Building a Dynamic Reputation System for DNS

Donika Mirdita

Department of Computer Science
Technical University Munich

Seminar: Large-Scale Malware Analysis
Outline

1. Problem Statement

2. Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3. Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
DNS and Malicious Domains

- DNS resolves domain names into IP addresses.
- Botnets, Spyware, Fast-flux networks etc. take advantage of DNS agility.
Not as inconspicuous as they think

Typical malware behaviour includes...

- randomly generated domain names
- domains that point at *too many* IPs
- unusual utilization of network resources
- "incriminating" DNS history
- failure to comply with DNS RFCs
Static Solution: Blacklists

Not good enough

- DNS distributes DNSBLs (DNS-based Block Lists)
- Publicly available blacklists
- ... delay between creation of malicious domain and blacklisting
Dynamic Solution: Notos

- Takes advantage of typical malware behaviour
- Dynamically assigns reputation to new domains in real time
Outline

1. Problem Statement

2. Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3. Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
Problem Statement
Notos: A Dynamic Reputation System
Performance and Observations

Data Collection Infrastructure

Notos

ISP Recursive DNS Server (Atlanta)
ISP Recursive DNS Server (SJC)

Internet

Subnet
Subnet
Subnet

Passive DNS Database
Reputation Engine

S.I.E

Dynamic Reputation Scores

Donika Mirdita
Building a Dynamic Reputation System for DNS
Set of domain names: \( D = \{d_1, d_2, \ldots, d_m\} \)

Set of addresses: \( A(D) = \{\text{IPs pointed by } d \mid \forall d \in D\} \)

Set of IPs within a BGP prefix:
\( BGP(A) = \{\bigcup_{k=1}^{m} BGP(a_k)\} \)

Set of IPs within an AS prefix:
\( AS(A) = \{\bigcup_{k=1}^{m} AS(a_k)\} \)
For a new domain $d$, find $A(d)$ then query pDNS for:

1. **Related Historic IPs (RHIPs):** $A(d) \cup A_{3LD}(d) \cup A_{2LD}(d)$

2. **Related Historic Domains (RHDNs):** all domains where $A(d_i) \cap AS(A(d)) \neq \emptyset$
Outline

1. Problem Statement

2. Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
     - Off-Line Mode
     - Online Mode

3. Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
Overview

Building a Dynamic Reputation System for DNS
Feature Categories

1. Network-based Features
2. Zone-based Features
3. Evidence-based Features
Reputation Engine

![Reputation Engine Diagram]

Off-Line "Training" Mode:
- Train Reputation Function
- Network & Zone Clustering
- Network Profile Modeling
- Passive DNS DB
- Compute Vectors
- d

On-Line Mode:
- Network & Zone Assignment
- Network Profile Classification
- Reputation Assignment
- Dynamic Reputation Rating
- f(d)

New RR
- DC(d)
- NM(d)
- EV(d)
- EV(d')
- DC(d')
- NM(d')
Outline

1. Problem Statement

2. Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3. Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
Off-Line Training Mode I

1. **Network Profiles Model:** properties of benign networks . . .
   1. Popular Domains
   2. Common Domains
   3. Akamai Domains (CDN)
   4. CDN Domains
   5. Dynamic DNS Domains
Off-Line Training Mode II

2 Domain Name Clustering:
1. Network-based Clustering
2. Zone-based Clustering
The Reputation Function

1. The reputation function is a statistical classifier.
2. Labelled dataset $L = \{(v(d_i), y_i)\}$ for $d_i \in$ Knowledge Base and $y_i = 0$ if malicious, 1 otherwise.
The Reputation Function cont.

Ground Truth for malware:

1. public blacklists for malicious domains
2. Sender Policy Block (SBL) from Spamhaus
3. Zeus tracker
The Reputation Function cont.

Ground Truth for benign domains and networks:

1. top 500 alexa.com domains
2. 18 most common 2LDs for various CDNs
3. 464 dynamic DNS 2LDs
Outline

1 Problem Statement

2 Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3 Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
The Reputation Engine


domain name \( d \)

- Evidence Features
- Zone Features
- Network Features

- Domain Clustering Module
- Network Profiling Module

- \( EV(d) \) → \( DC(d) \) → \( u(d) \) → Reputation Function \( f(d) \)

Donika Mirdita
Building a Dynamic Reputation System for DNS
On-Line Mode

1. Assigns reputation scores $S$ to new domains
2. $S \in [0, 1]$ where $S = 1 - f(d)$
Outline

1. Problem Statement

2. Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3. Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
Data Statistics

- 27,377,461 unique resolutions collected in 68 days (July-September 2009)
- SIE collected a volume of 200 Mbit/s resolutions
- ISP DNS Servers processed 30’000 requests /s during peak hours
Performance Overview 1

- overall TP 96.8% and FP 0.38%
computing 100k networks and a 15-days pDNS DB enough for a stable number of clusters
previously unknown Zeus botnets accurately detected
Outline

1 Problem Statement

2 Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3 Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
Where Notos excels

- Identification of malicious domains weeks/months before official blacklisting
- High TP rate 98.6% and low FP rate of 0.38%
- Good scalability
- Equally good performance even with a smaller pDNS DB
Where it falls short

1. bad neighbourhoods
2. will not operate as well once IPv6 becomes main protocol
3. requires a large pDNS DB and training time
4. not ideal as a standalone defence system
Outline

1. Problem Statement

2. Notos: A Dynamic Reputation System
   - System Overview
   - Features Extraction
   - Off-Line Mode
   - Online Mode

3. Performance and Observations
   - Data and Performance
   - The Good, The Bad and ...
   - Alternatives
Alternative Dynamic Solutions

1. SNARE
2. Spamsscatter
3. EXPOSURE
Sources I


Antonakakis M., Perdisci R., Dagon D., Lee W., Feamster N.
Building a Dynamic Reputation System for DNS.