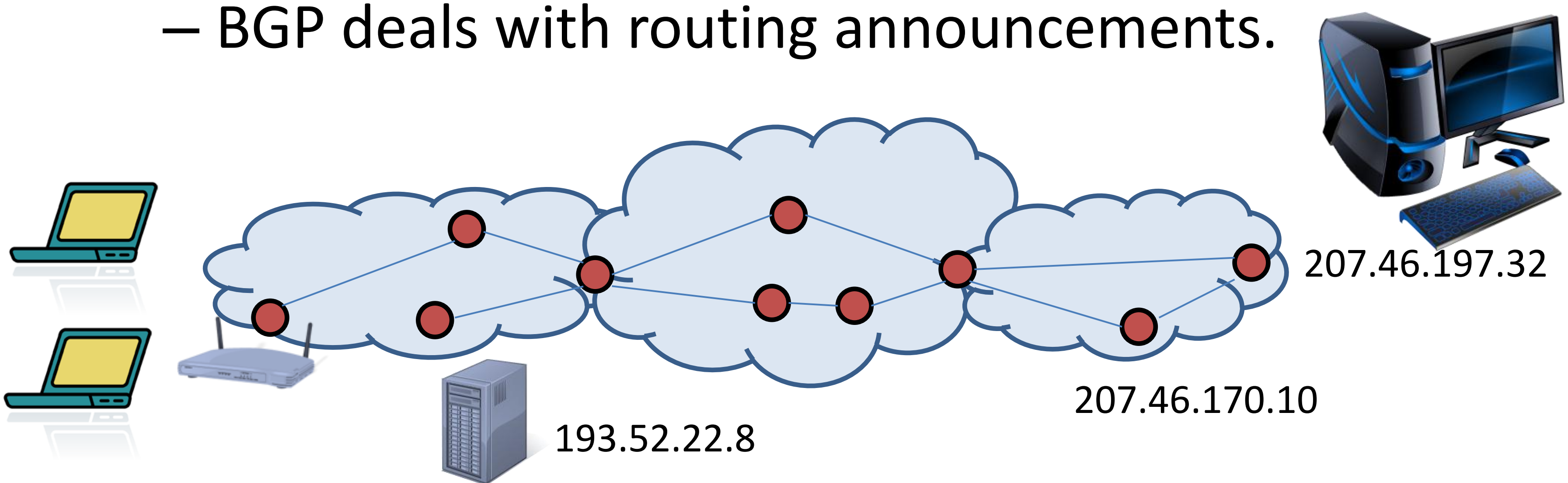
« On the Internet, nobody knows you are a dog » Twenty years later

Chaire Informatique et sciences numériques
Collège de France, cours du 4 mai 2011

Basics: weak authentication and its consequences

Infrastructure basics (brief review)

- Protocols for routing and communications work with IP addresses (e.g., 193.52.22.8).
  - IP delivers one packet.
  - Higher-level protocols, such as TCP, take care of multiple packets.
  - BGP deals with routing announcements.



Infrastructure basics (cont.)

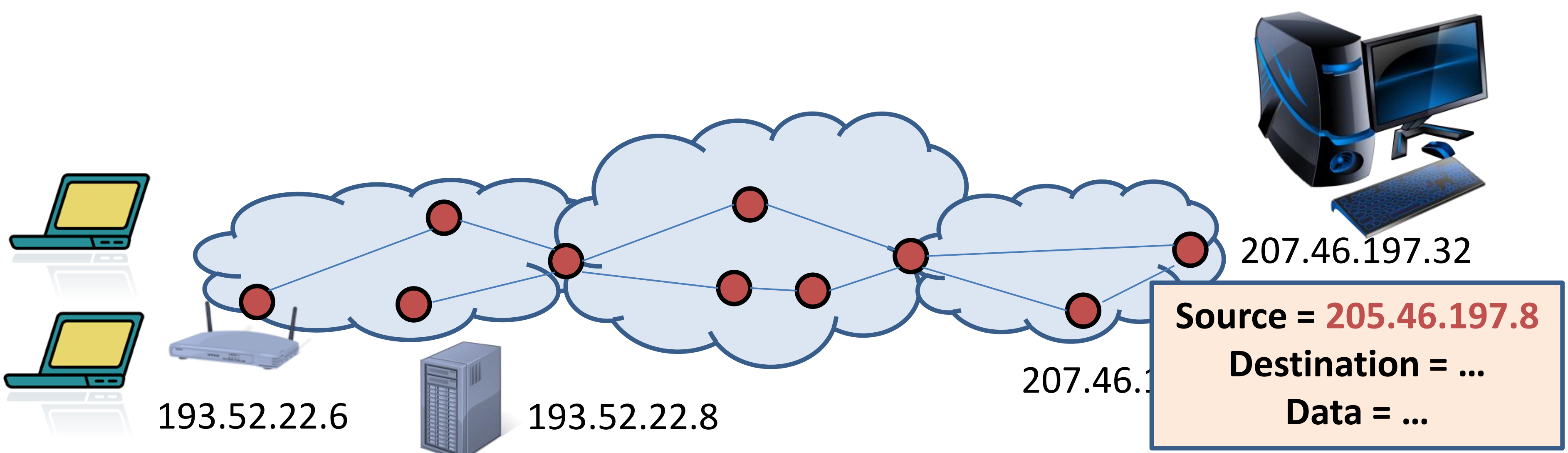
- The domain name system (DNS) associates symbolic names and IP addresses.
  - E.g., 193.52.22.8 is for www.college-de-france.fr.
  - The mapping is neither 1-1 nor constant.

Infrastructure basics (cont.)

- The domain name system (DNS) associates symbolic names and IP addresses.
  - E.g., 193.52.22.8 is for www.college-de-france.fr.
  - The mapping is neither 1-1 nor constant.
  - And there are also ***DNS lies*** (e.g., returning advertisements instead of NXDOMAIN for non-existent domains).

Problems: Authenticity

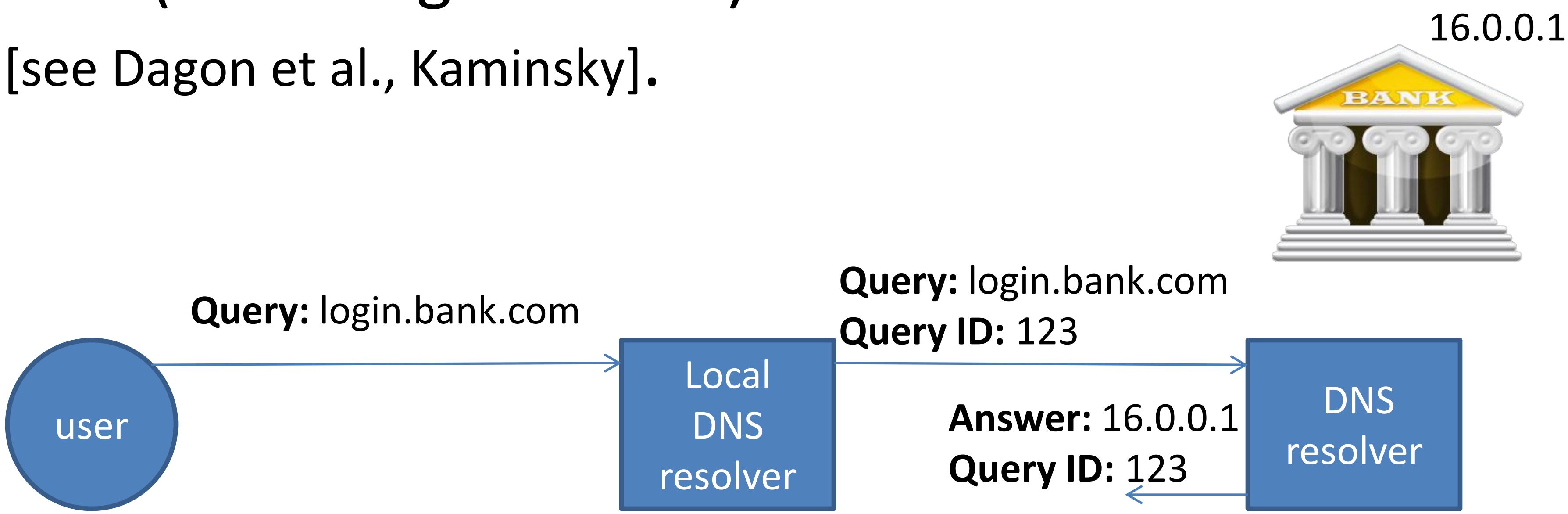
- Packets include source IP addresses.
- Those can be chosen arbitrarily by senders.
- Intermediaries may also tamper with packets.



Problems: Authenticity (cont.)

DNS (in its original form) is also vulnerable

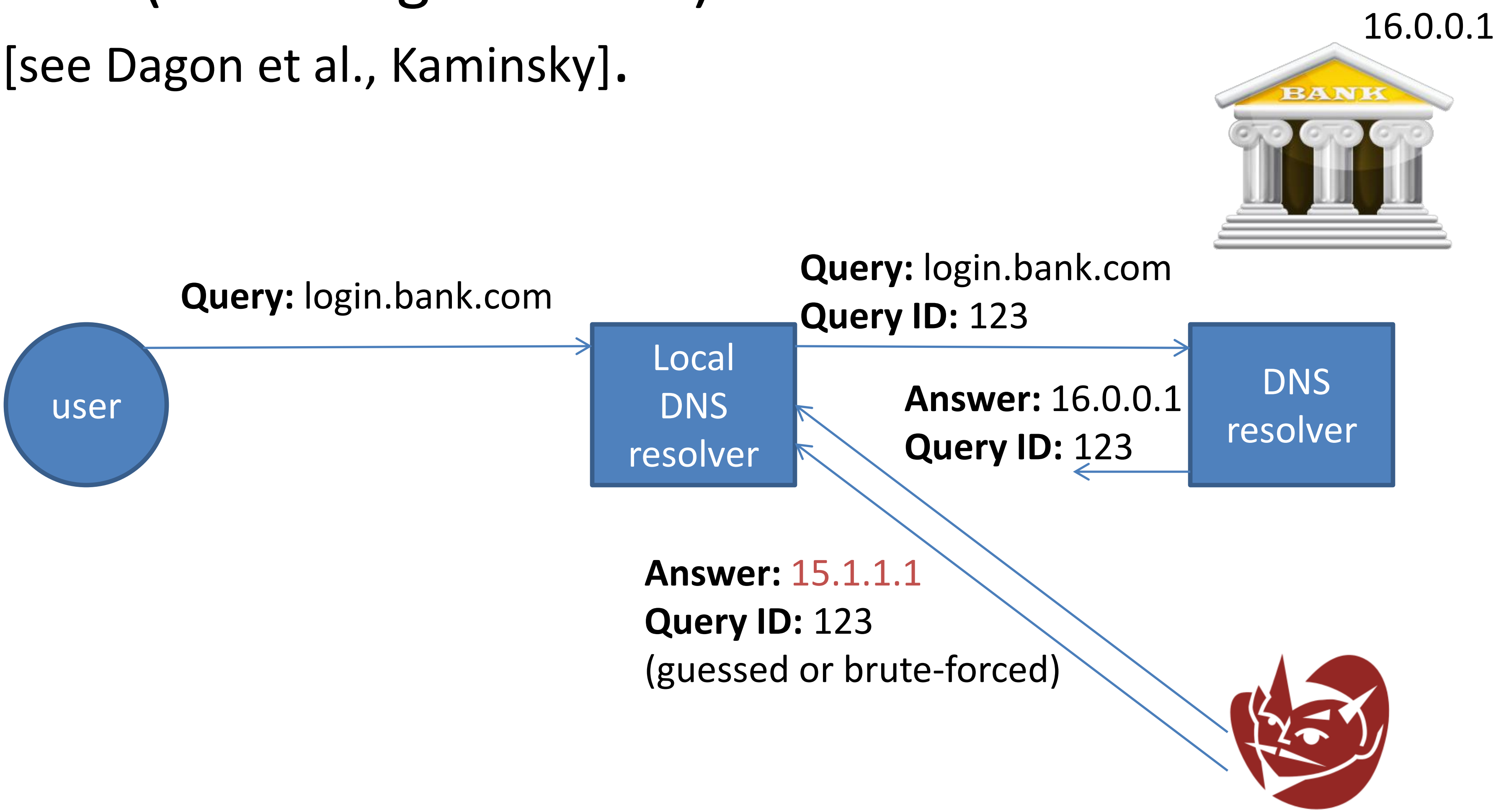
[see Dagon et al., Kaminsky].



Problems: Authenticity (cont.)

DNS (in its original form) is also vulnerable

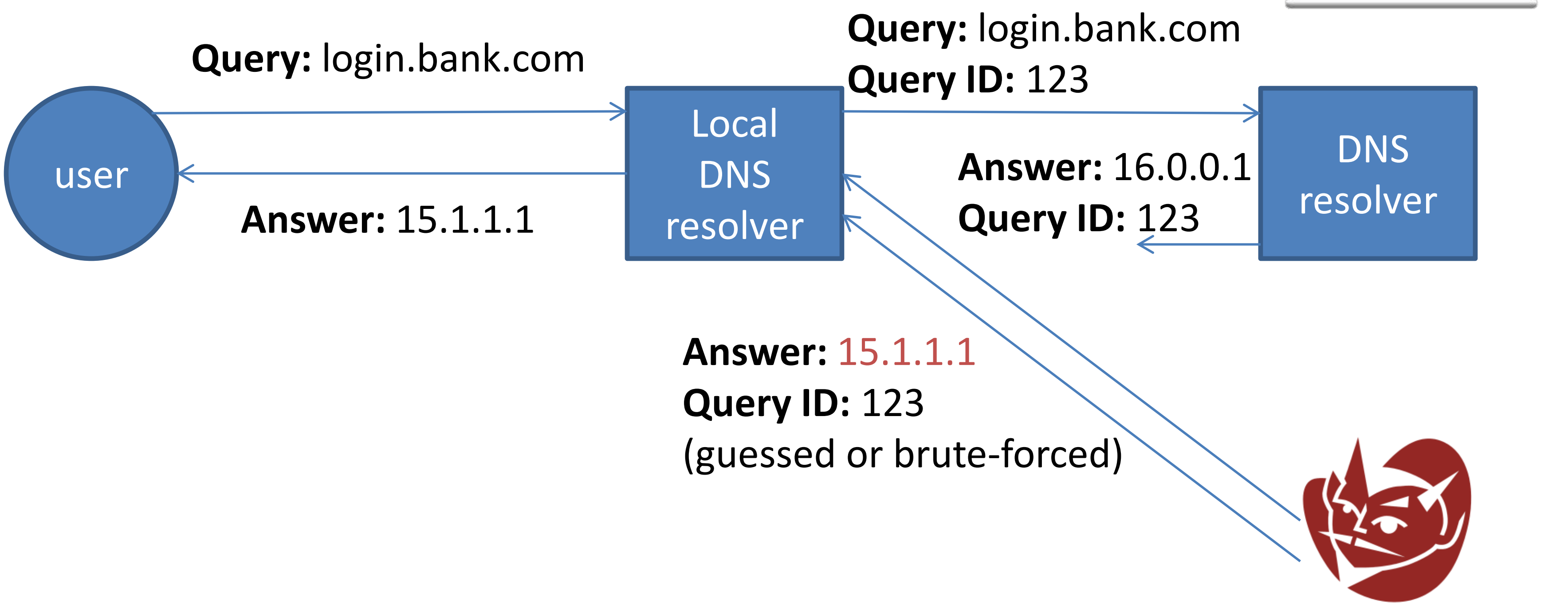
[see Dagon et al., Kaminsky].



Problems: Authenticity (cont.)

DNS (in its original form) is also vulnerable

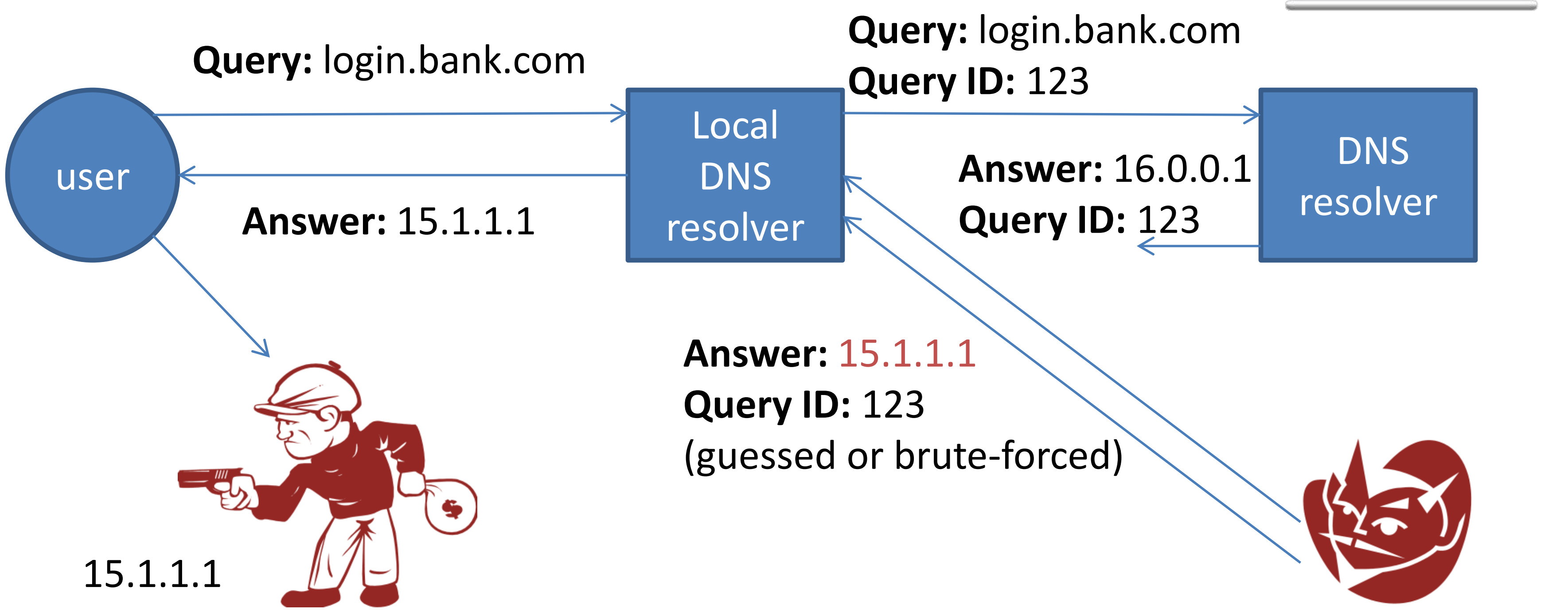
[see Dagon et al., Kaminsky].



Problems: Authenticity (cont.)

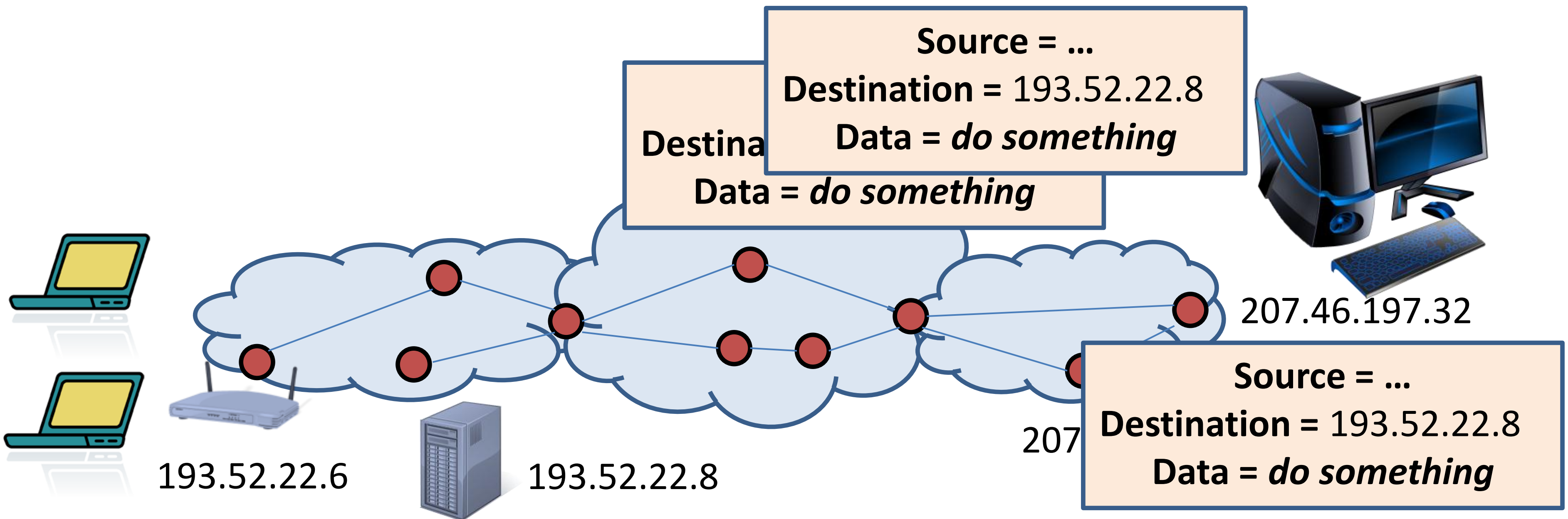
DNS (in its original form) is also vulnerable

[see Dagon et al., Kaminsky].



Problems: Availability

- Any sender (or group of senders, e.g., botnet) may be able to contact any potential target.
- It may cause the target to commit some resources and do some work.



Problems: Availability (cont.)

Corrigendum- Most Urgent

GOVERNMENT OF PAKISTAN
PAKISTAN TELECOMMUNICATION AUTHORITY
ZONAL OFFICE PESHAWAR
Plot-11, Sector A-3, Phase-V, Hayatabad, Peshawar.
Ph: 091-9217279- 5829177 Fax: 091-9217254
www.pta.gov.pk

NWFP-33-16 (BW)/06/PTA February ,2008

Subject: **Blocking of Offensive Website**

Problems: Availability (cont.)

- The blocking order focused on 208.65.153.238, 208.65.153.253, and 208.65.153.251.
- YouTube advertised the range 208.65.152.0**/22** (2^{10} IP addresses with top 22 bits in common).
- Pakistan telecom advertised the more specific range 208.65.153.0**/24** (2^{8} IP addresses).

Problems: Availability (cont.)

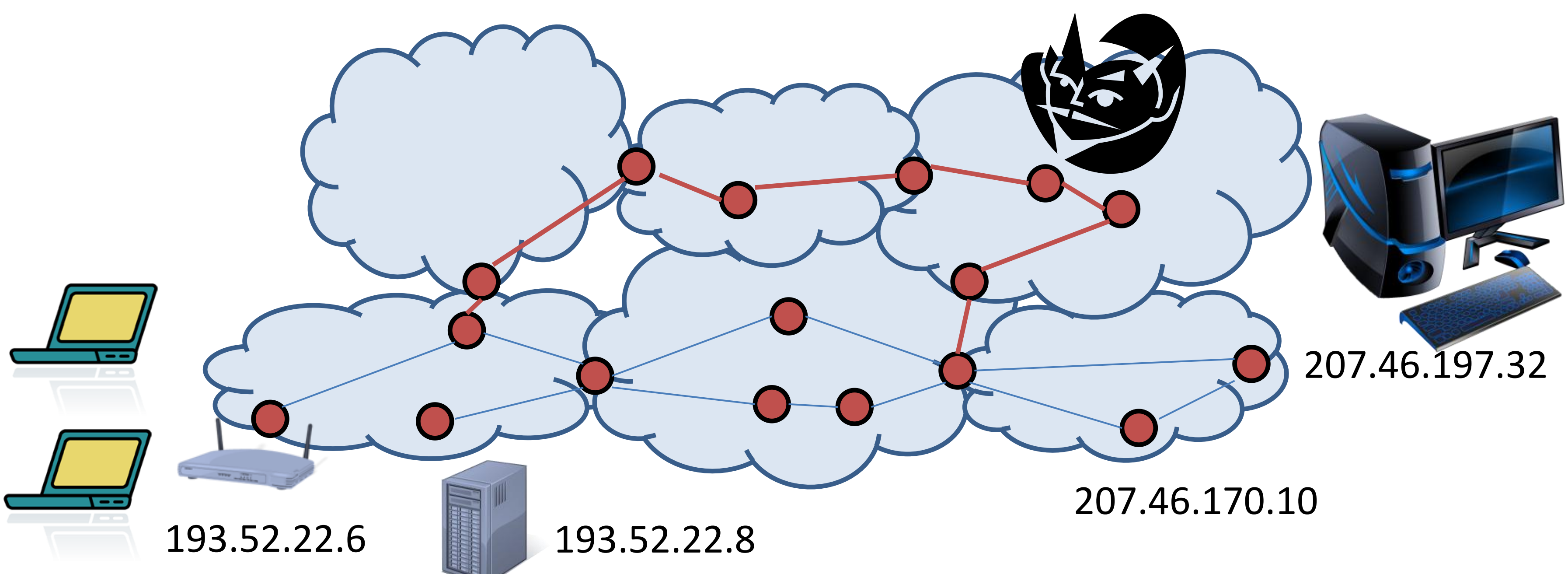
- The blocking order focused on 208.65.153.238, 208.65.153.253, and 208.65.153.251.
- YouTube advertised the range 208.65.152.0**/22** (2^{10} IP addresses with top 22 bits in common).
- Pakistan telecom advertised the more specific range 208.65.153.0**/24** (2^{8} IP addresses).

$\Rightarrow$ Within two minutes, *everyone* sent traffic for 208.65.153.238, 208.65.153.253, and 208.65.153.251 to Pakistan.

$\Rightarrow$ The outage lasted over two hours.

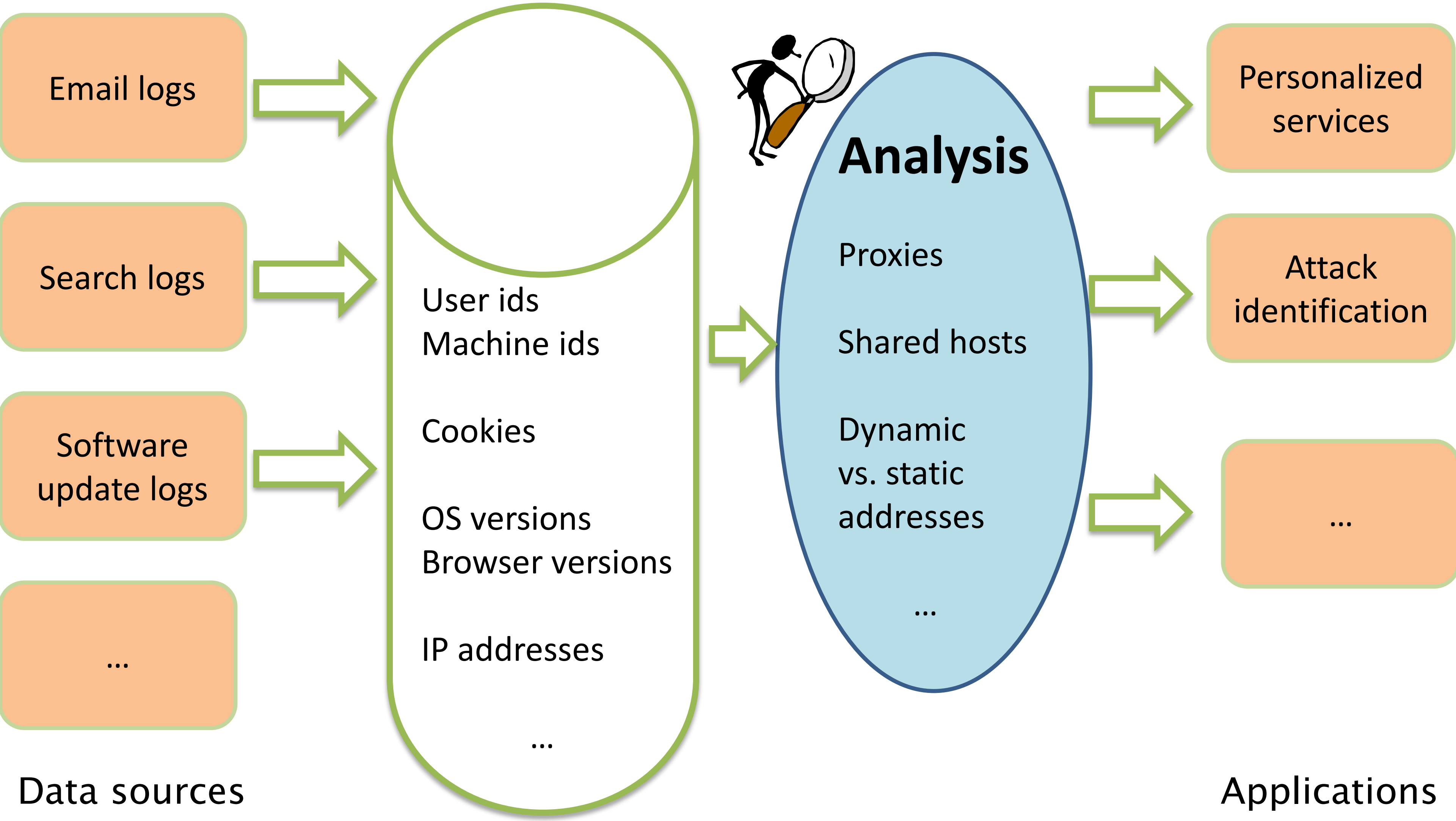
Problems: Secrecy

- Intermediaries see messages.
- Advertisement of false routes can allow unintended intermediaries.



Tracking

Lack of authenticity does not mean perfect anonymity!



A recent example: HostTracker

[with Xie and Yu]

Input: Hotmail user-login trace for one month.

- 550 million user IDs.
- Many of them botnet-created.

e_1: <Alice, IP_1, t_1>
e_2: <Alice, IP_1, t_2>
e_3: <Alice, IP_2, t_3>
…

A recent example: HostTracker

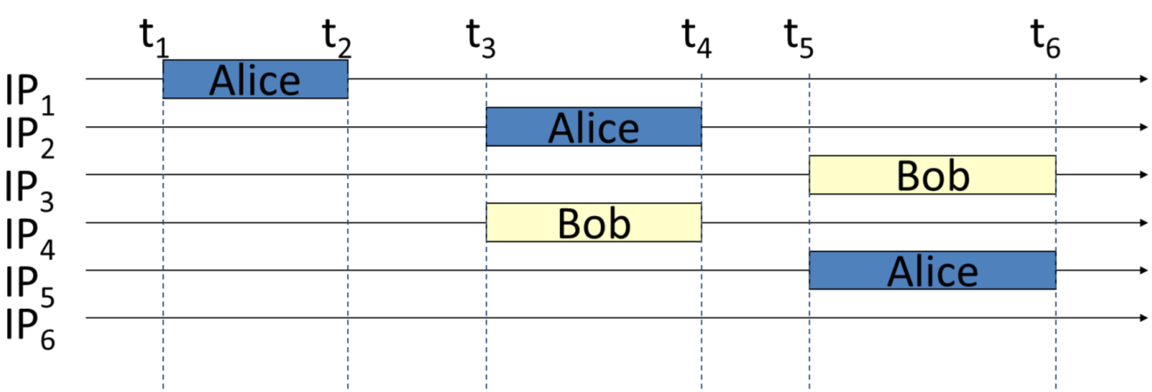
[with Xie and Yu]

Input: Hotmail user-login trace for one month.

- 550 million user IDs.
- Many of them botnet-created.

Output: host-IP bindings over time.

e_1: <Alice, IP_1, t_1>
e_2: <Alice, IP_1, t_2>
e_3: <Alice, IP_2, t_3>
…



Alice's host
IP_1: [t1, t2]
IP_2: [t3, t4]
IP_5: [t5, t6]

Bob's host
IP_4: [t3, t4]
IP_3: [t5, t6]

A recent example: HostTracker

[with Xie and Yu]

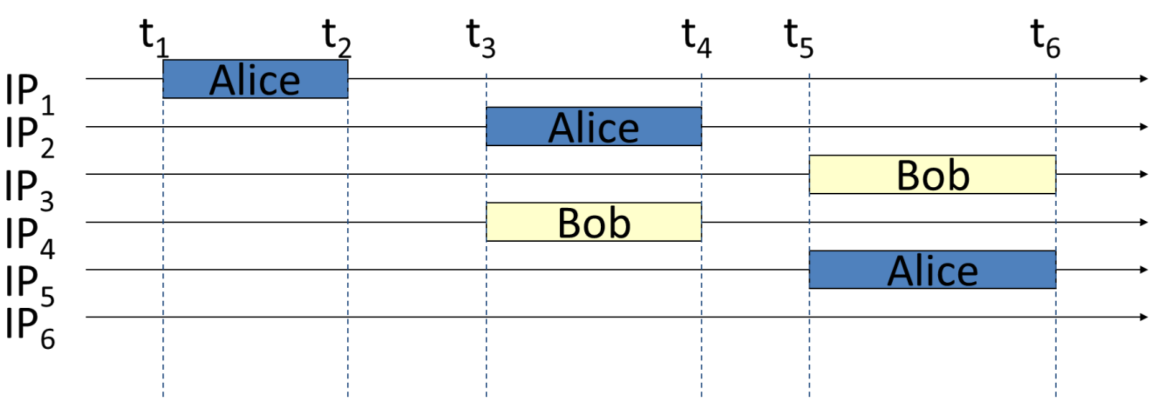
Input: Hotmail user-login trace for one month.

- 550 million user IDs.
- Many of them botnet-created.

e_1: <Alice, IP_1, t_1>
e_2: <Alice, IP_1, t_2>
e_3: <Alice, IP_2, t_3>
…

Output: host-IP bindings over time.

- Identified 220 million hosts.
- Validated accurate (~ 90%) against Windows Update data.
- 76% of login events attributed to hosts.



Alice's host
IP_1: [t1, t2]
IP_2: [t3, t4]
IP_5: [t5, t6]

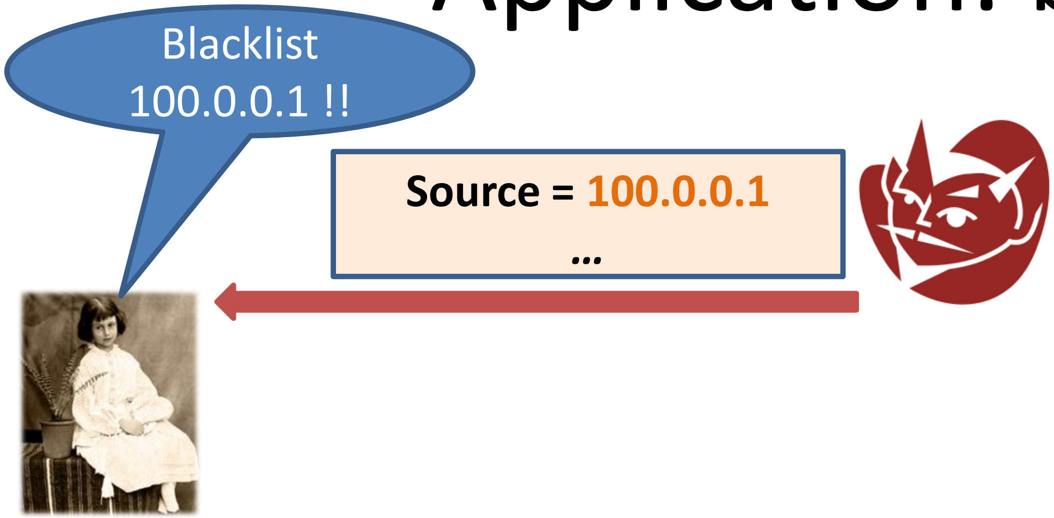
Bob's host
IP_4: [t3, t4]
IP_3: [t5, t6]

Application: blacklists

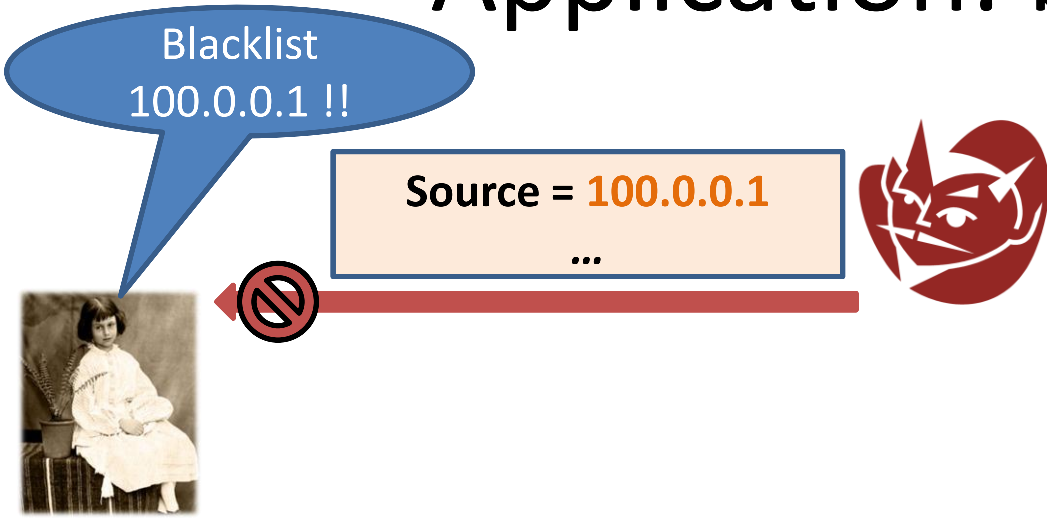
Source = **100.0.0.1**

…

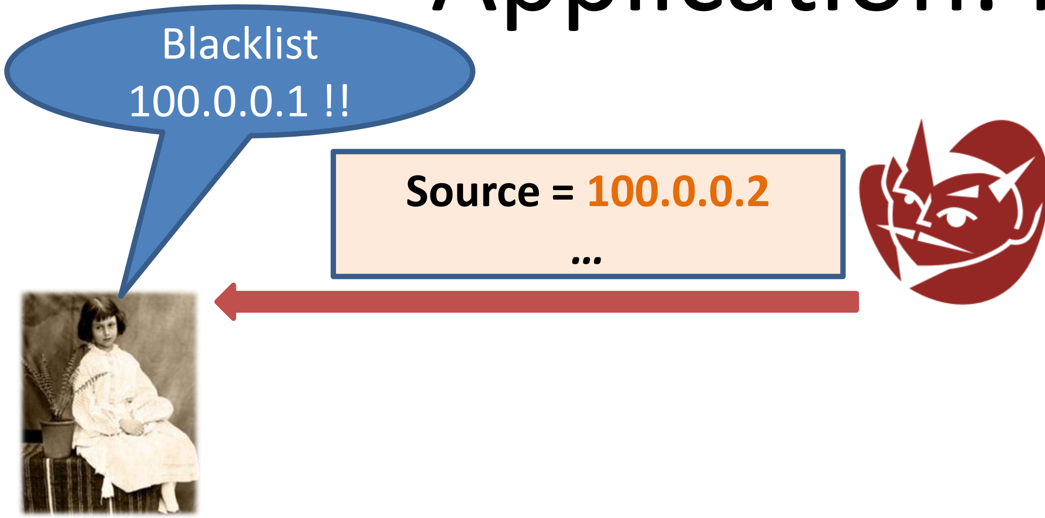
Application: blacklists



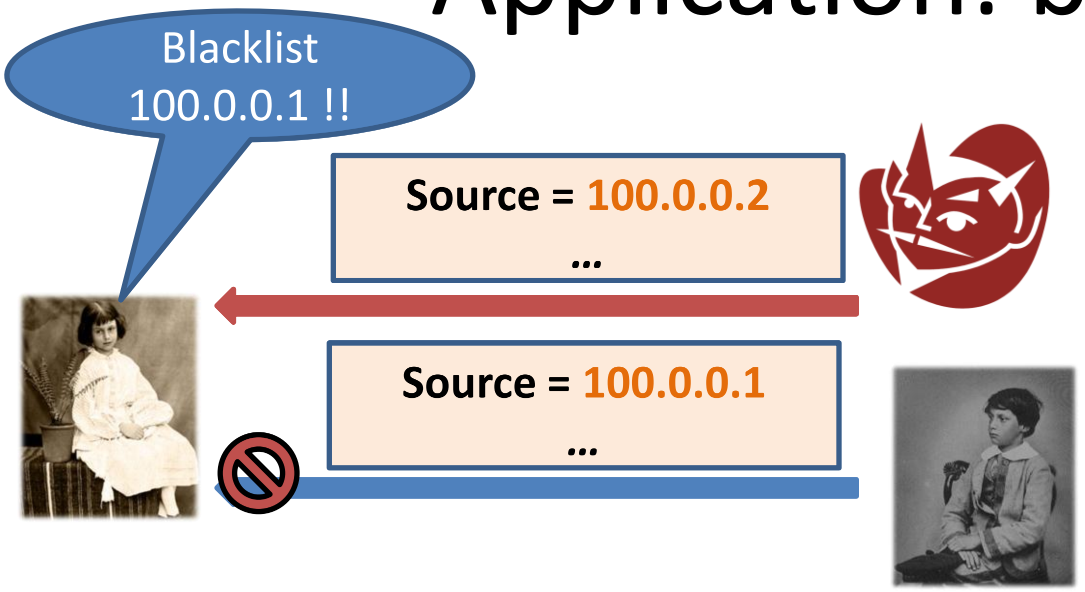
Application: blacklists



Application: blacklists



Application: blacklists

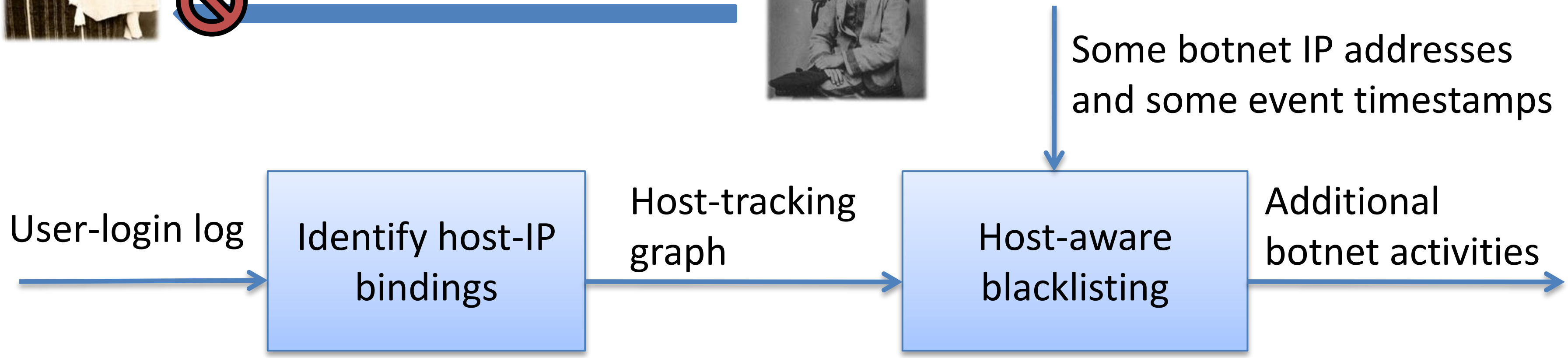


Application: blacklists

Tracking hosts can help reduce such false positives.

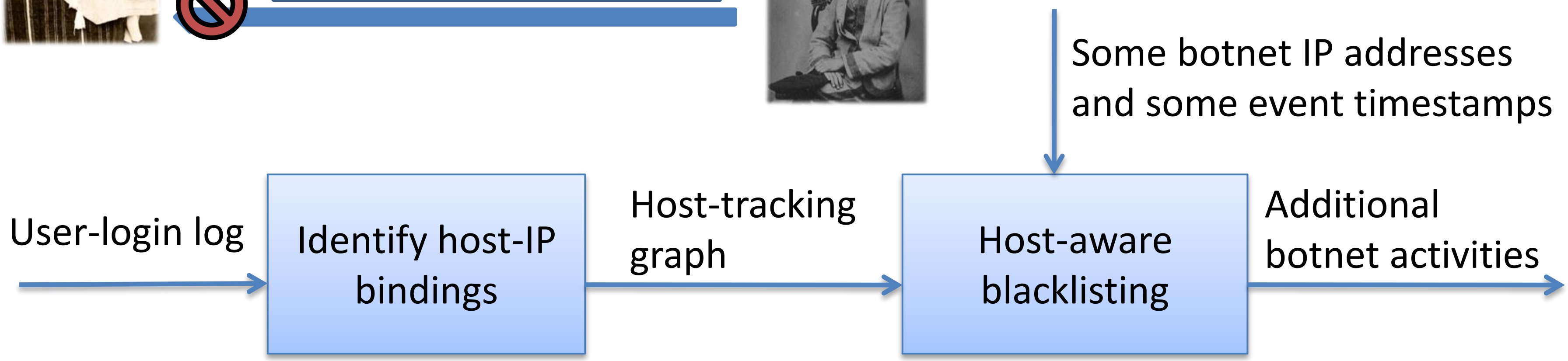
Application: blacklists

Tracking hosts can help reduce such false positives.



Application: blacklists

Tracking hosts can help reduce such false positives.



An experiment: Application to Hotmail bot blocking	**# of malicious blocked users**	**False positives**
Block IP / one hour	28 million	34%
Blacklist host / one hour	16 million	5%

Other fingerprints

Other information, besides logins, can identify users and hosts. E.g.:

- Cookies
- Browser user-agent strings
  - E.g., "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; InfoPath.3; MS-RTC LM 8; Zune 4.0)"
  - 19 million distinct ones seen in our logs [with Xie, Yen, and Yu].

These fingerprints are less secure but useful.

Other fingerprints (cont.)

Browser characteristics
have > 18 bits of entropy:

“if we pick a browser at random, at best we expect that only one in 286,777 other browsers will share its fingerprint”.

[Eckersley, EFF]
http://panopticlick.eff.org

Other fingerprints (cont.)

Browser characteristics have > 18 bits of entropy:

“if we pick a browser at random, at best we expect that only one in 286,777 other browsers will share its fingerprint”.

[Eckersley, EFF]
http://panopticlick.eff.org

Race Is On to “Fingerprint” Phones, PCs

December 1, 2010
by Julia Angwin and Jennifer Valentino-DeVries
Reporters, The Wall Street Journal

IP_1

Login doggy@kennel.com

Logout doggy@kennel.com

Search for “séminaire”

Login doggy@kennel.com

IP_1

Login doggy@kennel.com
User agent = Mozilla/4.0 (...)
Logout doggy@kennel.com

Search for “séminaire”
User agent = Mozilla/4.0 (...)

IP_1

Search for “nice dog food”
User agent = Mozilla/4.0 (...)

Search for “séminaire”
User agent = Mozilla/4.0 (...)

IP_1

Login doggy@kennel.com
User agent = Mozilla/4.0 (…)
Logout doggy@kennel.com

IP_2

Search for “séminaire”
User agent = Mozilla/4.0 (…)

IP_1

Search for “nice dog food”
User agent = Mozilla/4.0 (…)

IP_2

Search for “séminaire”
User agent = Mozilla/4.0 (…)

NOISE TO SIGNAL
Rob Cottingham

“On Facebook, 273 people know I’m a dog.
The rest can only see my limited profile.”

Source: socialsignal.com

Using cryptography

(preliminaries)

Cryptography to the rescue?

- Cryptography provides attractive techniques for improving network security.

But:

- Cryptography is not a panacea.
- It is not always perfect.
- It can be used inappropriately.
- And there are other techniques, such as firewalls and honeypots.

Communication and cryptography

- Many network protocols aim to achieve stronger security by the use of cryptography:
  - IPSec
  - S-BGP
  - DNSSEC
  - SSL (or TLS)
  - HTTPS
  - …

Communication with shared-key cryptography

For confidential messages

- The sender encrypts with a shared key K.
- The recipient decrypts with the same key K.

For messages with integrity

- The sender includes MACs with a shared key K.
- The recipient checks MACs with the same key K.

Communication with shared-key cryptography

For confidential messages

- The sender encrypts with a shared key K.
- The recipient decrypts with the same key K.

For messages with integrity

- The sender includes MACs with a shared key K.
- The recipient checks MACs with the same key K.

For both

- The proper order of signatures and encryptions is a subject of debate and confusion.
- And there are also *authenticated encryption* schemes.
- Encryption keys and MAC keys should be different.
- Each direction of communication may have its own keys.

Communication with public-key cryptography

For confidential messages

- The sender encrypts with the recipient’s public encryption key.
- The recipient decrypts with its secret decryption key.

For messages with integrity

- The sender signs with its secret signature key.
- The recipient checks with the corresponding public key.

Communication with public-key cryptography

For confidential messages

- The sender encrypts with the recipient's public encryption key.
- The recipient decrypts with its secret decryption key.

For messages with integrity

- The sender signs with its secret signature key.
- The recipient checks with the corresponding public key.

For both

- The proper order of signatures and encryptions is a subject of debate and confusion.
- If the sender should prove knowledge of the plaintext, sign before encrypting.
- Encryption keys and signature keys should be different.

Remaining problems (many!)

- Associating keys with principals
- Performance
- Correctness (e.g., signing the right fields)
- Many important specifics:
  - multiple messages, connections, and sessions,
  - timestamps, nonces, sequence numbers,
  - key identifiers,
  - compression and padding,
  - and peripheral concerns such as key storage.

See the next lecture.

Example: protecting search

https://encrypted.google.com/

Example: protecting search

https://encrypted.google.com/

with a key pair for
asymmetric encryption

Example: protecting search

https://encrypted.google.com/

with a key pair for
asymmetric encryption

A problem: how does Alice reliably learn Google's public key?

(more on this later)

Example: protecting search

(simplified, first take)

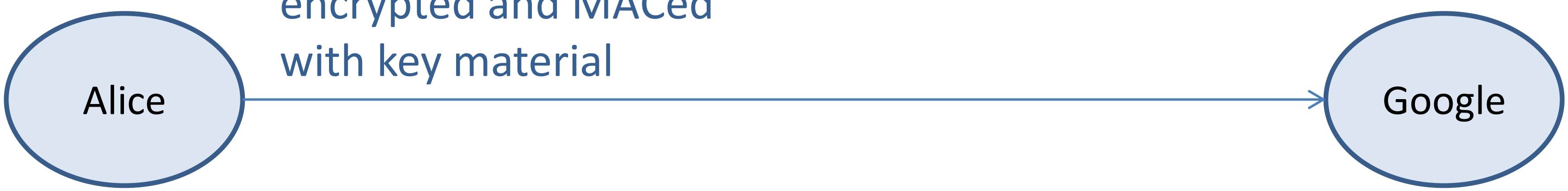
https://encrypted.google.com/

with a key pair for
asymmetric encryption

Example: protecting search
(simplified)

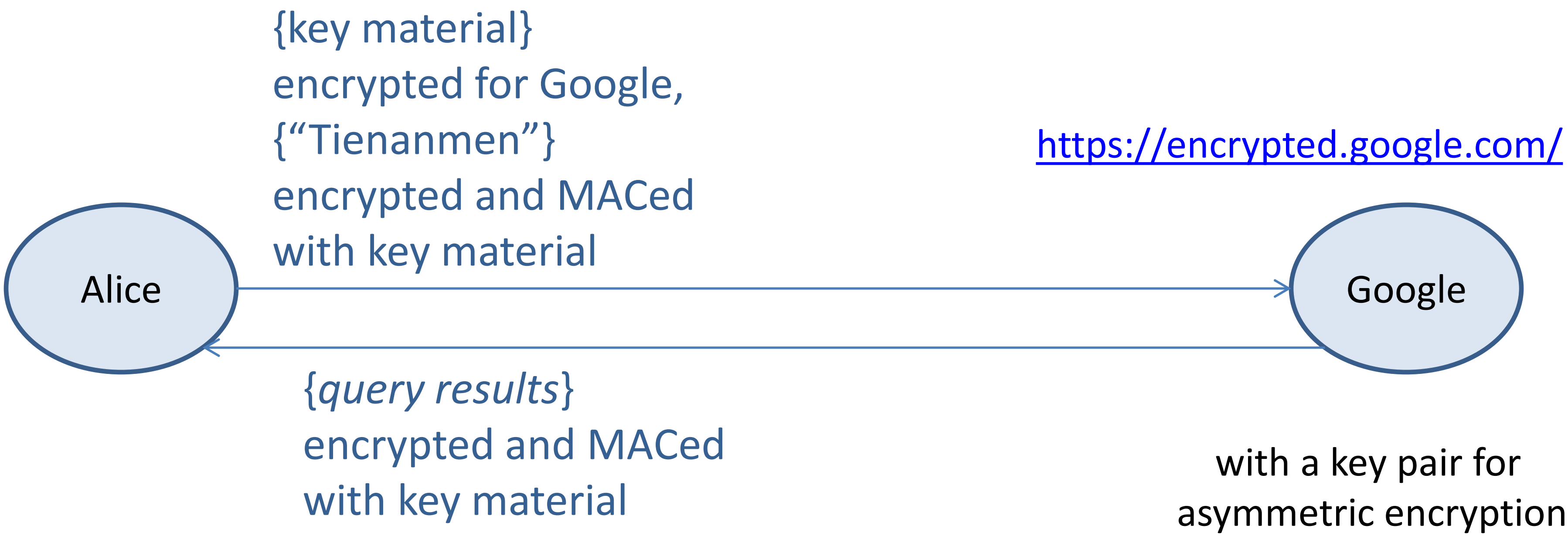
{key material}
encrypted for Google,
{“Tienanmen”}
encrypted and MACed
with key material

https://encrypted.google.com/

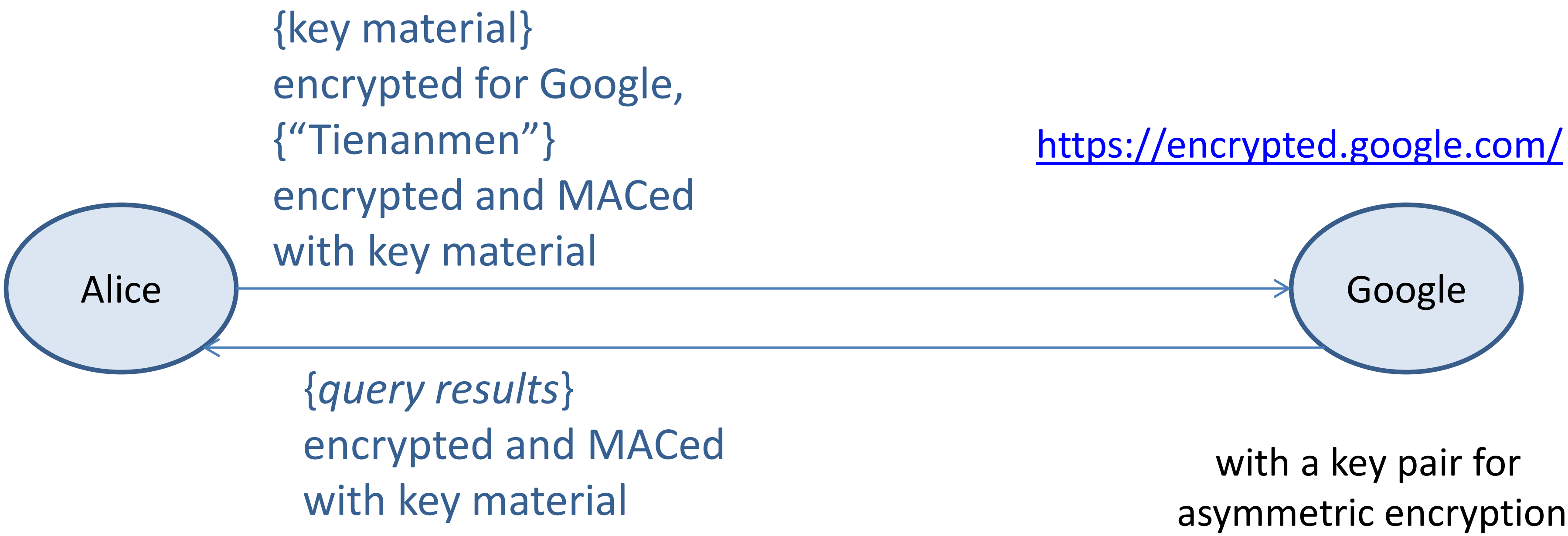


with a key pair for
asymmetric encryption

Example: protecting search
(simplified)

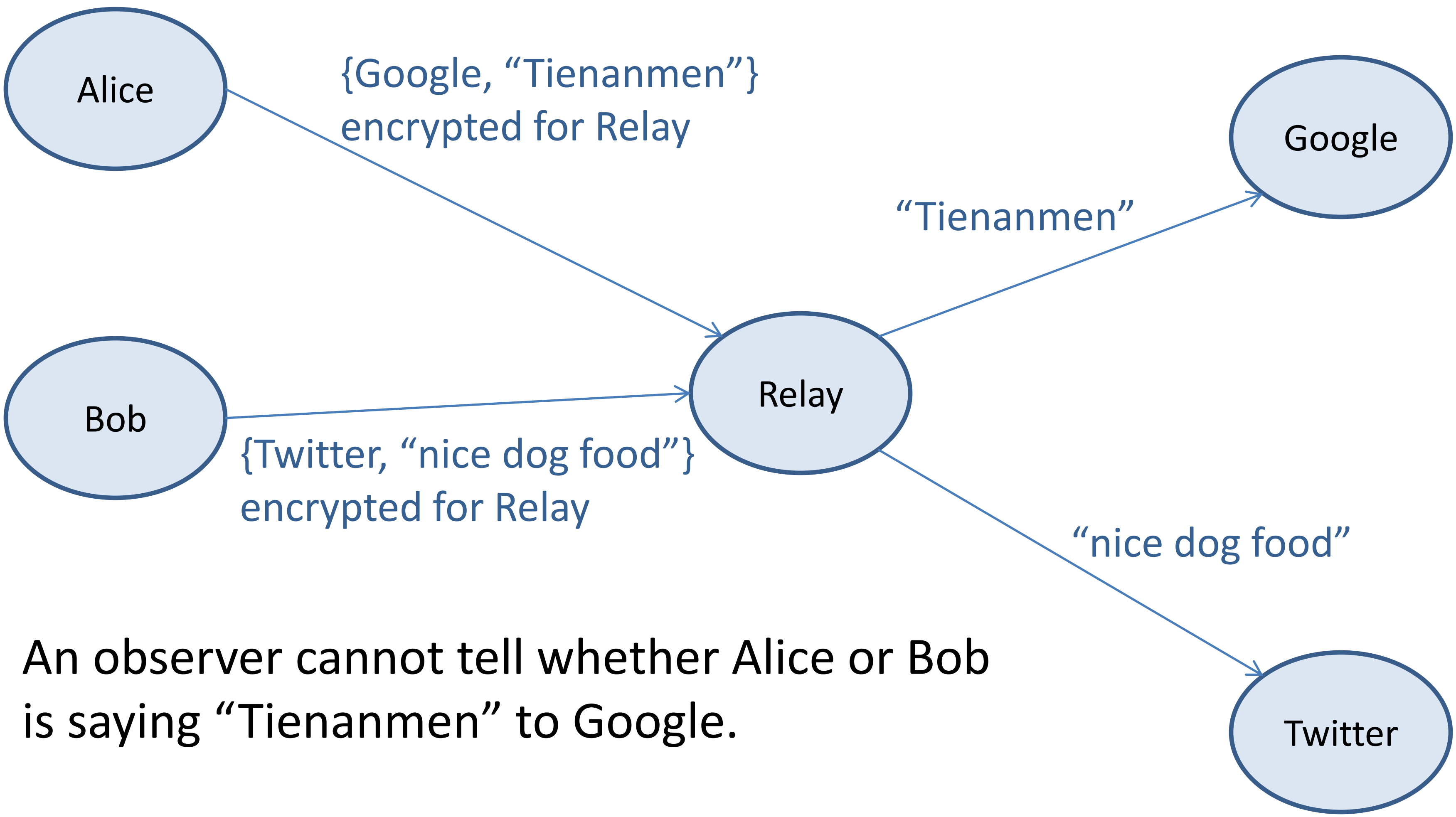


Example: protecting search
(simplified)



Still an issue: network operators and intermediaries may identify the interlocutors and analyze traffic.

Example: anomyzing by a relay
(simplified)

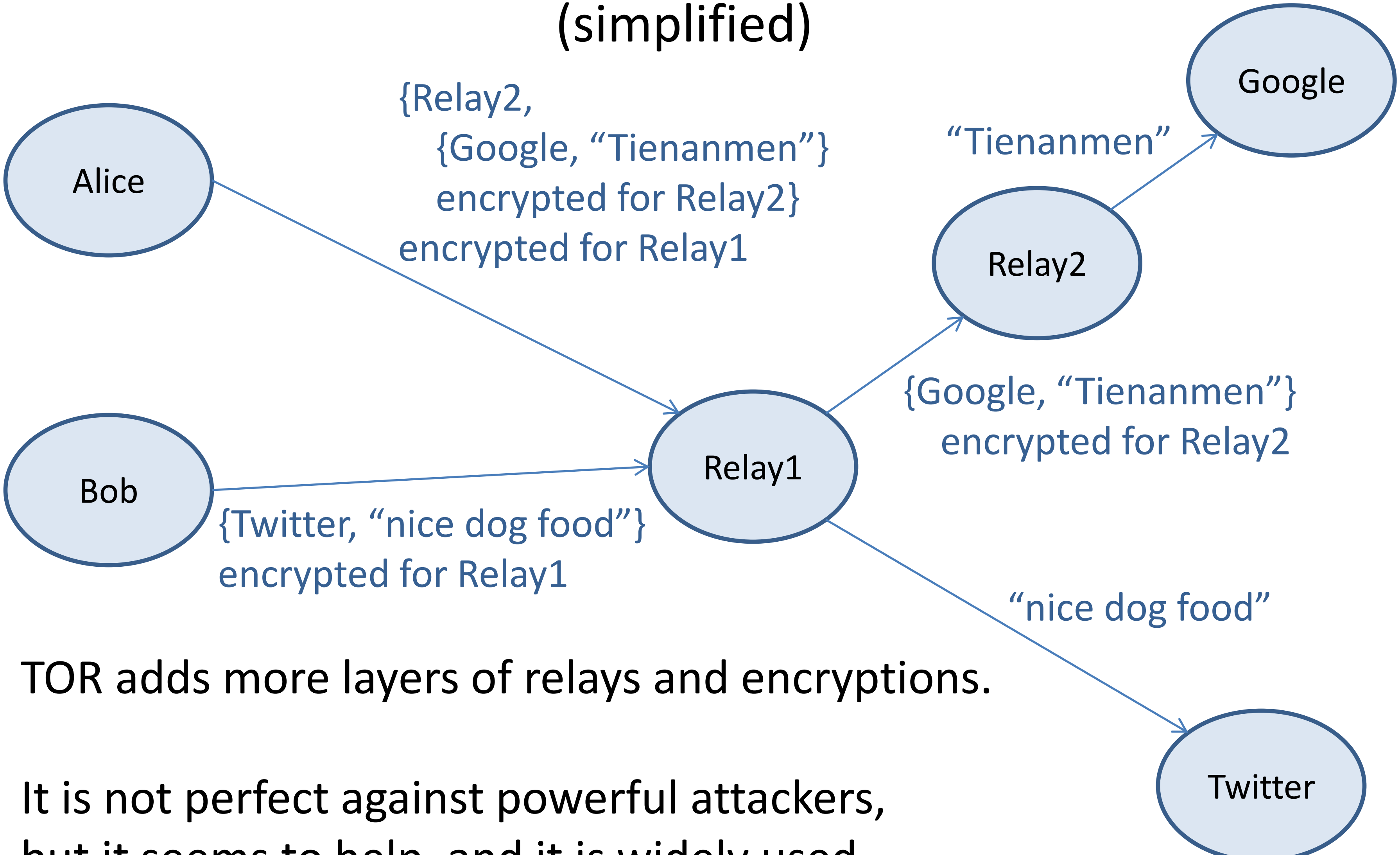


An observer cannot tell whether Alice or Bob is saying “Tienanmen” to Google.

But a corrupt Relay can reveal everything.

Example: anomyzing by TOR

(simplified)



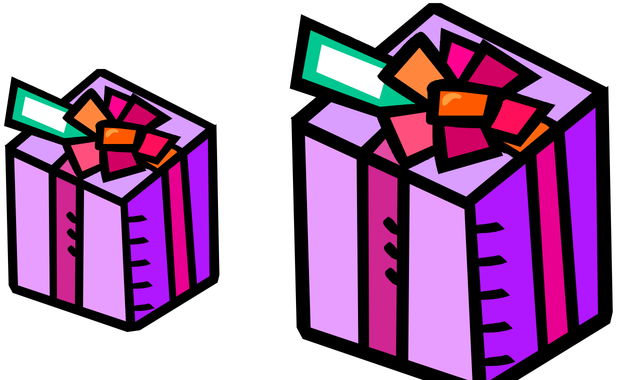
TOR adds more layers of relays and encryptions.

It is not perfect against powerful attackers,
but it seems to help, and it is widely used
(est. 500,000 daily users in 2010).

Side channels

Even with encryption, the timing, the number and size of packets, etc., may be exploited.

- E.g., Sun et al. (2002) identified (static) Web pages by their number of objects and their sizes.
- E.g., Chen et al. (2010) attacked several Web applications despite encryption:
  - search engines,
  - online health sites,
  - financial services.



Attack on investment service

[from Chen et al.]

Mutual Funds

Fund A

Price $52.85   09/09/2009

Value   Quantity
$12345   234

Fund B

Price $32.15   09/09/2009

Value   Quantity
$12330   384

Fund C

Price $28.80   09/09/2009

Value   Quantity
$11111   386

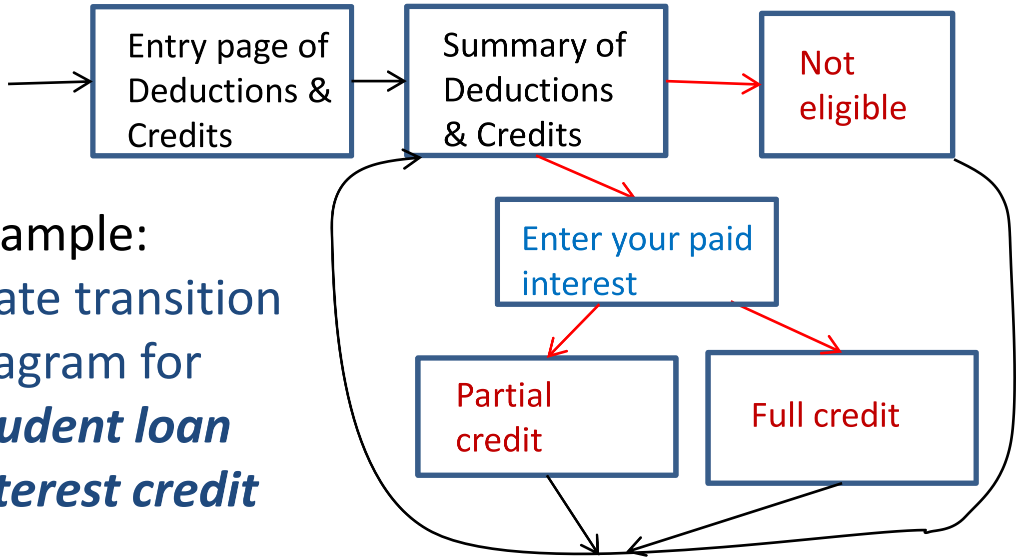
Each price history curve is a GIF from MarketWatch, which anyone can obtain.

⇒ *Just compare image sizes to identify the funds!*

Attack on tax-filing service

[from Chen et al.]

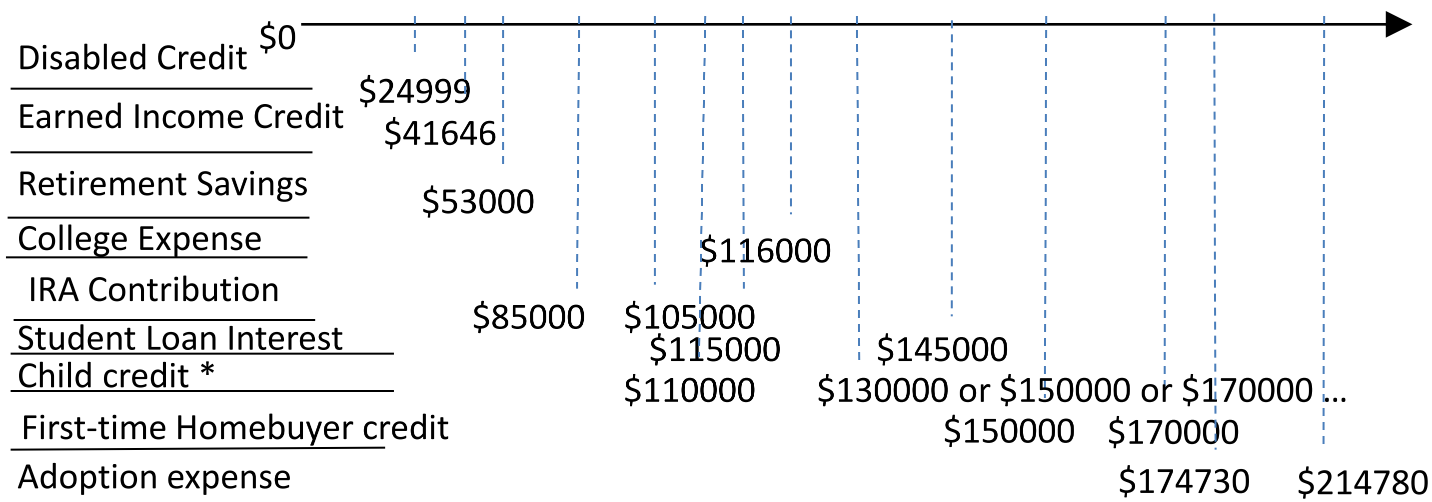
Example:
State transition diagram for ***Student loan interest credit***



According to tax laws:

- "Full credit" implies
  AGI < \$115,000
- "Partial credit" implies
  \$115,000 ≤ AGI < \$145,000
- "Not eligible" implies
  AGI ≥ \$145,000

A subset of identifiable income thresholds



Certification authorities

Certification authorities (CAs)

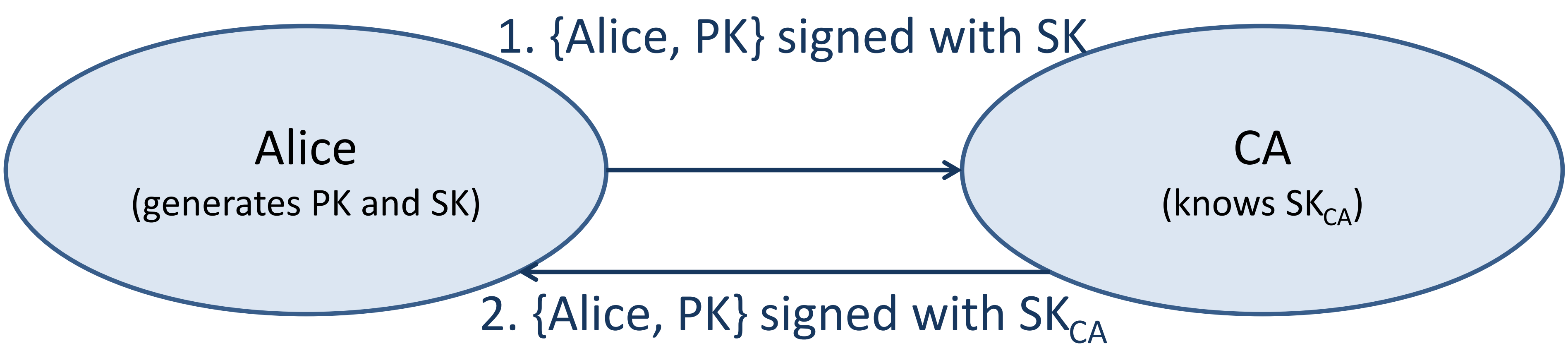
- If Alice sends its public key to Bob, how can Bob know that it is really Alice's?
- A CA is a trusted third party that solves this problem by signing Alice's public key.
- The key may be
  - a signature-verification key,
  - an encryption key,
  - both.

- Bob should check the certificate!

Obtaining a certificate

(one method)

- Alice generates a key pair (PK, SK).
- Alice signs PK and identity information with SK.
- The CA does some verifications.
  (It may refuse a certificate to A1ice.)
- The CA signs PK and the identity information.
- Alice checks CA's certificate.



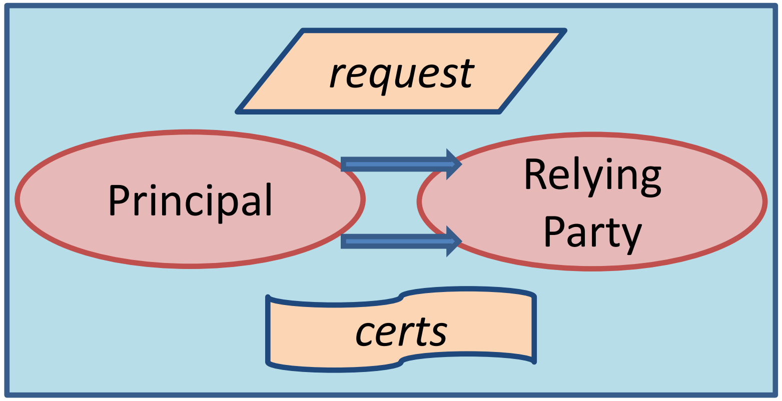
Certificate distribution

Alice may show (*push*) its certificate when it uses its keys.

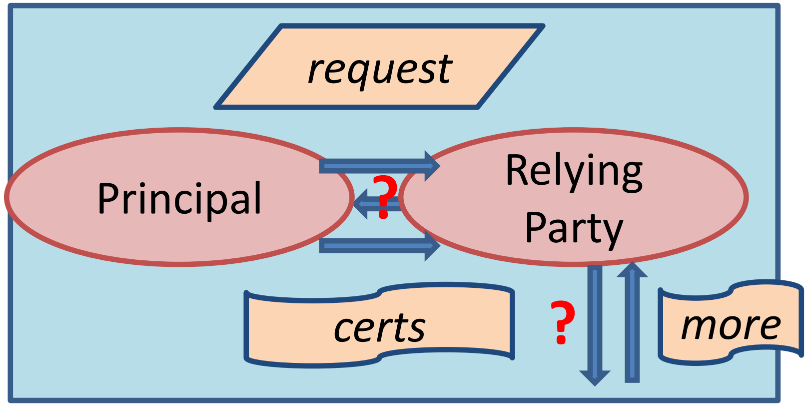
Or relying parties may request (*pull*) the certificate:

- from CA,
- from other directories,
- from Alice.

push: Principals present certificates proactively.



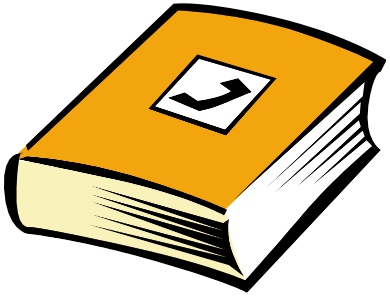
pull: Relying parties gather certificates.



“The phonebook CA”

Early on, it was hoped that a simple directory could associate public keys with names.

- The directory could be implemented as a set of certificates, signed with a CA key.
- The CA could be kept off-line, in a safe, most of the time.



Public-key infrastructures (PKIs)

The basic tasks of a PKI are:

- creation of certificates,
- dissemination of certificates,
- renewal of certificates,
- revocation of certificates,
- (sometimes) key escrow and archival.

Who are the CAs (and why)?
Why are the CAs trusted (for this purpose)?

Scaling: certificate chains

Having a single CA is unrealistic beyond small, closed organizations:

- No CA is trusted by everyone for everything.
- A single CA may be a bottleneck.

One solution is to have multiple CAs (perhaps a hierarchy), and to chain certificates:

CA1 certifies Alice
CA2 certifies CA1
...
Root certifies CAn

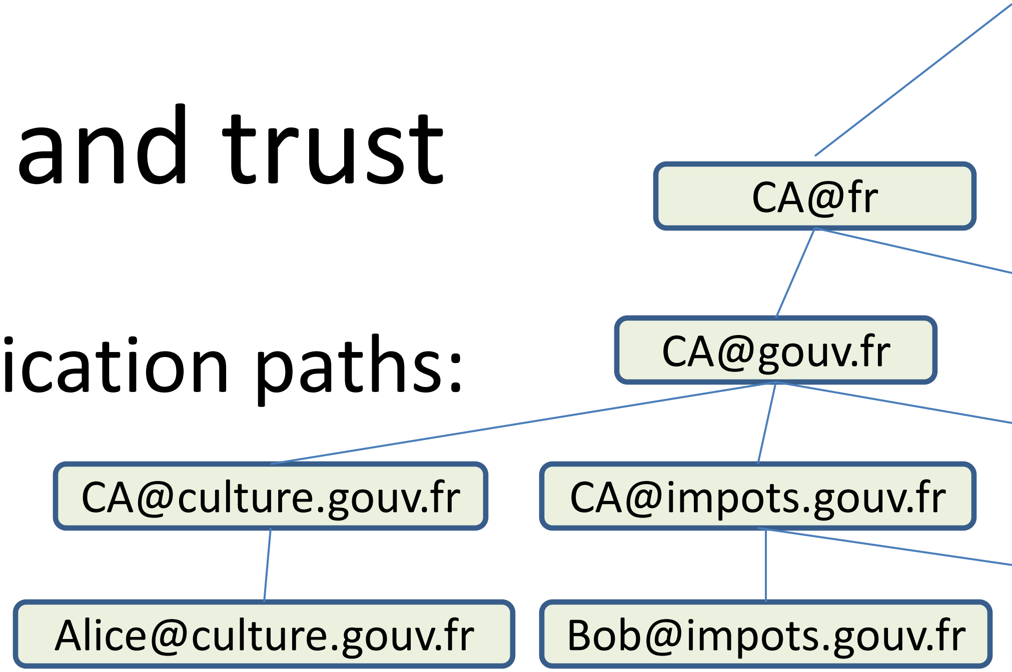
Scaling: names

- Ordinary naming is not a bijection.
  – Who is "John Smith"?
  – Who is "Prince"?
- Many names are not stable.
  $\Rightarrow$ *Early vs. late binding*
- Adding addresses, etc., complicates matters.
- UIDs and other possible forms of names have their own problems.

Names and trust

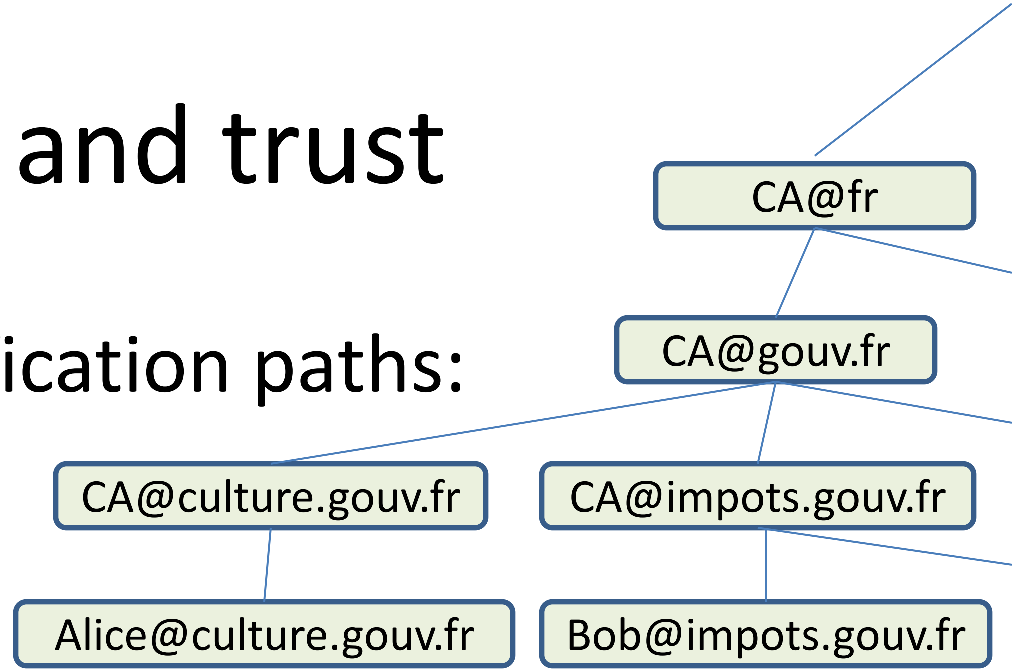
Names may yield certification paths:

E.g., for Alice@culture.gouv.fr,
CA1 is CA@culture.gouv.fr,
CA2 is CA@gouv.fr, and
Bob@impots.gouv.fr trusts it.



- Hierarchical names correspond to hierarchical CAs. (See Privacy Enhanced Email.)

Names and trust



Names may yield certification paths:

E.g., for Alice@culture.gouv.fr,
CA1 is CA@culture.gouv.fr,
CA2 is CA@gouv.fr, and
Bob@impots.gouv.fr trusts it.

- Hierarchical names correspond to hierarchical CAs. (See Privacy Enhanced Email.)
- In *web-of-trust* systems, without hierarchy, names may still relate to trust. (See SDSI.)
  E.g., Bob may be trusted on the key for Bob's attorney.

X.500

X.500 relies on the notion of distinguished names (DNs).
Everything should have a DN.

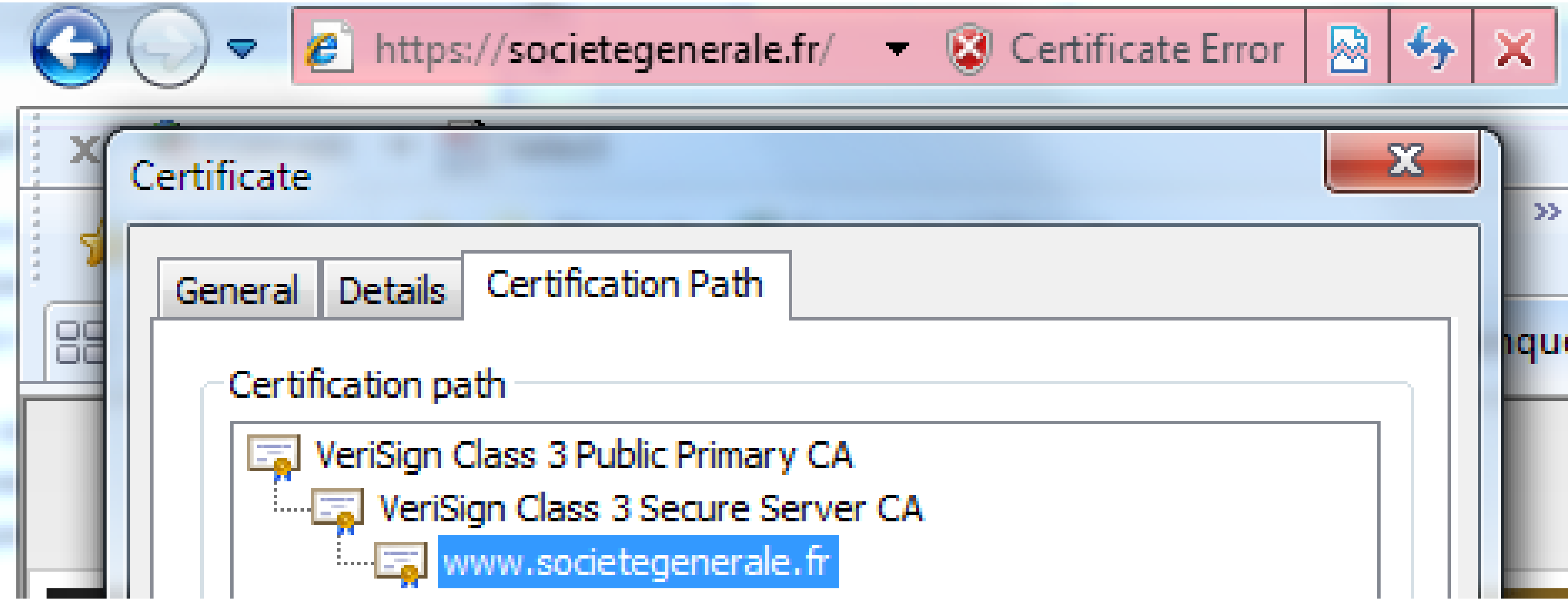
A DN includes:

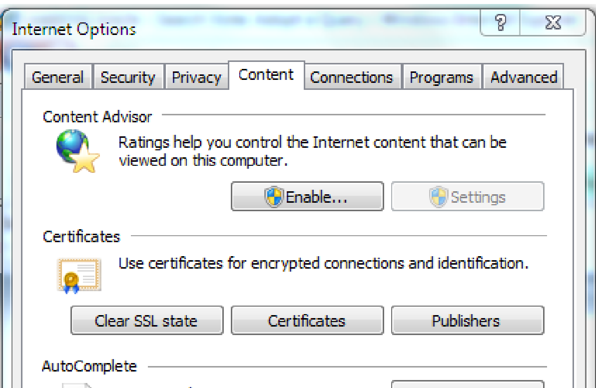
- country,
- state or province,
- locality,
- organization,
- organizational unit,
- common name,
- certificate type,
- email address,
- fields required by signature laws,
- …

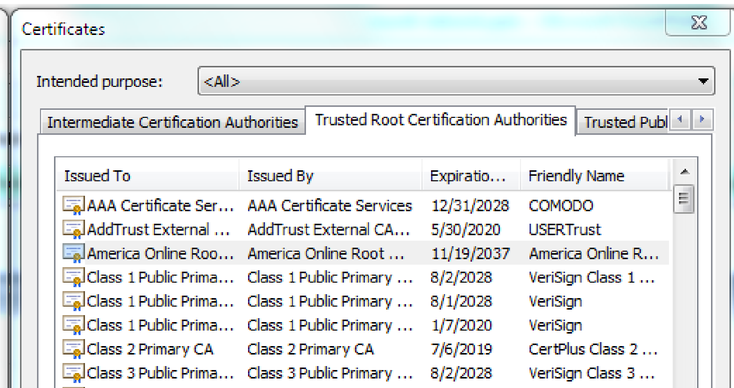
But:
There is no agreement on what these mean.
The specification is vague in various areas.
Implementations are not always consistent.

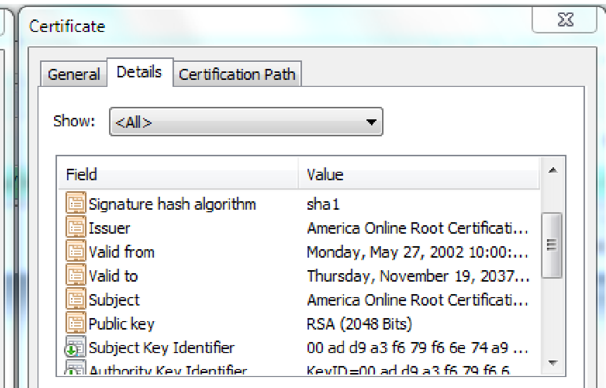
Nevertheless, X.500 is in widespread use.

X.500 in browsers *(go look!)*

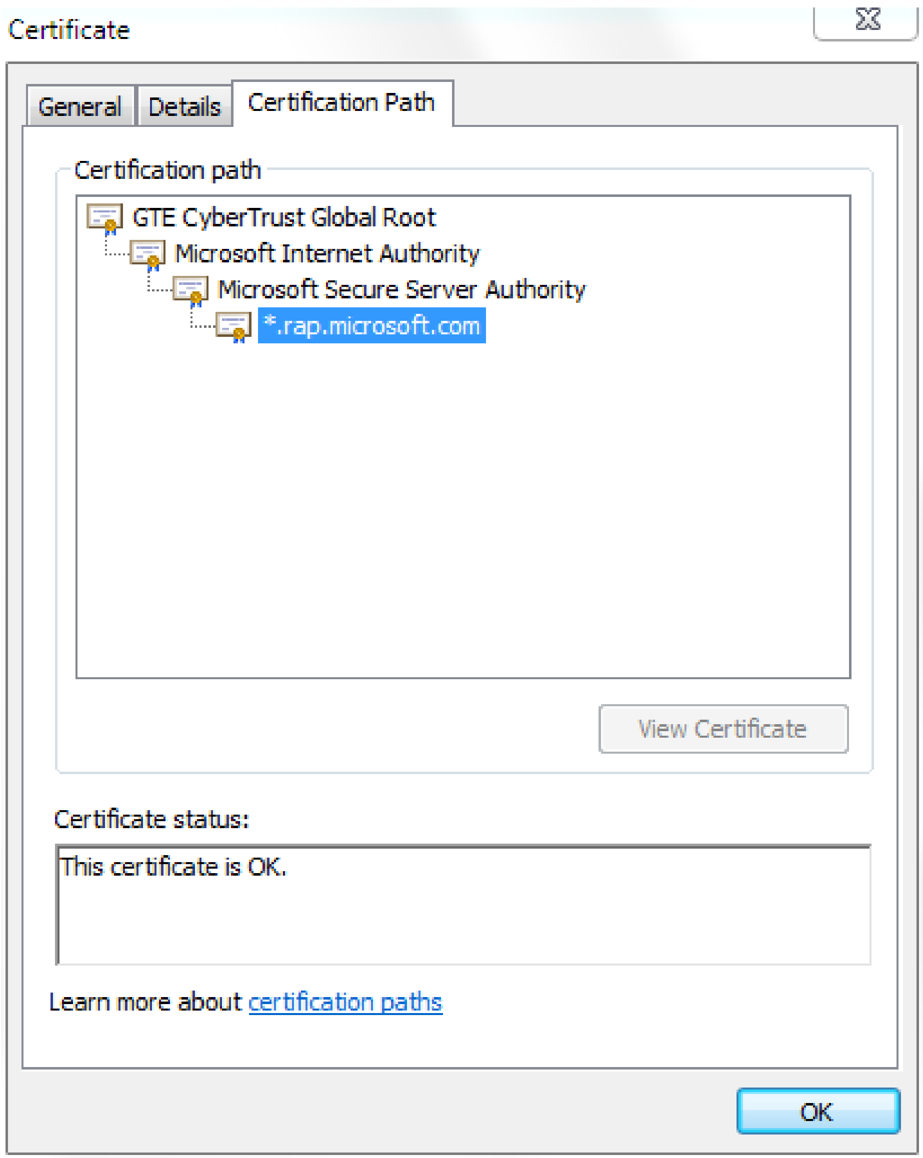








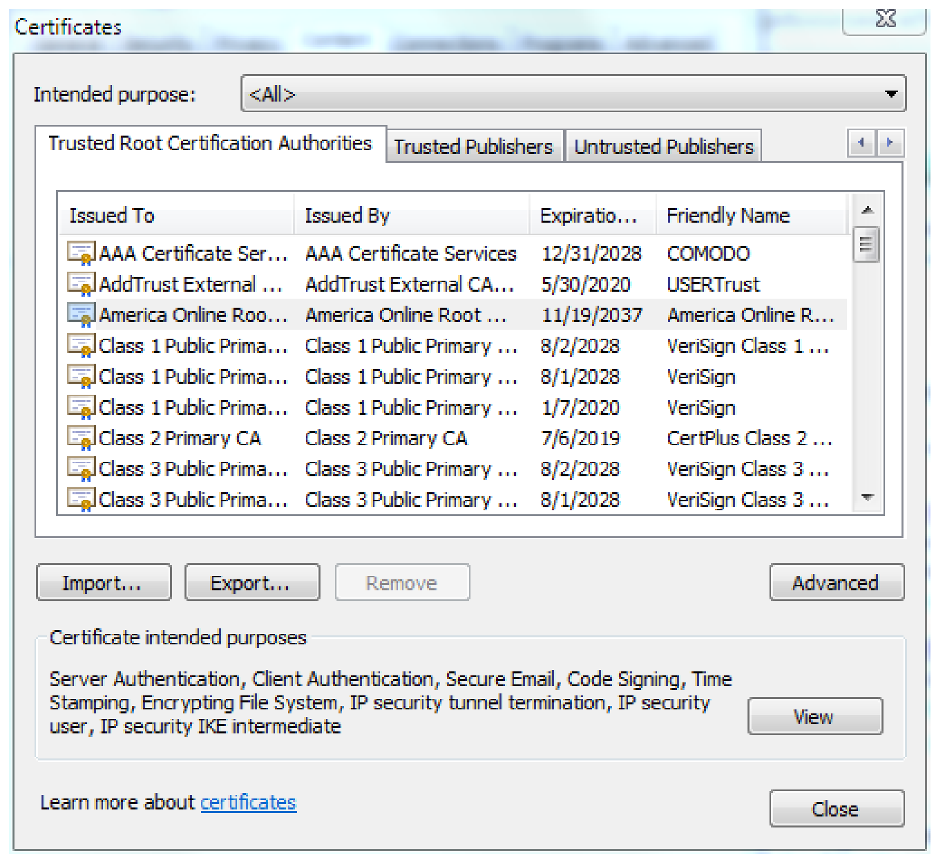
X.500 in browsers (cont.)



Note that a Microsoft browser seems to use GTE CyberTrust for authenticating a Microsoft server.

Some observations
[Eckersley and Burns]

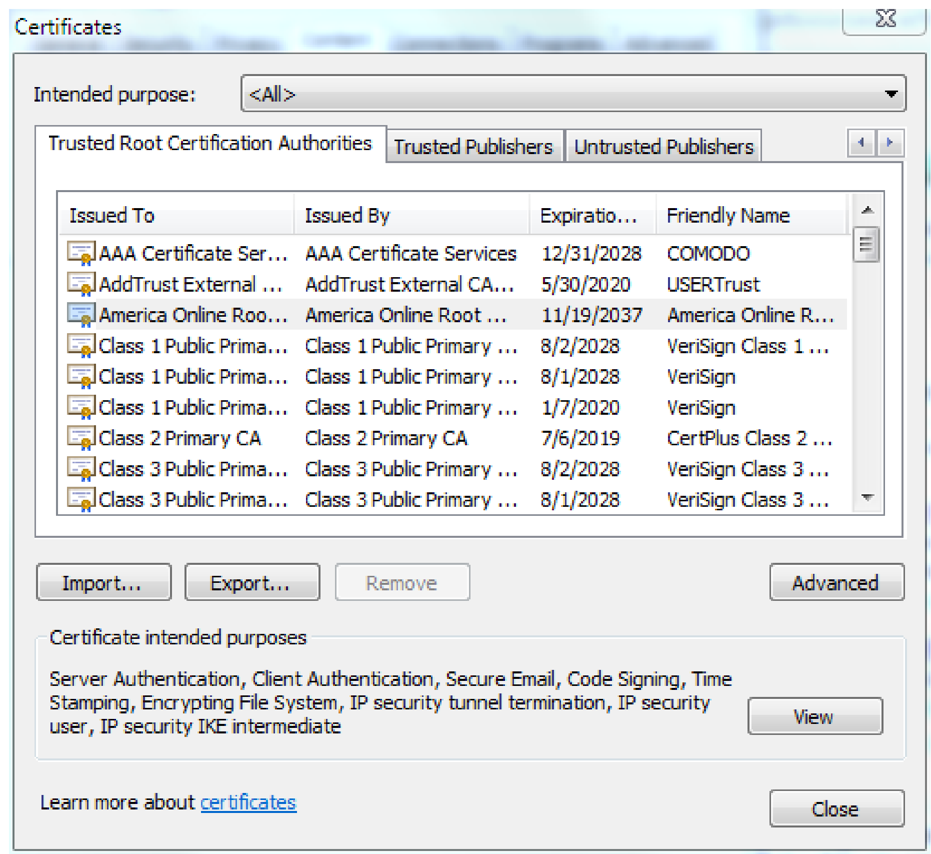
Browsers come with knowledge of some certification authorities and more get added.



Some observations
[Eckersley and Burns]

Browsers come with knowledge of some certification authorities and more get added.

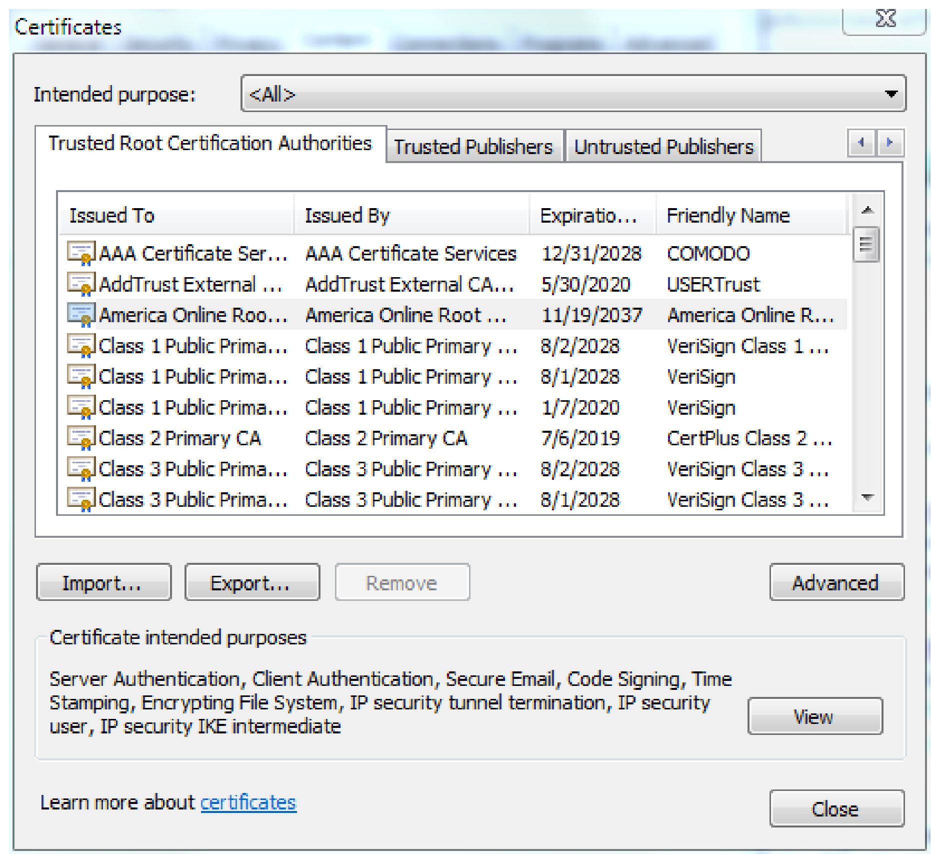
- Mozilla comes with 124 trust roots.
- IE in Win7 comes with 19 trust roots.



Some observations
[Eckersley and Burns]

Browsers come with knowledge of some certification authorities and more get added.

- Mozilla comes with 124 trust roots.
- IE in Win7 comes with 19 trust roots. *But silent updating can make this > 300!*



Some observations

[Eckersley and Burns]

- 16.2M IP addresses listened on port 443.
- 10.8M started an SSL handshake.
- 4.3+M used valid certificate chains.
- 1.3+M were *distinct* valid leaves.
- There are:
  - strange certificates (e.g., for “localhost”, “mail”),
  - vulnerabilities (e.g., 508-bit RSA keys).

Trusted but not trustworthy?

Experts Warn of a Weak Link in the Security of Web Sites

The New York Times
nytimes.com

By MIGUEL HELFT
Published: August 13, 2010

From EFF's open letter to Verizon:

> We are writing to request that Verizon investigate the security and privacy implications of the SSL CA certificate (serial number 0x40003f1) that Cybertrust (now a division of Verizon) issued to Etisalat on the 19th of December, 2005, and evaluate whether this certificate should be revoked.
>
> As you are aware, Etisalat is a telecommunications company headquartered in the United Arab Emirates. In July 2009, Etisalat issued a mislabeled firmware update to approximately 100,000 of its BlackBerry subscribers that contained malicious surveillance software [1]. Research In Motion subsequently issued patches to remove this malicious code [2].

Trusted but not trustworthy?

TECHNOLOGY | MARCH 24, 2011

THE WALL STREET JOURNAL.
Digital Network

Web Firm Suspects Iran Hacked Into It

Internet-Security Company Says It Was Tricked Into Authenticating Fake Sites, Opening Access to Data, Not Money

By CHRISTOPHER RHOADS

An Internet-security company said it was tricked into trying to lure Iranian users to fake versions of major websites, a sophisticated hack it suspects the Iranian government carried out.

Some reading

- Bellovin's "A Look Back at Security Problems in the TCP/IP Protocol Suite".
- Goldberg et al.'s "How Secure are Secure Interdomain Routing Protocols?".
- Dingledine et al.'s "Tor: The Second-Generation Onion Router".
- Chen et al.'s "Side-Channel Leaks in Web Applications".
- Xie et al.'s "De-anonymizing the Internet Using Unreliable IDs".
- Eckersley's "How Unique Is Your Web Browser?".
- Chapter 15 of Schneier's book *Secrets and Lies*.