Secure Intersection with MapReduce

R. Ciucanu[1] M. Giraud[2] **P. Lafourcade**[2] L. Ye[3]

[1]LIFO, INSA Centre Val de Loire
Université d'Orléans

[3]School of Computer Science and Technology
Harbin Institute of Technology, China

26 July 2019 @ SECRYPT, Prague

Big Data

Cloud Service Provider (CSP)

aws

Azure

Google Cloud

Model 1

Application

Avoid double submissions in conferences

Mutual Private Set Intersection (PSI)

		
Participants List	A	B
Result	$A \cap B$	$A \cap B$

Model 2

Application

FBI wants to detect suspicious passengers of an airline company

One-way PSI

		
Passengers List	A	B
Result	$A \cap B$	$\emptyset$

Model 3

Application

Interpol wants the most dangerous persons from FBI and MI6

Our PSI Model

			
Suspects Lists	A	B	
Result	$\emptyset$	$\emptyset$	$A \cap B$

Example

Suspects Lists

Alice
Cesar
Mallory

Oscar
Bob
Mallory

Intersection List

Mallory

Outline

Outline

MapReduce[1]

MapReduce Environment

Take care of

- Partitioning input data
- Scheduling program execution on a set of machines
- Handling machine failures

Programmer

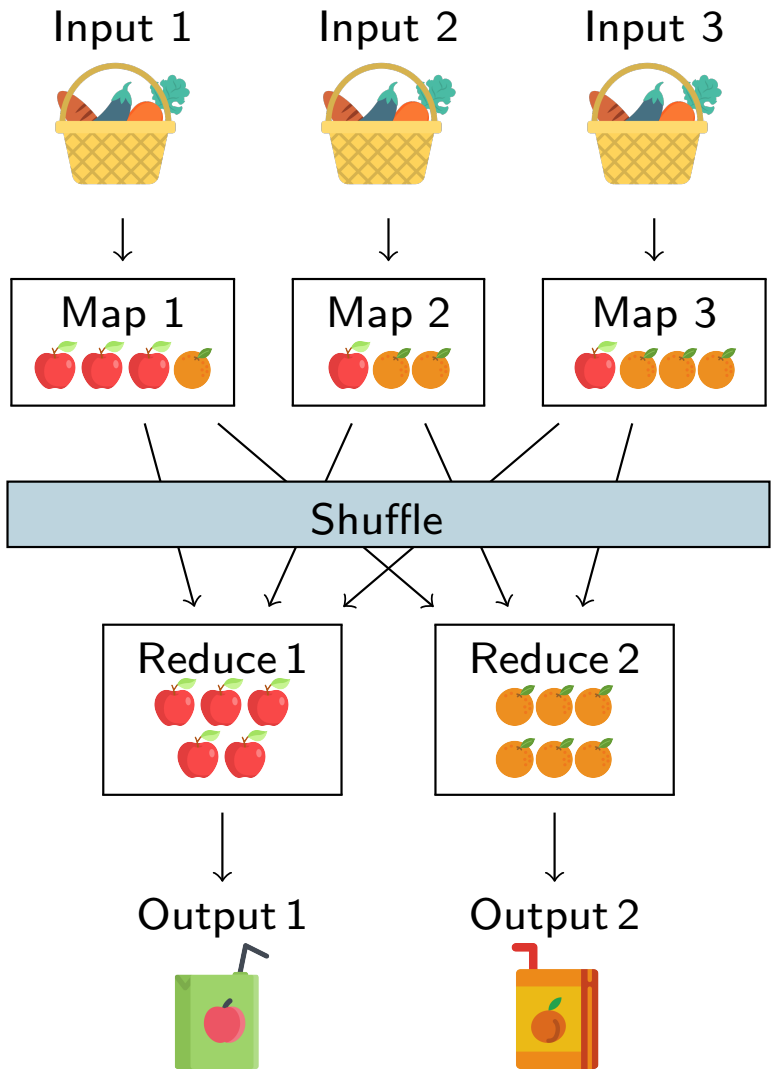
Specify

- Map and Reduce functions
- Input files

LIMOS LABORATOIRE D'INFORMATIQUE, DE MODÉLISATION ET D'OPTIMISATION DES SYSTÈMES

[1] J. Dean and S. Ghemawat. *MapReduce: Simplified Data Processing on Large Clusters*. In the proceedings of OSDI 2004.

MapReduce Example



MapReduce in 3 Steps

1. Map tasks

Input: ID of chunk
Output: *key-value* pairs

2. Master Controller

- Key-value pairs aggregated and sorted by key
- Pairs with same key sent to the same Reduce task

3. Reduce tasks

Input: One key
Output: Combine values associated to the key

Outline

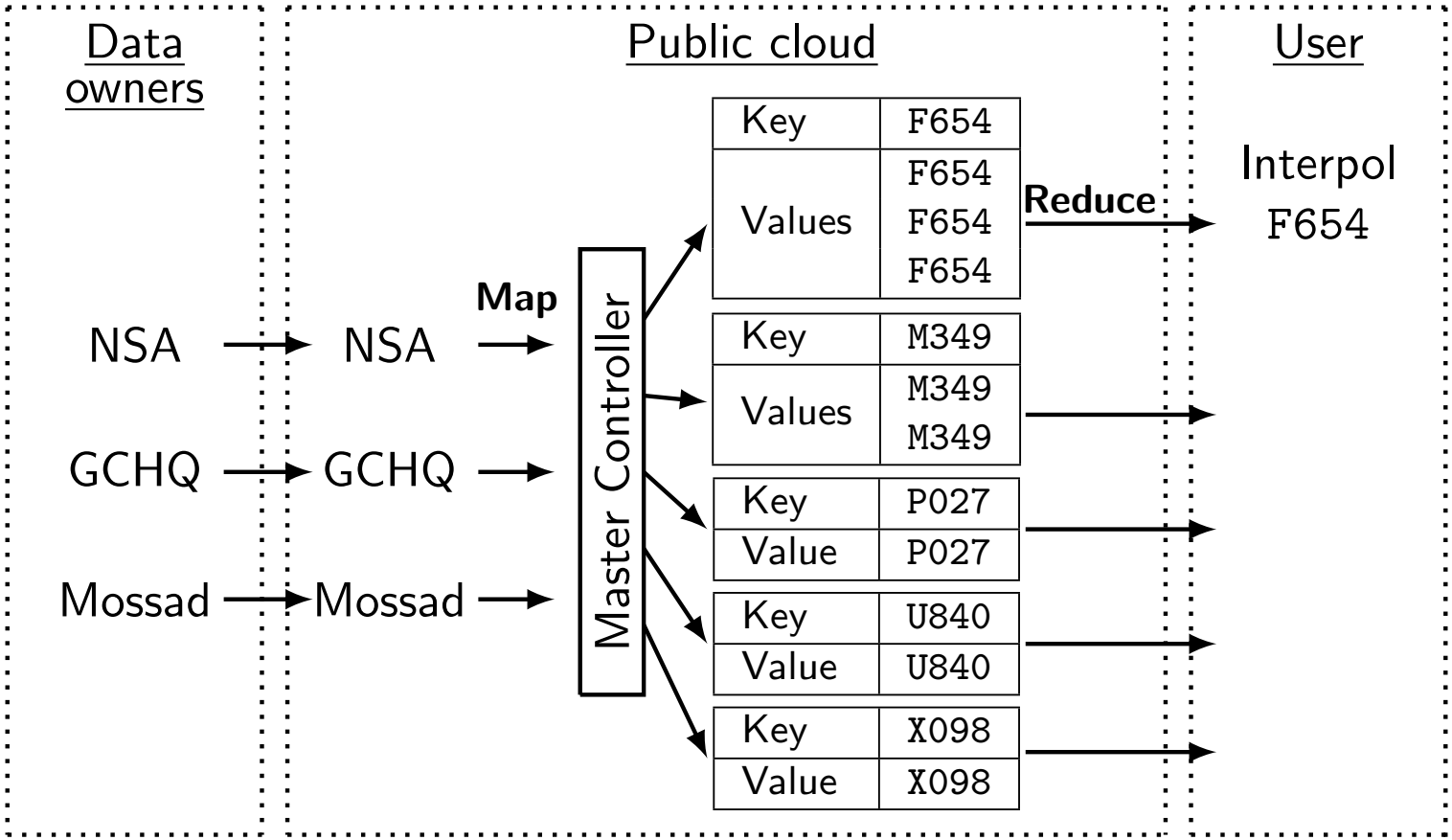
Intersection with MapReduce[2]

3 participants

NSA	GCHQ	Mossad
F654	F654	F654
U840	M349	M349
X098	P027	U840

[2] J. Leskovec, A. Rajaraman and J. D. Ullman. *Mining of Massive Datasets*. Cambridge University Press.

Intersection with MapReduce



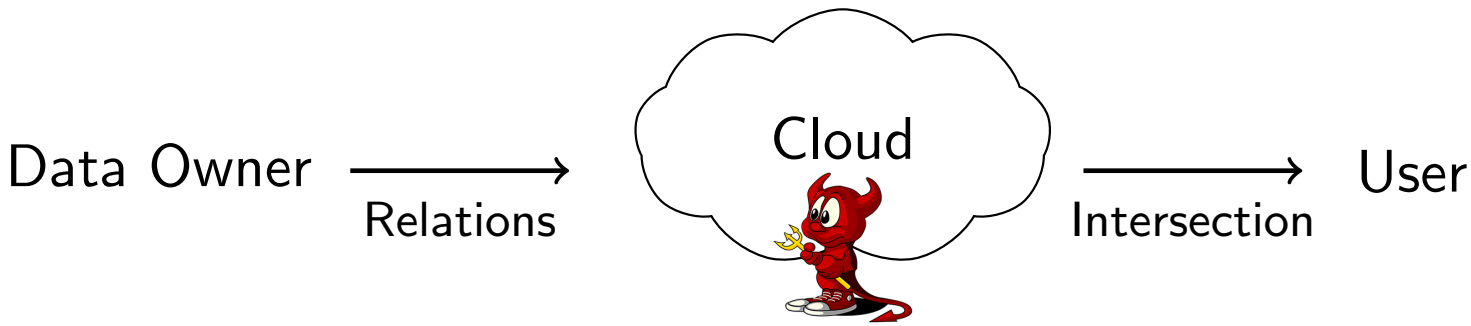
Reduce function

It returns value only if: #values = #participants

Outline

Security Model

Cloud is **honest-but-curious**



Without security, Cloud learns:

- Content of relations
- Intersection result

Cryptographic Tools

Pseudorandom function

$$f : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$$

- Deterministic
- Indistinguishable from a random function

Notation

$$[m]_k = f(k, m)$$

Cryptographic Tools

Asymmetric encryption scheme

- $(pk, sk) \leftarrow \mathcal{G}(\lambda)$
- $c \leftarrow \mathcal{E}(pk, m)$
- $m \leftarrow \mathcal{D}(sk, c)$

$$\mathcal{D}(sk, \mathcal{E}(pk, m)) = m$$

Notation

$$\{m\} = \mathcal{E}(pk, m)$$

Outline

Secure Intersection with MapReduce

Setting

- n relations: $R_1, \ldots, R_n$
- R_1 has: $k_1, \ldots, k_n$ PRF secret keys, and pk
- R_i (for $2 \leq i \leq n$) has: k_1 and k_i

Preprocessing

- One *main* relation using the public key of the final user
  For each element x, compute the key-value pair:

$$\left([x]_{k_1}, \left(\{x\} \oplus \left(\oplus_{i=2}^{i=n} [x]_{k_i}\right)\right)\right)$$

- Other relation compute the key-value pair:

$$([x]_{k_1}, [x]_{k_i})$$

Secure Intersection with MapReduce

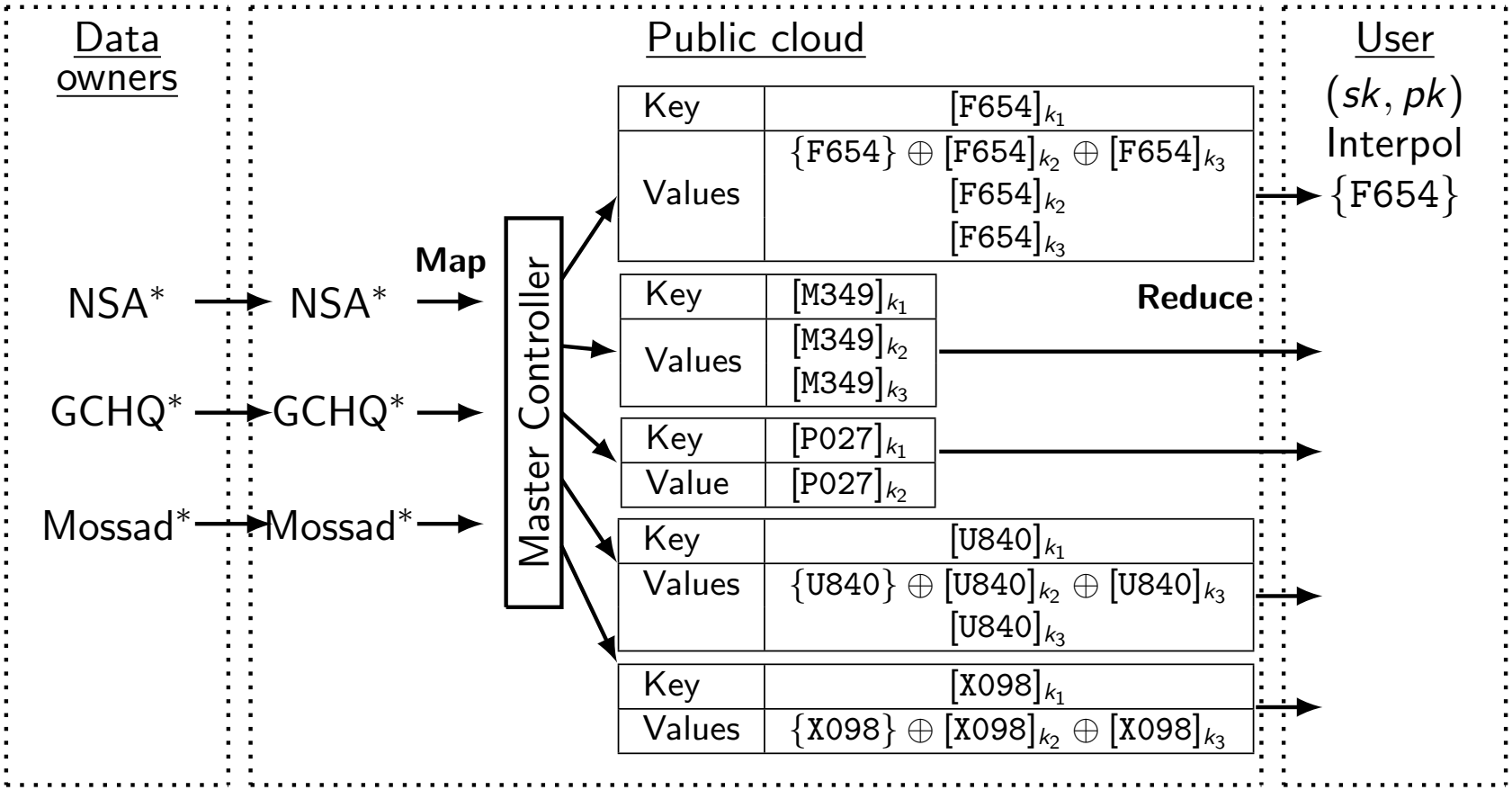
Processed relations

NSA*
$([\texttt{F654}]_{k_1}, (\{\texttt{F654}\} \oplus [\texttt{F654}]_{k_2} \oplus [\texttt{F654}]_{k_3}))$
$([\texttt{U840}]_{k_1}, (\{\texttt{U840}\} \oplus [\texttt{U840}]_{k_2} \oplus [\texttt{U840}]_{k_3}))$
$([\texttt{X098}]_{k_1}, (\{\texttt{X098}\} \oplus [\texttt{X098}]_{k_2} \oplus [\texttt{X098}]_{k_3}))$

GCHQ*
$([\texttt{F654}]_{k_1}, [\texttt{F654}]_{k_2})$
$([\texttt{M349}]_{k_1}, [\texttt{M349}]_{k_2})$
$([\texttt{P027}]_{k_1}, [\texttt{P027}]_{k_2})$

Mossad*
$([\texttt{F654}]_{k_1}, [\texttt{F654}]_{k_3})$
$([\texttt{M349}]_{k_1}, [\texttt{M349}]_{k_3})$
$([\texttt{U840}]_{k_1}, [\texttt{U840}]_{k_3})$

Secure Intersection with MapReduce



Outline

Experimental Results

Settings

-  3.2.0 / Standalone mode / Streaming
- ubuntu 16.04 LTS
- Map and Reduce functions in GO

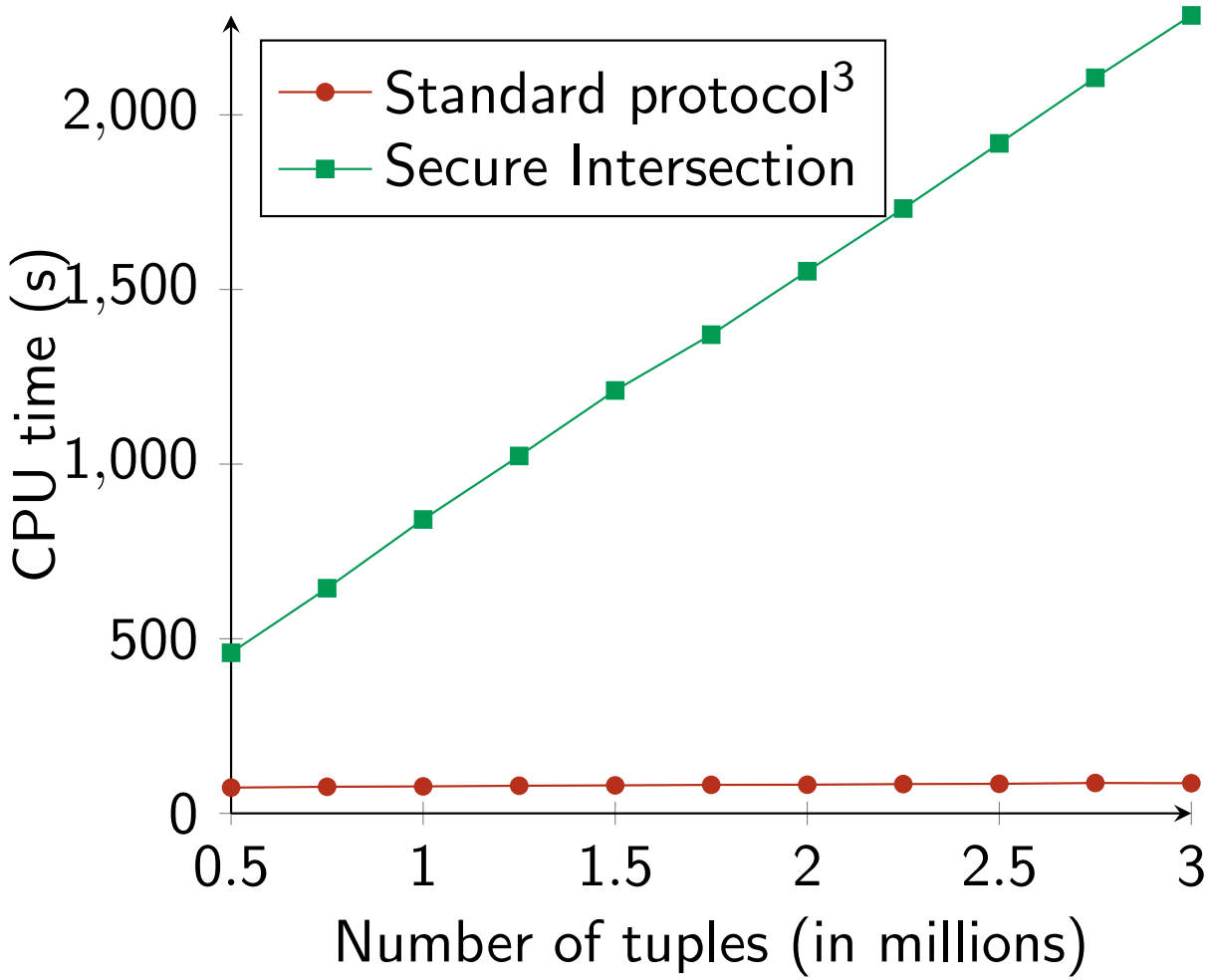
Hardware

- 4 CPU @ 2.4 GHz
- 80 Gb of disk
- 8 Gb of RAM

Experiments

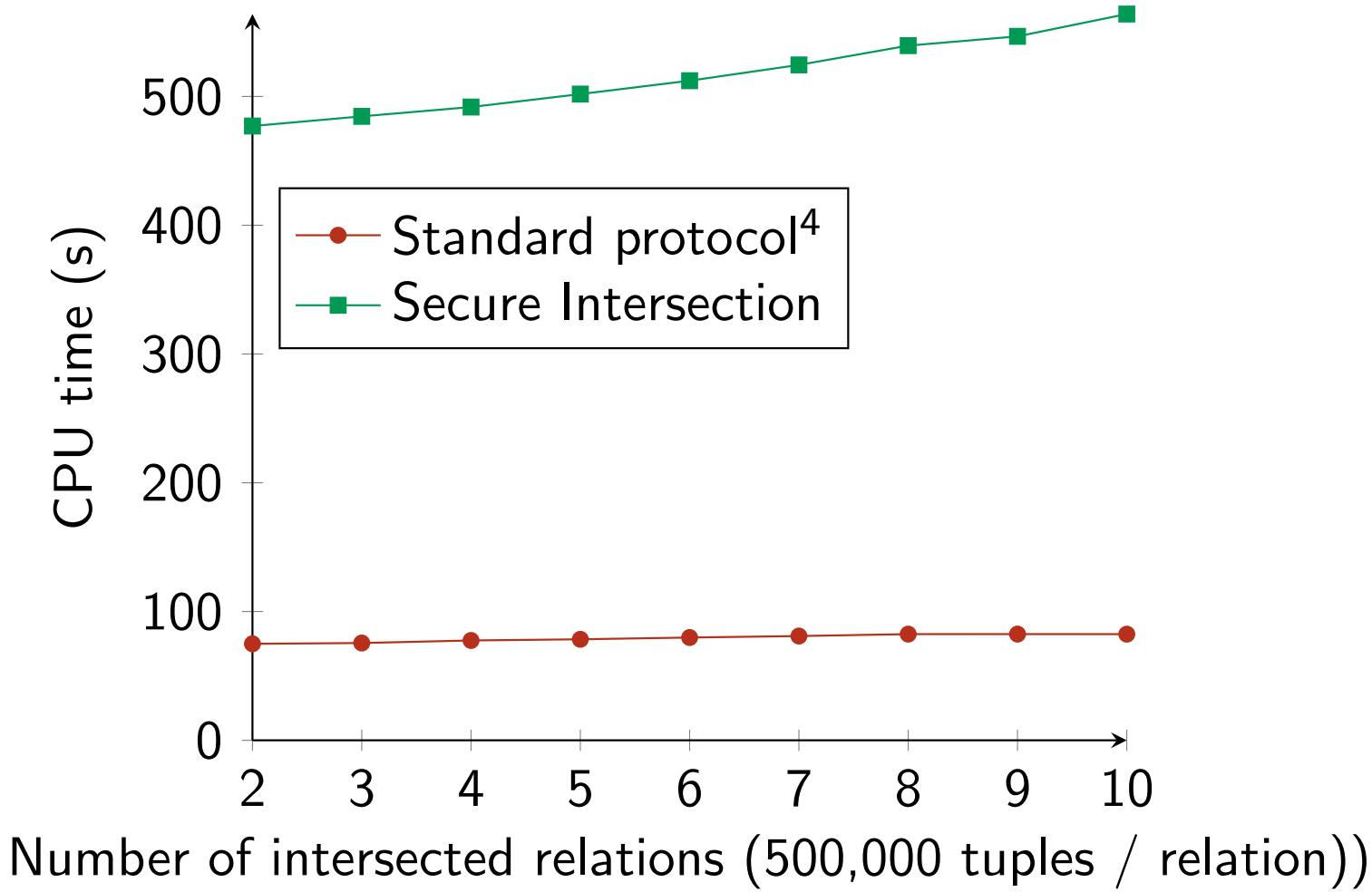
1. Varying the number of tuples
2. Varying the number of intersected relations

Results: Varying the Number of Tuples



[3] J. Leskovec, A. Rajaraman and J. D. Ullman. *Mining of Massive Datasets*. Cambridge University Press.

Results: Varying the Number of Intersected Relations



[4] J. Leskovec, A. Rajaraman and J. D. Ullman. *Mining of Massive Datasets*. Cambridge University Press.

Outline

Conclusion and Future Works

Conclusion

- Design of secure intersection with MapReduce
- Collision resistance
- Practical scalability

Future Works

- Apache Spark environment
- Malicious model

Thank you for your attention.

Any questions?

pascal.lafourcade@uca.fr

LIMOS
LABORATOIRE D'INFORMATIQUE,
DE MODÉLISATION ET D'OPTIMISATION DES SYSTÈMES