Building a Dynamic Reputation System for DNS

Donika Mirdita

Department of Computer Science
Technical University Munich

Seminar: Large-Scale Malware Analysis

Outline

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

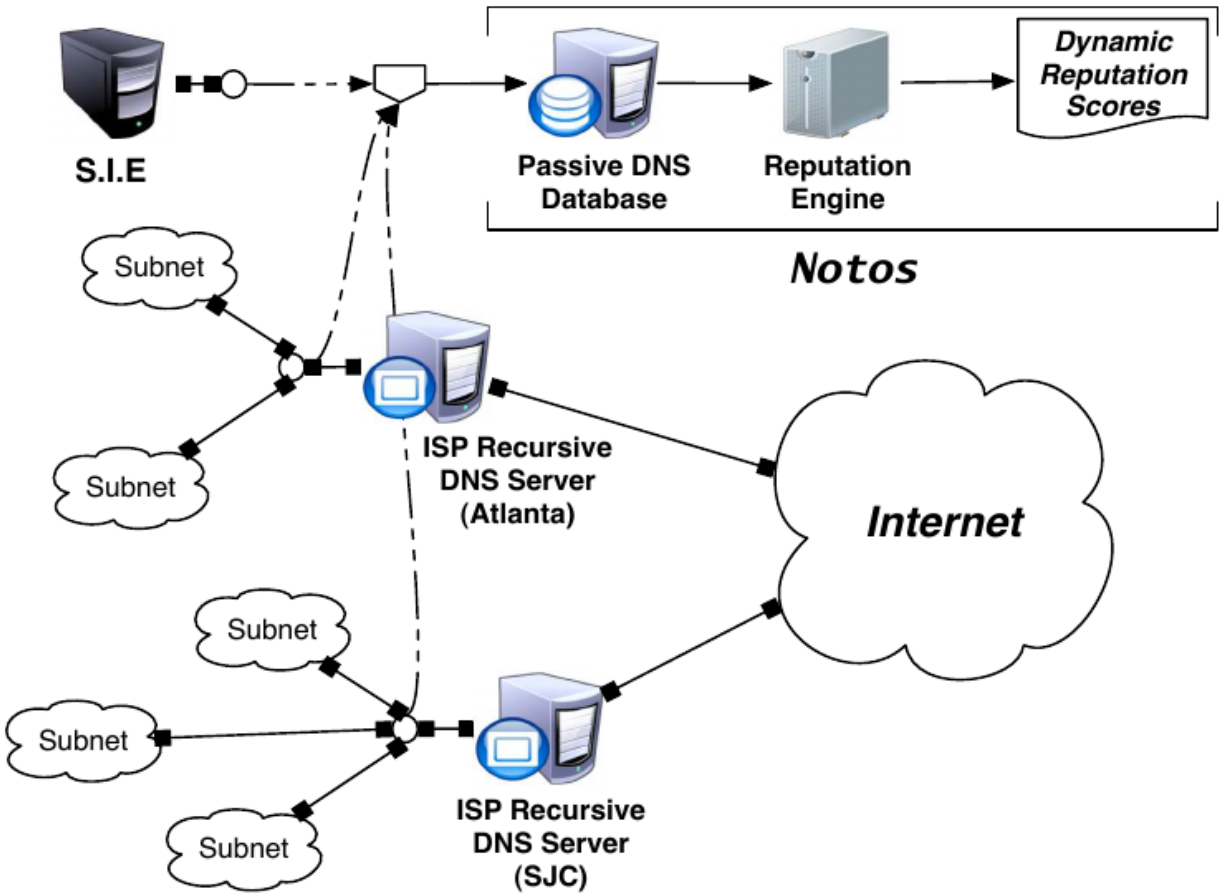
- DNS distributes DNSBLs (DNS-based Block Lists)
- Publicly available blacklists
- ... delay between creation of malicious domain and blacklisting

Dynamic Solution: Notos

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

Outline

Data Collection Infrastructure



pDNS Content

1. Set of domain names: $D = \{d_1, d_2, \ldots, d_m\}$
2. Set of addresses: $A(D) = \{\text{IPs pointed by d } | \forall d \in D\}$
3. Set of IPs within a BGP prefix: $BGP(A) = \{\bigcup_{k=1\ldots m} BGP(a_k)\}$
4. Set of IPs within an AS prefix: $AS(A) = \{\bigcup_{k=1\ldots m} AS(a_k)\}$

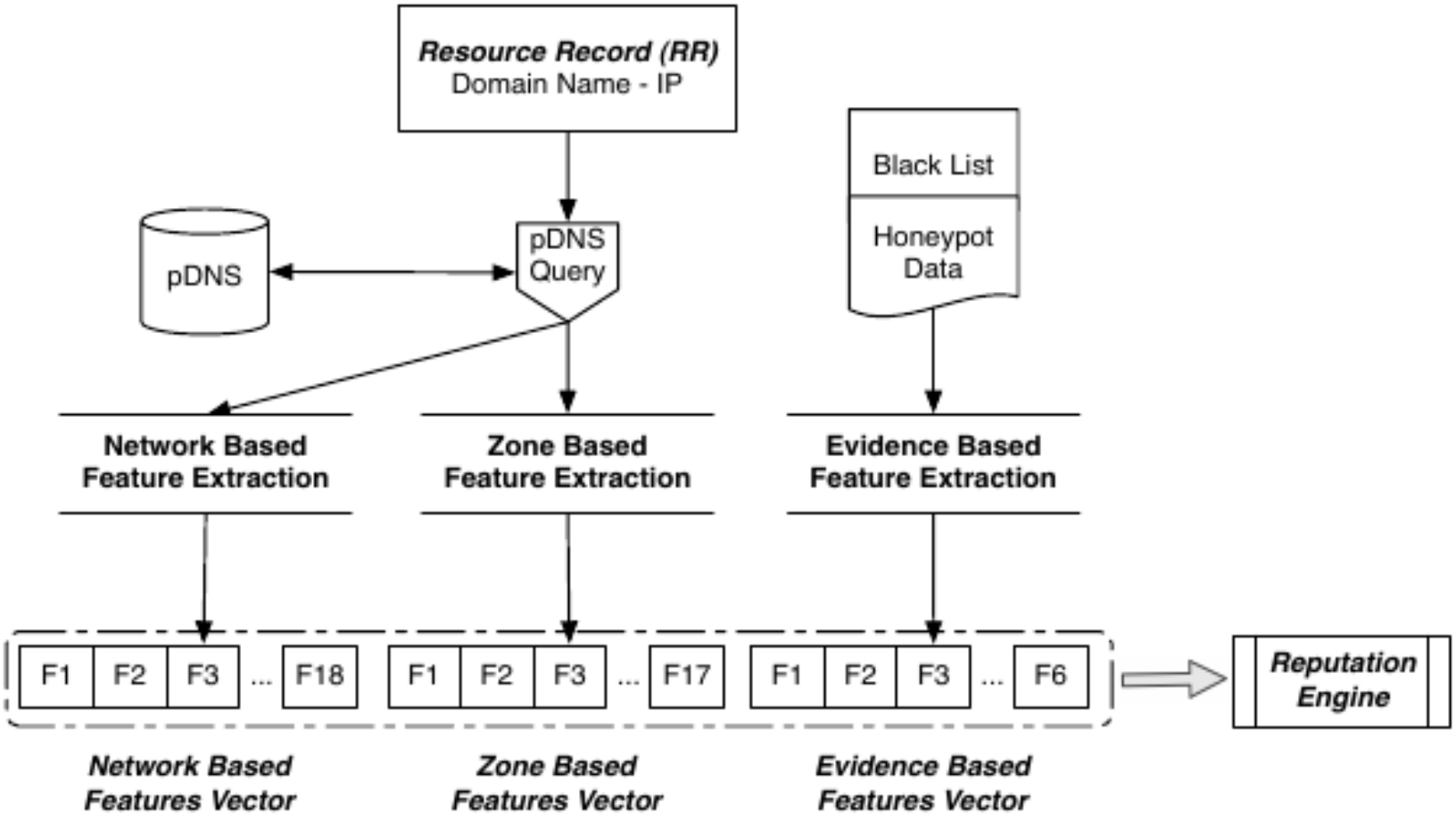
pDNS Query

For a new domain d, find $A(d)$ then query pDNS for:

1. *Related Historic IPs (RHIPs)*: $A(d) \bigcup A_{3LD}(d) \bigcup A_{2LD}(d)$
2. *Related Historic Domains (RHDNs)*: all domains where $A(d_i) \cap AS(A(d)) \neq \emptyset$

Outline

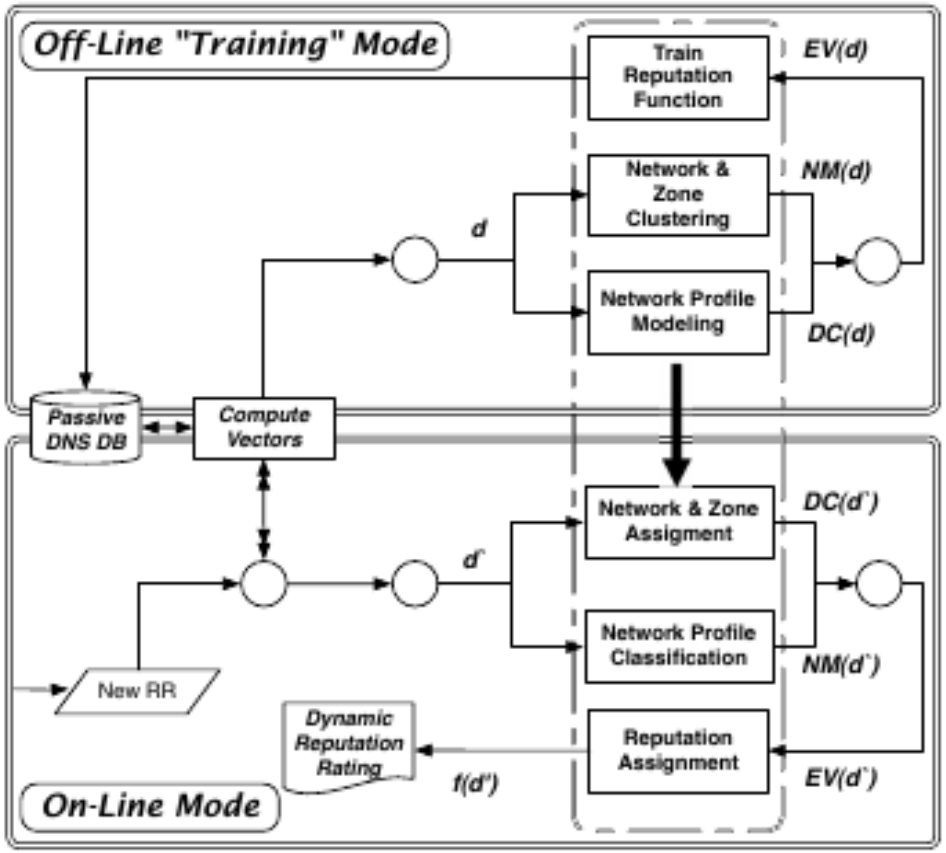
Overview


Resource Record (RR)
Domain Name - IP
pDNS
pDNS Query
Black List
Honeypot Data
Network Based Feature Extraction
Zone Based Feature Extraction
Evidence Based Feature Extraction
F1 F2 F3 ... F18
F1 F2 F3 ... F17
F1 F2 F3 ... F6
Reputation Engine
Network Based Features Vector
Zone Based Features Vector
Evidence Based Features Vector

Feature Categories

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

Reputation Engine



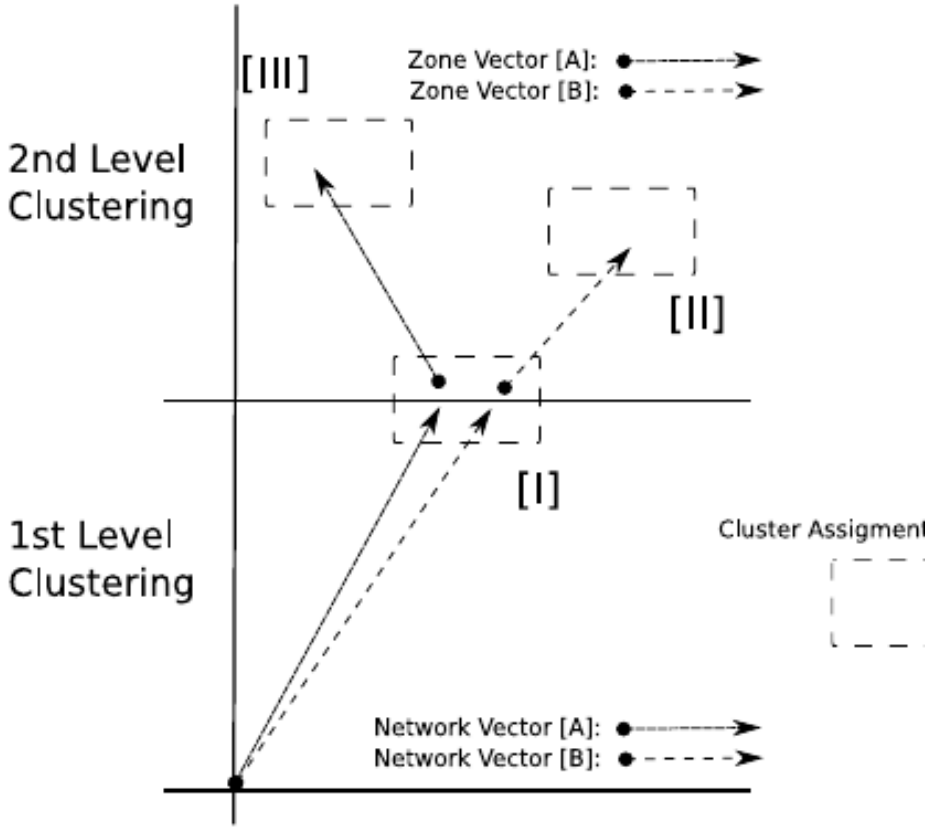
Outline

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

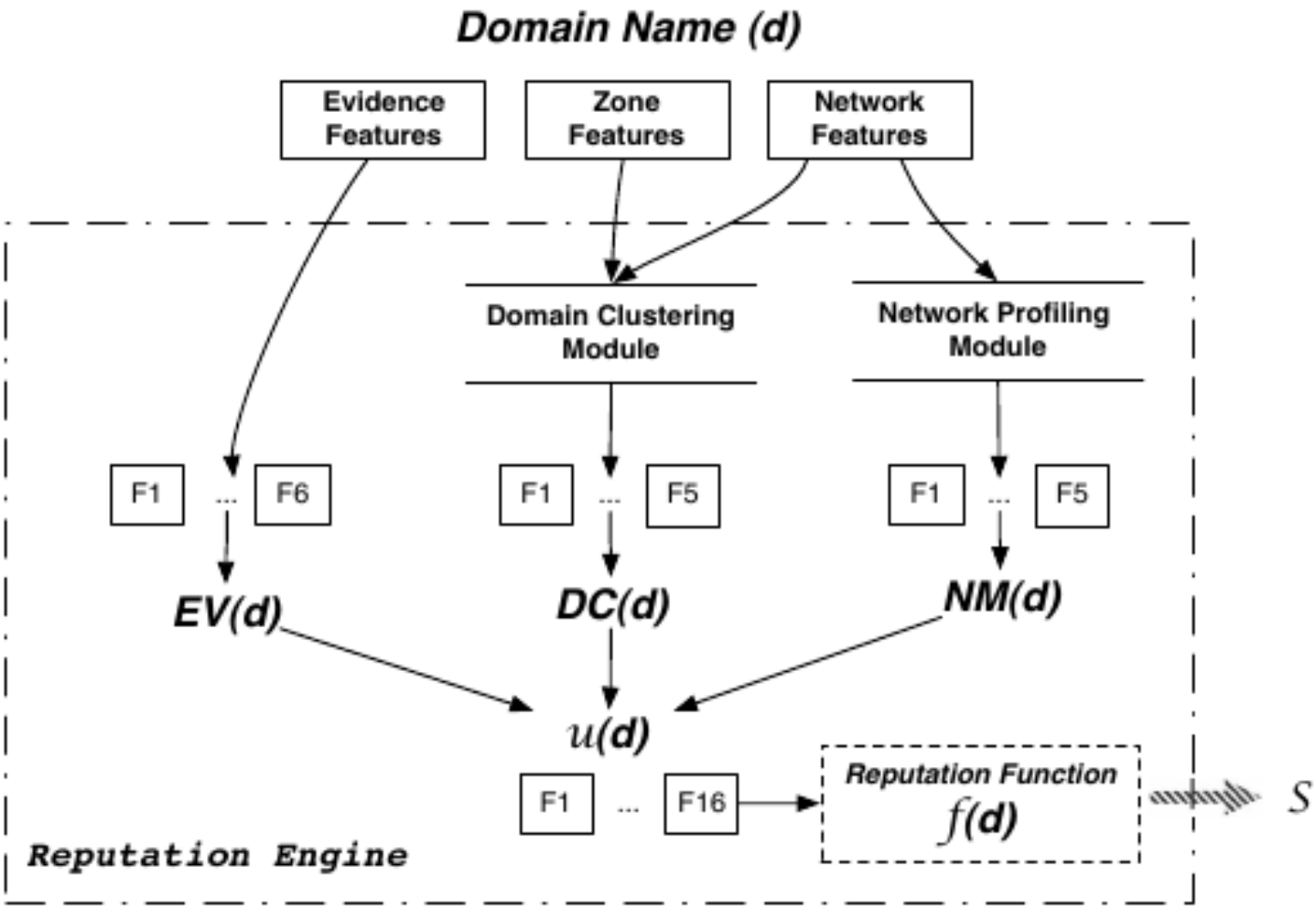
The Reputation Function cont.

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

The Reputation Engine



On-Line Mode

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

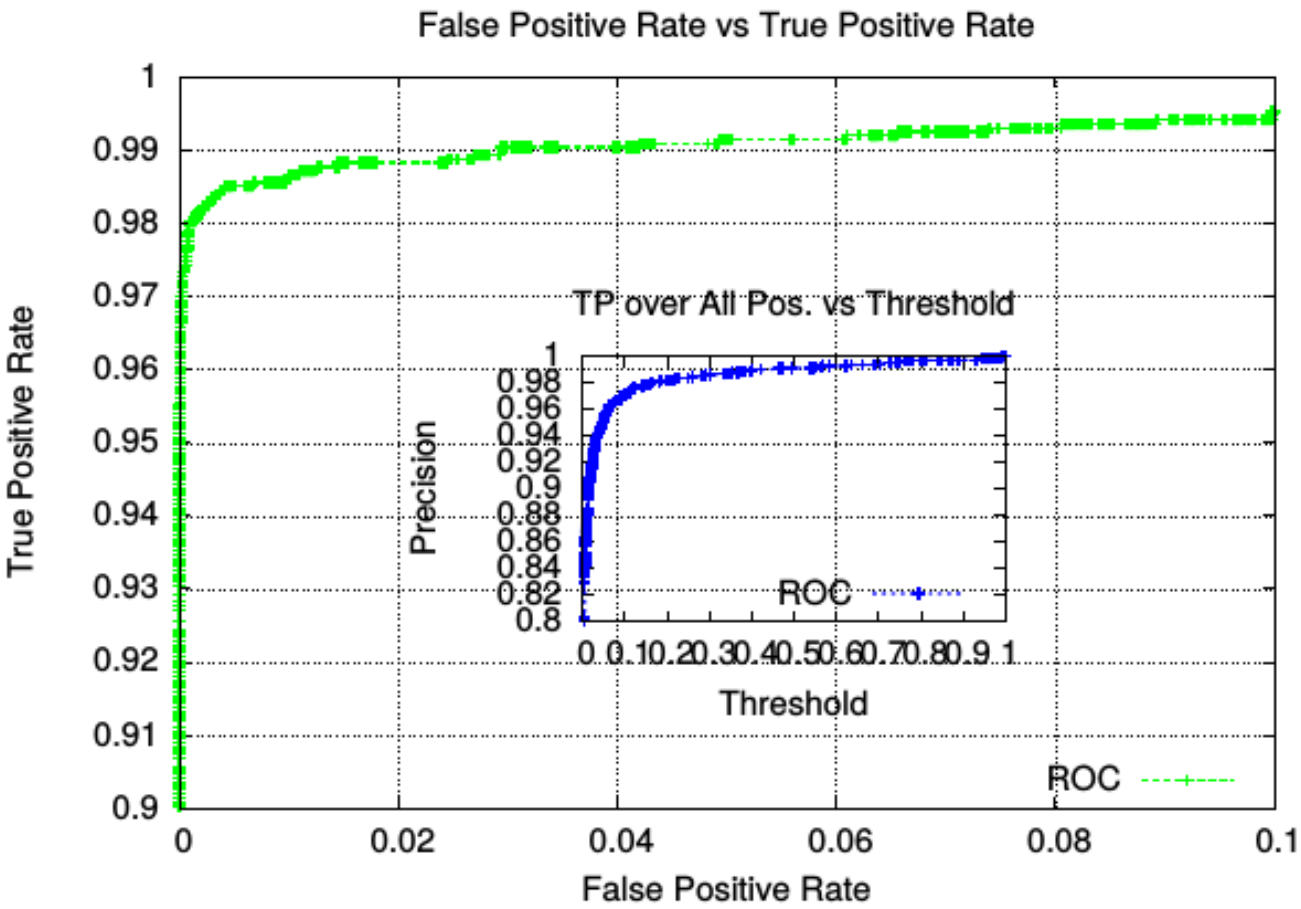
Outline

Data Statistics

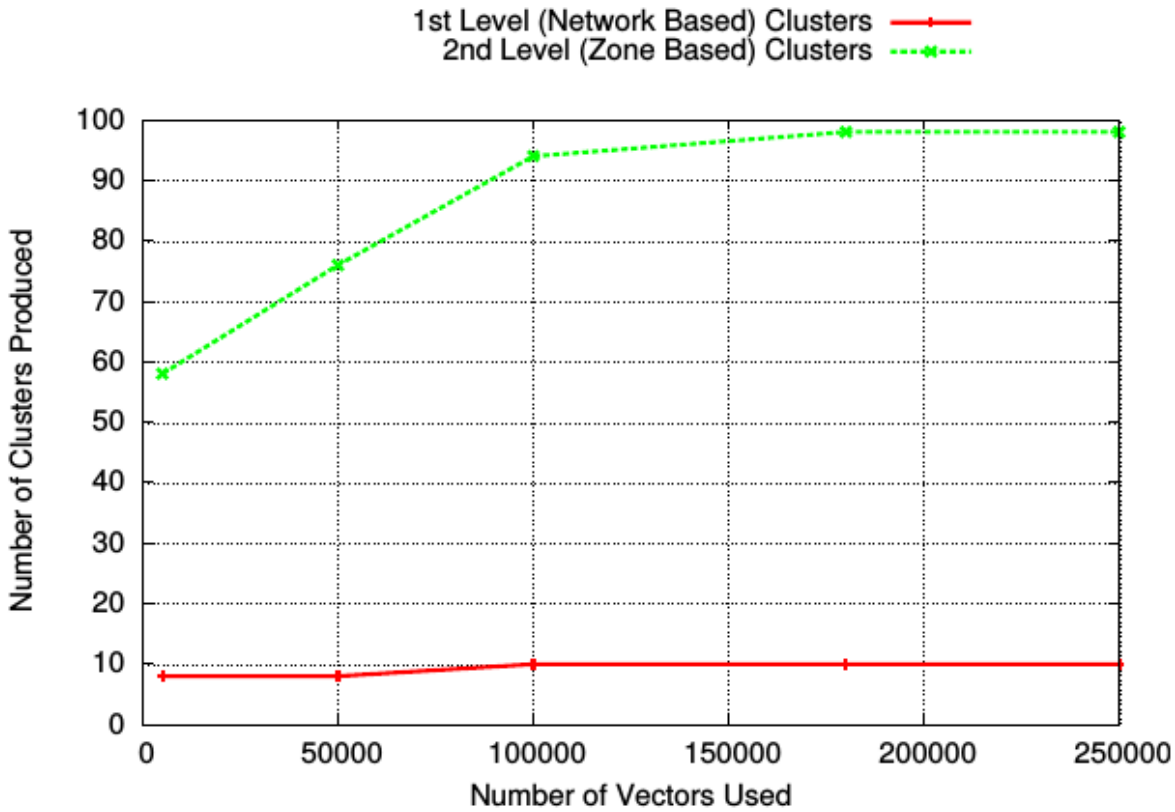
- 27,377,461 unique resolutions collected in 68 days (July-September 2009)
- SIE collected a volume of 200 Mbit/s resolutions
- ISP DNS Servers processed 30’000 requests /s during peak hours

Performance Overview I

- overall TP 96.8% and FP 0.38%

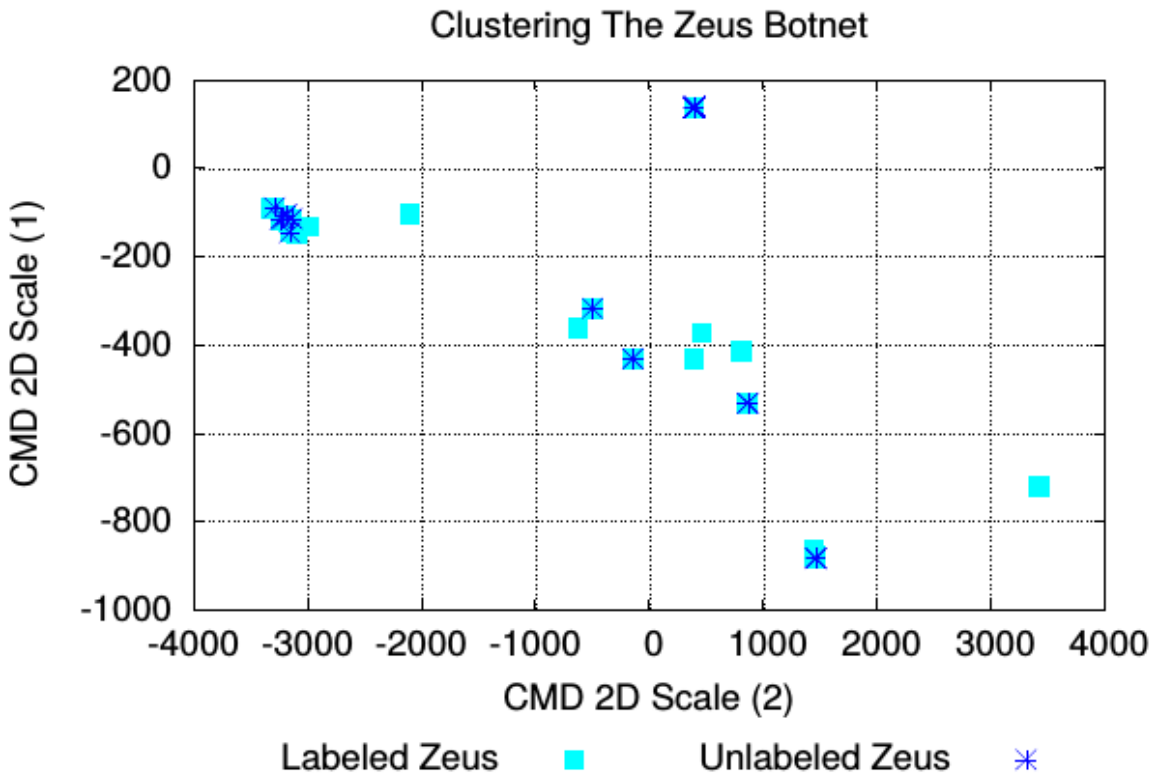


Performance Overview II



- computing 100k networks and a 15-days pDNS DB enough for a stable number of clusters

Performance Overview III



- previously unknown Zeus botnets accurately detected

Outline

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

Alternative Dynamic Solutions

1. *SNARE*
2. *Spamscatter*
3. *EXPOSURE*

Sources I

https://www.ietf.org/rfc/rfc1035.txt.

https://www.ietf.org/rfc/rfc1930.txt.

https://www.ietf.org/rfc/rfc4271.txt.

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