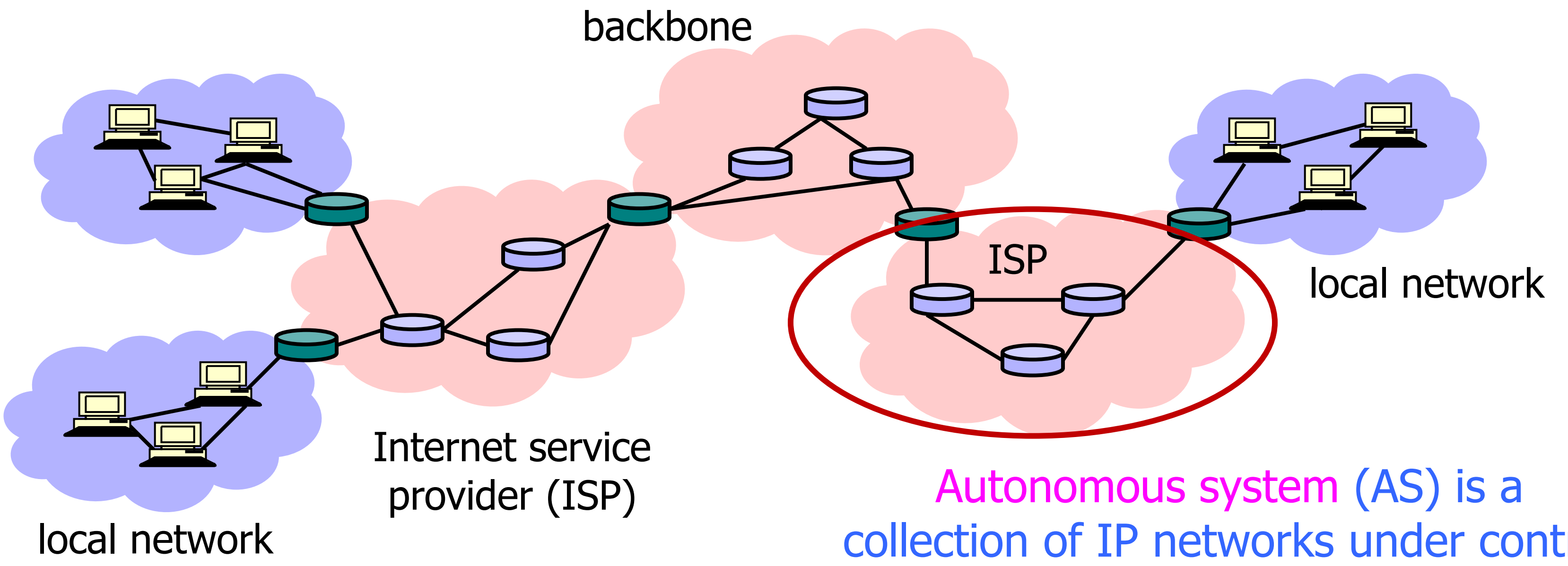
CS 6431

BGP and DNS Security

Vitaly Shmatikov

Internet Is a Network of Networks

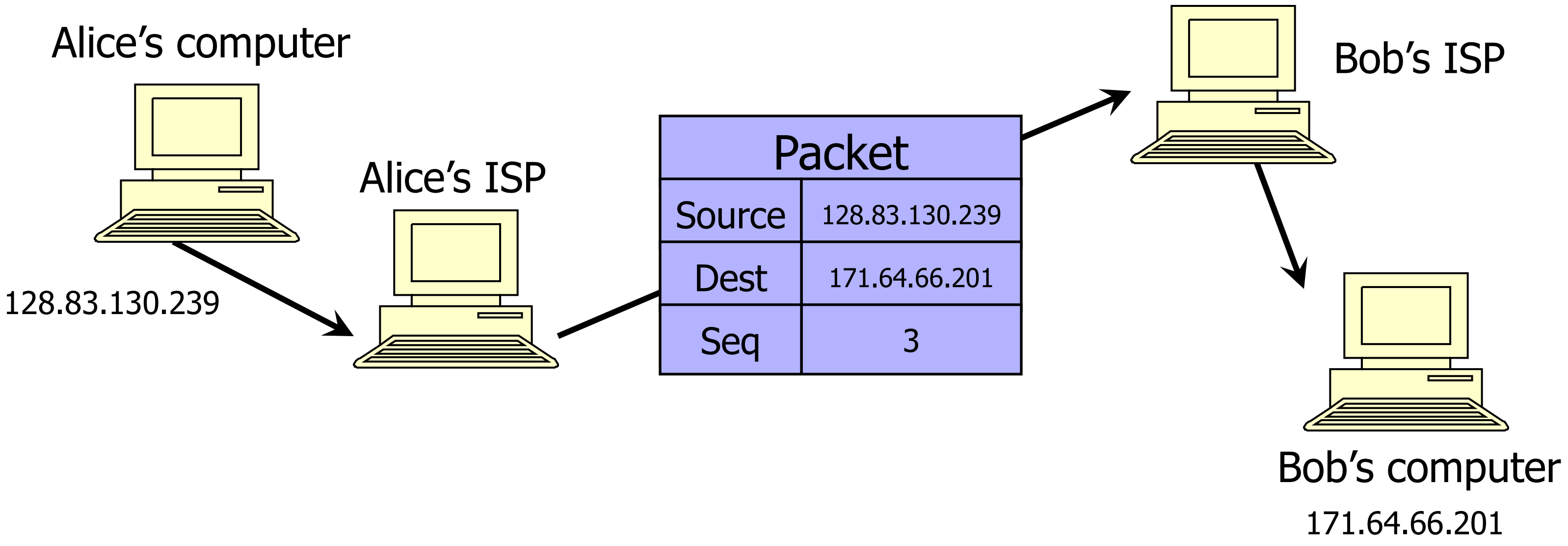


Autonomous system (AS) is a collection of IP networks under control of a single administrator (e.g., ISP)

- TCP/IP for packet routing and connections
- Border Gateway Protocol (BGP) for route discovery
- Domain Name System (DNS) for IP address discovery

IP (Internet Protocol)

- Connectionless
  - Unreliable, “best-effort” protocol
- Uses numeric addresses for routing
- Typically several hops in the route



Packet	
Source	128.83.130.239
Dest	171.64.66.201
Seq	3

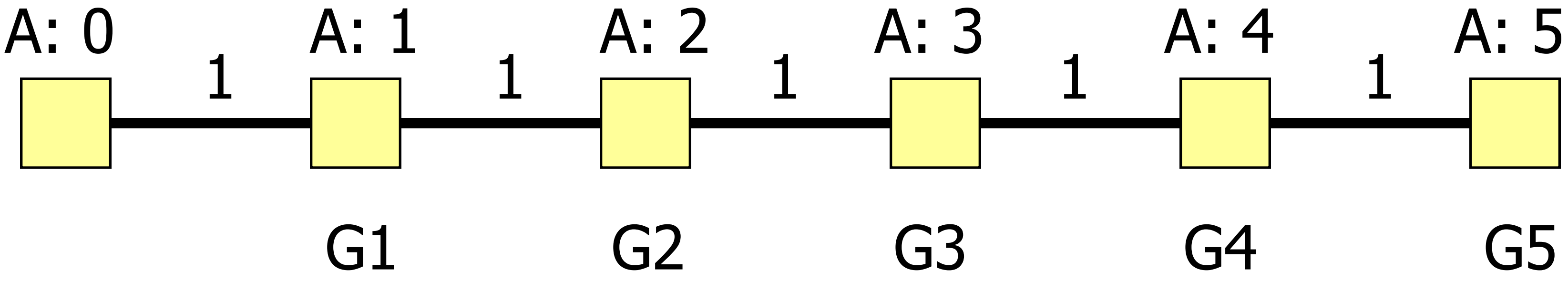
IP Routing

- Routing of IP packets is based on IP addresses
  - 32-bit host identifiers (128-bit in IPv6)
- Routers use a forwarding table
  - Entry = destination, next hop, network interface, metric
  - Table look-up for each packet to decide how to route it
- Routers learn routes to hosts and networks via routing protocols
  - Host is identified by IP address, network by IP prefix
- BGP (Border Gateway Protocol) is the core Internet protocol for establishing inter-AS routes

Distance-Vector Routing

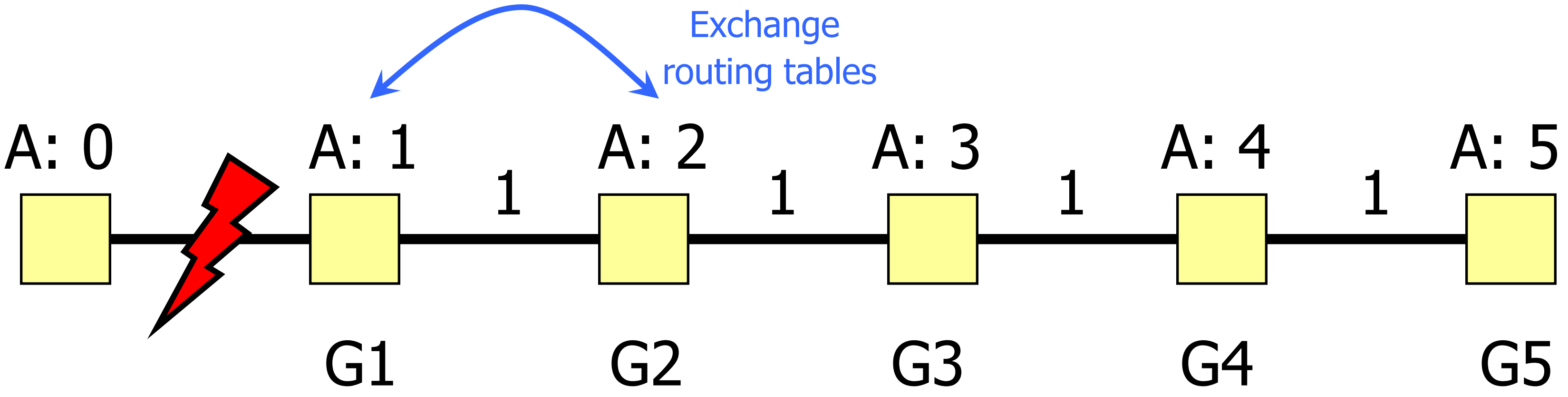
- Each node keeps vector with distances to all nodes
- Periodically sends distance vector to all neighbors
- Neighbors send their distance vectors, too; node updates its vector based on received information
  - Bellman-Ford algorithm: for each destination, router picks the neighbor advertising the cheapest route, adds his entry into its own routing table and re-advertises
  - Used in RIP (routing information protocol)
- Split-horizon update
  - Do not advertise a route on an interface from which you learned the route in the first place!

Good News Travels Fast



- G1 advertises route to network A with distance 1
- G2-G5 quickly learn the good news and install the routes to A via G1 in their local routing tables

Bad News Travels Slowly



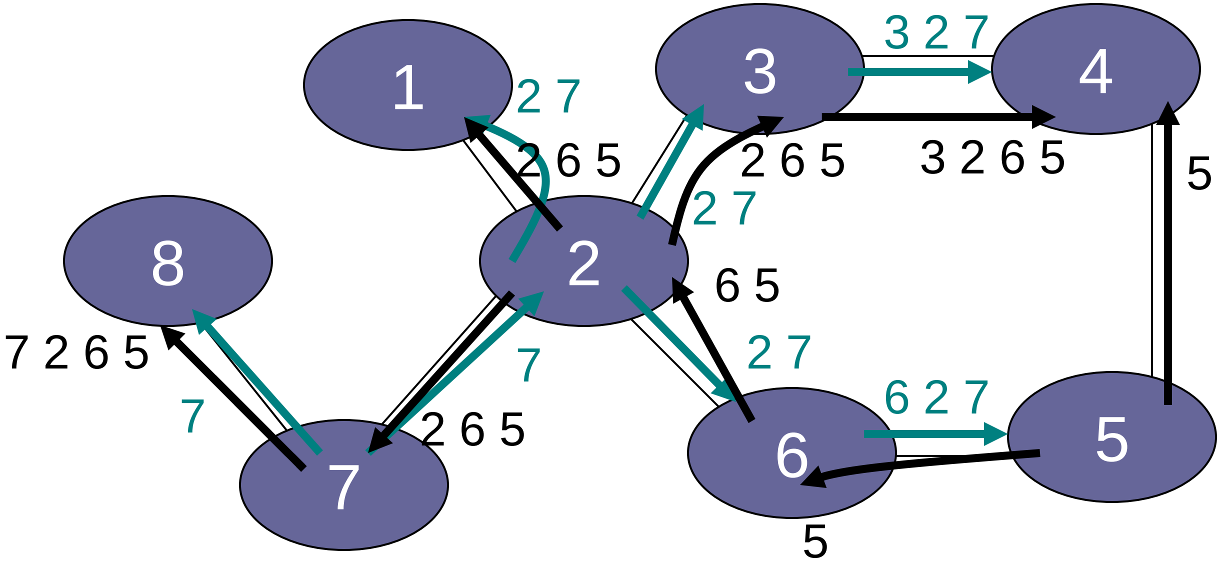
- G1’s link to A goes down
- G2 is advertising a pretty good route to G1 (cost=2)
- G1’s packets to A are forever looping between G2 and G1
- G1 is now advertising a route to A with cost=3, so G2 updates its own route to A via G1 to have cost=4, and so on
    - G1 and G2 are slowly counting to infinity
    - Split-horizon updates only prevent two-node loops

Overview of BGP

- BGP is a path-vector protocol between ASes
- Just like distance-vector, but routing updates contain an actual path to destination node
  - The list of traversed ASes and the set of network prefixes belonging to the first AS on the list
- Each BGP router receives update messages from neighbors, selects one “best” path for each prefix, and advertises this path to its neighbors
  - Can be the shortest path, but doesn’t have to be
    - “Hot-potato” vs. “cold-potato” routing
  - Always route to the most specific prefix for a destination

BGP Example

[Wetherall]



- AS 2 provides transit for AS 7
  - Traffic to and from AS 7 travels through AS 2

Some (Old) BGP Statistics

- BGP routing tables contain about 125,000 address prefixes mapping to about 17-18,000 paths
- Approx. 10,000 BGP routers
- Approx. 2,000 organizations own AS
- Approx. 6,000 organizations own prefixes
- Average route length is about 3.7
- 50% of routes have length less than 4 ASes
- 95% of routes have length less than 5 ASes

BGP Misconfiguration

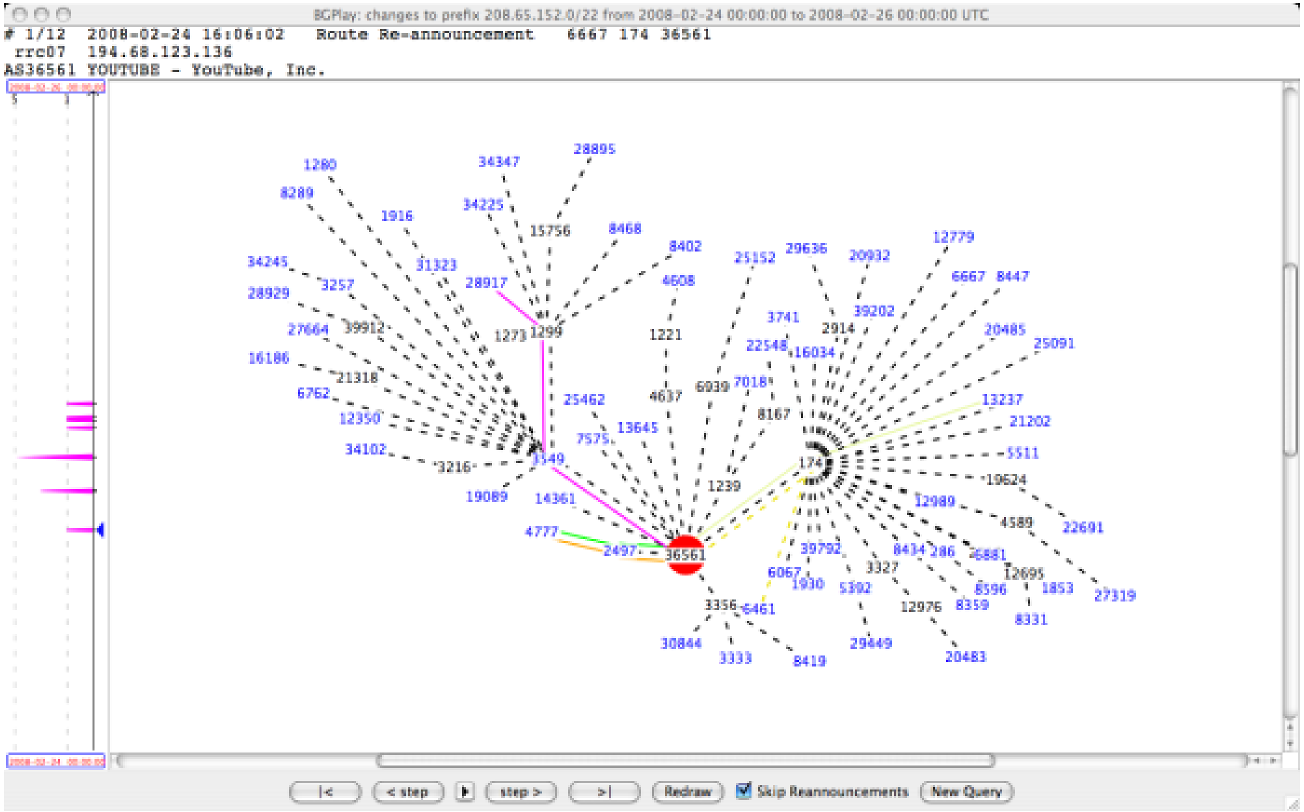
- Domain advertises good routes to addresses it does not know how to reach
  - Result: packets go into a network "black hole"
- April 25, 1997: "The day the Internet died"
  - AS7007 (Florida Internet Exchange) de-aggregated the BGP route table and re-advertised all prefixes as if it originated paths to them
    - In effect, AS7007 was advertising that it has the best route to every host on the Internet
  - Huge network instability as incorrect routing data propagated and routers crashed under traffic

BGP (In)Security

- BGP update messages contain no authentication or integrity protection
- Attacker may falsify the advertised routes
  - Modify the IP prefixes associated with a route
    - Can blackhole traffic to certain IP prefixes
  - Change the AS path
    - Either attract traffic to attacker's AS, or divert traffic away
    - Interesting economic incentive: an ISP wants to dump its traffic on other ISPs without routing their traffic in exchange
  - Re-advertise/propagate AS path without permission
    - For example, a multi-homed customer may end up advertising transit capability between two large ISPs

YouTube (Normally)

- AS36561 (YouTube) advertises 208.65.152.0/22

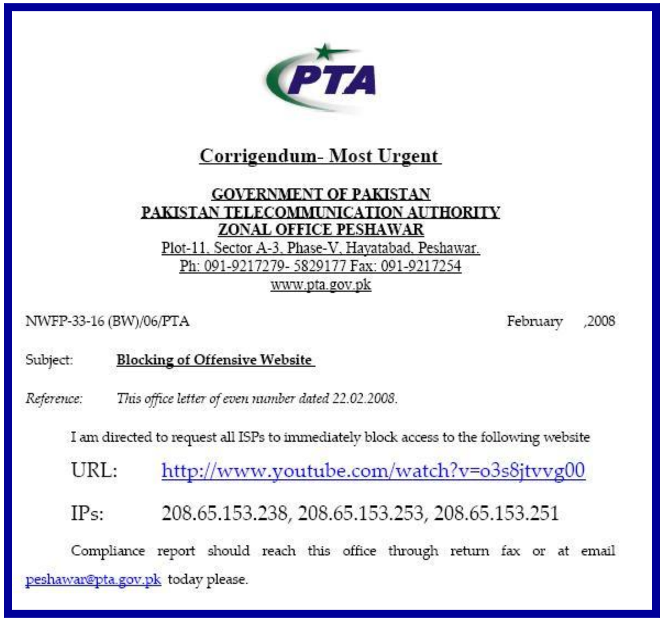


BGPlay: changes to prefix 208.65.152.0/22 from 2008-02-24 00:00:00 to 2008-02-26 00:00:00 UTC

```
# 1/12  2008-02-24 16:06:02   Route Re-announcement   6667 174 36561
 rrc07  194.68.123.136
AS36561 YOUTUBE - YouTube, Inc.
```

February 24, 2008

- Pakistan government wants to block YouTube



PTA

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**

Reference: This office letter of even number dated 22.02.2008.

I am directed to request all ISPs to immediately block access to the following website

URL: http://www.youtube.com/watch?v=o3s8jtvvg00

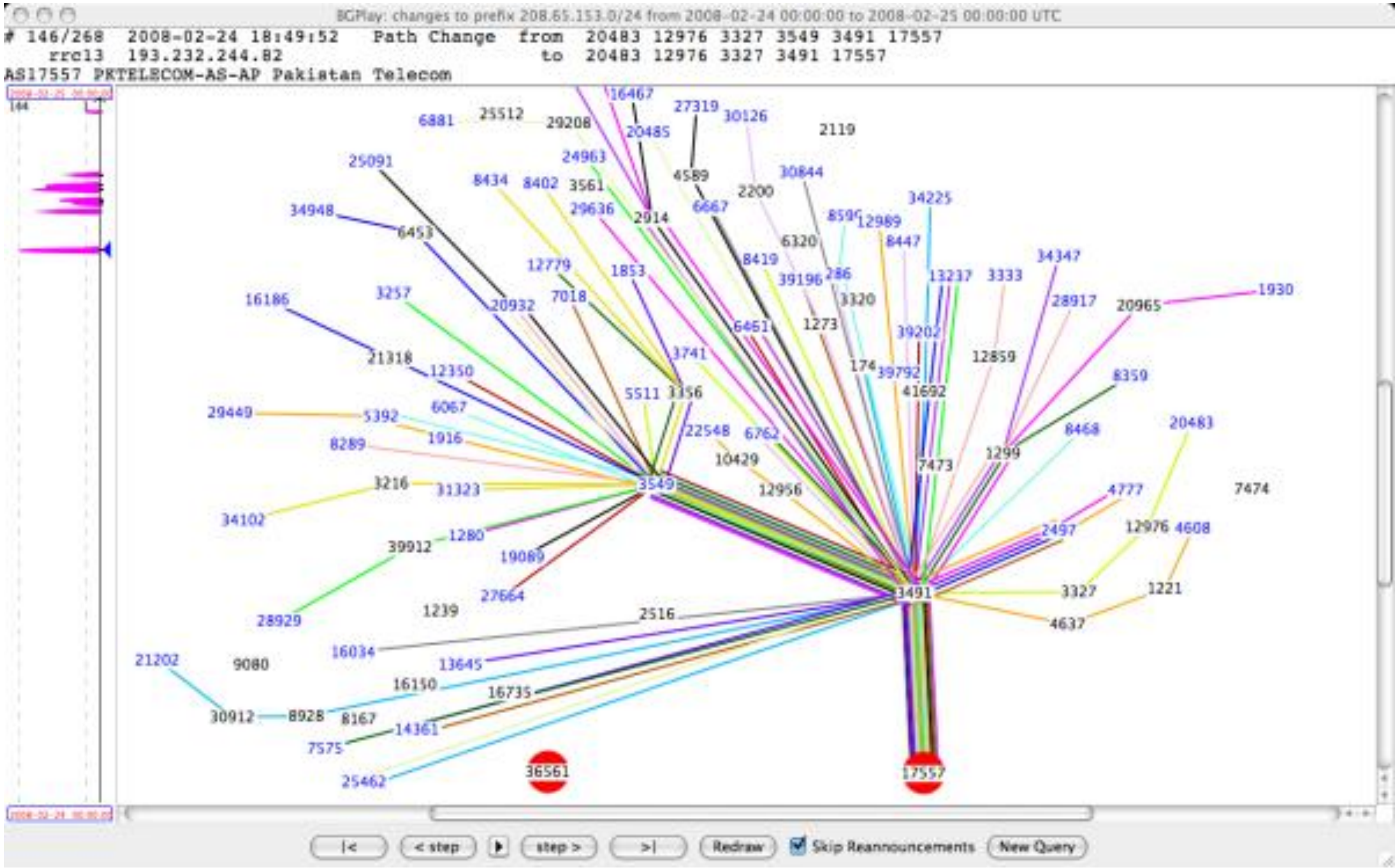
IPs: 208.65.153.238, 208.65.153.253, 208.65.153.251

Compliance report should reach this office through return fax or at email peshawar@pta.gov.pk today please.

- AS17557 (Pakistan Telecom) advertises 208.65.153.0/24 outwards
  - All YouTube traffic worldwide directed to AS17557

More specific than the /22 prefix advertised by YouTube itself

Two-Hour YouTube Outage



```
# 146/268  2008-02-24 18:49:52   Path Change  from  20483 12976 3327 3549 3491 17557
    rrc13  193.232.244.82                     to  20483 12976 3327 3491 17557
AS17557 PKTELECOM-AS-AP Pakistan Telecom
```

Other BGP Incidents

- May 2003: Spammers hijack unused block of IP addresses belonging to Northrop Grumman
  - Entire Northrop Grumman ends up on spam blacklist
  - Took two months to reclaim ownership of IP addresses
- Dec 2004: Turkish ISP advertises routes to the entire Internet, including Amazon, CNN, Yahoo
- Apr 2010: Small Chinese ISP advertises routes to 37,000 networks, incl. Dell, CNN, Apple
- Feb-May 2014: Someone uses BGP to hijack the addresses of Bitcoin mining-pool servers, steals $83,000 worth of Bitcoins

Preventing Prefix Hijacking

- Origin authentication
  - Secure database lists which AS owns which IP prefix
- soBGP
  - Digitally signed certificates of prefix ownership
- Prefix hijacking is not the only threat… in general, BGP allows ASes to advertise bogus routes
  - Remove another AS from a path to make it look shorter, more attractive, get paid for routing traffic
  - Add another AS to a path to trigger loop detection, make your connectivity look better

Securing BGP

- Dozens of proposals, various combinations of cryptographic mechanisms and anomaly detection
  - IRV, SPV, psBGP, Pretty Good BGP, PHAS, Whisper…
  - Example: Secure BGP (S-BGP)
    - Origin authentication + entire AS path digitally signed
    - Can verify that the route is recent, no ASes have been added or removed, the order of ASes is correct
- How many of these have been deployed?

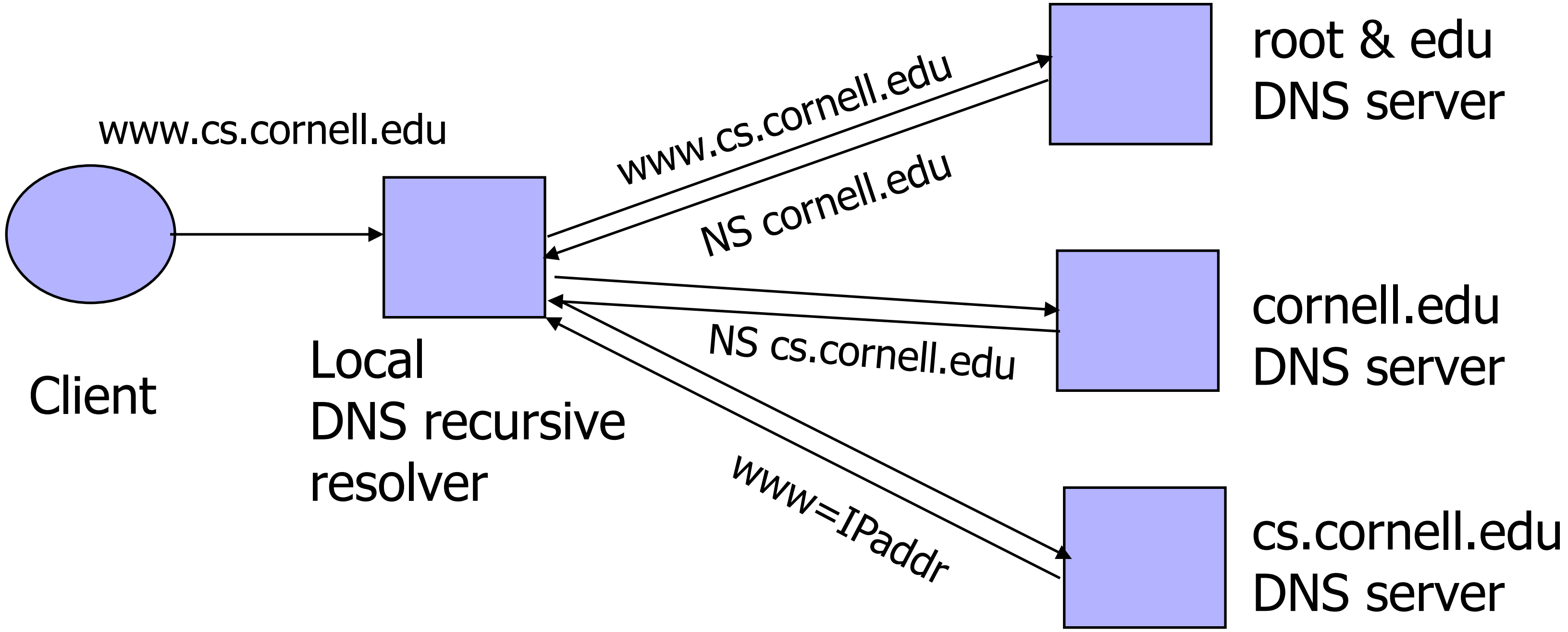
None

- No complete, accurate registry of prefix ownership
- Need a public-key infrastructure
- Cannot react rapidly to changes in connectivity
- Cost of cryptographic operations
- Not deployable incrementally

DNS: Domain Name Service

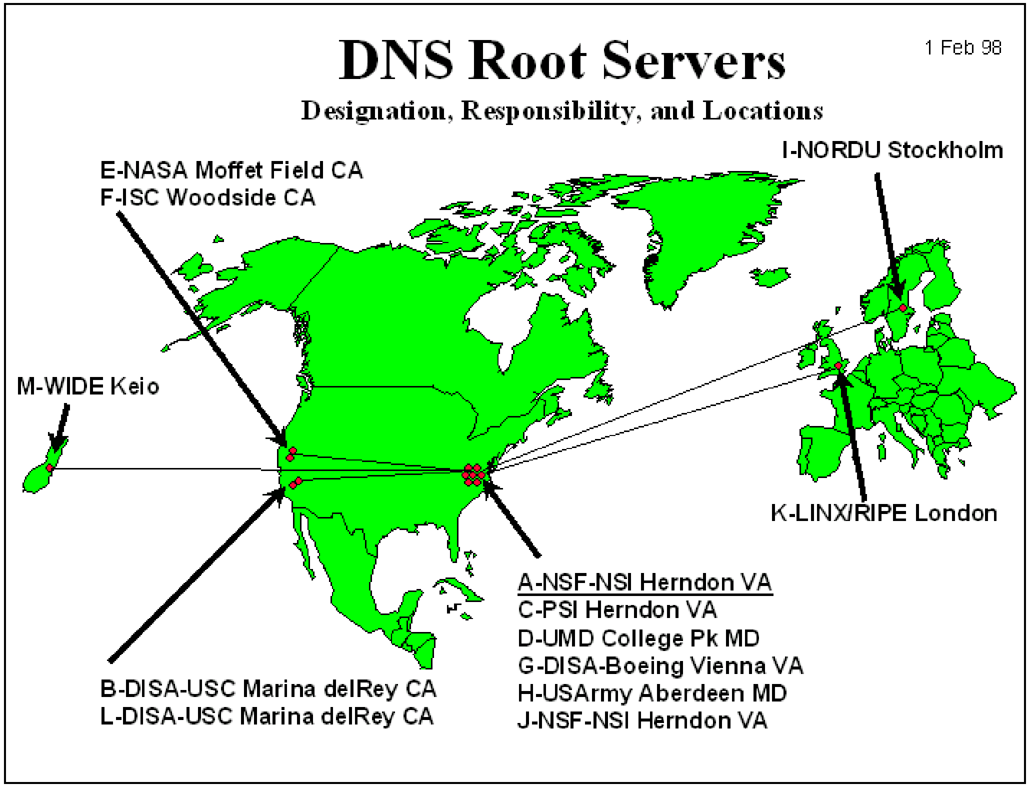
DNS maps symbolic names to numeric IP addresses

(for example, www.cs.cornell.edu ↔ 128.84.154.137)



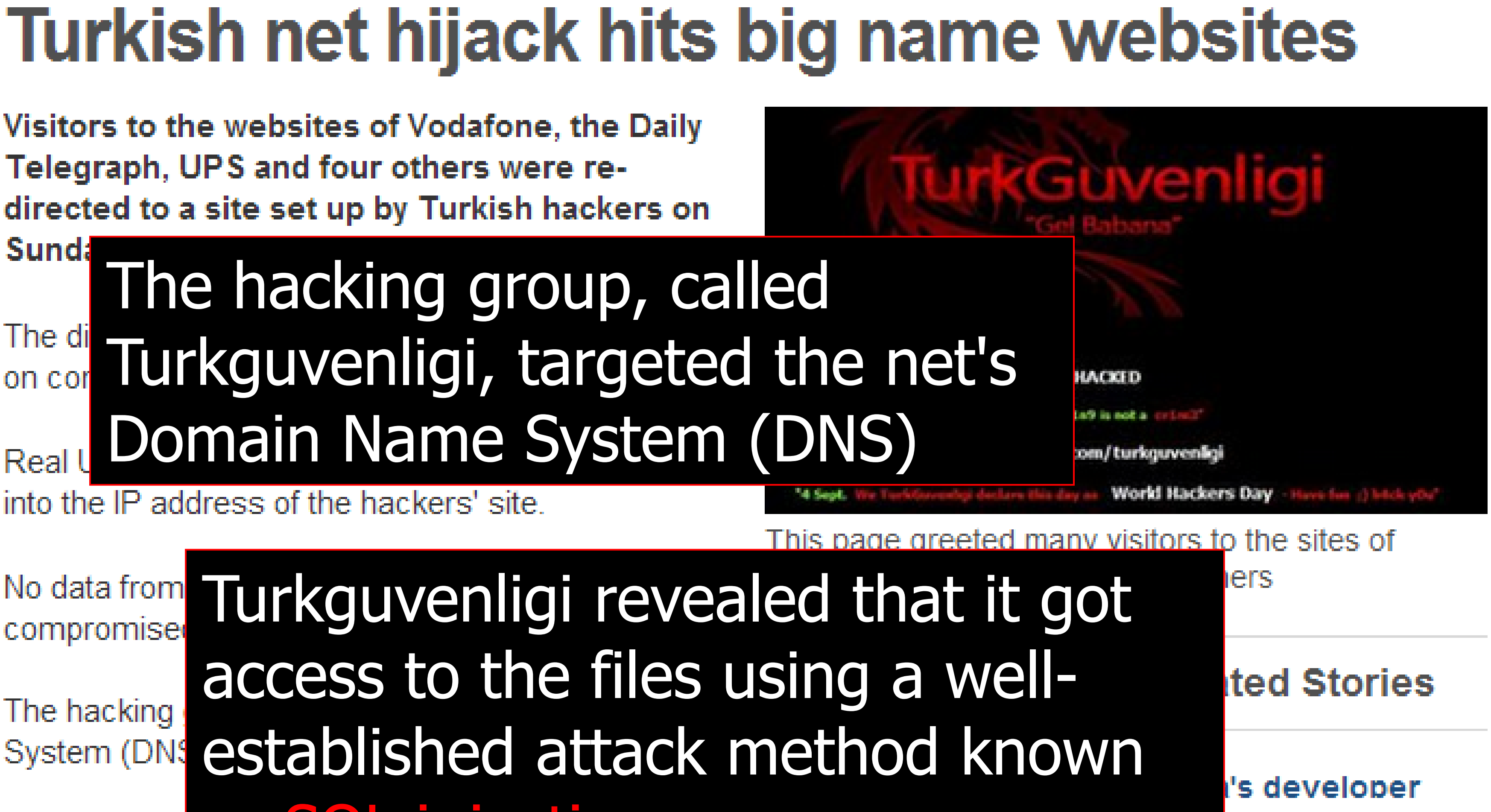
DNS Root Name Servers

- Root name servers for top-level domains
- Authoritative name servers for subdomains
- Local name resolvers contact authoritative servers when they do not know a name



Feb 6, 2007: Botnet DoS attack on root DNS servers

Turkish net hijack hits big name websites



Visitors to the websites of Vodafone, the Daily Telegraph, UPS and four others were re-directed to a site set up by Turkish hackers on Sunda

The di
on co

Real U
into the IP address of the hackers' site.

No data from
compromise

The hacking
System (DN

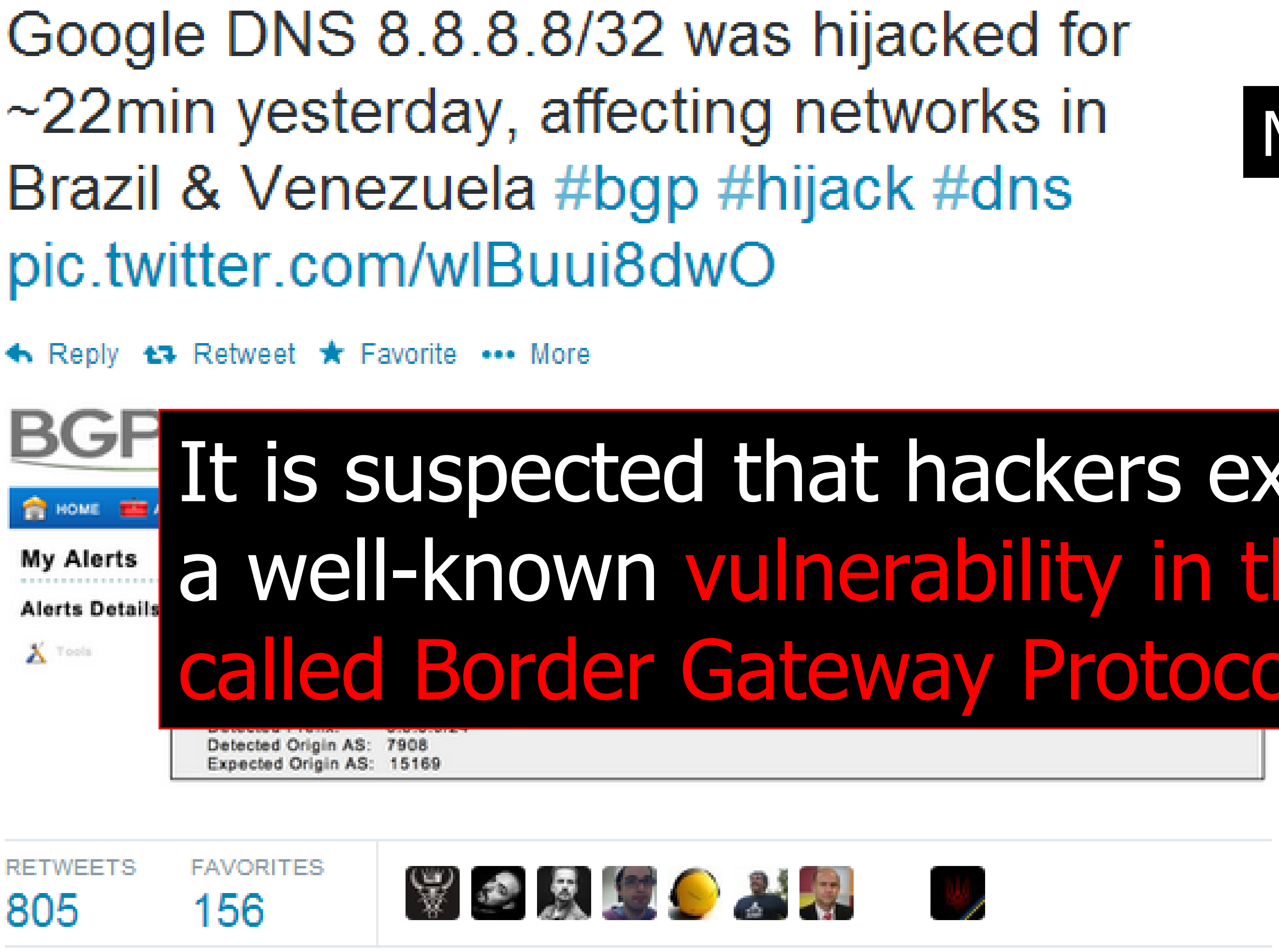
This page greeted many visitors to the sites of
ers

ted Stories

's developer

The hacking group, called Turkguvenligi, targeted the net's Domain Name System (DNS)

Turkguvenligi revealed that it got access to the files using a well-established attack method known as SQL injection



Google DNS 8.8.8.8/32 was hijacked for ~22min yesterday, affecting networks in Brazil & Venezuela #bgp #hijack #dns pic.twitter.com/wlBuui8dwO

Detected Origin AS: 7908
Expected Origin AS: 15169

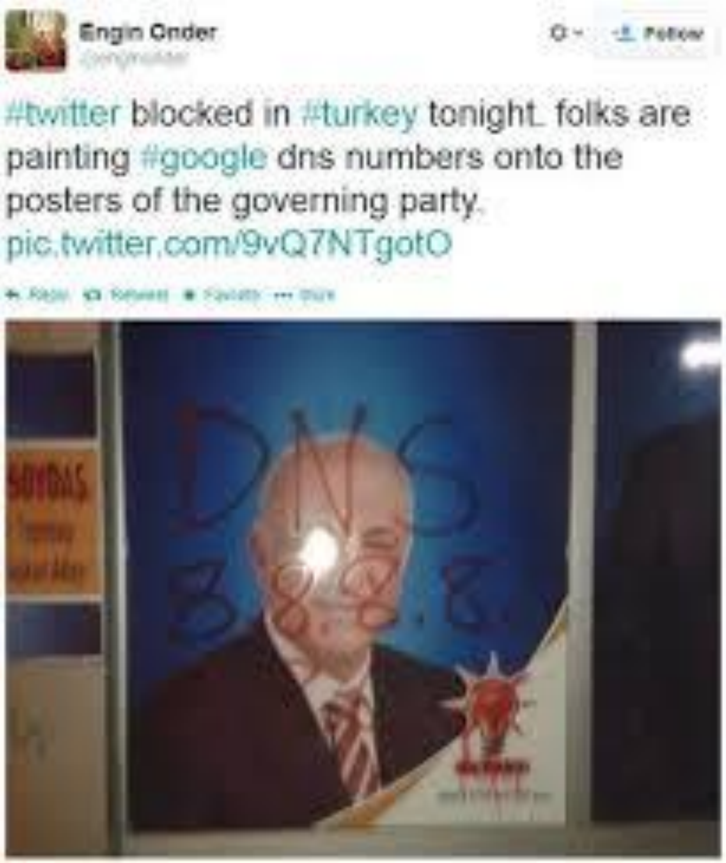
RETWEETS 805

FAVORITES 156

March 16, 2014

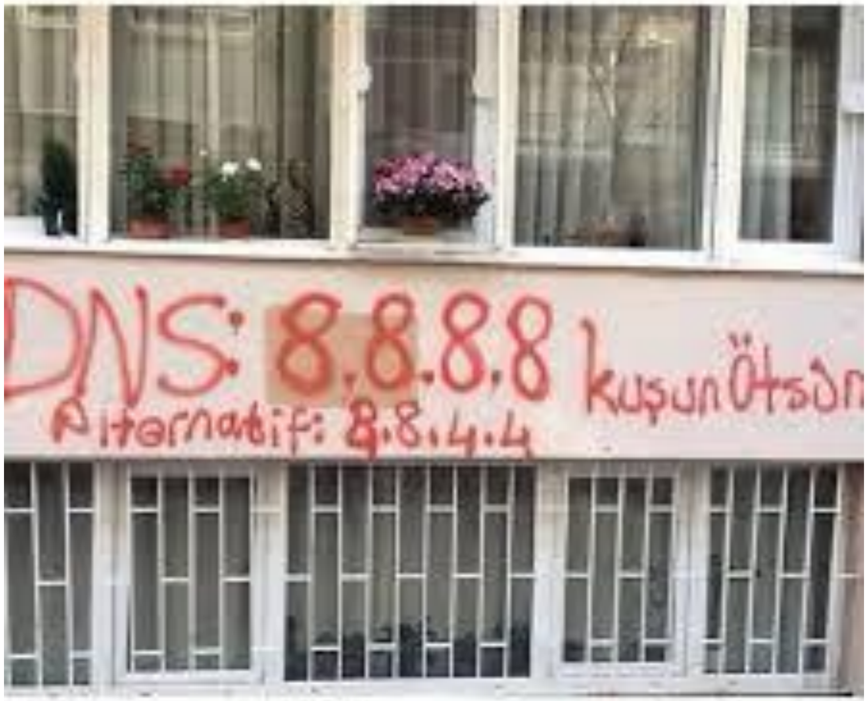
It is suspected that hackers exploited a well-known vulnerability in the so-called Border Gateway Protocol (BGP)

Turkey (2014)

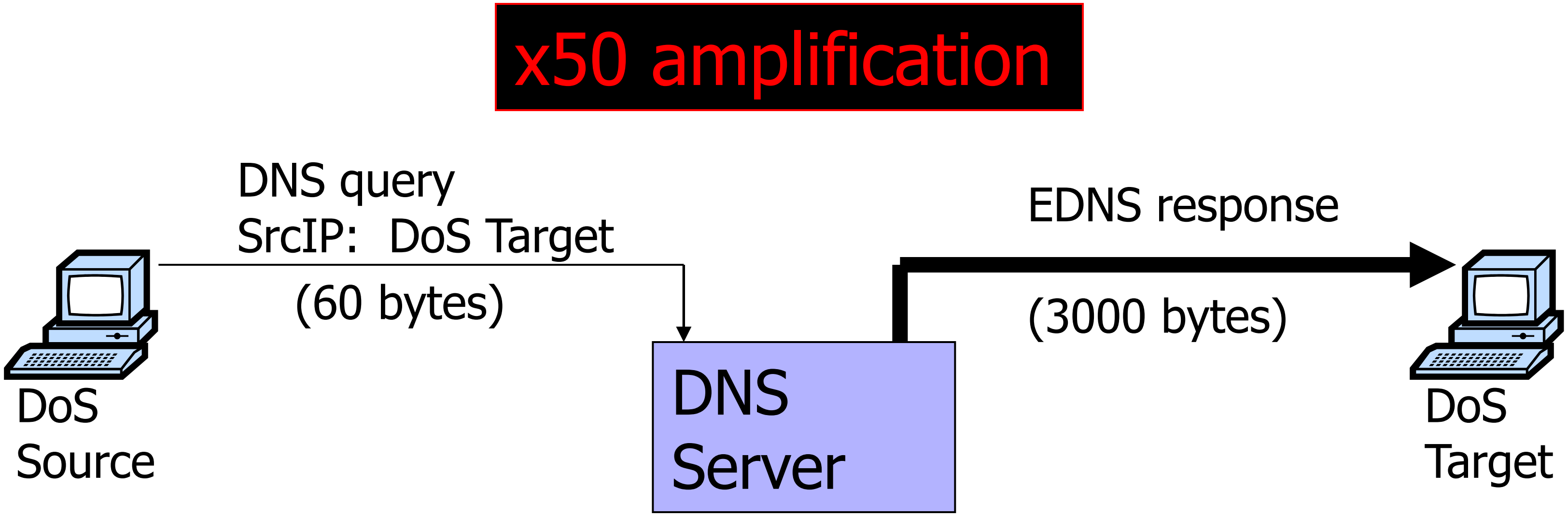


Engin Onder
@enginonder

#twitter blocked in #turkey tonight. folks are painting #google dns numbers onto the posters of the governing party. pic.twitter.com/9vQ7NTgotO



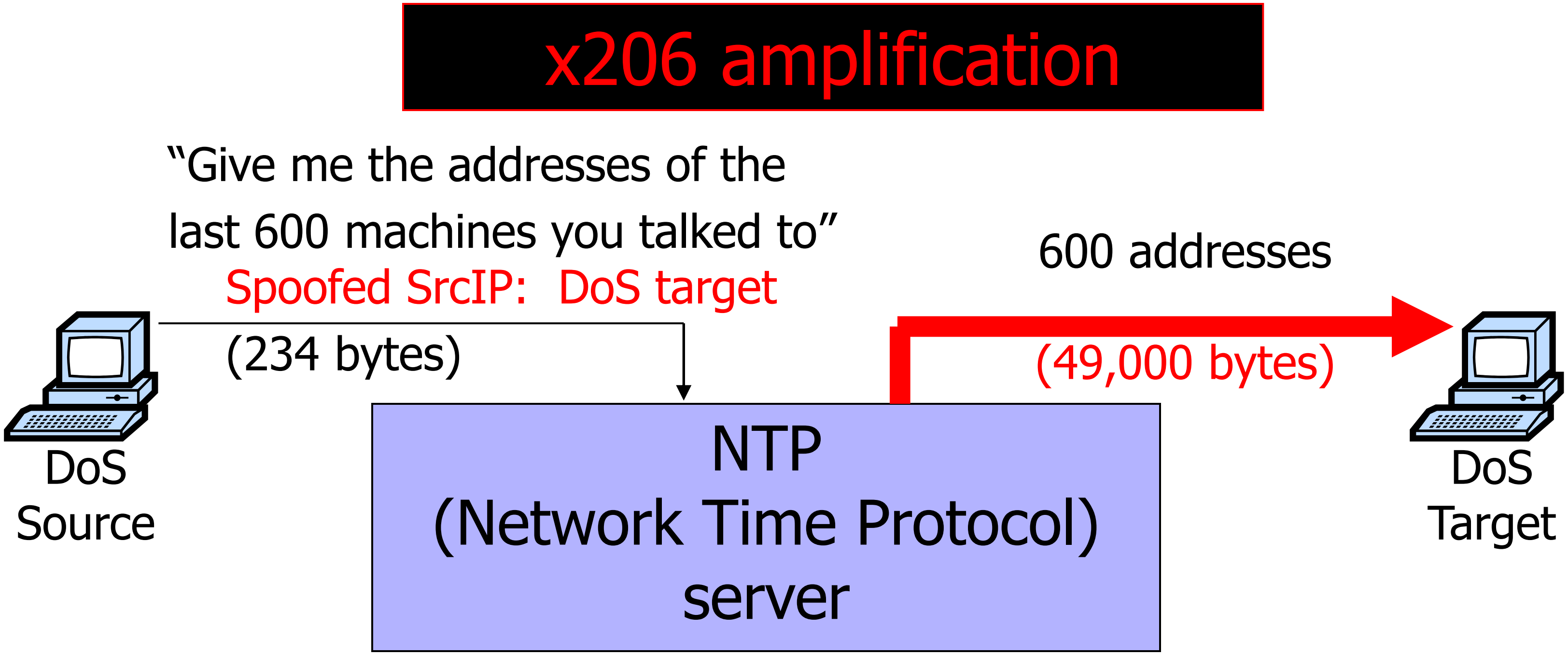
DNS Amplification Attack



2006: 0.58M open resolvers on Internet (Kaminsky-Shiffman)
2013: 21.7M open resolvers (openresolverproject.org)

March 2013: 300 Gbps DDoS attack on Spamhaus

(Not Just DNS)



x206 amplification

"Give me the addresses of the last 600 machines you talked to"

Spoofed SrcIP: DoS target

(234 bytes)

600 addresses

(49,000 bytes)

DoS Source

NTP (Network Time Protocol) server

DoS Target

December 2013 – February 2014:

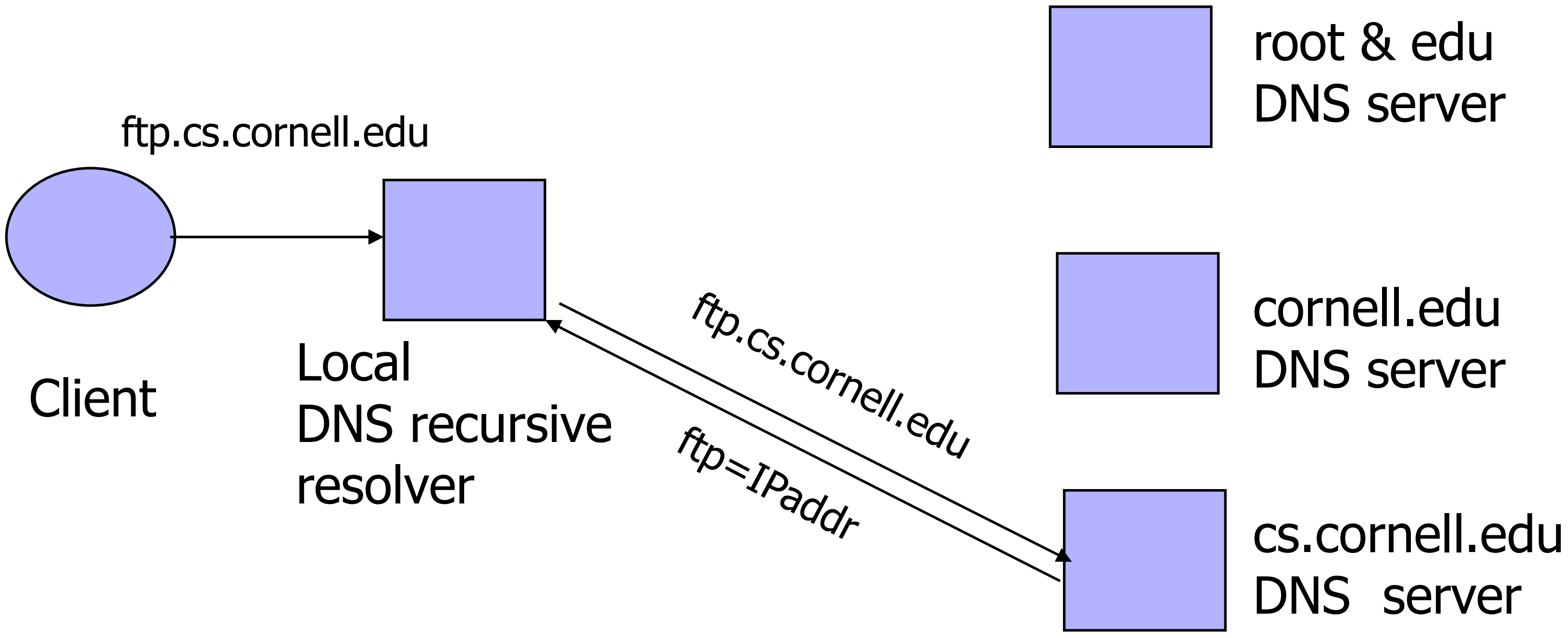
400 Gbps DDoS attacks involving 4,529 NTP servers

7 million unsecured NTP servers on the Internet (Arbor)

DNS Caching

- DNS responses are cached
  - Quick response for repeated translations
  - Other queries may reuse some parts of lookup
    - NS records identify name servers responsible for a domain
- DNS negative queries are cached
  - Don’t have to repeat past mistakes (misspellings, etc.)
- Cached data periodically times out
  - Lifetime (TTL) of data controlled by owner of data, passed with every record

Cached Lookup Example



ftp.cs.cornell.edu

Client

Local
DNS recursive
resolver

ftp.cs.cornell.edu

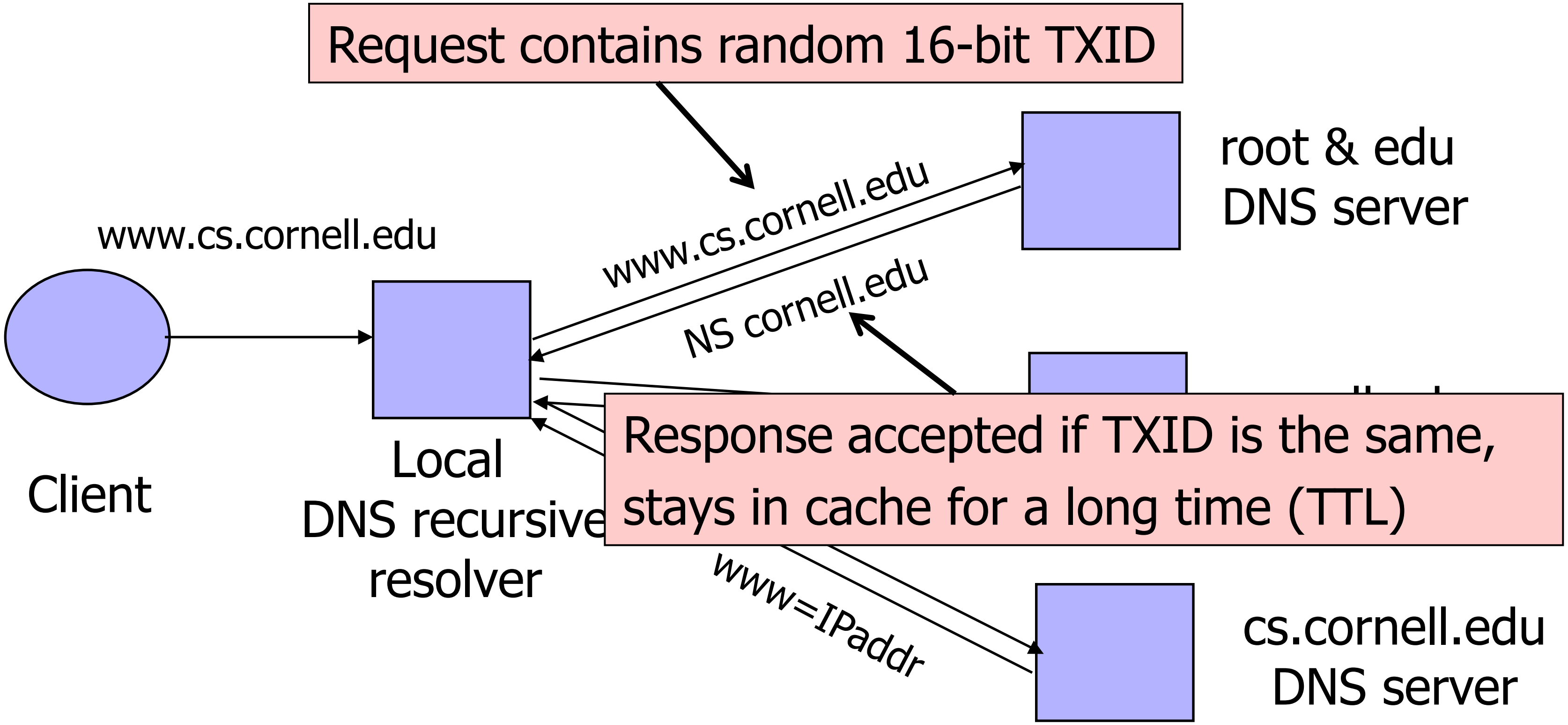
ftp=IPaddr

root & edu
DNS server

cornell.edu
DNS server

cs.cornell.edu
DNS server

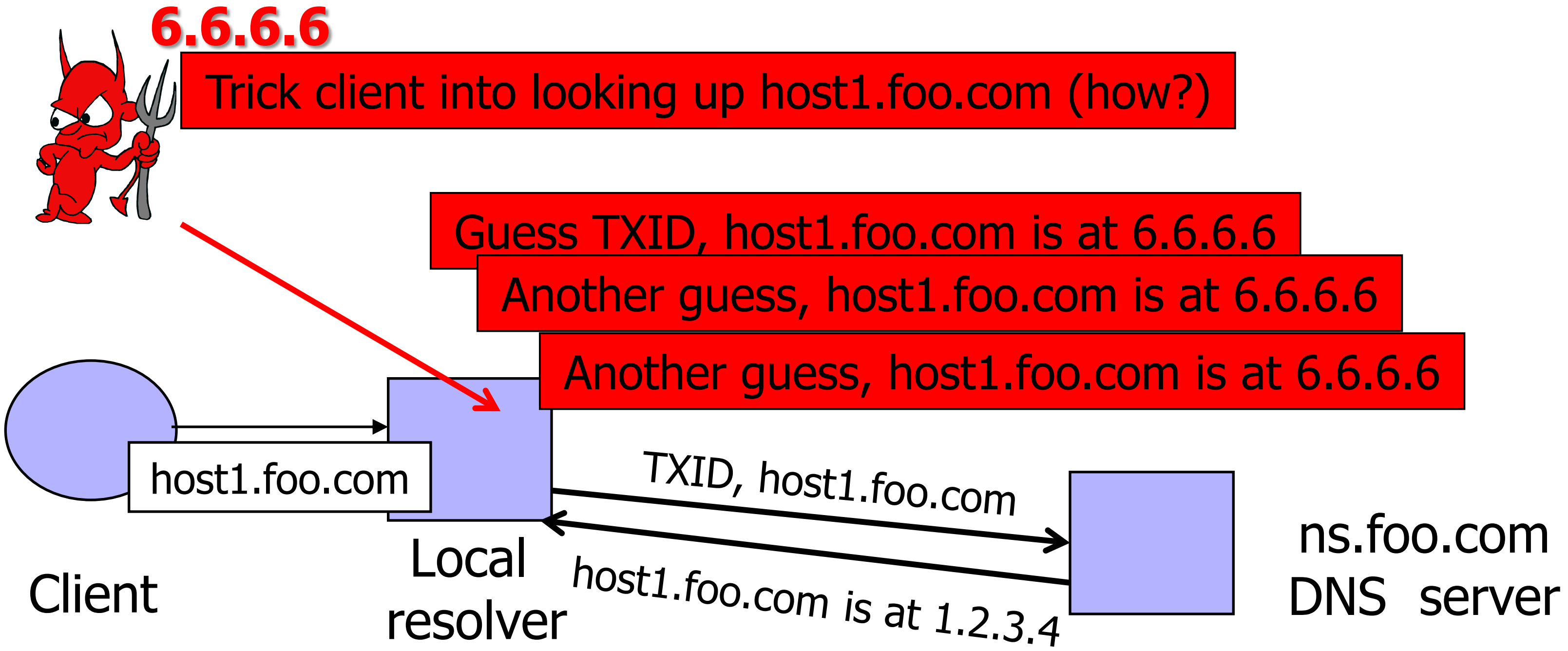
DNS “Authentication”



Request contains random 16-bit TXID

Response accepted if TXID is the same, stays in cache for a long time (TTL)

DNS Spoofing



6.6.6.6

Trick client into looking up host1.foo.com (how?)

Guess TXID, host1.foo.com is at 6.6.6.6

Another guess, host1.foo.com is at 6.6.6.6

Another guess, host1.foo.com is at 6.6.6.6

host1.foo.com

TXID, host1.foo.com

host1.foo.com is at 1.2.3.4

Client

Local resolver

ns.foo.com DNS server

Several opportunities to win the race.

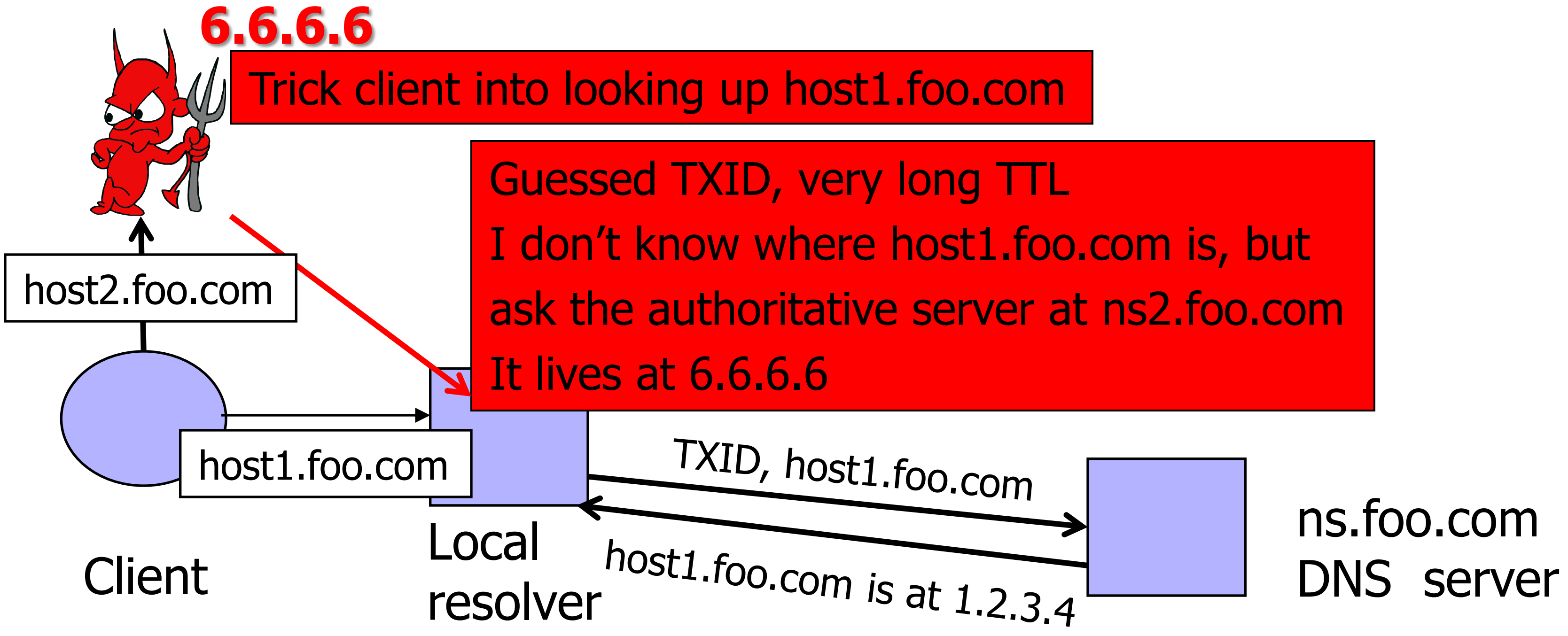
If attacker loses, has to wait until TTL expires…

… but can try again with host2.foo.com, host3.foo.com, etc.

… but what’s the point of hijacking host3.foo.com?

Exploiting Recursive Resolving

[Kaminsky]



Trick client into looking up host1.foo.com

Guessed TXID, very long TTL
I don’t know where host1.foo.com is, but
ask the authoritative server at ns2.foo.com
It lives at 6.6.6.6

If win the race, any request for XXX.foo.com will go to 6.6.6.6

The cache is poisoned… for a very long time!

No need to win future races!

If lose, try again with <ANYTHING>.foo.com

Triggering a Race

- Any link, any image, any ad, anything can cause a DNS lookup
  - No JavaScript required, though it helps
- Mail servers will look up what bad guy wants
  - On first greeting: HELO
  - On first learning who they're talking to: MAIL FROM
  - On spam check (oops!)
  - When trying to deliver a bounce
  - When trying to deliver a newsletter
  - When trying to deliver an actual response from an actual employee

Reverse DNS Spoofing

- Trusted access is often based on host names
  - Example: permit all hosts in .rhosts to run remote shell
- Network requests such as rsh or rlogin arrive from numeric source addresses
  - System performs reverse DNS lookup to determine requester's host name and checks if it's in .rhosts
- If attacker can spoof the answer to reverse DNS query, he can fool target machine into thinking that request comes from an authorized host
  - No authentication for DNS responses and typically no double-checking (numeric → symbolic → numeric)

Pharming

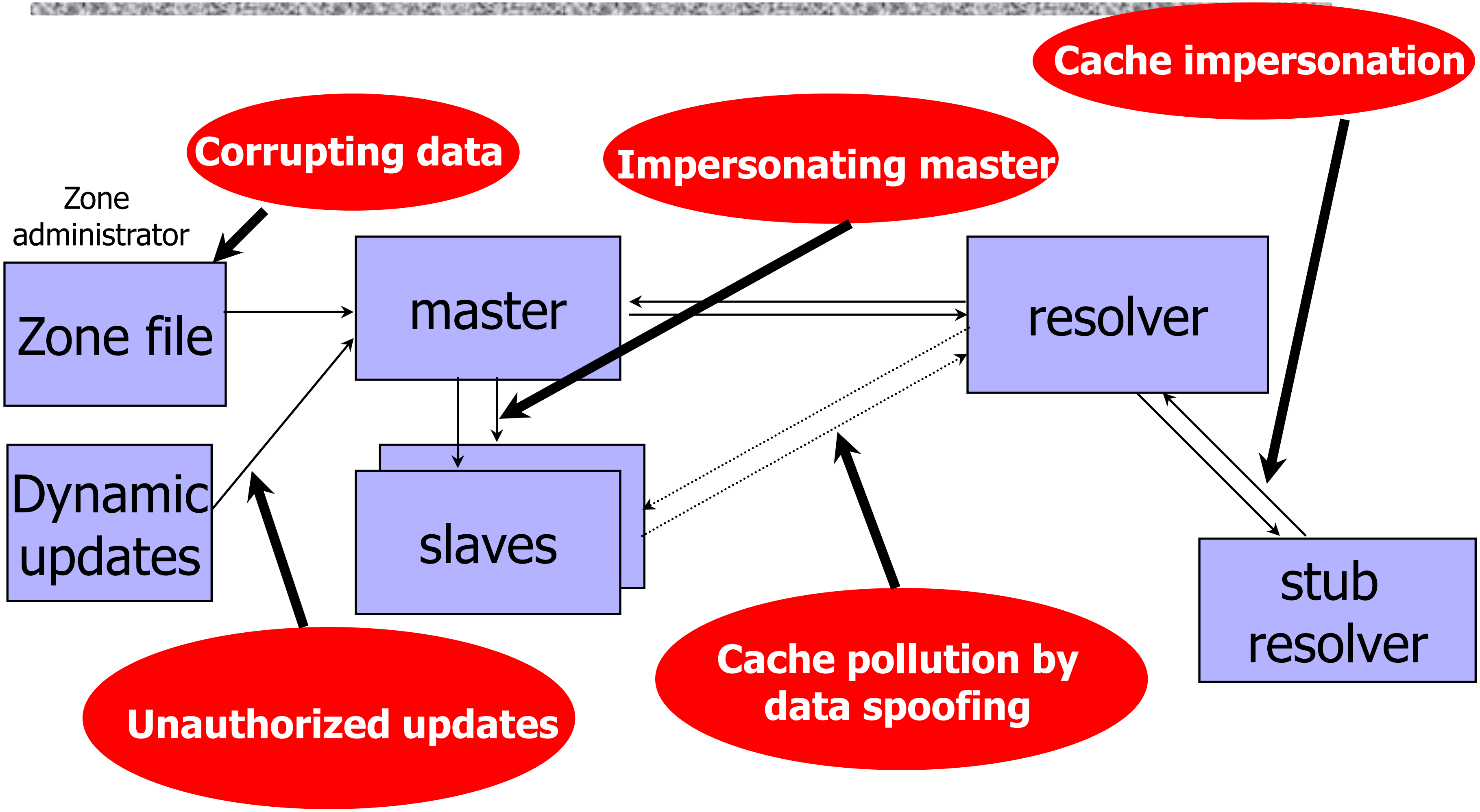
- Many anti-phishing defenses rely on DNS
- Can bypass them by poisoning DNS cache and/or forging DNS responses
  - Browser: “give me the address of www.paypal.com”
  - Attacker: “sure, it’s 6.6.6.6” (attacker-controlled site)
- Dynamic pharming
  - Provide bogus DNS mapping for a trusted server, trick user into downloading a malicious script
  - Force user to download content from the real server, temporarily provide correct DNS mapping
  - Malicious script and content have the same origin!

Other DNS Vulnerabilities

- DNS implementations have vulnerabilities
  - Multiple buffer overflows in BIND over the years
  - MS DNS for NT 4.0 crashes on chargen stream
- Denial of service
  - Oct ’02: ICMP flood took out 9 root servers for 1 hour
- Can use “zone transfer” requests to download DNS database and map out the network
  - “The Art of Intrusion”: NYTimes.com and Excite@Home

See http://cr.yp.to/djbdns/notes.html

DNS Vulnerabilities: Summary



Solving the DNS Spoofing Problem

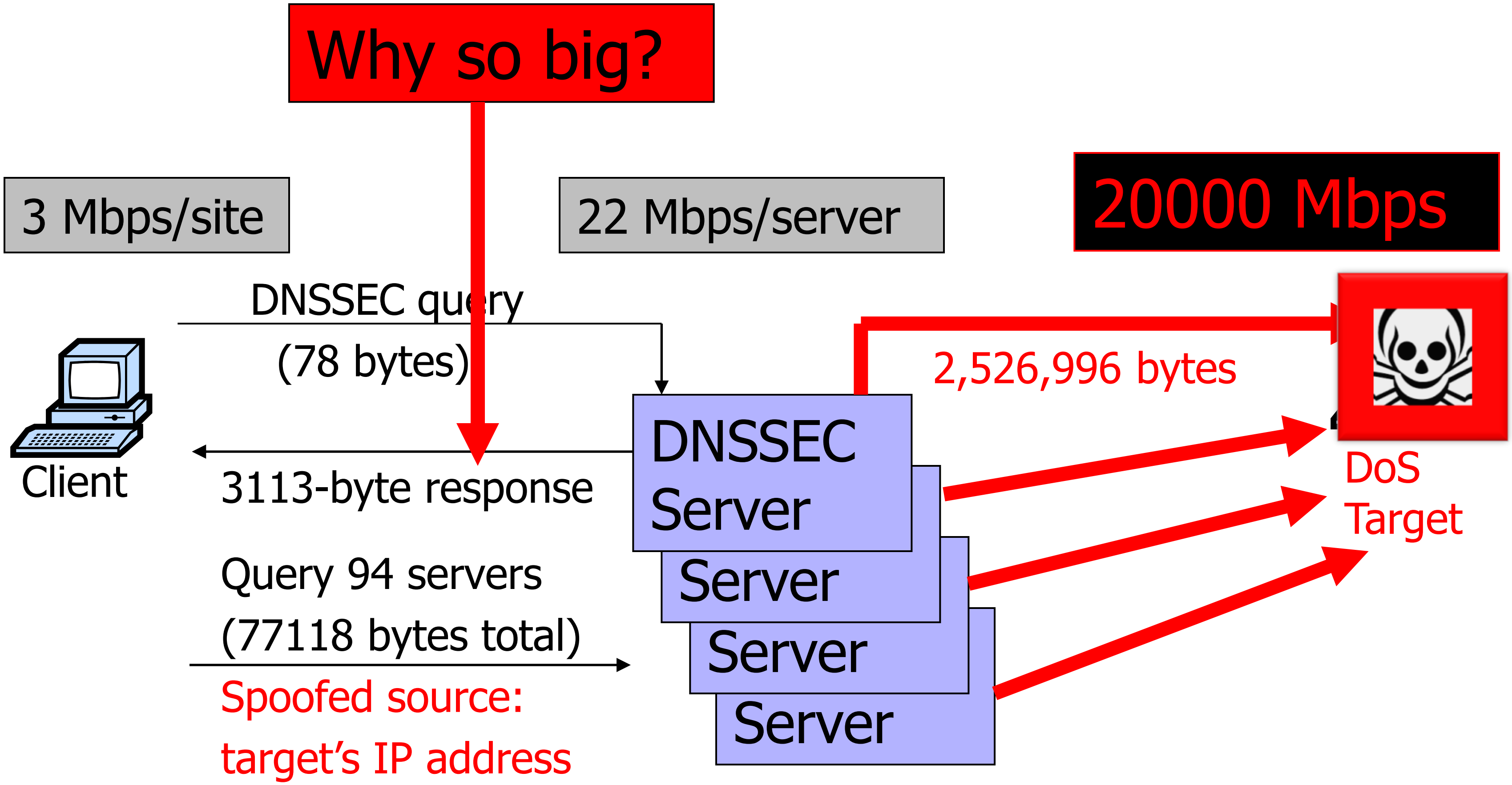
- Long TTL for legitimate responses
  - Does it really help?
- Randomize port in addition to TXID
  - 32 bits of randomness, makes it harder for attacker to guess TXID+port
- DNSSEC
  - Cryptographic authentication of host-address mappings

DNSSEC

- Goals: authentication and integrity of DNS requests and responses
- PK-DNSSEC (public key)
  - DNS server signs its data – done in advance
  - How do other servers learn the public key?
- SK-DNSSEC (symmetric key)
  - Encryption and MAC: $E_k(m, MAC(m))$
  - Each message contains a nonce to avoid replay
  - Each DNS node shares a symmetric key with its parent
  - Zone root server has a public key (hybrid approach)

Querying DNSSEC Servers

[Bernstein]



Why so big?

3 Mbps/site

22 Mbps/server

20000 Mbps

DNSSEC query

(78 bytes)

Client

3113-byte response

Query 94 servers

(77118 bytes total)

Spoofed source:

target’s IP address

DNSSEC Server

Server

Server

Server

2,526,996 bytes

DoS Target

5 times per second, from 200 sites

Using DNSSEC for DDoS

[Bernstein]

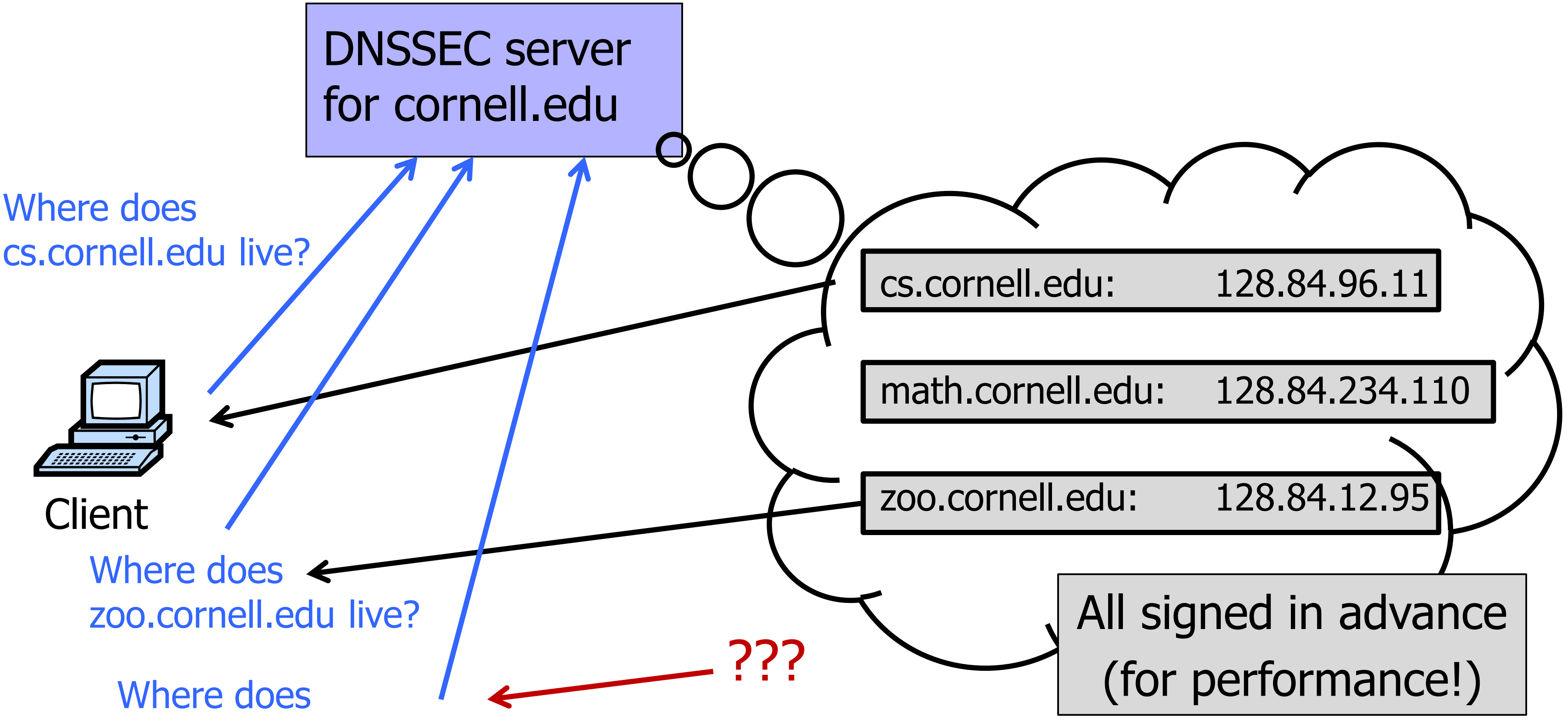
- RFC 4033 says:

"DNSSEC provides no protection against denial of service attacks"

- RFC 4033 doesn't say:

"DNSSEC is a remote-controlled double-barreled shotgun, the worst DDoS amplifier on the Internet"

DNSSEC In Action



DNSSEC server for cornell.edu

Where does cs.cornell.edu live?

Client

Where does zoo.cornell.edu live?

Where does DNSSECIsTehSuck.cornell.edu live?

???

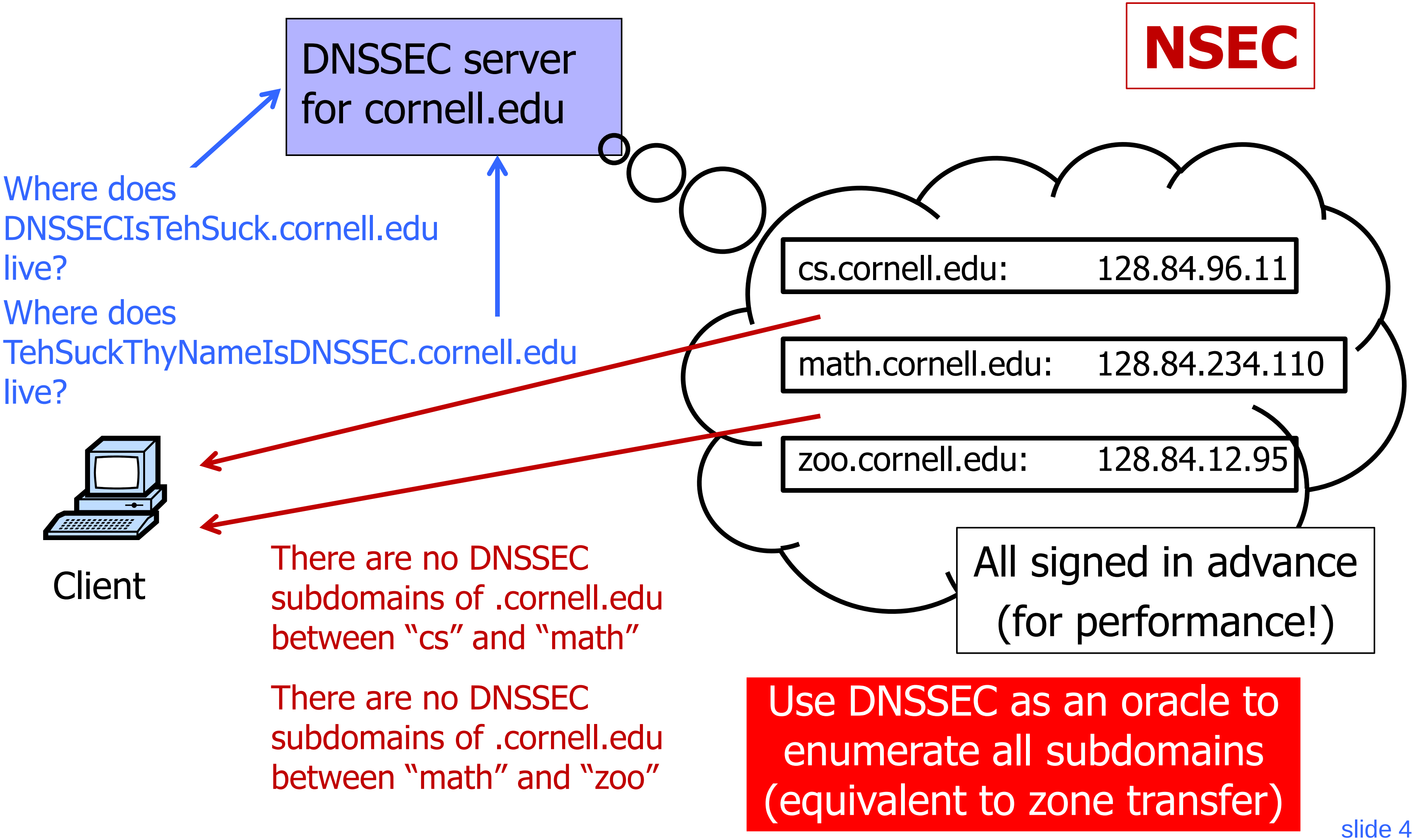
cs.cornell.edu:	128.84.96.11
math.cornell.edu:	128.84.234.110
zoo.cornell.edu:	128.84.12.95

All signed in advance (for performance!)

Each name has exactly one signed record

Why can’t the resolver simply send an empty record when queried for a domain that does not exist?

Authenticated Denial of Existence



NSEC

DNSSEC server for cornell.edu

Where does DNSSECIsTehSuck.cornell.edu live?

Where does TehSuckThyNameIsDNSSEC.cornell.edu live?

Client

cs.cornell.edu: 128.84.96.11

math.cornell.edu: 128.84.234.110

zoo.cornell.edu: 128.84.12.95

There are no DNSSEC subdomains of .cornell.edu between “cs” and “math”

There are no DNSSEC subdomains of .cornell.edu between “math” and “zoo”

All signed in advance (for performance!)

Use DNSSEC as an oracle to enumerate all subdomains (equivalent to zone transfer)

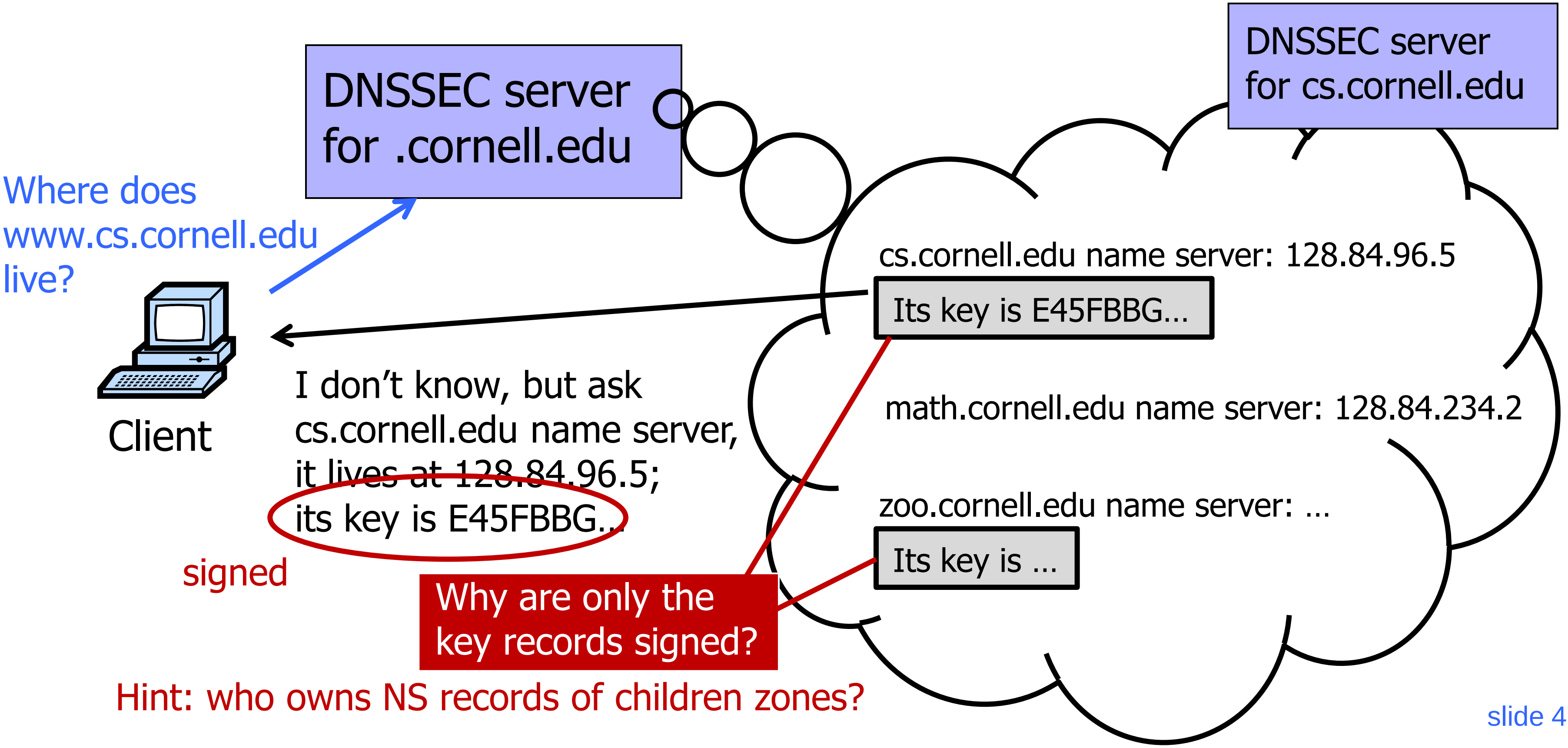
NSEC3

[Bernstein]

- Domain names hashed, hashes sorted in lexicographic order
- Denials of existence certify that there are no DNSSEC domains whose hash values fall into a certain interval
  - As opposed to actual domain names
- Are domain names random?
- Vulnerable to brute-force guessing attacks

Delegation in DNSSEC

- Delegation is essential for scalability
  - For example, there are 100,000,000 .com domains



Where does www.cs.cornell.edu live?

DNSSEC server for .cornell.edu

DNSSEC server for cs.cornell.edu

Client

I don't know, but ask cs.cornell.edu name server, it lives at 128.84.96.5; its key is E45FBBG...

signed

cs.cornell.edu name server: 128.84.96.5

Its key is E45FBBG…

math.cornell.edu name server: 128.84.234.2

zoo.cornell.edu name server: …

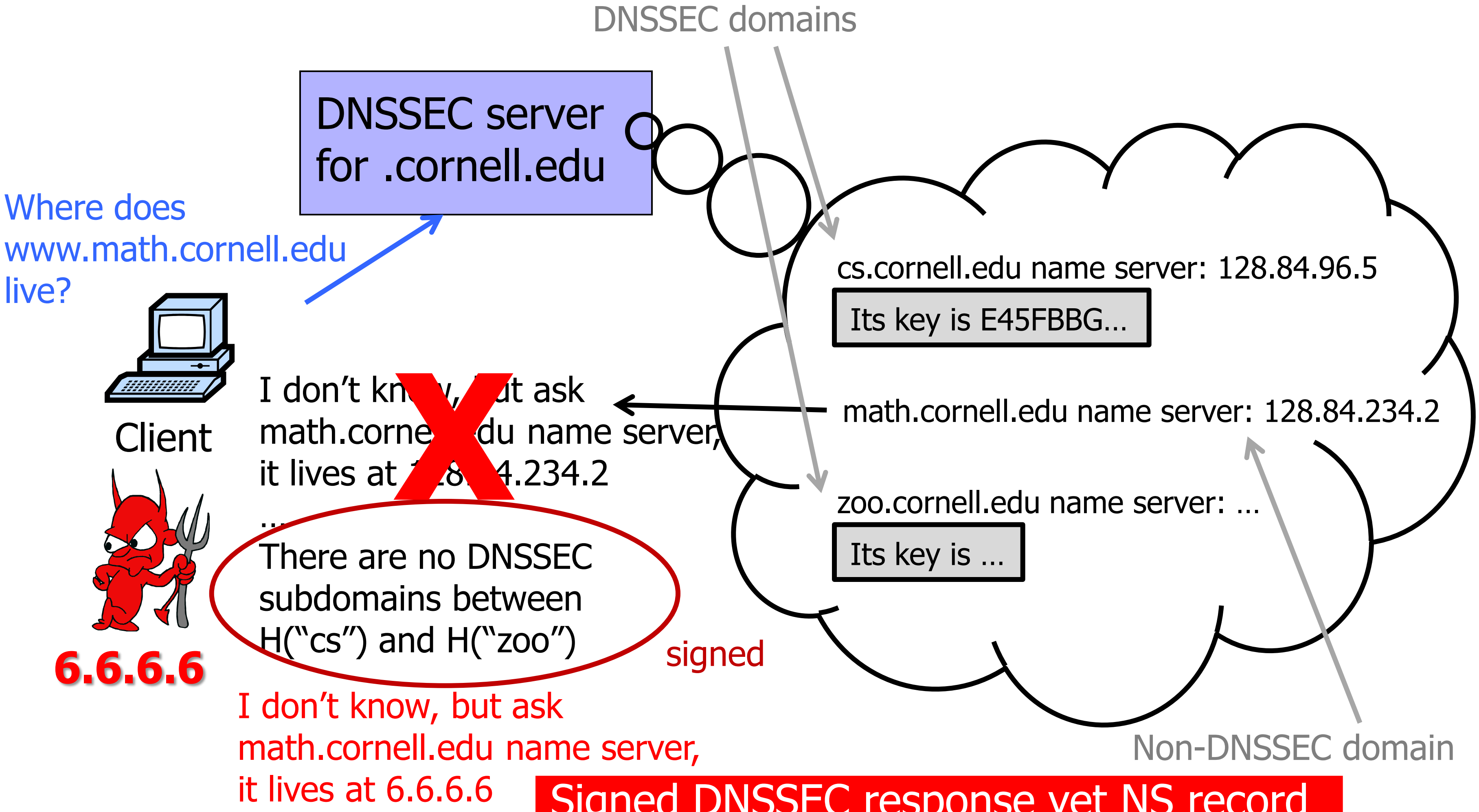
Its key is …

Why are only the key records signed?

Hint: who owns NS records of children zones?

Forging Delegation Responses

[Bernstein]



DNSSEC domains

DNSSEC server for .cornell.edu

Where does www.math.cornell.edu live?

Client

~~I don't know, but ask math.cornell.edu name server, it lives at 128.84.234.2~~

…

There are no DNSSEC subdomains between H("cs") and H("zoo")

signed

6.6.6.6

I don't know, but ask math.cornell.edu name server, it lives at 6.6.6.6

cs.cornell.edu name server: 128.84.96.5

Its key is E45FBBG…

math.cornell.edu name server: 128.84.234.2

zoo.cornell.edu name server: …

Its key is …

Non-DNSSEC domain

Signed DNSSEC response yet NS record has been forged… what happened??!!

Delegating to Secure Zones

- Q: When does verification of signatures on DNSSEC records actually happen?
- A: At the very end, when the resolver has the complete chain
- But the delegation record is not signed… what if it has been forged?
- Current DNSSEC deployments are only “secure” down to the ISP’s resolver
  - Stub resolvers on users’ machines only get an unsigned flag saying that the response is “secure”

DNSSEC “Features”

[Bernstein]

- Does nothing to improve DNS availability
- Allows astonishing levels of DDoS amplication, damaging Internet availability
  - Also CPU exhaustion attacks
- Does nothing to improve DNS confidentiality, leaks private DNS data (even with NSEC3)
- Does not prevent forgery of delegation records
- Does not protect the “last mile”
- Implementations suffered from buffer overflows