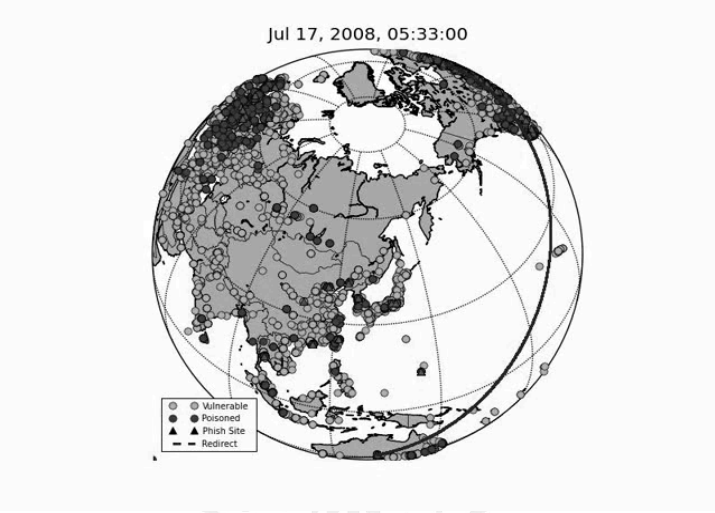
Internet Monitoring via DNS Traffic Analysis

Wenke Lee
Georgia Institute of Technology

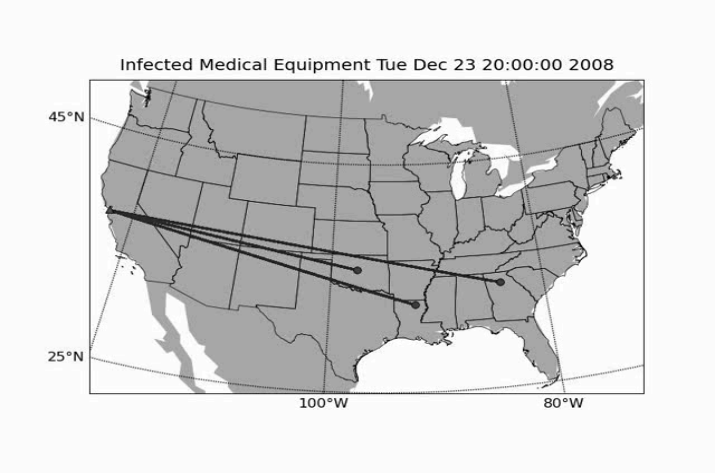
0

Malware Networks (Botnets)



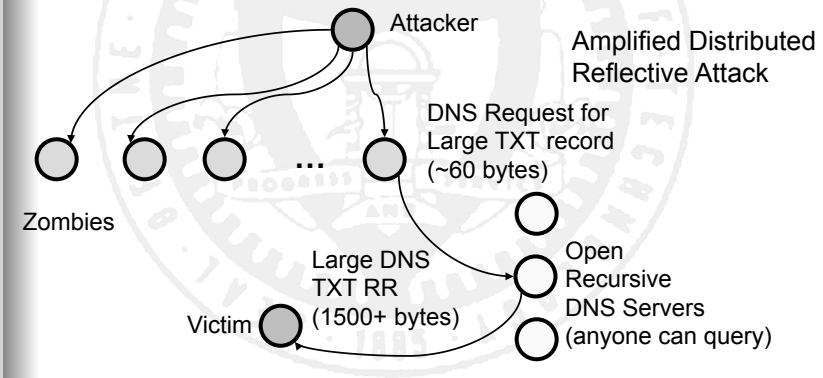
1

From General-Purpose to Targeted Attacks



Another Attack Example

- Botnets increasingly used for amplified distributed reflective attacks

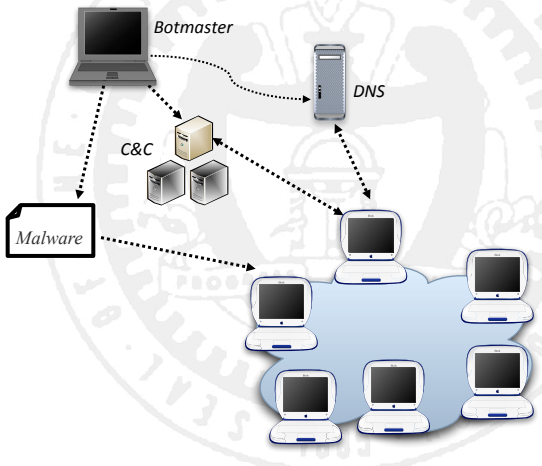


Command and Control

- Botnet design:
  - C&C is essential to a botnet
    - Without C&C, bots are just discrete, unorganized infections
  - Goal: robustness, or, no single point of failure
    - Mobility: Command and Control (C&C) can migrate to other networks
    - Stealth: difficult to detect

4

Use of DNS by Malicious Networks



5

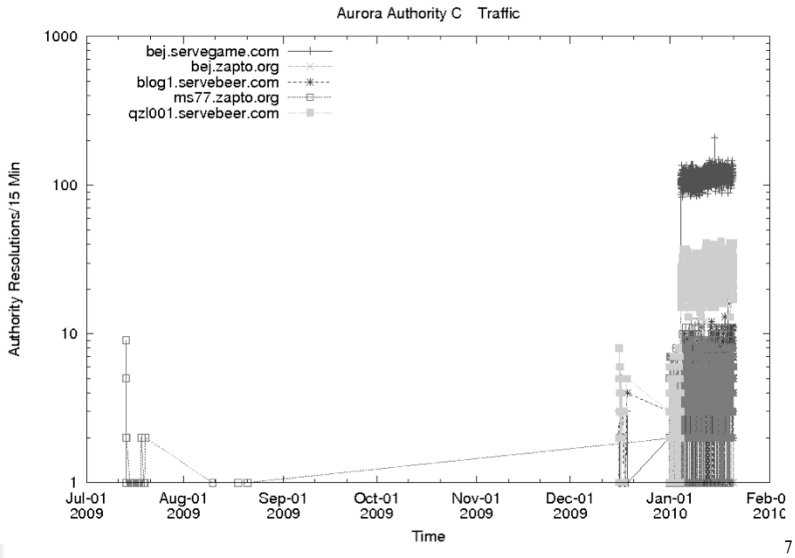
Example: Aurora

- Widely reported to be December 2009/ Jan. 2010 attack on Google
- Numerous C&C domains

Domain	Authority Creation Date (UTC)
bej.servegame.com	Dec. 15, 2009 21:26:22
bej.zapto.org	Dec. 15, 2009 21:25:48
ms77.zapto.org	July 13, 2009
qzl001.servebeer.com	Dec. 15, 2009 20:07:40
blog1.servebeer.com	Dec. 15, 2009: 21:28:55

6

Aurora C&C Resolutions



7

Dynamic DNS Reputation Across the DNS Hierarchy

8

Overview

- Motivation
  - Static DNSBL increasingly ineffective
  - Need a dynamic, comprehensive reputation system outputs reputation scores for domains
- Intuitions
  - Legitimate uses of domains/sites are different from botnet uses, and the differences can be observed in DNS query traffic
    - Patterns/reputation of Requesters, Resolved IPs, Network providers
- Approach
  - Extract temporal and statistical features from DNS traffic, compute/learn models

9

Notos

- Network and zone based features that capture the characteristics of resource provisioning, usages, and management of DNS domains
- Models of legitimate and malicious domains for computing reputation scores for new domains
- Accuracy: can correctly classify new domains with a very low FP% (0.3846%) and high TP% (96.8%)
- Predictability: able to detect and assign a low reputation score to fraudulent domain names, several days or even weeks before they appear on static blacklists
- 2010 USENIX Security Symposium

10

Kopis

- Passive monitoring in the upper levels of the DNS hierarchy; Internet-wide visibility
- Analyzes streams of DNS queries and responses at AuthNS or TLD servers, and extracts a set of statistical features and trains a model
- Accuracy: high detection rates (98.4%) and low false positive rates (0.3%)
- Predictability: able to identify newly created and previously unclassified malicious domain names weeks before they were listed in any blacklist
- Detected a DDoS botnet rising in networks within China almost one month before it propagated within other countries
- 2011 USENIX Security Symposium

11

Notation & Terminology

- Resource Record (RR)
  - www.example.com 192.0.32.10
- 2nd level domain (2LD) and 3rd level domain (3LD)
  - For the domain name www.example.com: 2LD is the example.com and 3LD is the www.example.com
- Related Historic IPs (RHIPs)
  - All "routable" IPs historically mapped with the domain name in the RR or any domain name under the 2LD and 3LD
- Related Historic Domains (RHDNs)
  - All fully qualified domain names (FQDN) that historically have been linked with the IP in the RR, its corresponding CIDR and AS
- Authoritative domain name tuple
  - The requester (or RDNS), the domain name and the RDATA

12

Passive DNS and Authoritative Data

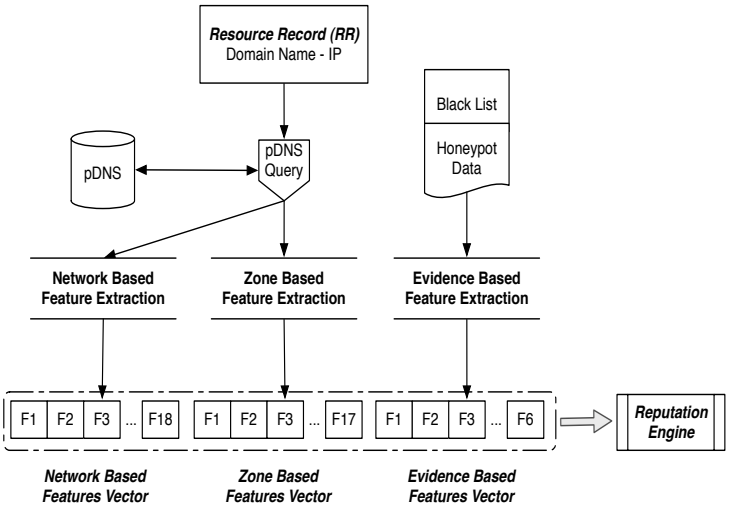
- Passive DNS (pDNS) data collection is the harvesting of successful DNS resolutions that can be observed in a given network
- Passive DNS database contain traffic from several ISP sensors and SIE
  - Observed that different classes of zones demonstrate different passive DNS behaviors
- Obtained authoritative DNS traffic from 2 large authoritative DNS servers (AuthNS) and the Canadian TLD

13

Notos

14

Statistical Features of Notos

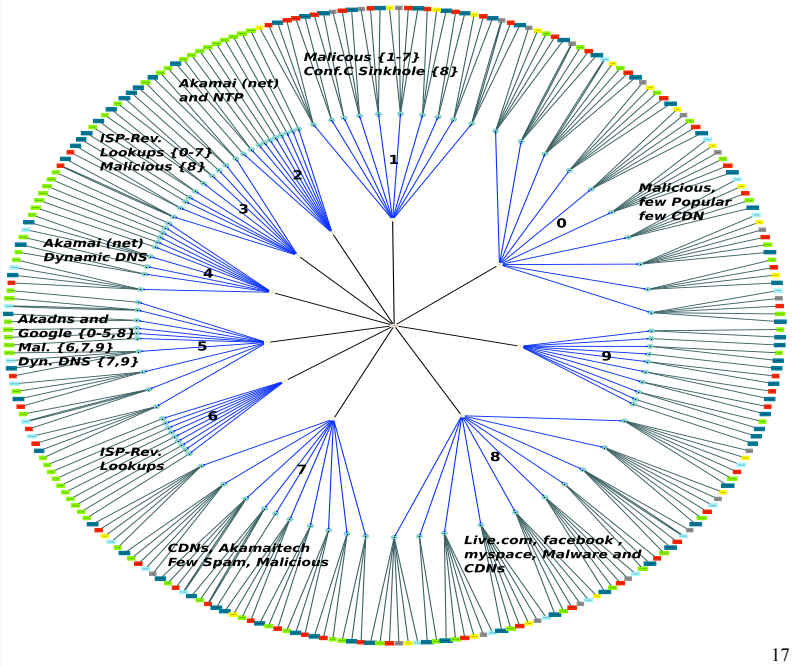


15

Statistical Features of Notos

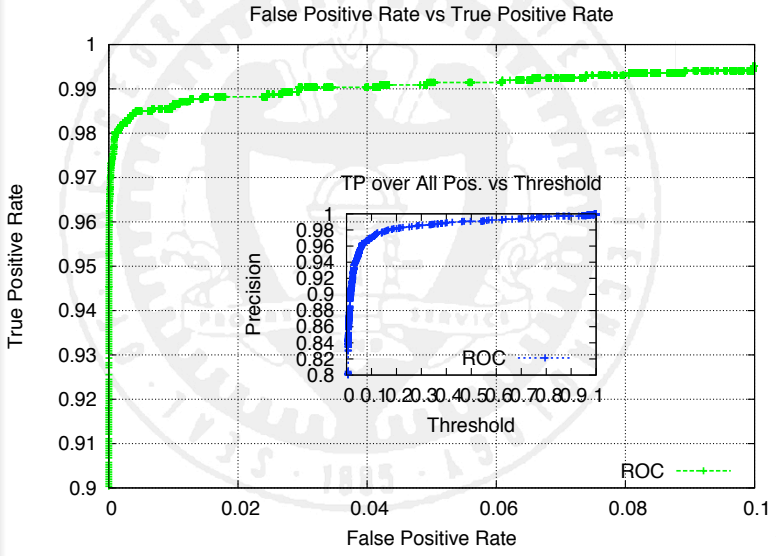
- Network-Based Features:
  - Extracted from the set RHIPs
  - E.g., the total number of IPs historically associated with a domain, the diversity of their geographical location, the number of distinct autonomous systems (ASs) in which they reside, etc.
- Zone-Based Features:
  - Extracted from the set RHDNs.
  - E.g., the average length of domain names in RHDNs, the number of distinct TLDs, the occurrence frequency of different characters, etc.
- Evidence-Based Features:
  - E.g., the number of distinct malware samples that contacted the domain, and the same for any of the resolved IPs, etc.

16



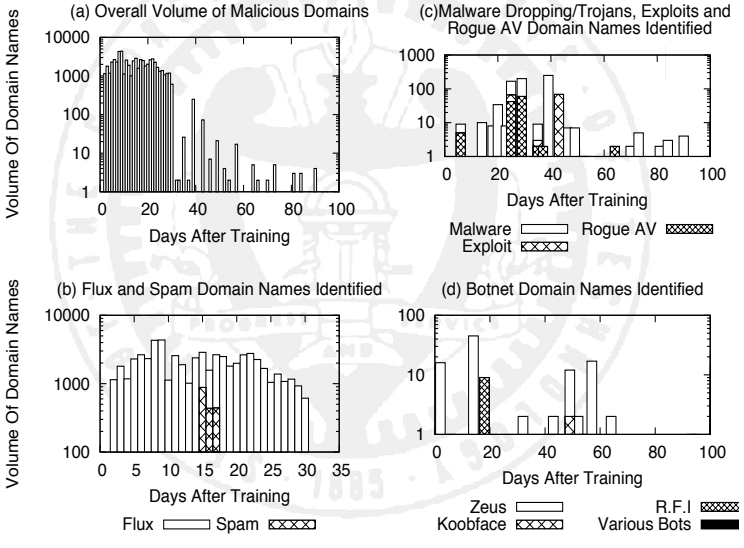
17

Notos' Reputation Function



18

Notos' Reputation Function

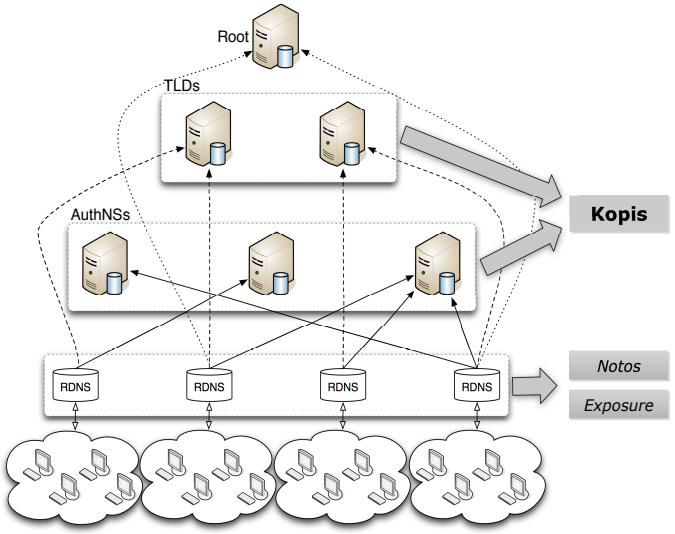


19

Kopis

20

Authoritative vs Recursive Monitoring



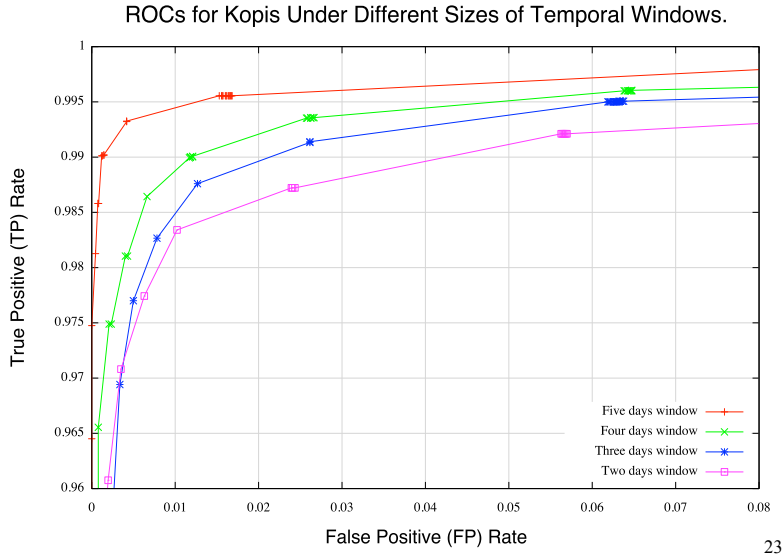
21

Statistical Features of Kopis

- Requester Diversity (RD)
  - Characterize if the machines (e.g., RDNS servers) that query a given domain name are localized or are globally distributed (based on BGP prefixes, AS numbers, country codes, etc.)
- Requester Profile (RP)
  - Distinguish between requesters located in ISP/small business and home networks
  - Assign a higher weight to RDNS servers that serve a large client population because a larger network would have a larger number of infected machines.
- Resolved-IPs Reputation (IPR)
  - Whether, and to what extent, the IP address space pointed to by a given domain has been historically linked with known malicious activities, or known legitimate services

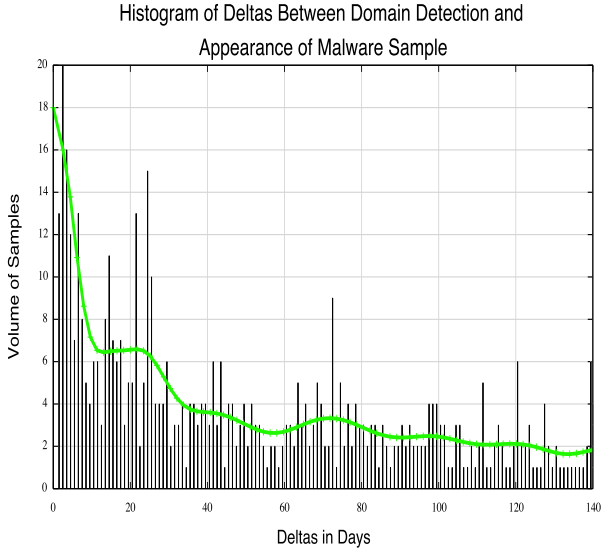
22

Kopis Detection Performance



23

Kopis Predictability



24

Analyzing Mobile DNS Traffic

25

Overview

- Motivation
  - Much work on mobile malware has been on analysis of (malicious) mobile apps
  - But, how prevalence are infections on mobile devices?
- Intuitions
  - The (malicious) mobile web is a part of the (malicious) web
  - Mobile malware uses similar infrastructure (C&C) techniques as non-mobile/Internet malware
- Approach
  - Obtain DNS traffic in cellular network and identify domains looked up by mobile apps
  - Analyze information related the Internet hosts pointed by these domains

26

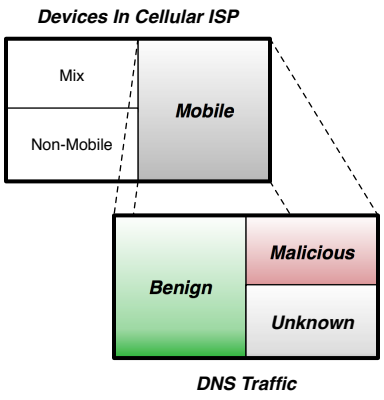
Data and Key Findings

- Three months of data from a major US cellular provider and a major US non-cellular ISP
- Known mobile malware samples are virtually unseen: only 6,585 out of 380,537,128 devices, or 0.002%
- iOS vs. Android and other devices: equally likely to connect to suspicious domains
- To appear in NDSS 2013

27

Methodology

- Identify mobile devices and attribute each DNS query to a device
- Analyze reputation of the RRs associated with the DNS queries



28

Reputation Analysis

- Use Notos to analyze the hosting infrastructures of the mobile domains
  - Obtain the host IPs pointed to by the mobile domains, for each IP, extract statistical features of
    - Related historic non-cellular domains
    - Related historic mobile domains
    - Malware association
    - URLs for phising and drive-by download
    - Blacklisting incidents

29

Tainted Hosts and Platforms

Device platform	% Total Requests by mobile device	% Population requesting tainted hosts	% Total tainted host requests
iOS	31.6%	8.8%	33.2%
All others (Android, etc.)	68.4%	8.2%	66.8%

30

Mobile Malware Prevalence

Malware Family	# Associated Domains	# Devices
DroidDreamLight	3	44
DroidKungFu	1	6
FakeDoc	1	2145
Fatakr	1	151
GGTrackers	3	1
NotCompatible	3	762
Planton	4	286
Malware β	1	1
WalkInWat	1	95
Gone60	1	1

31

Conclusion

- Malware networks rely on core Internet services, e.g., DNS, to maintain command-and-control infrastructures
  - Mobile or non-mobile
  - DNS traffic analysis can be used to identify malicious domains and infected devices
- With historical information, dynamic DNS reputation systems can even predict the maliciousness of a new domain resolution

32

Credits

- David Dagon
- Manos Antonakakis
- Roberto Perdisci
- Yacin Nadji
- Yizhen Chen
- Charles Lever
- Brad Reaves

- *Nick Feamster*
- *Patrick Traynor*

33

Government and Industry Support

- NSF, ARO, DARPA, DHS, ONR, AFRL, and IARPA
- Damballa, Equifax, Intel, Sandia, Symantec, Google, Microsoft, HP, and VeriSign

34