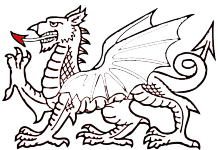


# Practical DNS Operations



John Kristoff [jtk@cymru.com](mailto:jtk@cymru.com)



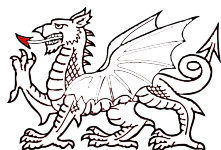
# DNS operational observations

- Flexibility as a virtue and scourge
- Expert pool is deep, but concentrated
- Best and common practices often undocumented
- Understanding of the deployed system is nascent
- Interest and innovation is ramping up



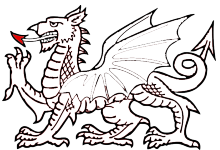
# One of two critical systems

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.

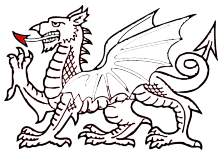


# DNS ops in three parts

- 1) Protocol and system overview
- 2) Best common practices (BCPs)
- 3) Introduction to advanced topics



# First a DNS resolution primer...

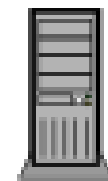


I need an IPv4 address for  
www.menog.net.

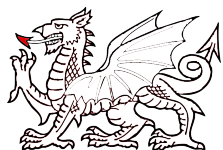
Please resolve it (recursion desired) for me?



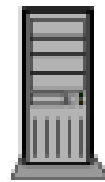
End User  
(stub resolver)



Local Caching Server  
(full resolver)



Check cache.  
If empty, ask a parent.  
Follow delegation if necessary.



Local Caching Server  
(full resolver)

parent zones: menog.net.  
net.  
.



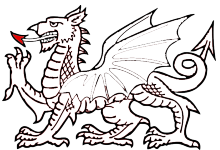
**Let's assume cache is empty, and  
all it knows about is (.) root.\***

**A.root-servers.net.**

**...**

**M.root-servers.net.**

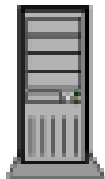
**\*Do you see why a reliable and trustworthy root is so important?**



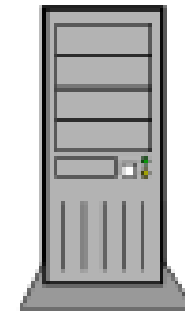


I need an IPv4 address for  
www.menog.net.

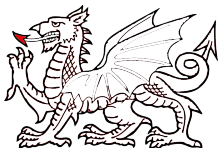
Can you tell me or refer me to someone?



Local Caching Server  
(full resolver)

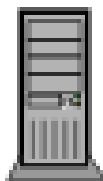


root (.) server

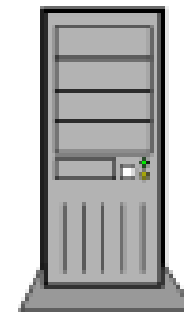


# Don't know. Try one of these .net servers:

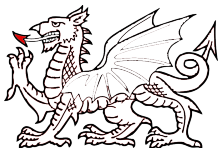
a.gtld-servers.net. b.gtld-servers.net.  
c.gtld-servers.net. d.gtld-servers.net.  
e.gtld-servers.net. f.gtld-servers.net.  
g.gtld-servers.net. h.gtld-servers.net.  
i.gtld-servers.net. j.gtld-servers.net.  
k.gtld-servers.net. l.gtld-servers.net.  
m.gtld-servers.net.



Local Caching Server  
(full resolver)

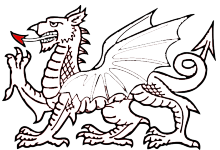


root (.) server

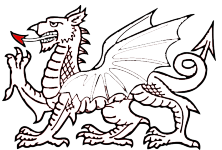


**Does the caching server have something in its cache now?**

**Raise your hand for yes.**



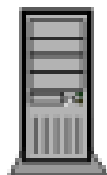
**Ultimately we should get here...**



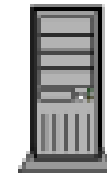
You've come to the right place.  
The authoritative answer is:

80.88.242.44

and that answer is valid  
for 7200 seconds



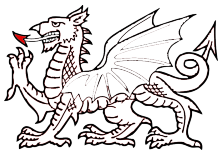
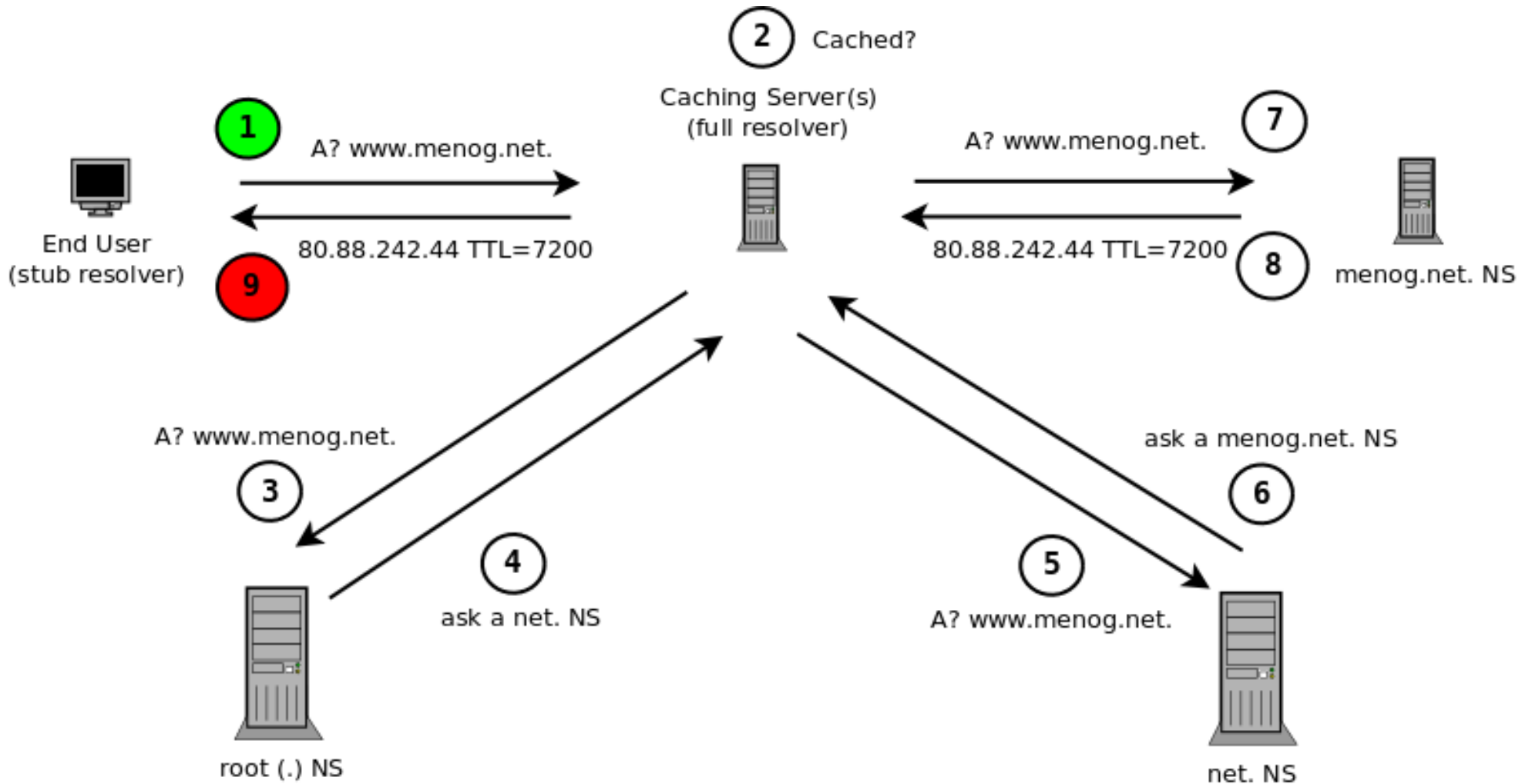
Local Caching Server  
(full resolver)



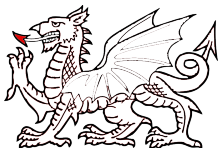
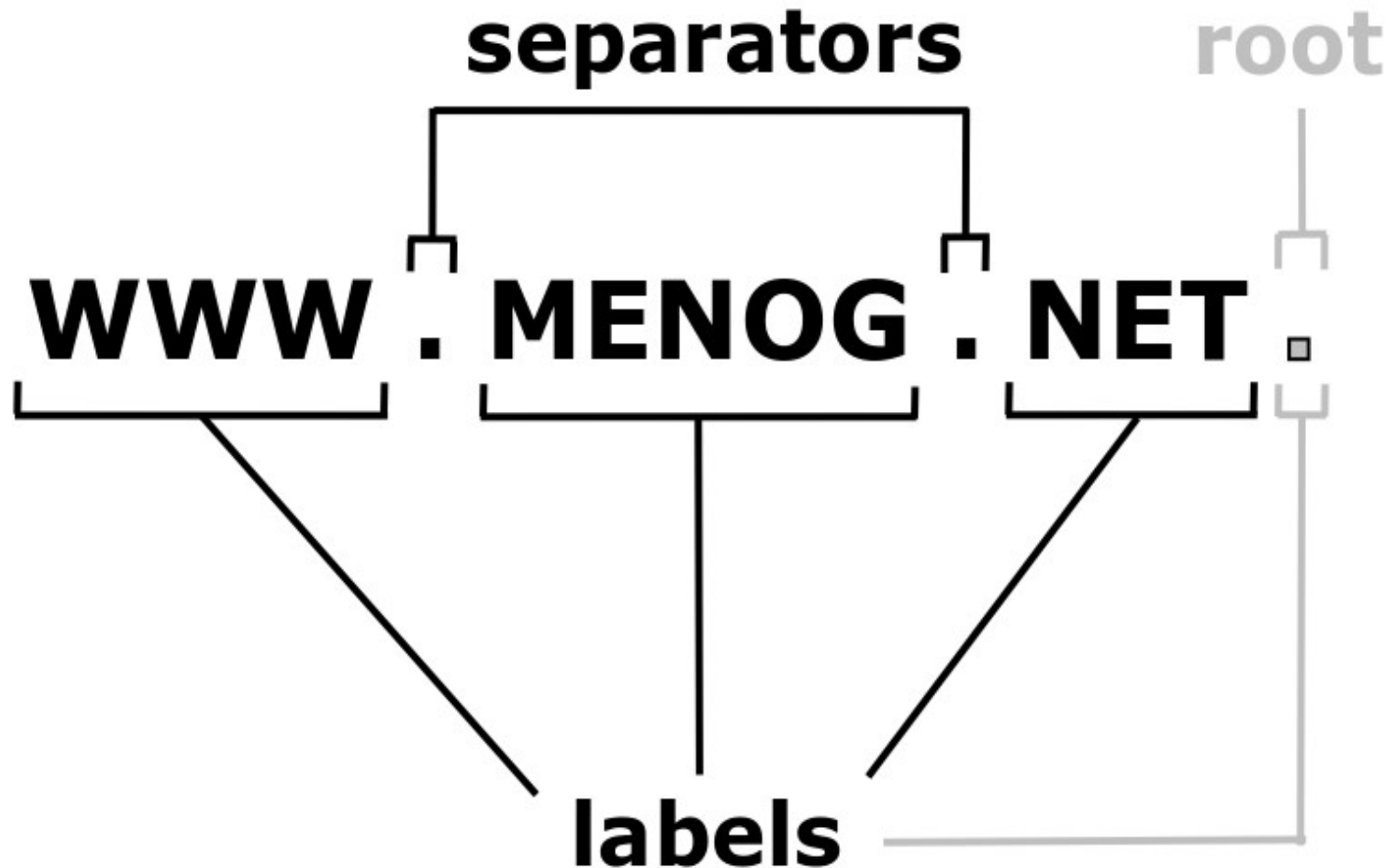
ns1.2connectbahrain.com.  
or  
ns2.2connectbahrain.com.



# Lookup summary

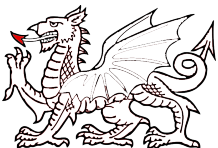


# Anatomy of a domain name



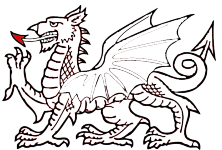
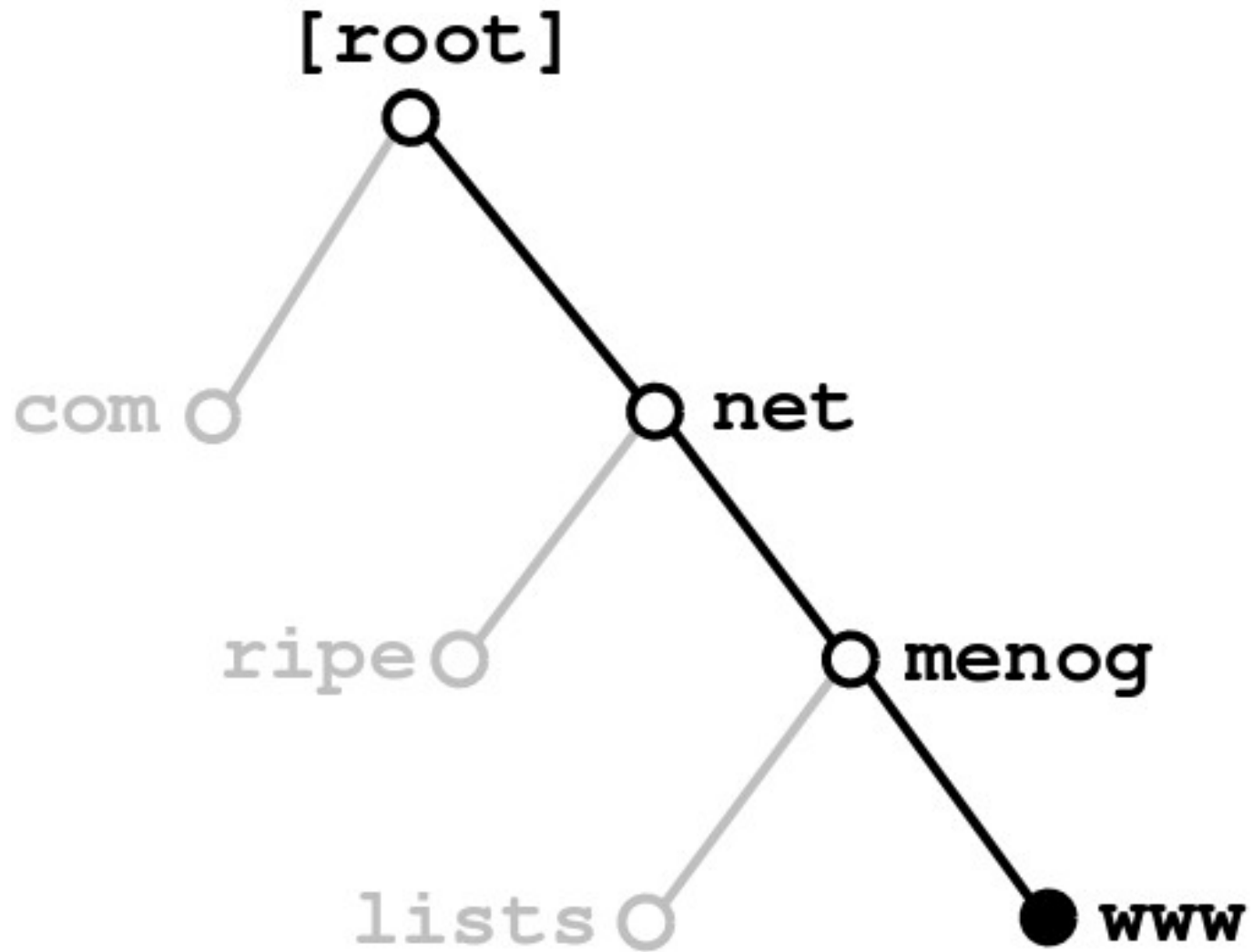
# What's in a name?

- 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/>



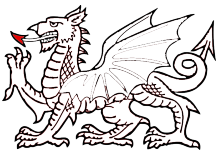


# Name space hierarchy



# 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.



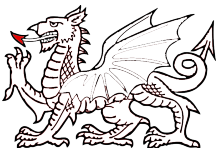
# Root zone

- 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



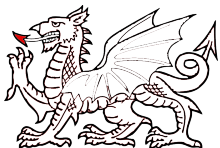
# Top-level domains (TLDs)

- 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>



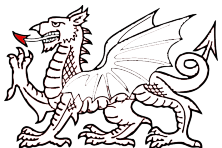
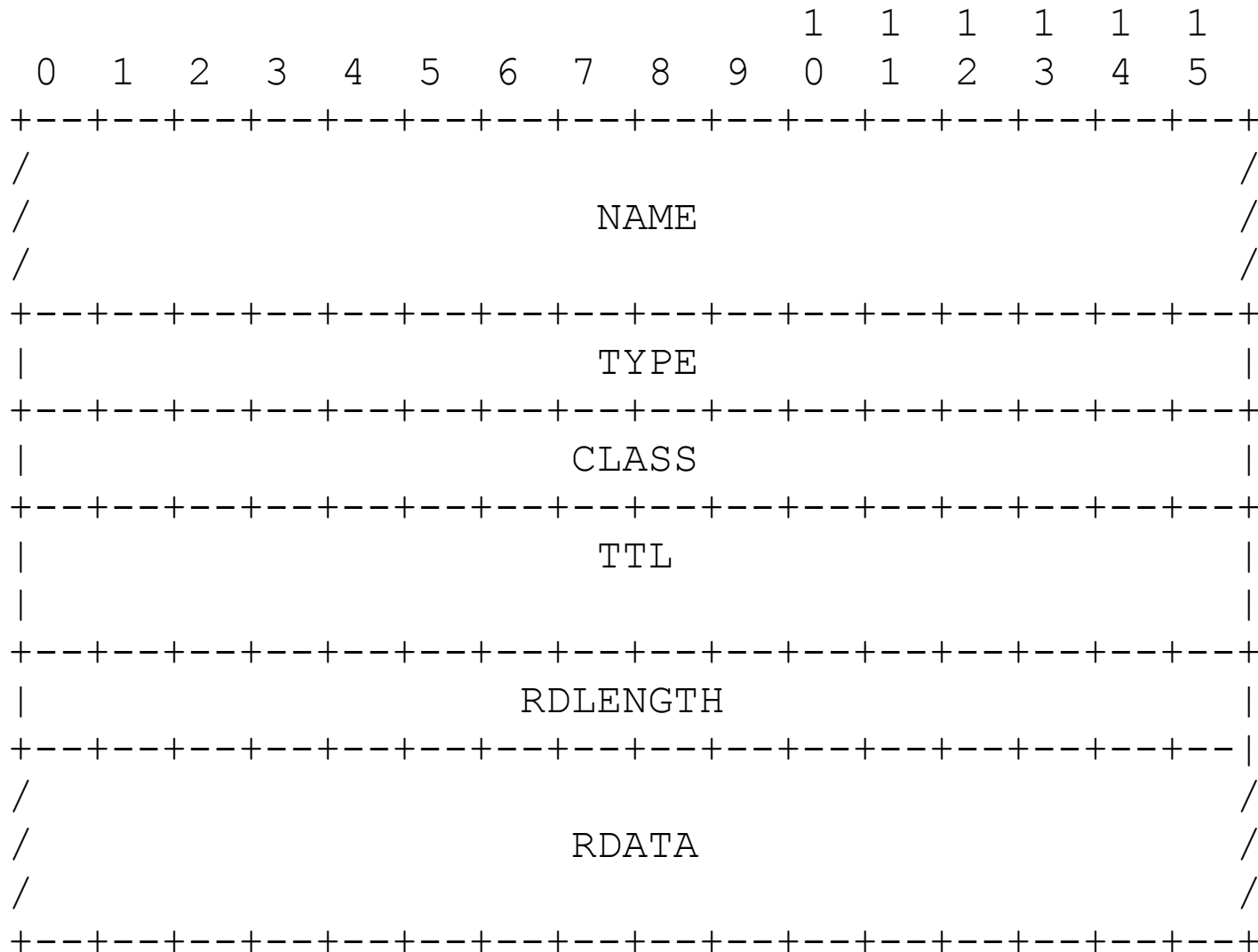
# DNS protocol message format

+-----+		
	Header	(see next slide)
+-----+		
	Question	the question for the name server
+-----+		
	Answer	RRs answering the question
+-----+		
	Authority	RRs pointing toward an authority
+-----+		
	Additional	RRs holding additional information
+-----+		



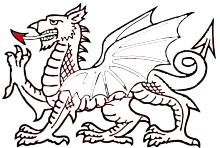


# DNS protocol RR format



# 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.”





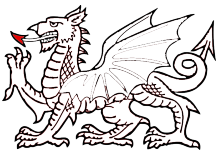
# Domain name registration

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

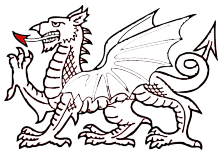


# WHOIS

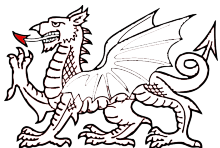
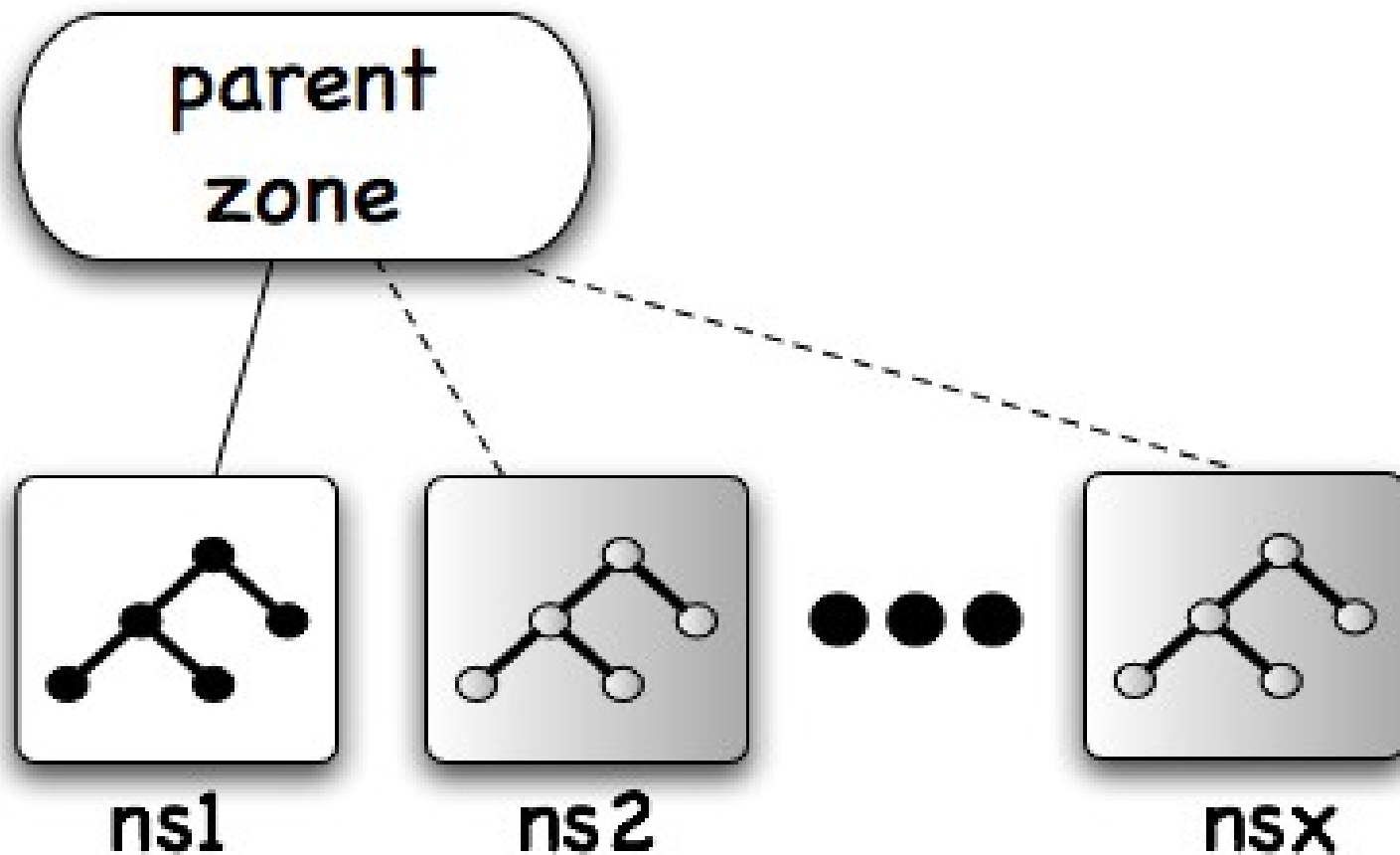
- Interface to assignees of Internet resources
  - e.g. domain names, IP addresses, ASNs
- Human readable text output
- Lacks modern design attributes
  - e.g. security, internationalization



# Best Common Practices (BCPs)



# How many NS RRs for your zone?

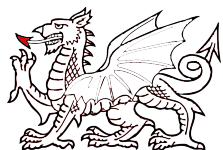
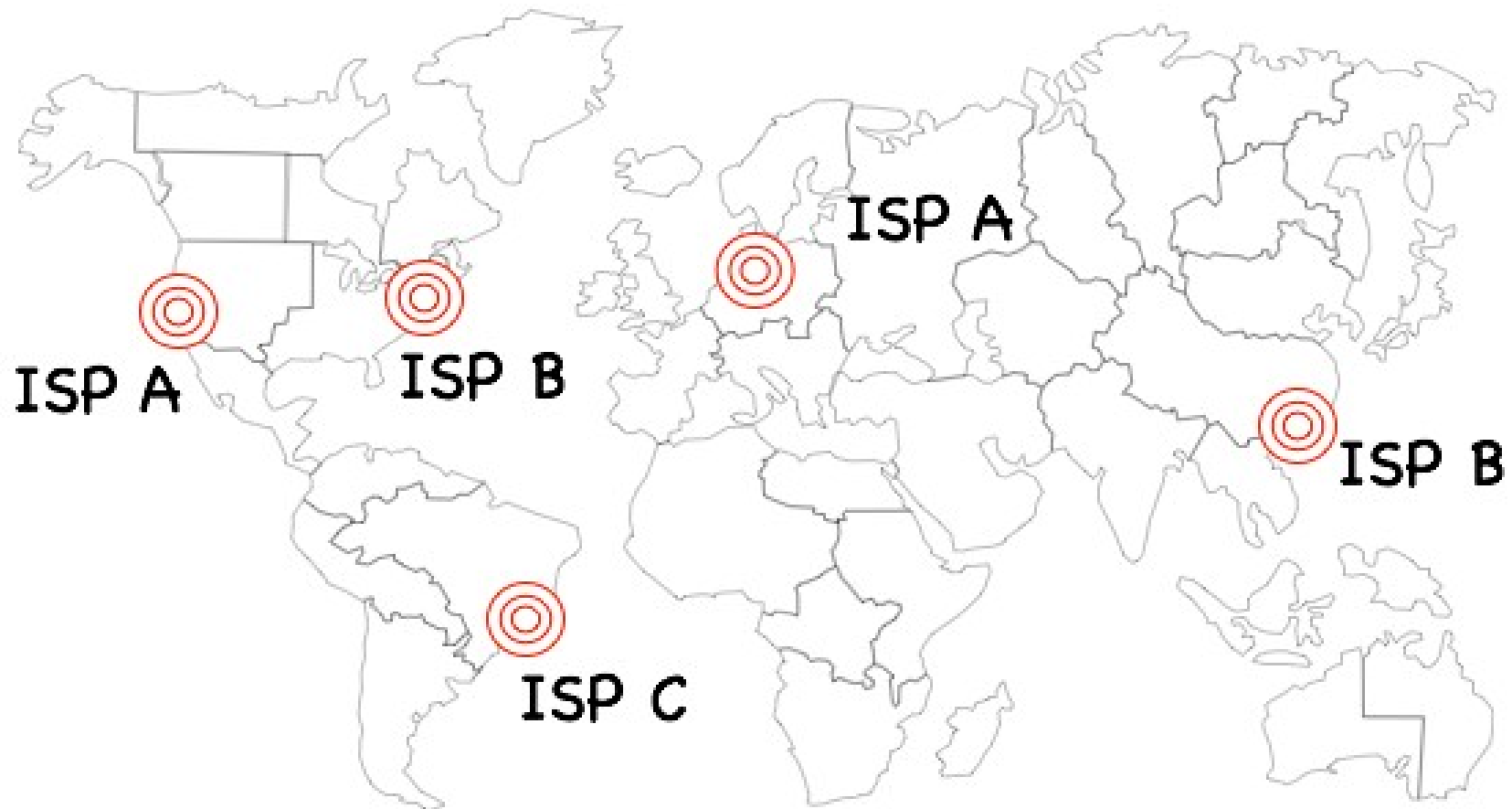


# Authoritative name server RRset

- 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...”

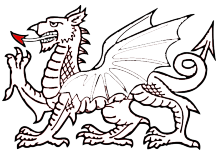


# Where are your name servers?



# 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



# Are parent and children consistent?

example. TLD



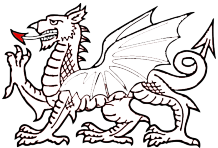
...  
foo NS ns1.foo.example.  
foo NS ns2.foo.example.  
foo NS bob.bar.example.  
...

---

ns1.foo.example.



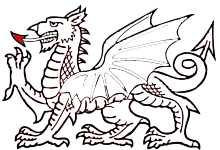
foo NS ns1.foo.example.  
foo NS ns2.foo.example.  
foo NS ns3.bar.example.



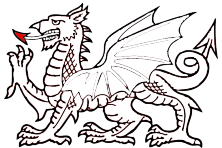
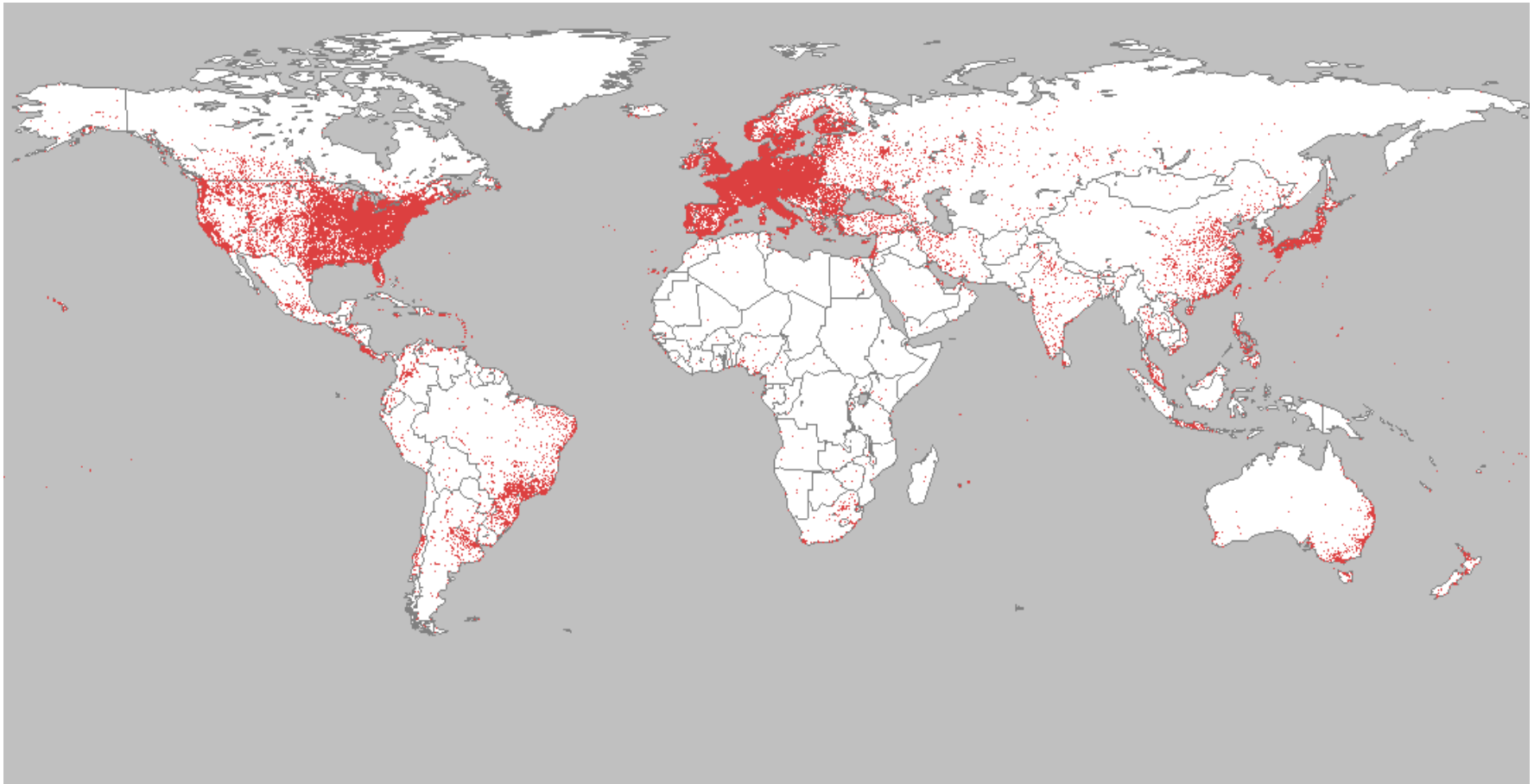


# Delegation Consistency

- 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



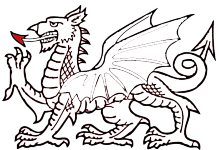
# Does your server answer anything from anyone?



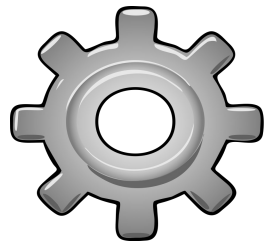
# Open Resolvers

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

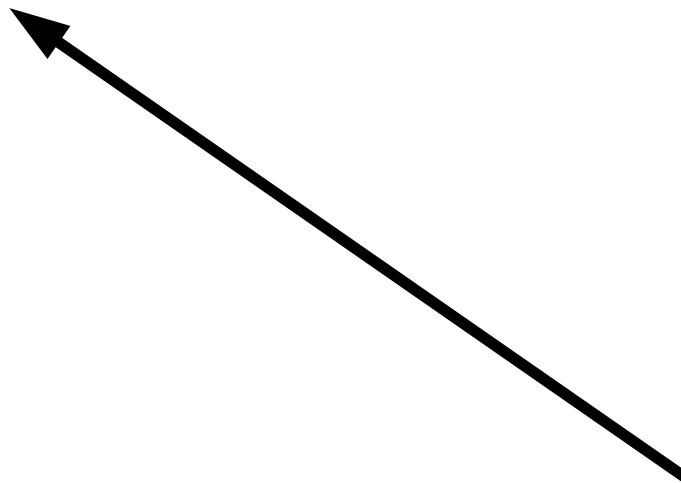
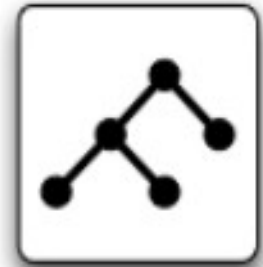
<http://www.team-cymru.org/Services/Resolvers/>



# How easily can returning answers be spoofed?



What is the rdata/ttl for ... ?

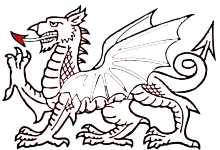


HERE IT IS!! Mmwuahaha...

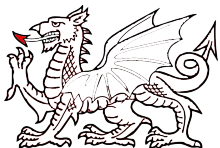
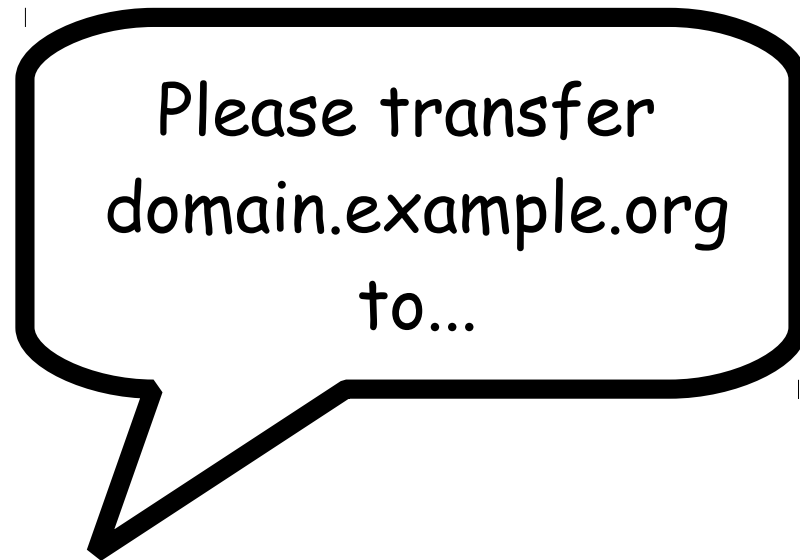


# Answer Spoofing Protection

- 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?

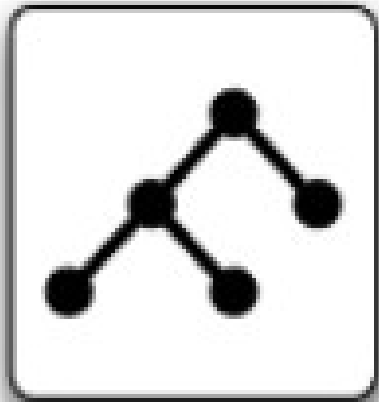


# Domain Name Registration

- 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



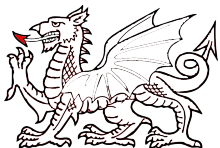
# What is on your name server?



+

httpd  
snmpd  
ftpd  
proxyd  
dhcpcd

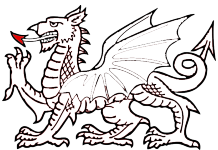
=





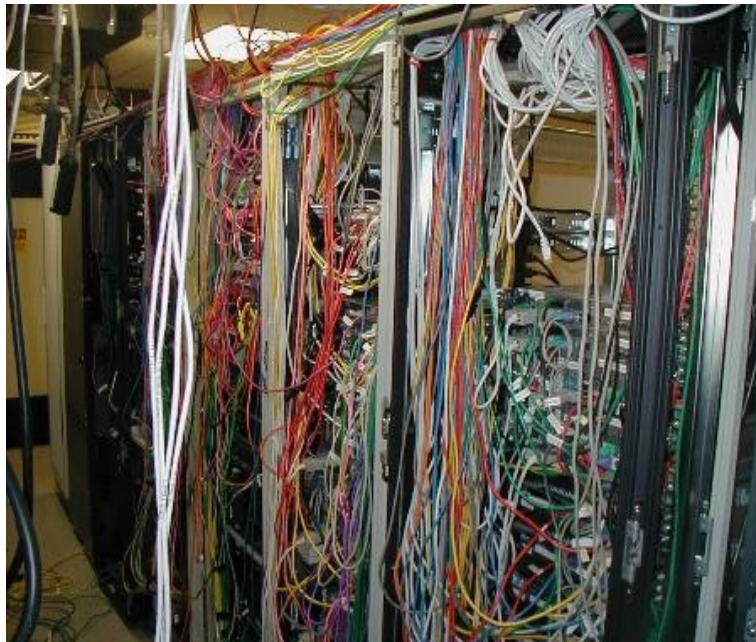
# Co-mingling Services

- 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

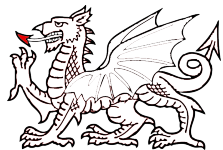


# How are servers administered?

pictures from techrepublic, Bill Detwiler



OR



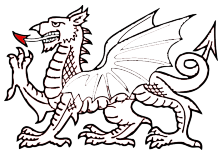
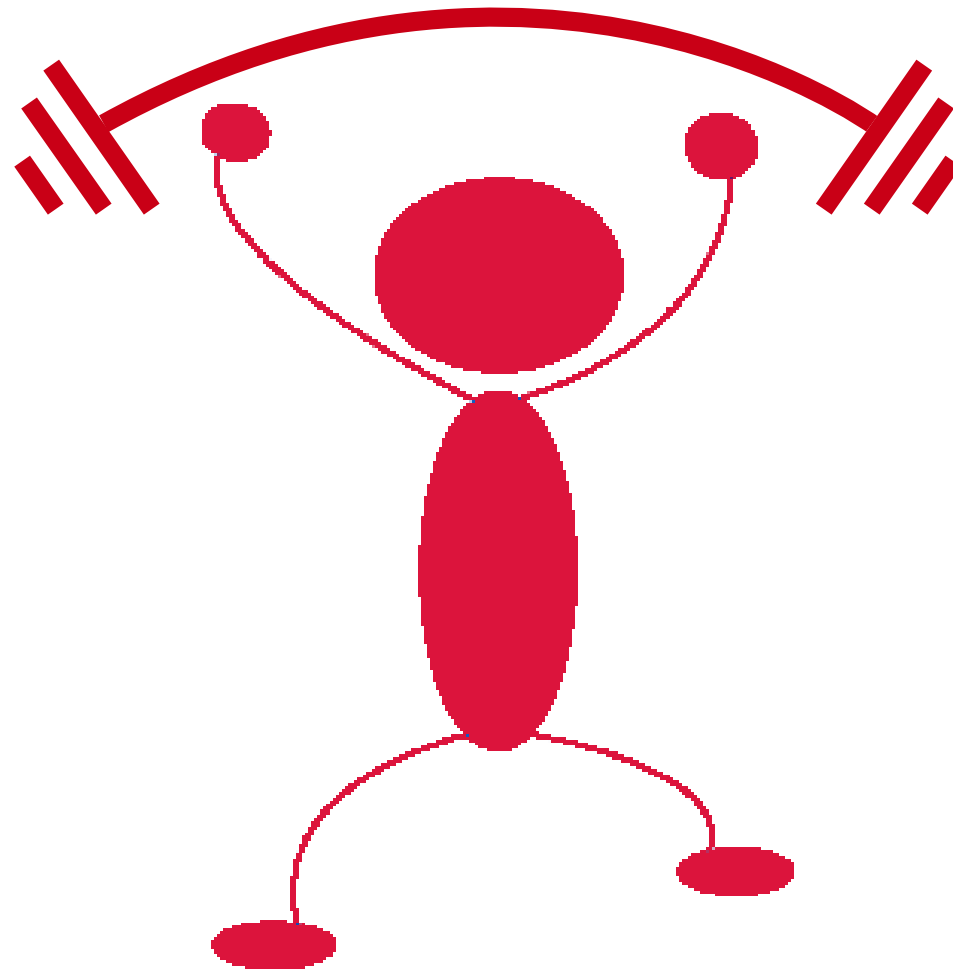
# Administrative Processes

- 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

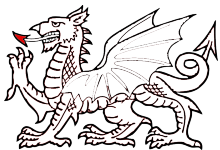
- Don't have enough, have way more than enough
- Resolvers can demand lots of RAM
- CPU may be important, especially for crypto
- Hard drives usually less important
  - Isolating partitions and directories may be useful
  - Try to offload data collection to another system
- Network capacity usually not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



# Are you filtering DNS over TCP?



OR



# TCP

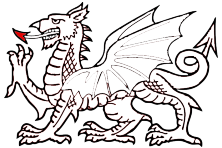
- Don't assume you have no DNS over TCP
- TCP isn't just for zone transfers
  - Large DNS messages may use TCP
  - Some operators may force TCP during DDoS
- TCP tuning may be required for some DoS threats



# What queries do you see/make?



<http://www.wordle.net>





# Monitoring and Auditing

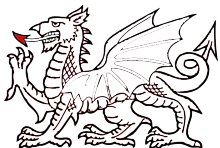
- Troubleshooting with query insight is very helpful
- Consider learning answers from the resolvers too
  - AKA passive DNS
- Minimally, trend DNS query/answer statistics
- Monitor servers, answers and routes from outside

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

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

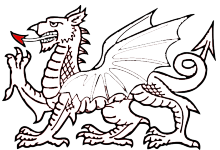


# Are name server clocks accurate?



# 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



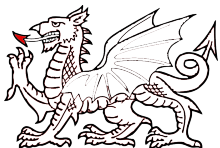
# Have you read IETF RFC 2870?



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

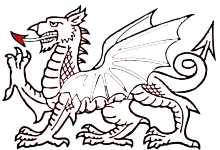
R. Bush  
Verio  
D. Karrenberg  
RIPE NCC  
M. Koster  
Network Solutions  
R. Plzak  
SAIC  
June 2000

Root Name Server Operational Requirements

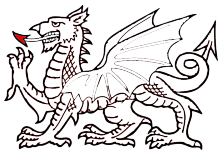


# IETF RFC 2870

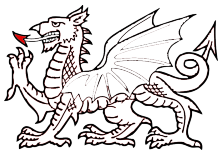
- Its a BCP, you should be familiar with it
- Its a bit dated and written for a specific audience
  - But it contains sound advice for most everyone
- A newer, generalized version may soon appear



# Advanced Topics

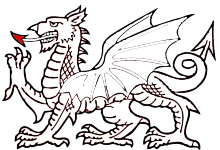


# Fast Flux DNS



# DNS terminology review

- resource record (RR)
  - database entry (row) about a domain name
- RRset
  - RRs with the same name, class, type (and TTL)
- A
  - DNS RR of type A, for IPv4 address record(s)
- NS
  - DNS RR of type NS, for name server records(s)





# dig output of an “A” query

```
$ dig @ns1.2connectbahrain.com. www.menog.net. IN A
```

```
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 12345
```

```
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 2, ADDITIONAL: 2
```

```
;; QUESTION SECTION:
```

```
;www.menog.net . IN A
```

```
;; ANSWER SECTION:
```

```
www.menog.net. 7200 IN A 80.88.242.44
```

```
;; AUTHORITY SECTION:
```

```
menog.net. 172800 IN NS ns1.2connectbahrain.com.  
menog.net. 172800 IN NS ns2.2connectbahrain.com.
```

```
;; ADDITIONAL SECTION:
```

```
ns1.2connectbahrain.com. 7200 IN A 46.2956.196  
ns2.2connectbahrain.com. 172800 IN A 80.88.242.4
```



# dig output of an “NS” query

```
$ dig @ns1.2connectbahrain.com. menog.net. IN NS
```

```
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 12345  
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 2
```

```
;; QUESTION SECTION:
```

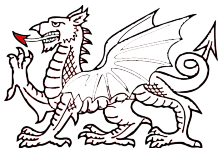
```
;menog.net.                IN      NS
```

```
;; ANSWER SECTION:
```

```
menog.net.                172800  IN      NS      ns1.2connectbahrain.com.  
menog.net.                172800  IN      NS      ns2.2connectbahrain.com.
```

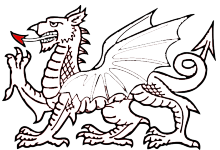
```
;; ADDITIONAL SECTION:
```

```
ns1.2connectbahrain.com.  7200    IN      A       46.29.56.196  
ns2.2connectbahrain.com.  172800  IN      A       80.88.242.4
```



# Affecting availability with DNS

- The RRs in an answer or NS RRset
- RRset TTL
- Unique answer based on origin (geoloc/views)
- Unique answer based on time
- Wildcards, answering authoritatively



# Different origin, different answer

```
$ dig www.google.com
```

```
;; ANSWER SECTION:
```

www.google.com.	604800	IN	CNAME	www.l.google.com.
www.l.google.com.	300	IN	A	74.125.95.147
www.l.google.com.	300	IN	A	74.125.95.99
www.l.google.com.	300	IN	A	74.125.95.103
www.l.google.com.	300	IN	A	74.125.95.104

```
$ dig www.google.com @4.2.2.2
```

```
;; ANSWER SECTION:
```

www.google.com.	43190	IN	CNAME	www.l.google.com.
www.l.google.com.	300	IN	A	209.85.171.103
www.l.google.com.	300	IN	A	209.85.171.104
www.l.google.com.	300	IN	A	209.85.171.147
www.l.google.com.	300	IN	A	209.85.171.99



# A RRset fast-fluxing

```
$ dig vqthe.cn
```

```
;; ANSWER SECTION:
```

vqthe.cn.	180	IN	A	89.46.127.47
vqthe.cn.	180	IN	A	123.237.100.126
vqthe.cn.	180	IN	A	123.237.108.142
vqthe.cn.	180	IN	A	190.191.142.122
vqthe.cn.	180	IN	A	196.202.6.66
vqthe.cn.	180	IN	A	71.239.64.226
vqthe.cn.	180	IN	A	78.92.180.208
vqthe.cn.	180	IN	A	85.254.64.153

```
;; AUTHORITY SECTION:
```

vqthe.cn.	180	IN	NS	ns2.krcrab.com.
vqthe.cn.	180	IN	NS	ns1.czwill.com.
vqthe.cn.	180	IN	NS	ns4.krcrab.com.
vqthe.cn.	180	IN	NS	ns2.czwill.com.

```
;; ADDITIONAL SECTION:
```

ns1.czwill.com.	172799	IN	A	78.92.180.208
ns2.czwill.com.	172799	IN	A	85.67.171.146
ns2.krcrab.com.	172799	IN	A	61.61.61.61
ns4.krcrab.com.	172799	IN	A	138.16.6.201



# Re-query, notice changes

```
$ dig vqthe.cn
```

```
;; ANSWER SECTION:
```

vqthe.cn.	180	IN	A	85.67.171.146
vqthe.cn.	180	IN	A	89.44.56.76
vqthe.cn.	180	IN	A	89.102.112.60
vqthe.cn.	180	IN	A	116.72.241.170
vqthe.cn.	180	IN	A	124.125.245.32
vqthe.cn.	180	IN	A	190.245.216.89
vqthe.cn.	180	IN	A	79.140.228.27
vqthe.cn.	180	IN	A	85.29.210.207

```
;; AUTHORITY SECTION:
```

vqthe.cn.	180	IN	NS	ns1.czwill.com.
vqthe.cn.	180	IN	NS	ns2.krcrab.com.
vqthe.cn.	180	IN	NS	ns2.czwill.com.
vqthe.cn.	180	IN	NS	ns4.krcrab.com.

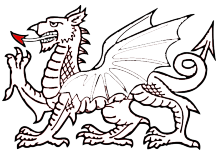
```
;; ADDITIONAL SECTION:
```

ns1.czwill.com.	172585	IN	A	78.92.180.208
ns2.czwill.com.	172585	IN	A	85.67.171.146
ns2.krcrab.com.	172585	IN	A	61.61.61.61
ns4.krcrab.com.	172585	IN	A	138.16.6.201



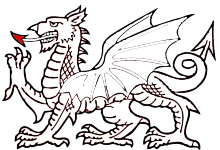
# Single-flux vs double-flux

- If you know two more buzzwords than the other guy, you're an expert
- Single-flux: the A RRs in the answer flux
- Double-flux: the A RRs for name server names flux
  - Note, name usually stays the same, but the name server IP addresses for the authoritative name servers at parent (usually registry) or child may be changing. This could be automated through a registrar's API or via a script at the child name server(s).



# Why is this cool? or... Why is this bad?

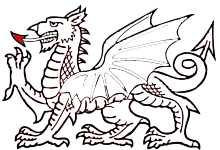
- Its a great way to ensure availability
- Taking away any single host has almost no impact
- How do you take down potentially dozens, if not hundreds of hosts in the A RRset?
- Take down the name?
  - Not all registrars or registries are willing and/or are able to support this whack-a-mole process





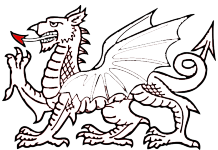
# Law enforcement problem

Fast-flux makes it literally impossible for them to bust bad guys. They are just not equipped to deal with these challenges and infrastructure. :-)



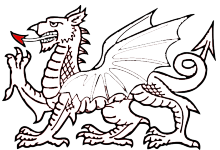
# Malicious use mitigation

- ICANN registry/registrar agreements can help
- Registrar and registry response capability is key
- Some success in detection and monitoring projects

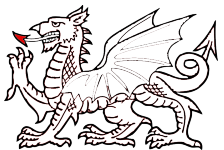


# An aside: Domain name generation algorithms in malware

- 1) worm generates pseudo-random name
- 2) attempts to contact server at random name
- 3) if not authentic, try another
- 4) if authentic, follow instructions
- Pool of names can be large and widely dispersed
  - i.e. many TLDs, registries, registrars affected
  - example worm: Conficker
- Problem: how do you mitigate so many names?

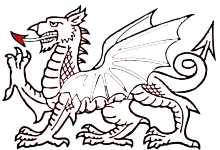


# Passive DNS



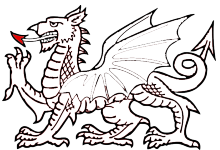
# History

- 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/>



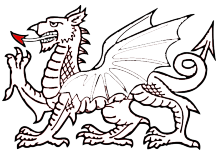
# Before passive DNS

- 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)



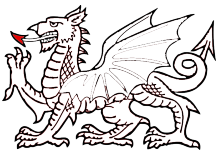
# After passive DNS

- 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



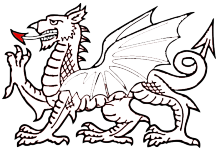
# Other passive DNS uses

- Cache poisoning detection
- Auditing and usage violation monitoring
  - e.g. our Netnames project (see TC Console)
- System and network profiling
- DNS hijacking analysis
- Other basic research



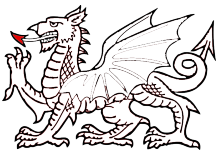


# BIND Administration Options



# Useful 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";`



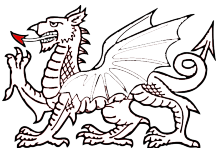
# Named pipe for query logging

- 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`



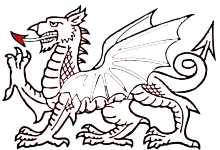
# Domain Name Hijacking

- 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



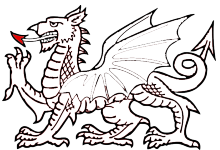
# 1) Create mitigating zone file

```
$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."
```



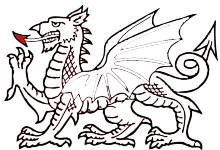
## 2) Add zone to named.conf

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

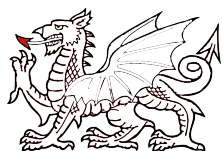


## 3) Load the new zone

```
rndc reconfig
```

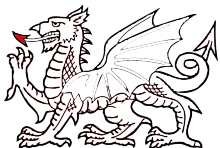
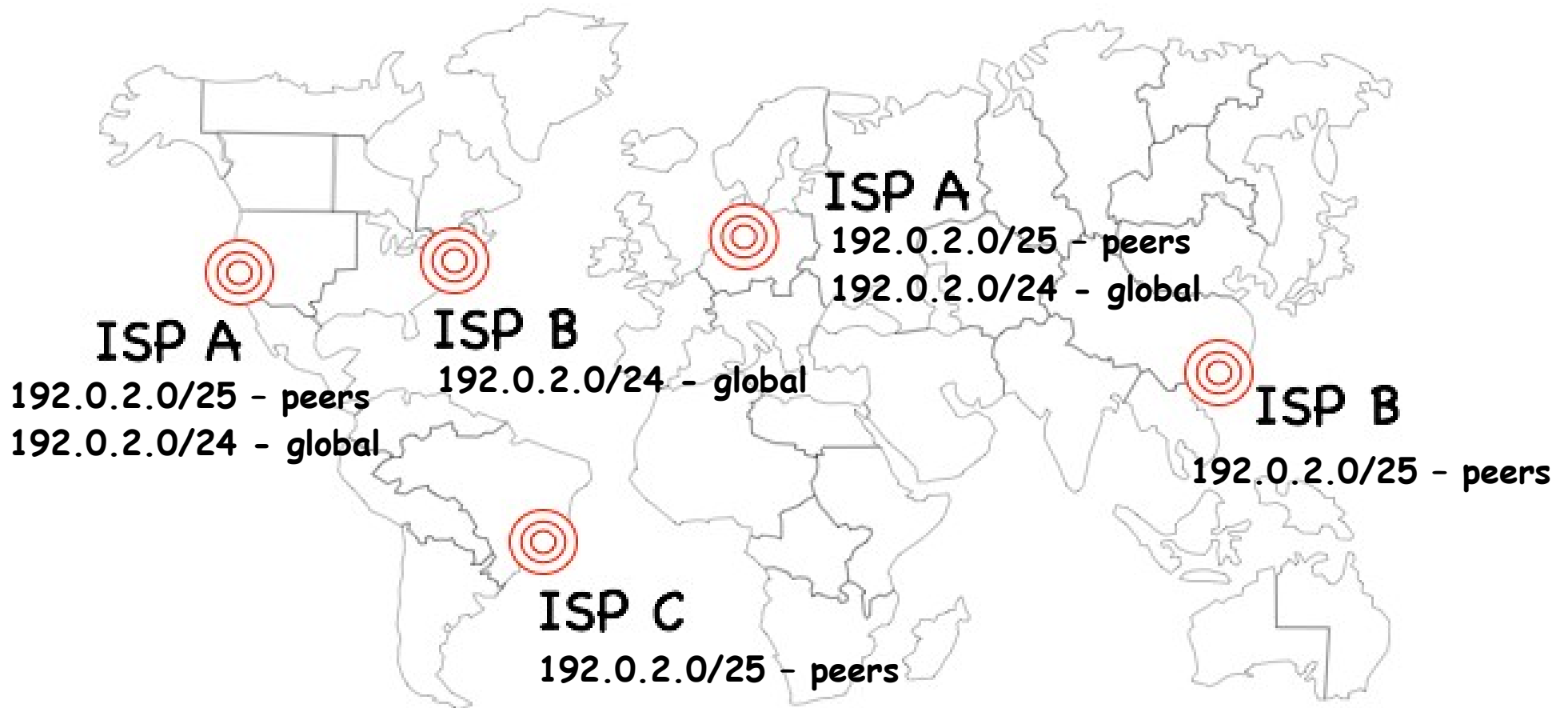


# Ancast





# Shared unicast addressing

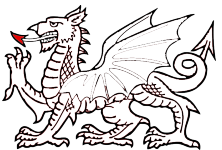


# Deployment

- 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 implementation notes

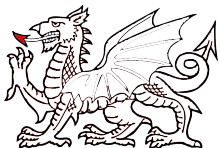
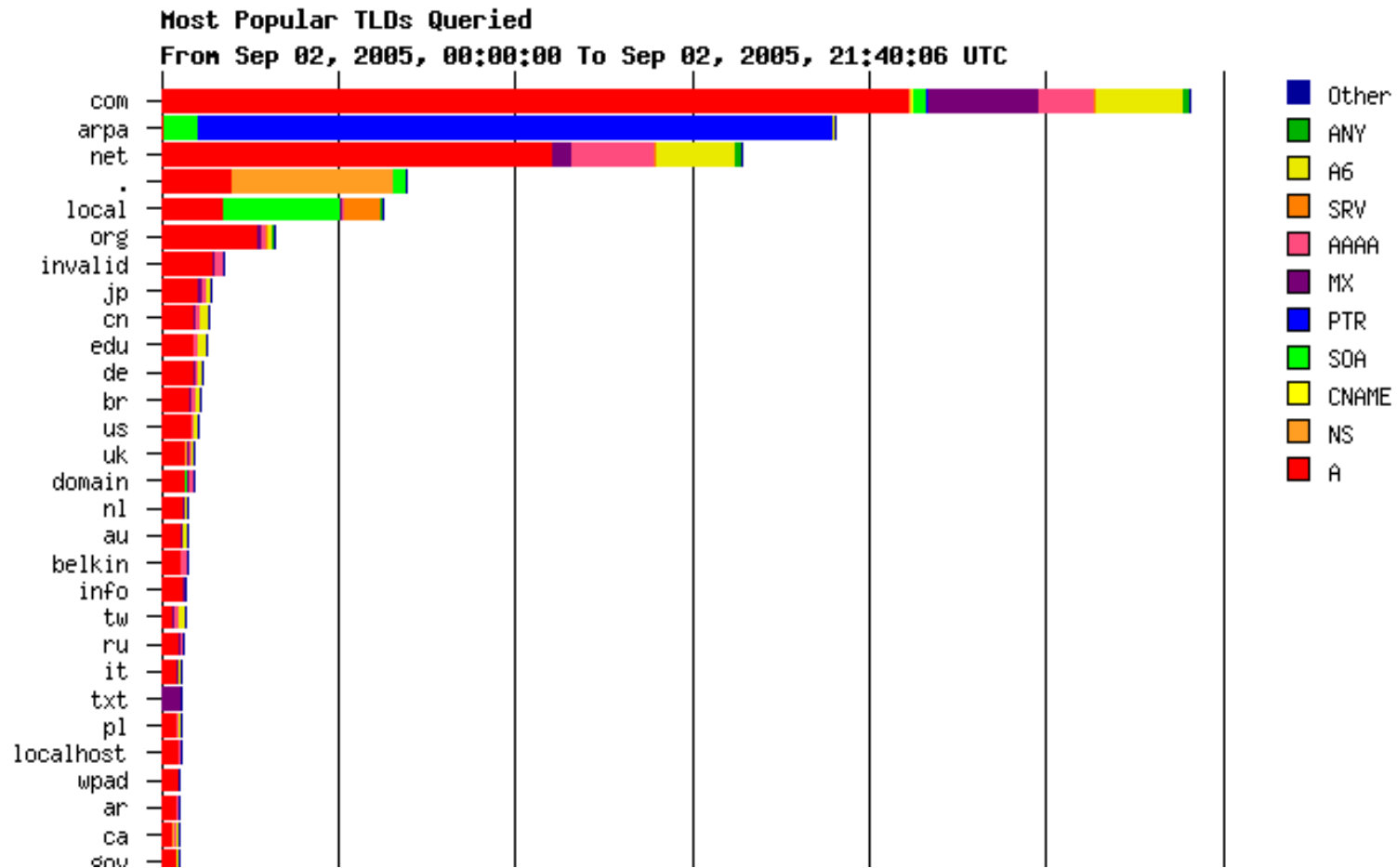


# Tools



# dsc

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

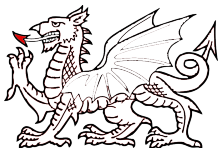


# dnstop

<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/>

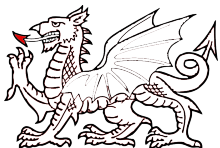
## ZoneCheck: menog.net

### Zone information

	menog.net	
	ns1.2connectbahrain.com	46.29.56.196
	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)



# How can Team Cymru help?

- Secure BIND template
- Open resolver feed
- DNS and BGP monitoring
- Passive DNS
- Returning soon: Lame delegation report
- Coming soon: DNS Report
- Feedback or questions to: [jtk@cymru.com](mailto:jtk@cymru.com)
  - <http://www.cymru.com/jtk/>
  - Everything else @ <http://www.team-cymru.org>

