DNS 101

John Kristoff jtk@cymru.com

Agenda

- DNS Basics
- Operational BCPs and FAQs
- Tools and Miscellaneous Topics

One of two critical systems

Routing (BGP) and naming (DNS) are by far the two most critical subsystems of the Internet infrastructure. And in the case of DNS, practically all Internet hosts participate directly in the DNS as a client, server or both. As a result, DNS is one of the most unencumbered protocols in use throughout the Internet. This can be good, bad or interesting depending on your perspective.

Subsystem control is power

- I can forward data, or not
- I can inspect data, and record it
- I can share audit data, or sell it
- I can study data, and build new products
- I can redirect data, or become the endpoint
- I can give access to friends, and your adversaries
- I could just pass bits

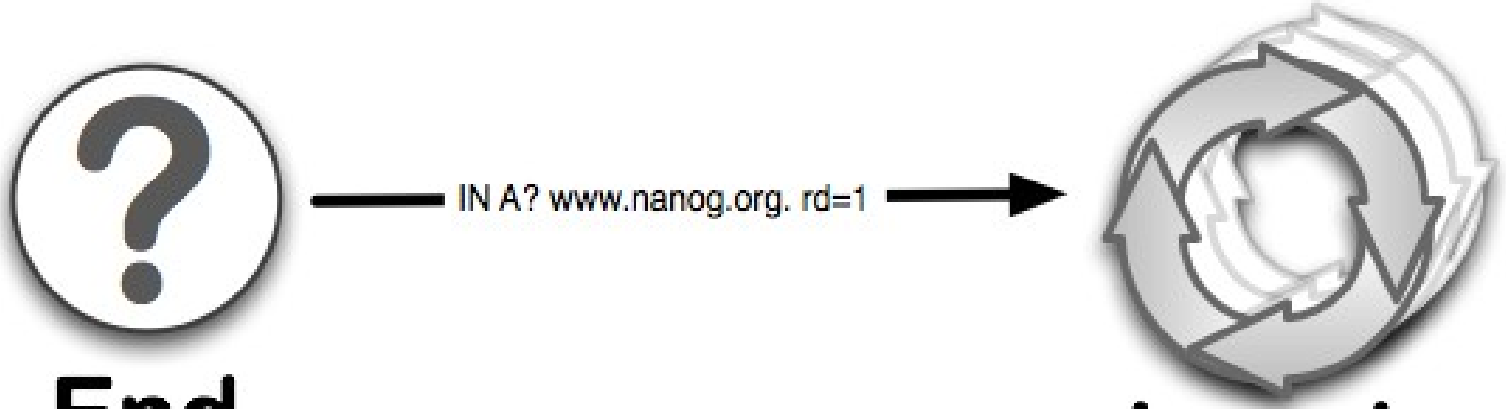
Flexibility as a boon and scourge

- DNS is largely invisible to users, it just works
- Or rather, in reality, “it works enough”
- DNS withstands a lot of the slop we put into it
- Nonetheless...
  - Poor performance is noticeable if not attributed
  - There are threats to availability and correctness
  - Interest and attention to DNS continues to grow

A DNS resolution primer...

What is the IPv4 address for www.nanog.org?

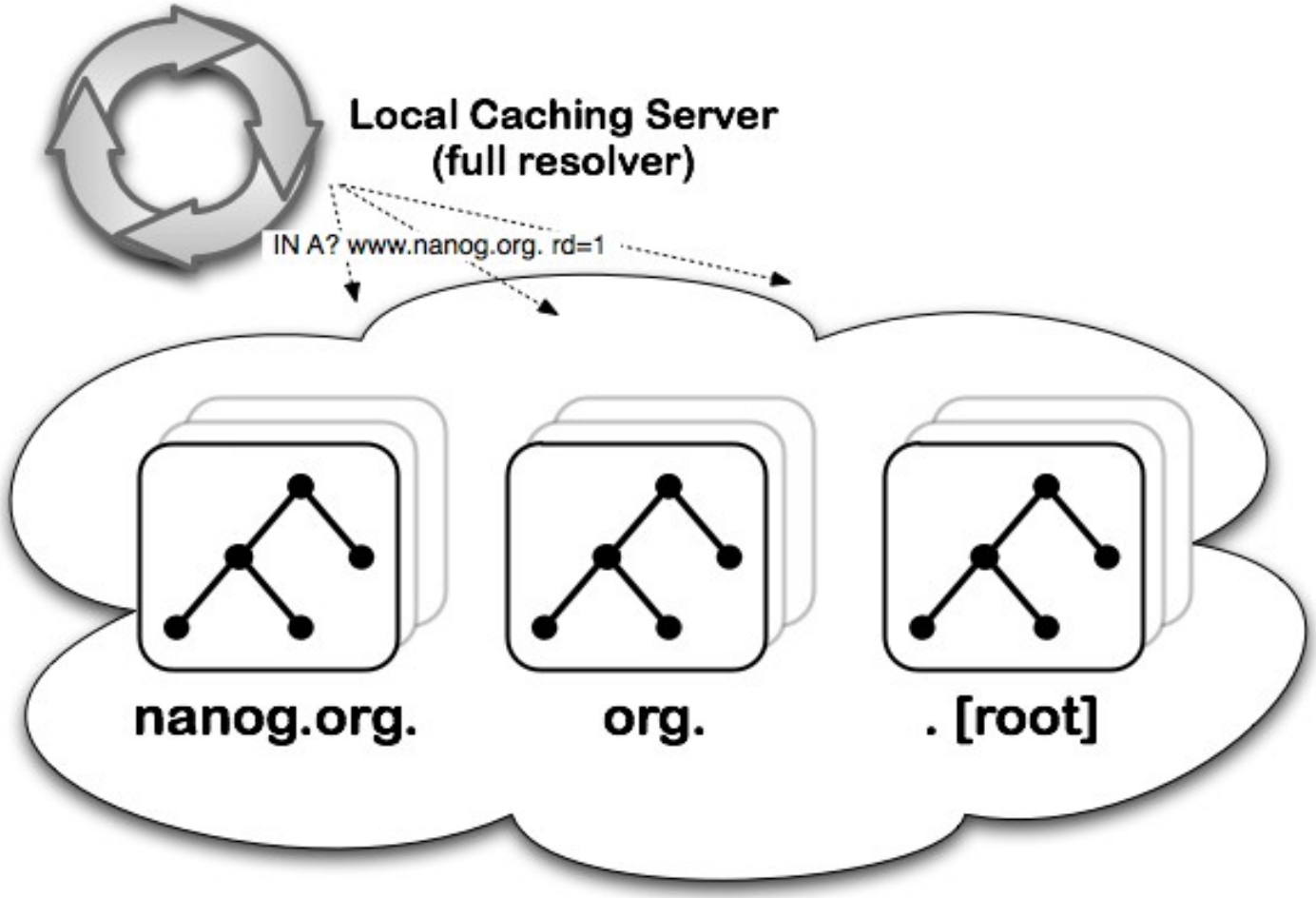
Do all the work for me (recursion desired).



End User
(stub resolver)

Local Caching Server
(full resolver)

1. Check cache, supply answer if available, or
2. Follow delegation from most specific cached parent, or
3. Start at root if cache is empty.



Let's assume cache is empty, and all it knows about is [.] root.*

A.root-servers.net.

...

M.root-servers.net.

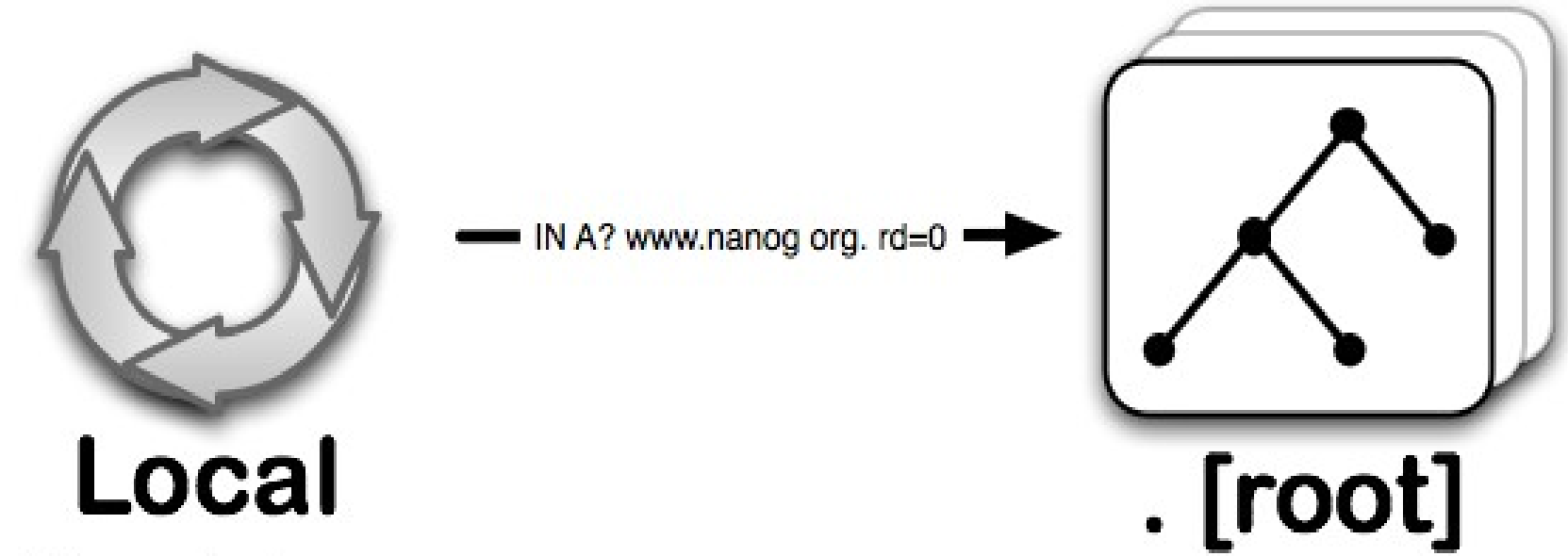
***Do you see why a reliable and trustworthy root is so important?**

What are the name servers for .org?

NO!!

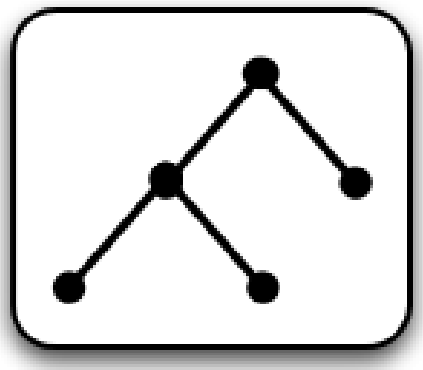
What is the IPv4 address for www.nanog.org?

I'll do the work myself (recursion disabled).



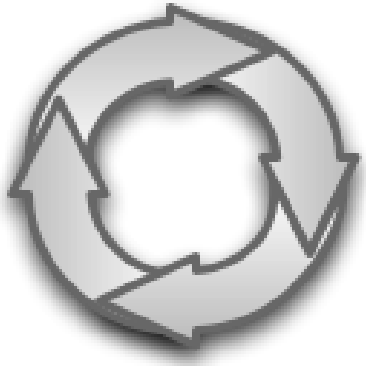
Dunno.
I refer you to .org NS RRset:

a0.org.afilias-nst.info.
a2.org.afilias-nst.info.
b0.org.afilias-nst.org.
b2.org.afilias-nst.org.
c0.org.afilias-nst.info.
d0.org.afilias-nst.org.



. [root]

NOERROR



Local Caching Server

Does the local caching server have something in its cache now?

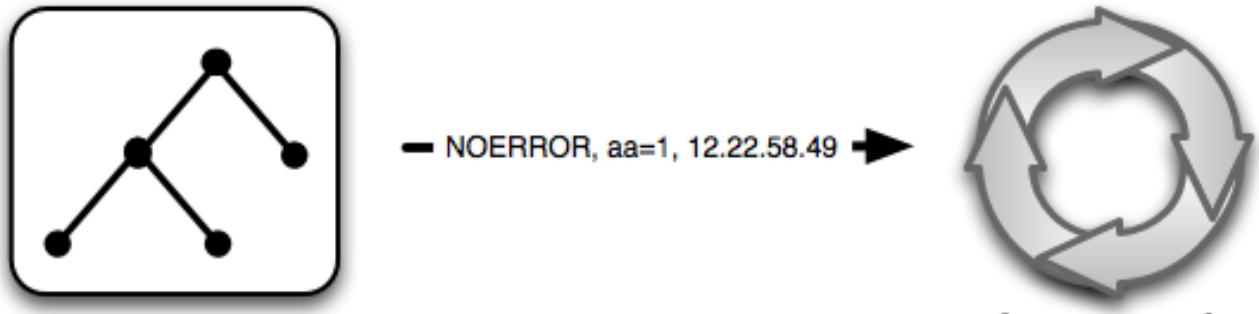
Raise your hand for yes.

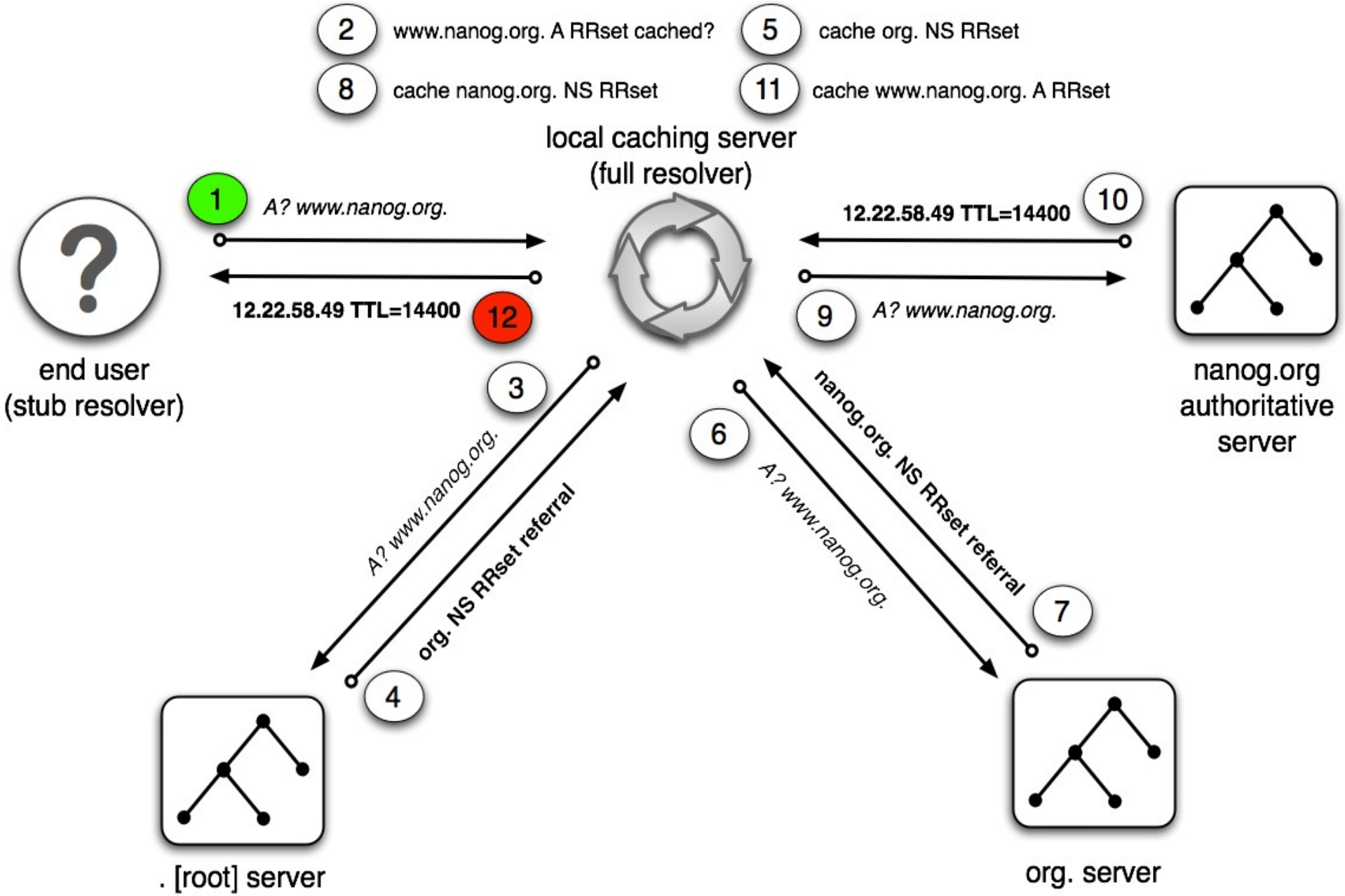
Ultimately we should get here...

You've come to the right place. The answer RRset is:

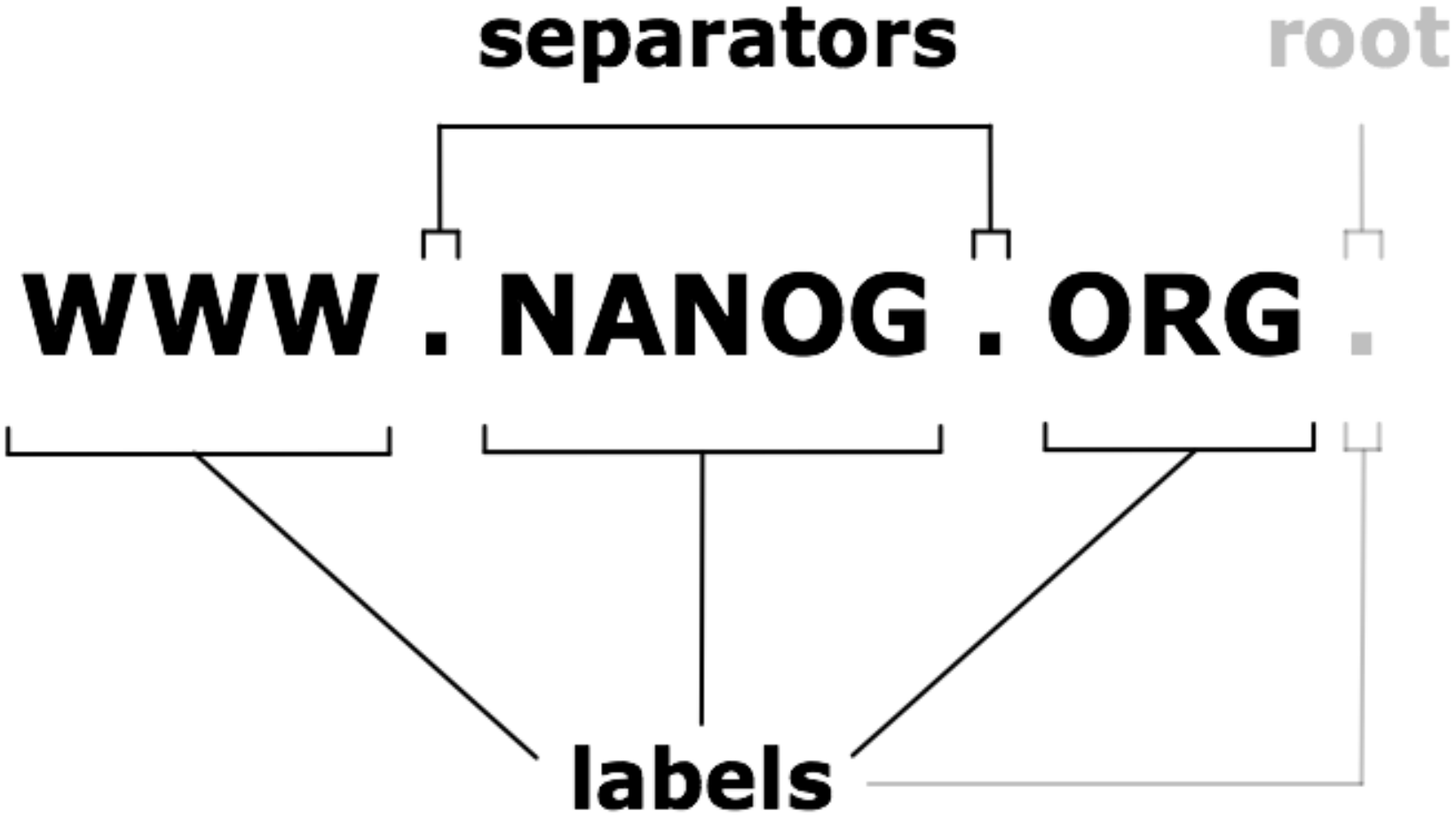
12.22.58.49 with TTL=14400

I am authoritative (aa bit is set).




2
www.nanog.org. A RRset cached?
5
cache org. NS RRset
8
cache nanog.org. NS RRset
11
cache www.nanog.org. A RRset
local caching server
(full resolver)
1
A? www.nanog.org.
12.22.58.49 TTL=14400
12
10
12.22.58.49 TTL=14400
9
A? www.nanog.org.
end user
(stub resolver)
nanog.org
authoritative
server
3
A? www.nanog.org.
org. NS RRset referral
4
6
A? www.nanog.org.
nanog.org. NS RRset referral
7
. [root] server
org. server

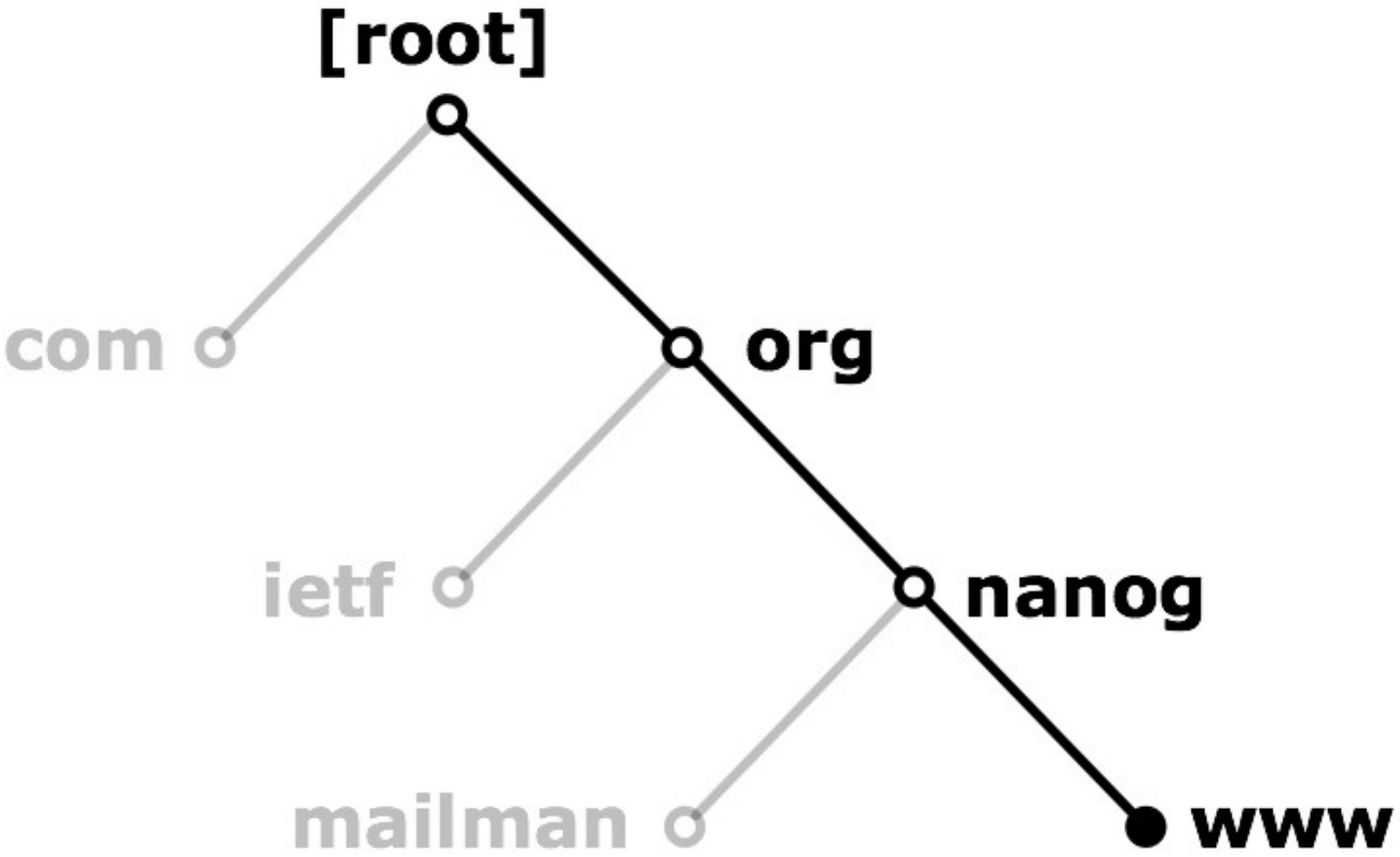
Anatomy of a domain name



What's in a name?

- As a domain name, any 8-bit value is valid
- For a host name, see IETF RFC 1123
  - [0-9a-zA-Z-]
  - underscore not strictly allowed, but often used
- On-wire max domain name length is 255 octets
  - max label length is 63 octets
- Some second-level domains behave like TLDs
  - e.g. co.uk.
  - related: http://publicsuffix.org/

Name space hierarchy



Distribution and delegation

There is no single all-encompassing DNS database server. Zone administration is delegated and zone data is distributed. This implies the desire and need for a single, authoritative, trustworthy and reliable root.

Root zone

- ICANN
  - US DoC contractor for IANA services
  - responsible for root zone contents
- VeriSign
  - data “mechanic”
- root-servers.org
  - 12 independent root server operators
  - 13 instances total, VeriSign runs two

Top-level domains (TLDs)

- All the first-level child labels of the root
- Various types (“marketing” terms)
  - gTLD, ccTLD, sTLD, uTLD and special TLDs
- Started with:
  - .arpa .com .edu .gov .int .mil .net .org
- Now approximately 300 (mostly ccTLDs), also see:
  - http://www.iana.org/domains/root/db/
  - https://www.dns-oarc.net/oarc/data/zfr/root

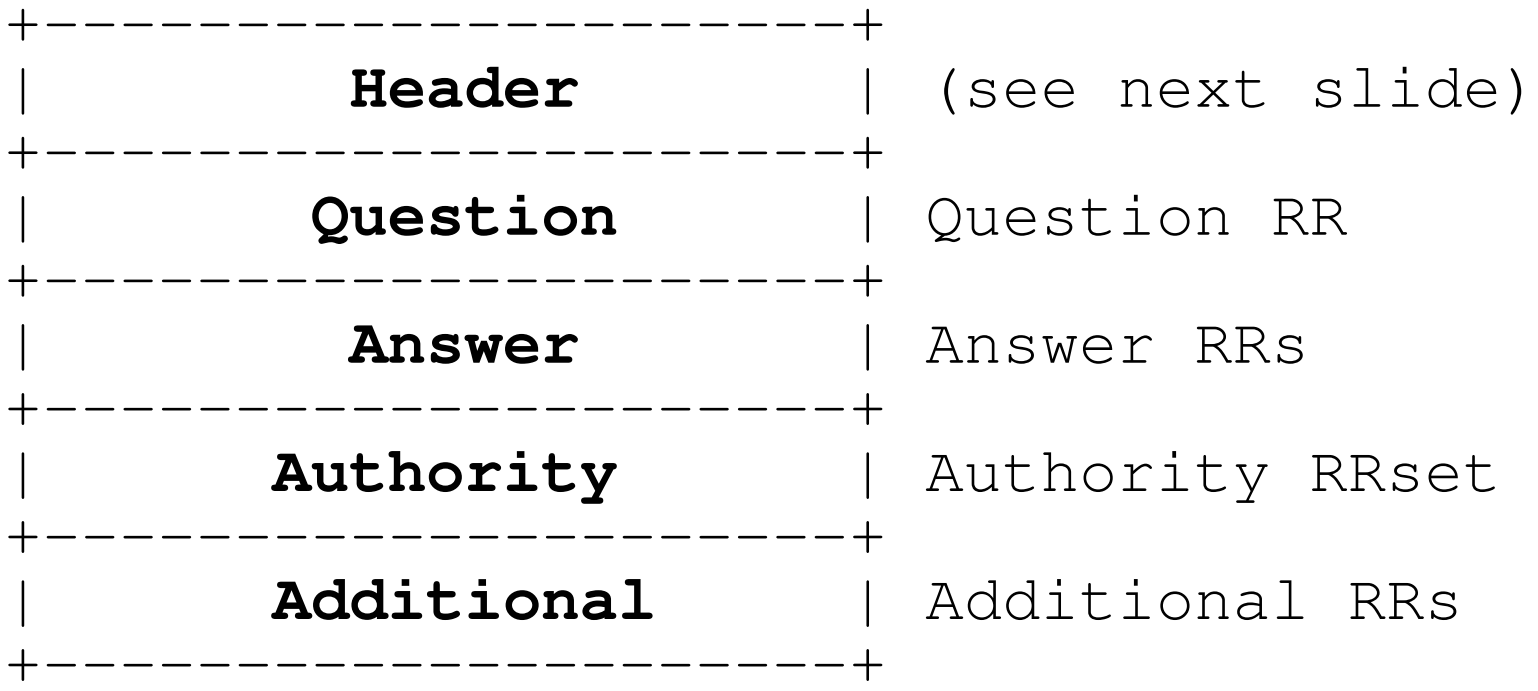
Domain name registration

- Registry
  - Keeper/maintainer of TLD zone data
- Registrar
  - Agent through which registrant obtains a name
- Registrant
  - Authorized user of name, customer of registrar

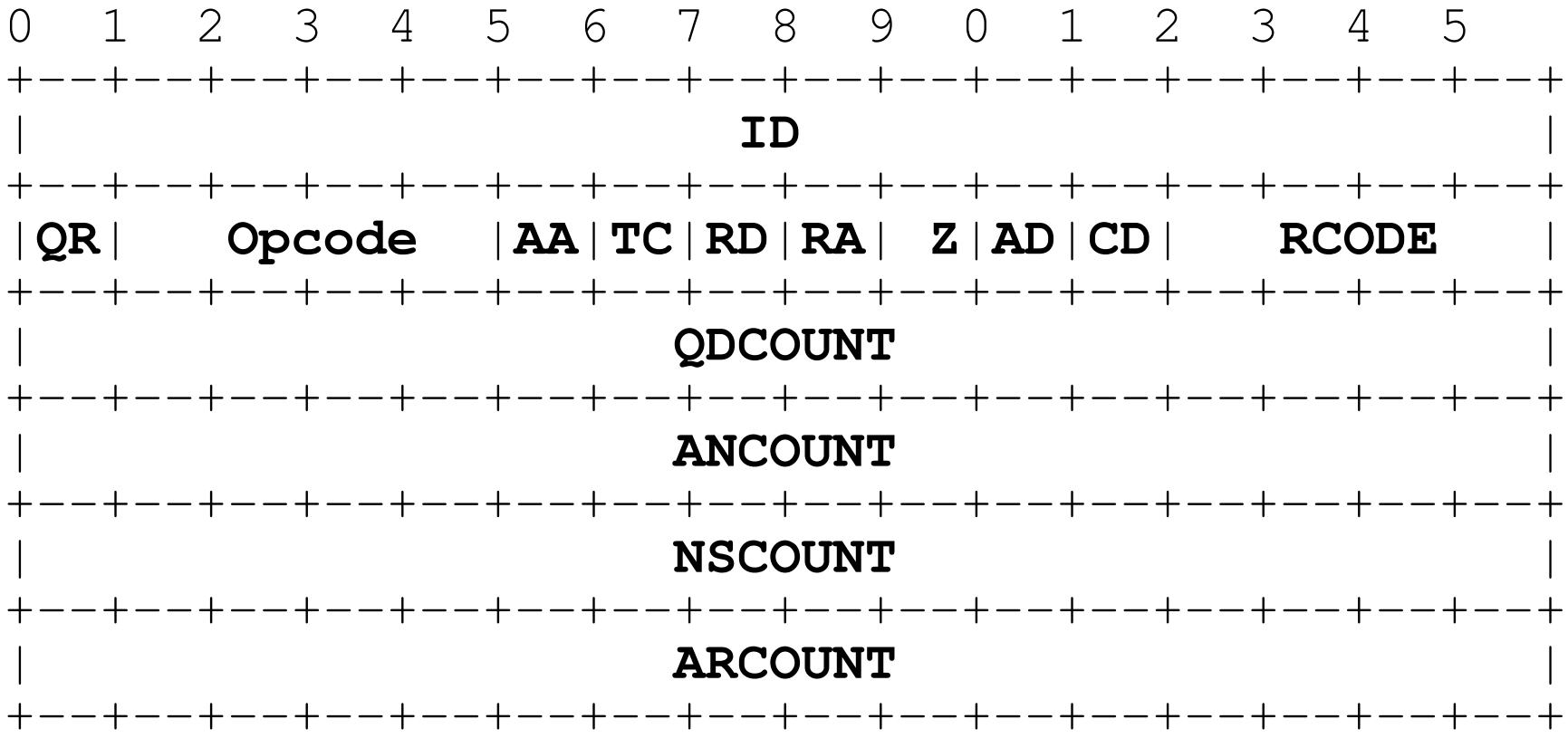
WHOIS

- Interface to assignees of Internet resources
  - e.g. domain names, IP addresses, ASNs
- Human readable text output
- Lacks modern design attributes
  - e.g. security, internationalization

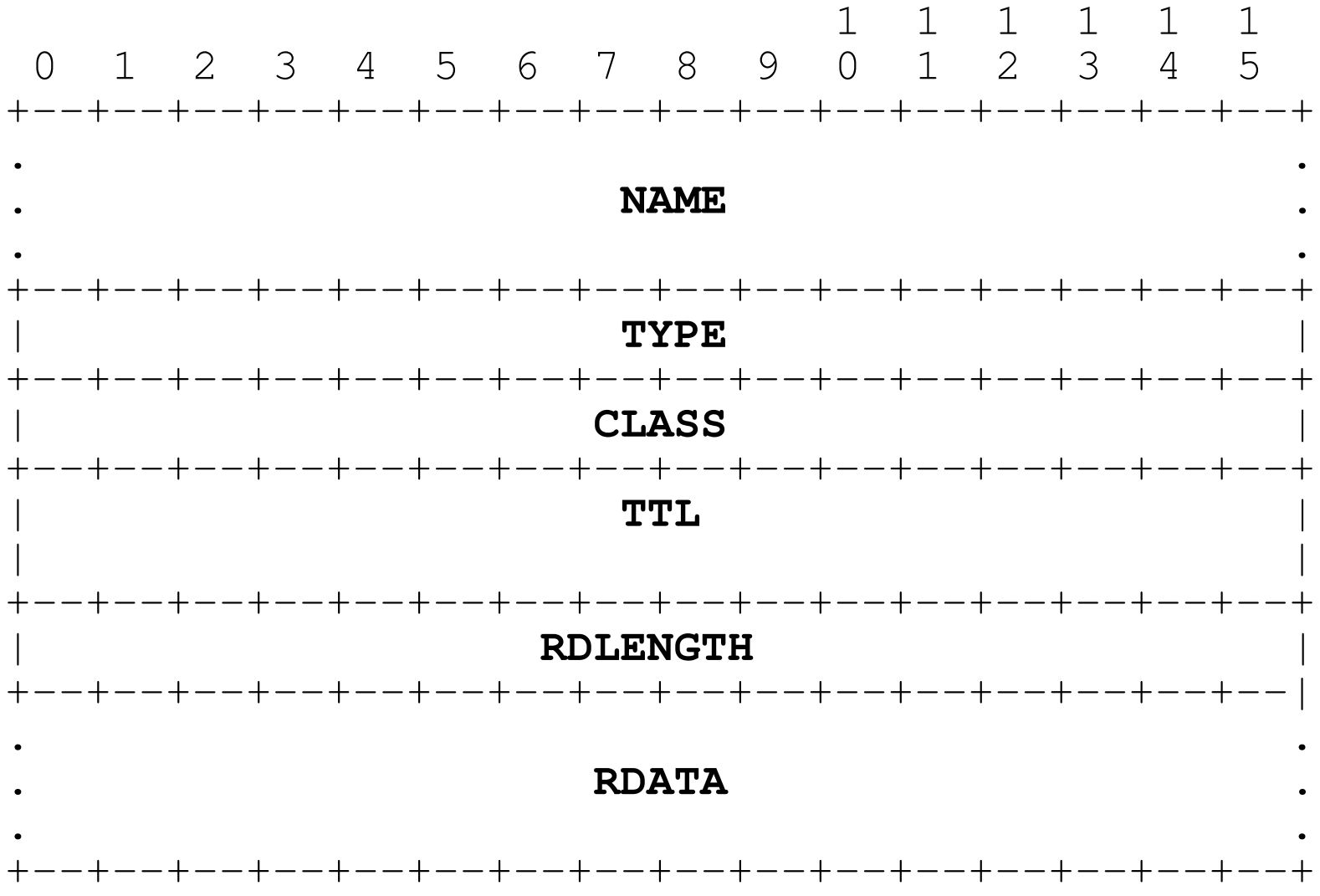
DNS protocol message format



DNS protocol header format



DNS protocol RR format



DNS transport

- DNS uses both UDP and TCP
- Well known port 53 reserved for server listener
- In practice, most queries/answers use UDP
- TCP is NOT just for zone transfers
  - DDoS mitigation hack
  - large RRsets (e.g. DNSSEC, TXT RRs)
  - RFC 5966, 2010-08, DNS Transport over TCP
    - “[...] TCP is henceforth a REQUIRED part of a full DNS protocol implementation.”

EDNS0

- Extension mechanism for DNS
- One OPT pseudo-RR added to additional section
- Example extension capabilities include:
  - signaling support for DNSSEC (DO bit)
  - indicating sender's max UDP payload size
  - including client query origin detail (draft)

DNSSEC

- Adds origin authenticity protection
- No encryption of DNS data
- What does this do?
  - Optimist: resists poison / replay / MITM attacks
  - Cynic: awkward mechanism for a non-problem
- Two of the original 3-bit Z field bits now defined:
  - AD – authentic data
  - CD – checking disabled

DNSCurve

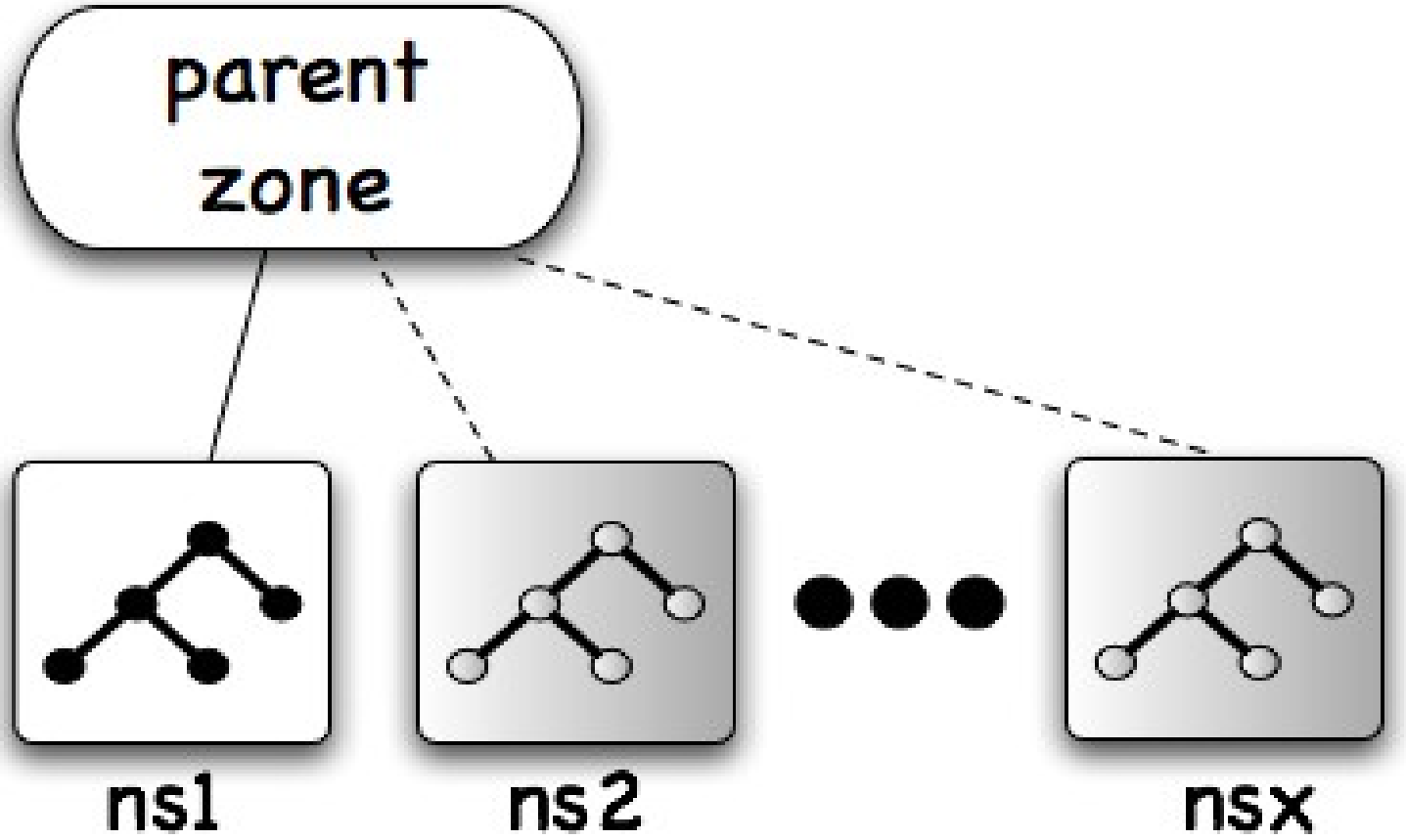
- Adds confidentiality to DNS messages
- What does this do?
  - Optimist: resists packet-level attacks
  - Cynic: insufficient end-to-end data protection
- Minimal changes to underlying DNS specifications

DNSSEC + DNSCurve or ?

- WARNING: here is where net theology comes in
- Bernstein's many criticisms of DNSSEC are valid
- There are threats and attacks, what matters?
- Can we have e2e trustworthiness and consistency?
  - ...or is it just users versus the network admins?
- Is DANE the right direction?
- How has this informed the BGPSEC / RPKI work?
- Passive monitoring with DNSSEC vs. DNSCurve

BCPs and FAQs

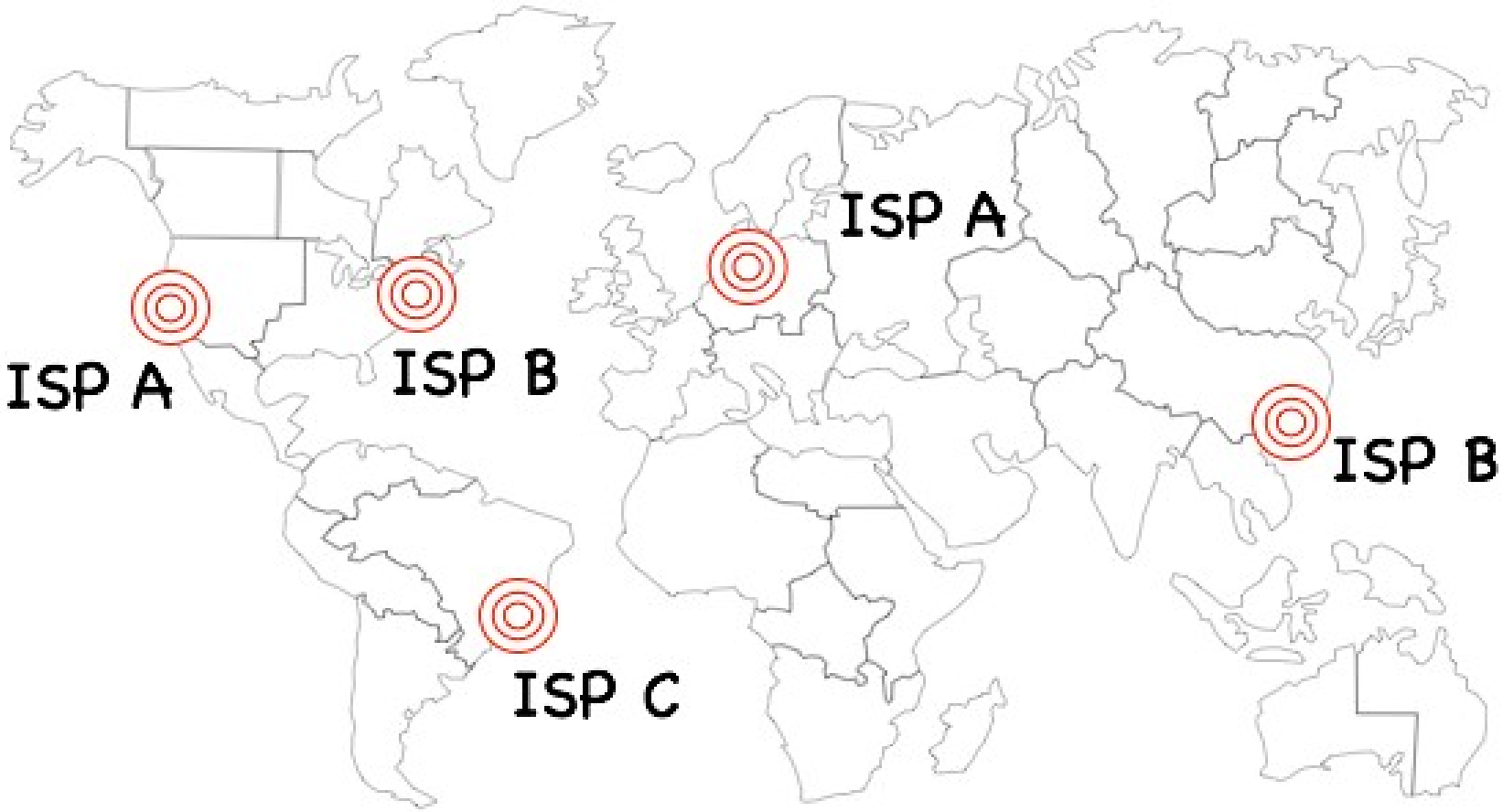
How many NS RRs for your zone?



Authoritative name server RRset

- Two is the de facto minimum
- Depending on design, more may be better
- Anycast service may be worth your consideration
- Some people use hardware-based load balancing
- Miscreants invented fast flux
  - Then legitimate providers said, “Hmm...”

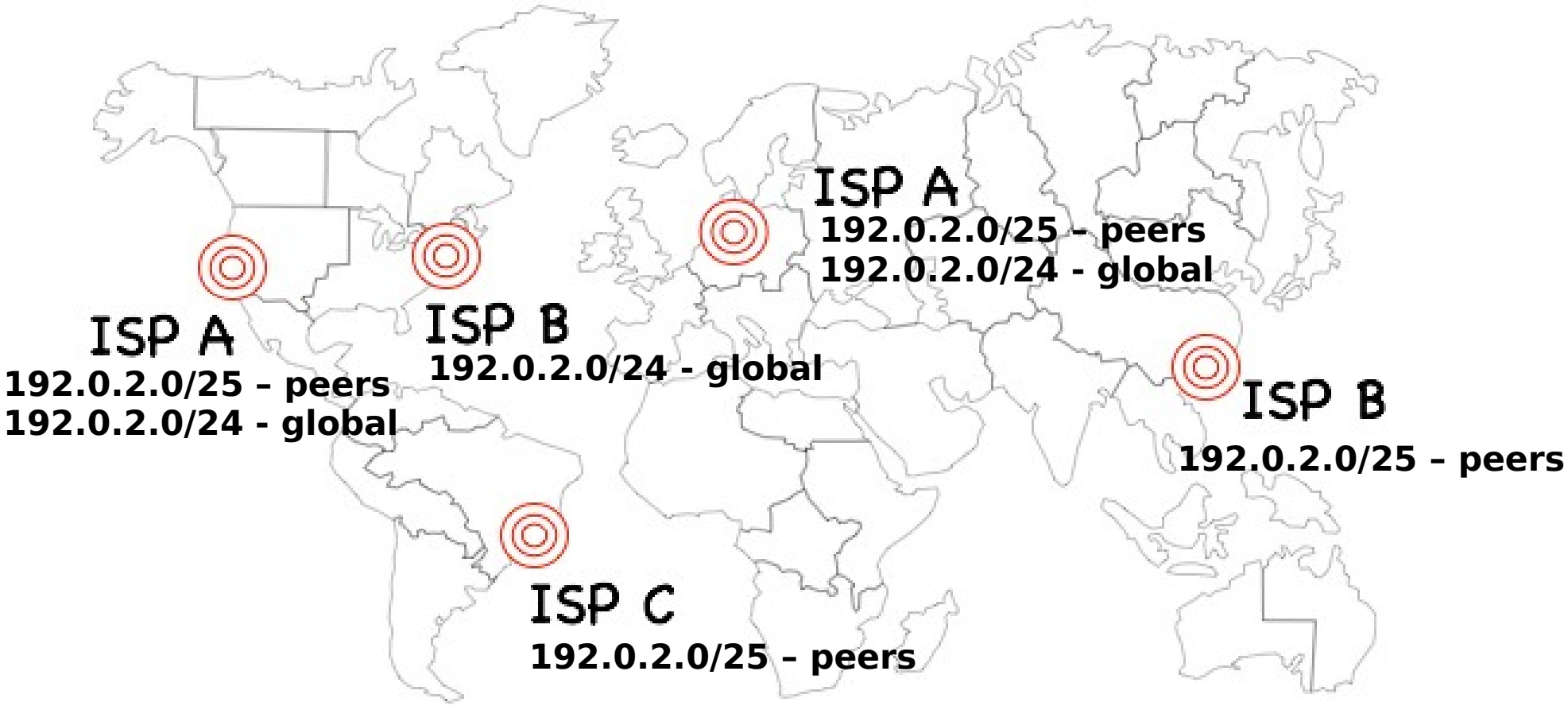
Where are your name servers?



DNS Server Diversity

- Consider physical and topological proximity
- All servers in the same building is suboptimal
  - As are all servers behind a shared upstream link
- Shorter prefixes mitigate route hijacks
- Diverse routing paths can improve resiliency
- Diverse origin AS for routes not strictly necessary
  - Just ask the DNS anycast service providers

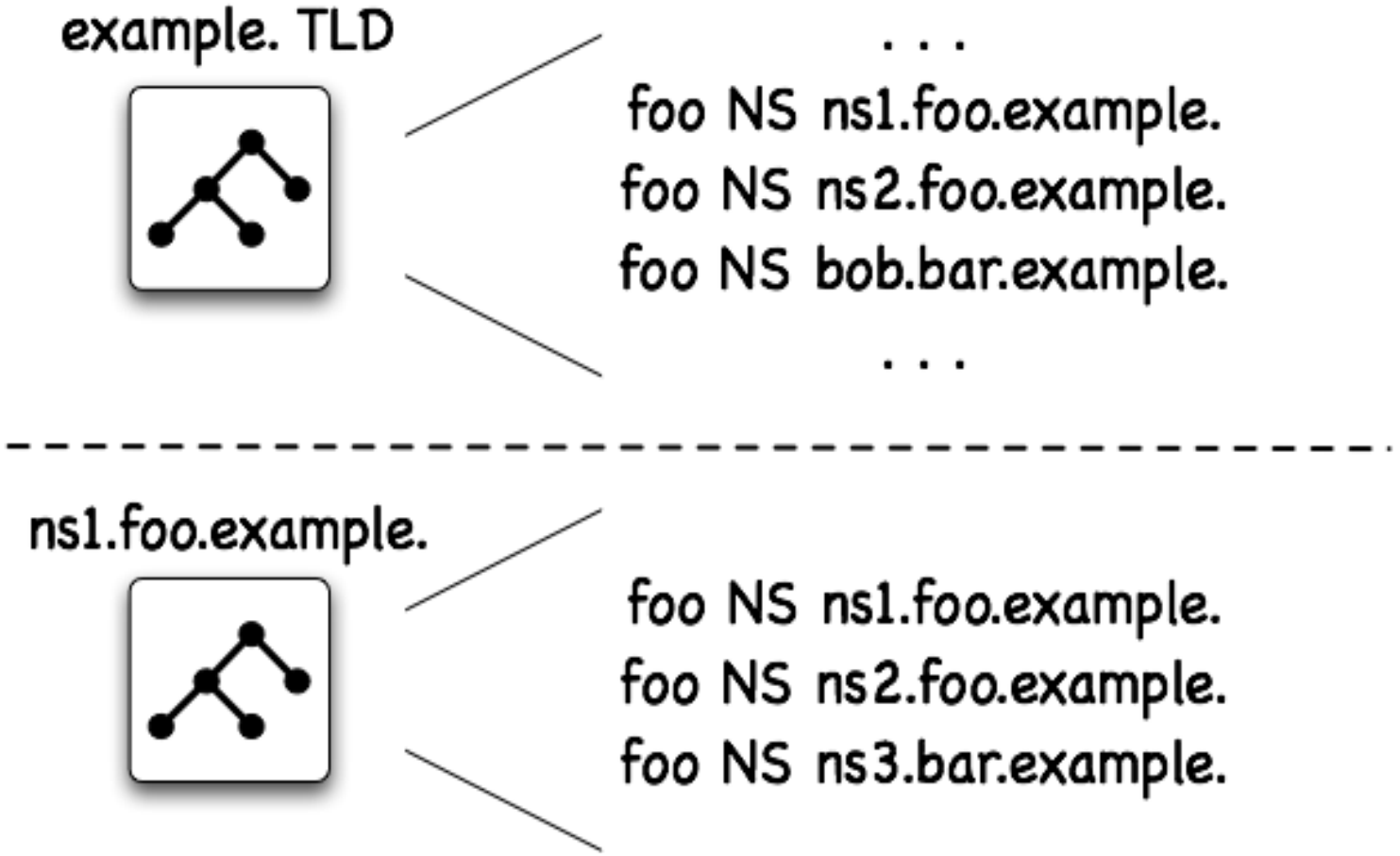
Shared unicast addressing



Deployment

- For both recursive and authoritative servers
- Widely implemented technique to spread the load
- Helps mitigate DDoS attacks
- Helps provide low latency service around the globe
- See IETF RFC 4786 for technical background
- See ISC-TN-2004-1 for implemenation notes

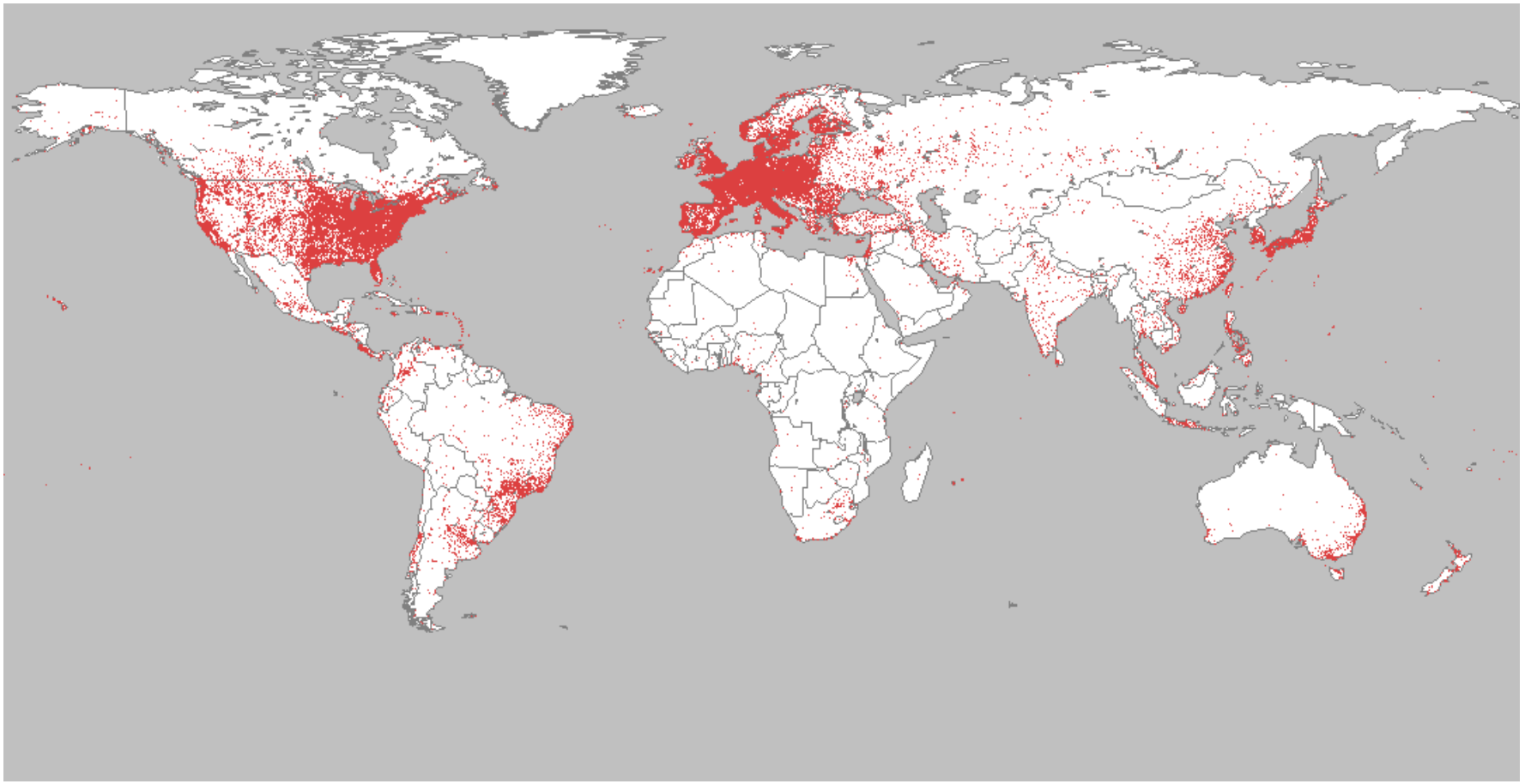
Are parent and children consistent?



Delegation Consistency

- Things may work if inconsistent, but sub-optimally
  - You're not getting full resiliency at best
  - Delays, timeouts and errors may be occurring
  - Domain name hijacks possible at worst
- Recent measurement showed:
  - 18% of domains in edu. have lame delegations
  - Only 0.1% were REN-ISAC institutions
  - Or less than 5% of all REN-ISAC institutions

Does your server answer anything from anyone?

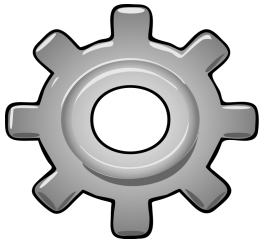


Open Resolvers

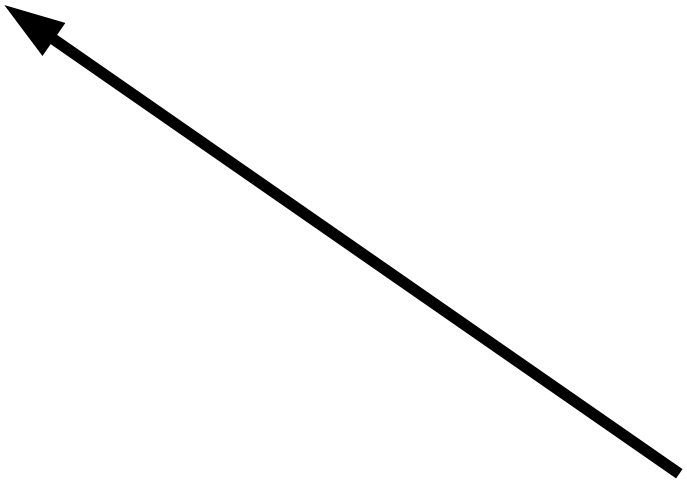
- Rarely necessary
- May be used for DDoS reflection and amplification
- Can facilitate cache poisoning attacks
- Can facilitate cache leaks
- Also see RFC 5358
- We'll tell you about open resolvers on your net:

  http://www.team-cymru.org/Services/Resolvers/

How easily can returning answers be spoofed?



What is the rdata/ttl for … ?

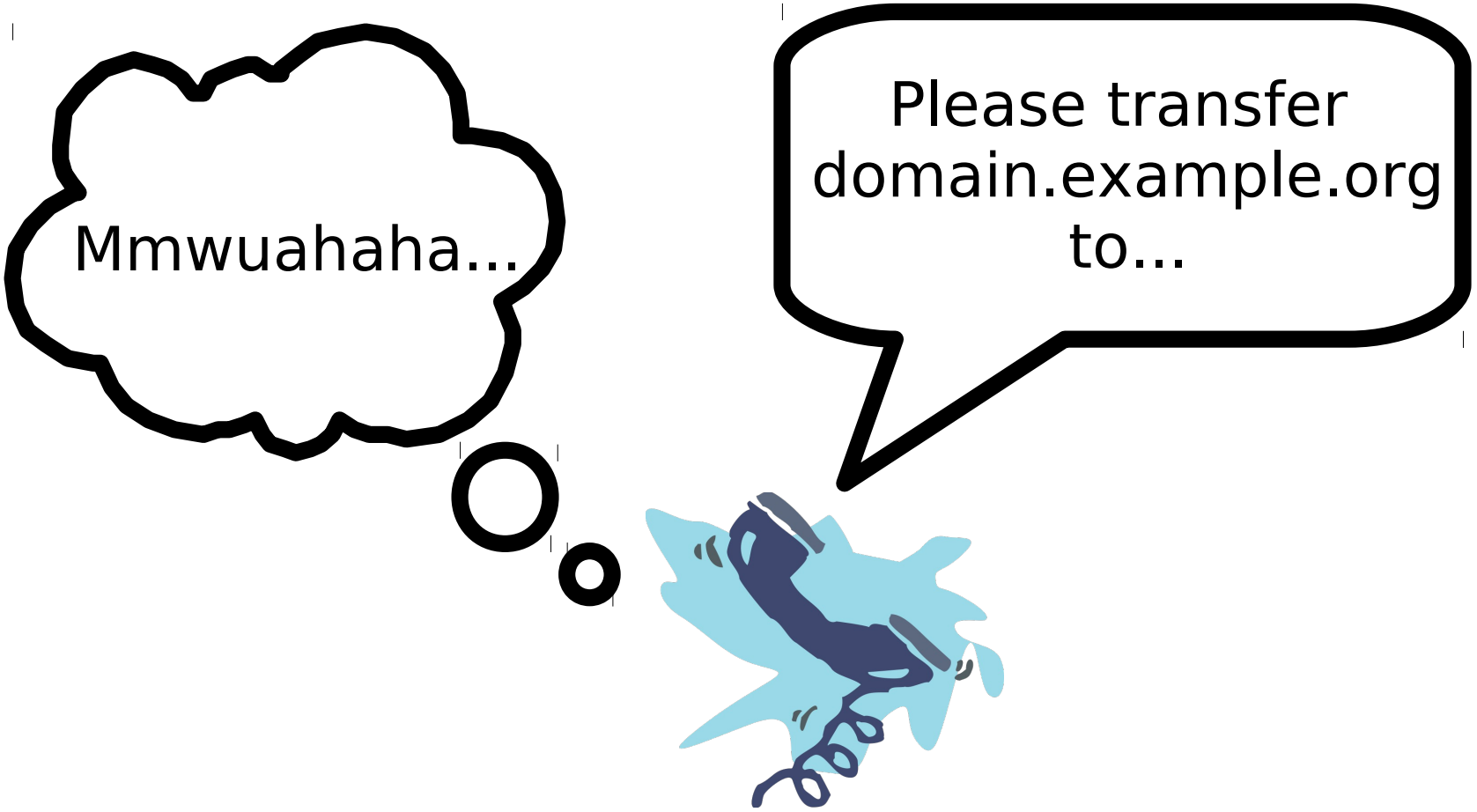


HERE IT IS!! Mmwuahaha…

Answer Spoofing Protection

- Implementations need to consider IETF RFC 5452
- Limit recursion (see the open resolvers slide)
- Ideally anti-spoofing is widely deployed
  - See IETF BCP 38 and IETF BCP 84

Is your name registration secure?



How long should my TTLs be?

- Recent advocates for long infrastructure TTLs
  - Using Long TTLs to Survive DNS Attacks, Duane Wessels, March 21, 2012 DNS-OARC
  - draft-pappas-dnsop-long-ttl (expired)
- May not be best for all in every circumstance
- Trade-offs:
  - availability, flexibility, traffic, hijack threat

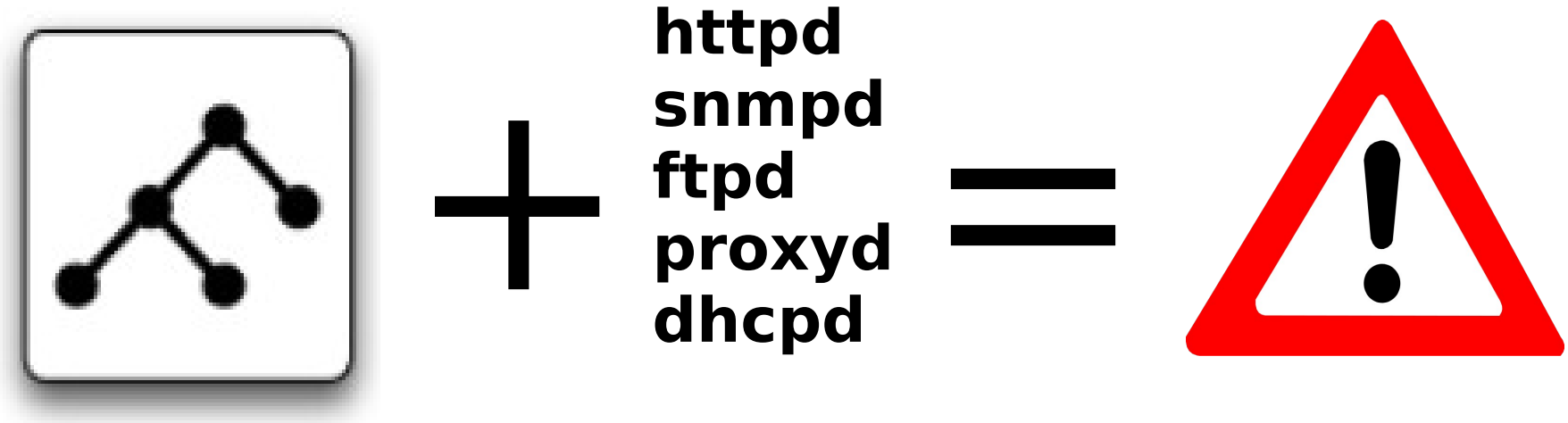
Separate aa and ra service?

- Many operators run aa+ra in a server
- Advantages to dual purpose server:
  - Apparent simplicity and cost
- Lack of separation problems:
  - Access control issues
  - Fetching remote answers can be expensive
  - Private zone and answer leaks
  - Corner case ambiguity

Domain Name Registration

- Do not let your name(s) expire needlessly
- Safeguard registrar accounts and passwords
- Some registrars offer additional safeguards
  - Ask about them, know what is available
- Make this part of a disaster recovery plan

What is on your name server?

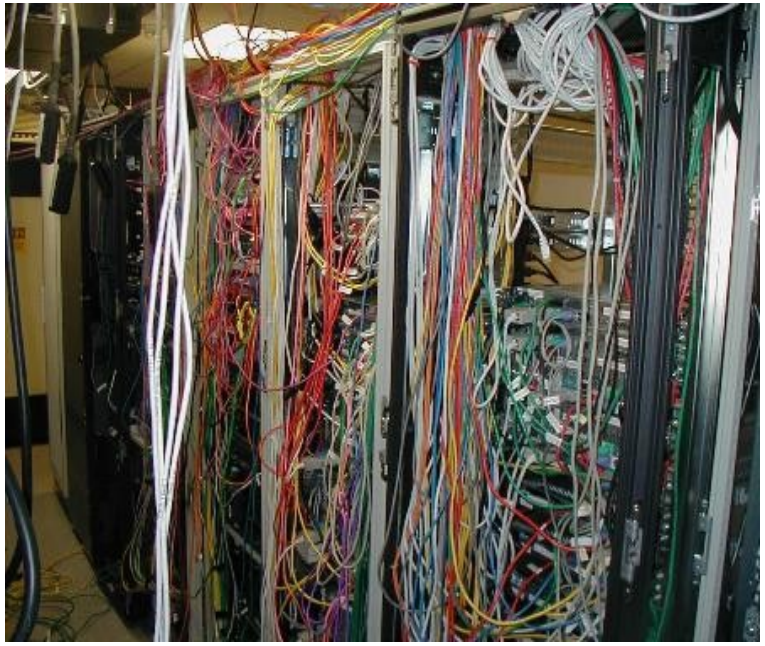


Co-mingling Services

- SSH and NTP are reasonable standard services
  - Most others are not
  - Even these should generally be inaccessible
- Consider isolating some zones from others
  - e.g. put DDoS risk zones on a separate platform
- Consider separating recursive/authoritative service

How are servers administered?

pictures from techrepublic, Bill Detwiler



OR

Administrative Processes

- We see a lot of successful SSH brute force attacks
- Limit physical access to facilities and hardware
- If it looks lousy, it probably is
- When in doubt, consult Occam's Razor
- Use revision control for configs and zone files
- As important as a backup plan is the restore plan
- Secure BIND Template

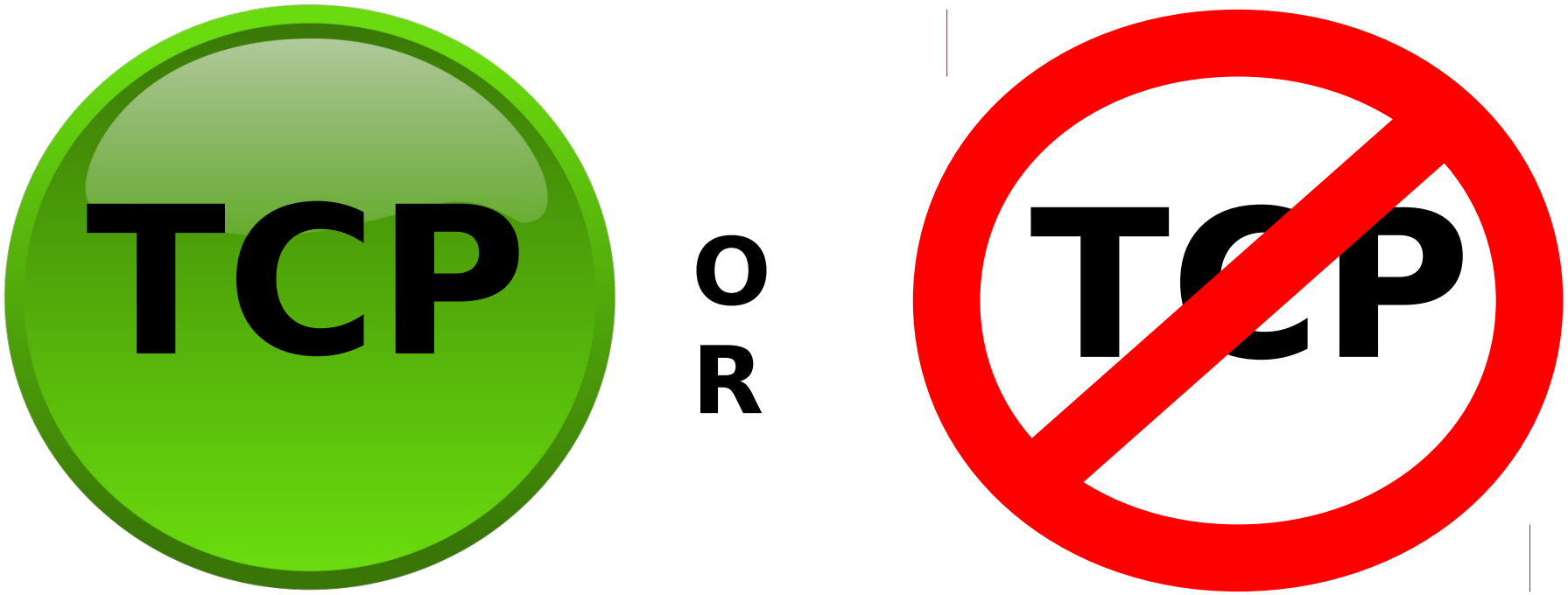
  http://www.team-cymru.org/ReadingRoom/Templates/

How much RAM, CPU, disk and network capacity is available?

Physical Resources

- Don't have enough, have way more than enough
- Resolvers can demand lots of RAM
- CPU may be important, especially for crypto
- Hard drives usually less important
  - Isolating partitions and directories may be useful
  - Try to offload data collection to another system
- Network capacity usually not an issue until DDoS

Are you filtering DNS over TCP?



TCP

- Don't assume you have no DNS over TCP
- TCP isn't just for zone transfers
  - Large DNS messages may use TCP
  - Some operators may force TCP during DDoS
- TCP tuning may be required for some DoS threats
- IETF RFC 5966 implementation requirement

What queries do you see/make?

Monitoring and Auditing

- Troubleshooting with query insight is very helpful
- Consider learning answers from the resolvers too
  - AKA passive DNS
- Minimally, trend DNS query/answer statistics
- Monitor servers, answers and routes from outside

  http://www.team-cymru.org/Monitoring/DNS/

  http://www.team-cymru.org/Monitoring/BGP/

Are name server clocks accurate?

Time Synchronization

- This probably means running NTP properly
- Troubleshooting works best with good timestamps
- Collected data is practically useless if time is off
- Some protocols require coordinated time
  - e.g. TSIG
- Consider setting clocks to UTC
  - Helpful for correlation across timezones

Have you read IETF RFC 2870?

Network Working Group
Request for Comments: 2870
Obsoletes: 2010
BCP: 40
Category: Best Current Practice

R. Bush
Verio
D. Karrenberg
RIPE NCC
M. Kosters
Network Solutions
R. Plzak
SAIC
June 2000

Root Name Server Operational Requirements

IETF RFC 2870

- Its a BCP, you should be familiar with it
- Its a bit dated and written for a specific audience
  - But it contains sound advice for most everyone
- A newer, generalized version may soon appear

Tools and Miscellanea

DNS troubleshooting with dig

- Query specific server, no recursion:

```
dig @nameserver query.example.net AAAA +norecurse
```

- Follow the delegation path:

```
dig query.example.net +trace
```

- Query using a specific saddr and sport:

```
sudo dig -b 0.0.0.0:53 @a.root-servers.net ns .
```

- Issue PTR query:

```
dig -x 192.0.2.1
```

Botnets and DNS Backstory

- 2004, meteoric rise in IRC botnets using DNS
  - widespread DNS insight/research efforts begin
  - “bad” names monitored and sinkholed
  - need way to uncover “bad” names
- Florian Weimer publishes Passive DNS Replication
  - basic idea: collect answers, learn namespace
  - immediately widely adopted and leveraged
- See: http://www.enyo.de/fw/software/dnslogger/

Before passive DNS

- Look for NetFlow involving “known bad” IP address
  - Look for related NetFlow records
- IP address changes, want to know DNS name
  - dnswatch Perl script
  - DNS recursive query correlation (query logging)

After passive DNS

- Quickly associate all addresses to names
  - and vice versa
- Find an IRC bot talking to 192.0.2.1?
  - check passive DNS...
  - botnet.example.org mapped to it YYYY-MM-DD
  - miscreant.example.net mapped there yesterday
  - miscreant.example.net now points to 192.0.2.2
  - 192.0.2.2 also maps to malware.example.org
  - and so on

Other passive DNS uses

- Cache poisoning detection
- Auditing and usage violation monitoring
- System and network profiling
- DNS hijacking analysis
- Other basic research

BIND Administration Options

Useful BIND named.conf options

- To enable query logging:
  - `logging { category queries { channel; }; };`
- To isolate and delegate changes with include:
  - `zone “a.example” { include “/etc/a.example”; };`
  - `acl “bogons” { include “/etc/bogons.named”;`

Named pipe for query logging

- Option for disk/log constrained environments
- Really only useful for real-time monitoring
  - `mknod /log/named.pipe`
  - `logging channel “pipe” { file “/etc/named.pipe”; }; };`
  - `tail /etc/named.pipe`
  - `grep 192.0.2.1 /etc/named.pipe`

Domain Name Hijacking

- Some names you may not want to resolve properly
  - e.g. malicious domain names
- You can set your resolvers to be authoritative for anything
- Reponse Policy Zones (RPZ) being put in BIND

1) Create mitigating zone file

```
$TTL 1D
@       IN  SOA  localhost.  Root (
                                    1970010100
                                    3H
                                    30M
                                    1W
                                    1D
)
        IN NS    localhost.
        IN A     127.0.0.1
        IN AAAA  ::1
        IN TXT   "Inquiries to security@localhost."
```

2) Add zone to named.conf

```
zone “malicious.example.org.” {
    type master;
    file “/etc/badnames.conf”;
};
```

3) Load the new zone

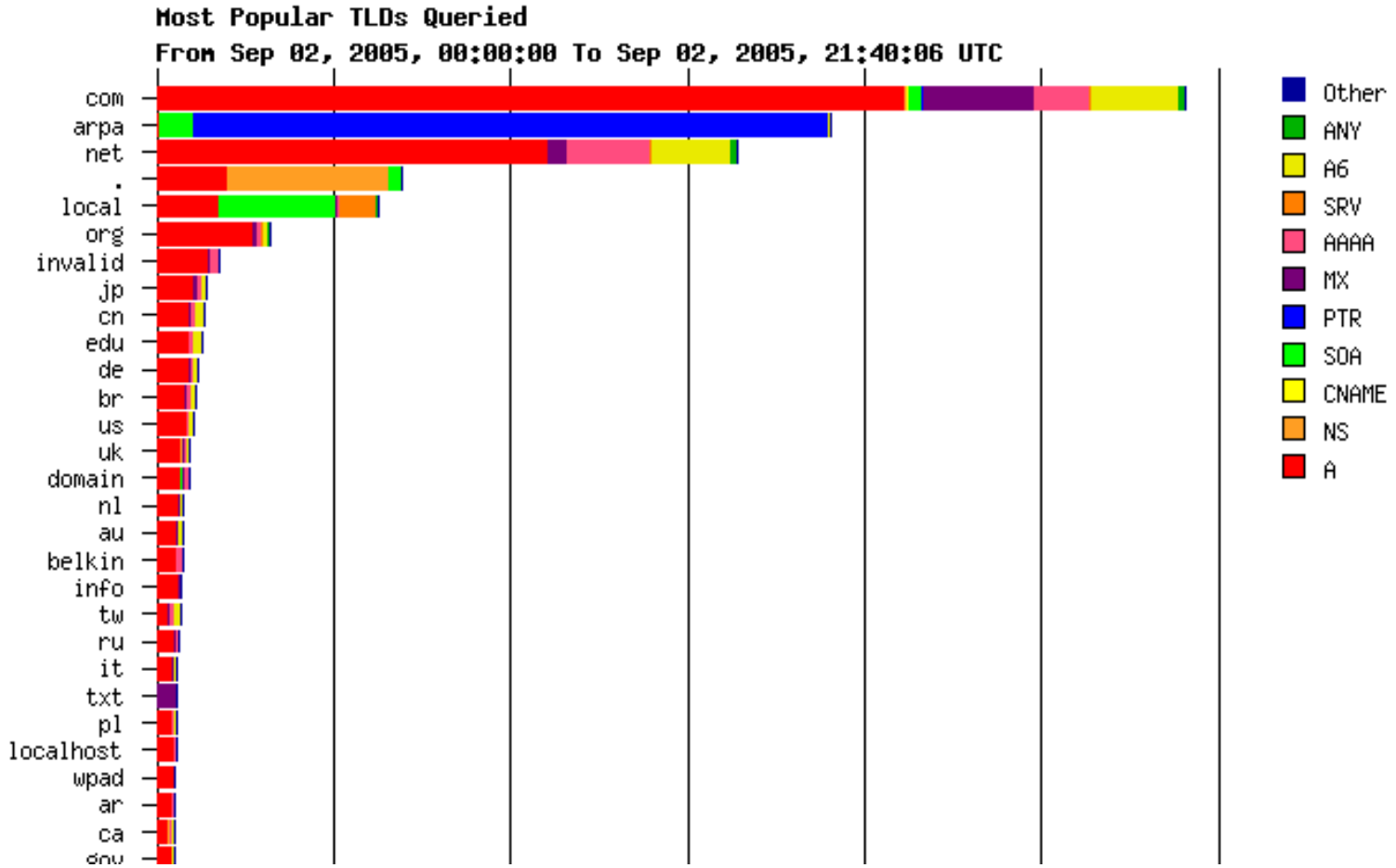
```
rndc reconfig
```

DNS DDoS Defenses

- Anycast
- Provisioning and scoping
  - Capacity, client separation, aa/rd separation
- Filtering, black holes and white lists
- Rate limiting and TCP switch-over (e.g. RRL)
- Upstream and peer cooperation
- Regulation, law enforcement
- Route poisoning

dsc

http://dns.measurement-factory.com/tools/dsc/



dnstop

http://dns.measurement-factory.com/tools/dnstop/

```
Queries: 0 new, 47 total

Query Name                Count       %
----------------- ---------- ------
example.org                25   53.2
example.edu                15   31.9
192.in-addr.arpa            6   12.8
ns1                         1    2.1
```

ZoneCheck
http://www.zonecheck.fr/

ZoneCheck: menog.net

Zone information

	menog.net	
1	ns1.2connectbahrain.com	46.29.56.196
2	ns2.2connectbahrain.com	80.88.242.4

Progress

- Testing: illegal symbols in domain name
- Testing: dash ('-') at start or beginning of domain name
- Testing: double dash in domain name
- Testing: one nameserver for the domain
- Testing: at least two nameservers for the domain
- Testing: identical addresses
- Testing: nameserver addresses are likely to be all on the same subnet
- Testing: nameservers belong all to the same AS
- Testing: delegation response fit in a 512 byte UDP packet
- Testing: delegation response with additional fit in a 512 byte UDP packet
- Testing: address in a private network (NS=ns1.2connectbahrain.com)

DNS Looking Glass

http://www.bortzmeyer.org/dns-lg.html

- DNS data inconsistency due to local policy
- Cached versus fetched answers
- Availability and delegation monitoring

Query for domain www.nanog.org., type A

- IP address: 12.22.58.49
- (Time-to-Live of this answer is 4 hours, 0 second)

Result obtained from resolver(s) ::1 at 2013-01-21 23:15:45Z. Query took 0:00:00.236909.

References

- DNS Tutorial @ IETF 80 – Gudmundsson, Koch
- Naming, DNS, & Security, DPU IT 263-901 - Lewis
- Securing DNS, EDUCAUSE SP – St Sauver
- DNS Debugging and monitoring, RIPE64 – Damas, Kerr
- An Introduction to DNSSEC, NANOG54 – Larson
- DNS(SEC) Troubleshooting, NANOG53 - Sinatra