

Next-Generation DNS Monitoring Tools



Cyber Security Division 2012 Principal Investigators' Meeting

October 9, 2012

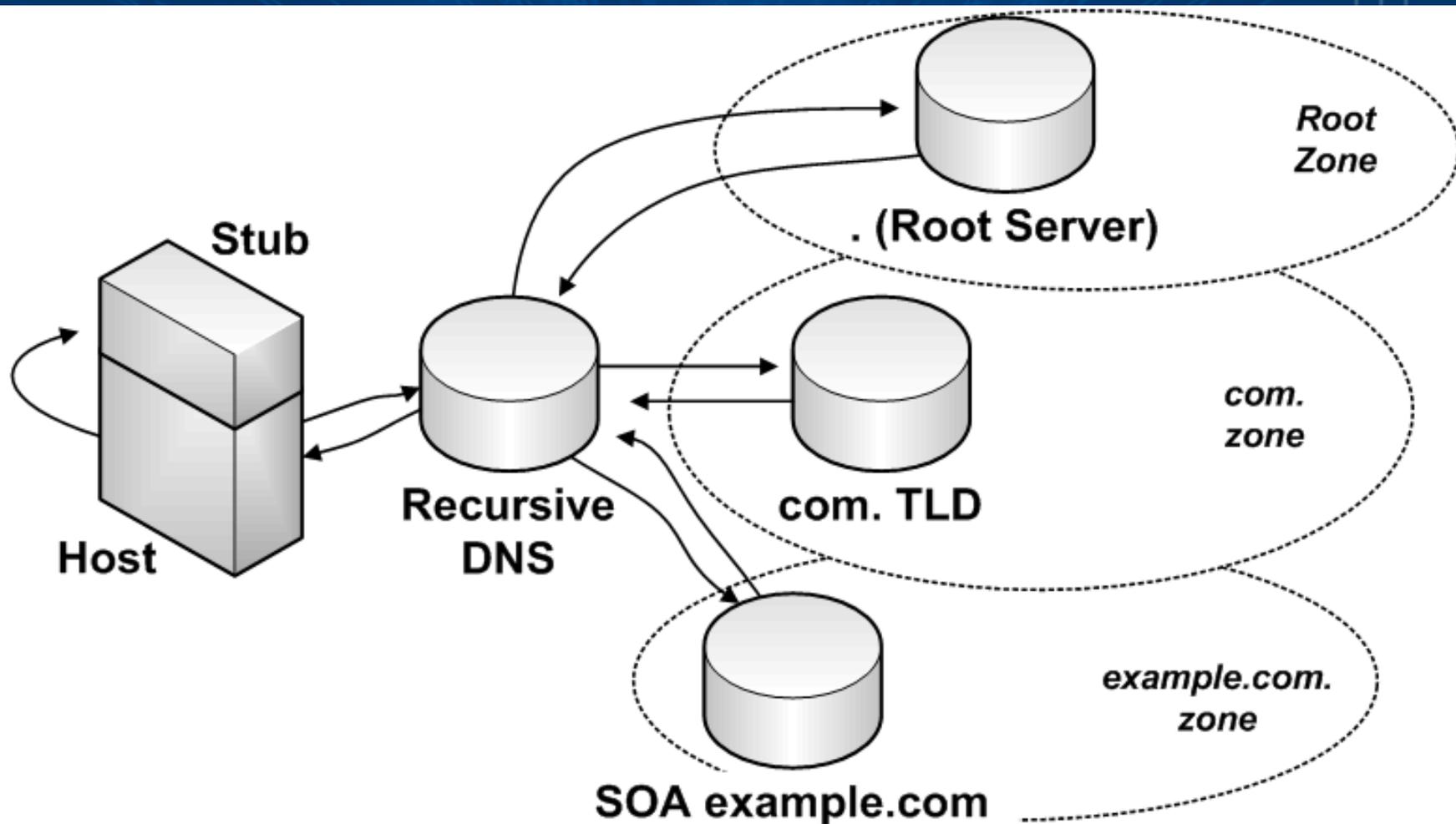
Wenke Lee and David Dagon
Georgia Institute of Technology
wenke@cc.gatech.edu
404-808-5172

Outline

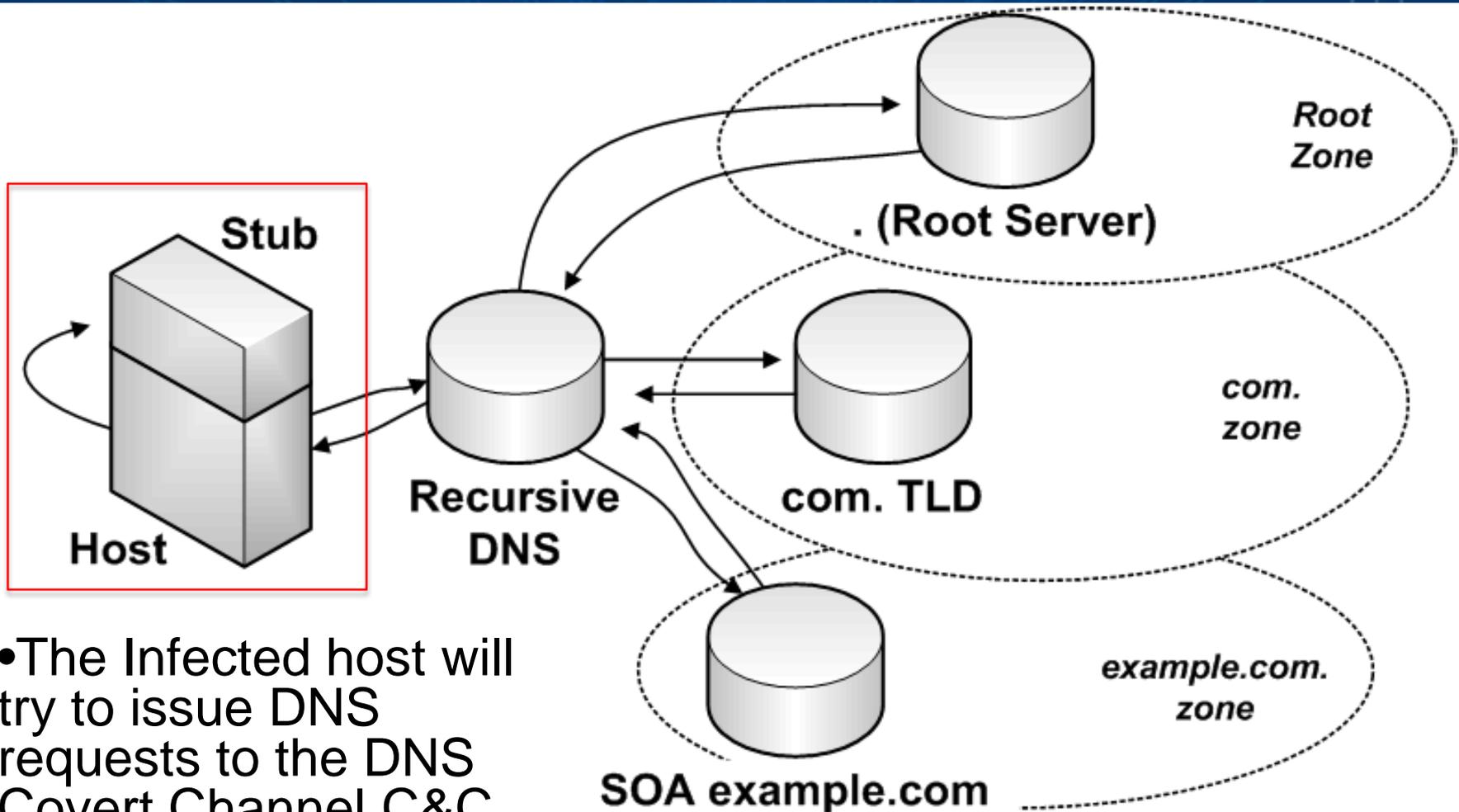
- DNS and Covert Channels
 - Storage-type DNS Covert Channel Detector
 - Real World DNS Covert Channel Detection
 - Morto DNS Trojan
- DNS-Based Traceback

Detection of DNS Covert Channels

DNS Resolution Process

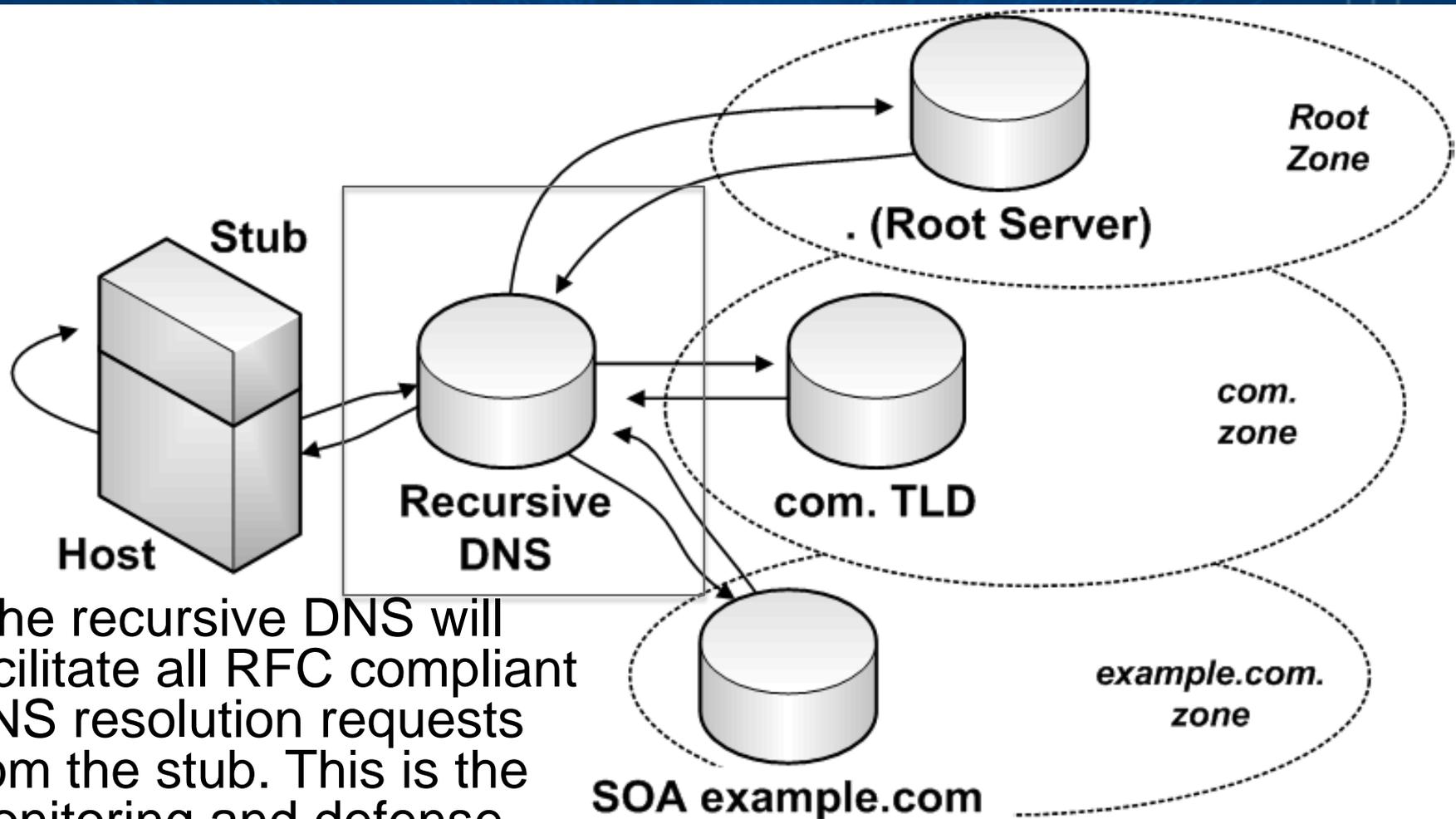


DNS Covert Channel Components



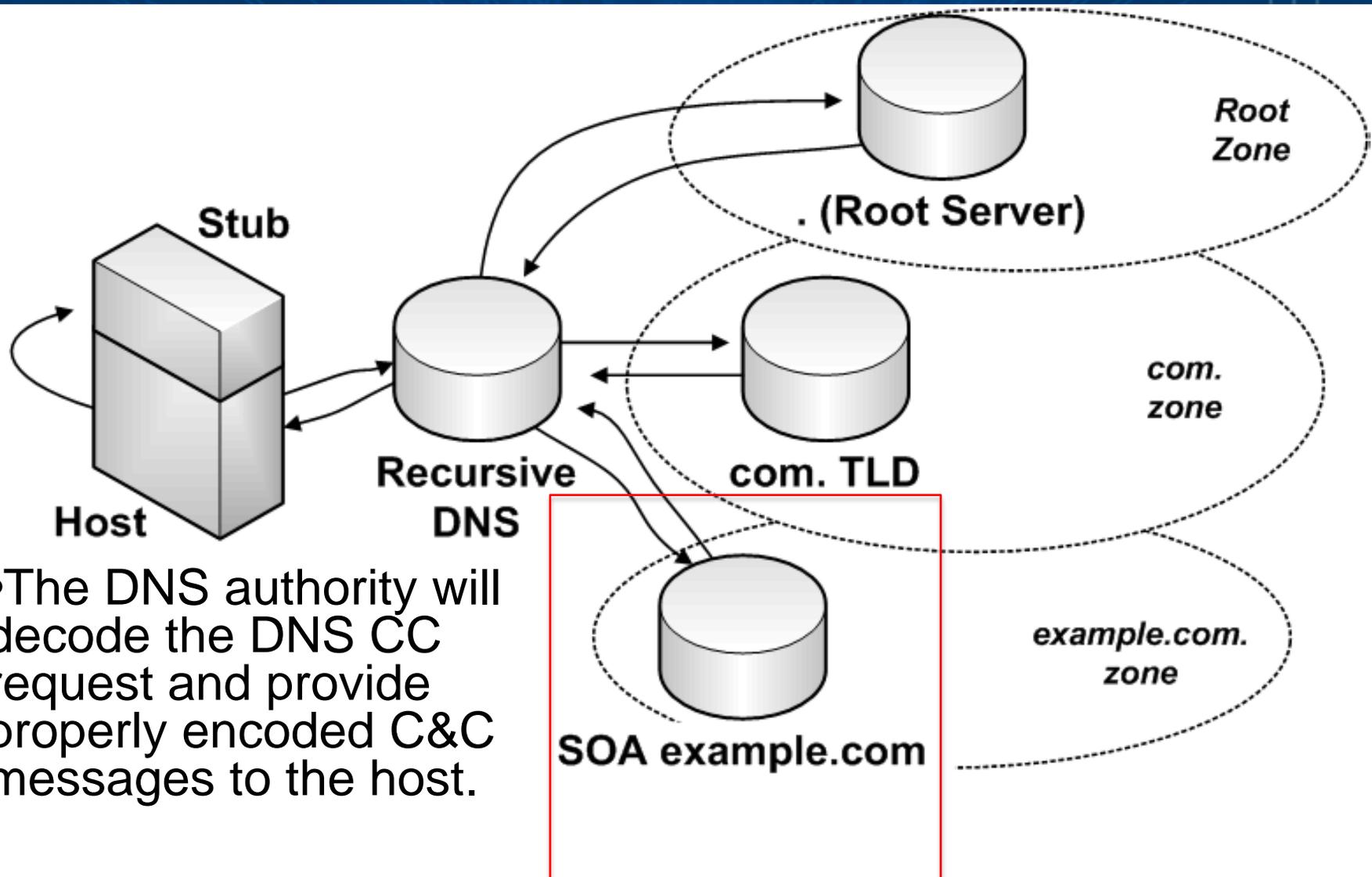
- The Infected host will try to issue DNS requests to the DNS Covert Channel C&C

DNS Covert Channel Components



- The recursive DNS will facilitate all RFC compliant DNS resolution requests from the stub. This is the monitoring and defense point.

DNS Covert Channel Components



- The DNS authority will decode the DNS CC request and provide properly encoded C&C messages to the host.

Main Types of DNS Covert Channels

Timing DNS Covert Channels

- Definition: Use timing properties of DNS requests to transmit messages to the C&C.
- Pros
 - Very stealthy
- Cons
 - Very low bandwidth
 - DNS caching and C&C protocol problems
 - Can be disturb easily by using pump-like network systems

Main Types of DNS Covert Channels

Timing DNS Covert Channels

- Definition: Use timing properties of DNS requests to transmit messages to the C&C.
- Pros
 - Very stealthy
- Cons
 - Very low bandwidth
 - DNS caching and C&C protocol problems
 - Can be disturb easily by using pump-like network systems

Storage DNS Covert Channels

- Definition: Overload the protocol to encapsulate encoded information (e.g., DNS-CC for C&C)
- Pros
 - High bandwidth
 - Easy to set up
 - Reliable (as part of DNS)
 - Hard to detect
- Cons
 - Passive DNS traces

Storage DNS-CC: Some Details

- (Infected) host:
 - Issues queries to domain names under the zone *.61020.pfwhereismyshoe.com
- DNS Covert Channel C&C (or external DNS authority):
 - The DNS authority acting as DNS-CC C&C sends answers back with encoded (in this case encrypted as well) messages as part of the TXT DNS answer fields

```
.....0.61020.pfwhereismyshoe.com.....
.....0.61020.pfwhereismyshoe.com.....fH...
VzZUVViL4GKsrqvLDqHdXVQP4RgesNo/EGdcXCKH2j1NgxR0AwMMcjhWU
JNobspmms+i+ML2wsWIOfpp4H+tPu31ojRyosONzmiJd9Tdqaay5/weAq
OFm0+VdWY1i3YqJGriLmvEla+5gUt3OFEF8el/qu+PM2C+sL25ccxyuBf
7VDvp+zrhn22YmXgPiGIQKVHMQv9+5iy2JGQlu/YUaltB1B.....
.....ns1.....ns2..
```

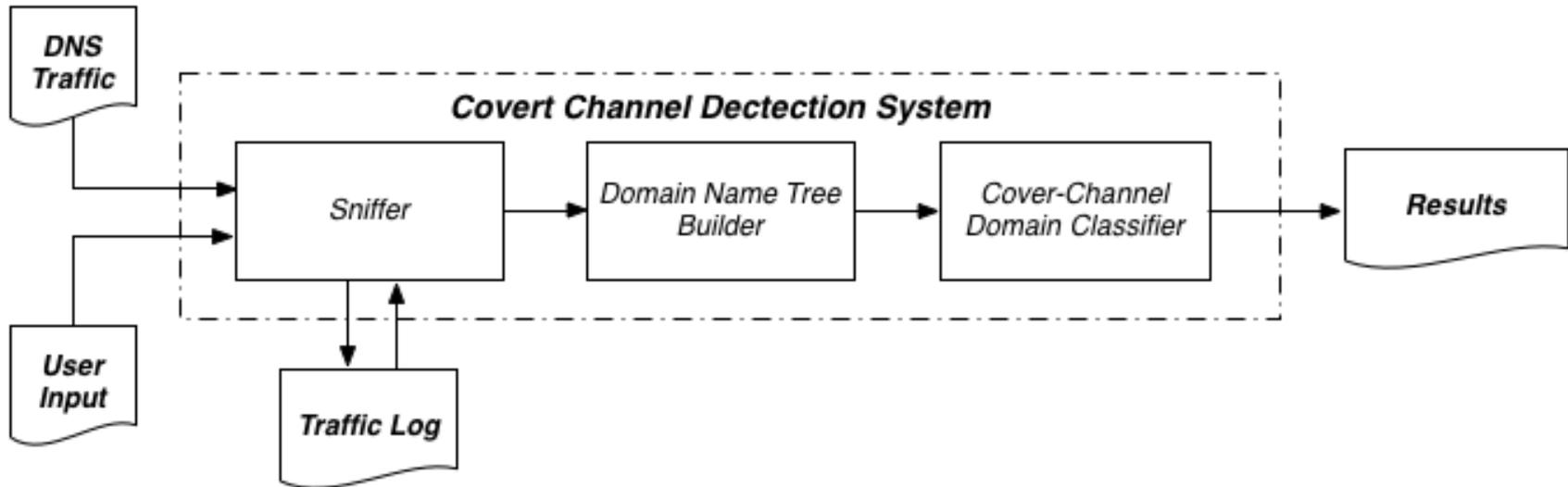
```
.....2.61020.pfwhereismyshoe.com.....
.....2.61020.pfwhereismyshoe.com.....fH...
b3GFbs+CTJ5ZjQLA3dBDaRe77I0gxM8TKVQRu+nHy1hS6Cfthyj8a0Qywy
SOTEAEAm429DygOALjlueRKw4AP1gIHppt21uJrka0AXCpPXXH+00gnRRS
iV0HB/4Ft+TzksfN0OBYKOGVH3nT8xasJ9ExZm9ZVafYrR0JTyi0WZRwS1
VdbQ3V59iWdxJz/AsgMiBZXZG/m/0ud0m/3wJXWOtuU/RB.....
.....ns1.....ns2..
```

.....

```
.\.....1.61020.pfwhereismyshoe.com.....
.\.....1.61020.pfwhereismyshoe.com.....fH...
B6KwFC+u1sjRM2+1xXKYJZTTejqxVoMpcbXXkHMUQDSolFUp21fNXbZYIc
Hpm7sL5iEbNkVFxgeq9UFHYV+DUZAmjlU0QY2Qkttl3yDVkj7qcEO60axW
Oc51JX1MB3LZrbVEEebZKoT0wbyX95q6lLE4JaAQ2jPIEWSpE2ZkP6R0xO
TOcPXTaJgoceRFp48bxFti9tyrSZwAg8evlyn4R14W3VMI.....
.....ns2.....ns1..
```

```
m.....566.61020.pfwhereismyshoe.com.....
m.....566.61020.pfwhereismyshoe.com.....g...
vbOoBv2QGtF+mZ3NgDulQVp9peX2RFJRlc+eZnfOaRKjKi2Mw5Q0Agz3AhGV
4hmUdoEqbkajGFra6mxiZ9NI4ks19J1exp35TPne9EpamrQUa/0zNVghJaYK
wSVByKQNAuHrFBZbQljGqjizO6gCFNBL+UImUszZhWi6/JgQ9iAMBAhxt8lz
2qC4TcAEiIDvXKd07LW/B/O3GUuMaHnBT/tNKqD7.....+...ns2...
.....+...ns1..
```

Detection of DNS-CC



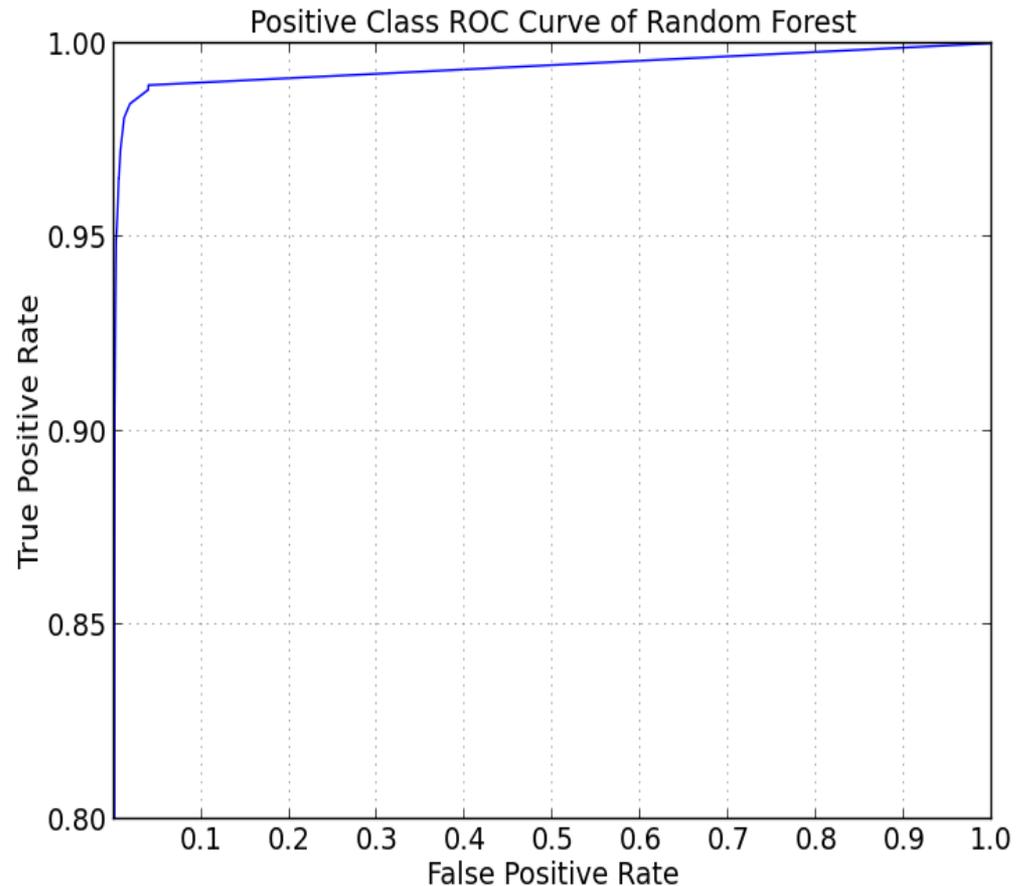
- Monitor DNS traces
- Measure statistical features according to the passive traces of each DNS zone observed
- Classify the zones as DNS covert channel candidates

The Statistical Features

- Given a group of strings s_i ($i = 1, \dots, n$), let the Shannon Entropy of characters in the string be $H(s_i)$, and construct a set with all the $H(s_i)$.
- Features: the cardinality of the set, the maximum, minimum, average, median, and variance of all $H(s_i)$ values.

Cover-Channel Domain Classifier

- Classifier: Random Forest
- True positive rate: 96.52%, false positive rate: 0.57%, using six statistical features



Real-World Evaluation

- Data
 - 6 days of Full Passive DNS data from the largest US-based ISP
 - Data format: <timestamp, requester, qname, qtype, rdata, ttl>
 - Days: 02/01 09/02 09/13 11/14 11/29 12/30 in 2011

Evaluation – Case Study

- In the news, “The Trojan is doing a *DNS TXT query* to *httpdconfig.com* every few minutes by using the current UNIX timestamp as subdomain(*unixtimestamp*.httpdsconfig.com). The C&C server replies with a encrypted string (seems to be always the same)...”
- <http://www.abuse.ch/?p=2740>

DNS Covert Channel Protocol

- If query for <timestamp>.httpxxx.xxx does not resolve, no command is replied. If it resolves, command is replied in the txt field of the DNS packet.
- 2LD is always changing
- We found a lot of DNS traffic of this C&C Covert Channel using our detection system.
- There is also another version of domains - <timestamp>.webxxx.xxx

C&C Covert Channel Zones Detected

- *.httpany.net
- *.httpbb.com
- *.httpbuyonline.net
- *.http-camerawebsoft.info
- *.httpcome.net
- *.httpdedicateserver.eu
- *.httpdigit.com
- *.httpdsconfig.com
- *.httpdump.net
- *.httpfastinternet.us
- *.httpfinalpack.us
- *.httpgo.net
- *.httphelp.us
- *.httplime.net
- *.httploading.com
- *.httplook.com
- *.httpmaindns.com
- *.httpmanagersuit.eu
- *.http-myprogramming.net
- *.httponlinewebsoft.com
- *.httppoint.net
- *.http-primetests.us
- *.httpput.net
- *.httpregion.net
- *.httprsrepoz.com
- *.httpsatellite.eu
- *.httpsdata.in
- *.httpsdsconf.com
- *.httpsea.com
- *.http-softlive.info
- *.httpsquer.com
- *.httptracking.net
- *.httpurl.net
- *.httpv1.com
- *.http-webhistorysite.us
- *.httpwebhost.eu
- *.httpwebinterface.info
- *.httpwebmail.us
- *.http-websignorg.info
- *.httpwebtech.info
- *.http-webtesting.us
- *.httpwebx.info
- *.httpwebz.info
- *.httpx4.com
- *.httpz0.com
- *.nhttpunique.net
- *.nhttpunit.net

Latest active C&C covert channel

- *.web-alm.net
- *.web-antispampc.com
- *.web-designe.us
- *.web-flashinfo.com
- *.web-freesite.com
- *.web-gamesinfo.us
- *.web-games.us
- *.web-htmlinfo.us
- *.web-mex.com
- *.web-myprice.us
- *.web-testhard.us
- *.web-testkeyboard.eu
- *.web-testprocessors.us

Status

- By the end of October 2012, we will deliver
 - A technical report describing the algorithm and features
 - The software DNS-CC Sniffer



DNS-Based Traceback

DNS-Based Traceback

- Goal:
 - Discover the origin of attacks
- Central insight:
 - The first N hosts to resolve a domain may have some connection to the domain

Approach Outline

- Only a few tens of thousands of domains are added each day to gTLDs like .com; fewer to others
- DNS monitoring tools can be extended to log IP/src of the first N-queries for a (new) qname
- Reducing false positives:
 - Multiple domains are often used by a single botnet
 - Higher confidence if first queries to several domains are from the same host/network
 - Partial take-down can force attacker into action – more monitoring/analysis opportunities

Status

- By the end of December 2012, we will deliver
 - A technical report describing the algorithm
 - The software prototype of the trace-back system

Credits

- Manos Antonakakis
- Yizhen Chen
- Jorge Villasmil