Surveillance, Censorship, and Countermeasures

Professor Ristenpart

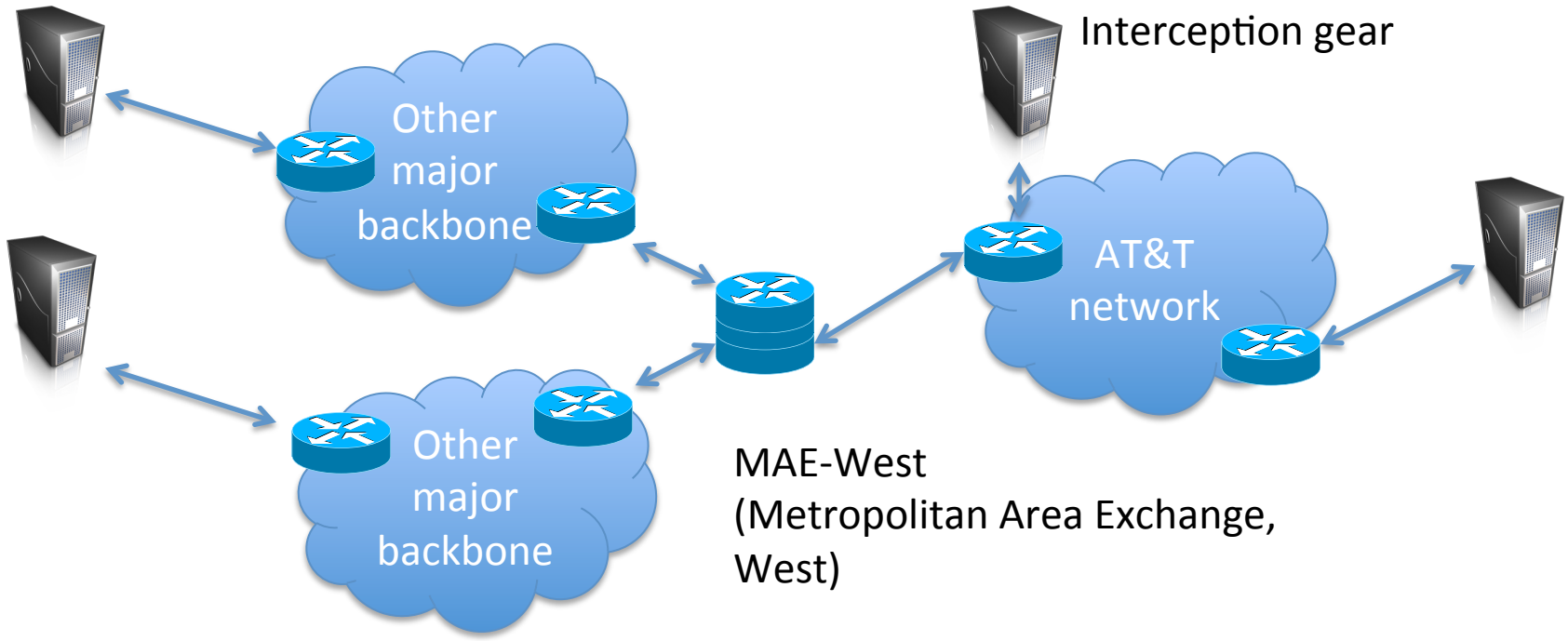
http://www.cs.wisc.edu/~rist/

rist at cs dot wisc dot edu

AT&T Wiretap case

- Mark Klein discloses potential wiretapping activities by NSA at San Francisco AT&T office
- Fiber optic splitter on major trunk line for Internet communications
  - Electronic voice and data communications copied to “secret room”
  - Narus STA 6400 device

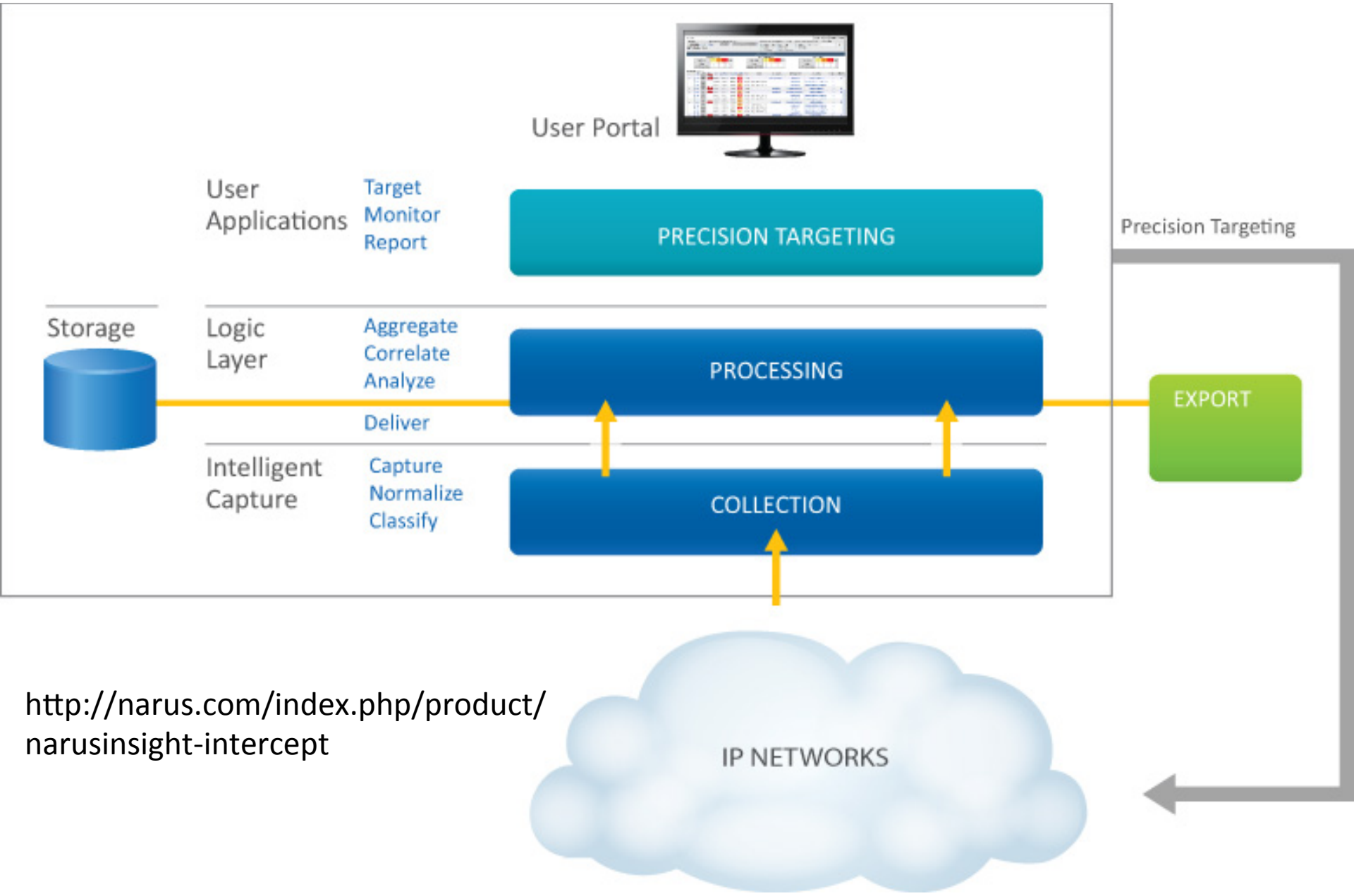
Wiretap surveillance



Large amounts of Internet traffic cross relatively few key points

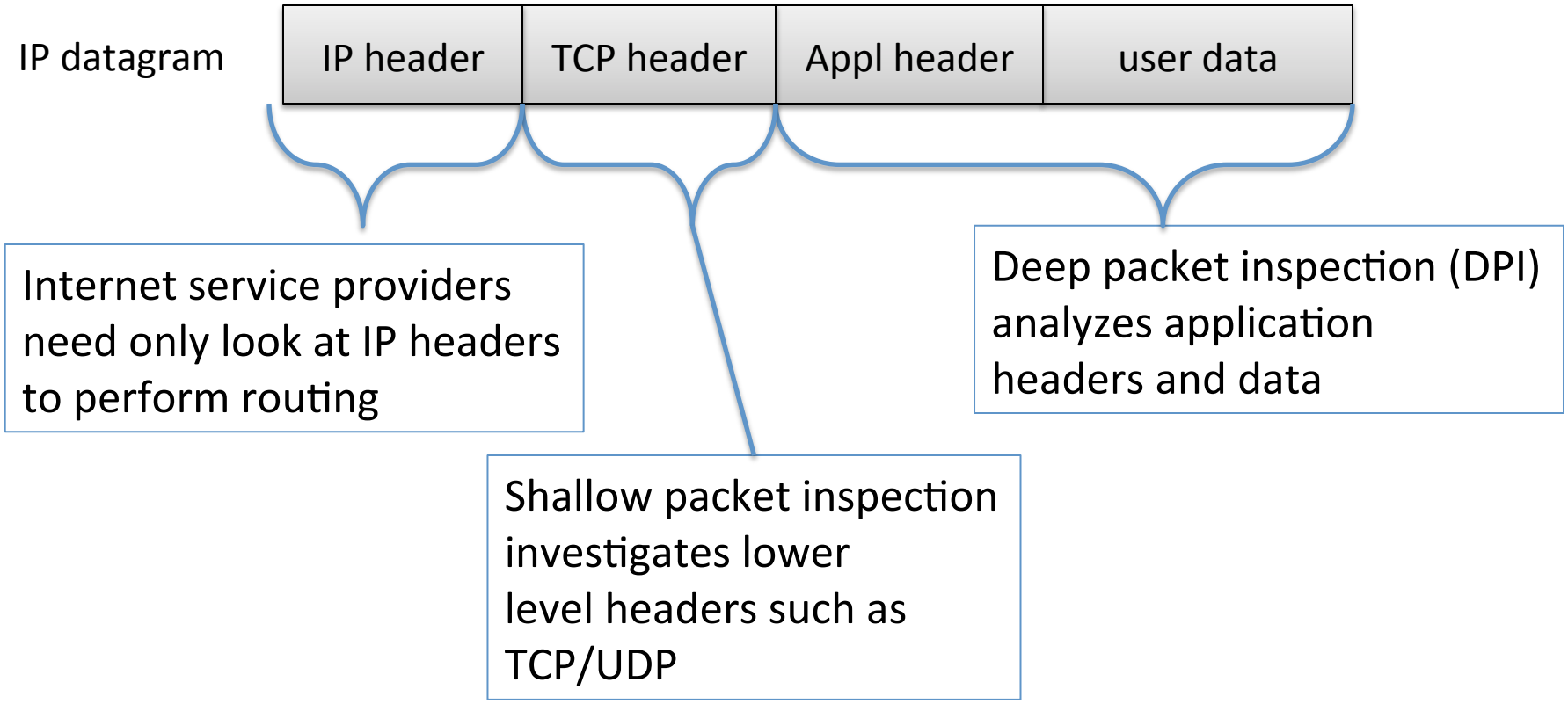
Interception technology

- From Narus' website (http://narus.com/index.php/product/narusinsight-intercept):
  - "Target by phone number, URI, email account, user name, keyword, protocol, application and more", "Service- and network agnostic", "IPV 6 ready"
  - Collects at wire speeds beyond 10 Gbps



http://narus.com/index.php/product/narusinsight-intercept

Types of packet inspection



Is dragnet surveillance technologically feasible?

- CAIDA has lots of great resources for researchers about traffic levels
- From their SanJoseA tier-1 backbone tap:

Application	Min	Avg	Max
HTTP	51.20M	2.20G	11.01G
UNKNOWN_UDP	4.08M	168.79M	711.57M
UNKNOWN_TCP	3.62M	136.02M	660.50M
HTTPS	3.96M	125.80M	543.15M
RTMP	2.00M	78.09M	314.79M
SMTP	289.75k	14.76M	55.82M
QUAKE	300.58k	8.31M	36.02M
SQUID	42.88k	7.25M	37.58M
IPSEC	213.15k	7.09M	23.97M
SSH	248.25k	6.73M	28.40M
WOW	72.88k	6.12M	34.40M
ABACAST	285.74k	3.43M	14.98M
NOPORTS_UDP	64.46k	2.04M	14.83M
other	1.23M	40.23M	161.56M

generated 2011-11-15 17:13 UTC

http://www.caida.org/data/realtime/passive/?monitor=equinix-sanjose-dirA

Key Features

From http://narus.com/index.php/product/narusinsight-intercept

Precision Targeting at Broadband Speeds

- Broad range of target types from Layer 2 through Layer 7, including ATM/MPLS/VPN support
- Target by phone number, URI, email account, user name, keyword, protocol, application and more
- Service- and network agnostic
- IPV 6 ready

Capture and Delivery

- Passive model collects from the line at wire speeds beyond 10 Gbps with support for asymmetric networks
- Efficient encoding of full packets and associated metadata for economical backhaul
- Flexible delivery for remote monitoring, retention or forwarding to alternate agencies

Reconstruction and Rendering

- Reconstruction and playback of captured traffic in near real time
- Integrated rendering of voice, video, email, Web mail, chat, and more
- Access to extensive metadata for all traffic types

Lawful intercept

- CALEA
  - Communications Assistance for Law Enforcement Act (1995)
- FISA
  - Foreign Intelligence Surveillance Act (1978)
  - Demark boundaries of domestic vs. foreign intelligence gathering
  - Foreign Intelligence Surveillance Court (FISC) provides warrant oversite
  - Executive order by President Bush suspend need for NSA to get warrants from FISC
- Almost all national governments mandate some kind of lawful intercept capabilities

Lots of companies

- Narus (originally Israeli company), now owned by Boeing
  - Partnered with Egyptian company Giza Systems
- Pen-Link (http://www.penlink.com/)
- Nokia, Nokia Siemens
- Cisco
- …

NarusInsight™ Selected To Save Pakistan's Telecommunications Networks Millions Of Dollars Per Year

NarusInsight™ Selected to Save Pakistan's Telecommunications Networks Millions of Dollars Per Year

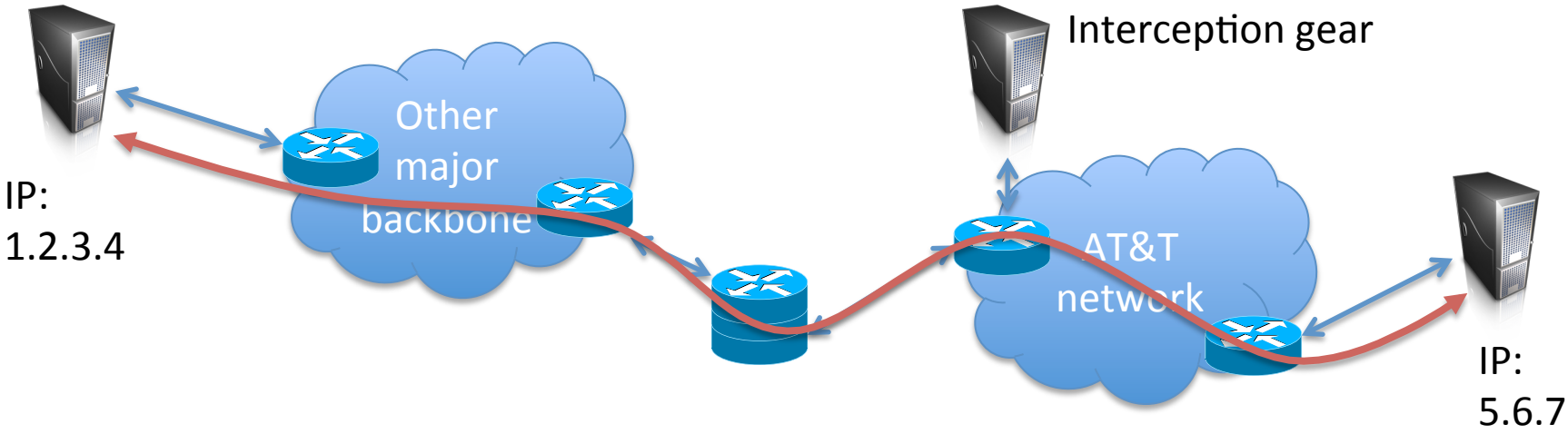
Narus System Chosen to Detect Rogue VoIP Traffic

MOUNTAIN VIEW, Calif.—September 21, 2007—Narus, Inc., the leader in carrier-class security for the world's largest IP networks, today announced that the company has teamed up with Inbox Business Technologies Pvt. Ltd, a leading total IT solution provider in Pakistan, to keep Pakistan's telecommunication networks clear of illegal, rogue and malicious IP traffic. NarusInsight was chosen by the Pakistan Telecommunication Authority (PTA) (the government administration responsible for regulating the establishment, operation and maintenance of telecommunication systems, and the provision of telecom services) to detect rogue VoIP traffic flowing through the telecommunications network in Pakistan.

http://www.narus.com/index.php/news/279-narusinsight-selected-to-save-pakistans-telecommunications-networks-millions-of-dollars-per-year

Preventing intercept

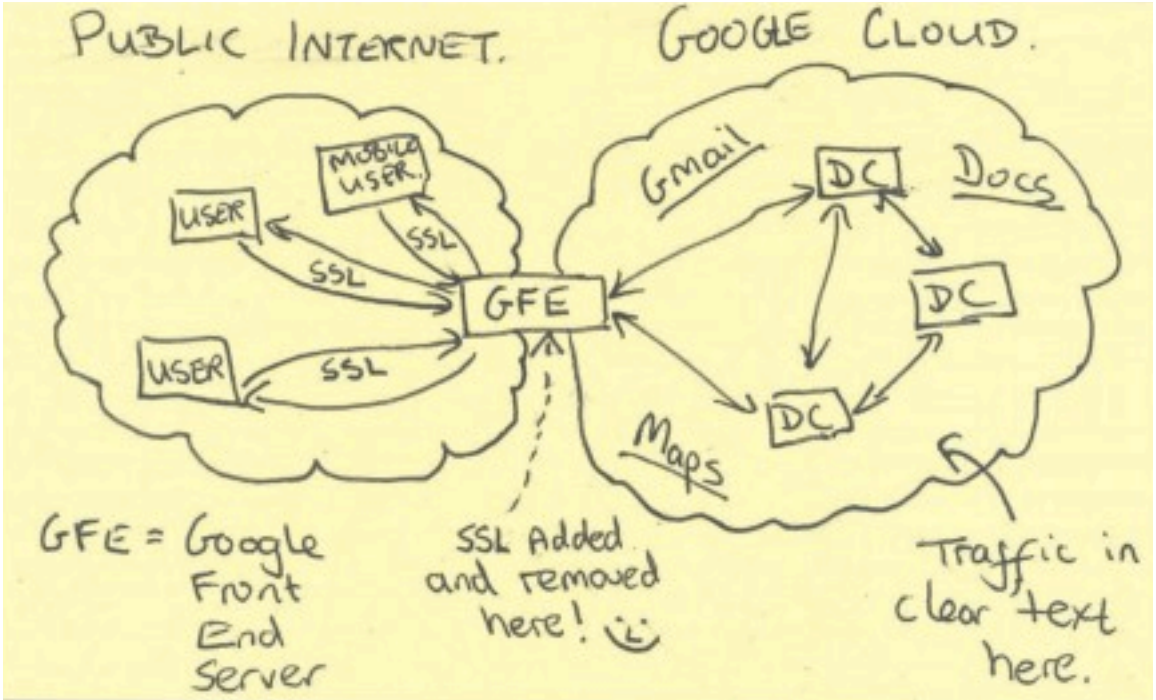
- End-to-end encryption (TLS, SSH)



- What does this protect? What does it leak?
- What can go wrong?

End-run around HTTPS

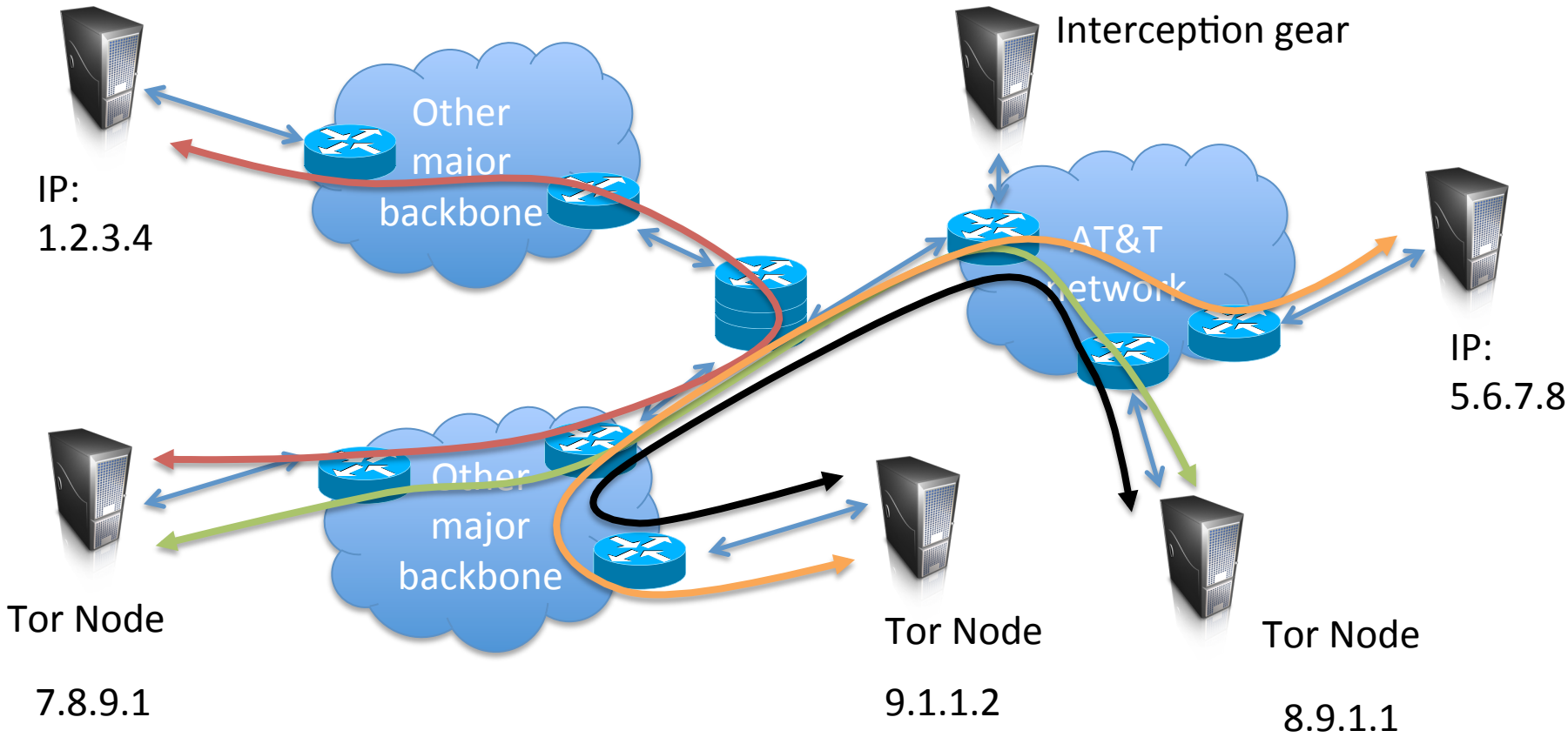
- HTTPS terminated at edge of Google networks
- Internal data center-to-data center communications on privately leased lines
  - No encryption up until last summer



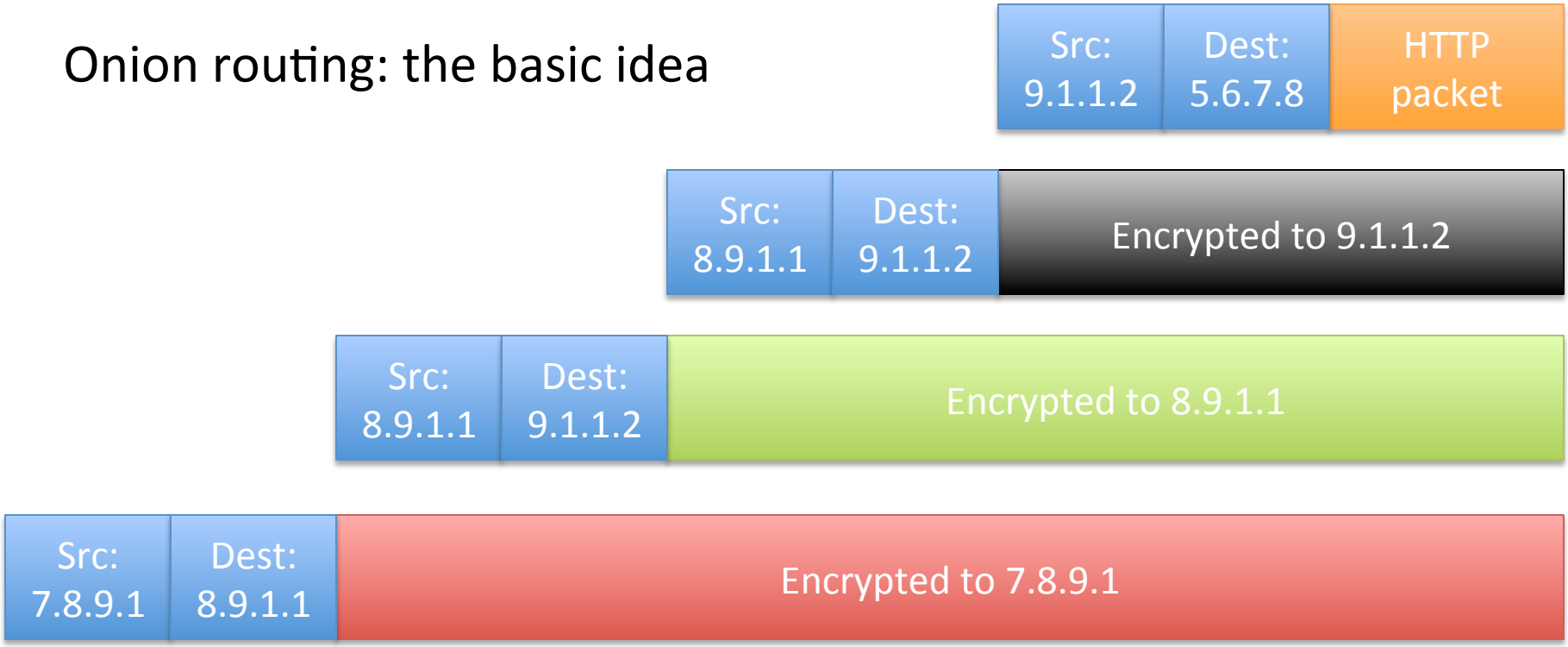
Hiding connectivity is harder

- IP addresses are required to route communication, yet not encrypted by normal end-to-end encryption
  - 1.2.3.4 talked to 5.6.7.8 over HTTPs
- How can we hide connectivity information?

Tor (The Onion Router)

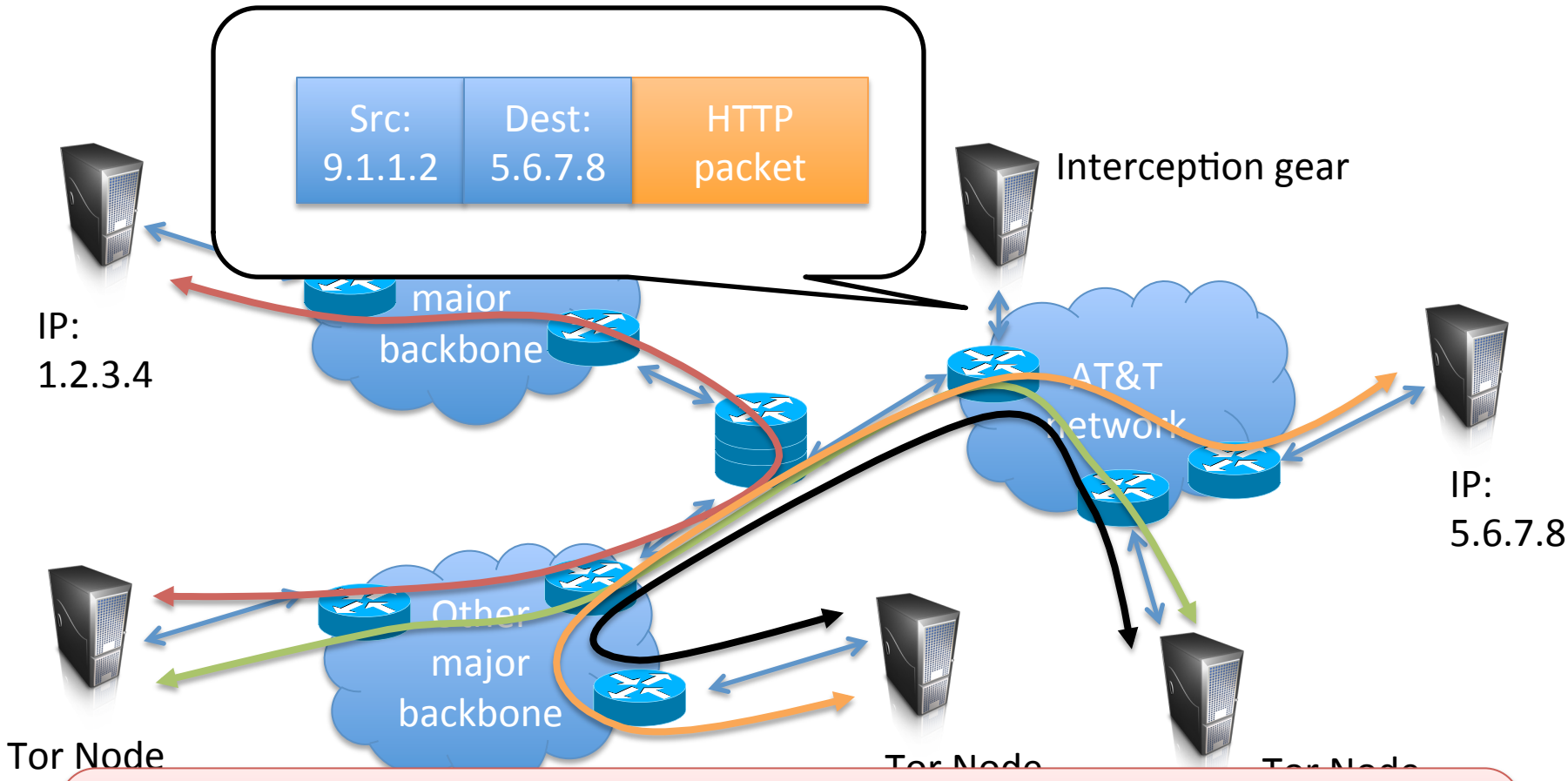


Onion routing: the basic idea

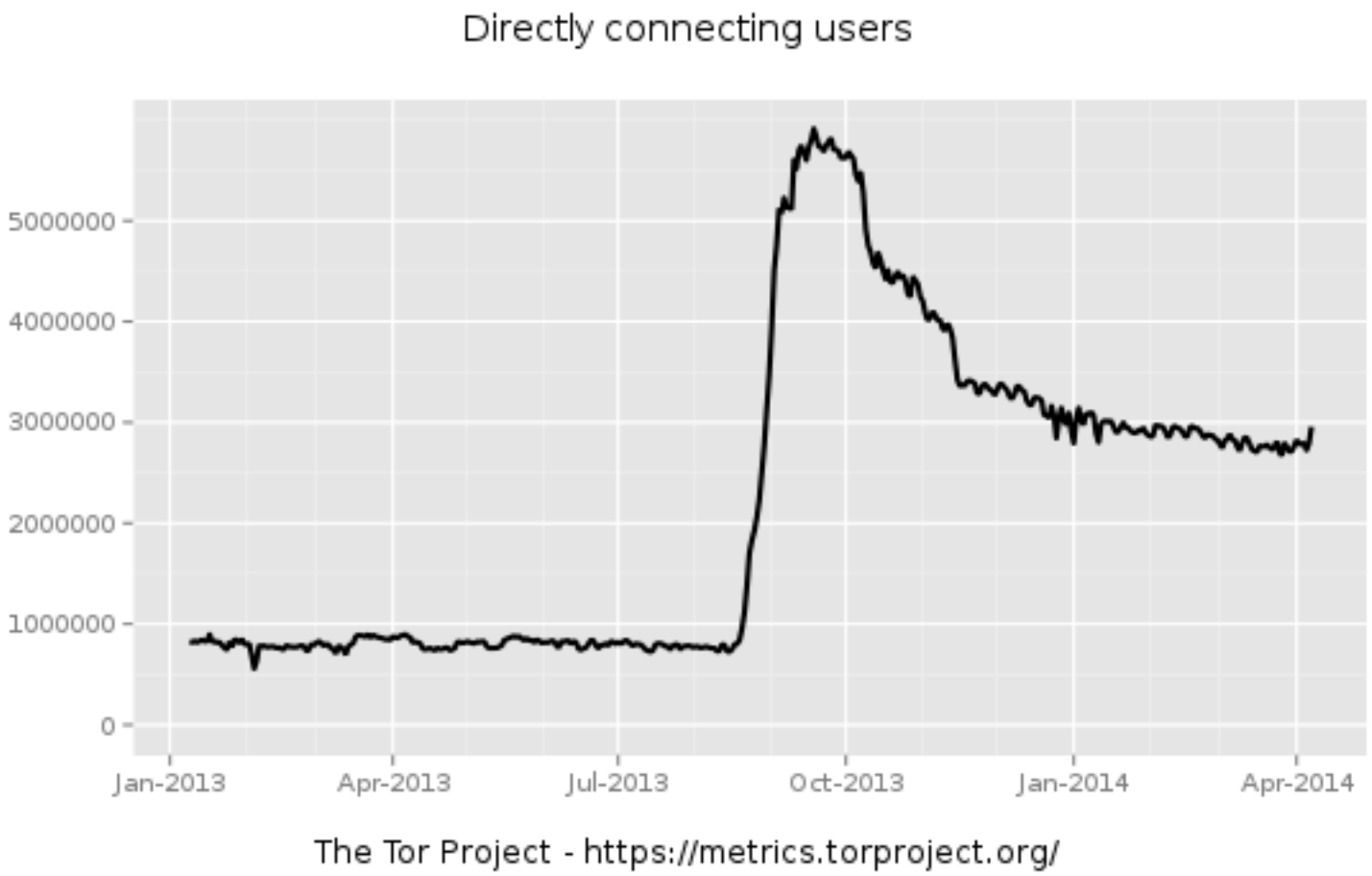


Tor implements more complex version of this basic idea

What does adversary see?



Tor obfuscates who talked to who, need end-to-end encryption (e.g., HTTPS) to protect payload


Directly connecting users
5000000
4000000
3000000
2000000
1000000
0
Jan-2013
Apr-2013
Jul-2013
Oct-2013
Jan-2014
Apr-2014
The Tor Project - https://metrics.torproject.org/

Other anonymization systems

- Single-hop proxy services

- JonDonym, anonymous remailers (MixMaster, MixMinion), many more…

Thursday, April 26, 2012

FBI seizes server used to anonymize e-mail

Jeffrey Brown 1 comment

Surveillance via third-party

- “Thus, some Supreme Court cases have held that you have no reasonable expectation of privacy in information you have "knowingly exposed" to a third party — for example, bank records or records of telephone numbers you have dialed — even if you intended for that third party to keep the information secret. In other words, by engaging in transactions with your bank or communicating phone numbers to your phone company for the purpose of connecting a call, you’ve "assumed the risk" that they will share that information with the government.”

From the EFF website

https://ssd.eff.org/your-computer/govt/privacy

Third-party legal issues

- Under Electronic Communications Privacy Act (ECPA) government has access via subpoena to:
  - Name, address
  - Length of time using service
  - Phone records (who you called, when, how long)
  - Internet records (what/when/how long services you used, your assigned IP address)
  - Info on how you pay your bill

Example: AT&T Hawkeye database

- All phone calls made over AT&T networks since approximately 2001
  - Originating phone number
  - Terminating phone number
  - Time and length of each call

Example: Google data requests

Country	▼ User Data Requests	Percentage of requests where some data produced	Users/Accounts Specified
United States	10,574	83%	18,254
France	2,750	51%	3,378
Germany	2,660	40%	3,255
India	2,513	66%	4,401
United Kingdom	1,397	69%	3,142
Brazil	1,085	49%	1,471
Italy	896	42%	1,084
Australia	780	70%	944
Singapore	755	68%	847
Spain	545	53%	761
Poland	502	23%	740
Taiwan	439	61%	580

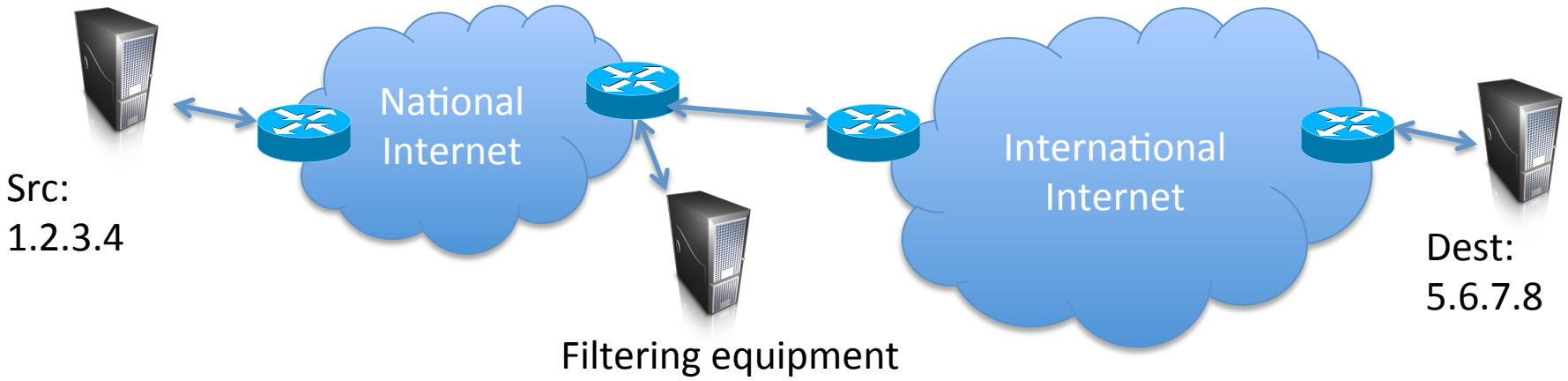
July to December 2013

From http://www.google.com/transparencyreport/governmentrequests/userdata/

Should we prevent? Can we?

- One can encrypt data that is stored, but no current way to protect data that needs to be used
- Companies are increasingly worried about perception of government surveillance
- Policy?
- Legal protections?

Censorship via Internet filtering

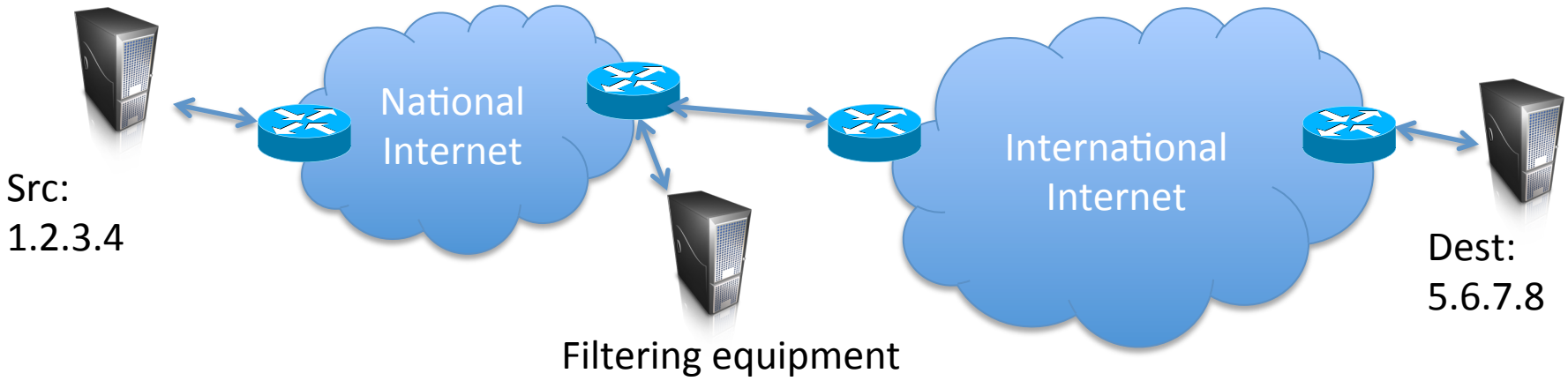


- Golden Shield Project most famous example
- But many other nations perform filtering as well including
  - Iran, Syria, Pakistan (YouTube anecdote)
  - Turkey (twitter ban recently)
  - Singapore, Australia (proposed legislation)
  - Other countries?

Big business

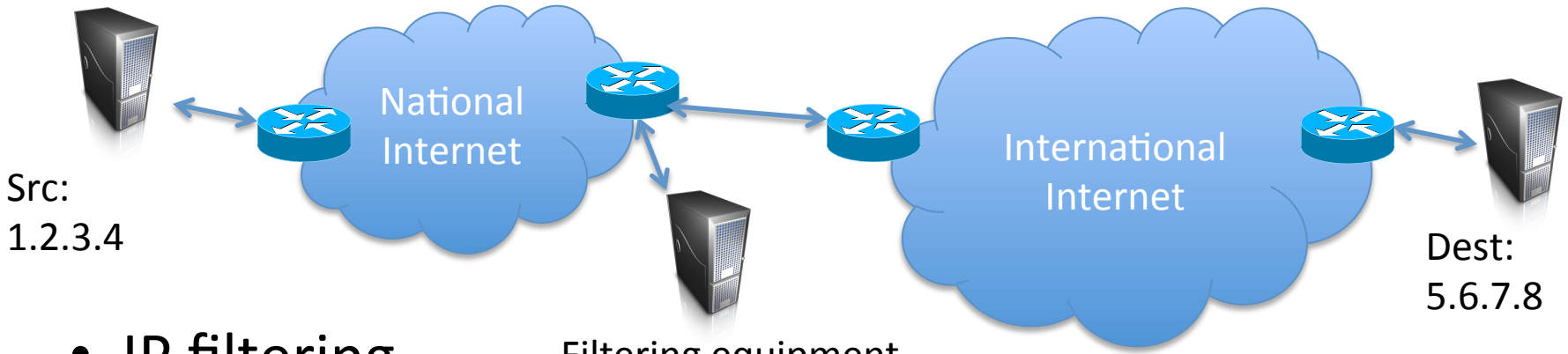
- Reports of products being used in Syria
  - Blue Coat (http://www.bluecoat.com/)
  - NetApp (http://www.netapp.com/)
- Iran, Saudi Arabia
  - Secure Computing's SmartFilter software
  - Secure Computing recently bought by McAffee
- Embargos prevent selling directly by USA companies, but resellers can do so

Filtering



- IP filtering
- DNS filtering / redirection
- URL filtering
- Packet filtering (search keywords in TCP packets)
- Protocol filtering (detect Tor protocol)

Circumvention of filtering



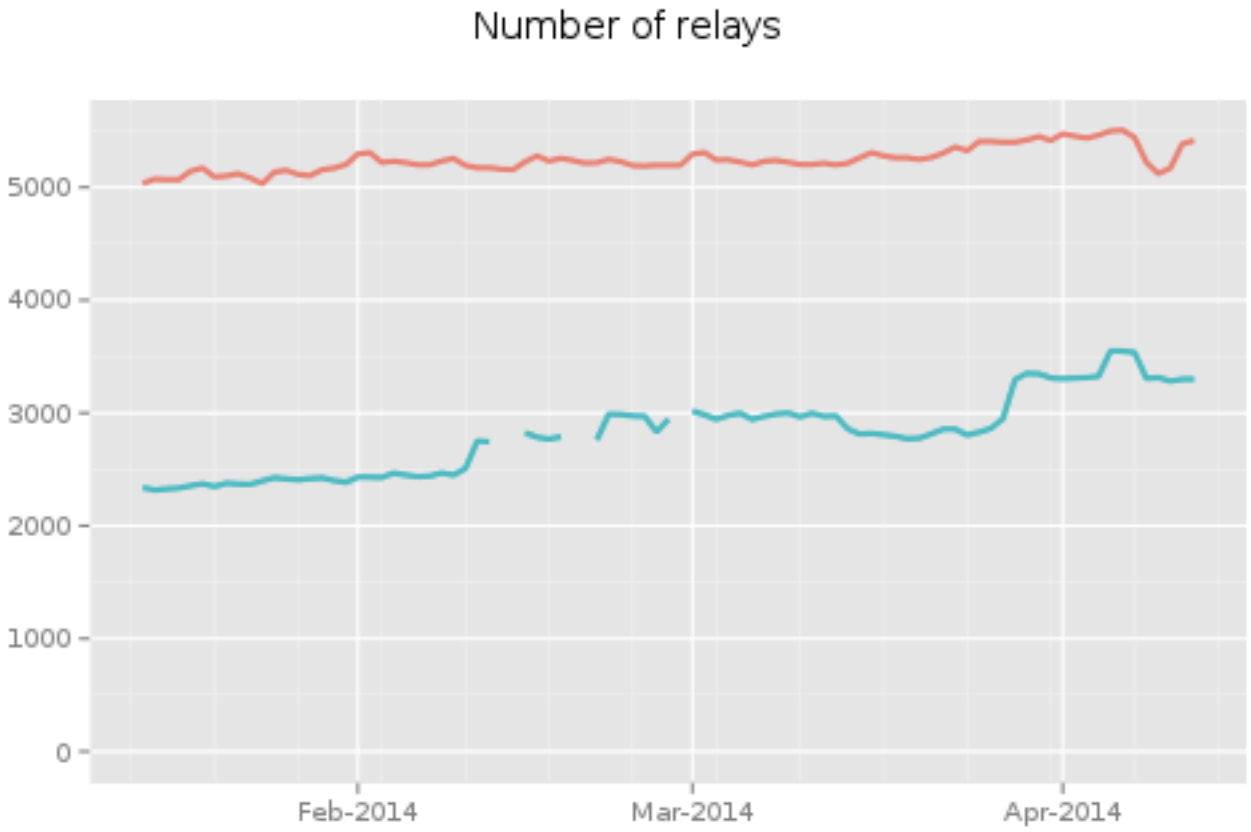
- IP filtering
  - Proxies
- DNS filtering / redirection
  - DNS proxy
- URL filtering or Packet filtering
  - Encryption / Tunneling / obfuscation
- Protocol filtering
  - Obfuscation techniques

Golden Shield Project (Great Firewall of China)

- IP filtering
- DNS filtering / redirection
- URL filtering
- Packet filtering (search keywords in TCP packets)
  - Send TCP FIN both ways
- Protocol filtering (Tor is shut down)

Great Firewall targeting of Tor (circa 2011 and before)

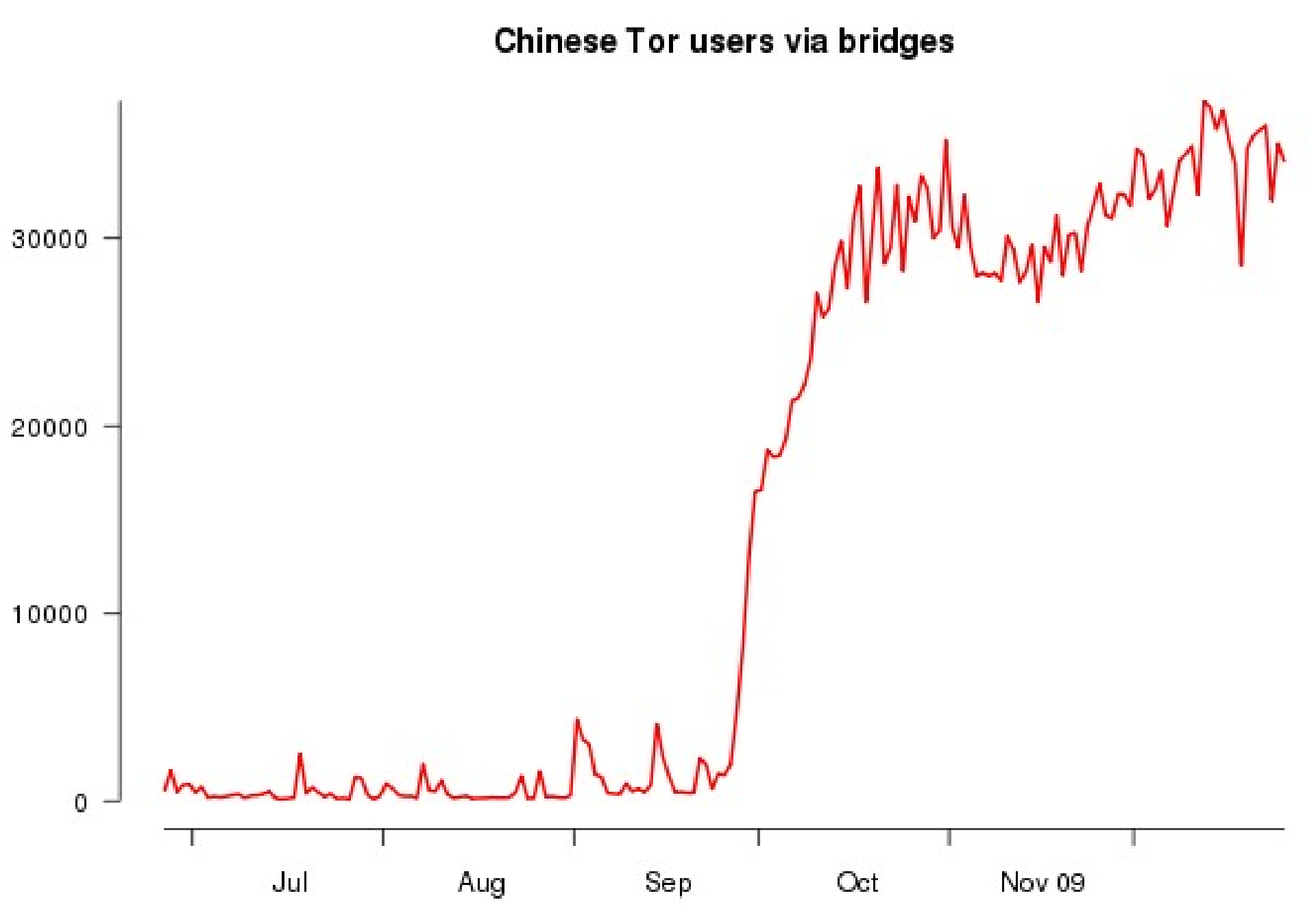
- Enumerate Tor relays and filter them



Relay is publicly listed Tor node

Bridge is Tor node not publicly listed

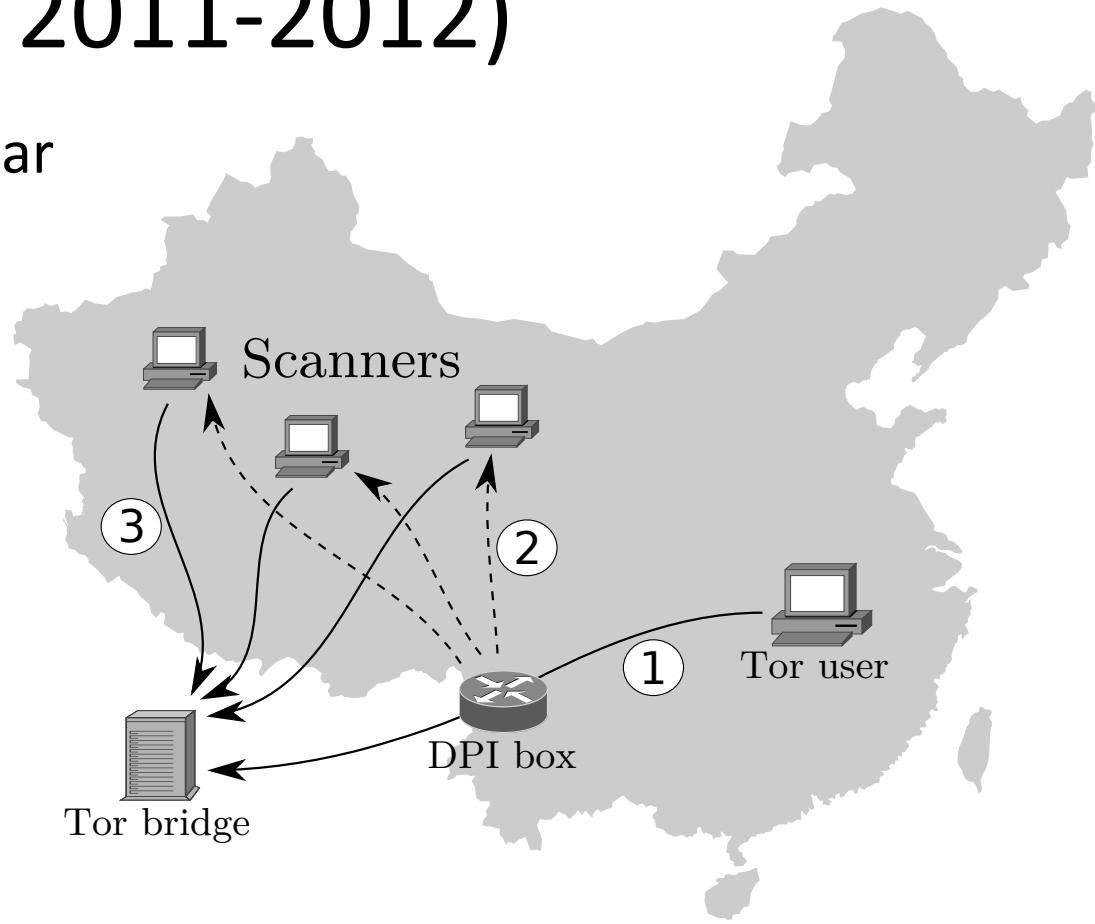

Number of directory requests to directory mirror trusted
10000
8000
6000
4000
2000
0
Sep 10
Sep 17
Sep 24
Oct 01
China
https://torproject.org



Tor project -- www.torproject.org

Great Firewall targeting of Tor (circa 2011-2012)

TLS connections with particular ciphersuites flagged



From [Winter, Lindskog 2012]

https://gist.github.com/da3c7a9af01d74cd7de7

Tor client

TLS Handshake

Tor bridge

Pick random Nc

ClientHello, MaxVer, Nc, Ciphers/CompMethods

ServerHello, Ver, Ns, SessionID, Cipher/CompMethod

Pick random Ns

Tor uses TLS for point-to-point communciations, including first hop

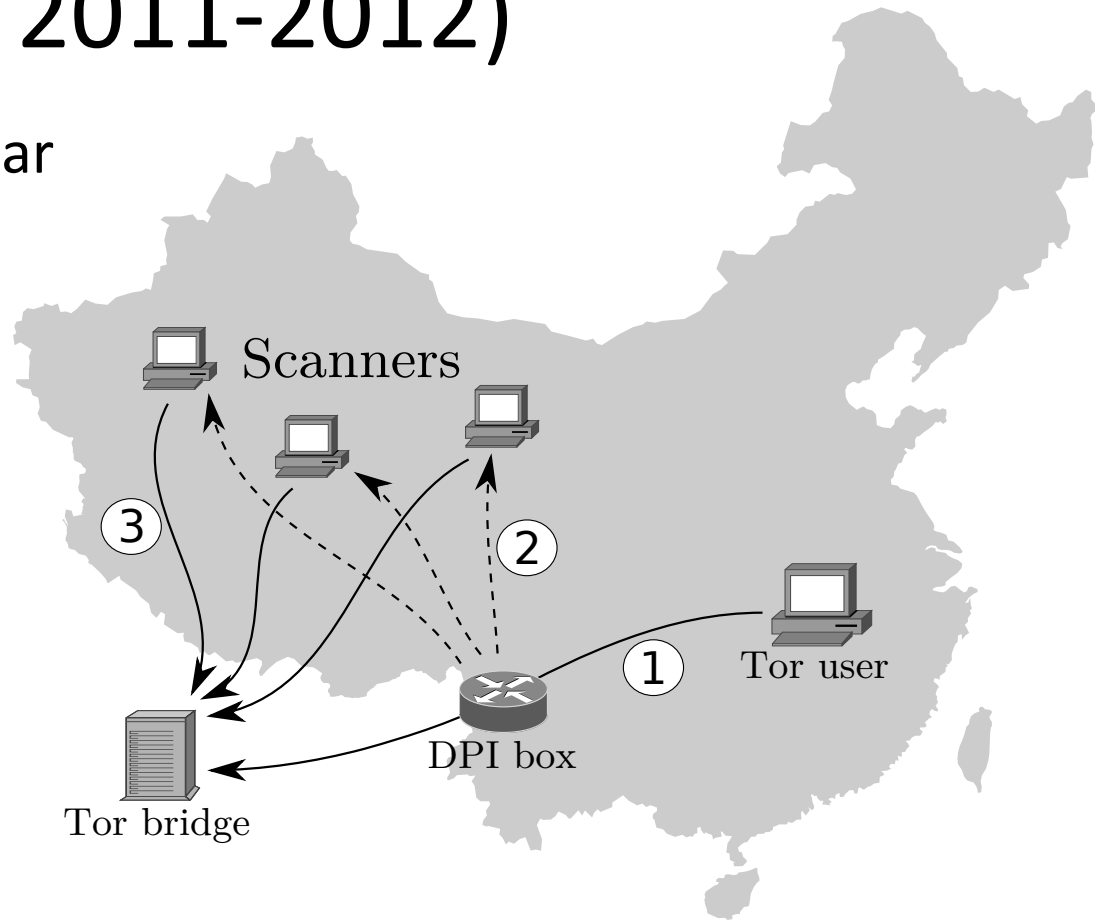
Tor clients used relatively non-standard Ciphers

Great Firewall targeting of Tor (circa 2011-2012)

TLS connections with particular ciphersuites flagged

Attempt to connect to dest IP by Tor client (source IP may be spoofed)

If server speaks Tor, then IP added to GFW black list

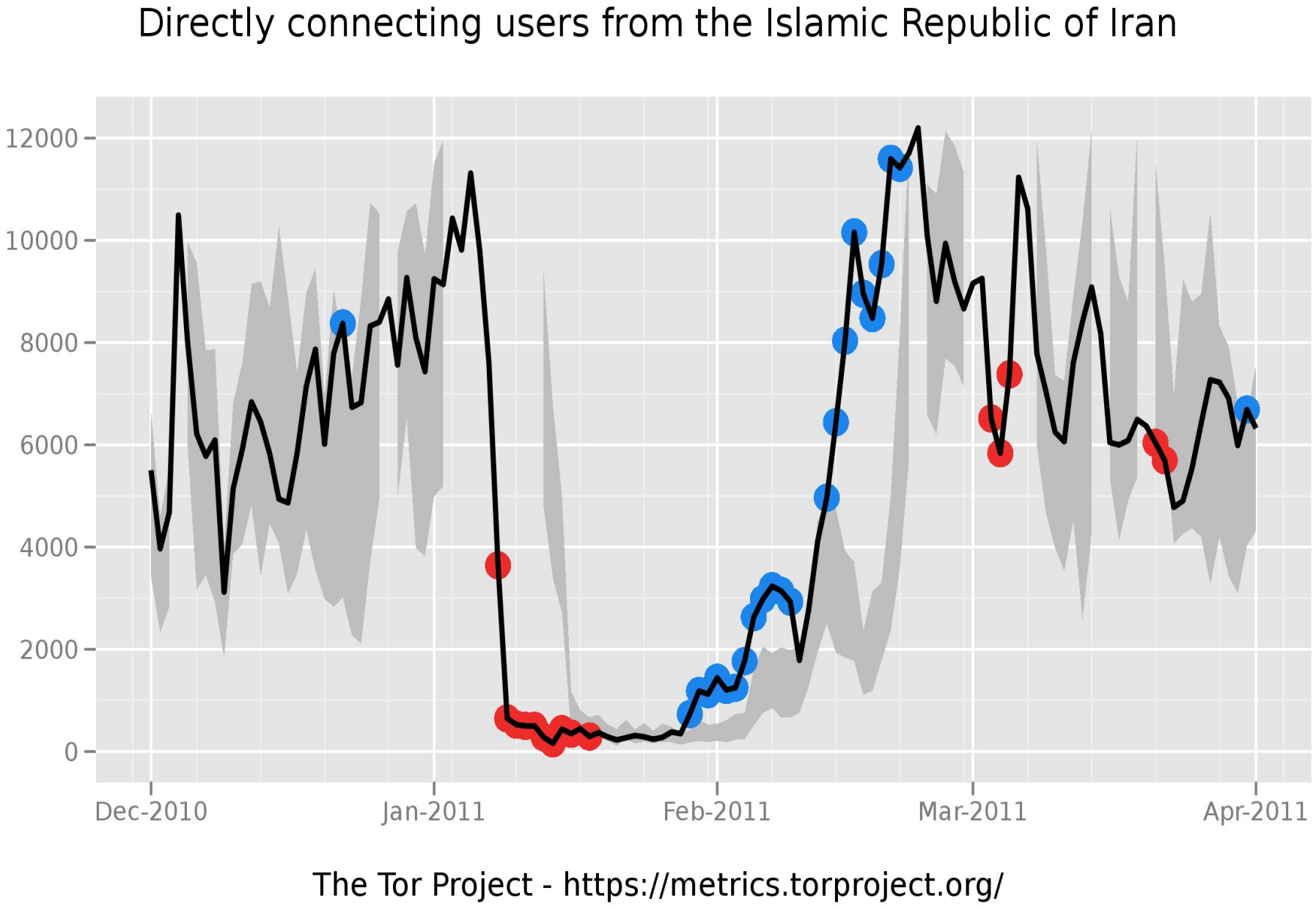


From [Winter, Lindskog 2012]

https://gist.github.com/da3c7a9af01d74cd7de7

Islamic Republic of Iran

- Every ISP must run “content-control software”
  - SmartFilter (up until 2009)
  - Nokia Siemens DPI systems
- According to wikipedia Facebook, Myspace, Twitter, Youtube, Rapidshare, Wordpress, BBC, CNN, all have been filtered
  - Big Web 2.0 security officer anecdote by way of Roger Dingledine (Tor project):
    - 10% (~10k) of traffic via Tor
    - 90% (~90k) of traffic via Amazon-hosted proxies


Directly connecting users from the Islamic Republic of Iran
12000
10000
8000
6000
4000
2000
0
Dec-2010
Jan-2011
Feb-2011
Mar-2011
Apr-2011
The Tor Project - https://metrics.torproject.org/

Iran DPI to shut down Tor

- Tor makes first hop look like TLS/HTTPS connection

Bank customer

TLS Handshake

Bank

Pick random Nc

ClientHello, MaxVer, Nc, Ciphers/CompMethods

ServerHello, Ver, Ns, SessionID, Cipher/CompMethod

Pick random Ns

Check CERT using CA public verification key

CERT = (pk of bank, signature over it)

C

Pick random PMS
C <- E(pk,PMS)

PMS <- D(sk,C)

ChangeCipherSpec,
{ Finished, PRF(MS, “Client finished” || H(transcript)) }

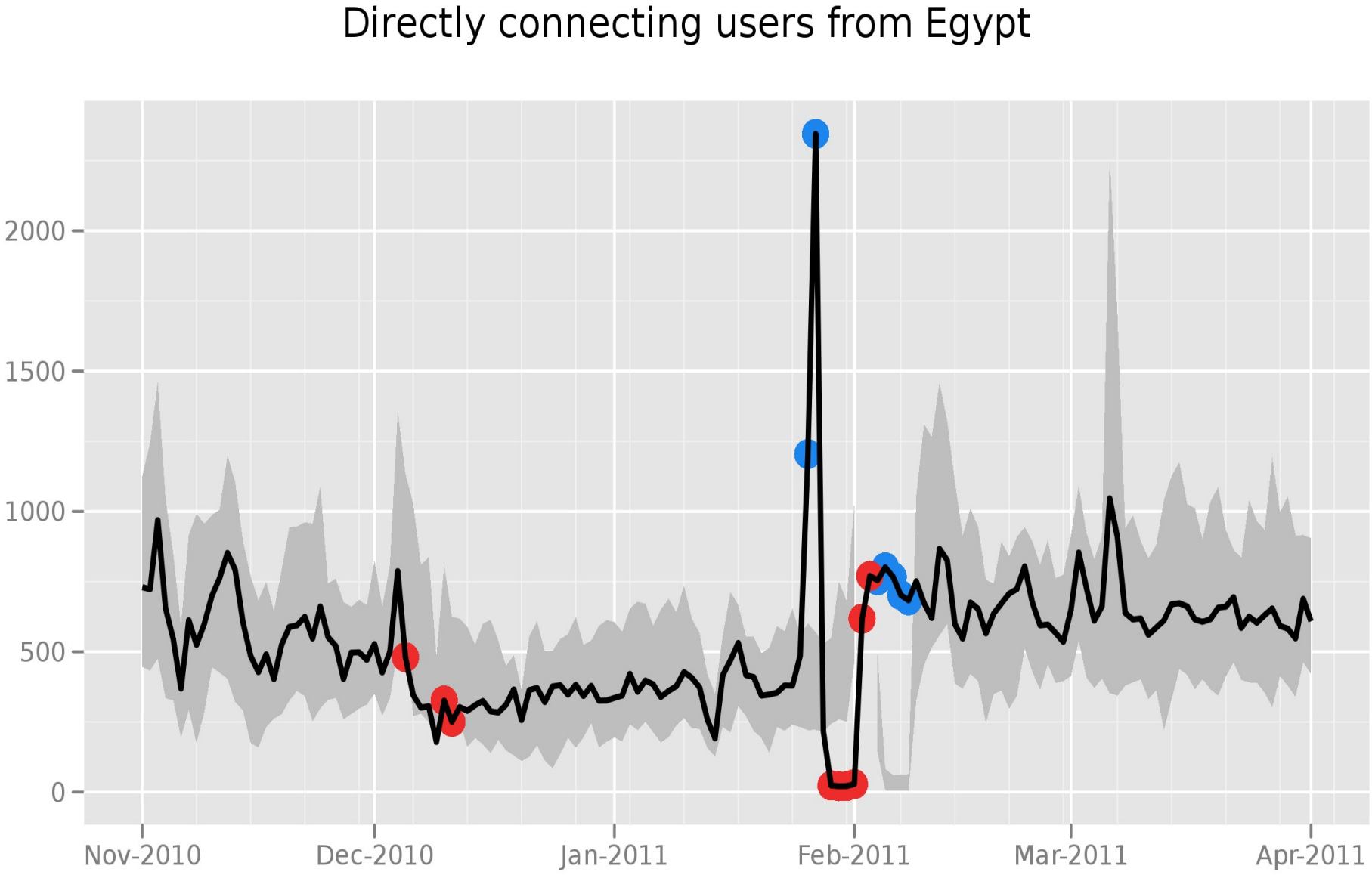
ChangeCipherSpec,
{ Finished, PRF(MS, “Server finished” || H(transcript’)) }

Bracket notation means contents encrypted

MS <- PRF(PMS, “master secret” || Nc || Ns)

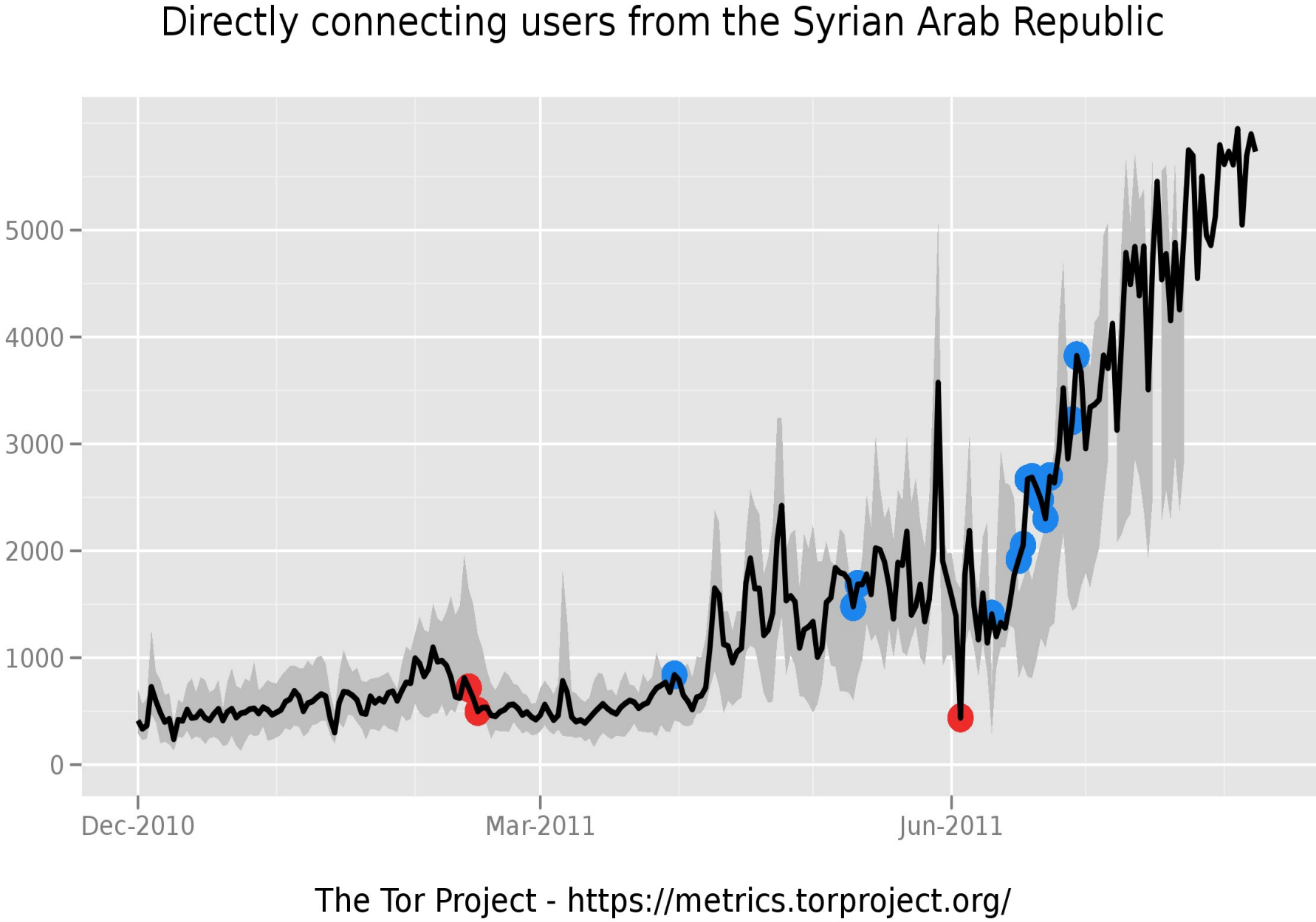
Iran DPI to shut down Tor

- Tor makes first hop look like TLS/HTTPS connection
- Use DPI to filter Tor connections:
  - Tor certificates have short expiration date
  - Most websites have long expiration date
  - Shut down those connections with short expiration dates
- Tor fixed via longer expiration dates
- Later in 2012: blocking/degrading all TLS connections

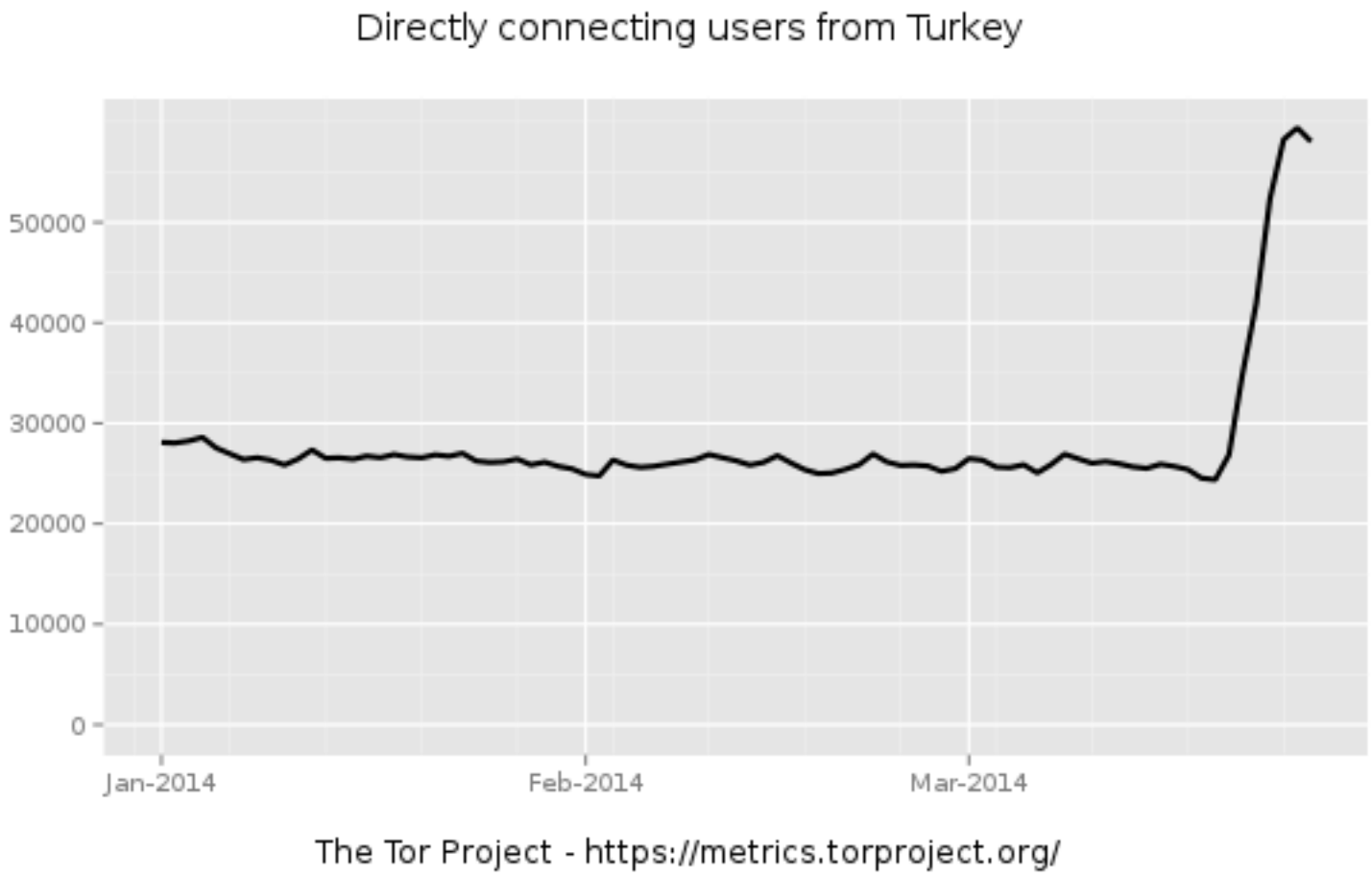

Directly connecting users from Egypt
2000
1500
1000
500
0
Nov-2010
Dec-2010
Jan-2011
Feb-2011
Mar-2011
Apr-2011
The Tor Project - https://metrics.torproject.org/

From BlueCoat:

- Our awareness of the presence of these ProxySG appliances in Syria came from reviewing online posts made by so-called “hacktivists” that contained logs of internet usage which appear to be generated by ProxySG appliances. We believe that these logs were obtained by hacking into one or more unsecured third-party servers where the log files were exported and stored. We have verified that the logs likely were generated by ProxySG appliances and that these appliances have IP addresses generally assigned to Syria. We do not know who is using the appliances or exactly how they are being used. We currently are conducting an internal review and also are working directly with appropriate government agencies to provide information on this unlawful diversion.


Directly connecting users from the Syrian Arab Republic
5000
4000
3000
2000
1000
0
Dec-2010
Mar-2011
Jun-2011
The Tor Project - https://metrics.torproject.org/

“Twitter, mwitter!”



Protocol identification via deep-packet inspection (DPI)

Check packet contents against ***regular expressions***

/^(\x16\x03[\x00\x01\x02]..\x02...\x03[\x00\x01\x02]|...? .*/

Free translation: Does packet include "I'm TLS 1.1" ?

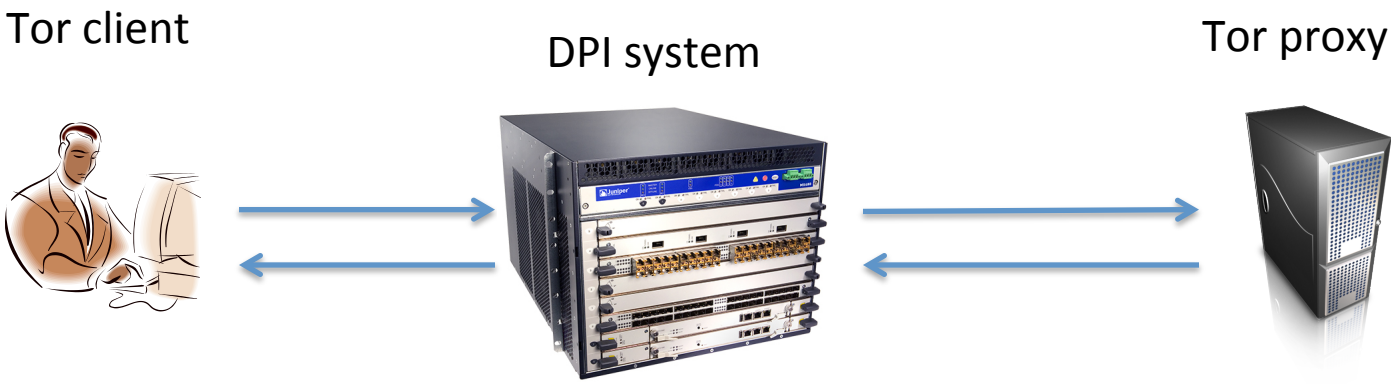
DPI users want to identify protocol X

X = TLS or Tor then throttle connection

X = HTTP then leave it alone

X = ??? then throttle traffic

Scenario:
DPI system only allows HTTP traffic unfettered



Stegonagraphy (e.g., Stegotorus): embed bits into HTTP messages

- Too slow for practical use (56k modem anyone?)

Obsfproxy (built into Tor): encrypt all bits sent over network (no plaintext bits)

- Really fast
- But DPI will flag traffic as ???

Want way to force DPI to classify traffic incorrectly as HTTP
So-called “misclassification attacks” against DPI

Surveying modern DPI systems

System	Look at ports?	TCP stream reassembly	Uses regex's	Use's C/C++
AppID	Yes	No	Yes	No
L7-filter	Yes	No	Yes	No
Yaf	Yes	Yes	Yes	No
Bro	Yes	Yes	Yes	Yes
nProbe	No	Yes	Not explicitly	Yes
Proprietary*	Yes	Yes	?	?

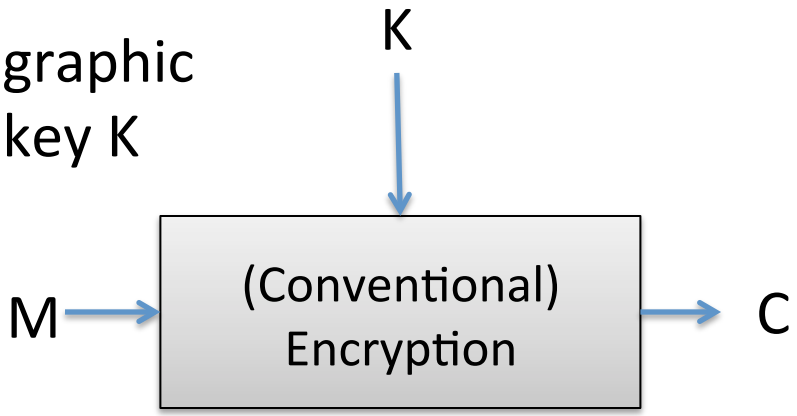
* Hint: it's a serious product (~$10k) and similar ones seem to be used in Iran.

Can we build encryption schemes that fool regex-based systems?

Attacking DPI

[Dyer, Coull, R., Shrimpton – CCS 2013]

Cryptographic secret key K



C looks like random junk. Won't look like HTTP

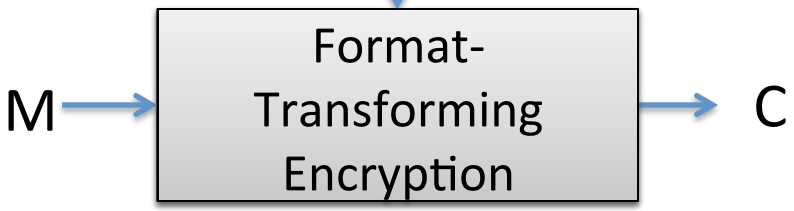
Attacking DPI

[Dyer, Coull, R., Shrimpton – CCS 2013]

Cryptographic secret key K

K , R

Regular expression R specifying desired ***ciphertext format***



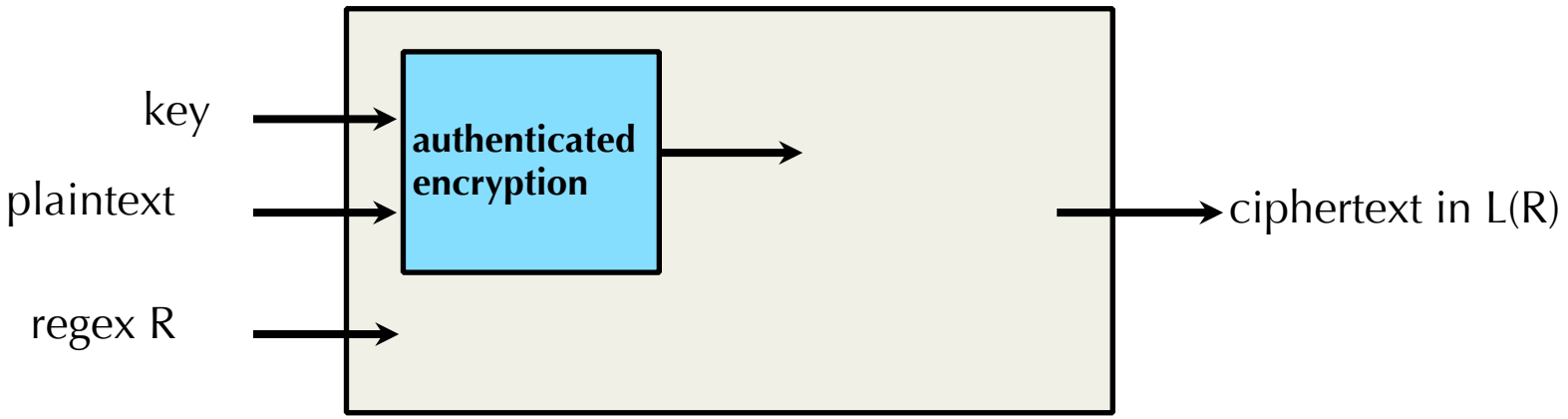
C is guaranteed to match against R

Realizing regex-based FTE

How should we realize regex-based FTE?

We want:
- Cryptographic protection for the plaintext
- Ciphertexts in L(R)

Realizing regex-based FTE

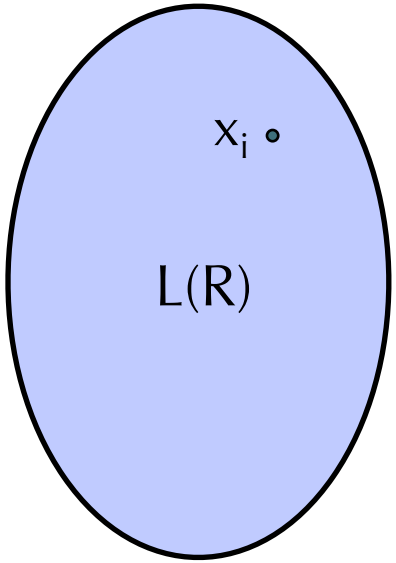


How should we realize regex-based FTE?

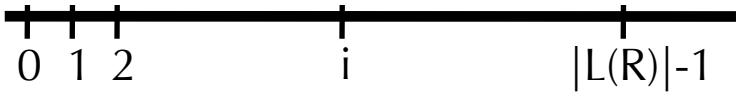
We want:
- Cryptographic protection for the plaintext
- Ciphertexts in L(R)

Ranking a Regular Language

[Goldberg, Sipser ’85]
[Bellare et al. ’09]



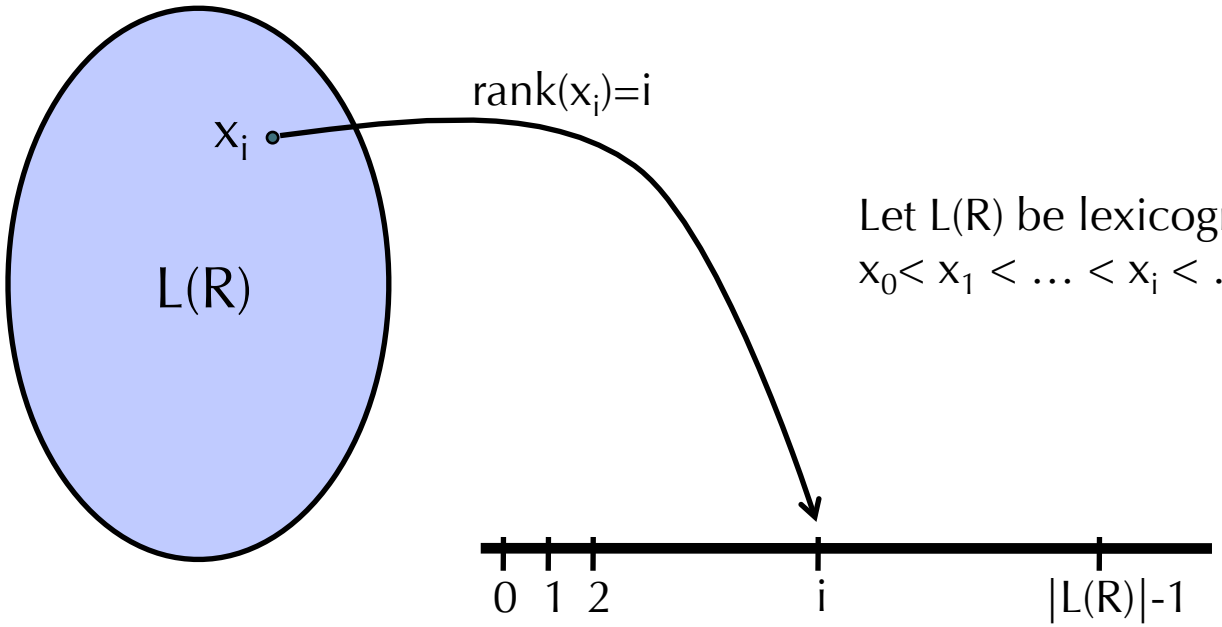
Let L(R) be lexicographically ordered
$x_0 < x_1 < \ldots < x_i < \ldots < x_{|L(R)-1|}$



Given a **DFA** (deterministic finite automaton) for L(R),
there are efficient algorithms

Ranking a Regular Language

[Goldberg, Sipser ’85]
[Bellare et al. ’09]



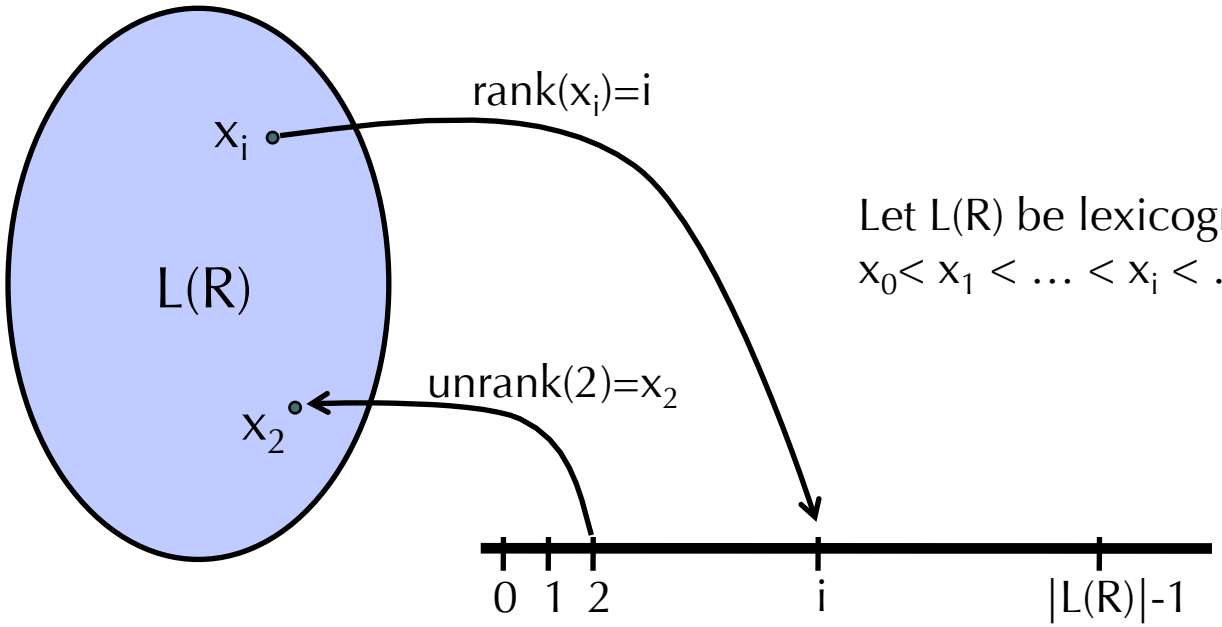
Let L(R) be lexicographically ordered
$x_0 < x_1 < \ldots < x_i < \ldots < x_{|L(R)-1|}$

Given a **DFA** (deterministic finite automaton) for L(R), there are efficient algorithms

rank: L(R) $\longrightarrow$ {0,1,…,|L(R)|-1}

Ranking a Regular Language

[Goldberg, Sipser ' 85]
[Bellare et al. ' 09]



Let L(R) be lexicographically ordered
$x_0 < x_1 < \ldots < x_i < \ldots < x_{|L(R)-1|}$

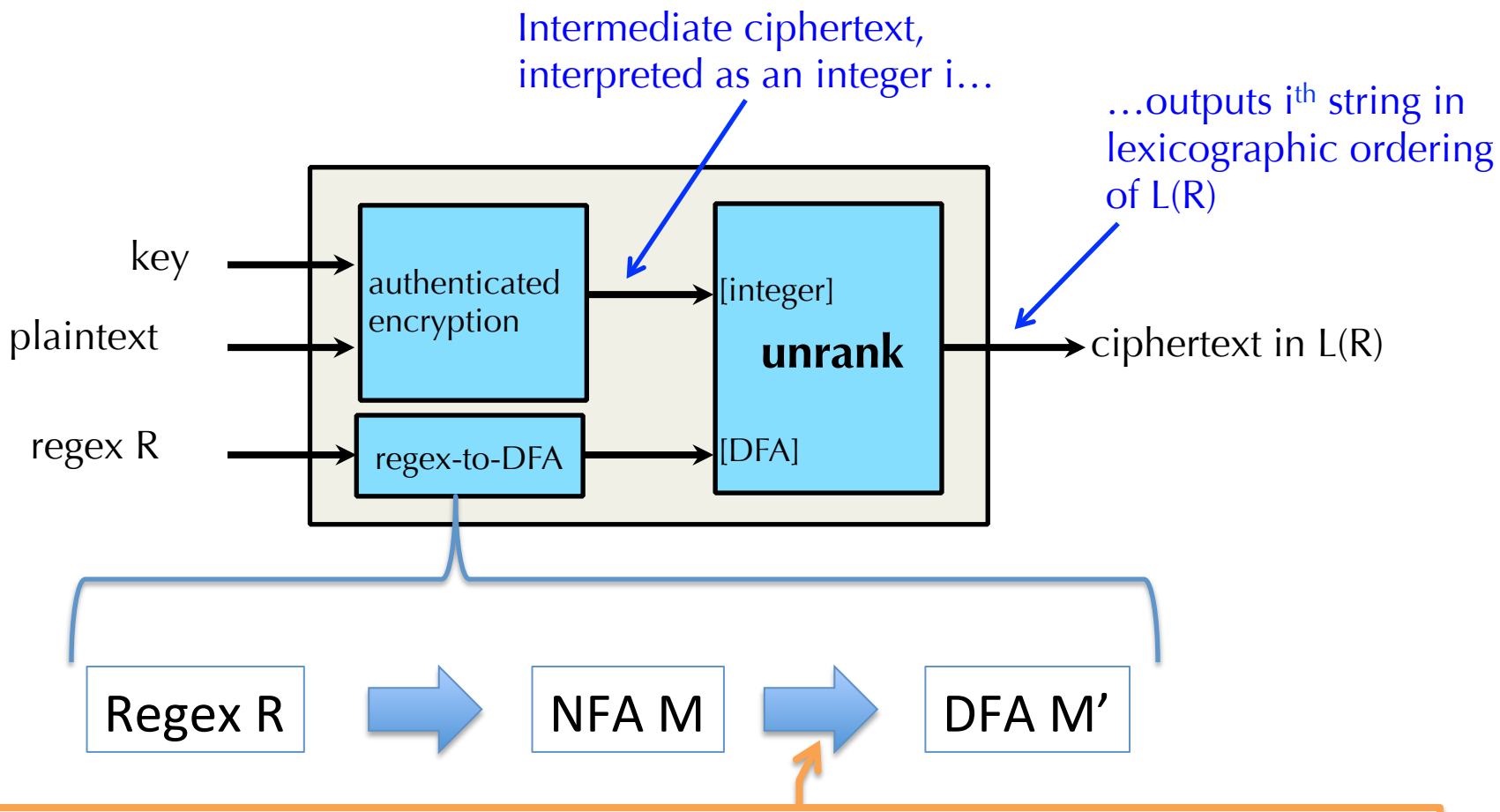
Given a **DFA** (deterministic finite automaton) for L(R), there are efficient algorithms

$$\text{rank: } L(R) \longrightarrow \{0,1,\ldots,|L(R)|-1\}$$
$$\text{unrank: } \{0,1,\ldots,|L(R)|-1\} \longrightarrow L(R)$$

such that rank(unrank(i)) = i
and unrank(rank(x_i)) = x_i

With precomputed tables, rank, unrank are *O*(n)

Realizing regex-based FTE

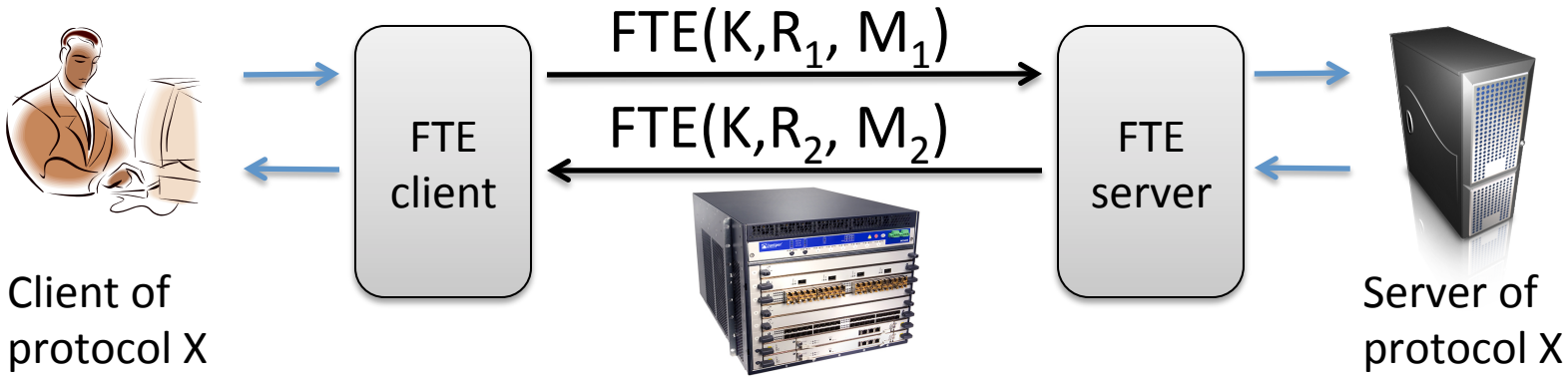


Exponential blow-up in worst case. Regexes we needed avoid this.

FTE using NFAs directly

[Luchaup, Dyer, Jha, R., Shrimpton – In submission 2014]

We built a complete FTE record layer and proxy system

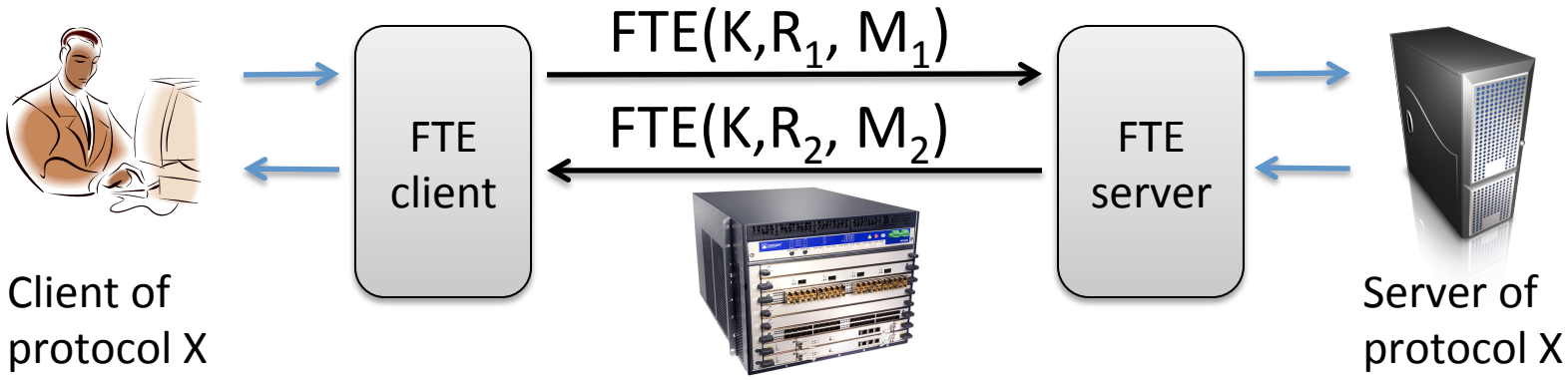


Want to trick DPI into thinking we're protocol Y != X
Where do we get R_1 and R_2?

(1) Get from DPI themselves
(2) Easy to manually craft
(3) Learn from traffic samples

We built regexes for variety of "cover" protocols:
Y = HTTP, SSH, SMB, SIP, RTSP

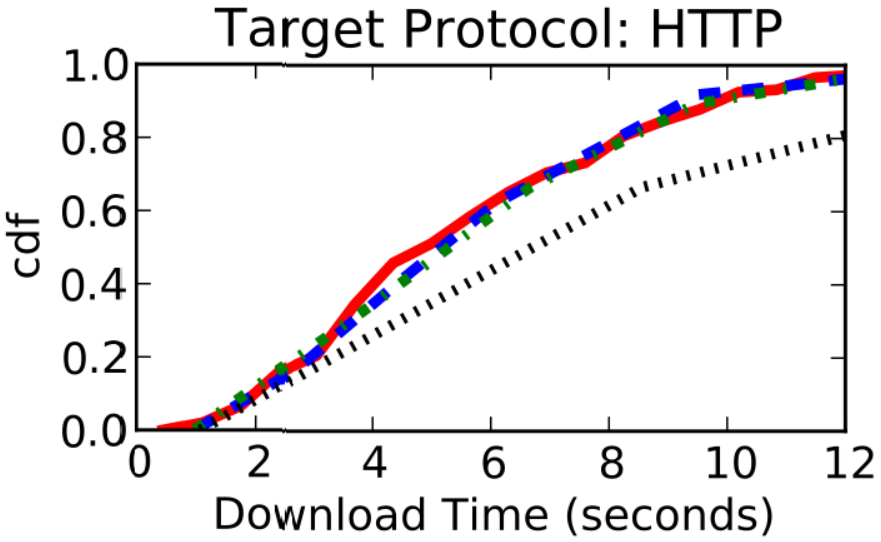
Evaluating FTE



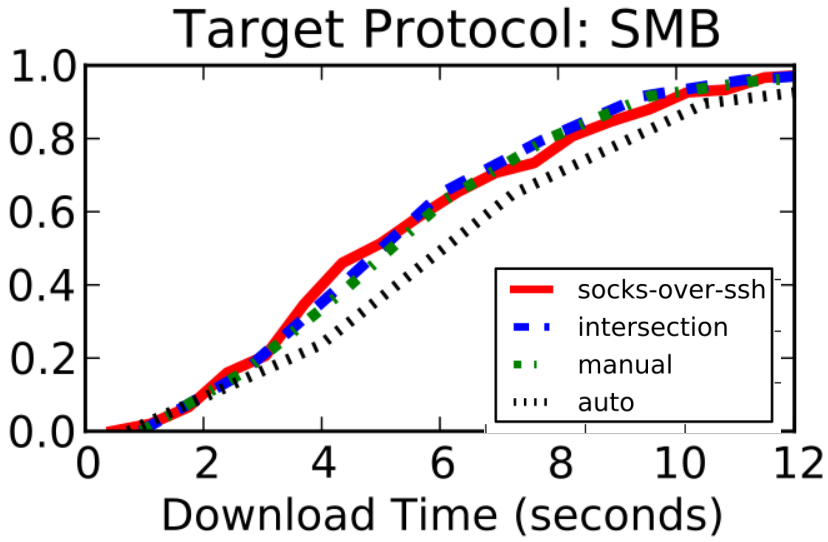
Tests with gets on Alexa Top 50 sites (X = mix of HTTPS/HTTP)
R_1 R_2 set to HTTP, SSH, SMB, and more. When do we trick DPI ?

System	DPI-derived regex's	Manual regex's	Learned regex's
AppID	Always	Always	Always
L7-filter	Always	Always	Always
Yaf	Always	Always	Always
Bro	Sometimes	Always	Always
nProbe	Never	Always	Almost always
Proprietary	Always	Always	Always

Web-browsing performance



Top 50 Alexa websites



Top 50 Alexa websites

Punchline: FTE or SSH tunnel result in the same user web-browsing experience

A field test…

Without FTE tunnel, we tried Facebook, YouTube, Tor website, banned search queries…

With FTE tunnel, we tried Facebook, YouTube, Tor website, banned search queries…

Used FTE to download Tor bundle:

Tor without FTE: "active blacklisting" attack on proxy
Tor through FTE: no problems

Ran various tests every 5 minutes for one month,
no sign of detection in logs. (We shut it down after that.)

FTE is open source,
runs on multiple platforms/OS, and
fully integrated into Tor.

http://fteproxy.org