DNS Tampering and Root Servers

AMS-IX: 24 Nov 2010

Renesys Corporation
Martin A. Brown
Doug Madory
Alin Popescu
Earl Zmijewski

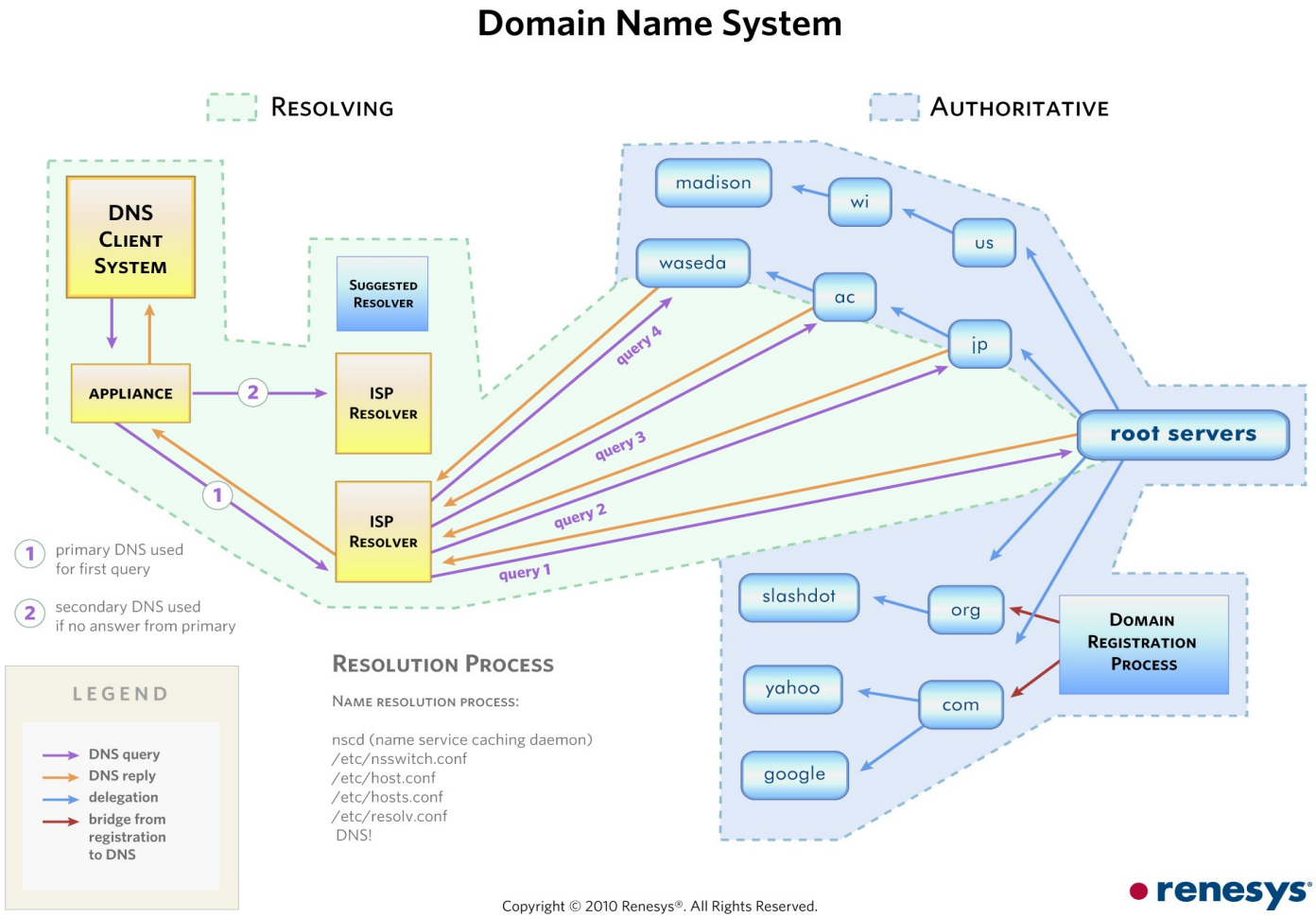
Overview

- Brief overview of Domain Name System (DNS)
- Demonstrating Great Firewall DNS tampering
- Root servers in Beijing, China
- Tampered reply from Beijing I-root (in March)
- Who could have been affected?
- The DNS Stars rating system

Brief (remedial) tour of DNS

- Application looks up a name (e.g. www.ams-ix.net.)
- A library/the IP stack calls out to a resolver
- The resolver returns a cached answer or …
- The resolver starts a series of queries
  - root server (.)
    - Q: www.ams-ix.net.
    - A: refer a.gtld-servers.net., b.gtld-servers.net. […]
  - gtld or cctld server (net.)
    - Q: www.ams-ix.net.
    - A: refer nemix1.ams-ix.net., nemix2.ams-ix.net., ns2.surfnet.nl.
  - an authoritatitive zone server (ams-ix.net.)
    - Q: www.ams-ix.net.
    - A: answer IPv4 91.200.16.42, IPv6 2001:67c:1a8:100::7
- The resolver caches ultimate answer (and intervening delegations) and returns the answer to the client.

Domain name system



renesys

Chinese firewall

- The Great Firewall (GFW) is a national technical control designed to implement policy and is reported to ...
  - Blackhole access to certain IPs and entire prefixes
  - Intercept and return incorrect DNS responses
  - Intercept TCP connections, possibly injecting TCP resets
- In particular, DNS queries (and answers) passing through the GFW can ...
  - Return bogus answers
  - Affect users outside the Chinese Internet

Note: GFW is a term of convenience for the non-point-source effects of a distributed technical control.

Try the Chinese firewall yourself ...

- Try the following sample query several times...
  - dig @dns1.chinatelecom.com.cn. www.facebook.com. A
- Answers will vary …
  - www.facebook.com. 11556 IN A 37.61.54.158
  - www.facebook.com. 24055 IN A 78.16.49.15
  - www.facebook.com. 38730 IN A 203.98.7.65
- Obviously bogus results—none of these answers is in a prefix originated by AS 32934 (Facebook, Inc.)
  - 37.0.0.0/8 is one of the few remaining unallocated /8s
  - 78.16.0.0/14 is originated by AS 2110 (BT Ireland, IE)
  - 203.98.0.0/18 is originated by AS 4768 (TelstraClear, NZ)
- Queries are to a resolver inside China Telecom (but may or may not ever get there; check your packet capture).

GFW Tampering in action!

- Request for 'www.facebook.com' through GFW

1290077908.155737 IP (tos 0x0, ttl 64, id 42149, offset 0, flags [none], proto UDP (17), length 62)
SOURCE_IP.59340 > 211.100.35.136.53: [udp sum ok] 1824+
A? www.facebook.com. (34)

- First response, GFW replies with bogus response

1290077908.464369 IP (tos 0x20, ttl 55, id 24309, offset 0, flags [none], proto UDP (17), length 78)
211.100.35.136.53 > SOURCE_IP.59340: [udp sum ok] 1824 q:
A? www.facebook.com. 1/0/0 www.facebook.com. [5m] A 78.16.49.15 (50)

- Second response, real server (?) refuses

1290077908.494594 IP (tos 0x34, ttl 48, id 8933, offset 0, flags [none], proto UDP (17), length 62)
211.100.35.136.53 > SOURCE_IP.59340: [udp sum ok] 1824 Refused- q:
A? www.facebook.com. 0/0/0 (34)

- Third response, another bogus GFW reply

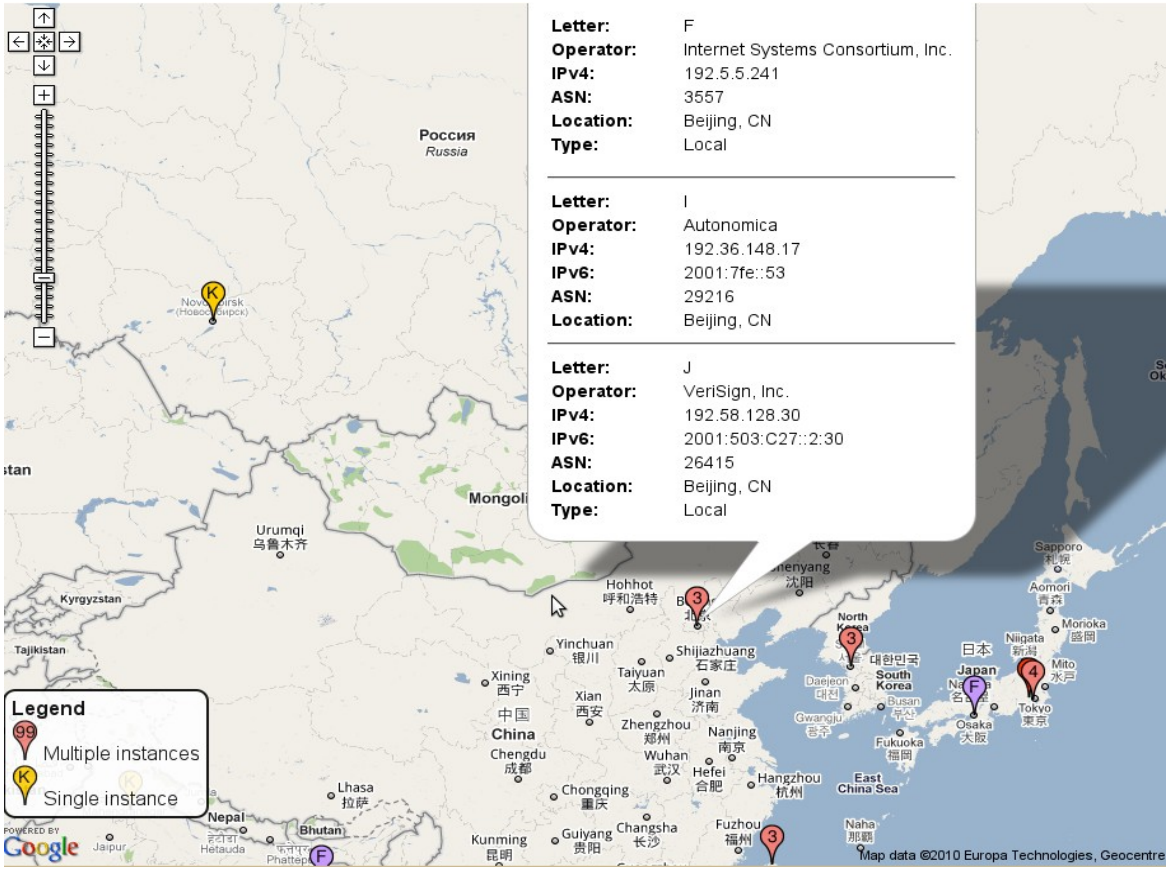
1290077908.496146 IP (tos 0x34, ttl 227, id 18283, offset 0, flags [none], proto UDP (17), length 94)
211.100.35.136.53 > SOURCE_IP.59340: [udp sum ok] 1824* q:
A? www.facebook.com. 1/0/0 www.facebook.com. [3h14m3s] A 159.106.121.75 (66)

What is the GFW doing?

- GFW uses the time-honored tradition of racing to get an answer to the client. [This well-known race was famously exploited in conjunction with authority records in the so-called Kaminsky DNS vulnerability.]
- The returned tampered results show that the set of bogus IPs returned all fall within a small set (in one day of lookups from a single location there were only 9 distinct bogus IPs).
- More details of behavior in 'The Great DNS Wall of China' http://cs.nyu.edu/~pcw216/work/nds/final.pdf (2007, Graham Lowe, Patrick Winters and Michael L. Marcus)
- General summary of the technical behavior of the Great Firewall in http://www.certmag.com/read.php?in=3906 by Shawn Conaway.

Root servers in China

There are three Beijing root server nodes: F-root, I-root and J-root.
Source: http://www.root-servers.org/map/



F-root: Just the facts

- IP address: 192.5.5.241
- Prefixes: 192.5.4.0/23 & 192.5.5.0/24
- Origin: AS 3557 (Dedicated to F-root)
- Primary upstream: AS 1280 (ISC)
  - AS 3557 has ~22 BGP adjacencies
  - AS 1280 has ~20 BGP adjacencies
  - F-root is run by ISC
  - F-root is *anycast* from around the world (47 instances)
    - 16 instances in EMEA
    - 14 in Asia Pacific
    - 17 in the Americas

I-root: Just the facts

- IP address: 192.36.148.17
- Prefixes: 192.36.148.0/23 & 192.36.148.0/24
- Origin: AS 29216 (Dedicated to I-root)
- Single Upstream: AS 8674 (Netnod)
  - AS 8674 has ~100 BGP adjacencies
  - I-root is run by Autonomica
  - Subsidiary of Sweden's Netnod
  - I-root is *anycast* from around the world (36 instances)
    - 15 instances in EMEA
    - 14 in Asia Pacific
    - 7 in the Americas

J-root: Just the facts

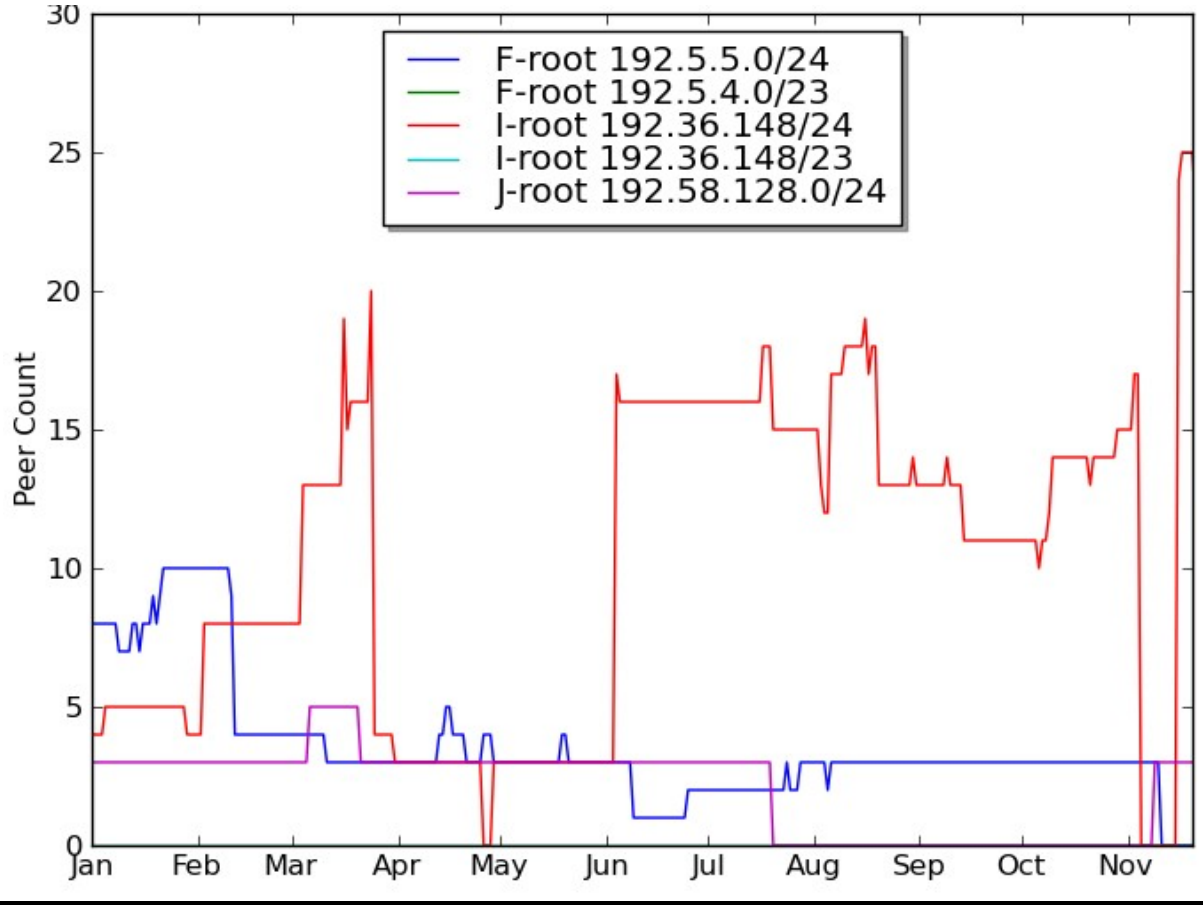
- IP address: 192.58.128.30
- Prefixes: 192.58.128.0/24
- Origin: AS 26415 (Verisign Global Registry)
- Diverse Upstreams
  - AS 26415 has ~43 BGP adjacencies
  - J-root is run by Verisign
  - J-root is *anycast* from around the world (70 instances)
    - 31 instances in EMEA
    - 14 in Asia Pacific
    - 25 in the Americas

Anycasting (BGP)

- Many DNS services (not just root servers) are provided from networks which are BGP anycasted to the Internet.
- This works very well for UDP-based services (primary transport protocol for DNS is UDP).
- Root server operators install equipment in multiple geographies and advertise the same prefix from each location—that logical network is available in many locations.
- Clients will use best path route selection to reach the nearest location (by network topology) with that prefix.
- The benefits of good server placement include good load distribution, improved average response latency and resiliency in the event of individual site failures.

Global visibility of anycasted routes...

Non-Chinese peers preferring routes which transit Chinese ASNs for the Beijing F-, I- and J-root instances.



DNS-Operations Report (24 March 2010)

Hi there! A local ISP has told us that there's some strange behavior with at least one node in i.root-servers.net (traceroute shows mostly China) It seems that when you ask A records for facebook, youtube or twitter, you get an IP and not the referral for .com

It doesn't happen every time, but we have confirmed this on 4 different connectivity places (3 in Chile, one in California)

This problem has been reported to Autonomica/Netnod but I don't know if anyone else is seeing this issue.

This is an example of what are we seeing:

$ dig @i.root-servers.net www.facebook.com A ;
….
ANSWER SECTION: www.facebook.com. 86400 IN A 8.7.198.45

Mauricio Vergara Ereche
Santiago CHILE

Chinese Client – Tampering (19 Nov 2010)

```
; <<>> DiG 9.2.4 <<>> @192.36.148.17 www.facebook.com.
;; global options:  printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 48526
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
;www.facebook.com.              IN      A

;; ANSWER SECTION:
www.facebook.com.       4974    IN      A       46.82.174.68      ← Deutsche Telekom

;; Query time: 2 msec
;; SERVER: 192.36.148.17#53(192.36.148.17)
;; WHEN: Fri Nov 19 17:52:25 2010
;; MSG SIZE  rcvd: 66
```

Chinese Client – Good Result (19 Nov 2010)

```
; <<>> DiG 9.2.4 <<>> @192.36.148.17 www.facebook.com.
;; global options:  printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 45919
;; flags: qr rd; QUERY: 1, ANSWER: 0, AUTHORITY: 13, ADDITIONAL: 14

;; QUESTION SECTION:
;www.facebook.com.              IN      A

;; AUTHORITY SECTION:
com.                    172800  IN      NS      a.gtld-servers.net.
com.                    172800  IN      NS      i.gtld-servers.net.
  [many correct answers omitted]

;; ADDITIONAL SECTION:
a.gtld-servers.net.     172800  IN      A       192.5.6.30
a.gtld-servers.net.     172800  IN      AAAA    2001:503:a83e::2:30
  [many correct answers omitted]

;; Query time: 50 msec
;; SERVER: 192.36.148.17#53(192.36.148.17)
;; WHEN: Fri Nov 19 17:52:43 2010
;; MSG SIZE  rcvd: 506
```

Chinese Client Packet Capture – Bad Result (10 Jun 2010)

18:06:17.581240 IP SRC-IP.57520 > 192.36.148.17.53: 54947+ A? www.facebook.com. (34)

18:06:17.585669 IP 192.36.148.17.53 > SRC-IP.57520: 54947 1/0/0 A 59.24.3.173 (50) ← Bad answer

18:06:17.600736 IP 192.36.148.17.53 > SRC-IP.57520: 54947* 1/0/0 A 243.185.187.39 (66) ← Another bad answer (for good measure)

18:06:17.600778 IP SRC-IP.128 > 192.36.148.17: icmp 102: SRC-IP.128 udp port 57520 unreachable ← 2nd bad answer, ICMP go away! 192.36.148.17

- This is completely expected behavior.
  - The GFW is known to tamper with DNS packets.
  - The client is inside of China.
  - You can see the exact same behavior querying *any other* root name server from inside China.

Explanation

- The global advertisements for 192.36.148.0/24 include AS 29216 (I-root) and AS 8674 and then traversed several Chinese ASNs (in red).
- Inbound packets on this path would traverse AS 10026 (PacNet), AS 7497 (Computer Network Information Center), AS 24151 (CNNIC) before reaching AS 29216 and 8674:
  - […] 10026 7497 7497 24151 8674 29216
- Peers selecting this path would clearly be sending their queries to the Beijing node.
- The results reported by Mauricio Vergara Ereche on the dns-operations mailing list are consistent with GFW behavior.

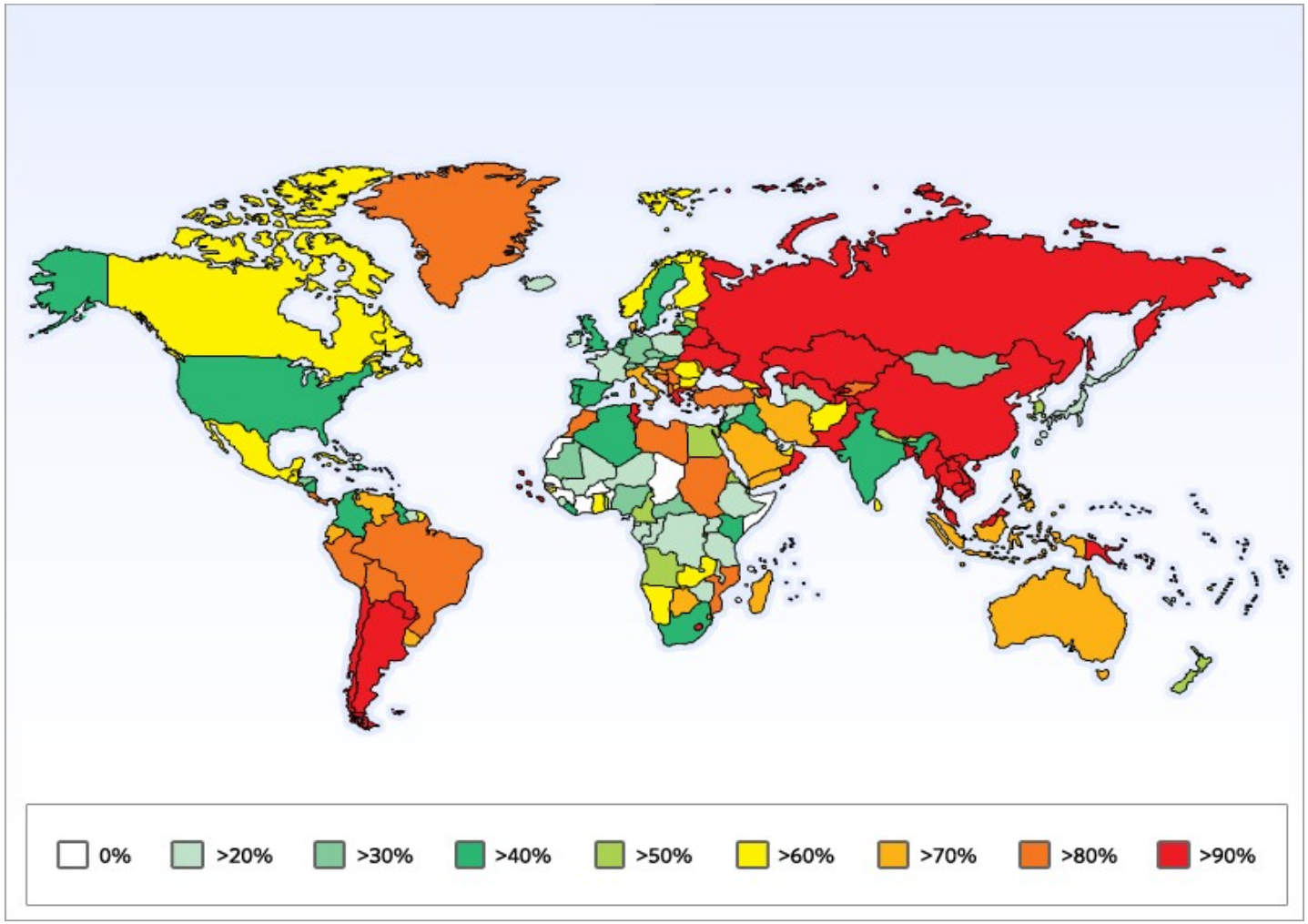
An unlikely series of events ...

- Resolver request for 'www.facebook.com' (or any other GFW-tampered name)
- The name is not already in cache
- The 'com.' zone (NS) is not cached either (48-hour TTL)
- Ask F-, I- or J-root instances
- Get directed to a root instance in Beijing
- Let the race begin!
  - Query should return the 'com.' zone servers (NS)
  - GFW may return an incorrect A record
  - If GFW wins the race, put A record in cache
  - Your DNS cache is now poisoned

National policy escapes containment

- Again, we learn that IP traffic does not respect jurisdictional boundaries (national or otherwise).
- This event demonstrates the collision of
  - a technical resiliency tool (anycast)
  - a technical policy enforcement control (GFW)
- Who was affected? Who could have been affected?
- Calculate the percentage of routed networks in each country which saw the path to any of the Beijing root servers at any point in 2010.

Who could have been affected?



Netnod serves Chinese market

- Netnod intends the Beijing node to be globally visible.
- Netnod employs TSIG and routinely checks serial numbers of the data at each of their root server instances against Verisign/IANA root zone data to ensure validity.
- The tampering of replies from the Beijing I-root was completely consistent with and almost irrefutably the GFW.
- Netnod withdrew their anycasted routes until their host (CNNIC) could secure assurances that the tampering would not recur.
- Netnod serves a large Internet user base in China and its Beijing node is one of its top 5 busiest instances.

Thanks to Kurt Erik Lindqvist, CEO Netnod, for info on I-root operations and this incident.

The Stars DNS Rating System

*****	**safe**	Providers respect the integrity of DNS and rarely (if ever) rewrite NXDomain responses.
****	**mostly safe**	Providers might use NXDomain to make money.
***	**caution**	Providers may be required to modify resolver responses to comply with local content laws.
**	**strong caution**	Routing/switching providers may be required to modify resolver responses in flight in order to comply with local laws.
*	**danger**	Routing/switching providers are required to modify all root and resolver responses in flight without warning.

Example ratings

- US would currently earn a 4 star rating
  - 5 star rating, in the years before NXDomain monetization
  - Recent proposed legislation, Combating Online Infringements and Counterfeits Act (COICA), could lower this rating to 3 or 2 stars.
- Many European nations would earn a 4 or 3 star rating
  - Increasingly strict anti-pornography laws are being enforced (often voluntarily) with blacklists distributed and supported in ISP resolvers.
- The Middle East and Australia would earn a 3 star rating.
- China is the only country known to have earned a 1 star rating.

Open questions

- In an effort to protect against the Four Horsemen of the Infocalypse, will governments continue to encourage technical control via DNS? Probably.
- Could DNS-based technical controls be limited to resolvers, preventing tampering with answers from authoritative servers?
- What confusion ensues if tampered results are employed for further technical controls?
- What new technical controls will follow the wider deployment of DNSSEC, which was built to withstand in flight tampering and ensure chain of authority?
- What is the next leaky technical control that will affect users far from the target jurisdiction?

Conclusions

- **Enable DNSSEC.**
- **Don't pass your queries across the GFW (if you can help it).**
- **If your government requires DNS-based technical controls, install them at the resolver.**

Thank You

Martin A. Brown **mabrown@renesys.com**
Doug Madory **dmadory@renesys.com**
Alin Popescu **alin@renesys.com**
Earl Zmijewski **earl@renesys.com**