DNS Data Science

Internet Week, Tokyo Japan

Bruce Van Nice
Nominum

About Nominum

- Paul Mockapetris
- Engineers architected BIND 9
- Deployed in over 40 countries
- Used by >400M subscribers daily
- Processes >1.8 trillion transactions daily

About Me

- Director of Product Marketing at Nominum
- Focus on DNS security
- Active participant at industry events
- About 30 years in networking industry

Why Do DNS Data Research?

- Existing threat analysis isn’t good enough
  - No coverage of DNS DDoS
  - Poor coverage of bots and malware
- Existing feeds have limitations:
  - Unacceptably high false positive rates
  - Unacceptable for providers
- Combining feeds didn't help

BAD + BAD ≠ GOOD!

DNS Data Science

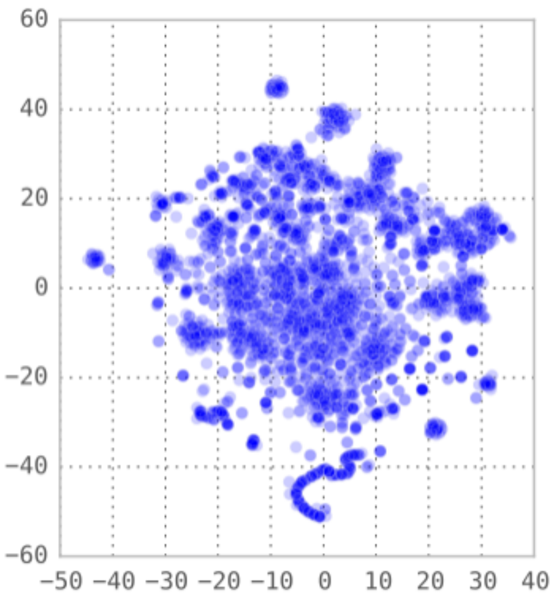
Discovering patterns and relationships

- > 3.5 Terabytes/data/day
- Est ~3% of ISP DNS traffic

Data is anonymized to protect privacy

Correlation Technology

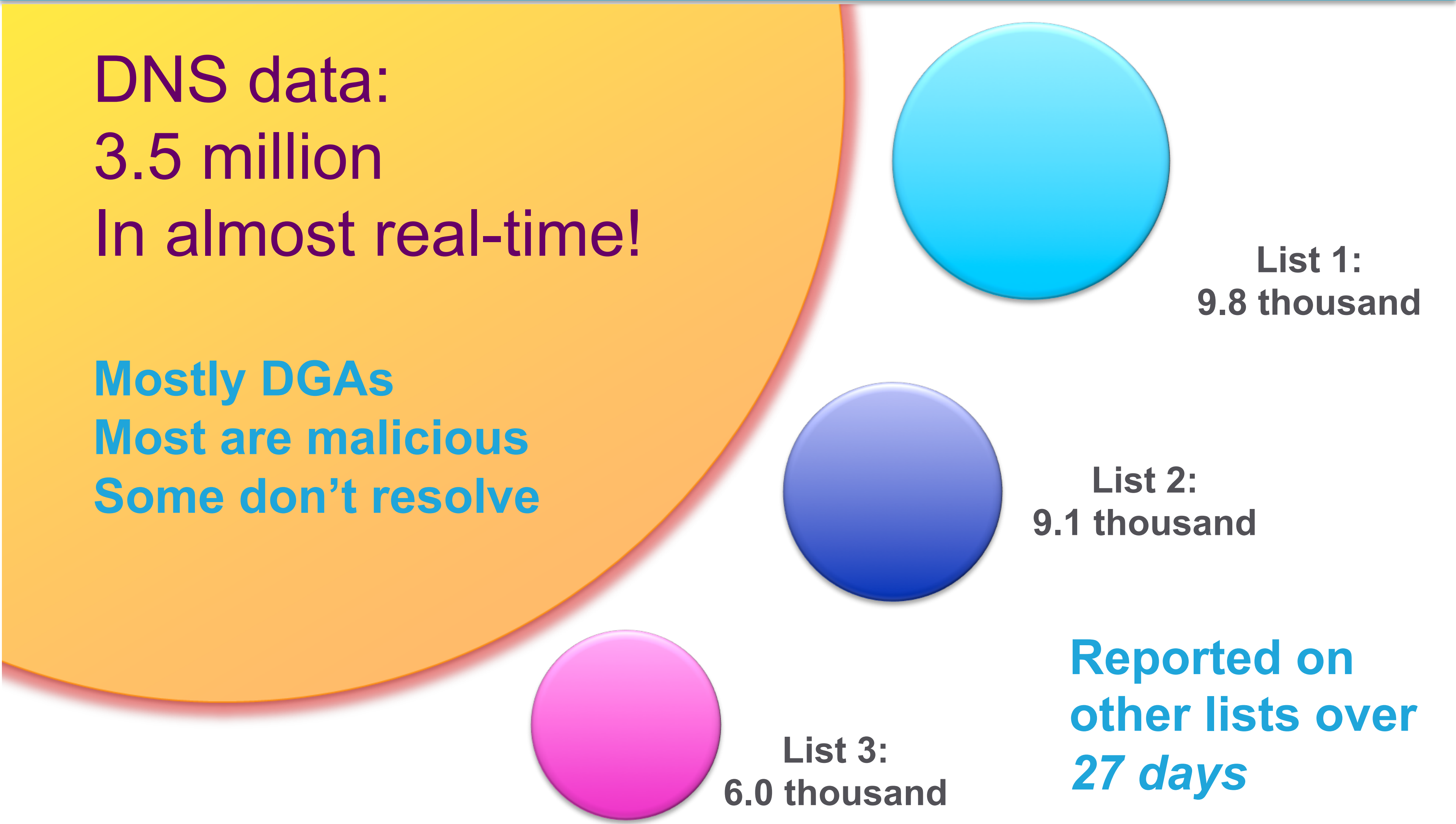
- Machine learning - search for similarities
  - sfgiants.com google.com redsox.com mlb.com facebook.com…
  - botnet_cnc1.com google.com botnet_cnc2.biz facebook.com yahoo.com botnet_cnc3.ru
- Visualization groups similar domains together



Correlation technology finds more malware faster

Uncovers other relationships - different apps infected with same malware

Example: New Core Domains



DNS data:
3.5 million
In almost real-time!

Mostly DGAs
Most are malicious
Some don't resolve

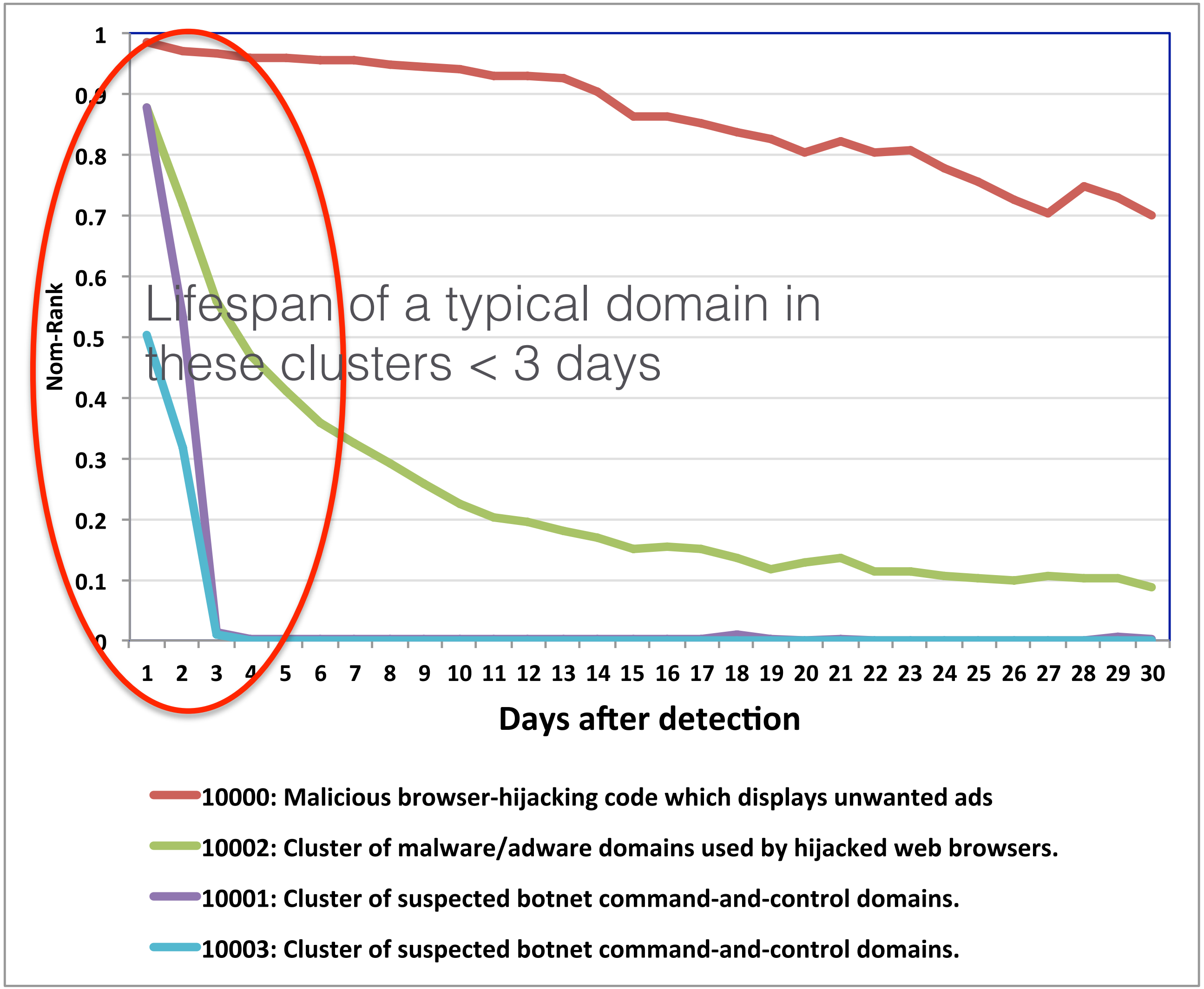
List 1:
9.8 thousand

List 2:
9.1 thousand

List 3:
6.0 thousand

Reported on other lists over *27 days*

Example: "Clusters Detected"



Lifespan of a typical domain in these clusters < 3 days

- 10000: Malicious browser-hijacking code which displays unwanted ads
- 10002: Cluster of malware/adware domains used by hijacked web browsers.
- 10001: Cluster of suspected botnet command-and-control domains.
- 10003: Cluster of suspected botnet command-and-control domains.

Example: Browser Hijacking

- Malware patterns with 1% to 20% infection rate depending on network
- Browser hijacking + unknown malicious activity
- Actively querying CNC multiple times/day
- Algorithm identifies infected PCs
  - CNC traffic can be blocked or monitored
- These domains appear on threat lists days after they are dormant

Example Domains

acapulcosonars.com.
enticingsuperpower.com.
envelopspunnet.com.
evocationsmotliest.com.
heronsquadrupled.com.
infernalbrazing.com.
joininguncoils.com.
mumbleinterim.com.
pancakeskennels.com.
pesterlipid.com.

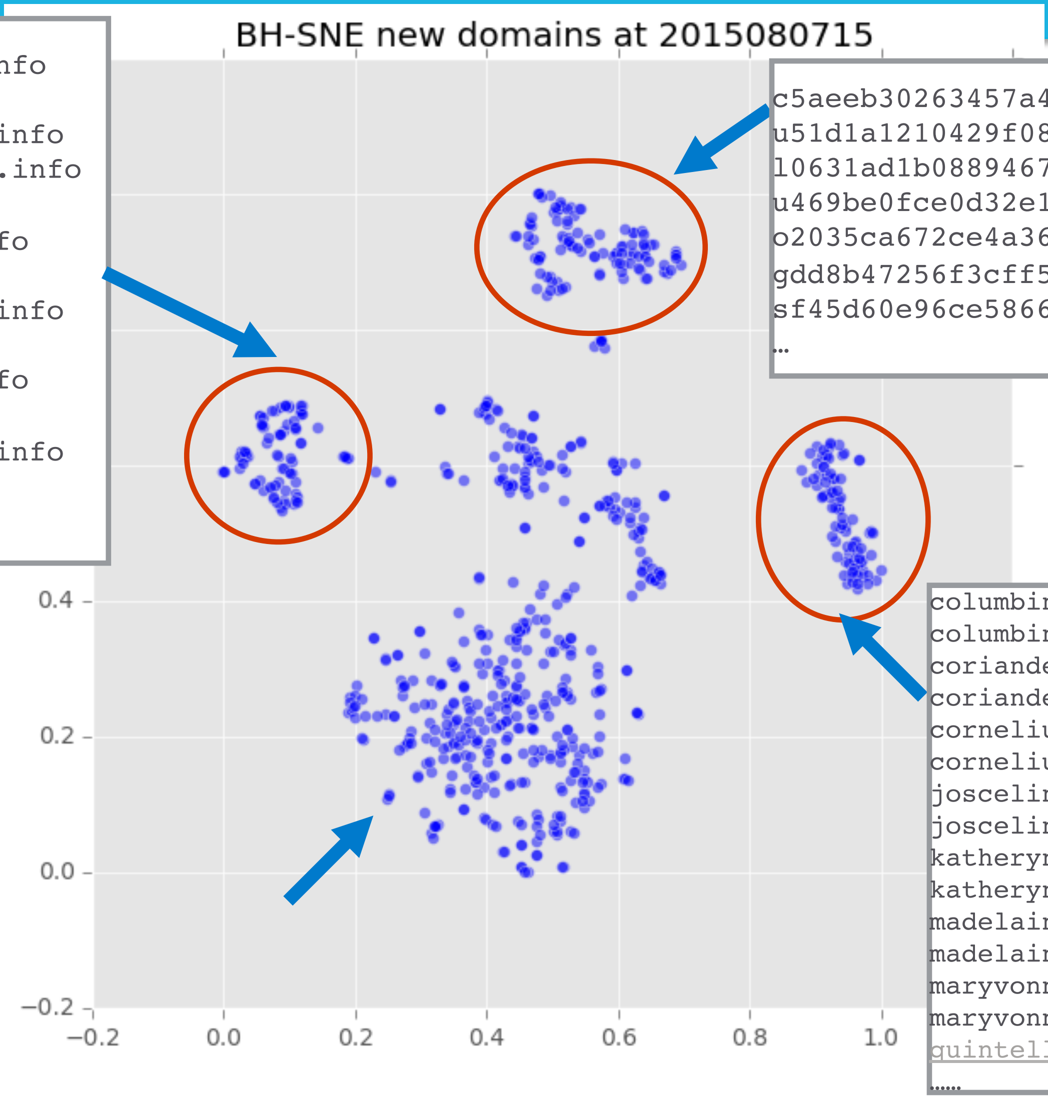
Example: Ad Trackers

- Websites use many ad trackers
- Seed with known ad trackers
- Look for correlated query patterns

Example Domains

2082.info.
6485.info.
7228.info.
daap.info.
drab.info.
fiiy.info.
flob.info.

Example: Botnet Clusters

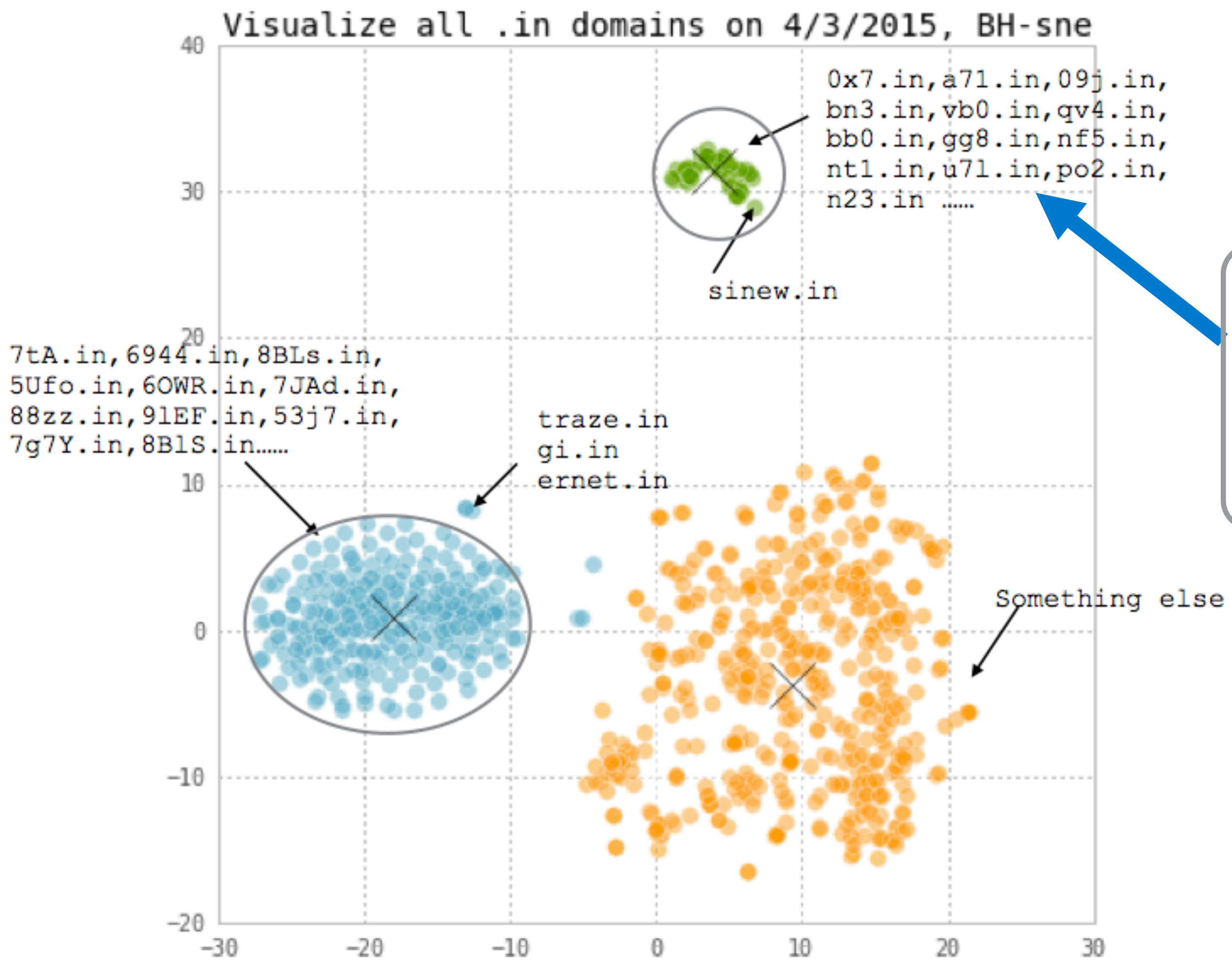


shjldview.info
zyuvl.info
rpcbfgrxcs.info
oxhgeuxxkit.info
oftguz.info
fatgdctn.info
zvnvlr.info
paougcwdvg.info
pxihq.info
fbbvzswf.info
nwvomw.info
clwkxxfoum.info
kthmu.info
……

c5aeeb30263457a4fd91a6bb9abad118f5.cc
u51d1a1210429f081f18a8b7be709d874e.tk
l0631ad1b088946744d0ee6633ee49ab60.so
u469be0fce0d32e1afa59ad8f3541fed47.tk
o2035ca672ce4a36df1b8e1537e6089791.cc
gdd8b47256f3cff5133a72de1fde331109.tk
sf45d60e96ce5866098841b1b1f6fbaea8.tk
…

columbinechristophers.net
columbinemiddlesworth.net
corianderchristophers.net
coriandermiddlesworth.net
corneliuschristophers.net
corneliusmiddlesworth.net
joscelinechristophers.net
joscelinemiddlesworth.net
katherynechristophers.net
katherynemiddlesworth.net
madelainechristophers.net
madelainemiddlesworth.net
maryvonnechristophers.net
maryvonnemiddlesworth.net
quintellachristophers.net
……

Example: DNS Tunneling



cross-checked and confirmed with the tunneling detection algorithm

Summary

- DNS data reveals many interesting things!
- DDoS
- Bots/Malware
- Adware
- Tunnels

- There will be *many* more insights

Thank You!