Exposure: Finding Malicious Domains Using Passive DNS Analysis

Leyla Bilge, Engin Kirda, Christopher Kruegel and Marco Balduzzi

DNS

- One of the core and most important component of Internet

- Besides being used for benign purposes, DNS is popular for malicious use as well
  - Botnet C&C
  - Dropzones
  - Phishing Sites
  - Spamming

Abusing DNS for malicious activity

- Attackers are faced with the same engineering challenges that global enterprises do
  - maintaining a large, distributed and reliable service infrastructure
- Leveraging DNS,
  - They acquire the flexibility to change the IP address of the malicious server
  - They can hide their critical servers behind proxy services
  - They get the flexibility of migrating their malicious servers by offering “fault-tolerant” services

Motivation

- As malicious services are often as dependent on DNS as benign services, being able to identify malicious domains would significantly help mitigate many Internet threats

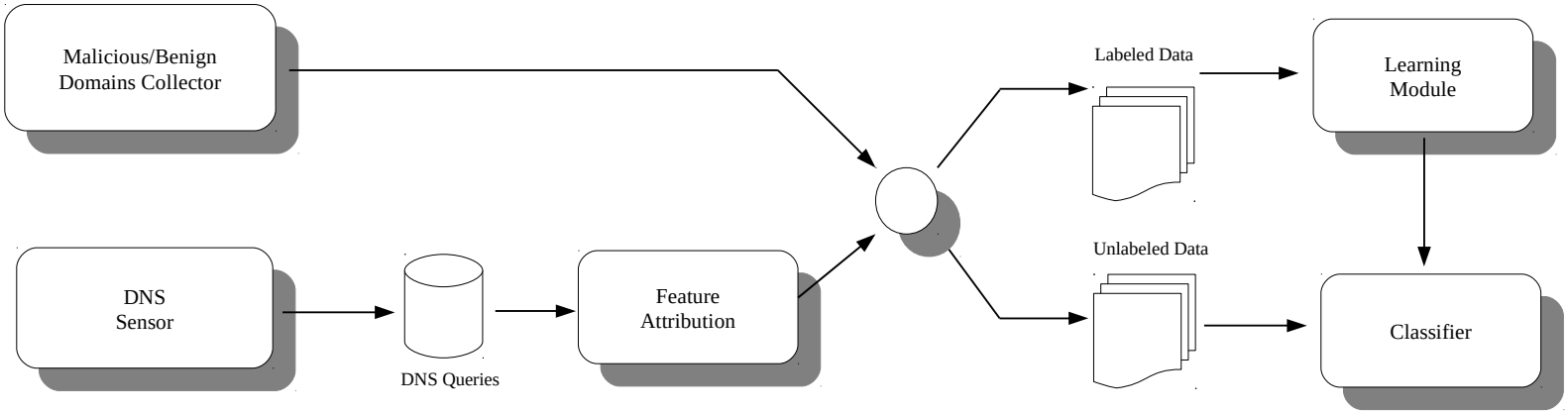
- When looking at large volumes of data, DNS requests for benign and malicious domains should exhibit enough differences in behavior that they can automatically be distinguished

Malware detection through DNS

- Is it possible to find distinguishable features for malicious and benign domains?
  - Fast-flux detectors
  - DNS reputation (Notos)

- Is it possible to build a live malware detection scheme by monitoring the DNS traffic of a network?

EXPOSURE : The Approach



DNS Sensor – Data Collectr

- Need for a large amount of training data
  - Identifying features that are able to distinguish malicious and benign DNS behavior
- We analyzed 2.5 months of DNS data produced by the SIE@ISC sensors
- The DNS traffic : DNS answer returned to the DNS servers
  - Time
  - TTL
  - DNS answer list
  - Domain name queried

Malicious Domains Collector

- A comprehensive list of malicious domains gathered from serveral sources
  - malwaredomains.com
  - Zeus Block List
  - Malware Domains List
  - Anubis reports
  - Wepawet
  - Phishtank
  - Domains list generated by DGAs of Conficker and Mebroot

Benign Domains Collector

- A list of benign domains that is representative for benign DNS usage
  - Alexa top 1000
  - Domains older than 1 year

- Two-way verification step
  - Cross-checked with the sources we gathered our malware domains list
  - Open Directory Project

    (i.e., a human-reviewed dictionary of web)

Feature Selection

- Time-based features
  - Short life, daily similar behavior, regular-irregular behavior
- DNS answer-based features
  - Fast-flux features, shared ip addresses
- TTL value-based features
  - Avg TTL, std TTL, TTL change
- Domain name-based features
  - Automatically generated domains

Feature Selection

- Time-based features
  - Short life, daily similar behavior, regular-irregular behavior
- DNS answer-based features
  - Fast-flux features, shared ip addresses
- TTL value-based features
  - Avg TTL, std TTL, TTL change
- Domain name-based features
  - Automatically generated domains

Time-based Features

- The time of an individual request is not very useful by itself
  - Requests to a particular domain over time may constitute different patterns on malicious and benign domains
- To analyze the changes of the number of requests for a domain, the collection of DNS queries targeting a domain were converted into time series

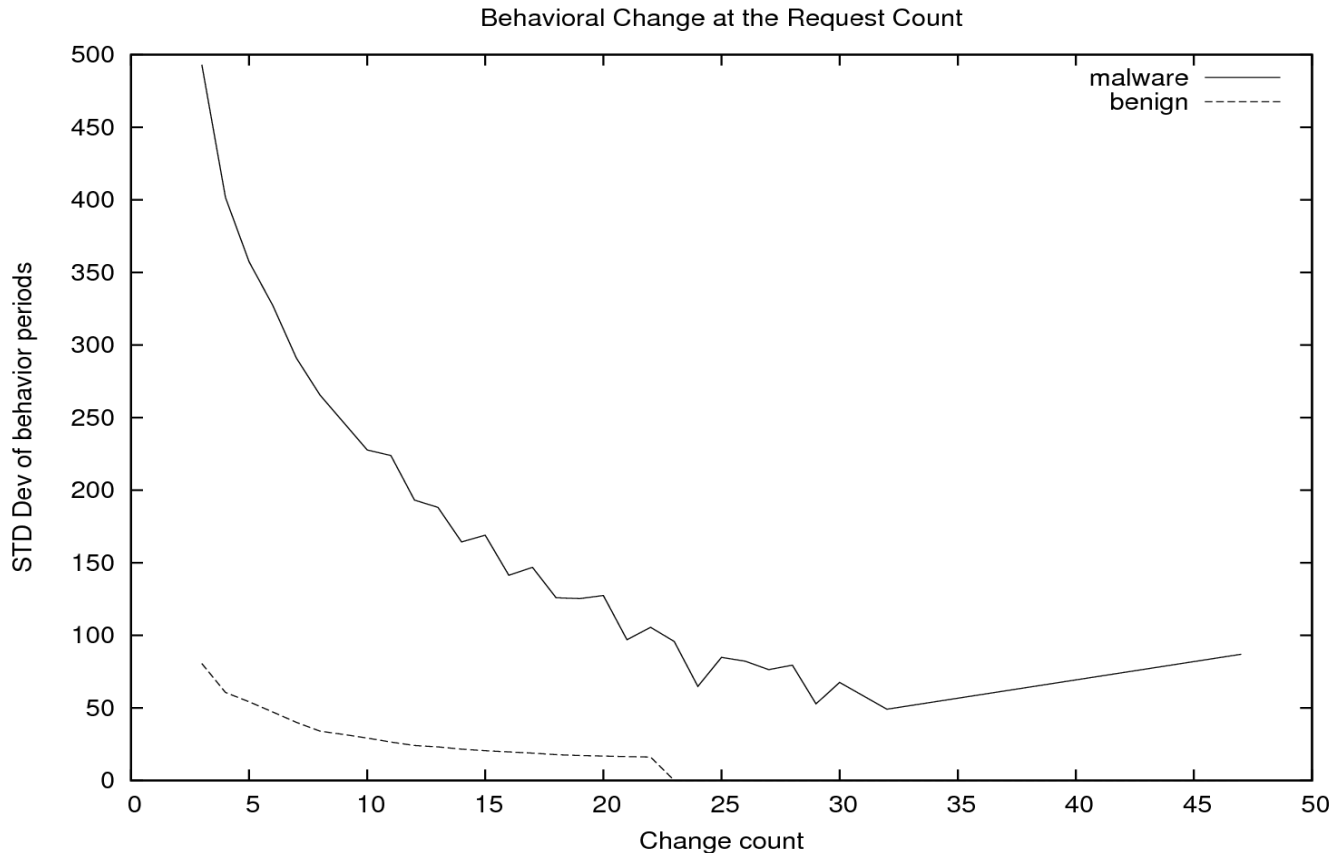
Time-based Features

- Malicious services that use the technique named *domain flux* show a sudden increase followed by a sudden decrease on the time series
  - Torpig
  - Conficker
- The problem of detecting *short-lived* domains can be treated as a change point detection (CPD) problem

Time-based Features

- CPD algorithm can also be used for detecting behavioral characteristics of a domain by zooming into its life time
- CPD algorithm outputs the points in time the changes are detected and the average behavior for each duration
- Features extracted from CPD algorithm
  - Number of changes
  - Average behavior
  - Standard deviation of the behavioral changes
  - Average behavior duration
  - Standard deviation of the behavior durations

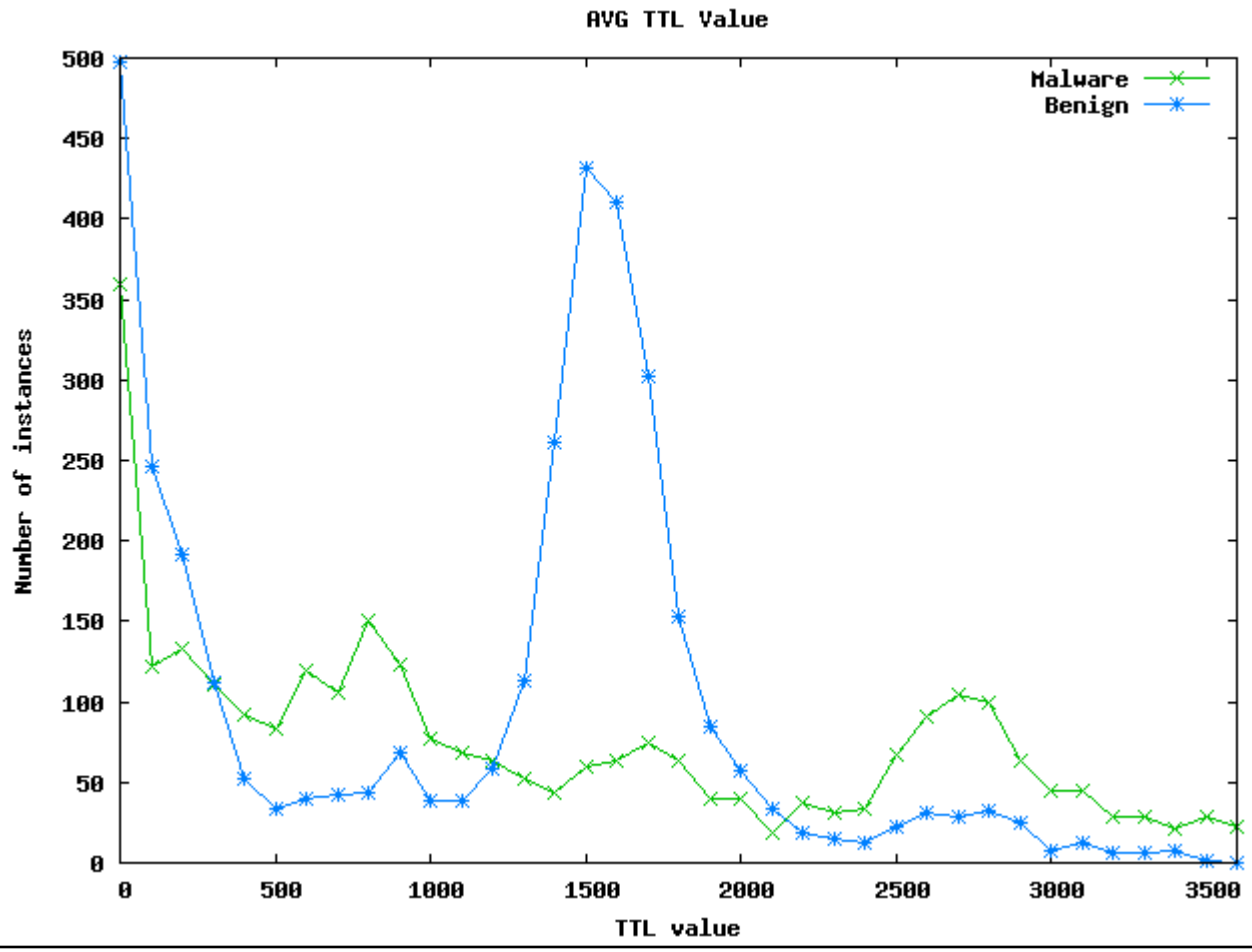
Time-based Features



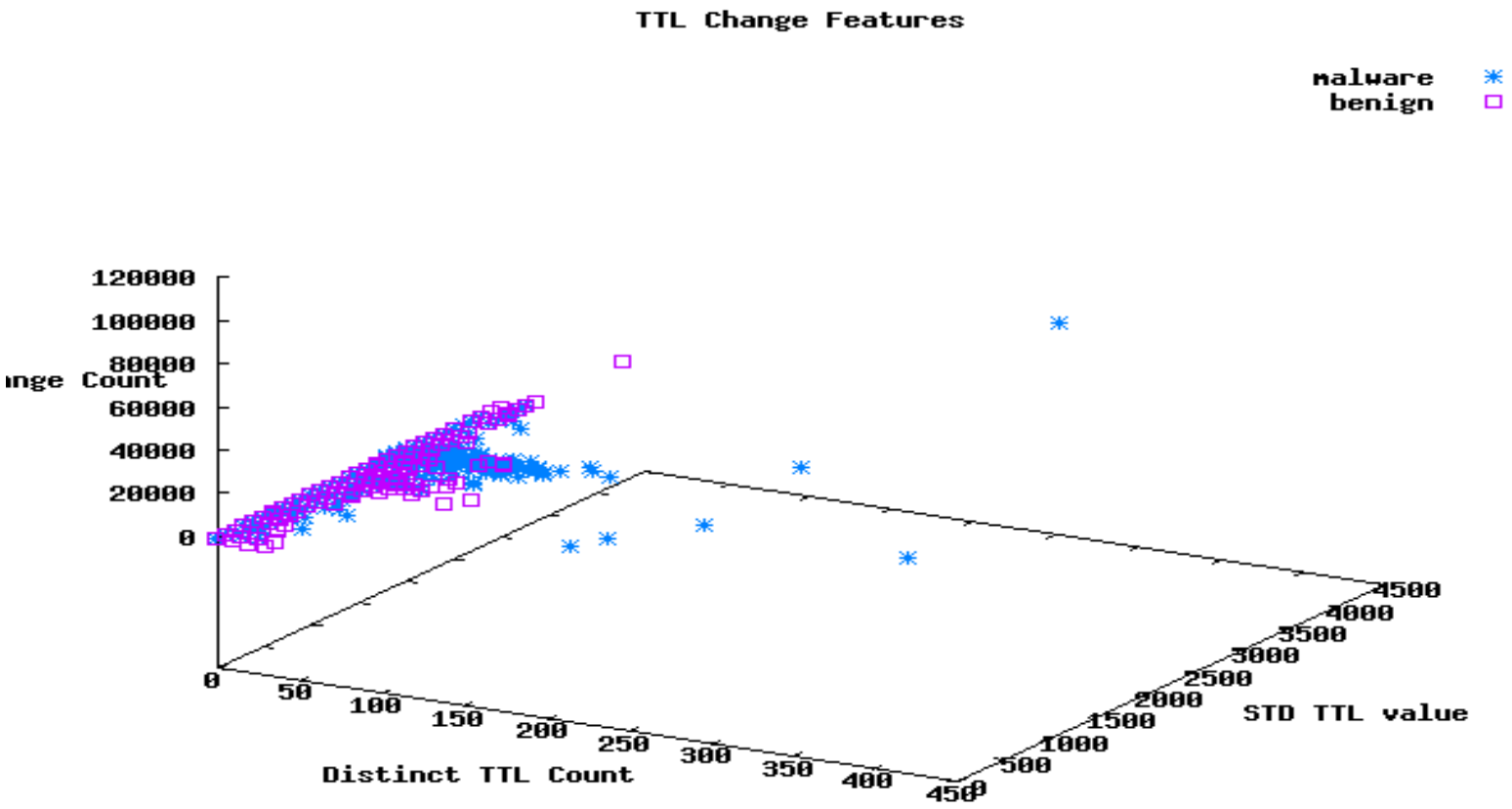
TTL-Based Features

- Every DNS record has a *Time To Live* (TTL)
- It is recommended that the TTL is set between 1 and 5 days so that the name servers can benefit from DNS caching
- However:
  - Systems that aim for high availability often set low TTL values to benefit from Round Robin DNS
  - A representative example for such systems are Content Delivery Networks (CDNs)
- Unfortunately:
  - Low TTL and Round Robing DNS is useful for the attackers as well. e.g. Fast-Flux Service Networks

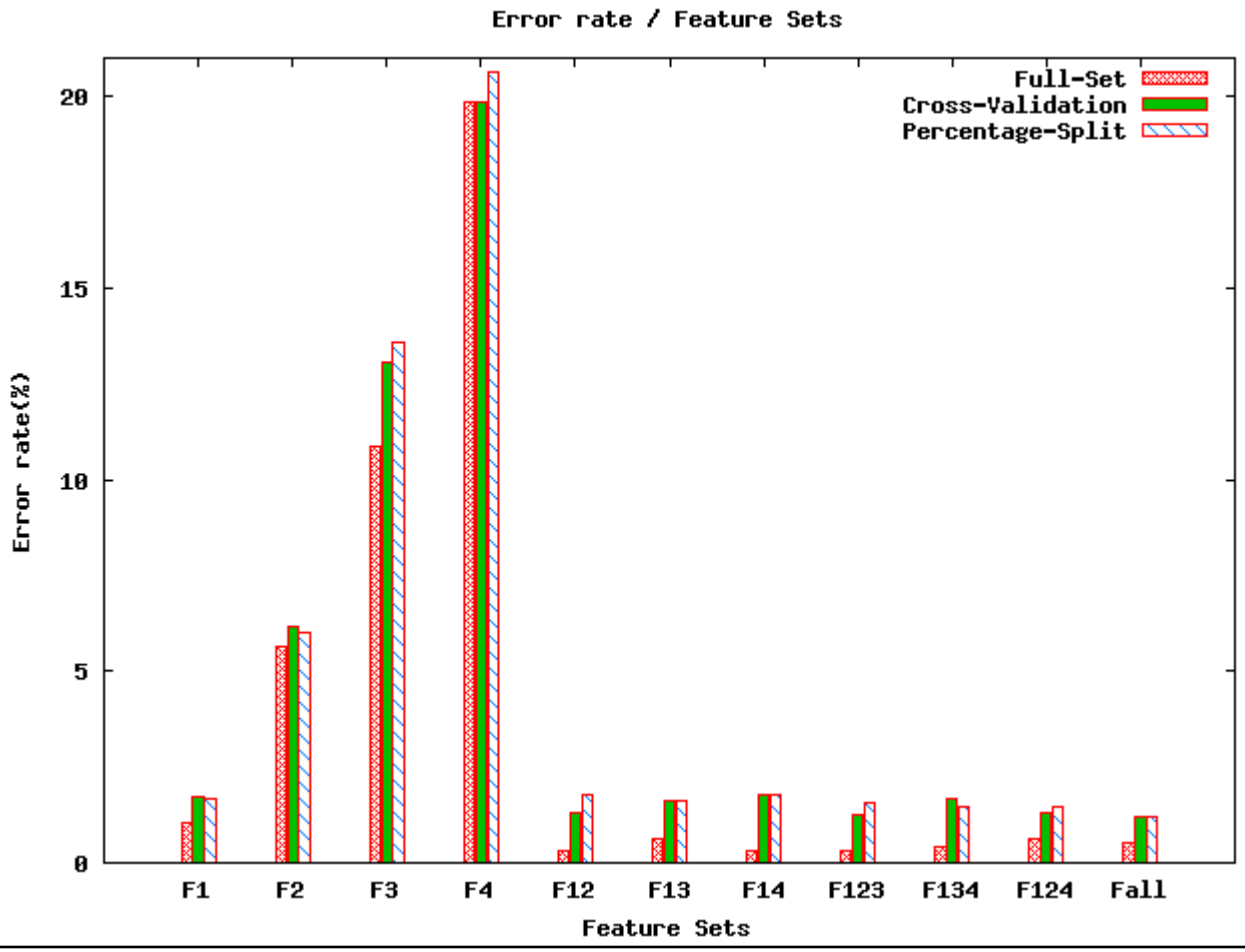
TTL Value


AVG TTL Value
Malware
Benign
Number of instances
TTL value
0
50
100
150
200
250
300
350
400
450
500
0
500
1000
1500
2000
2500
3000
3500

TTL Change


TTL Change Features
malware
benign
120000
100000
80000
60000
40000
20000
0
nge Count
0
50
100
150
200
250
300
350
400
450
Distinct TTL Count
0
500
1000
1500
2000
2500
3000
3500
4000
4500
STD TTL value

Learning Module - Classifier



Evaluation – SIE Data

- During a period of 2.5 months, we monitored 25 billion DNS queries
  - Since such an amount of data is not feasible in practice to be processed, we applied some filtering policies
    - Alexa TOP 1000
    - Domains that are older than 1 year
- After filtering, our system recorded 4.5 million distinct domains that were queried by real users

Evaluation – SIE Data

- Time series analysis produces accurate results only when the sampling count is high enough
  - Based on the empirical results, we set the threshold to 20 queries
- In our experiments, we focused on 300,000 domains that received more than 20 DNS requests
  - 17,686 out of 300,000 domains detected as malicious

Evaluation – SIE Data / DR

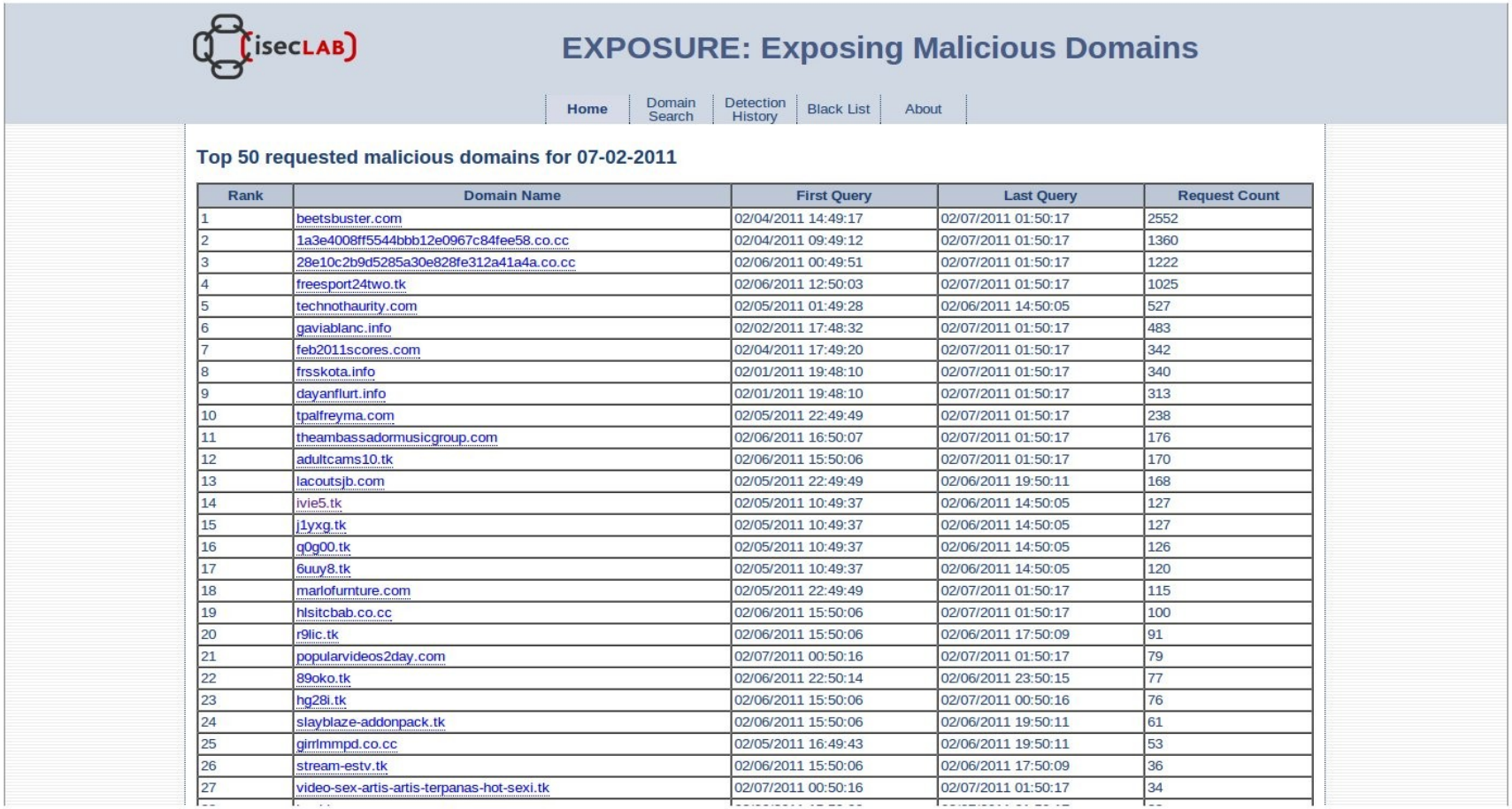
- The percentage split and cross-validation evaluations on the training set show that the detection rate of our classifier is around 98%.
- Can we also detect the malicious domains that do not exist in our training set?
  - During the period of our experiments, `malwareurls.com` reported 569 domains as being malicious
  - Our system observed 216 of them in the DNS traffic provided by SIE
  - 211 domains detected as malicious by our system

Evaluation – Real-Time Detection

- We deployed our system on an ISP network with 30,000 clients
  - No filtering was applied to the data
- During two weeks of the experiments, we detected 3117 malicious domains
  - 2821 of these domains fall into the category of domains that are generated by DGA, therefore they were all short-lived domains.
  - 5 out of remaining 396 domains were identified as malicious later by some malware analysis tools
  - The rest were cross-checked with McAfee Site Advisor

exposure.iseclab.org

iSecLab @ Eurecom



iSecLAB

EXPOSURE: Exposing Malicious Domains

Home | Domain Search | Detection History | Black List | About

Top 50 requested malicious domains for 07-02-2011

Rank	Domain Name	First Query	Last Query	Request Count
1	beetsbuster.com	02/04/2011 14:49:17	02/07/2011 01:50:17	2552
2	1a3e4008ff5544bbb12e0967c84fee58.co.cc	02/04/2011 09:49:12	02/07/2011 01:50:17	1360
3	28e10c2b9d5285a30e828fe312a41a4a.co.cc	02/06/2011 00:49:51	02/07/2011 01:50:17	1222
4	freesport24two.tk	02/06/2011 12:50:03	02/07/2011 01:50:17	1025
5	technothaurity.com	02/05/2011 01:49:28	02/06/2011 14:50:05	527
6	gaviablanc.info	02/02/2011 17:48:32	02/07/2011 01:50:17	483
7	feb2011scores.com	02/04/2011 17:49:20	02/07/2011 01:50:17	342
8	frsskota.info	02/01/2011 19:48:10	02/07/2011 01:50:17	340
9	dayanflurt.info	02/01/2011 19:48:10	02/07/2011 01:50:17	313
10	tpalfreyma.com	02/05/2011 22:49:49	02/07/2011 01:50:17	238
11	theambassadormusicgroup.com	02/06/2011 16:50:07	02/07/2011 01:50:17	176
12	adultcams10.tk	02/06/2011 15:50:06	02/07/2011 01:50:17	170
13	lacoutsjb.com	02/05/2011 22:49:49	02/06/2011 19:50:11	168
14	ivie5.tk	02/05/2011 10:49:37	02/06/2011 14:50:05	127
15	j1yxg.tk	02/05/2011 10:49:37	02/06/2011 14:50:05	127
16	q0g00.tk	02/05/2011 10:49:37	02/06/2011 14:50:05	126
17	6uuy8.tk	02/05/2011 10:49:37	02/06/2011 14:50:05	120
18	marlofurnture.com	02/05/2011 22:49:49	02/07/2011 01:50:17	115
19	hlsitcbab.co.cc	02/06/2011 15:50:06	02/07/2011 01:50:17	100
20	r9lic.tk	02/06/2011 15:50:06	02/06/2011 17:50:09	91
21	popularvideos2day.com	02/07/2011 00:50:16	02/07/2011 01:50:17	79
22	89oko.tk	02/06/2011 22:50:14	02/06/2011 23:50:15	77
23	hg28i.tk	02/06/2011 15:50:06	02/07/2011 00:50:16	76
24	slayblaze-addonpack.tk	02/06/2011 15:50:06	02/06/2011 19:50:11	61
25	girrlmmpd.co.cc	02/05/2011 16:49:43	02/06/2011 19:50:11	53
26	stream-estv.tk	02/06/2011 15:50:06	02/06/2011 17:50:09	36
27	video-sex-artis-artis-terpanas-hot-sexi.tk	02/07/2011 00:50:16	02/07/2011 01:50:17	34

Conclusion

- As DNS is critical service for the functioning of benign services, it plays an important role for malicious activities as well.
- Monitoring the use of DNS on a large-scale allows us to find distinguishable features for malicious and benign domains.
- A real-time malicious domains detection system can be realized using these features.

Thanks...

?

FP Estimation

MW-Group	Rand 50	Malicious	MW-Group	Rand 50	Malicious
Spam	18	3691	Adult	3	1716
Black-List	8	1734	Risky	-	788
FastFlux	-	114	Phishing	3	-
Malware	6	979	No Info	5	2854
`Conficker	4	3693	FP	3 (6%)	1408 (7.9%)

Evaluation - Real-Time Detection

Groups	Avg Life Time	Most Freq Life Time	# of infected machines
DGA domains	1.2 days	0.99 days	49
Iksmas Worm	11.9 days	11.9 days	70
Worm:Win32/Slenping	12.0 days	12.0 days	253
Trojan-Generic.dx	11.9 days	11.9 days	70
Other	10.8 days	11.9 days	391

Detecting short-lived domains as malicious after a long time passes is useless.

Malicious Activity on the Internet

- Malicious activities performed on Internet pose a big threat to the users

- Increasing number of large scale malicious activities
  - Collections of remotely controlled hosts that are often used to launch DoS, steal sensitive information etc.
  - attackers set up a phishing website and lure unsuspecting users into entering sensitive information