Restructuring the NSA Metadata Program

Seny Kamara

Microsoft Research

Thanks to: Timothy Edgar, Matt Green, Noah Kunin, Payman Mohassel, Kurt Rohloff, Chris Soghoian and Marcy Wheeler

TOP SECRET//SI//NOFORN

UNITED STATES

FOREIGN INTELLIGENCE SURVEILLANCE COURT

WASHINGTON, D.C.

IN RE APPLICATION OF THE
FEDERAL BUREAU OF INVESTIGATION
FOR AN ORDER REQUIRING THE
PRODUCTION OF TANGIBLE THINGS
FROM VERIZON BUSINESS NETWORK SERVICES,
INC. ON BEHALF OF MCI COMMUNICATION
SERVICES, INC. D/B/A VERIZON
BUSINESS SERVICES.

Docket Number: BR 13-80

SECONDARY ORDER

This Court having found that the Application of the Federal Bureau of Investigation (FBI) for an Order requiring the production of tangible things from **Verizon Business Network Services, Inc. on behalf of MCI Communication Services Inc., d/b/a Verizon Business Services (individually and collectively "Verizon")** satisfies the requirements of 50 U.S.C. § 1861,

IT IS HEREBY ORDERED that, the Custodian of Records shall produce to the National Security Agency (NSA) upon service of this Order, and continue production

TOP SECRET//SI//NOFORN

Derived from: Pleadings in the above-captioned docket
Declassify on: 12 April 2038

June 5th, 2013

1st Snowden document published

Verizon Court Order

Top secret court order

Compels Verizon to give NSA metadata of *every*

- US to Foreign call
- US to US call
- Foreign to US call

On a daily basis!

Similar arrangement with Sprint and AT&T


NOBAMA
PRISM
THOSE WHO SACRIFICE FREEDOM 4 SECURITY DESERVE NEITHER
TRANSPARENTER STAAT STATT GLÄSERNER BÜRGER!
change.org/prism

Why the Outrage?

Most Americans believed

- NSA could only spy on foreigners
- A warrant was required to access someone's data

The meta-data program

- Includes US-to-US calls
- NSA gets *everyone's* meta data with a *single* court order
- Order provided by a *secret* court

Is the Metadata Program Legal?

Is it Constitutional?

4th Amendment

Gov. cannot search your home without a warrant

1967

Supreme court says 4th Amendment protects *people*

Whenever they have a "*reasonable expectation of privacy*"

1970's

3rd Party Doctrine

Metadata not protected by 4th Amendment

Customers have no "reasonable expectation of privacy" about metadata

Is it Consistent with FISA/Patriot Act?

Sec. 501 of Foreign Intelligence Surveillance Act (FISA)

Amended by Sec. 215 of PATRIOT Act

Says a provider can be compelled to hand over data

> "if there are reasonable grounds to believe that the tangible things sought are ***relevant*** to an authorized investigation"

The FISA court interpreted "**relevant**" so as to include *every* record

"… I believe *we need a new approach*. I am therefore ordering a transition that will end the Section 215 bulk metadata program as it currently exists and establish a mechanism that *preserves the capabilities we need without the government holding this bulk metadata*."

"I have instructed the intelligence community … to develop options for *a new approach* that can match the capabilities and fill the gaps that the Section 215 program was designed to address, *without the government holding this metadata itself*."

January 17th, 2014

Obama speech on NSA reform

How do we design such a system ?

Outline

Motivation

MetaDB (current NSA system)

- How does it work?
- Security analysis

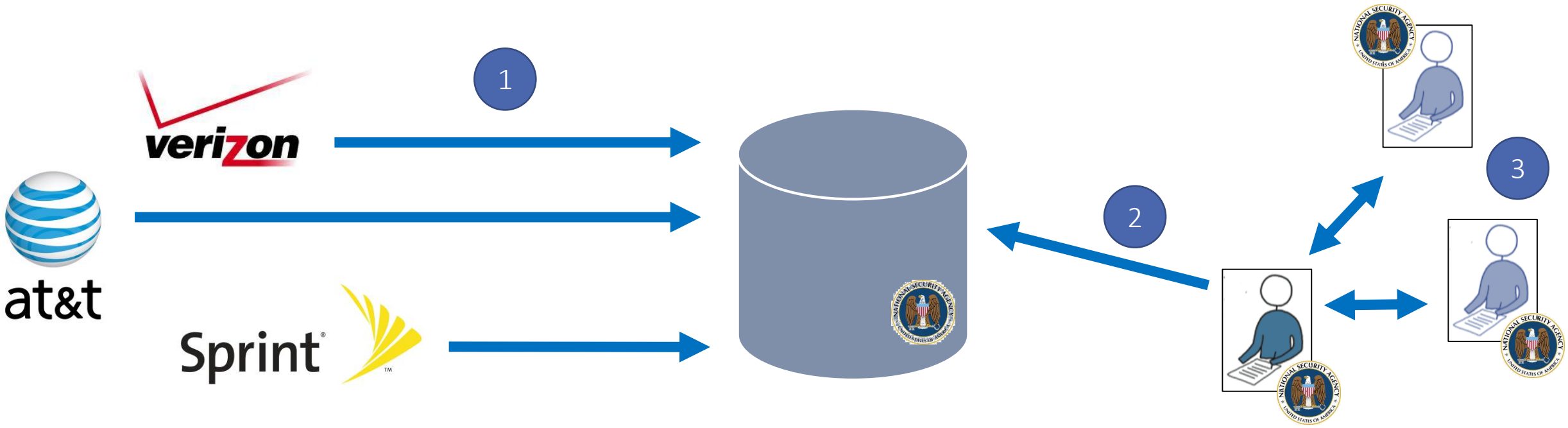
Possible Solutions

- The OB protocol
- The IARPA protocols

MetaCrypt

- Secure multi-party computation
- Structured encryption

How Does MetaDB Work?



1. To & from numbers, time of call, duration for all US-to-US, US-to-Foreign and Foreign-to-US calls
2. MDB can only be queried by individual phone number (seed)
3. Analyst queries must be approved by small number of NSA officials

Functionality of MetaDB

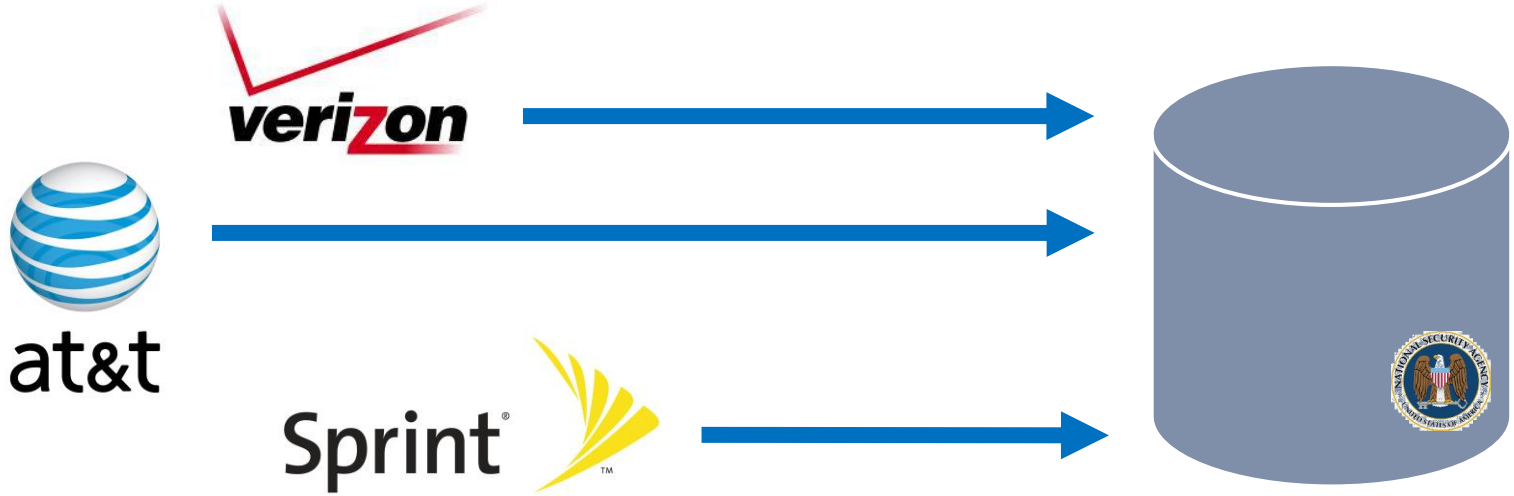
Includes data from (at least) 3 parties

Supports 3-hop queries

- reduced to 2 hops by Obama
- Hops include incoming & outgoing calls

Holds data for at least 5 years

- Data deleted after that



Security Mechanisms of MetaDB

Few analysts can query MetaDB

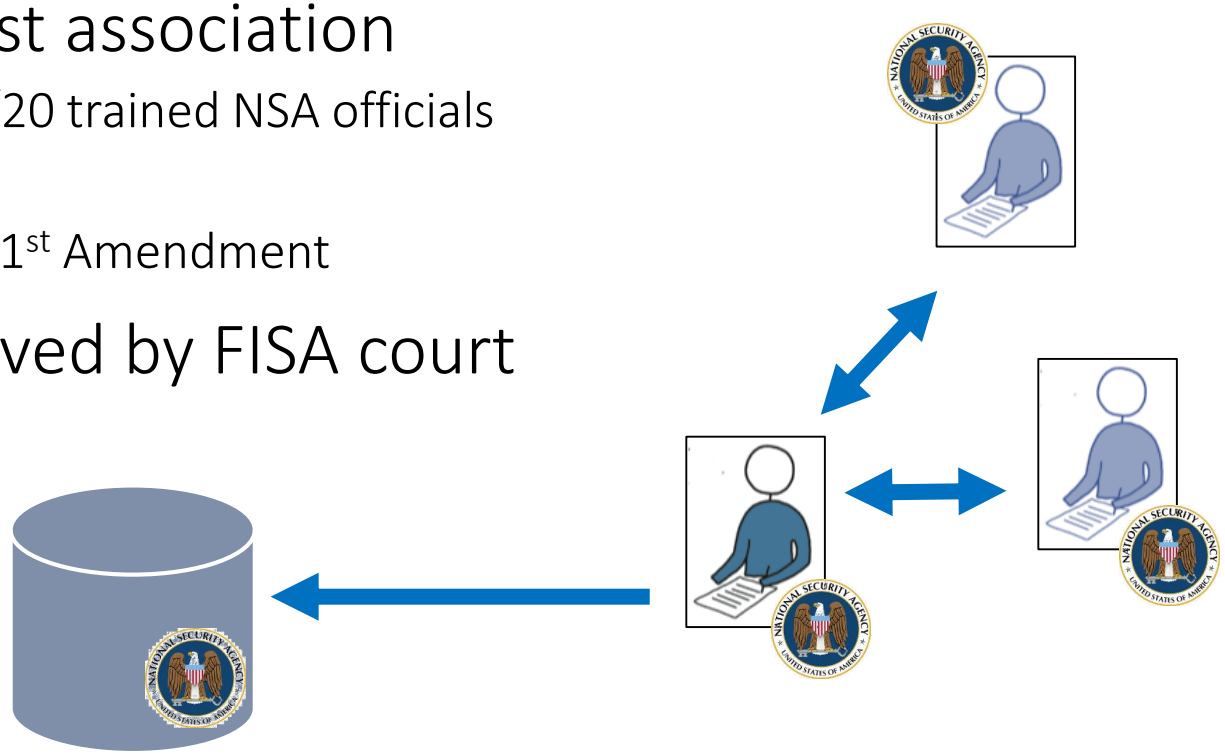
- Each one receives “appropriate & adequate” training
- Only for foreign intelligence information

Seed has to be *suspected* of terrorist association

- Suspicion decided independently by at least 2/20 trained NSA officials
- Approved by 1/2 trained NSA supervisors
- Suspicion not based on activities protected by 1st Amendment

List of terrorist organizations approved by FISA court

Access is logged and audited



What Security Properties do We Want?

Isolation

MetaDB should be protected from outsiders

Query Certification

Only certified queries can be executed

Data privacy

Analysts learn at most query response

Query privacy

Telcos learn nothing about NSA queries

Security Analysis of MetaDB

Let’s assume (best-case)

- Process is enforced at the system level
- e.g., supervisors use credentials to certify seed query, etc…

Security of current design relies on following assumptions

- Isolation under secure systems assumption
- Query cert. under secure systems assumption & non-collusion b/w analysts & supervisors
- Data privacy under secure systems assumption
- Query privacy without assumptions

Can we do better ?

Options Under Consideration

Office of Director of National Intelligence & Justice Department

Discontinue program completely

- Not going to happen…

Non-NSA government agency holds MetaDB (e.g., FBI…)

- Who?

Private 3rd-party holds MetaDB

- Who? Would be filling a government function with less oversight

Telcos hold data

- Telcos do not want to hold data
- Liability, cost, bad PR, …

A Modest Proposal [Kamara13]

"Are Privacy and Compliance Always at Odds" from Outsourcedbits.org

Solution with following properties

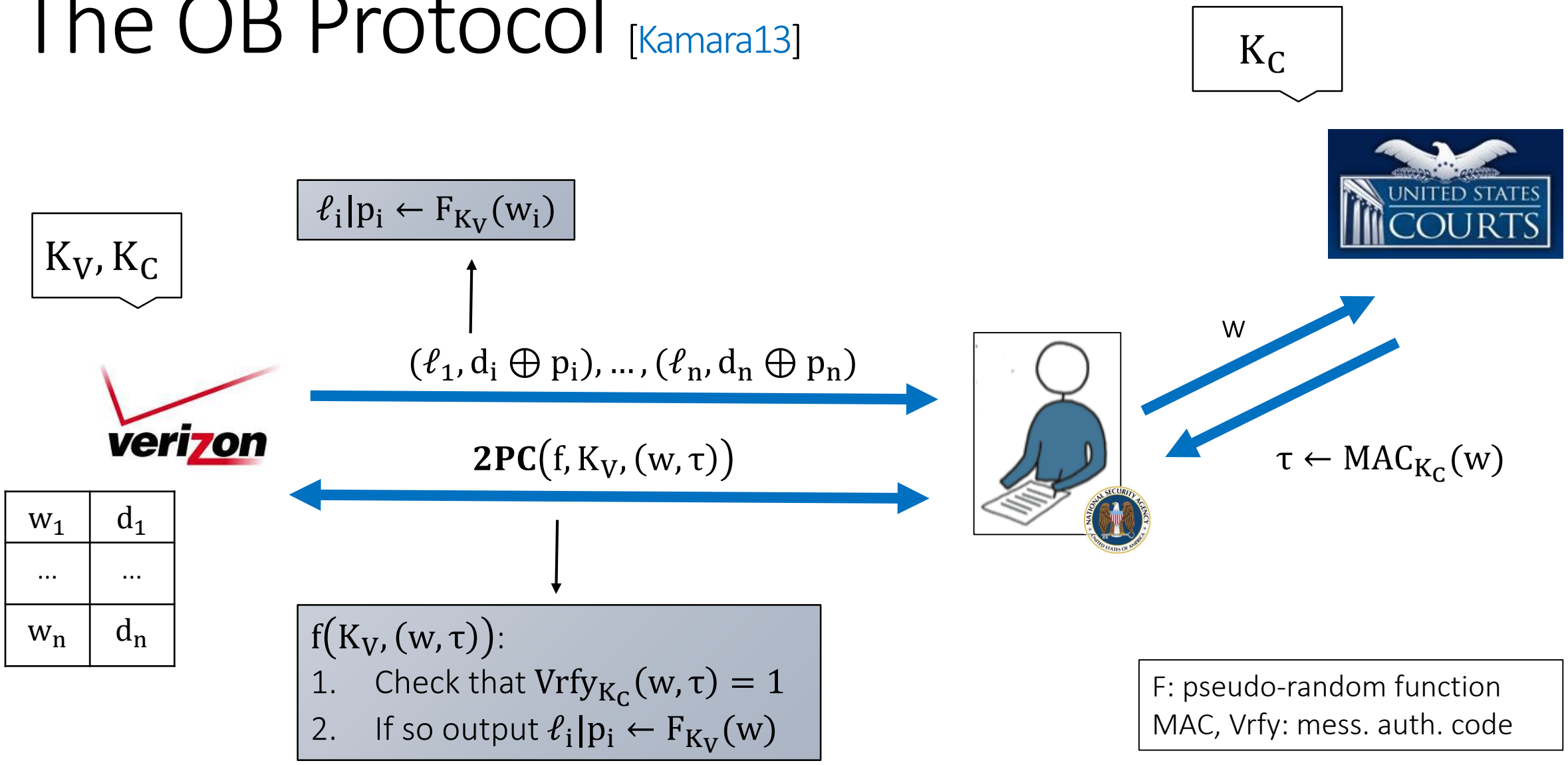
- Isolation
- Data privacy
- Certified queries
- Query privacy

} Existence of symmetric-key encryption, public-key encryption and pseudo-random functions

Design based on combination of

- Keyword OT [Freedman-Ishai-Pinkas-Reingold05]
- Secure two-party computation [Yao82]
- Message authentication codes (MACs)

The OB Protocol [Kamara13]



IARPA

Intelligence Advanced Research Projects Activity

"invests in high-risk, high-payoff research programs that have the potential to provide the United States with an overwhelming intelligence advantage over future adversaries"

Security and Privacy Assurance Research (SPAR)

Started in 2011

Program manager: Konrad Vesey

Two teams: IBM Research & Columbia University

[Cash-Jarecki-Jutla-Krawczyk-Rosu-Steiner13]

[Jarecki-Jutla-Krawczyk-Rosu-Steiner14]

[Cash-Jarecki-Jutla-Krawczyk-Rosu-Steiner14]

[Krell-Pappas-Vo-Choi-Bellovin-Keromitis-Kolenikov-Malkin14]

"efficient cryptographic protocols for querying a database that keep the query confidential, yet still allow the database owner to determine if the query is authorized and, if so, return only those records that match it"

Outsourced Symmetric PIR [JJKRS14]

[Jarecki-Jutla-Krawczyk-Rosu-Steiner14]

- Based on [...,Cash-JJKRS13,CJJKRS14]
- Similar (*at a very high-level*) to OB protocol
- Much more challenging due to support for Boolean queries!
- Uses Oblivious PRFs and homomorphic signatures

Security

- Isolation
- Data privacy
- Certified queries
- Query privacy

} Existence of random oracles, one-more gap Diffie-Hellman groups, symmetric-key encryption, authenticated encryption

Can We Use OB or OSPIR ?

OB & OSPIR rely on following assumptions

- OB relies on standard crypto assumptions ☺☺
- OSPIR relies on reasonable crypto assumptions ☺
- Crypto can be securely implemented ☺
- Keys can be protected ☺

Functionality ☹

- OB & OSPIR are encrypted *text* databases that support *keyword* search
- MetaDB is a *graph* database that supports *2-hop neighbor queries*!

Certification ☹☹

- OB & OSPIR support only basic query certification
- OB query certification by single human party
- OSPIR query certification by “format” (full version will include certification by single “human” party)
- MetaDB requires certification by multiple (human) parties

A New Design: MetaCrypt

The MetaCrypt Protocol

N+6 parties

- N Telcos
- 1 server which can be *an untrusted cloud*!
- 2 NSA analysts, 2 NSA supervisors, 1 NSA party

Two phases

- Store phase between Telcos & server
- Query phase between Telcos & NSA parties

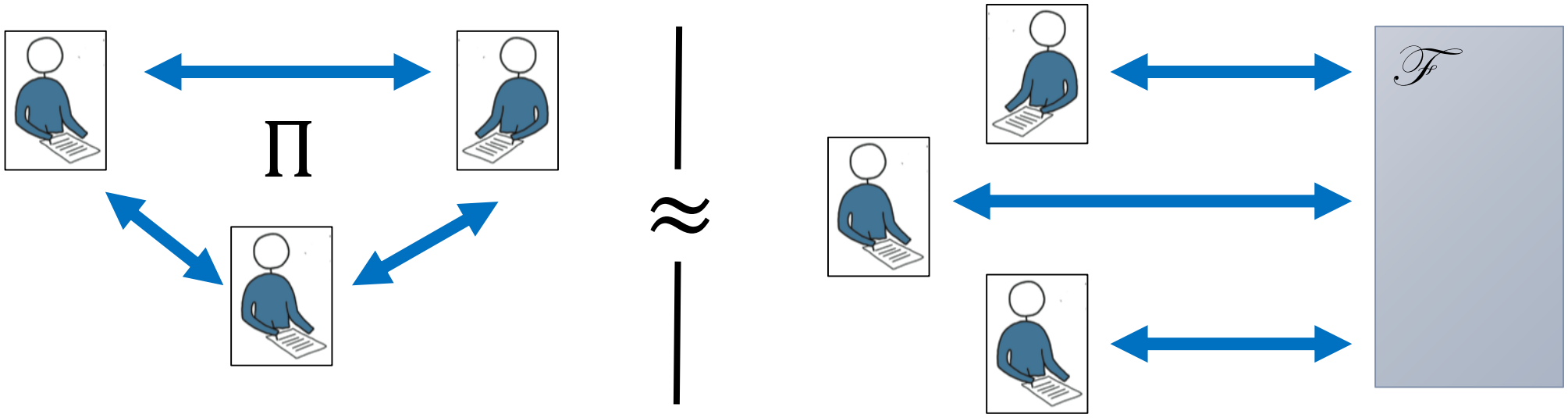
Formalizing Security Goals of MetaCrypt

Ideal/real-world paradigm [..., Canetti01]

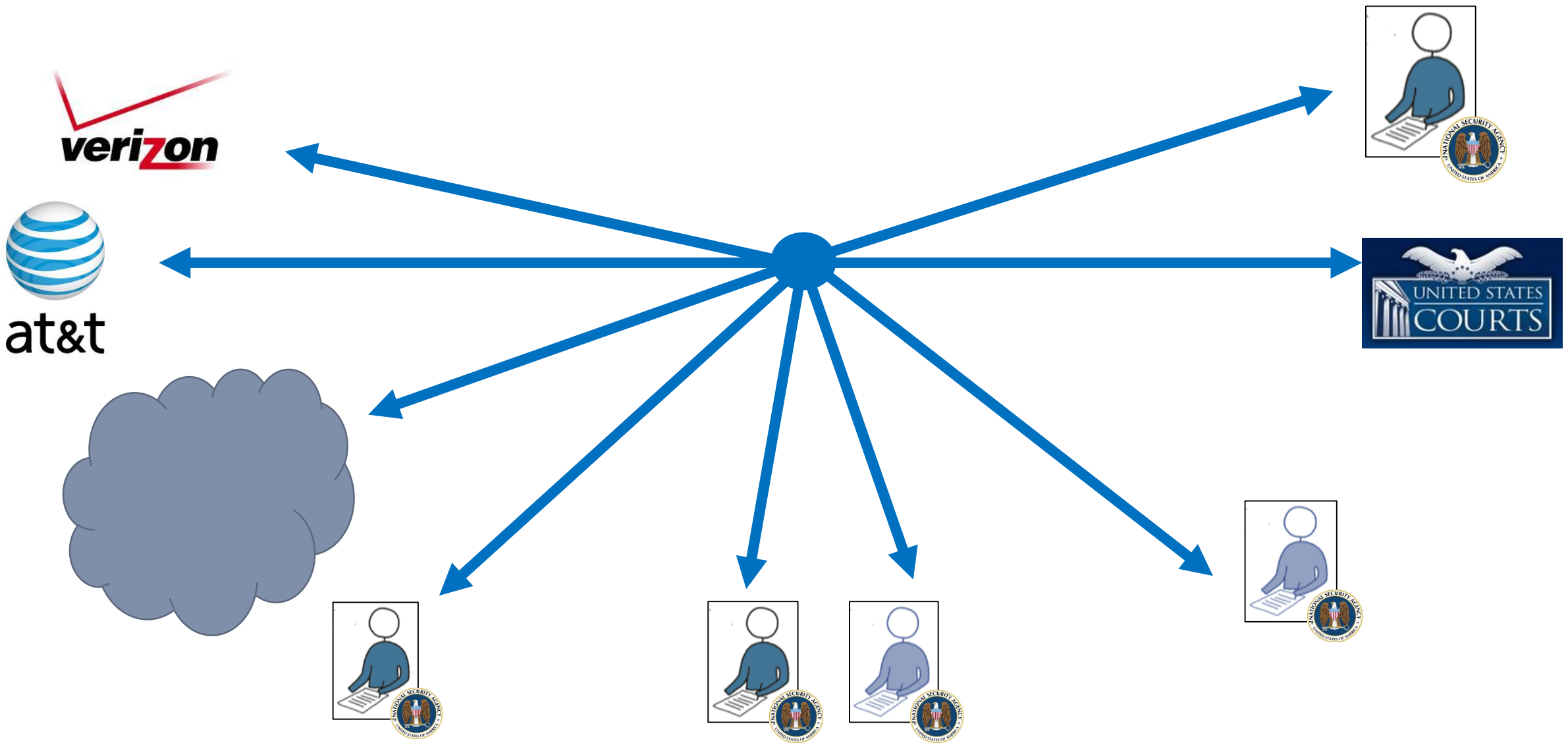
- Secure multi-party computation type definition

Indistinguishability of two worlds

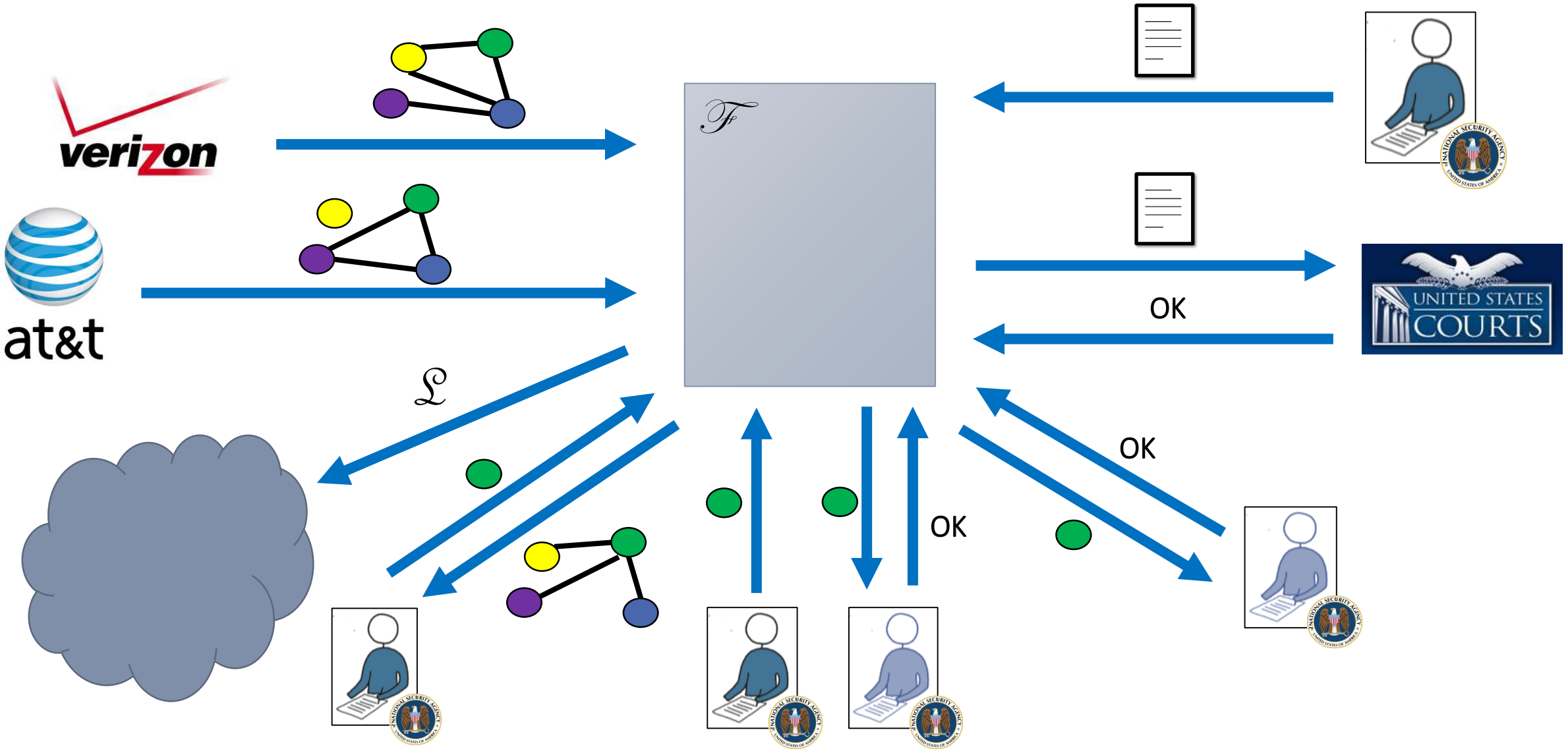
- In real-world parties execute protocol Π
- In ideal-world parties interact with ideal functionality $\mathcal{F}$
- If real-world execution is indistinguishable from ideal-world then Π is secure



Formalizing Security Goals of MetaCrypt



Formalizing Security Goals of MetaCrypt



MetaCrypt Building Blocks

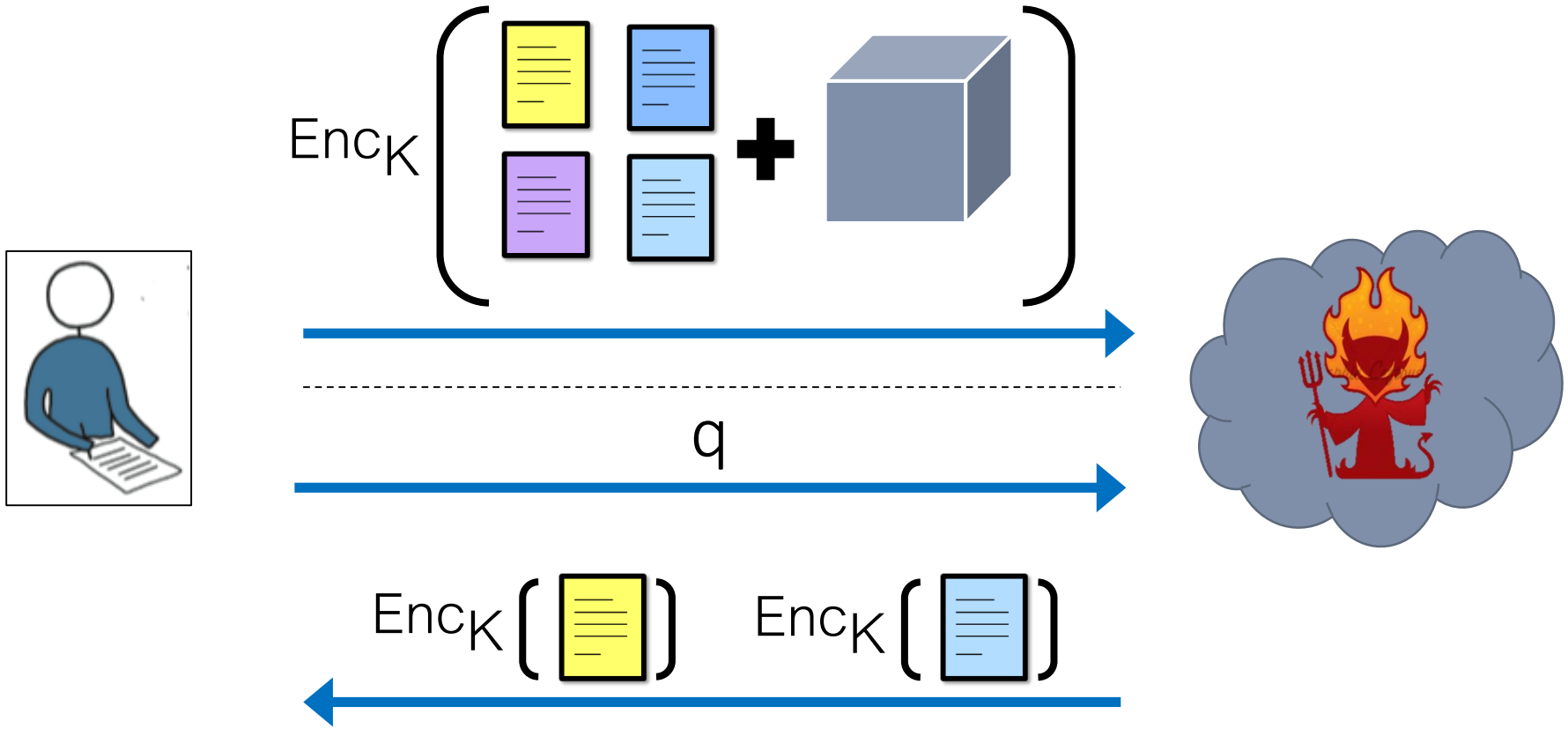
Structured encryption [Chase-Kamara10]

- New graph encryption scheme with support for 2-hop neighbor queries
- Combination of two graph encryptions with support for 1-hop neighbor queries

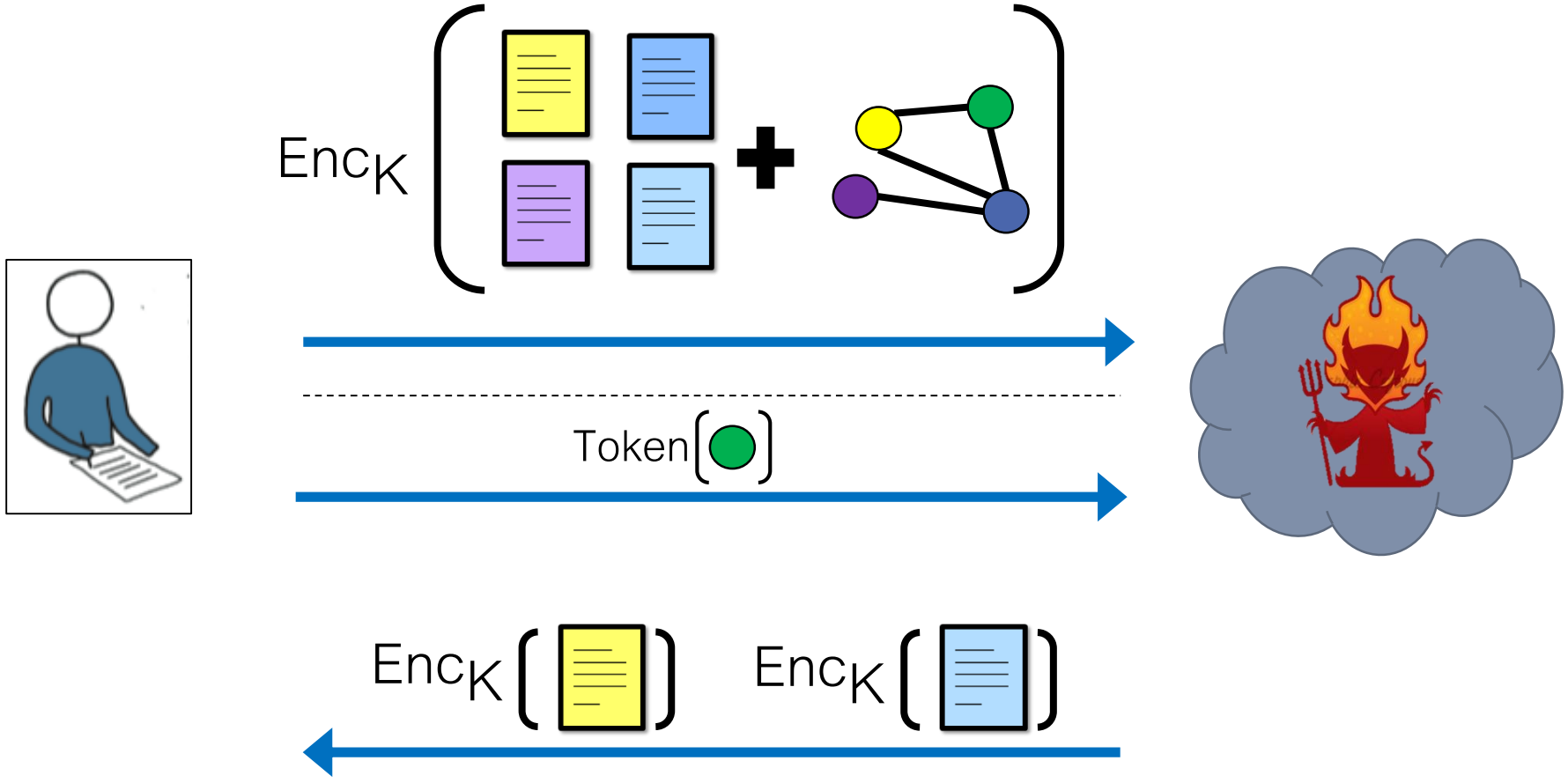
Secure multi-party computation [Yao82,Goldreich-Micali-Wigderson87]

- N telcos, 2 NSA analysts, 2 NSA supervisors, 1 NSA party

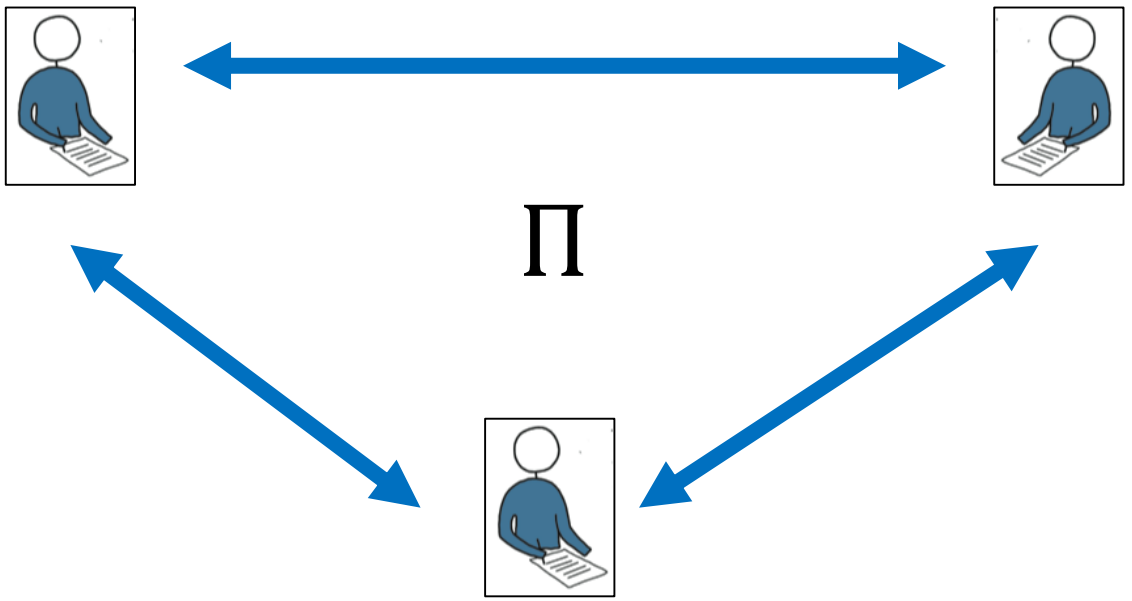
Structured Encryption [Chase-Kamara10]



Graph Encryption [Chase-Kamara10]



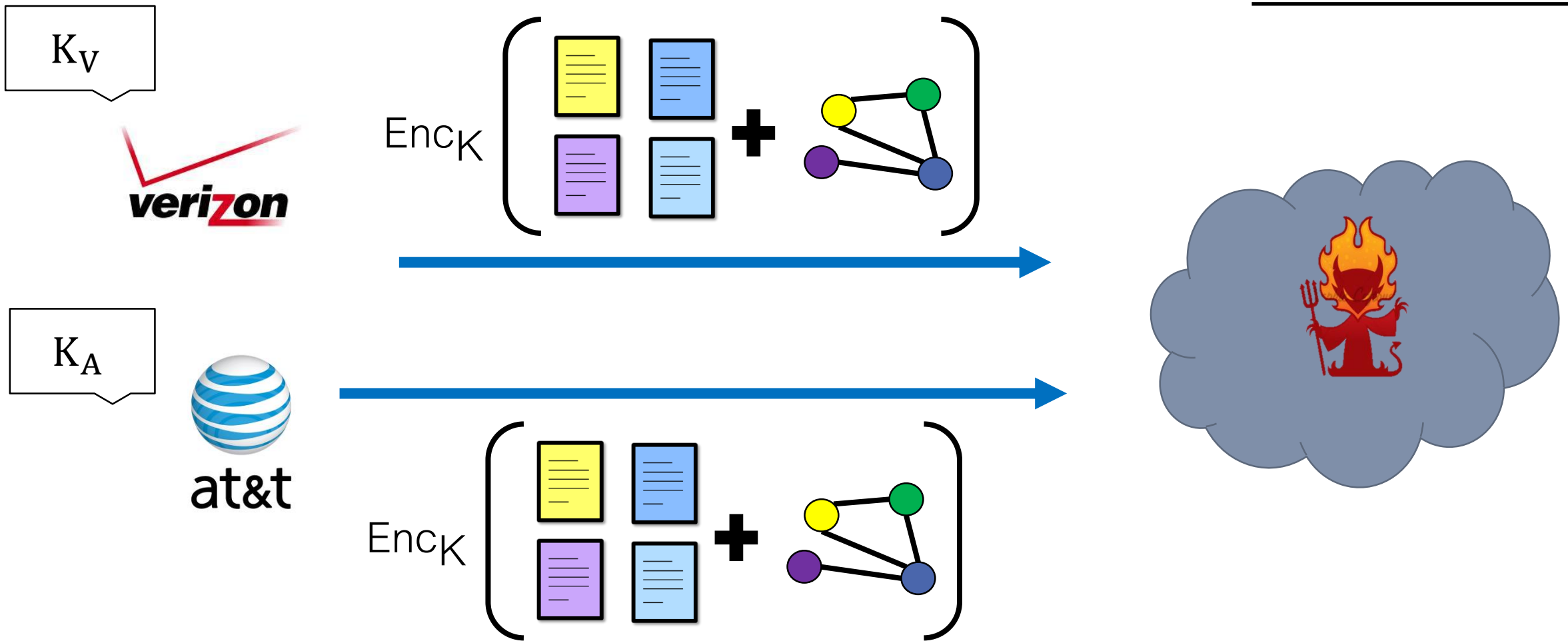
Secure Multi-Party Computation [Yao82,GMW87]



- Allows N parties to compute privately
  - The parties learn only their prescribed output
  - Nothing about other parties' inputs
  - Except what they can infer from their output
- Computation can be any arbitrary function
- Result is guaranteed to be correct
  - Else parties abort

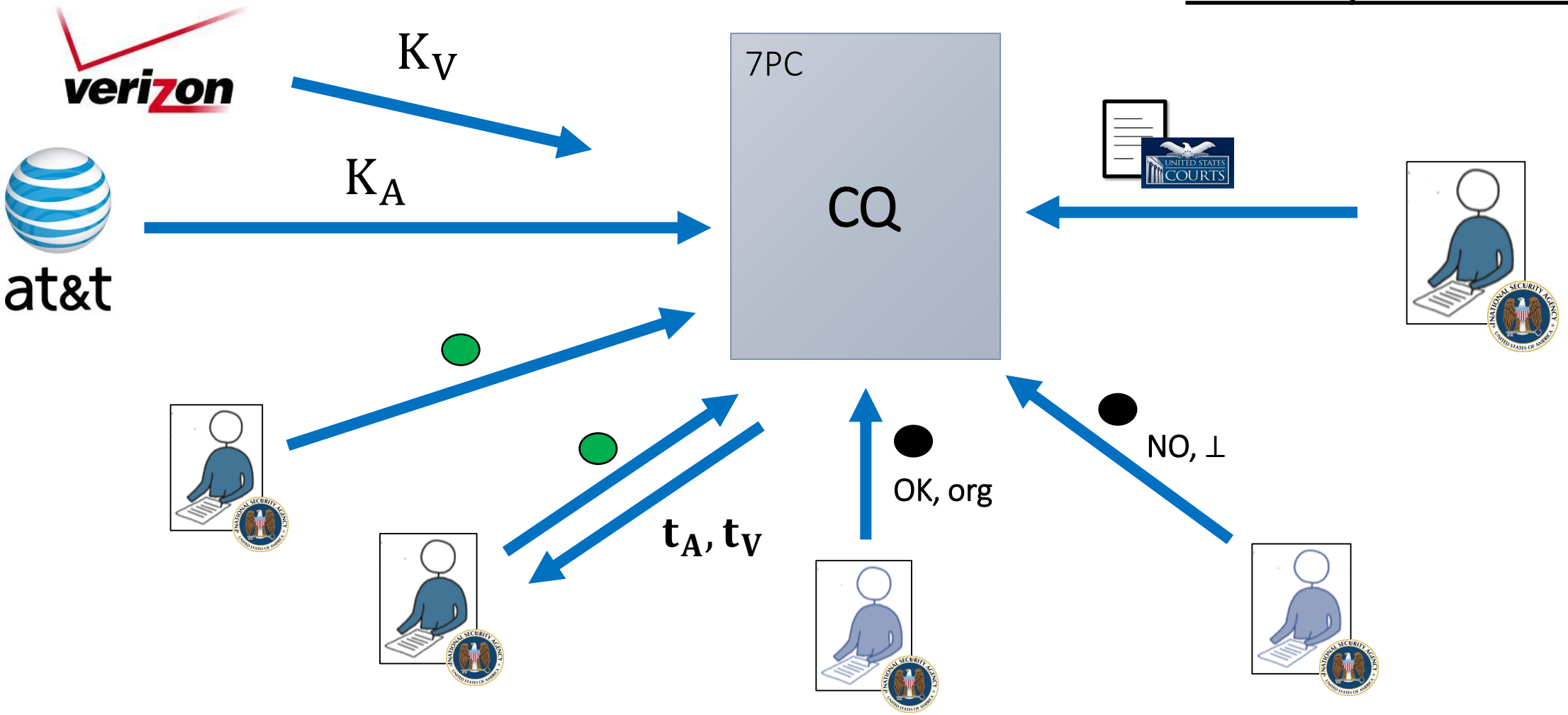
The MetaCrypt Protocol

Store Phase



The MetaCrypt Protocol

Query Phase #1



The Certification Functionality

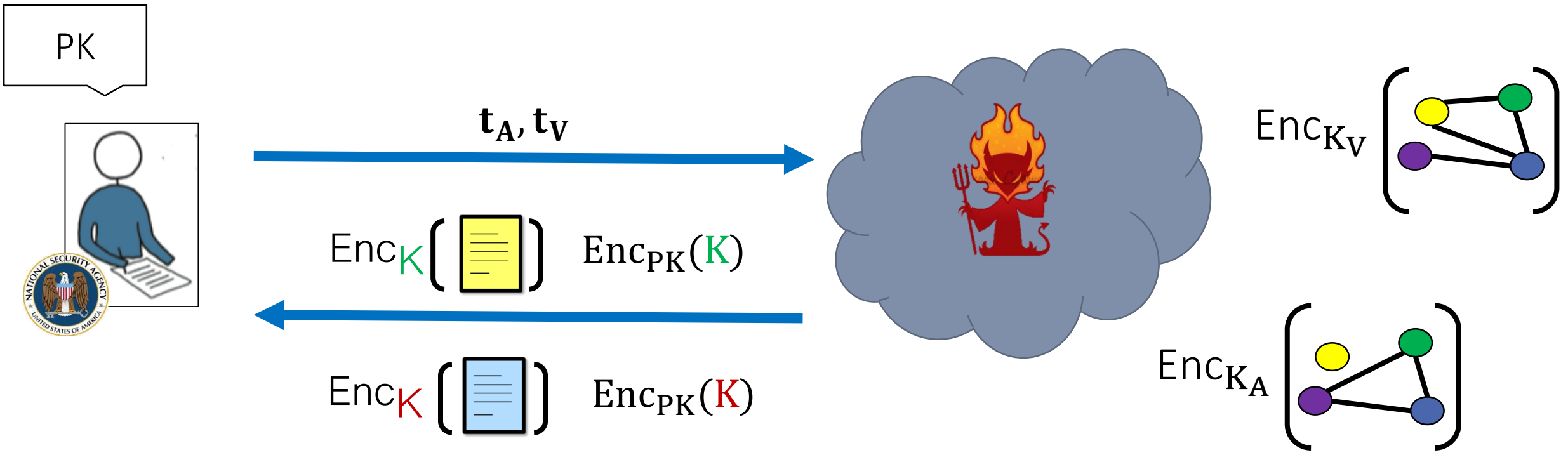
7PC

$$\mathsf{CQ}(\mathrm{K_V}, \mathrm{K_A}, \mathrm{q_1}, \mathrm{q_2}, (\mathrm{q_3}, \mathrm{m_3}, \mathrm{org_3}), (\mathrm{q_4}, \mathrm{m_4}, \mathrm{org_4}), (\mathrm{TL}, \sigma))$$

if $q_1 \neq q_2$ abort;
if $\mathrm{Vrfy}(\mathrm{TL}, \sigma) = \mathrm{false}$ abort;
if $(\mathrm{m_3} = \mathrm{NO} \wedge \mathrm{m_4} = \mathrm{NO})$ abort;
if $(\mathrm{q_i} \neq \mathrm{q_1} \mathrm{Vorg_i} \notin \mathrm{TL})$ abort, where i is accepting SV;
Output to Analyst

$\mathrm{t_A} \leftarrow \mathrm{Token_{K_A}}(\mathrm{q_1})$ and $\mathrm{t_V} \leftarrow \mathrm{Token_{K_V}}(\mathrm{q_1})$

The MetaCrypt Protocol



Query Phase #2

The MetaCrypt Protocol

Underlying 2-hop graph encryption scheme

- Too complex to describe here
- Can be built from symmetric-key encryption, public-key encryption & pseudo-random permutations
- Combines two instances of a construction from [Chase-Kamara10]
- Will appear in the paper

Thouhgts

Motivation

If metadata program is preserved we need

- A **privacy-preserving** solution
- That is computationally-**efficient** at **scale**
- With security & privacy based on **weak** assumptions

The original MetaDB design does not achieve this

The solutions being considered by White House do not achieve this

As crypto & security researchers it is our responsibility to work on this

Roadmap

Need to understand NSA requirements & procedures

- ex: understanding basic process pointed to limitations of OB & OSPIR protocols
- Graph vs. text DBs, complex query certification vs. naïve single-party certification

Need to understand the scale of the data

Need to design more protocols

- More efficient
- Better functionality
- Stronger security definitions
- Weaker assumptions
- Etc…

Need to implement systems to improve designs

Limitations

The problem cannot be addressed by crypto alone!

- Crypto is only a tiny piece of the puzzle

A comprehensive solution requires ideas from

- Policy, software security, systems security, traffic analysis, data mining, databases, …

What's the ETA?

MetaCrypt is a first pass

But based on efficient building blocks

- Secure multi-party computation (with ≈ 8 parties)
- Graph encryption
- Question is: how far will they scale?

Still lots of room for

- More efficient protocol designs
- Low-level crypto optimizations
- Hardware optimizations
- Systems optimizations

Paper coming soon!

The End