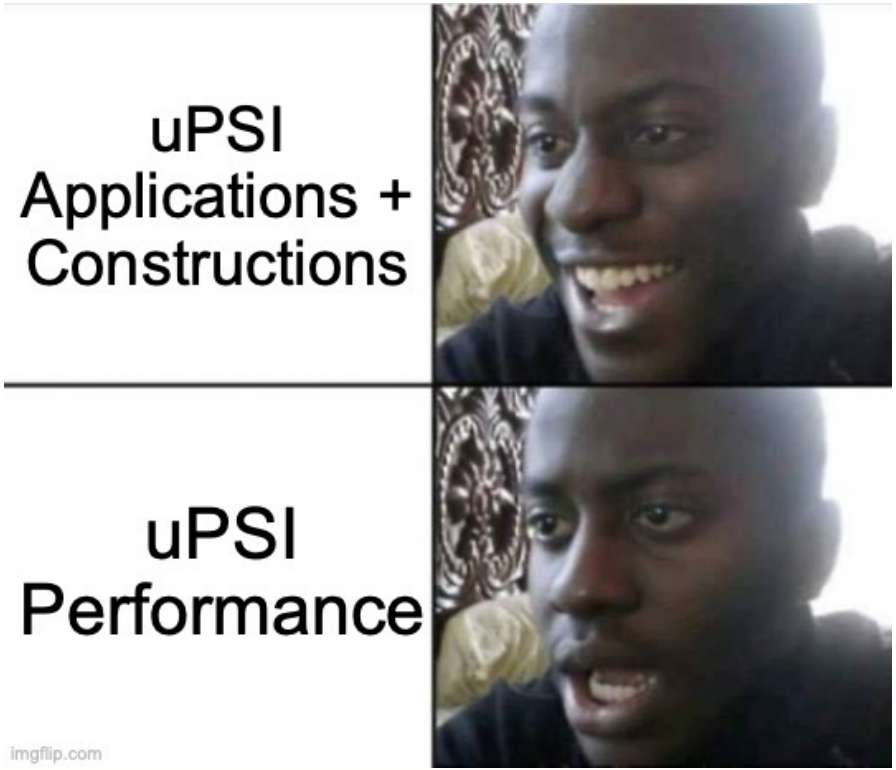
ROYAL HOLLOWAY UNIVERSITY OF LONDON

		Database Size			
		46,376	142,506	15,020,334	185,250,786
[OPPW24a]	Total Storage [TB]	0.02	2.04	873.98	411,745.17
	Total Queries [$\times 2^{30}$]	73.2	45	3154	3175
	Run-Time [min]	8	459	(impossible to benchmark)	
[OPPW24b]	Total Storage [TB]	0.04	2.93	750.92	394,474.94
	Total Queries [$\times 2^{30}$]	7	5	134	335
	Run-Time [min]	18	104	(impossible to benchmark)	

Summary

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Conclusion


uPSI
Applications +
Constructions
uPSI
Performance
imgflip.com

Thank You!

Get in touch: christian.weinert@rhul.ac.uk

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Literature (1/3)

[App21] Apple Inc.: "CSAM Detection - Technical Summary". URL: https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf

[BGI16] E. Boyle, N. Gilboa, Y. Ishai: "Function Secret Sharing: Improvements and Extensions". In: *CCS'16*.

[CHLR18] H. Chen, Z. Huang, K. Laine, P. Rindal: "Labeled PSI from Fully Homomorphic Encryption with Malicious Security". In *CCS'18*.

[CLR17] H. Chen, K. Laine, P. Rindal: "Fast Private Set Intersection from Homomorphic Encryption". In *CCS'17*.

[CMG+21] K. Cong, R. Moreno, M. da Gama, W. Dai, I. Iliashenko, K. Laine, M. Rosenberg: "Labeled PSI from Homomorphic Encryption with Reduced Computation and Communication". In *CCS'21*.

[DPT20] T. Duong, D. Phan, N. Trieu: "Catalic: Delegated PSI Cardinality with Applications to Contact Tracing". In *ASIACRYPT'20*.

[DRRT18] D. Demmler, P. Rindal, M. Rosulek, N. Trieu: "PIR-PSI: Scaling Private Contact Discovery". In *PoPETS'18*.

[HHC+23] A. Henzinger, M. M. Hong, H. Corrigan-Gibbs, S. Meiklejohn, V. Vaikuntanathan: "One Server for the Price of Two: Simple and Fast Single-Server Private Information Retrieval". In *USENIX Sec'23*.

[HSW23] L. Hetz, T. Schneider, C. Weinert: "Scaling Mobile Private Contact Discovery to Billions of Users". In *ESORICS'23*.

[JL10] S. Jarecki, X. Liu: "Fast Secure Computation of Set Intersection". In *SCN'10*.

[KC21] D. Kogan, H. Corrigan-Gibbs: "Private Blocklist Lookups with Checklist". In *USENIX Sec'21*.

Literature (2/3)

[KKRT16] V. Kolesnikov, R. Kumaresan, M. Rosulek, N. Trieu: "Efficient Batched Oblivious PRF with Applications to Private Set Intersection". In: *CCS'16*.

[KLS+17] Á. Kiss, J. Liu, T. Schneider, N. Asokan, B. Pinkas: "Private Set Intersection for Unequal Set Sizes with Mobile Applications". In *PoPETS'17*.

[KRS+19] D. Kales, C. Rechberger, T. Schneider, M. Senker, C. Weinert: "Mobile Private Contact Discovery at Scale". In *USENIX Sec'19*.

[KU11] K. S. Kedlaya, C. Umans: "Fast Polynomial Factorization and Modular Composition". In: *SIAM J. Comput.'11*.

[LLMT24] C. Lin, Z. Liu, P. Miao, M. Tromanhauser: "Finding Balance in Unbalanced PSI: A New Construction from Single-Server PIR". URL: https://cs.brown.edu/media/filer_public/32/f1/32f1278e-e603-43ba-9097-3f9ae39ea09d/maxtromanhauser.pdf

[LMW23] W. Lin, E. Mook, D. Wichs: "Doubly Efficient Private Information Retrieval and Fully Homomorphic RAM Computation from Ring LWE". In: *STOC'23*.

[MCR21] M. H. Mughees, H. Chen, L. Ren: "OnionPIR: Response Efficient Single-Server PIR". In: *CCS'21*.

[Mic21] K. Lauter, S. Kannepalli, K. Laine, R. C. Moreno (Microsoft): "Password Monitor: Safeguarding passwords in Microsoft Edge". URL: https://www.microsoft.com/en-us/research/blog/password-monitor-safeguarding-passwords-in-microsoft-edge/

Literature (3/3)

[MLE+24] R. Mahdavi, N. Lukas, F. Ebrahimianghazani, T. Humphries, B. Kacsmar, J. Premkumar, X. Li, S. Oya, E. Amjadian, F. Kerschbaum: "PEPSI: Practically Efficient Private Set Intersection in the Unbalanced Setting". In *USENIX Sec'24*.

[OPPW24a] H. Okada, R. Player, S. Pohmann, C. Weinert: "Towards Practical Doubly-Efficient Private Information Retrieval". In *FC'24*.

[OPPW24b] H. Okada, R. Player, S. Pohmann, C. Weinert: "On Algebraic Homomorphic Encryption and its Applications to Doubly-Efficient PIR". URL: https://ia.cr/2024/1307

[RA18] A. Resende, D. Aranha: "Faster Unbalanced Private Set Intersection". In *FC'18*.

[Sig22] G. Connell (Signal): "Technology Deep Dive: Building a Faster ORAM Layer for Enclaves". URL: https://signal.org/blog/building-faster-oram/

[TSS+20] N. Trieu, K. Shehata, P. Saxena, R. Shokri, D. Song: "Epione: Lightweight Contact Tracing with Strong Privacy". In *IEEE Data Eng. Bull.'20*.

[WY23] M. Wu, T. Yuen: "Efficient Unbalanced Private Set Intersection Cardinality and User-friendly Privacy-preserving Contact Tracing". In *USENIX Sec'23*.

[ZLTS23] M. Zhou, W. Lin, Y. Tselekounis, E. Shi: "Optimal Single-Server Private Information Retrieval". In: *EUROCRYPT'23*.

London Crypto Day 2024

Date: Friday 15th November 2024

Location: Enigma Room, The Alan Turing Institute (British Library)

Confirmed Speakers:

- Andrew Mendelsohn, *Imperial College*
- Dave Buckley, *OpenMined*
- Eamonn Postlethwaite, *KCL*
- Maria Corte-Real Santos, *UCL*
- Lydia Garms, *EY*
- Saqib Kakvi, *RHUL*

https://sites.google.com/view/london-crypto-day/