Faster Private Set Intersection based on OT Extension

TECHNISCHE UNIVERSITÄT DARMSTADT

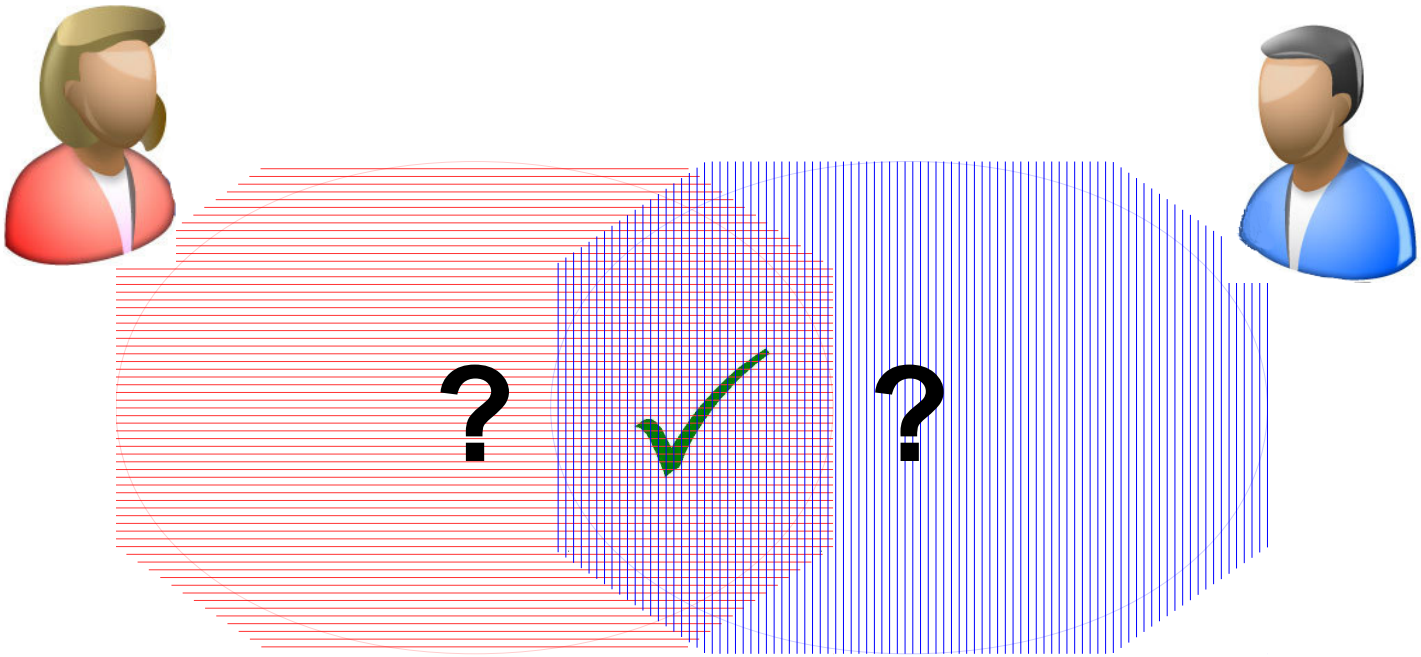
Michael Zohner (TU Darmstadt)

Joint work with
Benny Pinkas (Bar Ilan University)
Thomas Schneider (TU Darmstadt)

EC SPRIDE

Private Set Intersection (PSI)

TECHNISCHE UNIVERSITÄT DARMSTADT



EC SPRIDE

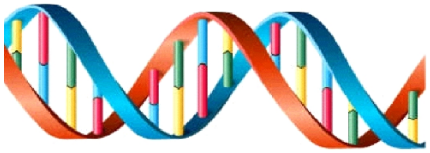
Applications

Secure database join

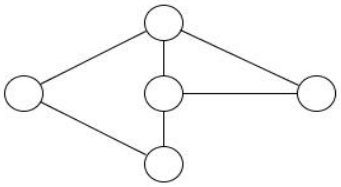
Common contacts

Botnet detection

Cheater detection in online games



Testing human genomes



Relationship path discovery

A naïve PSI protocol

TECHNISCHE UNIVERSITÄT DARMSTADT

Input: $x_1, \ldots, x_n$

Input: $y_1, \ldots, y_n$

$H(x_1), \ldots, H(x_n)$

$H(y_1), \ldots, H(y_n)$ ◄

$H(y_1), \ldots, H(y_n)$

$H(x_i) \stackrel{?}{=} H(y_j)$, for $0 < i, j < n$

- Pro: fast, little communication

- Con: can leak privacy of Bob's inputs

Our Contributions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Survey major results on semi-honest PSI

- Optimize existing PSI protocols

- Present a new PSI scheme

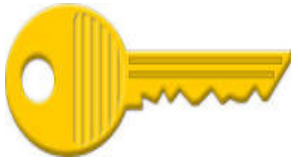
- Compare performance of all schemes

EC SPRIDE

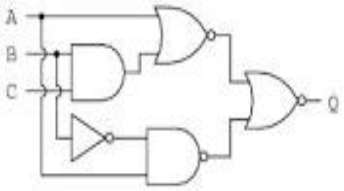
Existing PSI Protocols

TECHNISCHE UNIVERSITÄT DARMSTADT

- Public-key Cryptography
  - DH-based Protocol [M86], O(n) pk-crypto
  - Blind RSA Protocol [CT10], O(n) pk-crypto



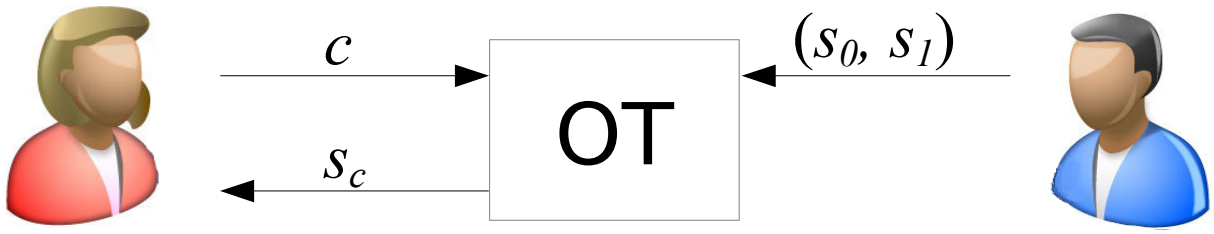
- Generic Secure Computation
  - Based on Yao's garbled circuits, GMW
  - Circuit in [HEK12], O($n \log n$) sym-crypto



- Oblivious Transfer
  - Only sym-crypto via OT extension [IKNP03]
  - Bloom-filter [DCW13], O(n) sym-crypto

EC SPRIDE

Oblivious Transfer (OT)



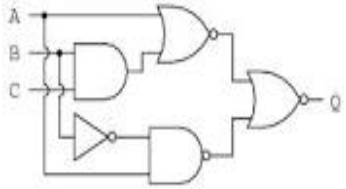
- **Input**: Bob holds two strings (s_0, s_1), Alice holds a choice bit c

- **Output**: Alice receives s_c but learns nothing about s_{1-c},
  Bob learns nothing about c

Optimizing Existing Protocols

TECHNISCHE UNIVERSITÄT DARMSTADT

- Improve circuit-based PSI of [HEK12] using GMW
  - Multiplexer complexity independent of bit-length
  - Reduce computation / communication by factor 2
  - Also applicable to other functionalities



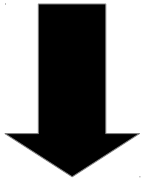
- Randomize Garbled Bloom filter of [DCW13]
  - Reduce computation by factor 3
  - Reduce communication by factor 4
  - Whole protocol can be parallelized

EC SPRIDE

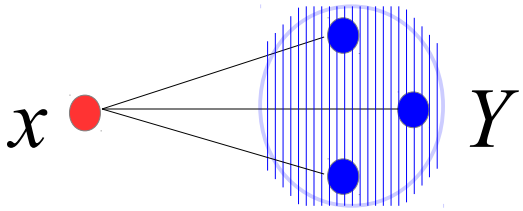
PSI based on OT

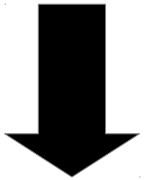
TECHNISCHE UNIVERSITÄT DARMSTADT

Private Equality Test:

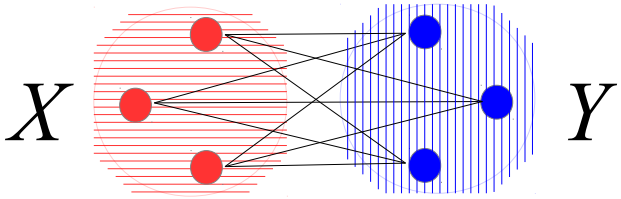


Private Set Inclusion:





Private Set Intersection:



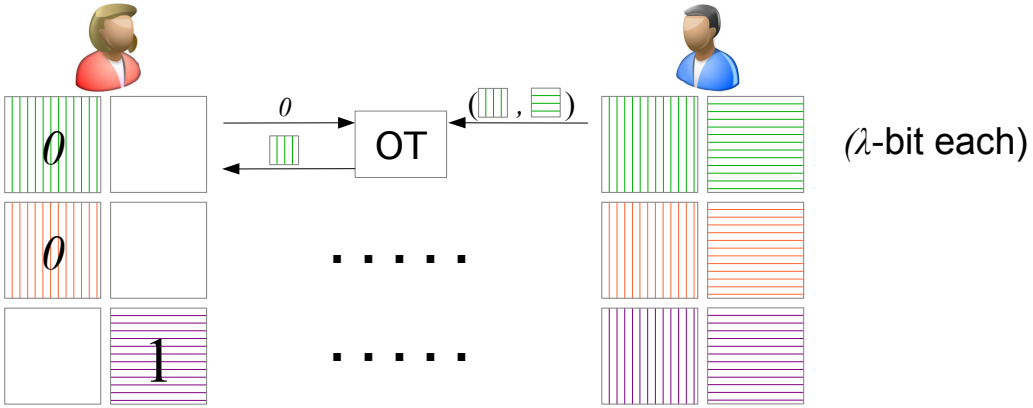
EC SPRIDE

PSI based on OT (Equality Test)

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Input:** Alice has x, Bob has y. **Output:** $x \stackrel{?}{=} y$

- **Example:** $x = 001, y = 011$



- Bob sends λ-bit mask $0 \oplus 1 \oplus 1$ to Alice

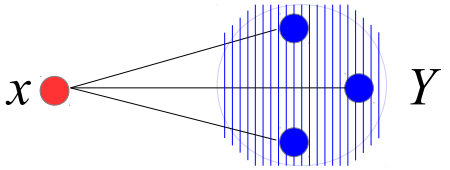
- Alice computes $0 \oplus 0 \oplus 1$ and compares

EC SPRIDE

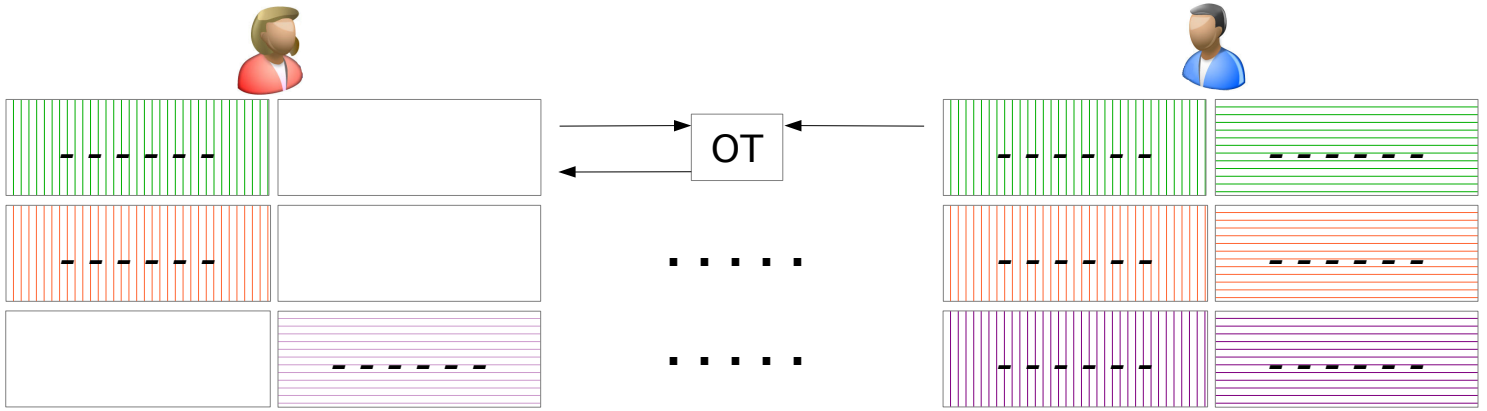
PSI based on OT (Set Inclusion)

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Input:** Alice has x, Bob has $Y = \{y_1, \ldots, y_n\}$. **Output:** $x \overset{?}{\in} Y$



- Run n Private Equality Tests in parallel
  - Alice's OT choices for all $y_1, \ldots, y_n$ are the same
  - Send $n\lambda$ bits from Bob to Alice



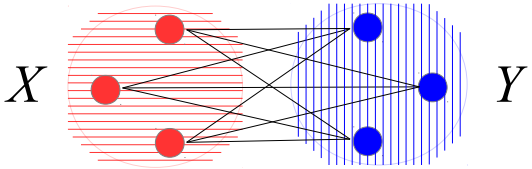
EC SPRIDE

PSI based on OT (Set Intersection)

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Input:** Alice has $X = \{x_1, \ldots, x_n\}$, Bob has $Y = \{y_1, \ldots, y_n\}$.

Output: $X \cap Y$.



- Run n Private Set Inclusions in parallel
  - Requires n^2 comparisons, hence not an option

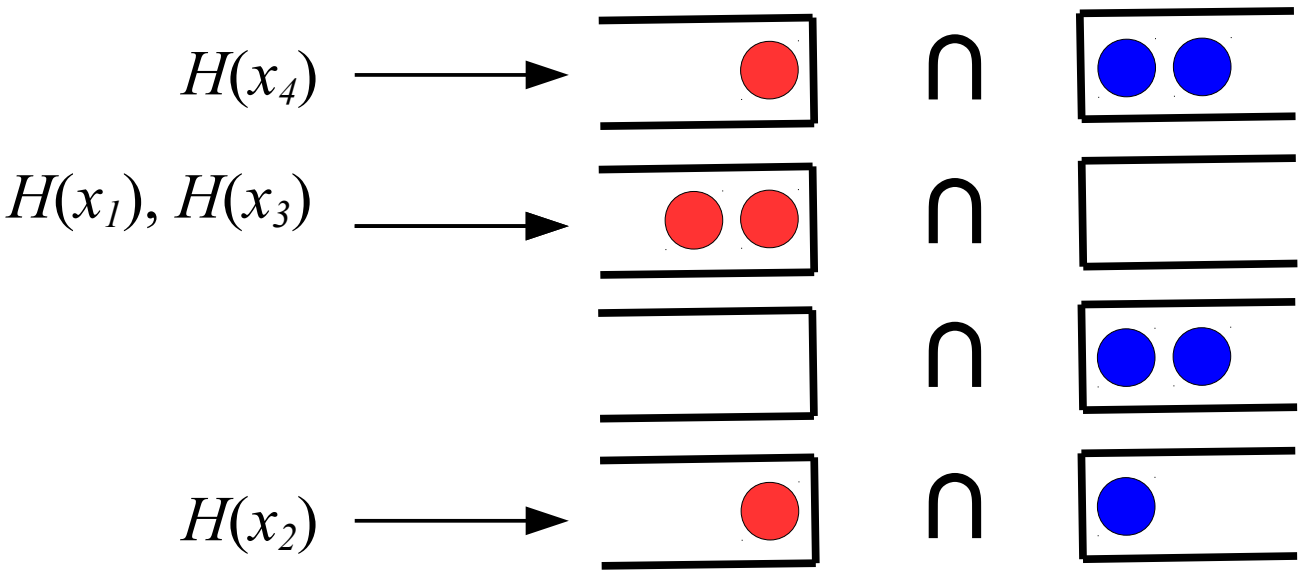
EC SPRIDE

Hashing

TECHNISCHE UNIVERSITÄT DARMSTADT

- Hash elements to bins to reduce comparisons

- **Example:** Alice holds $X=\{x_1, \ldots, x_4\}$, Bob holds $Y=\{y_1, \ldots, y_4\}$



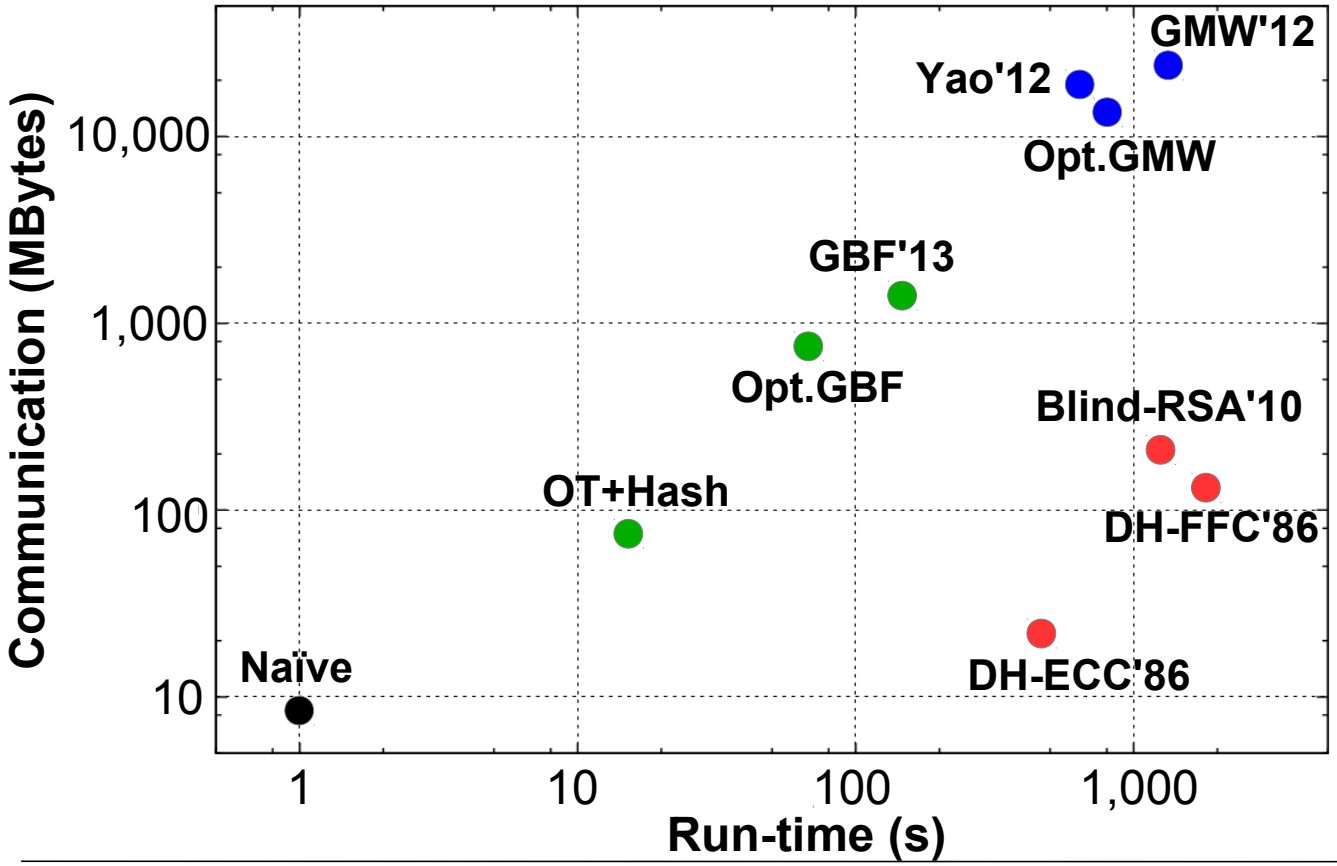
- Reduces comparisons from n^2 to $O(n \log n)$

EC SPRIDE

Comparison Results

TECHNISCHE UNIVERSITÄT DARMSTADT

- PSI on $n = 2^{18}$ elements of 32-bit length for 128-bit security on Gbit LAN

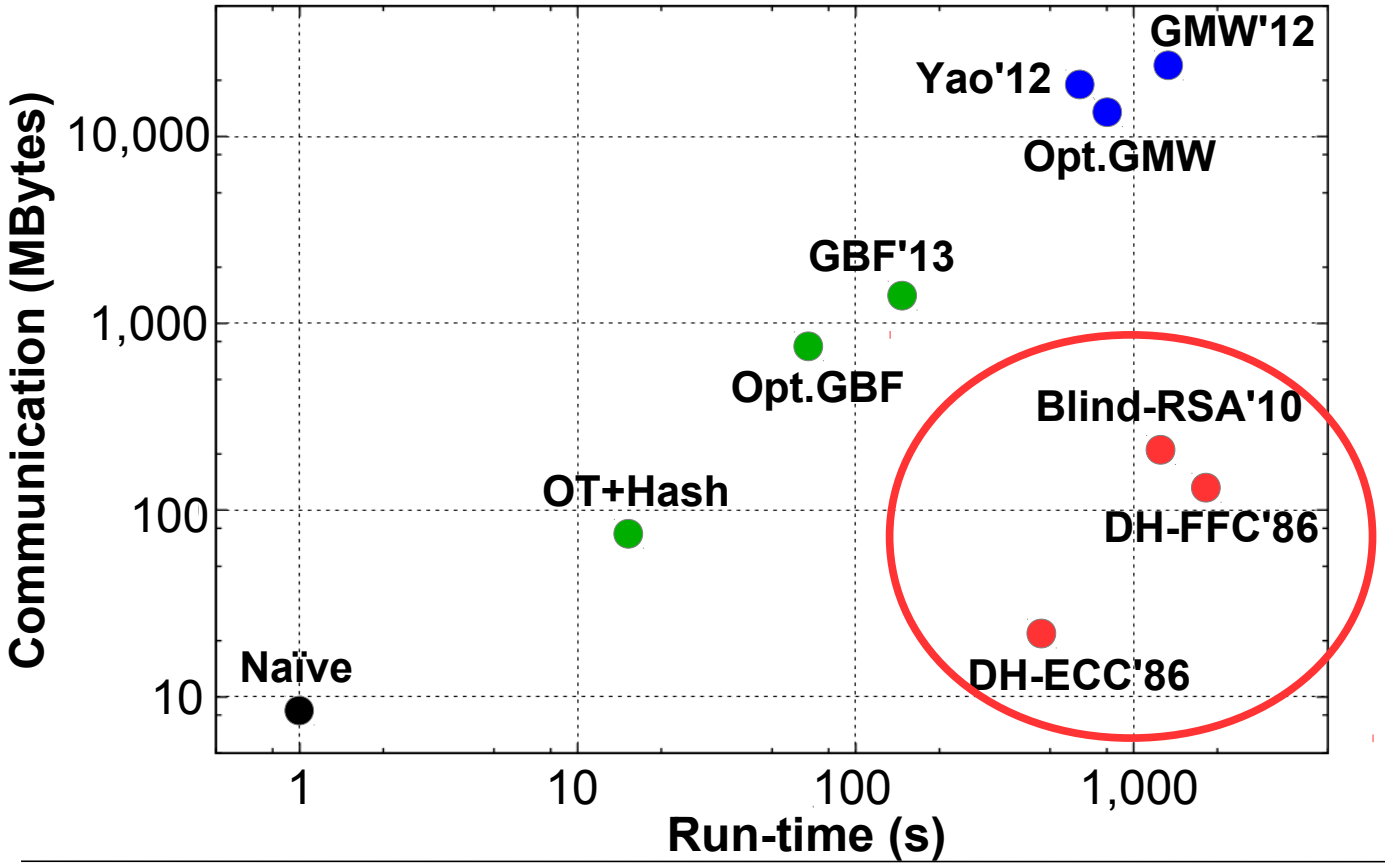


EC SPRIDE

Comparison Results

TECHNISCHE UNIVERSITÄT DARMSTADT

- PSI on $n = 2^{18}$ elements of 32-bit length for 128-bit security on Gbit LAN



PK-Based:

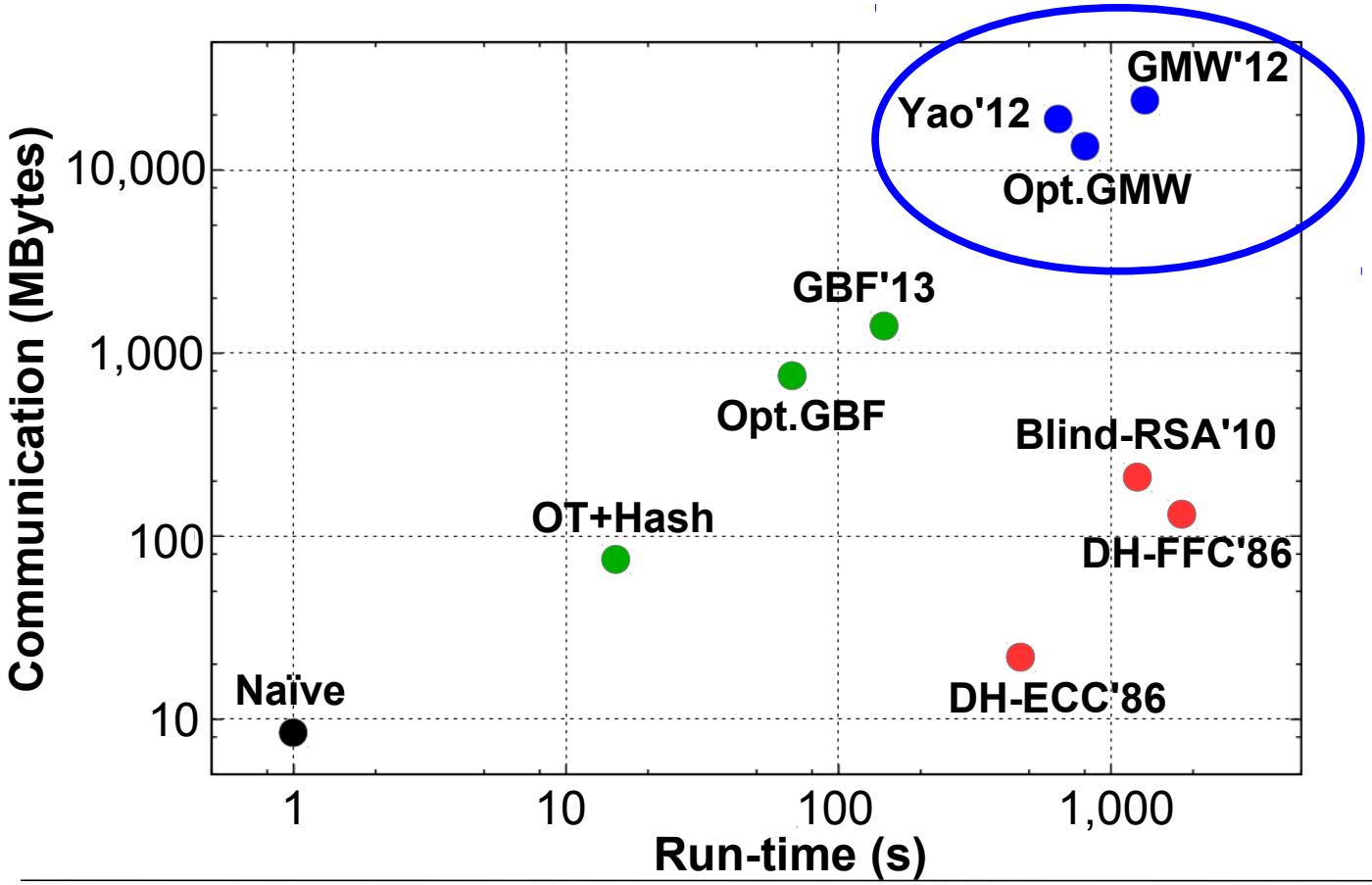
- high run-time for large security parameters
+ best communication

EC SPRIDE

Comparison Results

TECHNISCHE UNIVERSITÄT DARMSTADT

- PSI on $n = 2^{18}$ elements of 32-bit length for 128-bit security on Gbit LAN



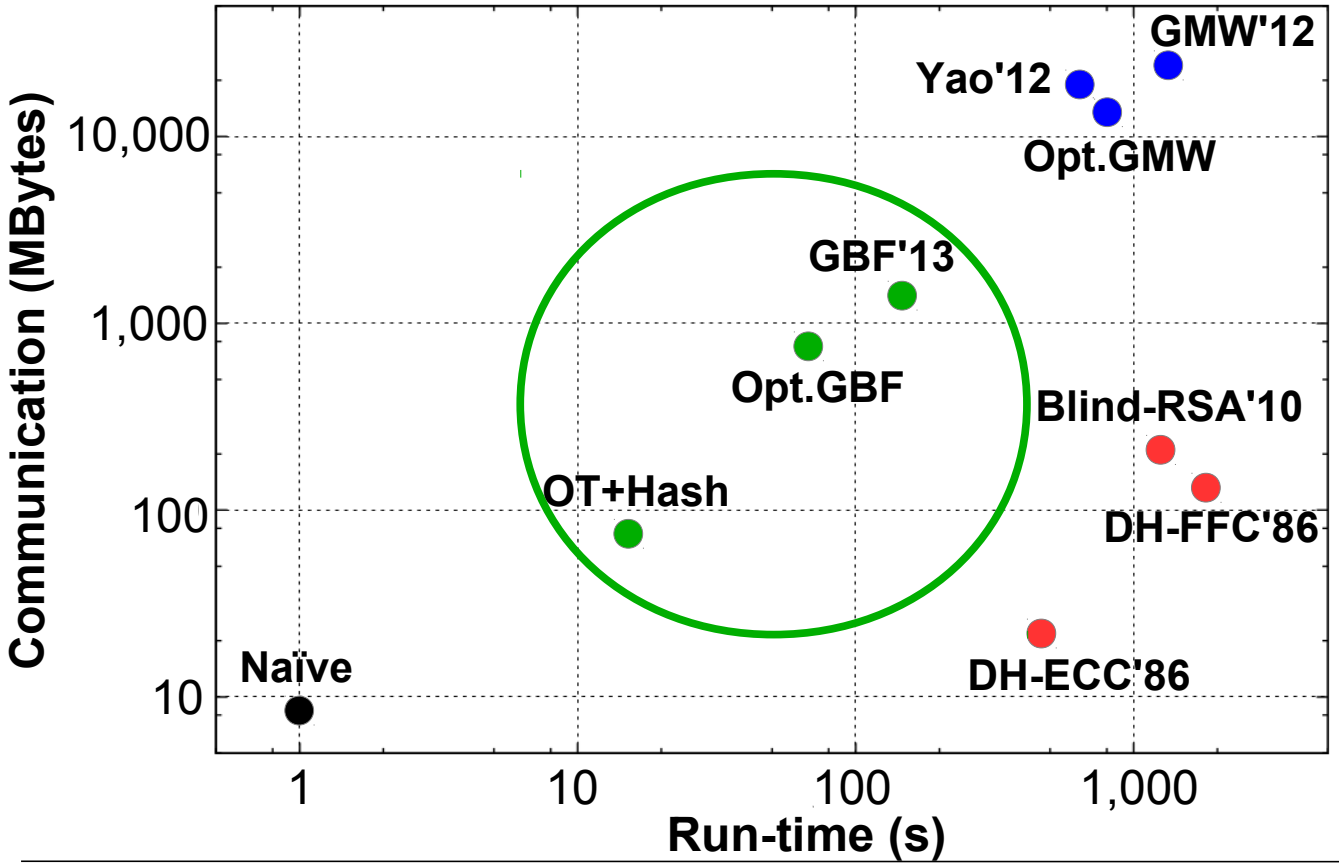
Circuit-Based:
- high run-time & communication
+ easily extensible to arbitrary functions

EC SPRIDE

Comparison Results

TECHNISCHE UNIVERSITÄT DARMSTADT

- PSI on $n = 2^{18}$ elements of 32-bit length for 128-bit security on Gbit LAN



OT-Based:
+ good communication and run-time

EC SPRIDE

Conclusion

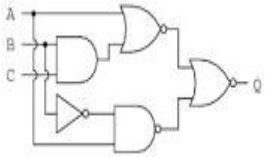
TECHNISCHE UNIVERSITÄT DARMSTADT

Rule of Thumb:

- OT-based protocols in general case

- DH-based ECC if communication is bottleneck

- Circuit-based protocols for easy extension



Goal: PSI on million element sets in less than 1 second

Faster Private Set Intersection based on OT Extension

TECHNISCHE UNIVERSITÄT DARMSTADT

Thank you for your attention

References

- [M86] C. Meadows: A more efficient cryptographic matchmaking protocol for use in the absence of a continuously available third party. In IEEE S&P 86.

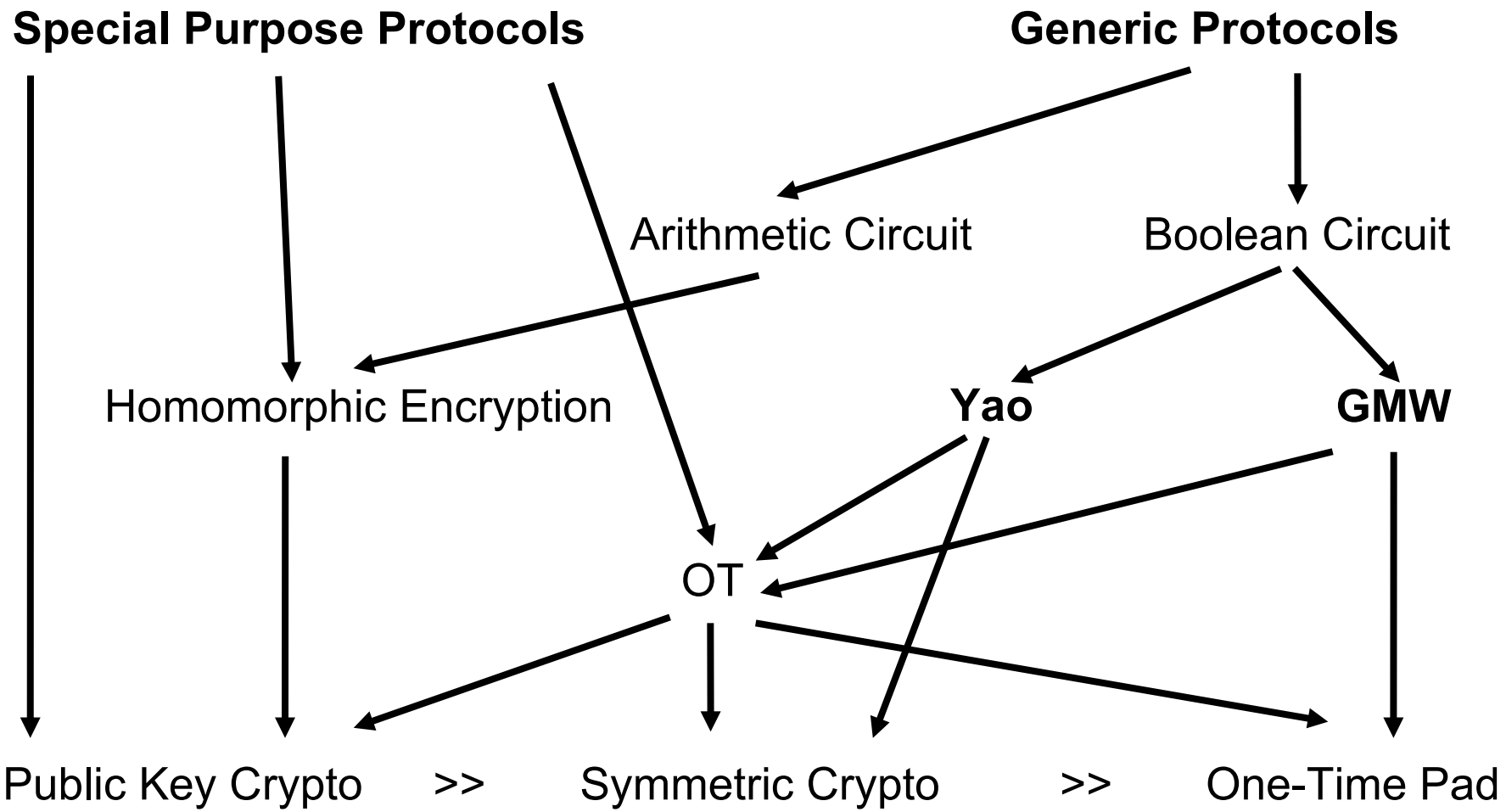
- [CT10] E. De Cristofaro and G. Tsudik: Practical private set intersection protocols with linear complexity. In FC'10.

- [HEK12] Y. Huang, D. Evans, and J. Katz: Private set-intersection: Are garbled circuits better than custom protocols? In NDSS'12.

- [DCW13] C. Dong, L. Chen, and Z. Wen: When private set intersection meets big data: An efficient and scalable protocol. In ACM CCS'13.

- [IKNP03] Y. Ishai, J. Kilian, K. Nissim, E. Petrank: Extending Oblivious Transfers Efficiently. In CRYPTO'03.

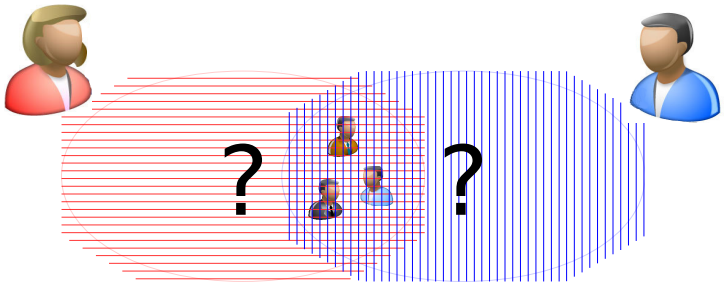
Protocol Overview



Faster Private Set Intersection based on OT Extension

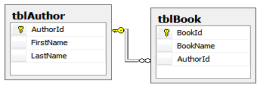
TECHNISCHE UNIVERSITÄT DARMSTADT

Private Set Intersection



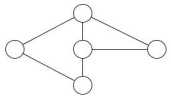
Applications

- Secure Database Join



- Common Contacts

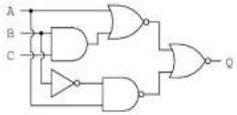
- Relationship Path Discovery



Techniques

- Public-key Cryptography

- Generic Secure Computation



- Oblivious Transfer

Summary

- Optimize & Implement

- New Protocol

- Comparison

1 Mio. elements in 20 seconds on PCs

EC SPRIDE

Results

TECHNISCHE UNIVERSITÄT DARMSTADT

Protocol	Communication	Computation
Naive Hashing	8	1
DH FFC	192	1224
DH ECC	26	416
Blind RSA	132	1982
Circuit + GMW	23400	1304
Circuit + Opt GMW	14040	762
Yao	20736	609
Garbled Bloom	1393	154
Opt. GBF	740	68
OT + hashing	78	14