Global cyber surveillance: how to defend against?

A sample of a few cryptographic techniques

Giuseppe Bianchi

May 6, 2021

What we will learn today?

➔Perfect forward Secrecy

⇨Must-have post-Snowden requirement

➔Certificate transparency

⇨blockchain-type data structures for public verification

➔«trivial» Secure Multiparty Computation

⇨How to compute over encrypted data

mass surveillance: who?

List of government mass surveillance projects

From Wikipedia, the free encyclopedia

Main article: Mass surveillance

This list is incomplete; you can help by adding missing items with reliable sources.

This is a **list of known government surveillance projects** and related databases throughout the world.

Contents [hide]

1 International
1.1 European Union
2 National
2.1 Australia
2.2 China
2.3 France
2.4 Germany
2.5 India
2.6 Russia
2.7 Sweden
2.8 Switzerland
2.9 United Kingdom
2.10 United States
3 Unclear origin
4 Recently discontinued
5 See also
6 References

mass surveillance: who?

WIKIPEDIA
The Free Encyclopedia

Main page
Contents

Australia [edit]

Main article: Mass

- In August 2014 it
without a warrant.
- It was reported[3]

China [edit]

Main article: Mass

- **Golden Shield P**
division of the gov
- The **Integrated J**
gathers biometric
- **Monitoring Bure**
- **Public Informati**
- **Social Credit Sy**

France [edit]

- **Frenchelon**: A da

List of government mass surveillance projects

Russia [edit]

- **SORM**: A technical system used by the Federal Security Service of the Russian Federation to monitor internet and telephone communication.
- **Yarovaya Law** is a piece of anti-terrorist legislation that includes a requirement to store all phone call and text messaging data, as well as providing cryptographic backdoors for security services.

Sweden [edit]

- **Titan traffic database**: A database established by the Swedish National Defence Radio Establishment (Swedish: Försvarets radioanstalt, FRA) where call detail records (CDRs) of telephony and internet traffic and transaction data (IPDRs) concerning international telecommunications are stored.[13]
- **X-Keyscore**: A system used by the United States National Security Agency for searching and analysing internet data about foreign nationals. FRA has been granted access to the program.[14]

Switzerland [edit]

- **Onyx**: A data gathering system maintained by several Swiss intelligence agencies to monitor military and civilian communications, such as e-mails, telefax and telephone calls. In 2001, Onyx received its second nomination for the ironically-named "Big Brother Award".[15]

United Kingdom [edit]

Further information: Mass surveillance in the United Kingdom

- **Impact Nominal Index**: The Impact Nominal Index or INI is a computer system that enables the UK police force to establish whether other relevant authorities are holding information regarding a person of interest.[16]
- **Interception Modernisation Programme**: An initiative to extend the UK government's capability to lawfully intercept and store communications data in a central database.[17]
- **Mastering the Internet** (MTI): A clandestine mass surveillance program led by the British intelligence agency GCHQ. Data gathered by the GCHQ include the contents of email messages, entries on the social networking platform Facebook and the web browsing history of internet users.[18]
- **UK National DNA Database** (NDNAD): It is also the oldest national DNA database in the world.[19] Since its establishment in 1995, the database has grown to include DNA samples from 2.7 million individuals, or 5.2% of the UK's population, many of whom have neither been charged with, or convicted of, any offence.[19]
- **Tempora**: Launched in the autumn of 2011, this initiative allows the GCHQ to set up a large-scale buffer that is capable of storing internet content for 3 days and

Remember Snowden's revelations, 2013

WIKIPEDIA
The Free Encyclopedia

Main page
Contents

Australia [edit]

Main article: Mass

- In August 2014 it
without a warrant.
- It was reported[3]

China [edit]

Main article: Mass

- **Golden Shield P**
division of the gov
- The Integrated J
gathers biometric
- **Monitoring Bure**
- **Public Informati**
- **Social Credit Sys**

France [edit]

- **Frenchelon**: A da

United States [edit]

Further information: Mass surveillance in the United States

- **Boundless Informant**: A system deployed by the National Security Agency to analyze global electronic information. In March 2013, Boundless Informant gathered 14 billion data reports from Iran, 6.3 billion from India, and 2.8 billion from the United States.[23]
- **BULLRUN**: a highly classified U.S. National Security Agency program to preserve its ability to eavesdrop on encrypted communications by influencing and weakening encryption standards, by obtaining master encryption keys, and by gaining access to data before or after it is encrypted either by agreement, by force of law, or by computer network exploitation (hacking).
- **Carnivore**: A system implemented by the Federal Bureau of Investigation that was designed to monitor email and electronic communications. Apparently replaced by commercial software such as NarusInsight.
- **Comprehensive National Cybersecurity Initiative**
- **DCSNet**: The Federal Bureau of Investigation (FBI)'s point-and-click surveillance system that can perform instant wiretaps on any telecommunications device located in the United States.[24]
- **Fairview**: A mass surveillance program directed at foreign mobile phone users.
- **Financial Crimes Enforcement Network**: A bureau of the Department of the Treasury that collects and analyzes financial transactions in order to combat financial crimes.
- **ICREACH**: Surveillance frontend GUI that is shared with 23 government agencies, including the CIA, DEA, and FBI, to search illegally collected personal records.
- **Magic Lantern**: A keystroke logging software deployed by the FBI in the form of an e-mail attachment. When activated, it acts as a trojan horse and allows the FBI to decrypt user communications.[25]
- **Main Core**: A personal and financial database storing information of millions of U.S. citizens believed to be threats to national security.[26] The data mostly comes from the NSA, FBI, CIA, as well as other government sources.[26]
- **MAINWAY**: NSA database containing metadata for hundreds of billions of telephone calls made through the four largest telephone carriers in the United States.

contents of email messages, entries on the social networking platform Facebook and the web browsing history of intern

- **UK National DNA Database** (NDNAD): It is also the oldest national DNA database in the world.[19] Since its establishm
include DNA samples from 2.7 million individuals, or 5.2% of the UK's population, many of whom have neither been cha
- **Tempora**: Launched in the autumn of 2011, this initiative allows the GCHQ to set up a large-scale buffer that is capable

National Security Agency surveillance

Map of global NSA data collection, with countries subject to the most data collection shown in red

Programs [hide]

Pre-1978
ECHELON · MINARET · SHAMROCK · PROMIS

Since 1978
Upstream collection · BLARNEY · FAIRVIEW · Main Core · ThinThread · Genoa

Since 1990
RAMPART-A

Since 2001
OAKSTAR · STORMBREW · Trailblazer · Turbulence · Genoa II · Total Information Awareness · President's Surveillance Program (Terrorist Surveillance Program)

Since 2007
PRISM · Dropmire · Stateroom · Bullrun · MYSTIC · MonsterMind (alleged)

Databases, tools etc.
PINWALE · MARINA · Main Core · MAINWAY · TRAFFICTHIEF · DISHFIRE · XKeyscore ·

And of course surveillance from big OTT as well!

"Surveillance is the business model of the Internet."

Everyone is under constant surveillance by many companies, ranging from social networks like Facebook to cellphone providers. This data is collected, compiled, analyzed, and used to try to sell us stuff. Personalized advertising is how these companies make money, and is why so much of the internet is free to users. **We're the product, not the customer.**

- **Bruce Schneier**

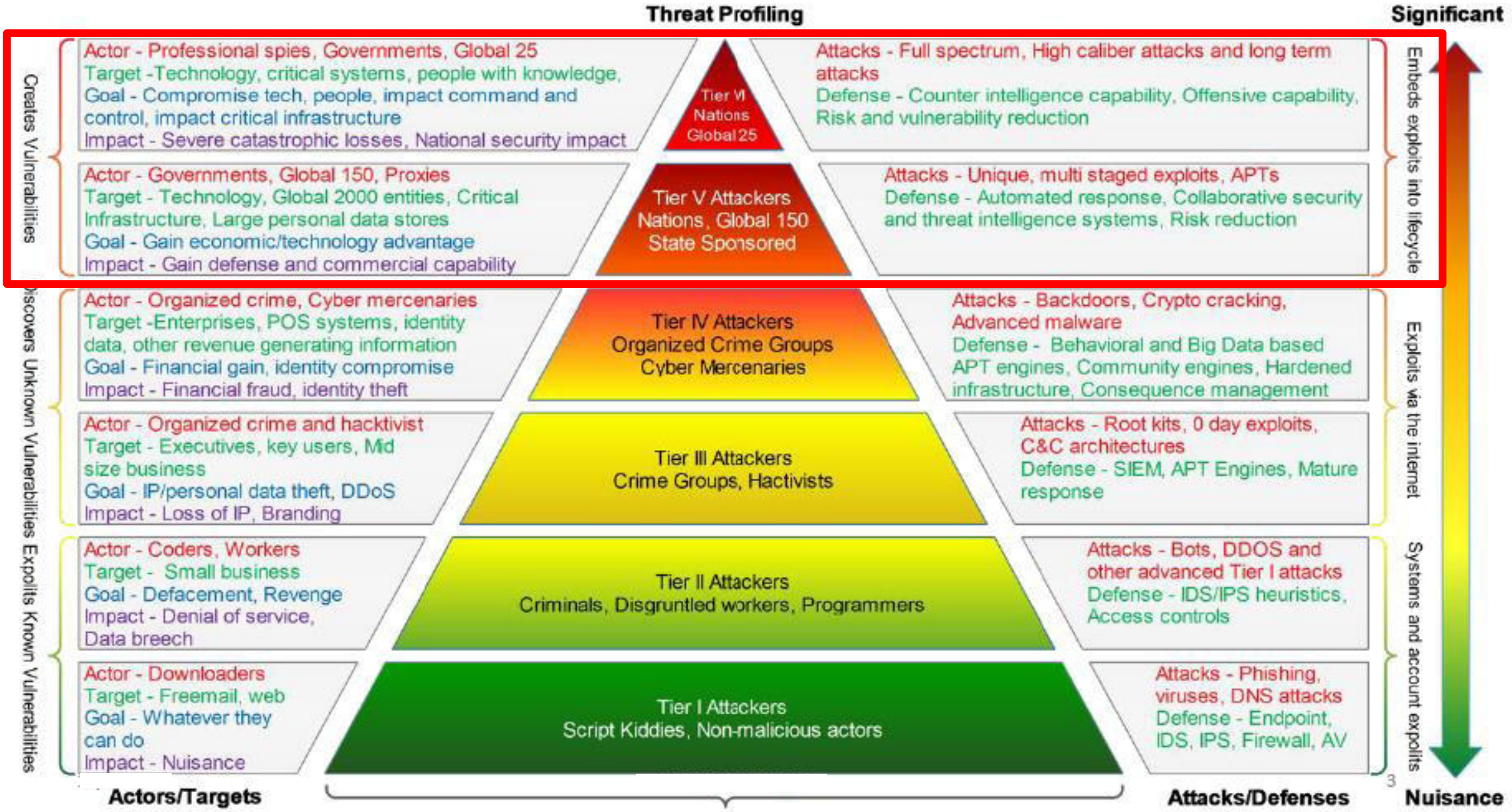
·privacy-be-very-afraid-analyst-suggests/

Mobile Handset Privacy: Measuring The Data iOS and Android Send to Apple And Google

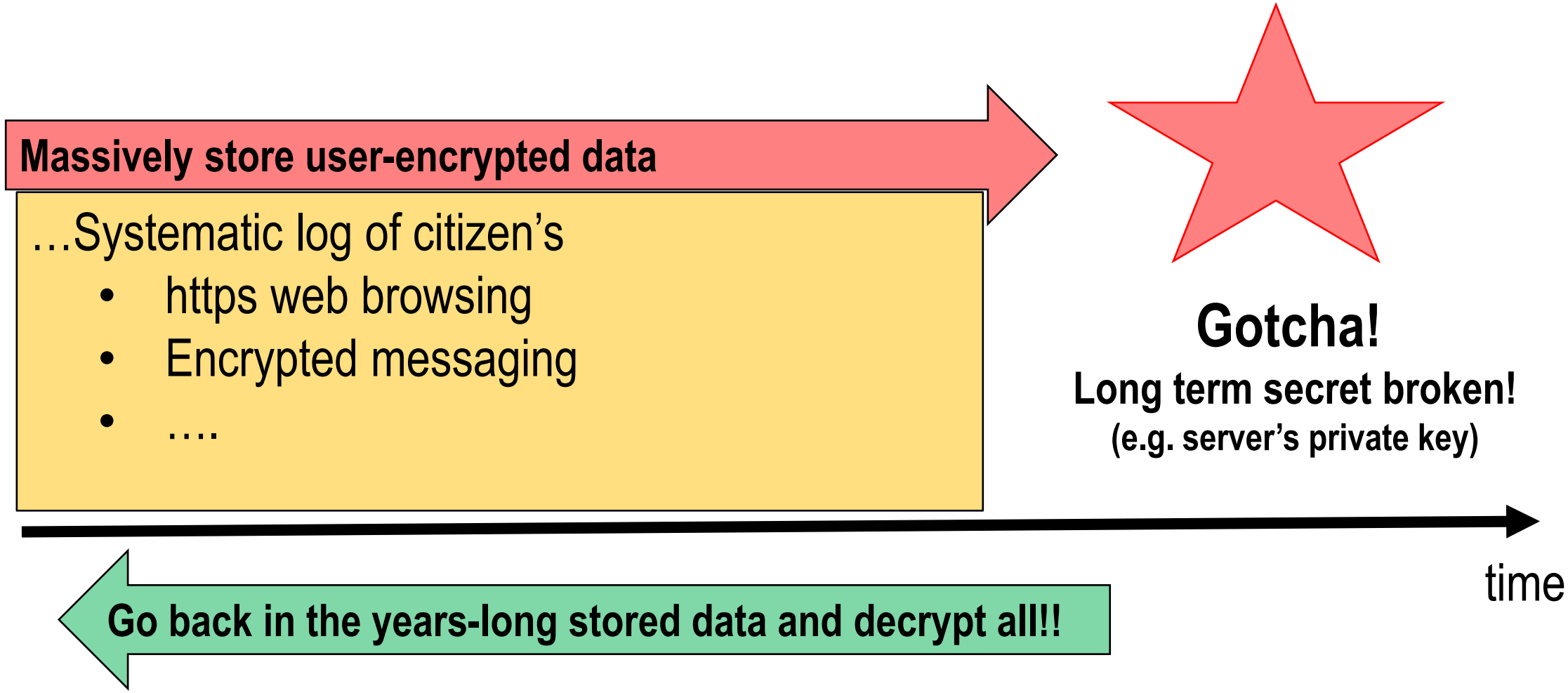
Douglas J. Leith
School of Computer Science & Statistics,
Trinity College Dublin, Ireland
25th March, 2021

***Abstract*—We investigate what data iOS on an iPhone shares with Apple and what data Google Android on a Pixel phone shares with Google. We find that even when minimally configured and the handset is idle both iOS and Google Android share data with Apple/Google on average every 4.5 mins. The phone IMEI, hardware serial number, SIM serial number and IMSI, handset phone number etc are shared with Apple and Google. Both iOS and Google Android transmit telemetry, despite the user explicitly opting out of this. When a SIM is inserted both iOS and Google Android send details to Apple/Google. iOS sends the MAC addresses of nearby devices, e.g. other handsets and the home gateway, to Apple together with their GPS location. Users have no opt out from this and currently there are few, if any, realistic options for preventing this data sharing.**

High tier adversary's model!



Time not an issue for «them»
a.k.a.: standard encryption is NOT enough



A must today: (Perfect) Forward Secrecy

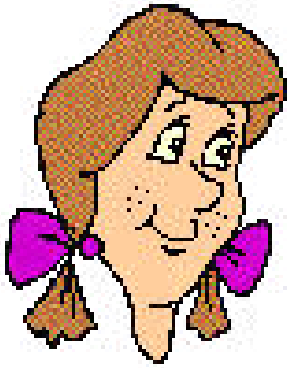
Informally:

if a long-term private key (e.g. of the server) is compromised at some time, this should NOT affect data delivered **BEFORE** such time!

More technically:

session keys will not be compromised even if long-term secrets used in the session key exchange are compromised

Traditional TLS: RSA key transport, no PFS



pk_{bank}

sk_{bank}

Hello, let's talk HTTPS (HTTP over TLS)

Here is my public key certificate: $\{Bank, pk_{bank}\}_{CA}$

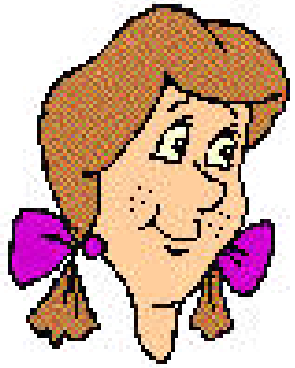
choose Random Pre-master secret K

And here is the RSA-encrypted secret key we'll use in our session: $RSA(K)_{pk_{bank}}$

Got it: K

Data exchange: encrypted with symmetric key K

Traditional TLS: RSA key transport, no PFS



choose Random Pre-master secret K

Here is

And here is the RSA

D

Session 12/12/2018: **RSA(k1)$_{pk}$**
(--- DATA ---)$_{k1}$
Session 08/01/2019: **RSA(k2)$_{pk}$**
(--- DATA ---)$_{k2}$
Session 11/05/2019: **RSA(k3)$_{pk}$**
(--- DATA ---)$_{k3}$
Session 09/01/2020: **RSA(k4)$_{pk}$**
(--- DATA ---)$_{k4}$
Session 12/07/2020: **RSA(k5)$_{pk}$**
(--- DATA ---)$_{k5}$
Session 01/04/2021: **RSA(k6)$_{pk}$**
(--- DATA ---)$_{k6}$

Breach 06/05/2021→ sk exposed

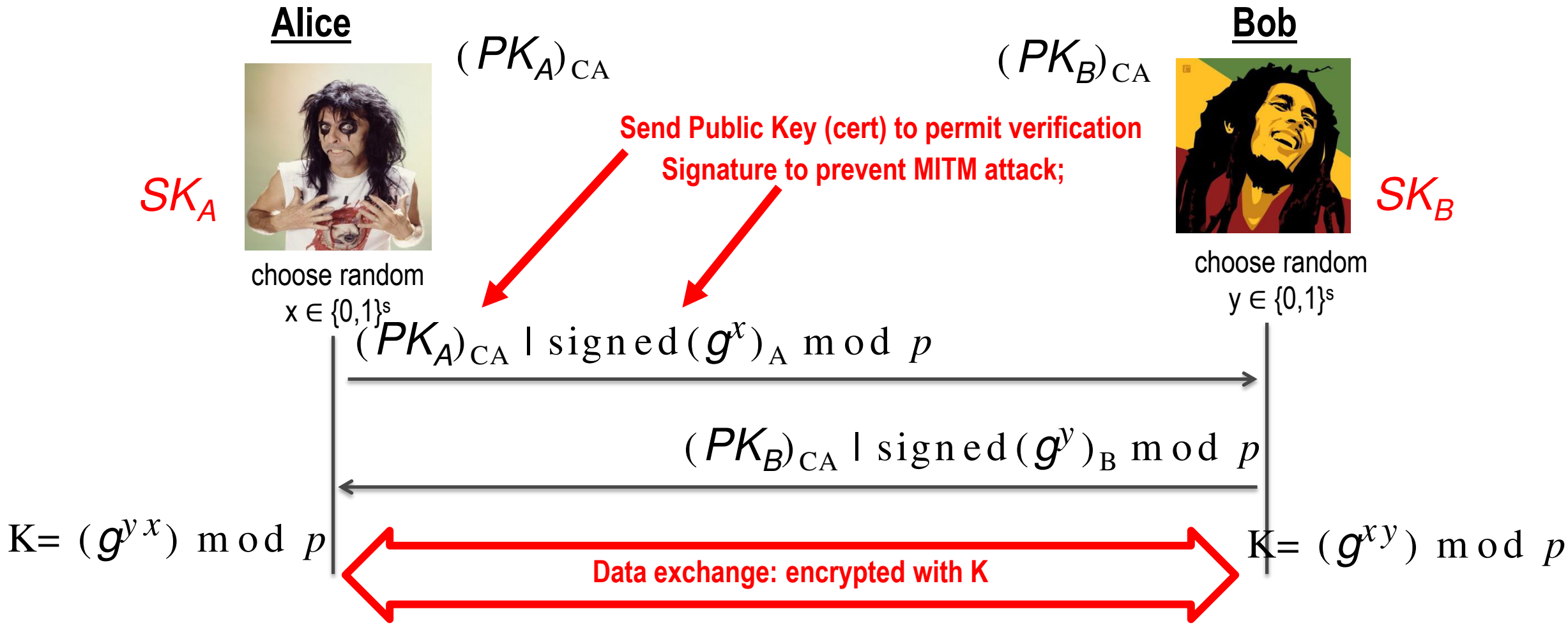
pk_{bank}

sk_{bank}

RSA(K)pk_{bank}

Got it: K

PFS - how to? Ephemeral Diffie-Hellman!



If SK breaks, (ephemeral) session keys are not revealed! Forward Secrecy!

Why you should upgrade to TLSv1.3? Perfect Forward Secrecy becomes mandatory!

➔**TLS1.2: four key handshake methods supported**

- ⇨RSA key transport (most common, no PFS))
- ⇨Diffie-Hellman Anonymous (vulnerable to MITM)
- ⇨Diffie-Hellman Fixed (no PFS)
- ⇨Diffie-Hellman Ephemeral

➔**TLS1.2: removed all, except one!**

- ⇨Diffie-Hellman Ephemeral
- ⇨Moreover, Authenticated Encryption mandatory as well
  - →To clear the original MAC-then-Encrypt flaw (e.g. POODLE attack)

What about PFS vs pre-shared key?

WPA2 key
K = 31EF21456

WPA2 key
K = 31EF21456

Session 1 encrypted with K

Session 2 encrypted with K

We must use same static K… how to get PFS then???

$g^x \bmod p$

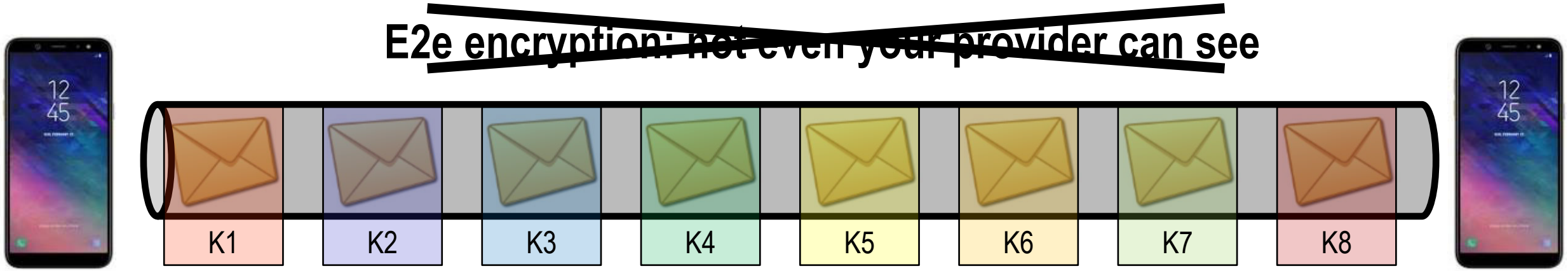
$g^y \bmod p$

Session 3 encrypted with KE = $HKDF_K(g^{xy})$

Session 4 encrypted with KE = $HKDF_K(g^{ab})$

… … …

Does your messaging app guarantee PFS?



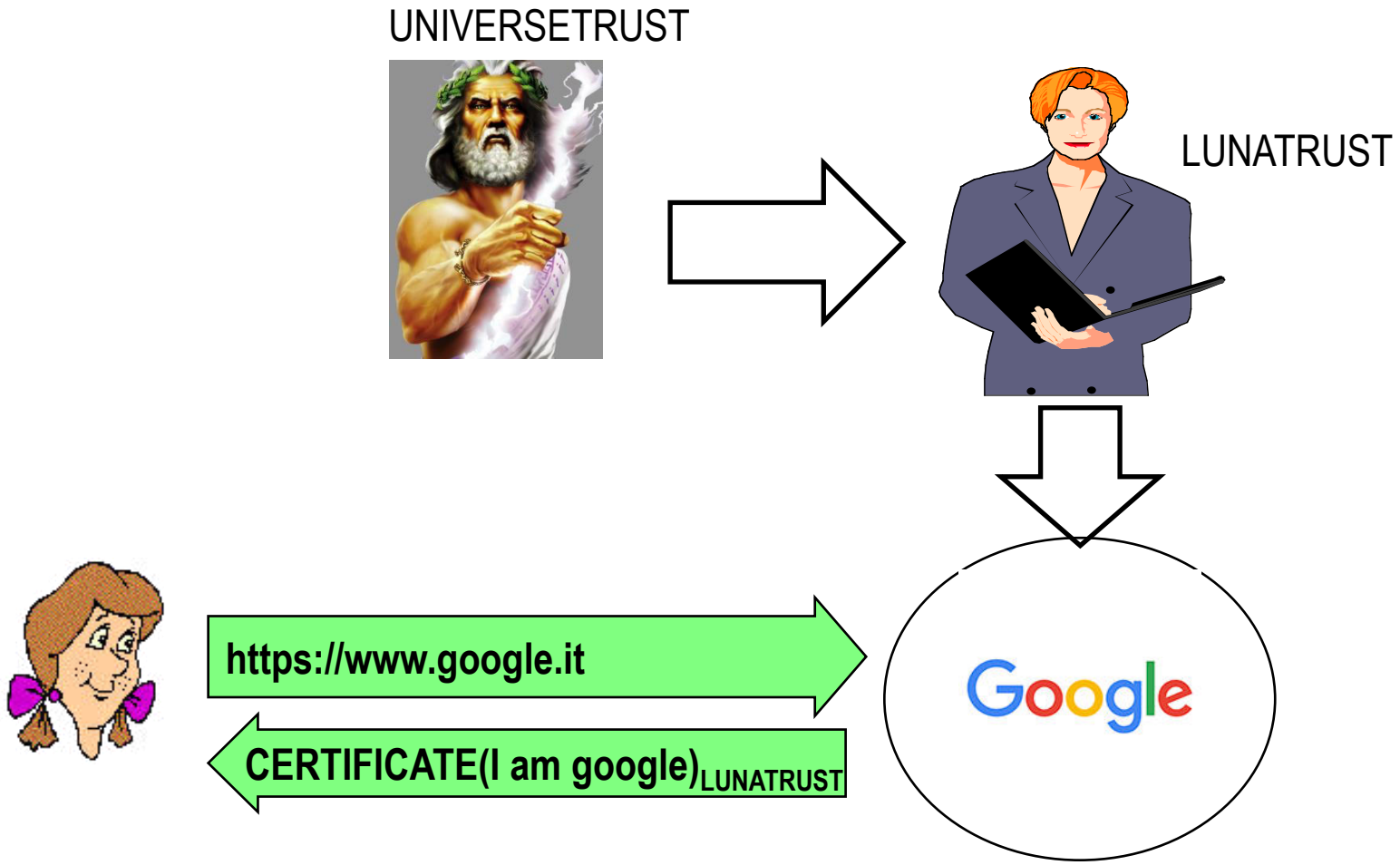
E2e encryption with PFS: Not only your provider cannot see,
But even if it logs all data and gets access later on to your key, he cannot decrypt past messages!

Signal's Double Ratchet: brilliant solution!
(too tech for today)

Back to mass surveillance: What about fake certificates?

a.k.a.: how transparency can help solving security problems

Web security pillar: Certificate Authorities ARE trusted!



Fact: trusted CA assumption at stake

With powerful threats (governments),
and many players which can «make mistakes»
the security of the PKI model is getting weaker and weaker

- ➔ **Google's VALID fake Certificates mistakenly (?) issued**
  - ➔ by TurkTrust (2012), ANSSI France (2013), etc
- ➔ **Smaller CAs: compromised**
  - ⇨ Holland: Dgnotar
  - ⇨ Malaysia: DigiCert sdn. Bhd.
  - ⇨ etc


Published: 09 Jan 2015

Netcraft has found dozens of fa
Some of these certificates may
customers. Successful attacks
and forwarding it to the bank.
authentication credentials, or r

The fake certificates bear com
As the certificates are not sign

Sections

Why 'b
SSL ce

Online banking apps for mobile
is far from trivial, and mobile
of iOS-based banking apps test
authenticity of SSL certificates
manual tests by Leibniz Univer
may also be vulnerable if a use

Enhancing digital certificate security

Posted: Thursday, January 3, 2013

+1 183 Tweet 300 Mi piace

Engineer

cted and blocked an unauthorized digital certificate for the "*.google.com"
ly and found the certificate was issued by an intermediate certificate authority
a Turkish certificate authority. Intermediate CA certificates carry the full
has one can use it to create a certificate for any website they wish to

certificate revocation metadata on December 25 to block that intermediate CA,
other browser vendors. TURKTRUST told us that based on our information,

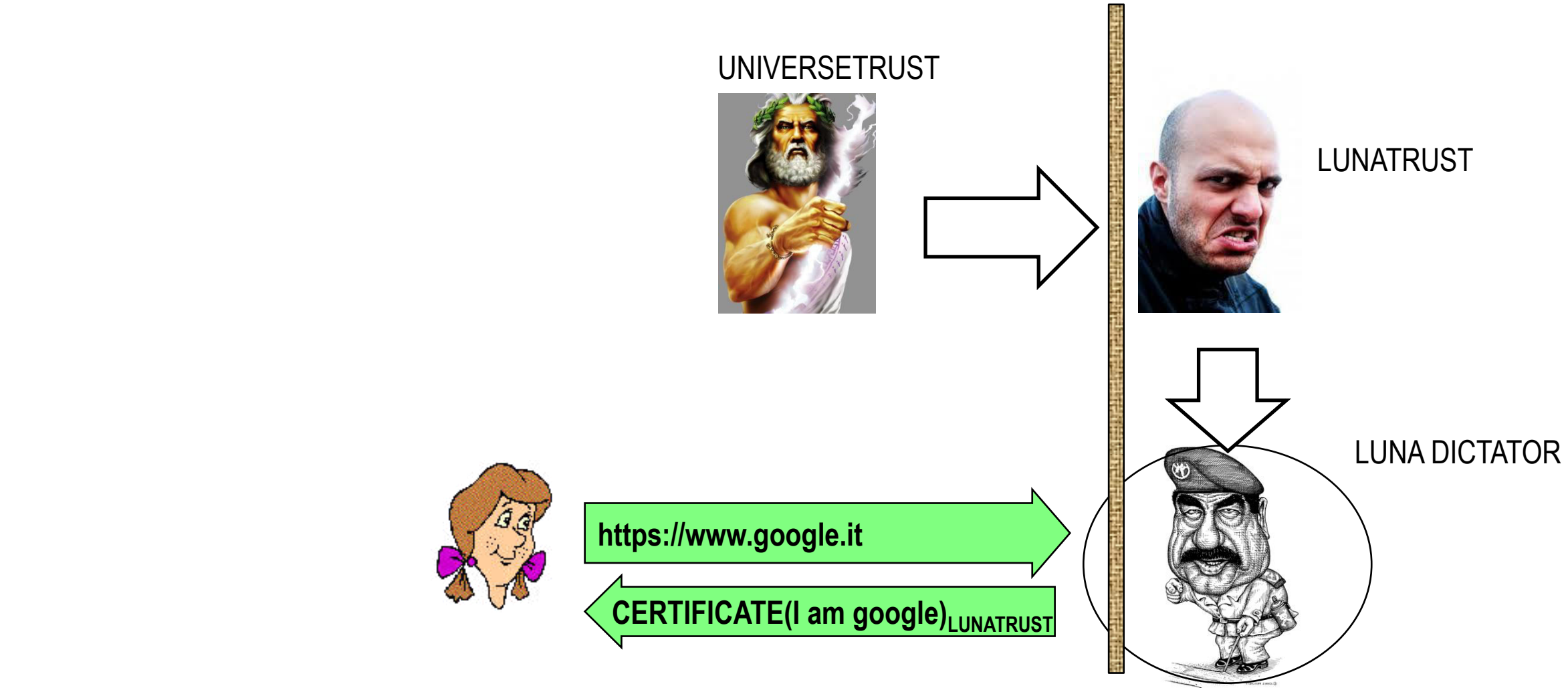
Our actions add
Chrome again in
though connecti

Since our priority
further discussion and careful consideration.

TLS Proxies: Friend or Foe?

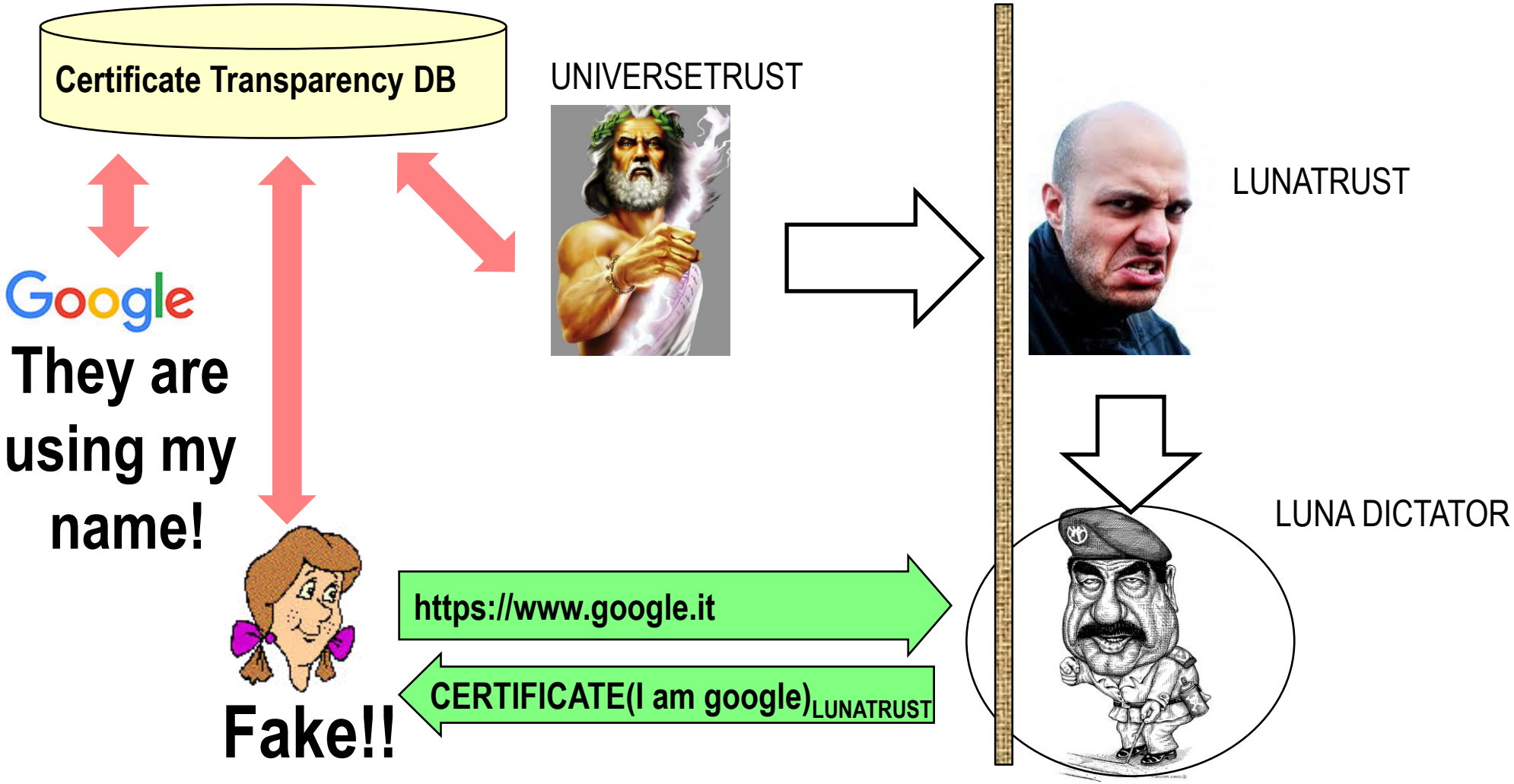
Mark O'Neill, Scott Ruoti, Kent Seamons, Daniel Zappala
Brigham Young University
Department of Computer Science
Provo, UT 84602

Mass surveillance with fake certificates…

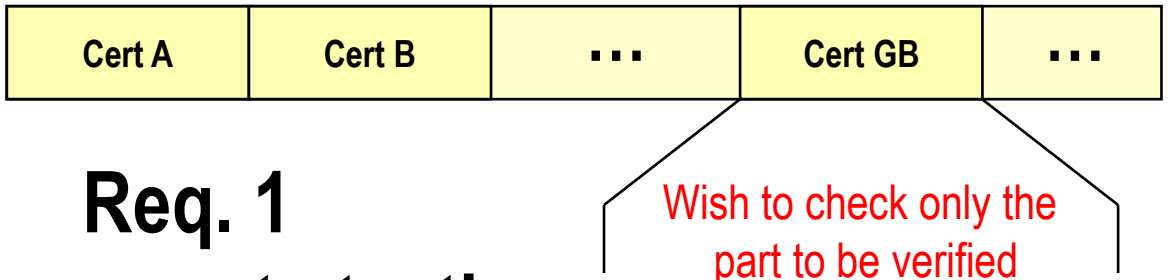


How to cope with malicious CAs?

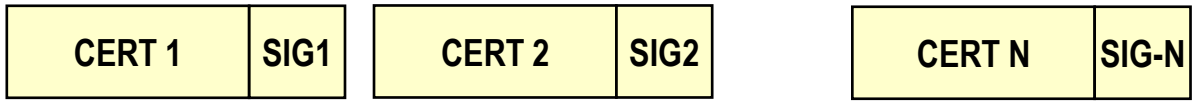
Idea: gigantic worldwide DB which anyone can check!



How to implement such gigantic Database of certificates?



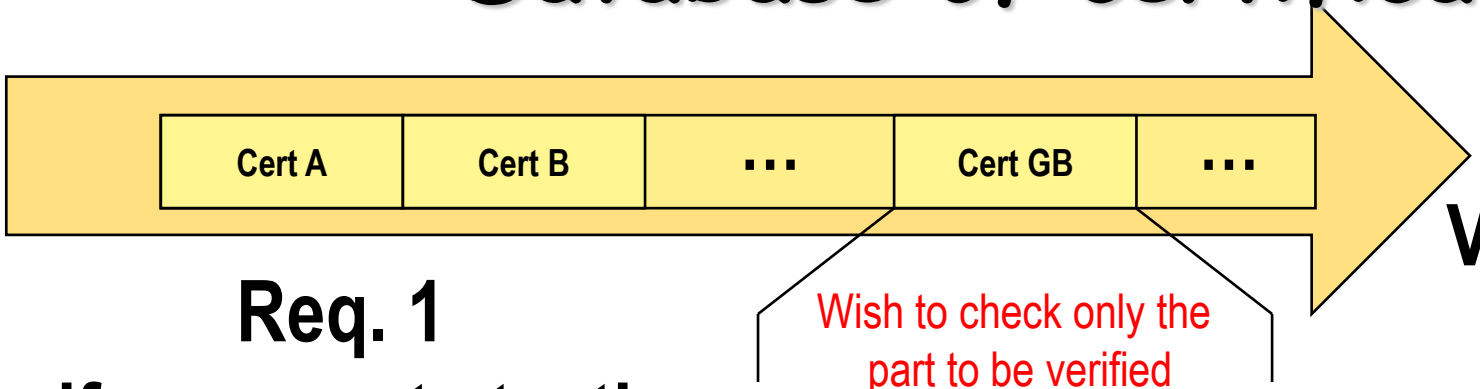
Req. 1
Verify one cert at a time



Easy! Add SIG @ each cert

Well… not really!
Strip-type attack

How to implement such gigantic Database of certificates?

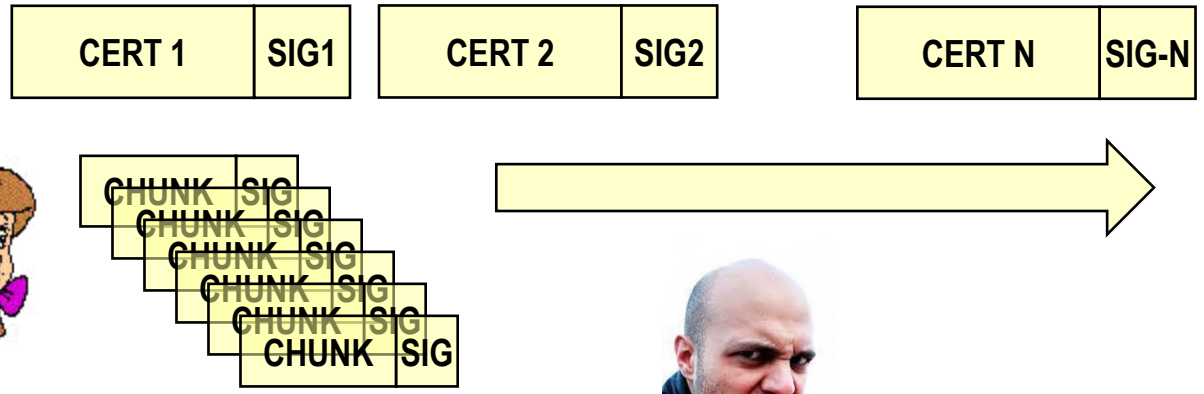


Req. 1
Verify one cert at a time

Wish to check only the part to be verified

Req. 2
Verify as-a-whole integrity!

Need to verify – hence read the WHOLE database!

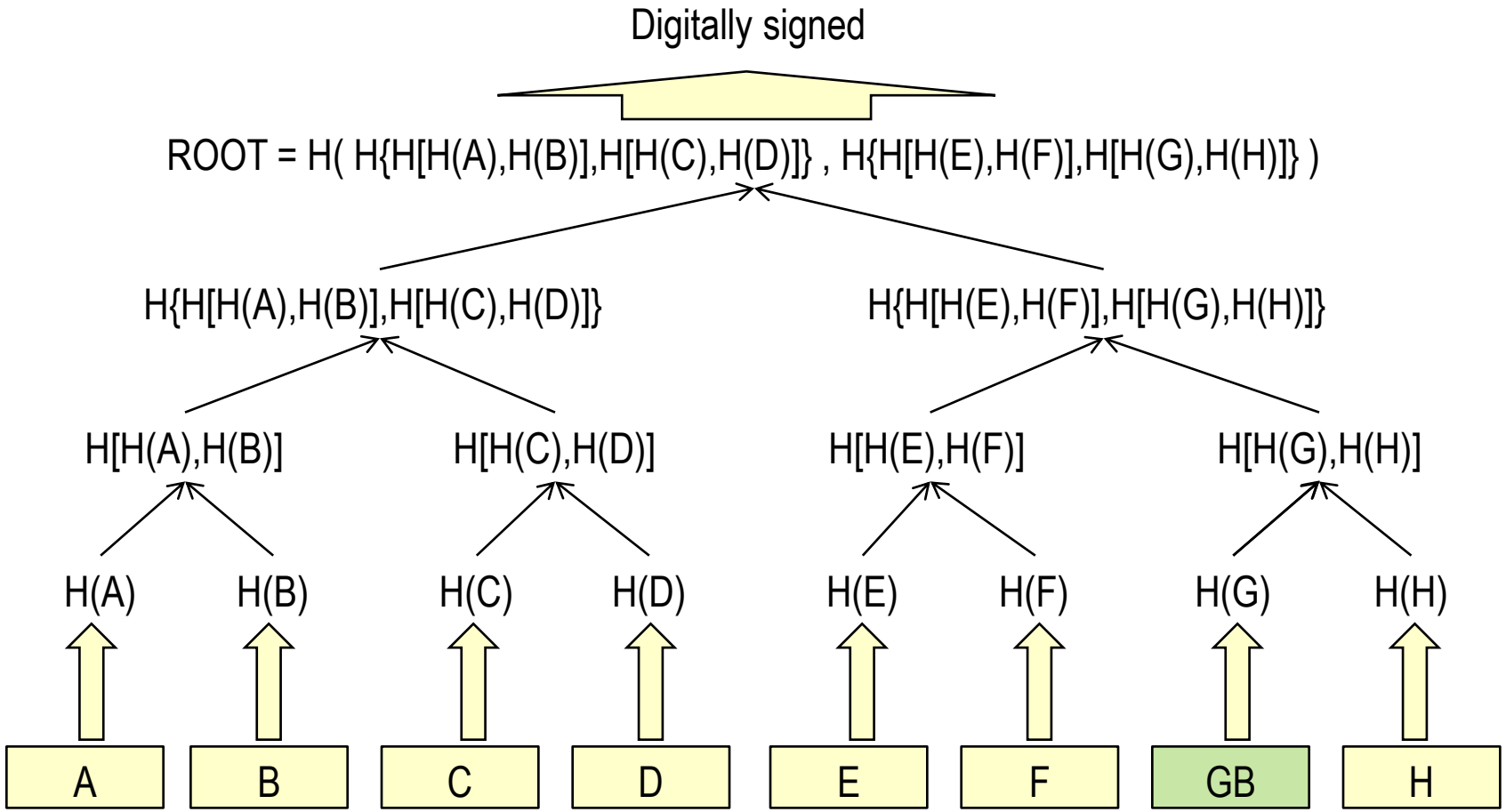


Easy! Add SIG @ each cert

Ops… did I receive all?

Well… not really!
Strip-type attack

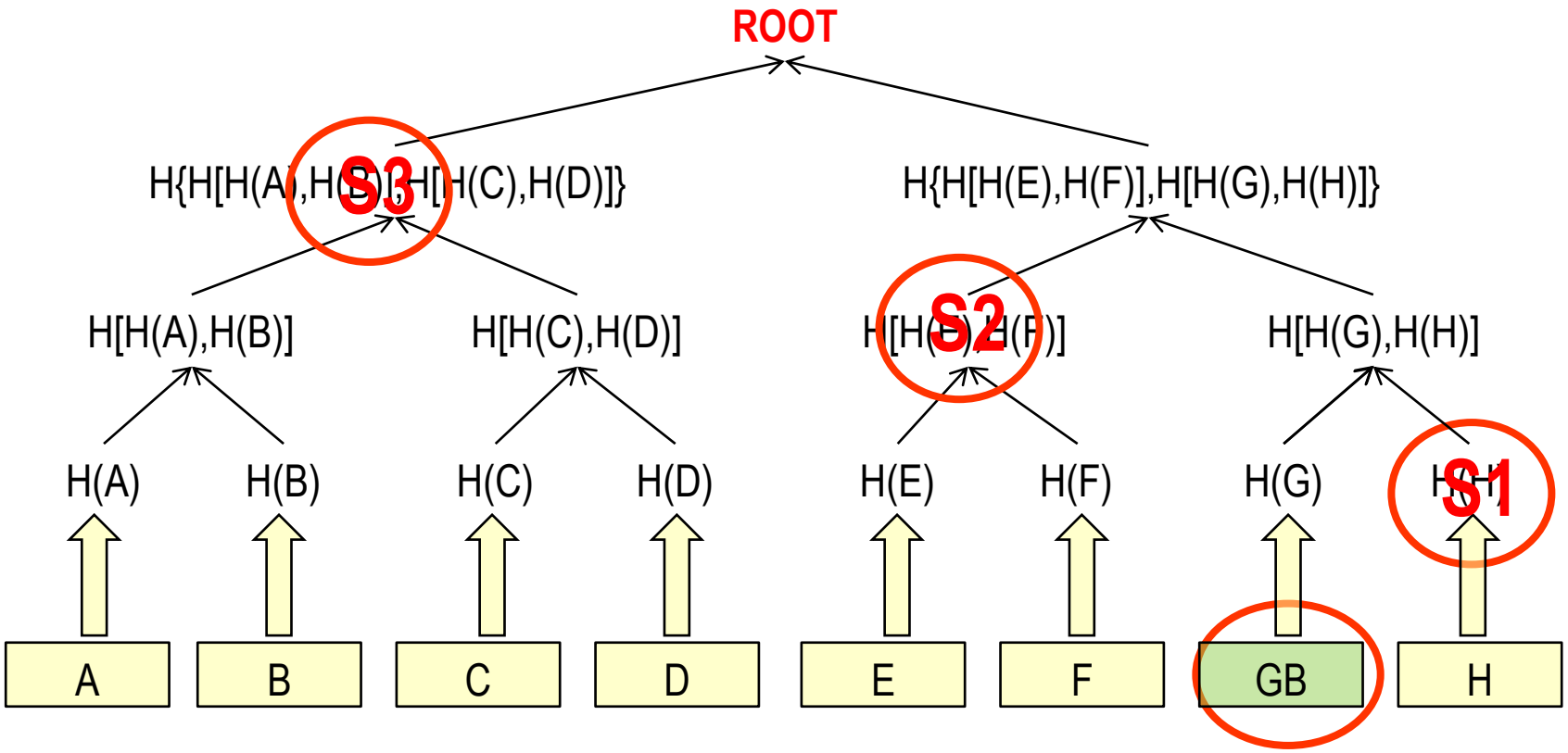
You can have them both! Merkle's trees (1979)



Single CERT verification: fast with “siblings”!

Give me:

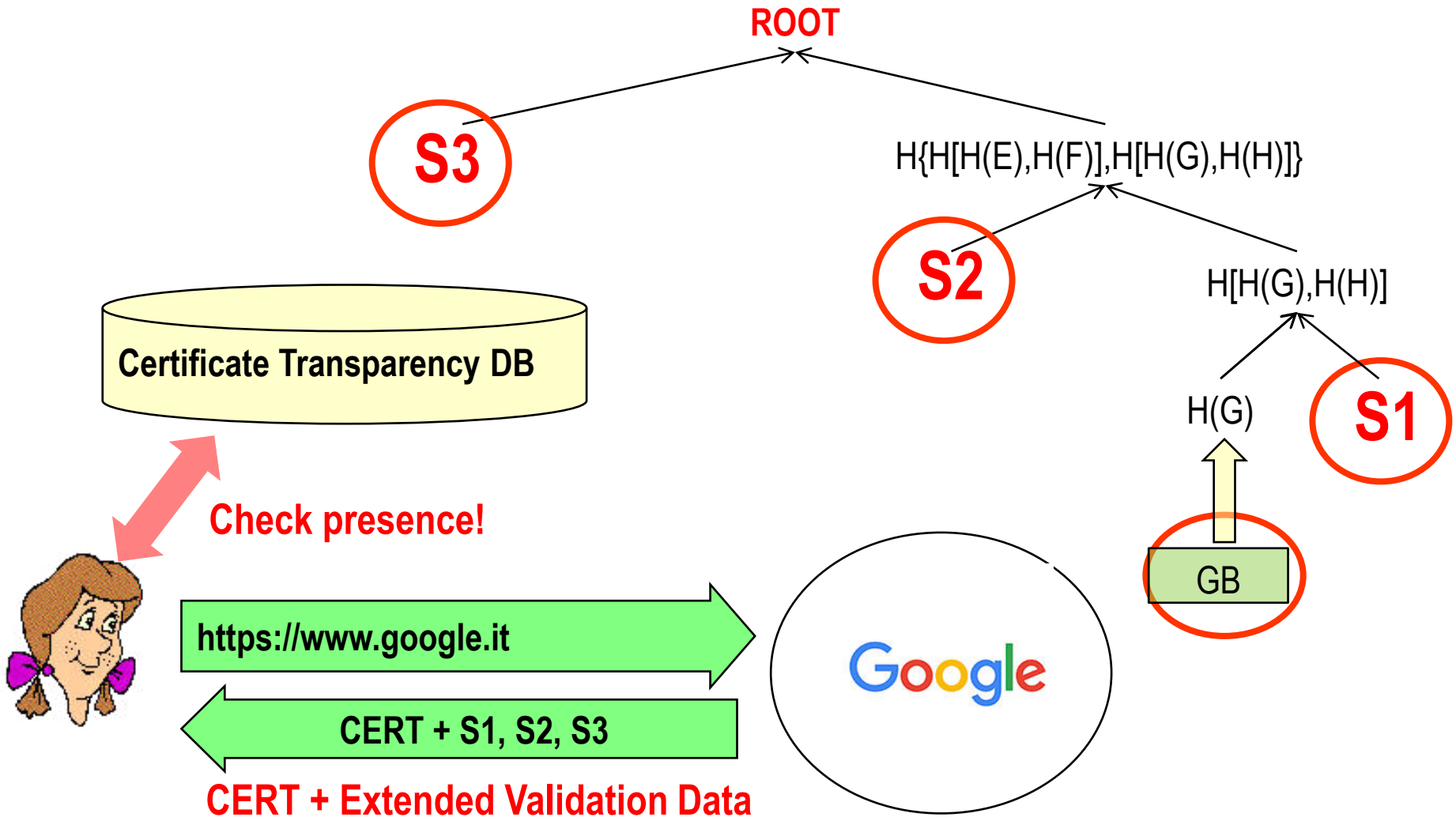
- Cert GB
- Sibling S1
- Sibling S2
- Sibling S3



Single CERT verification: fast with “siblings”!

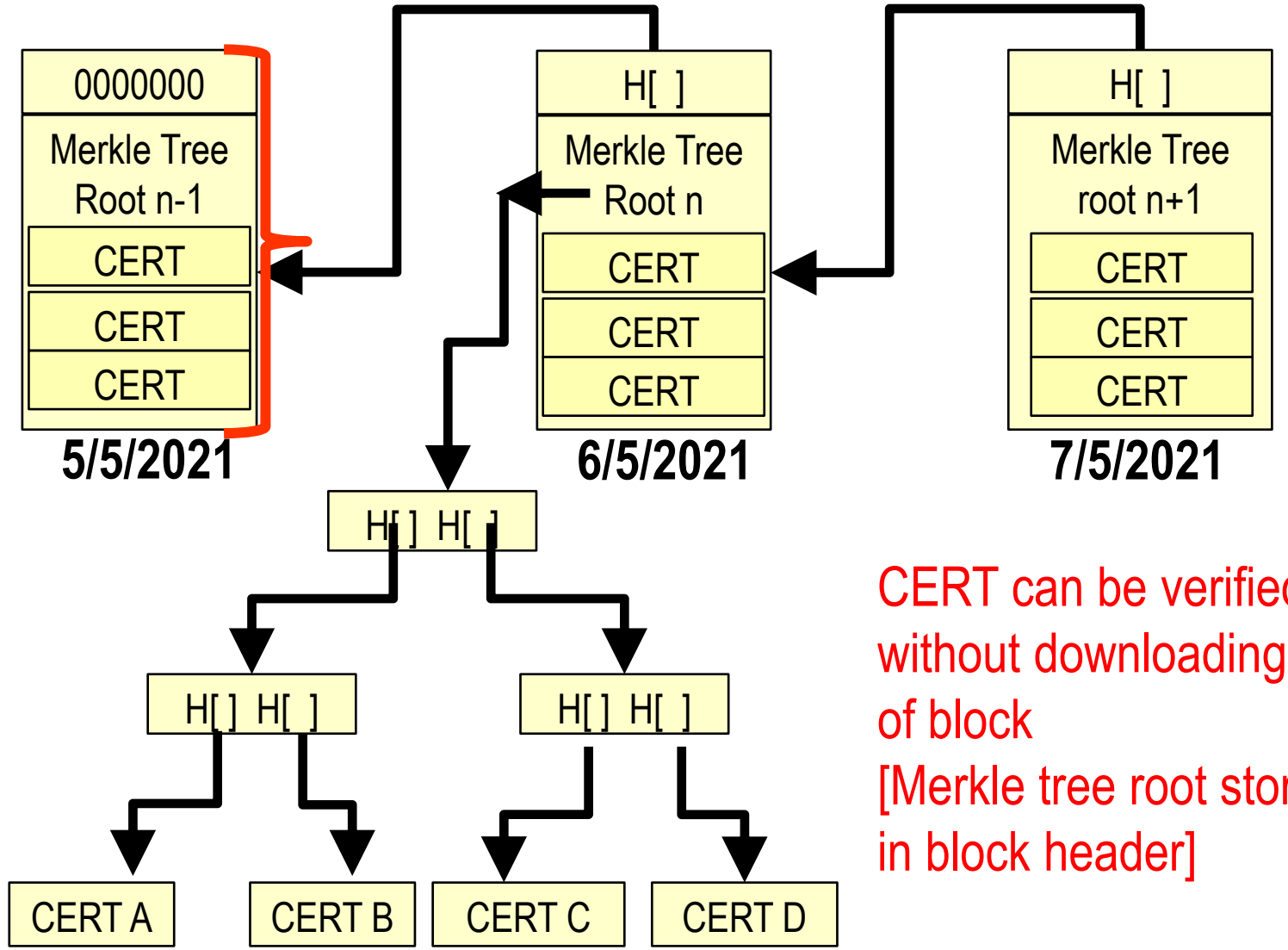
Give me:

- Cert GB
- Sibling S1
- Sibling S2
- Sibling S3



Merkle trees vs time? Blockchain-type DB!

Hash pointers:
Integrity of the
Whole history
From block 0!



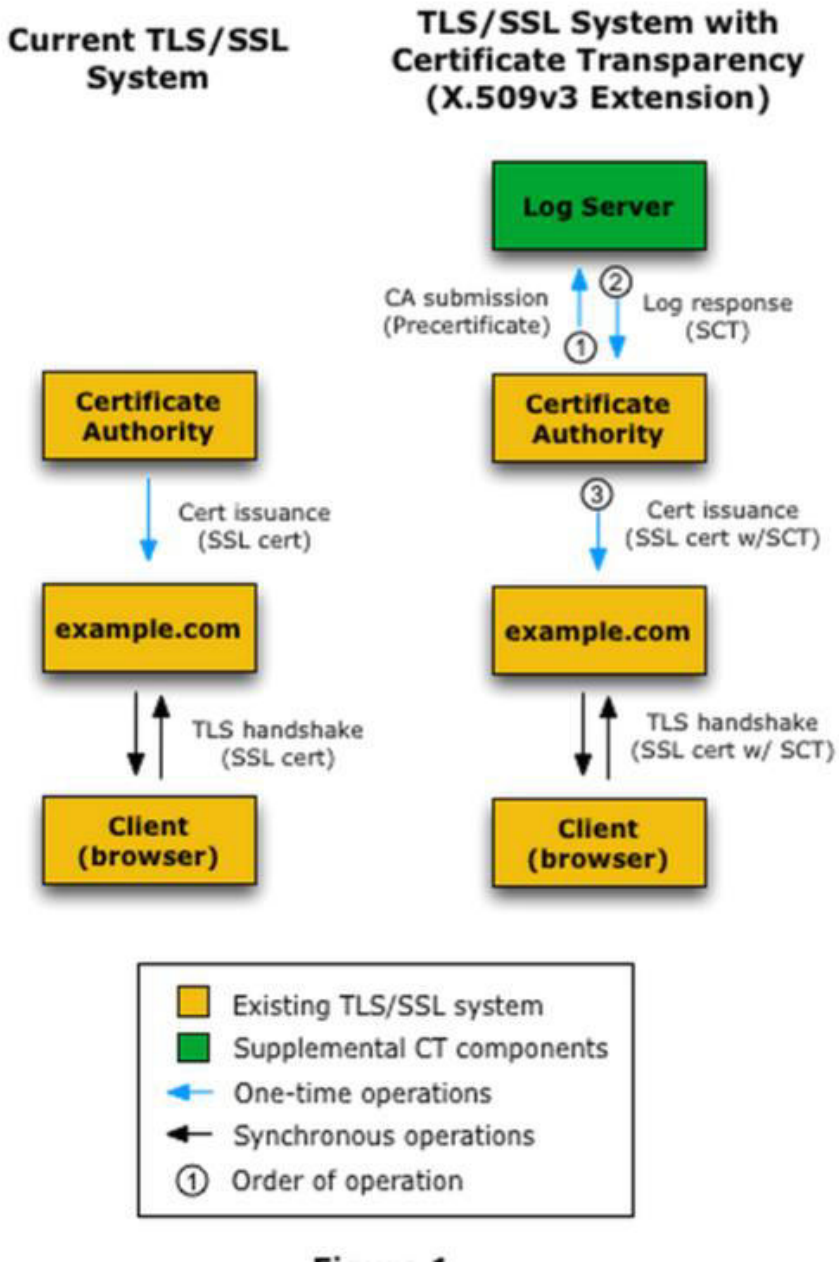
Does this
look familiar?

☺

CERT can be verified
without downloading content
of block
[Merkle tree root stored
in block header]

Done for real: Certificate Transparency

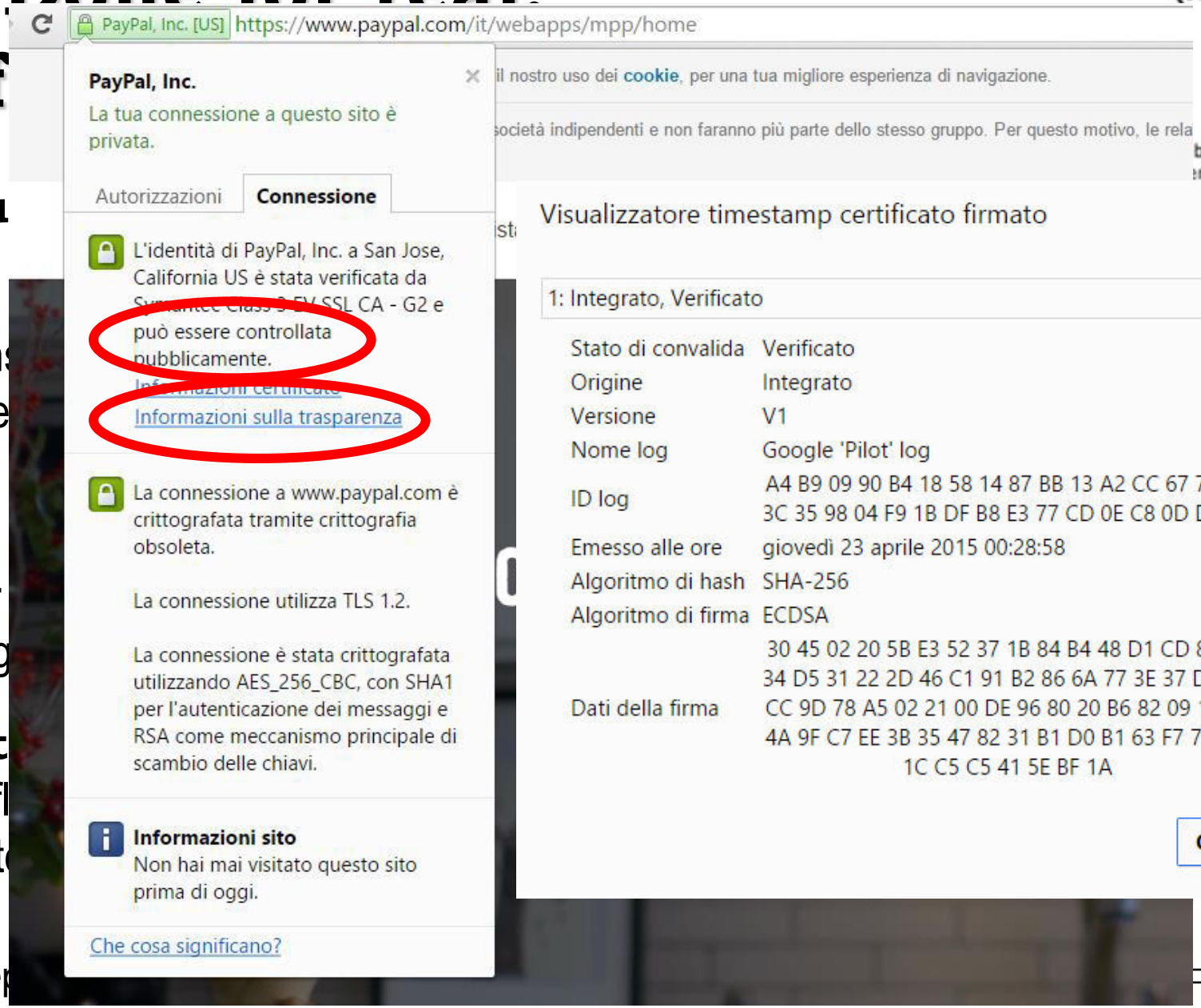
- **Lauched in july 2013 by Google**
- **experimental IETF RFC 6962**
  - IETF WG «trans»
  - Specified processes and protocols
- **Integrated in all major browsers**
- **Supported by major sites**
  - Paypal, CertSign, etc
- **Multiple pilot logs**
  - Google, Cloudflare, DigiCert, etc
  - https://certificate.transparency.dev/logs/

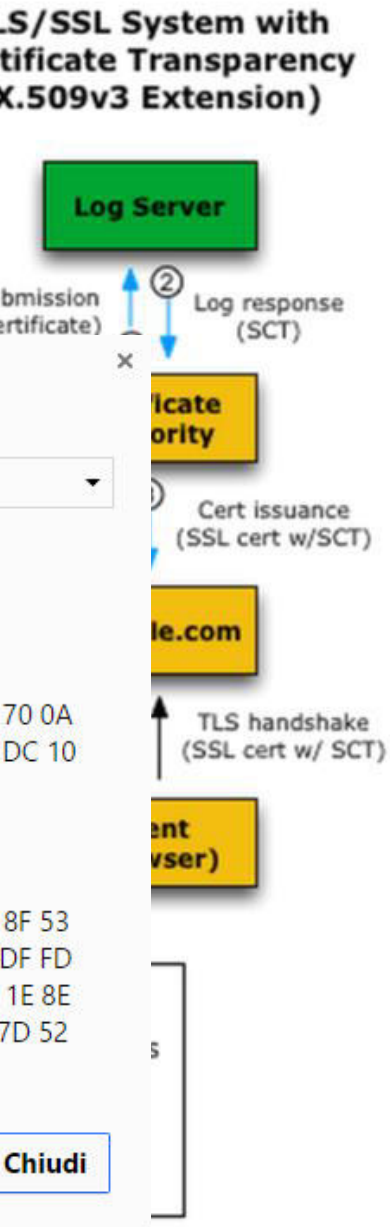


Done for real:

Certif

- ➔ **Lauched in ju**
- ➔ **experimental**
  - ➔IETF WG «trans
  - ➔Specified proce
- ➔ **Integrated in**
- ➔ **Supported by**
  - ➔Paypal, CertSig
- ➔ **Multiple pilot**
  - ➔Google, Cloudfl
  - ➔https://certificat





Cert Trans → is a Blockchain?

→It definitely looks like

⇨Identical (!) architecture as Bitcoin's ledger

→It could be distributed as well

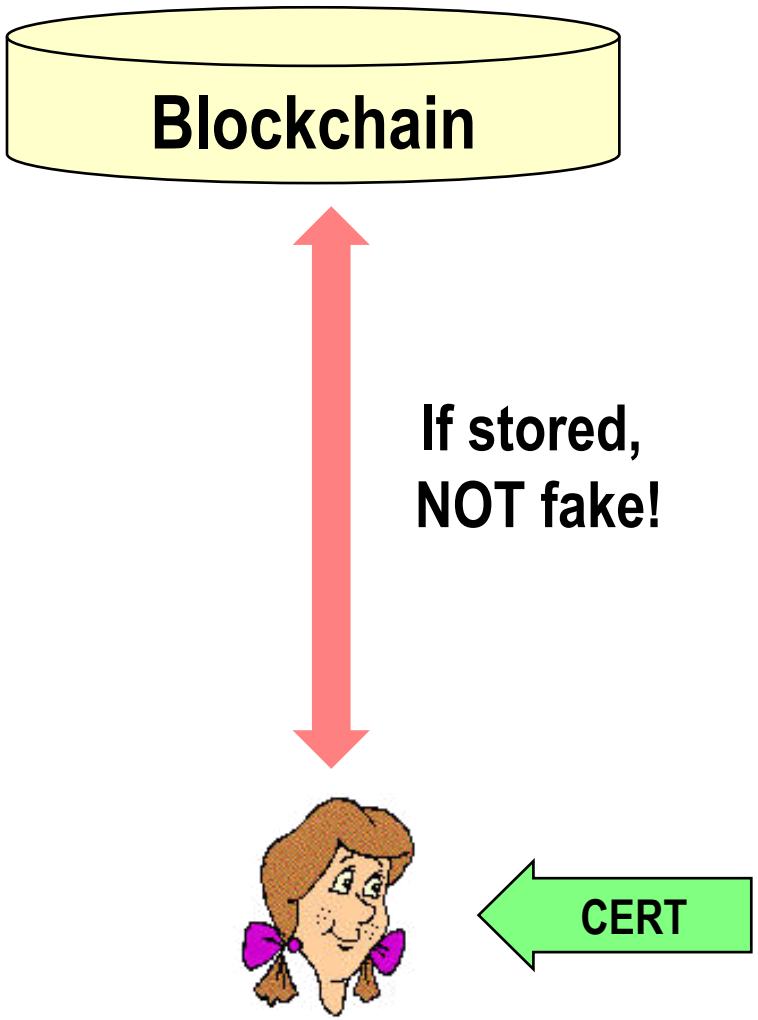
⇨Instead of different logs, make it a same «consensus» log

→But it is NOT a Blockchain

⇨It does NOT necessarily contain only VALID certificates!

→As a Blockchain mandates; also Google being crystal clear on this!

→



Cert Trans → is a Blockchain?

➔It definitely looks like

⇨Identical (!) architecture as Bitcoin's ledger

➔It could be distributed as well

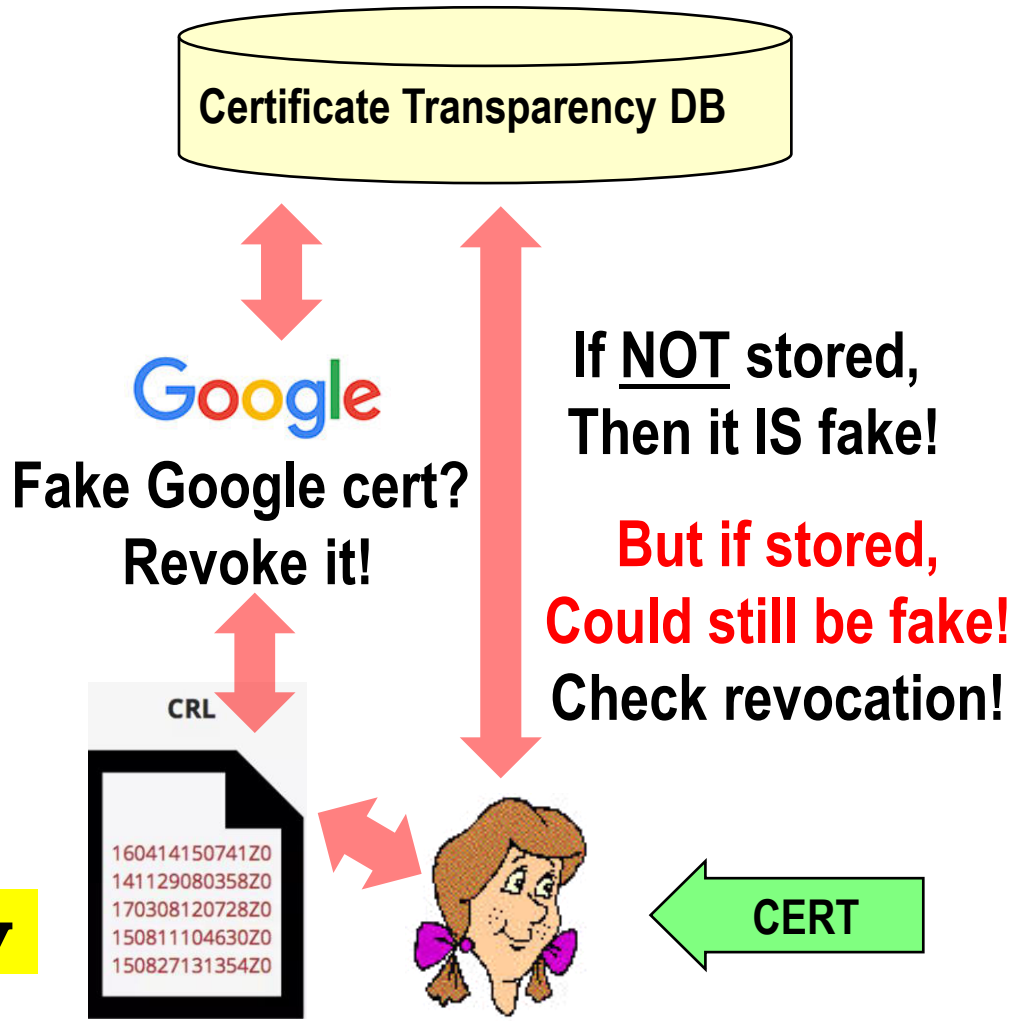
⇨Instead of different logs, make it a same «consensus» log

➔But it is NOT a Blockchain

⇨It does NOT necessarily contain only VALID certificates!

→As a Blockchain mandates; also Google being crystal clear on this!

➔Security comes from transparency



And what about cloud privacy? We must get rid of it?

a.k.a.: how to compute on encrypted data - ultra-brief intro to SECURE MULTIPARTY COMPUTATION

(next slides in Italian)

Cominciamo dalle conclusioni!

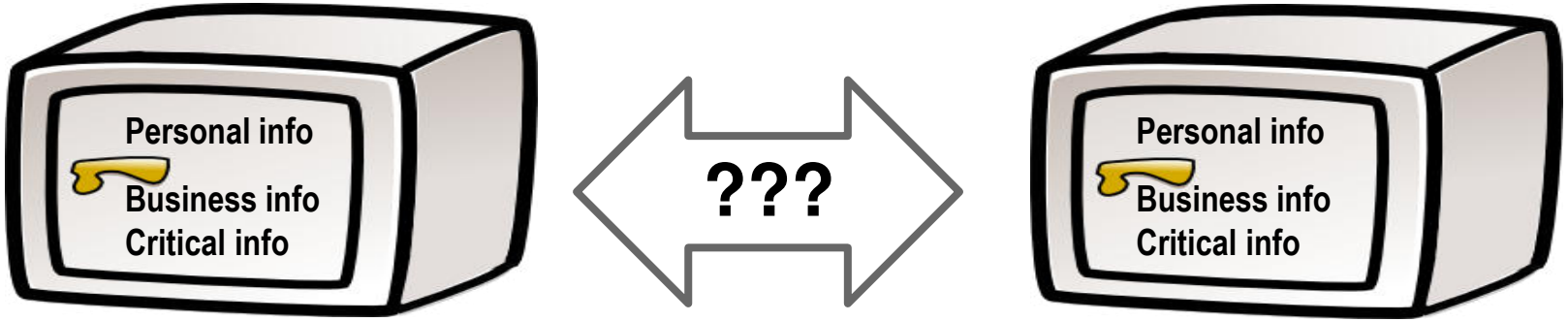
Elaborazione di dati cifrati: 

ovvero: elaborare dati condivisi da più persone
senza che nessuno possa «vedere» i dati di origine

Risultati scientifici disponibili da quasi 40 anni!
Falsità della dicotomia
confidenzialità ←→ utilità dei dati

Ma perché pochi sono informati? E perché le applicazioni sono praticamente inesistenti? potremmo fare molto di più…

«The sharing dilemma»
Condividere o non condividere?

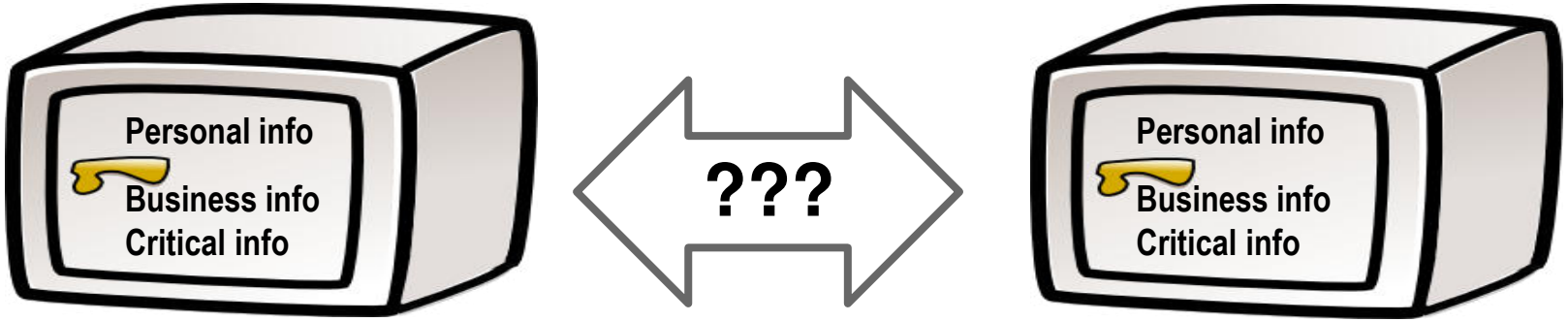


Caso critico:
NO autorità o intermediario fidato

Internet AS, Stati, Aziende

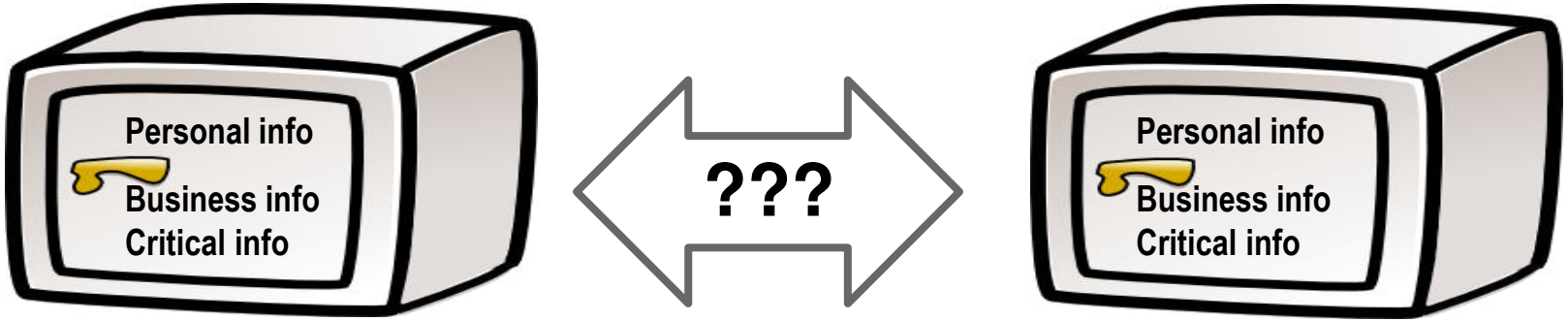
…

«The sharing dilemma»
Condividere o non condividere?



Condividere (minacce, attacchi, dati):
collaborazione → più sicurezza,
più efficienza, servizi migliori

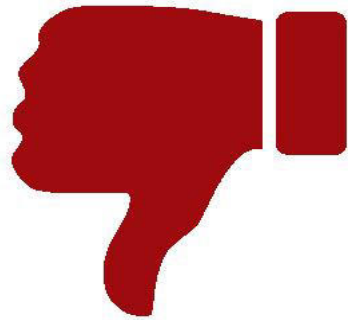
«The sharing dilemma»
Condividere o non condividere?



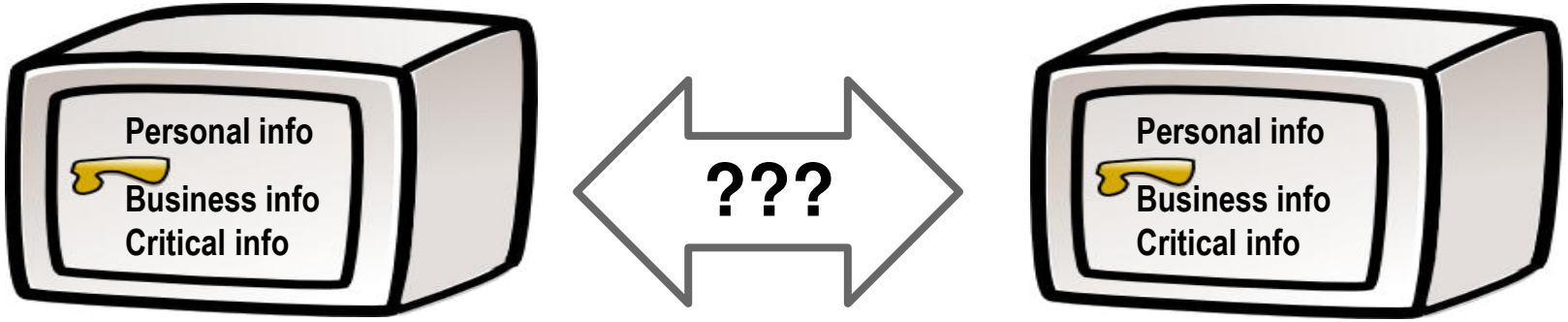
Condividere (minacce, attacchi, dati):

Meno privacy, meno confidenzialità

del business, maggior esposizione



«The sharing dilemma»
Condividere o non condividere?



Percezione da parte del «non tecnico»:

DEVE *essere un gioco a somma zero*

Dobbiamo ***per forza*** *cercare un trade-off*

«The sharing dilemma»
Una terza via: Win-Win!

Homeland Security Department, stato 1	Airline, stato 2
Lista Sospetti -BNCGPP75T10H5091 -RSSFRC82S21X5091 -…	**Lista Passeggeri** -VRDRBT65T10B157X -BNCGPP75T10H5091 -…

➔ **C'è qualche sospetto su questo volo?**

➔ **Soluzione 1: fornire lista sospetti a compagnia aerea**

⇨ No way!!

➔ **Soluzione 2: obbligare airline a fornire lista passeggeri**

⇨ In mancanza di meglio… ma.... (passeggeri in transito? privacy?)

«The sharing dilemma»
Una terza via: Win-Win!

Homeland Security Department, stato 1	Airline, stato 2
Lista Sospetti -BNCGPP75T10H5091 -RSSFRC82S21X5091 -…	**Lista Passeggeri** -VRDRBT65T10B157X -BNCGPP75T10H5091 -…

➔**C'è qualche sospetto su questo volo?**

➔**Soluzione Win-Win: Private Set Intersection!**

⇨Freedman et. Al. 2004, Ateniese et al. 2011, GB 2016, etc

Nulla di nuovo! (per i crittografi)

➔**Secure Multi Party Computation (e variazioni sul tema)**

- ⇨*private information retrieval,*
- ⇨*Homomorphic encryption*
- ⇨*Conditional encryption*
- ⇨*etc*

➔**SOLO interessato al risultato, NON ai dati di input?**

- ⇨**MPC is for you!** (dati input rimangono confidenziali)

➔**Ricerca estremamente consolidata, da quasi 40 anni!**

- ⇨Andrew Yao, 1982, problema del Milionario
- ⇨Migliaia (!) di lavori scientifici e di applicazioni, da allora

Si, ma…

➔Troppo difficile, per nulla scalabile

⇨Inizialmente vero

⇨*garbled circuits, oblivious transfer, Pailler Encryption, …*

➔Ma ora è BEN diverso! Tecniche moderne ESTREMAMENTE efficienti e pratiche!

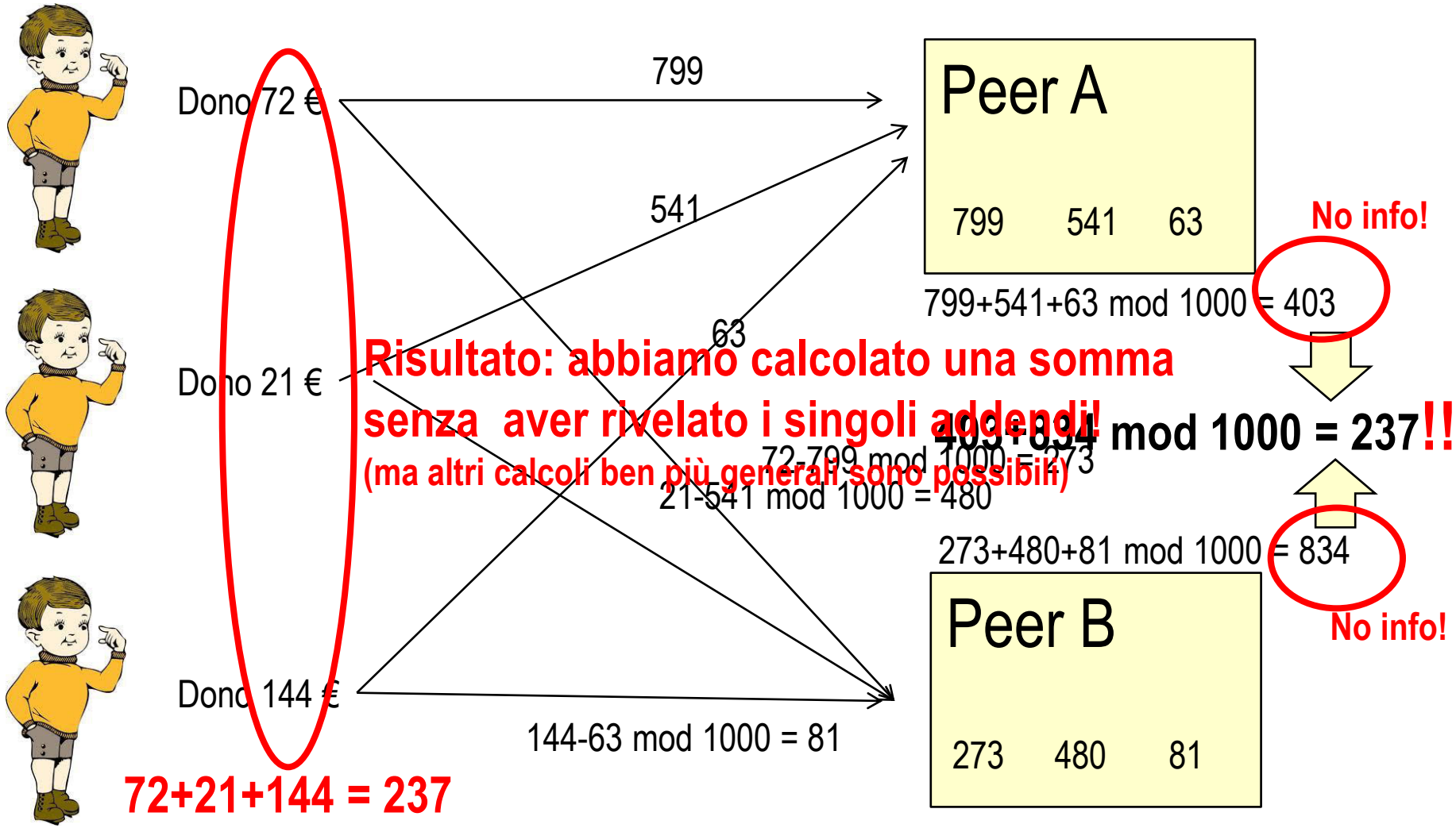
⇨Post-2010, basate su secret sharing

⇨Nigel Smart: ERC →Dyadic → Unbound

➔E molto flessibili

⇨Gentry's doubly-homomorphic encryption

Estremamente pratiche?! Un semplicissimo esempio!



Altri «layman examples» (casi realmente accaduti!!)

- **Valutazione docenti: da cartacea a web**
  - Studenti: non ci fidiamo!
  - Centro di calcolo: garantisco io!
    - (ma se qualcuno poi attacca il server?)
  - server docenti + server studenti + MPC:
    - Fiducia non più necessaria!

- **Commissione valutazione best PhD award**
  - 5 commissari, 1 aggregatore sapeva voti di tutti
  - MPC distribuito!
    - Nessuno saprà nulla
    - Ma tutti conosceremo media voti.

Applicazioni: tante! Esempi:

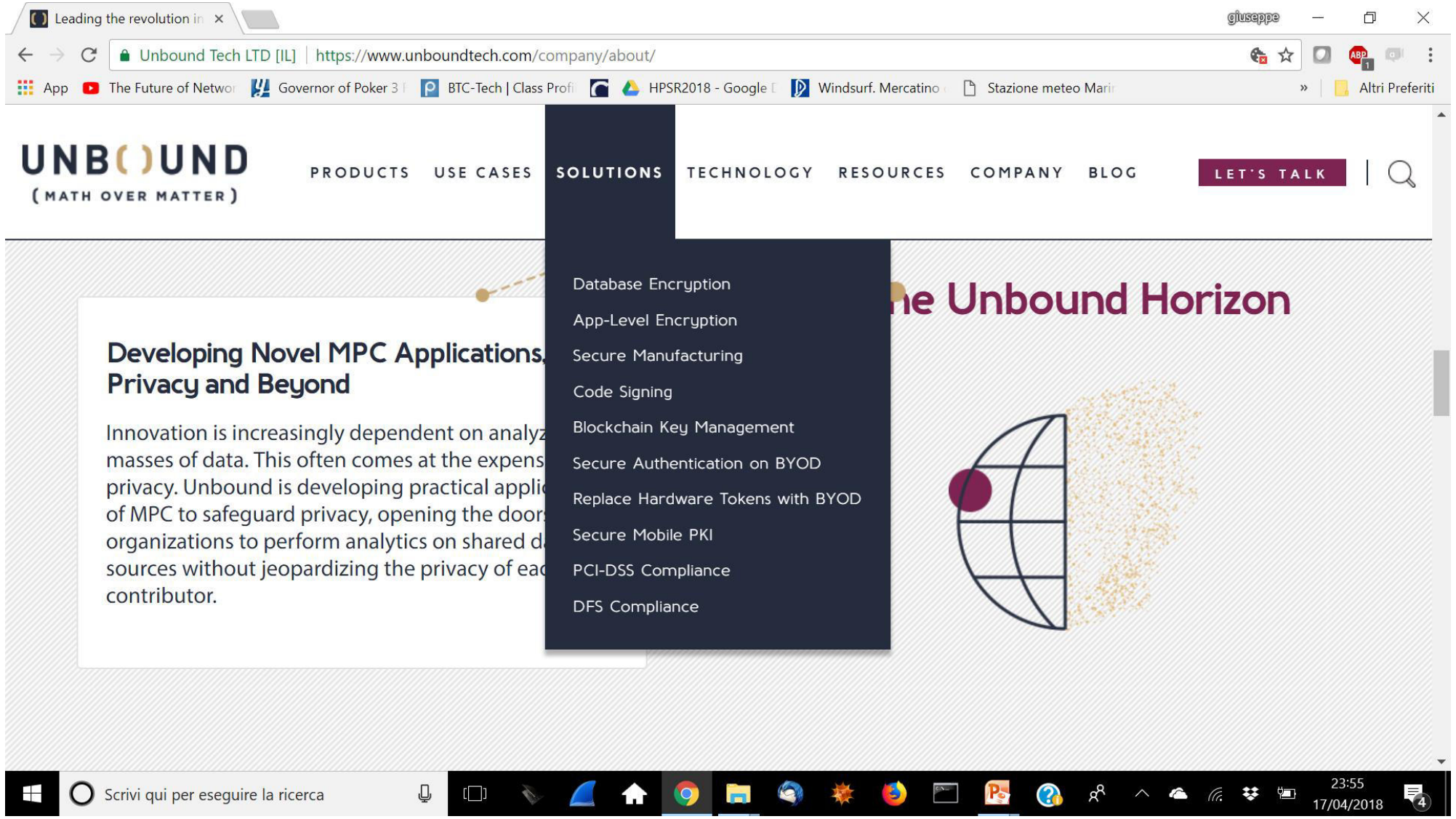
- **Key segregation/isolation, signature-splitting, …**
  - Segreti MAI usati in chiaro → più sicurezza, anti-tampering, …
  - **Modello di business principale di Unbound Security – Israele!**
- **Inter-domain data sharing**
  - Scambio dati tra unità operative in domini differenti
  - monitoring & security → più privacy, business confidentiality
- **Cloud Systems, auction bid systems**
  - Permettere al cloud di operare sui miei dati → senza rivelarli
- **Bio-informatica, genetica, e-health, …**
  - Elaborazione su dati sensibili cifrati, minima disclosure → privacy
- …

Ma perché MPC è ancora una nicchia?

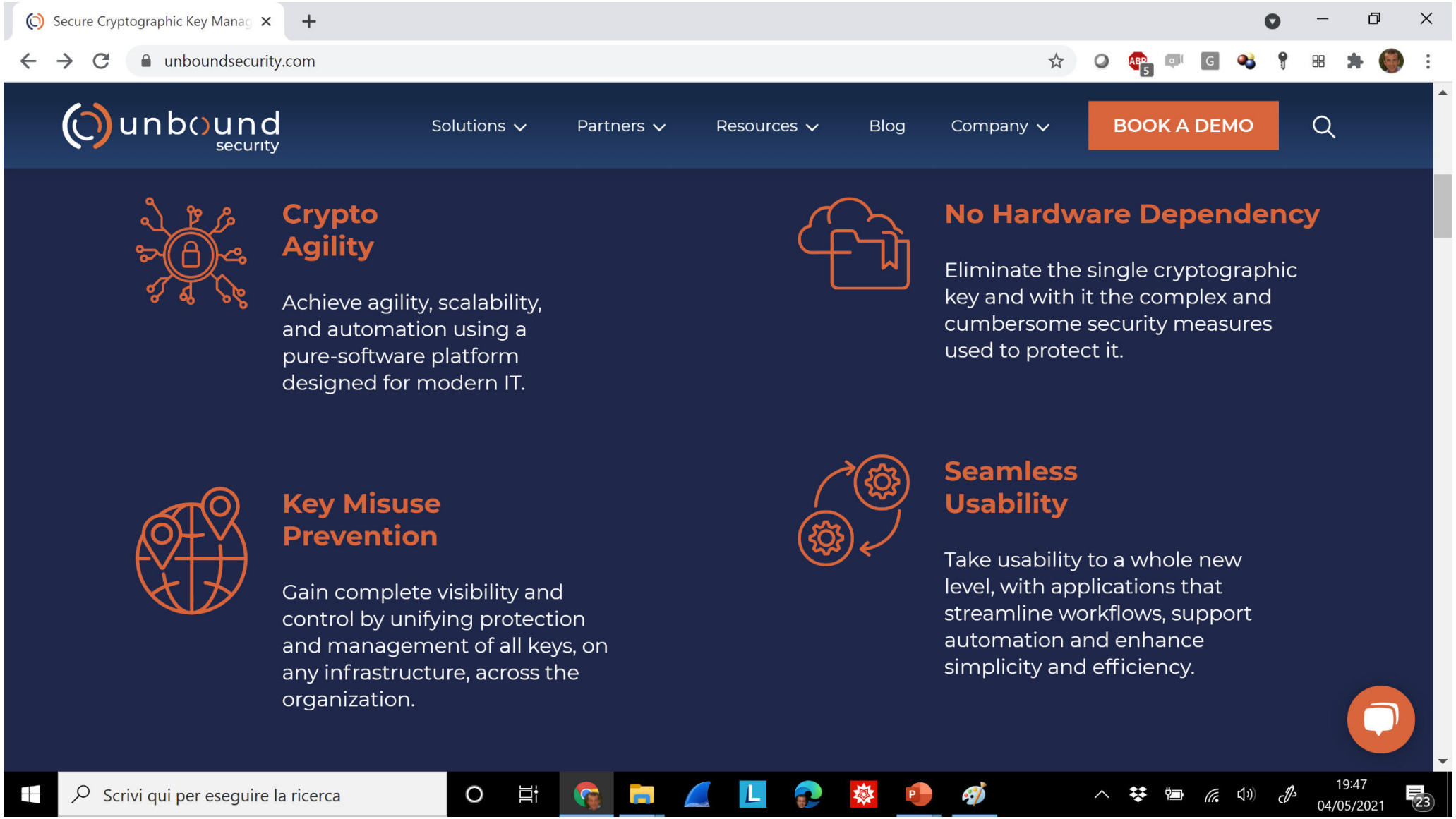
A maggior ragione visto che è in giro da quasi 40 anni…

- **Non ne sapevo nulla!!**
  - MPC non nei corsi base; crittografi poco divulgatori
- **Ma è «solo» privacy… e privacy non ci interessa (ooops)**
  - **Falso!** E' anche business confidentiality, data protection, system security hardening, key protection in BYOD, etc!
- **Scarsa interdisciplinarietà, gap tra problemi e soluzioni**
  - *"…Most research papers give imaginary applications…"*
  - Ma qualcosa sta cambiando (bio-informatica, start-ups, …)
- **MPC non scala, costa, richiede riprogrammazione**
  - Non più (dal 2010 in poi): prestazioni + buona flessibilità
    - SS-MPC per algo non banali quali PCA, SVD, vari ML, K-means, etc

Commercial MPC in 2018...



Commercial MPC in 2021… security!



Università di Roma Tor Vergata

cnit consorzio nazionale interuniversitario per le telecomunicazioni

Thank you

Giuseppe Bianchi

Professor of networking and network security
University of Roma «Tor Vergata»

giuseppe.bianchi@uniroma2.it