Optimized Invariant Representation of Network Traffic for Detecting Unseen Malware Variants

Karel Bartos

(kbartos@cisco.com)

Michal Sofka

Vojtech Franc

Motivation – Network Security Challenges

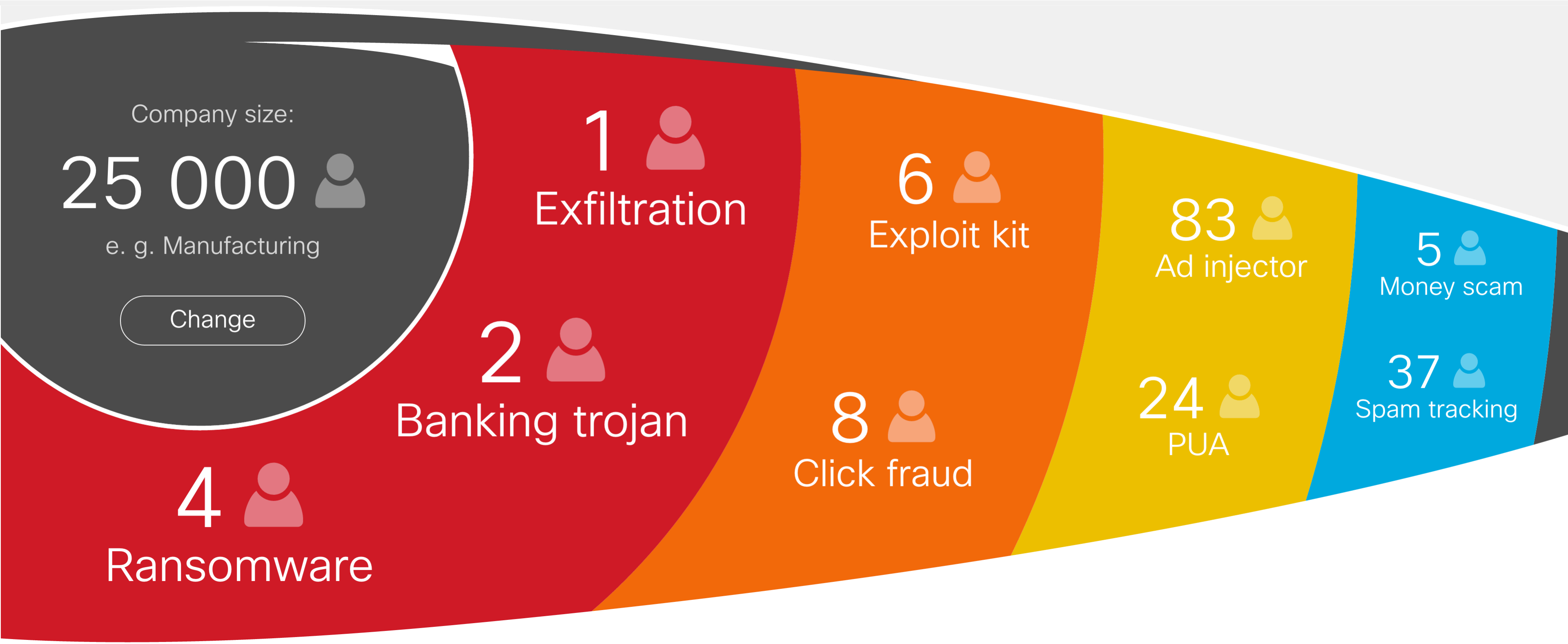
- Large variability of malicious samples
  - 100k new or modified malware samples every day

- Lack of labeled data (obtaining additional labels is costly)
  - Most of existing methods rely on signature matching or feeds

  ↑ High precision ↓ Low recall (detect only known threats)

- Behavior changes introduce problems when training detectors
  - Attackers change the behavior frequently to remain undetected

Malicious Traffic and HTTP(S)



Company size:

25 000

e. g. Manufacturing

Change

1 Exfiltration

2 Banking trojan

4 Ransomware

6 Exploit kit

8 Click fraud

83 Ad injector

24 PUA

5 Money scam

37 Spam tracking

Our Goal

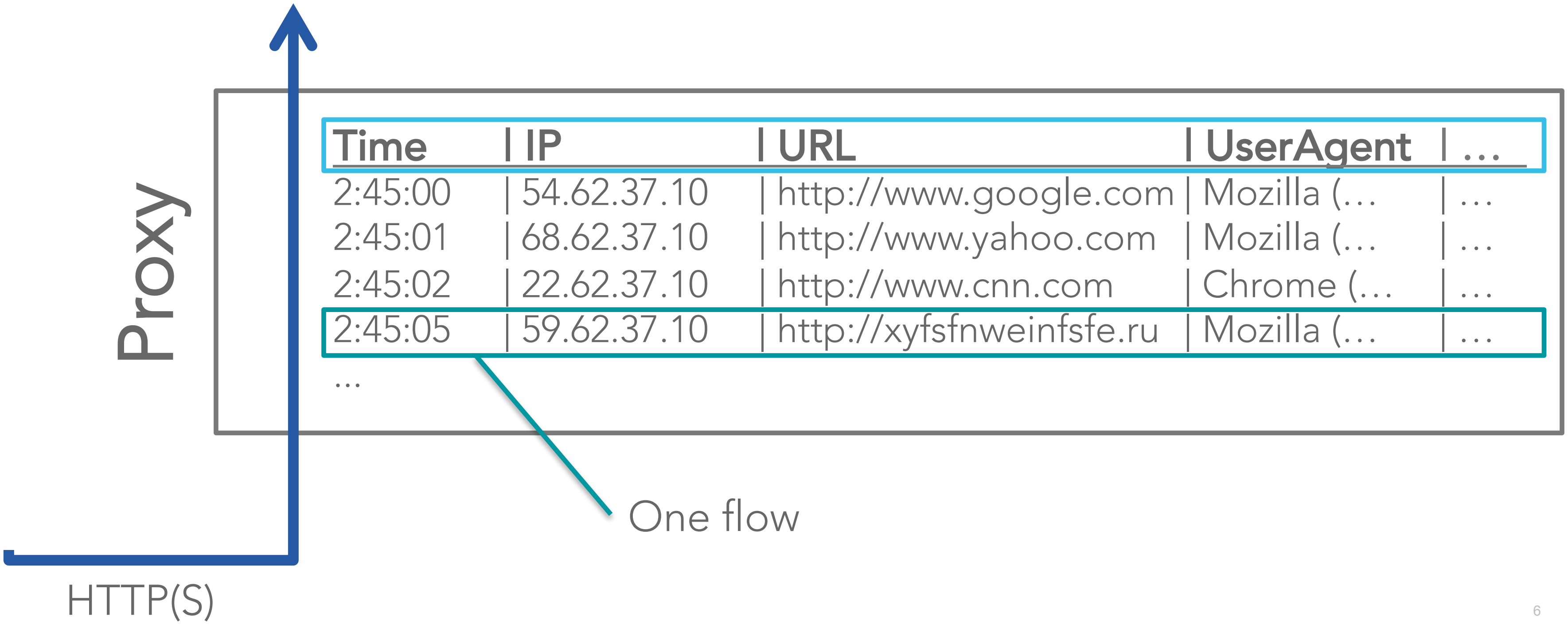
Build a representation of malware behavior robust to most of the modifications done by the attacker in the future:

Our Goal

Build a representation of malware behavior robust to most of the modifications done by the attacker in the future:

- Change in malicious code, payload, obfuscation
- Change in hostname or server IP address
- Change in the intensity
- Change in timing
- Change in URL path, parameters, etc.

Input data – proxy log records



Time	IP	URL	UserAgent	…
2:45:00	54.62.37.10	http://www.google.com	Mozilla (…	…
2:45:01	68.62.37.10	http://www.yahoo.com	Mozilla (…	…
2:45:02	22.62.37.10	http://www.cnn.com	Chrome (…	…
2:45:05	59.62.37.10	http://xyfsfnweinfsfe.ru	Mozilla (…	…
…				

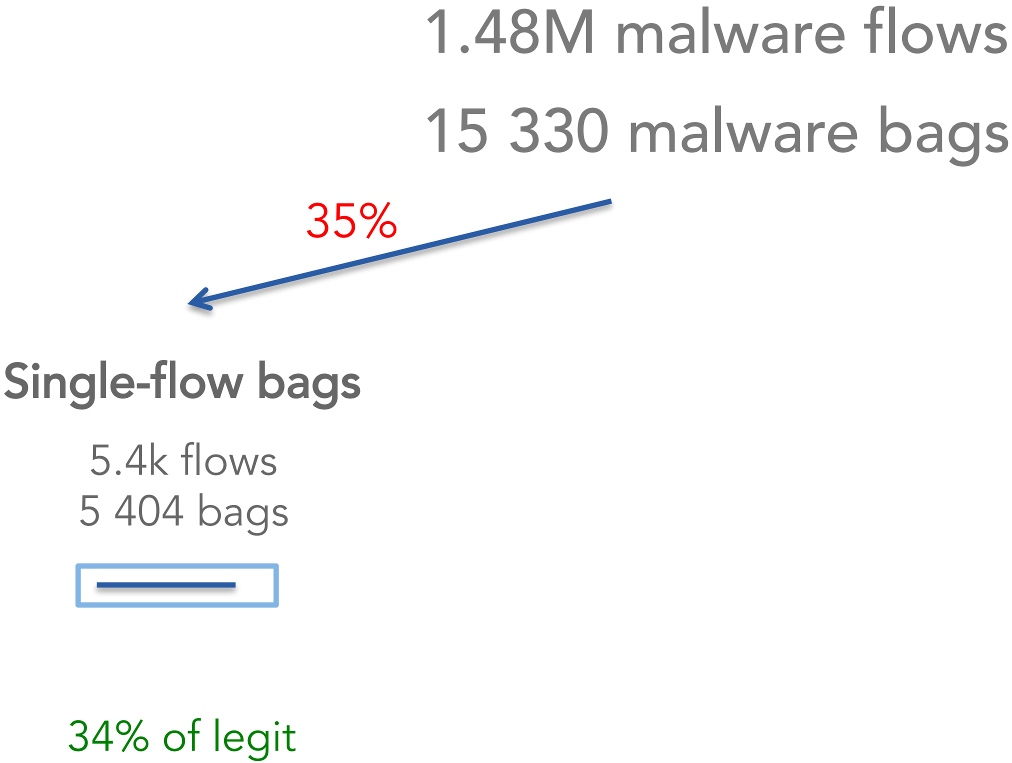
Flows are Grouped into Bags

BAG

= Flows from one user/device to one hostname in the given time interval

Contains user-hostname communication

Malware Bags



1.48M malware flows

15 330 malware bags

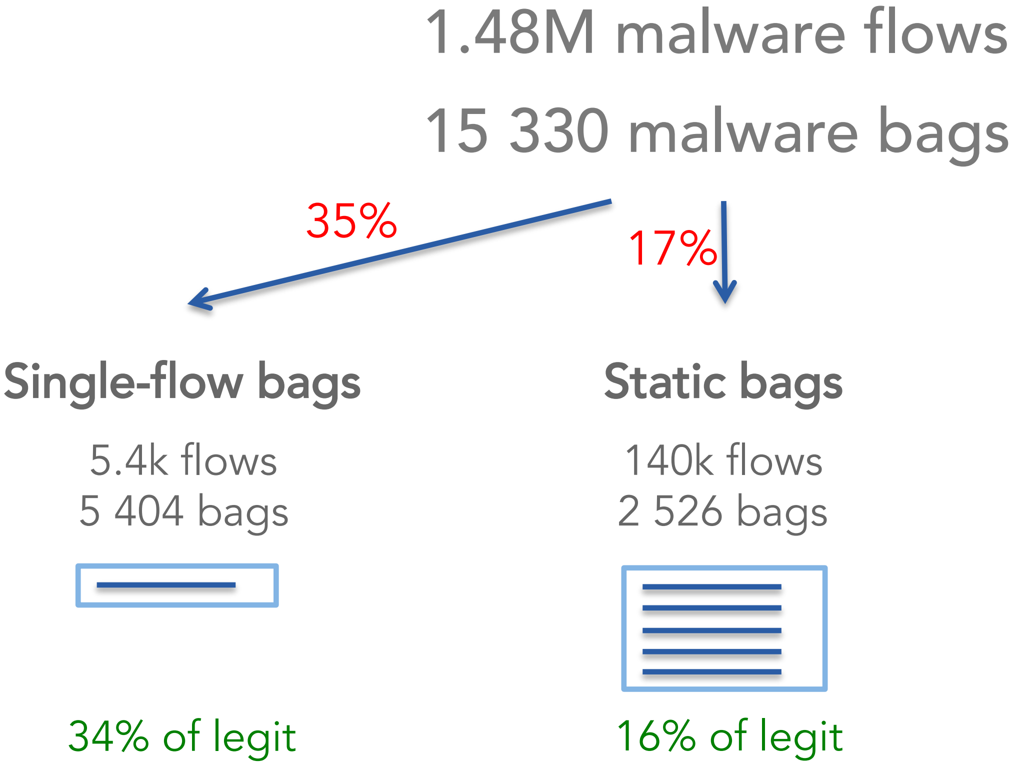
35%

Single-flow bags

5.4k flows
5 404 bags

34% of legit

Malware Bags



1.48M malware flows

15 330 malware bags

35%

17%

Single-flow bags

5.4k flows
5 404 bags

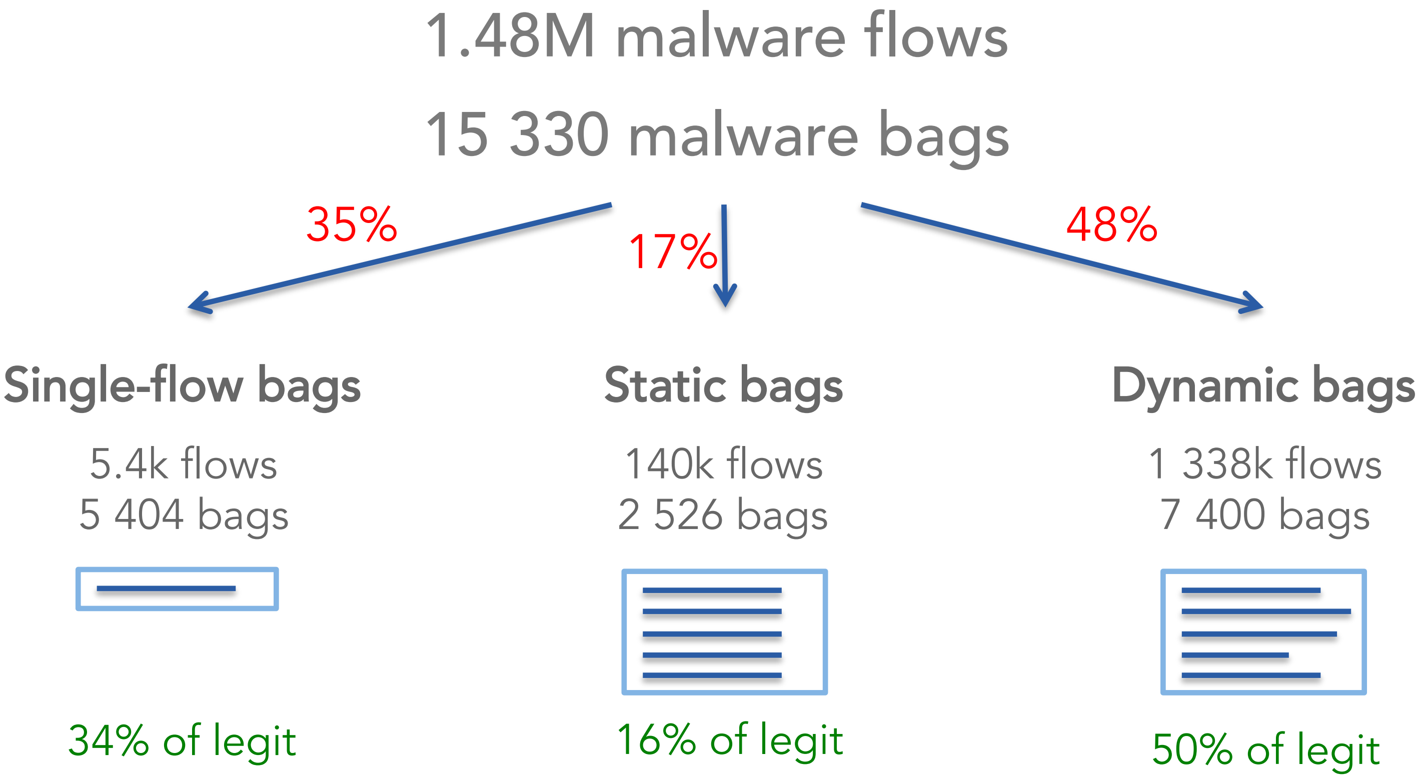
34% of legit

Static bags

140k flows
2 526 bags

16% of legit

Malware Bags



1.48M malware flows

15 330 malware bags

35%

17%

48%

Single-flow bags

5.4k flows
5 404 bags

34% of legit

Static bags

140k flows
2 526 bags

16% of legit

Dynamic bags

1 338k flows
7 400 bags

50% of legit

Malware Bags

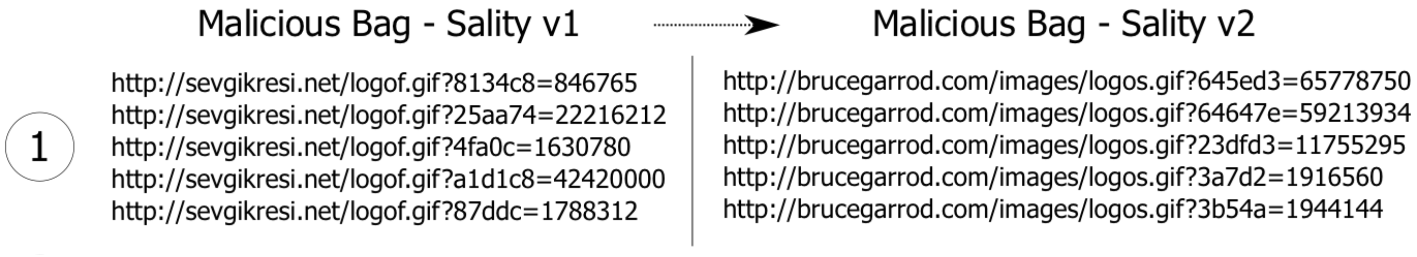
	Single-flow bags	Static bags	Dynamic bags
	5.4k flows 5 404 bags	140k flows 2 526 bags	1 338k flows 7 400 bags
Percent of malware bags:	35%	17%	48%
Features:	Flow-based	Flow-based	Flow-based

Malware Bags

	Single-flow bags	Static bags	Dynamic bags
	5.4k flows 5 404 bags	140k flows 2 526 bags	1 338k flows 7 400 bags
Percent of malware bags:	35%	17%	48%
Features:	Flow-based	Flow-based	~~Flow-based~~ **Bag-based Features**

Malware Changes – Example

Network traffic of two malware bags of the same type



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

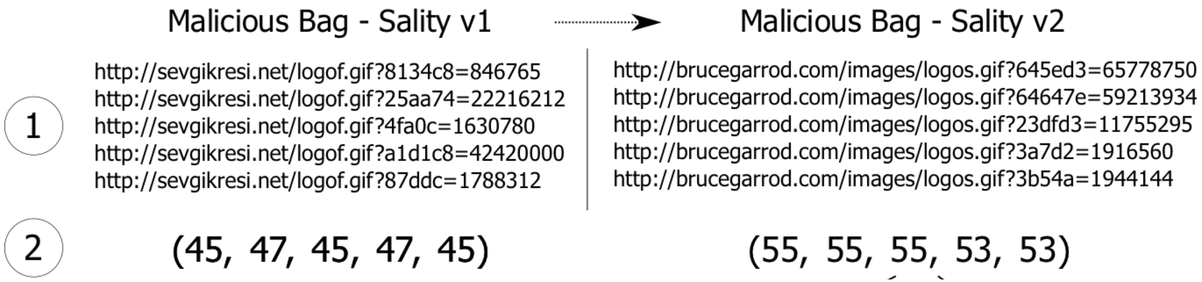
1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

Similar of different?

Flow-based features



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

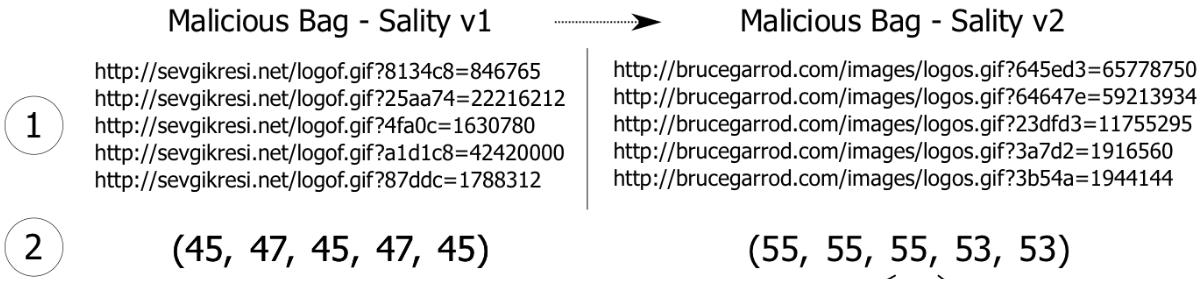
http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

1 feature: URL length

Flow-based features



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

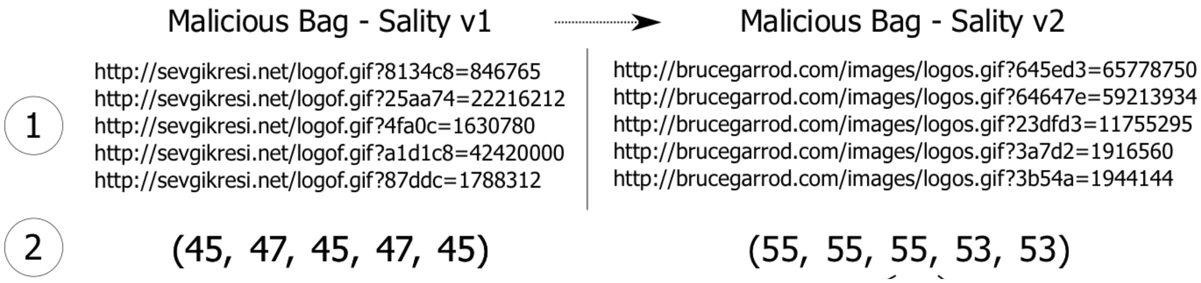
1 feature: URL length

Training bag

Testing bag

Flow-based features

1 feature: URL length



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

```
http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312
```

```
http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144
```

2

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

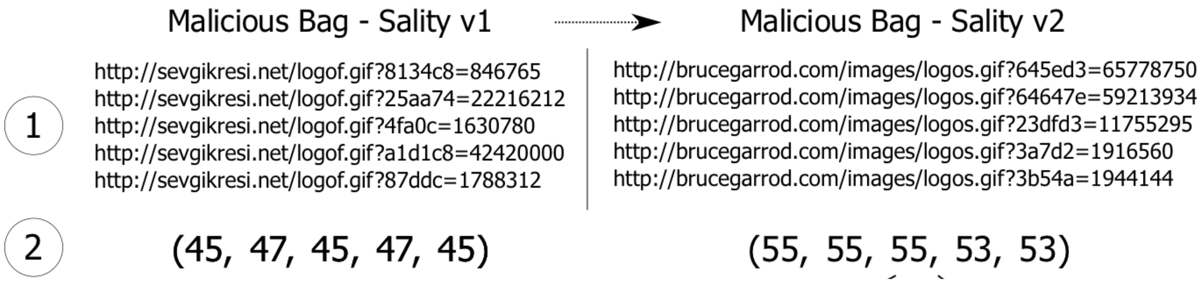
Training bag

Testing bag

Model: 44 < length(URL) < 48

Flow-based features

1 feature: URL length



Malicious Bag - Sality v1

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

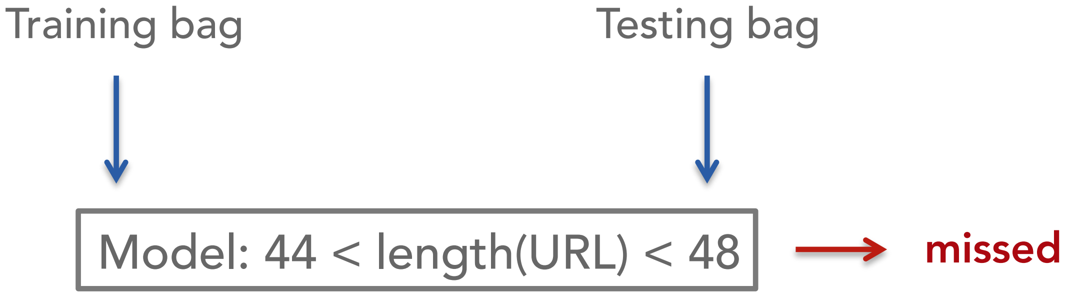
2

(45, 47, 45, 47, 45)

Malicious Bag - Sality v2

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

(55, 55, 55, 53, 53)



Training bag

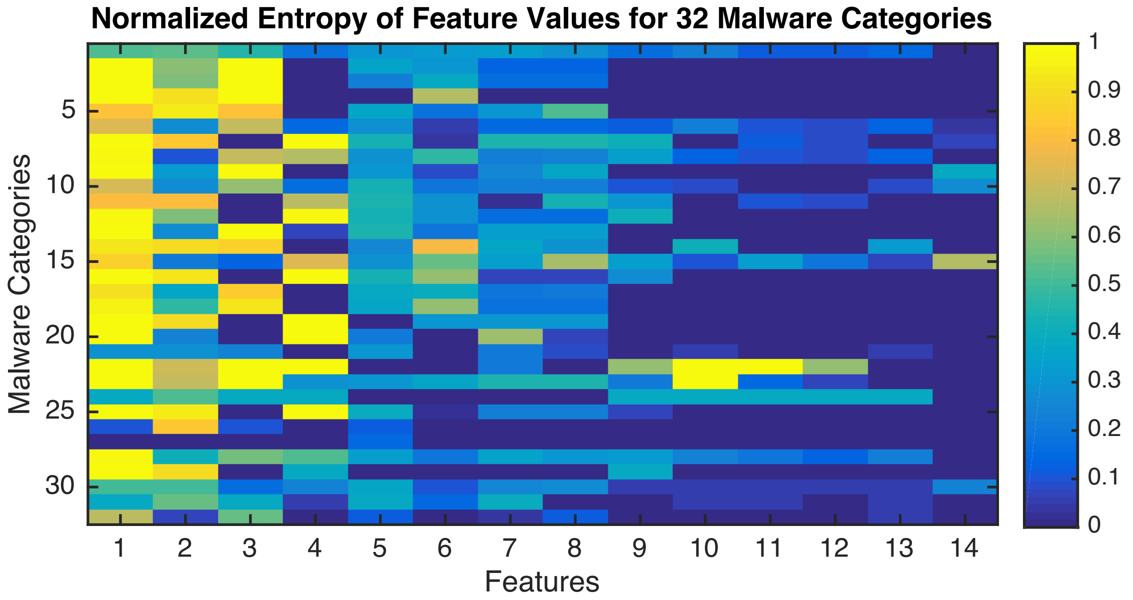
Testing bag

Model: 44 < length(URL) < 48

missed

Intuition: Flow-based features are not suitable for dynamic bags.

High Variability of Flow-based Features



Yellow = high variability

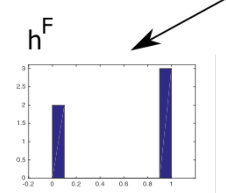
Features:
1 – URL string
2 – Thinking time
3 – URL query values
4 – URL path
5 – Number of flows
6 – SC bytes
7 – Server IP
8 – Hostname
9 – URL path length
10 – URL query names
11 – URL filename
12 – URL filename length
13 – Number of URL query params
14 – Cs bytes

Categories: Asterope, Bedep, Dridex, Gamarue, InstallCore, Mudrop, MultiPlug, Poweliks, Rerdom, Ramnit, Rovnix, Sality, Tempedreve, Upatre, Vawtrak, Wowliks, …

Histogram

	Malicious Bag - Sality v1	Malicious Bag - Sality v2
1	http://sevgikresi.net/logof.gif?8134c8=846765 http://sevgikresi.net/logof.gif?25aa74=22216212 http://sevgikresi.net/logof.gif?4fa0c=1630780 http://sevgikresi.net/logof.gif?a1d1c8=42420000 http://sevgikresi.net/logof.gif?87ddc=1788312	http://brucegarrod.com/images/logos.gif?645ed3=65778750 http://brucegarrod.com/images/logos.gif?64647e=59213934 http://brucegarrod.com/images/logos.gif?23dfd3=11755295 http://brucegarrod.com/images/logos.gif?3a7d2=1916560 http://brucegarrod.com/images/logos.gif?3b54a=1944144
2	(45, 47, 45, 47, 45)	(55, 55, 55, 53, 53)

3



histogram

4 bins

4

(0.4, 0, 0, 0.6

4 feature values

Histogram

Malicious Bag - Sality v1

Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

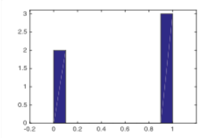
2

(45, 47, 45, 47, 45)

(55, 55, 55. 53. 53)

3

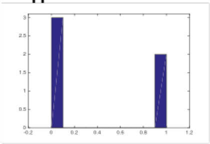
h^F



histogram

4 bins

h^F



4

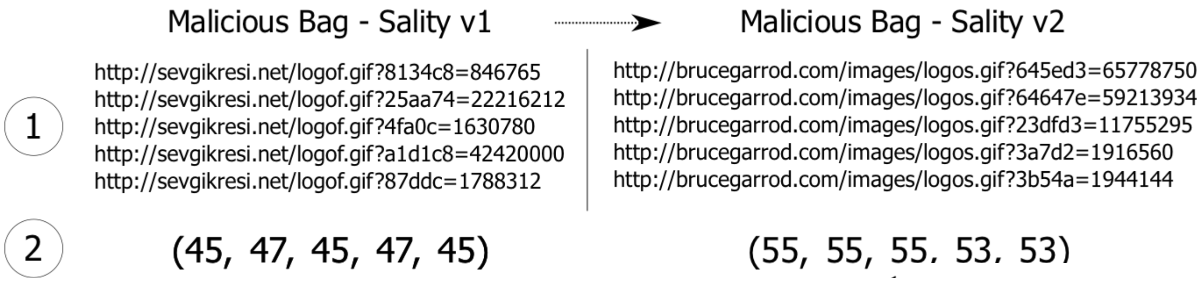
(0.4, 0, 0, 0.6

4 feature values

(0.6, 0, 0, 0.4

missed?

What is better?



Malicious Bag - Sality v1

1
```
http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312
```

2 (45, 47, 45, 47, 45)

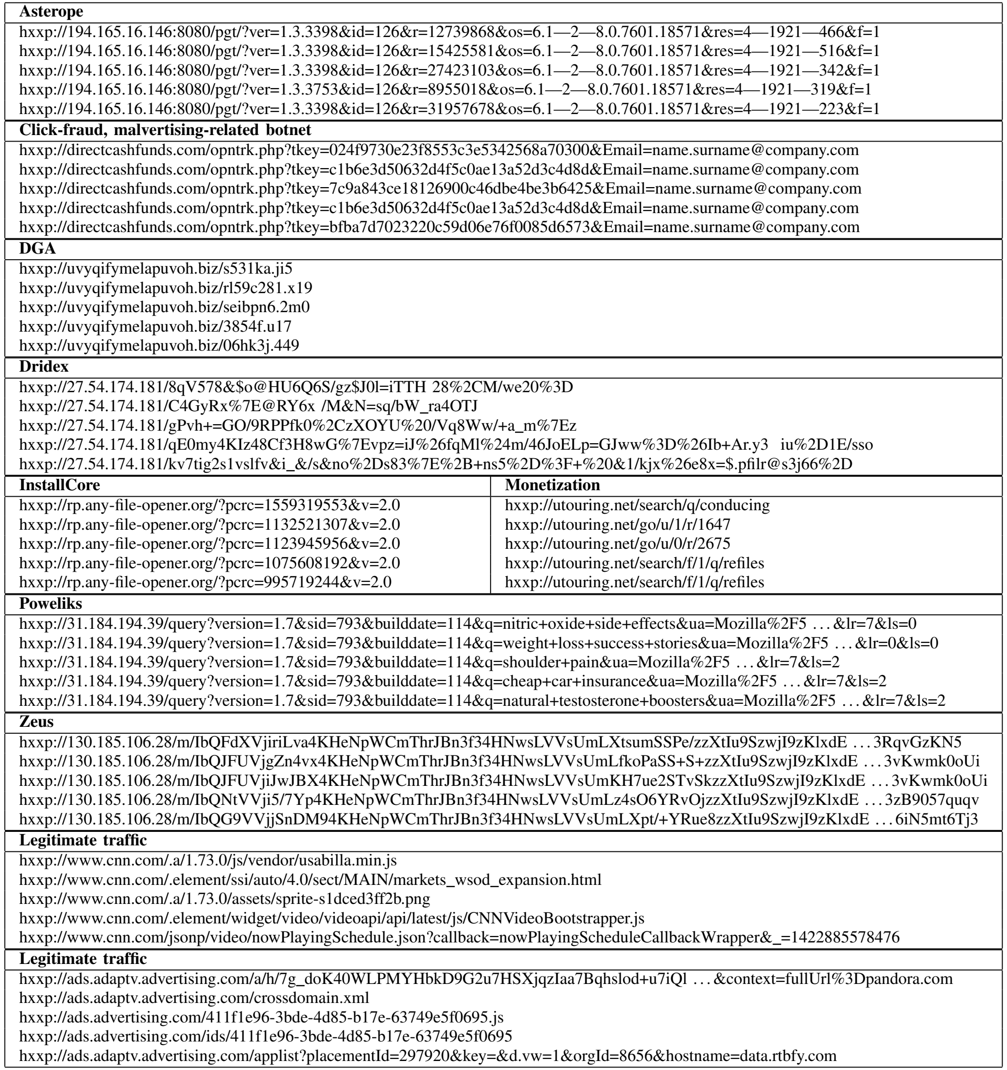
Malicious Bag - Sality v2

1
```
http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144
```

2 (55, 55, 55, 53, 53)

What do we want to represent?

What is common across malware categories?

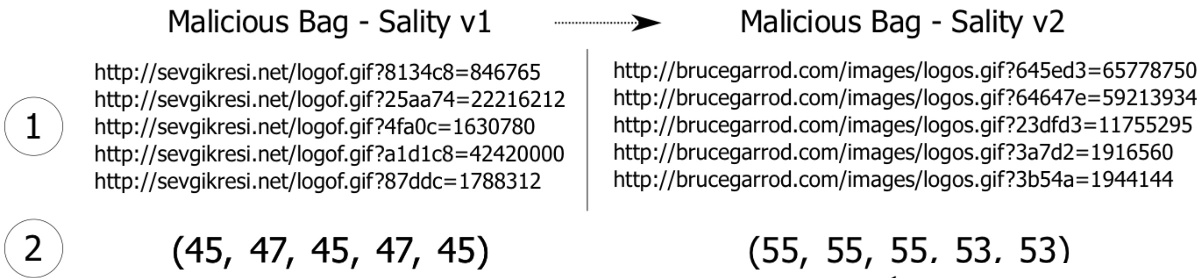


Asterope	
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=12739868&os=6.1—2—8.0.7601.18571&res=4—1921—466&f=1	
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=15425581&os=6.1—2—8.0.7601.18571&res=4—1921—516&f=1	
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=27423103&os=6.1—2—8.0.7601.18571&res=4—1921—342&f=1	
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3753&id=126&r=8955018&os=6.1—2—8.0.7601.18571&res=4—1921—319&f=1	
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=31957678&os=6.1—2—8.0.7601.18571&res=4—1921—223&f=1	
Click-fraud, malvertising-related botnet	
hxxp://directcashfunds.com/opntrk.php?tkey=024f9730e23f8553c3e5342568a70300&Email=name.surname@company.com	
hxxp://directcashfunds.com/opntrk.php?tkey=c1b6e3d50632d4f5c0ae13a52d3c4d8d&Email=name.surname@company.com	
hxxp://directcashfunds.com/opntrk.php?tkey=7c9a843ce18126900c46dbe4be3b6425&Email=name.surname@company.com	
hxxp://directcashfunds.com/opntrk.php?tkey=c1b6e3d50632d4f5c0ae13a52d3c4d8d&Email=name.surname@company.com	
hxxp://directcashfunds.com/opntrk.php?tkey=bfba7d7023220c59d06e76f0085d6573&Email=name.surname@company.com	
DGA	
hxxp://uvyqifymelapuvoh.biz/s531ka.ji5	
hxxp://uvyqifymelapuvoh.biz/rl59c281.x19	
hxxp://uvyqifymelapuvoh.biz/seibpn6.2m0	
hxxp://uvyqifymelapuvoh.biz/3854f.u17	
hxxp://uvyqifymelapuvoh.biz/06hk3j.449	
Dridex	
hxxp://27.54.174.181/8qV578&$o@HU6Q6S/gz$J0l=iTTH 28%2CM/we20%3D	
hxxp://27.54.174.181/C4GyRx%7E@RY6x /M&N=sq/bW_ra4OTJ	
hxxp://27.54.174.181/gPvh+=GO/9RPPfk0%2CzXOYU%20/Vq8Ww/+a_m%7Ez	
hxxp://27.54.174.181/qE0my4KIz48Cf3H8wG%7Evpz=iJ%26fqMl%24m/46JoELp=GJww%3D%26Ib+Ar.y3 iu%2D1E/sso	
hxxp://27.54.174.181/kv7tig2s1vslfv&i_&/s&no%2Ds83%7E%2B+ns5%2D%3F+%20&1/kjx%26e8x=$.pfilr@s3j66%2D	
InstallCore	**Monetization**
hxxp://rp.any-file-opener.org/?pcrc=1559319553&v=2.0	hxxp://utouring.net/search/q/conducing
hxxp://rp.any-file-opener.org/?pcrc=1132521307&v=2.0	hxxp://utouring.net/go/u/1/r/1647
hxxp://rp.any-file-opener.org/?pcrc=1123945956&v=2.0	hxxp://utouring.net/go/u/0/r/2675
hxxp://rp.any-file-opener.org/?pcrc=1075608192&v=2.0	hxxp://utouring.net/search/f/1/q/refiles
hxxp://rp.any-file-opener.org/?pcrc=995719244&v=2.0	hxxp://utouring.net/search/f/1/q/refiles
Poweliks	
hxxp://31.184.194.39/query?version=1.7&sid=793&builddate=114&q=nitric+oxide+side+effects&ua=Mozilla%2F5 …&lr=7&ls=0	
hxxp://31.184.194.39/query?version=1.7&sid=793&builddate=114&q=weight+loss+success+stories&ua=Mozilla%2F5 …&lr=0&ls=0	
hxxp://31.184.194.39/query?version=1.7&sid=793&builddate=114&q=shoulder+pain&ua=Mozilla%2F5 …&lr=7&ls=2	
hxxp://31.184.194.39/query?version=1.7&sid=793&builddate=114&q=cheap+car+insurance&ua=Mozilla%2F5 …&lr=7&ls=2	
hxxp://31.184.194.39/query?version=1.7&sid=793&builddate=114&q=natural+testosterone+boosters&ua=Mozilla%2F5 …&lr=7&ls=2	
Zeus	
hxxp://130.185.106.28/m/IbQFdXVjiriLva4KHeNpWCmThrJBn3f34HNwsLVVsUmLXtsumSSPe/zzXtIu9SzwjI9zKlxdE …3RqvGzKN5	
hxxp://130.185.106.28/m/IbQJFUVjgZn4vx4KHeNpWCmThrJBn3f34HNwsLVVsUmLfkoPaSS+S+zzXtIu9SzwjI9zKlxdE …3vKwmk0oUi	
hxxp://130.185.106.28/m/IbQJFUVjiJwJBX4KHeNpWCmThrJBn3f34HNwsLVVsUmKH7ue2STvSkzzXtIu9SzwjI9zKlxdE …3vKwmk0oUi	
hxxp://130.185.106.28/m/IbQNtVVji5/7Yp4KHeNpWCmThrJBn3f34HNwsLVVsUmLz4sO6YRvOjzzXtIu9SzwjI9zKlxdE …3zB9057quqv	
hxxp://130.185.106.28/m/IbQG9VVjjSnDM94KHeNpWCmThrJBn3f34HNwsLVVsUmLXpt/+YRue8zzXtIu9SzwjI9zKlxdE …6iN5mt6Tj3	
Legitimate traffic	
hxxp://www.cnn.com/.a/1.73.0/js/vendor/usabilla.min.js	
hxxp://www.cnn.com/.element/ssi/auto/4.0/sect/MAIN/markets_wsod_expansion.html	
hxxp://www.cnn.com/.a/1.73.0/assets/sprite-s1dced3ff2b.png	
hxxp://www.cnn.com/.element/widget/video/videoapi/api/latest/js/CNNVideoBootstrapper.js	
hxxp://www.cnn.com/jsonp/video/nowPlayingSchedule.json?callback=nowPlayingScheduleCallbackWrapper&_=1422885578476	
Legitimate traffic	
hxxp://ads.adaptv.advertising.com/a/h/7g_doK40WLPMYHbkD9G2u7HSXjqzIaa7Bqhslod+u7iQl …&context=fullUrl%3Dpandora.com	
hxxp://ads.adaptv.advertising.com/crossdomain.xml	
hxxp://ads.advertising.com/411f1e96-3bde-4d85-b17e-63749e5f0695.js	
hxxp://ads.advertising.com/ids/411f1e96-3bde-4d85-b17e-63749e5f0695	
hxxp://ads.adaptv.advertising.com/applist?placementId=297920&key=&d.vw=1&orgId=8656&hostname=data.rtbfy.com	

MALICIOUS

LEGITIMATE

What is better?



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

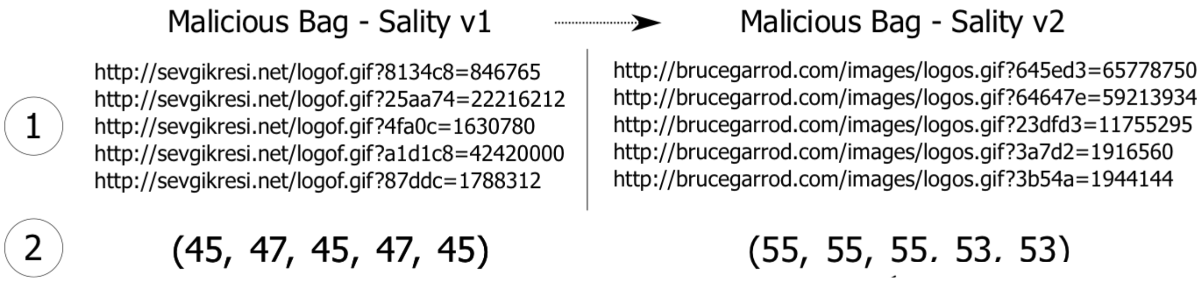
(55, 55, 55, 53, 53)

What do we want to represent?

What is common across malware categories?

Usually parameter names and subdomains are not stable, but the **URL structure** usually remains the same.

Example



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

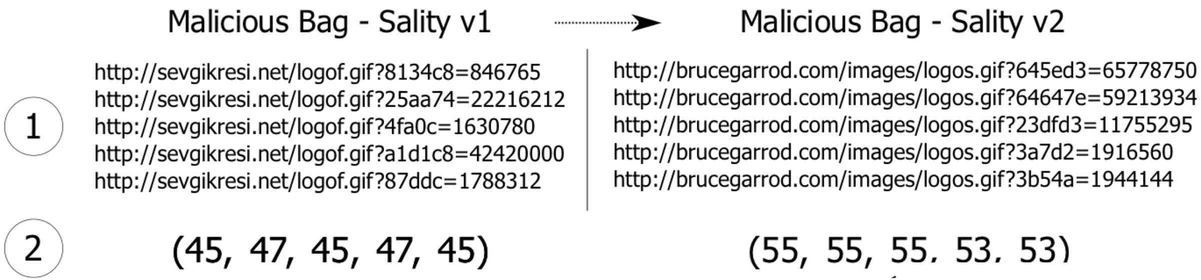
(45, 47, 45, 47, 45)

(55, 55, 55. 53. 53)

Malware dynamics:
It’s common for many mw categories and different from most of legitimate traffic

How?

Example



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

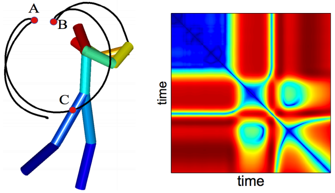
2

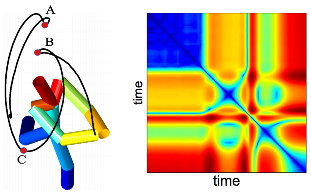
(45, 47, 45, 47, 45)

(55, 55, 55. 53. 53)

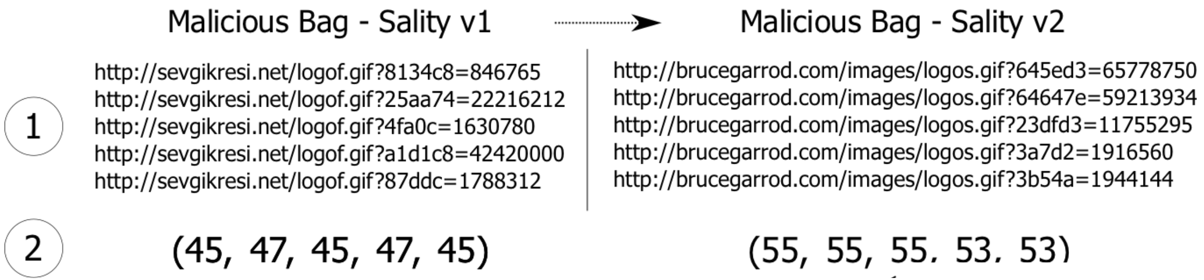
Malware dynamics:
It’s common for many mw categories and different from most of legitimate traffic

Parallel to action recognition:





Example



	Malicious Bag - Sality v1	Malicious Bag - Sality v2
1	http://sevgikresi.net/logof.gif?8134c8=846765 http://sevgikresi.net/logof.gif?25aa74=22216212 http://sevgikresi.net/logof.gif?4fa0c=1630780 http://sevgikresi.net/logof.gif?a1d1c8=42420000 http://sevgikresi.net/logof.gif?87ddc=1788312	http://brucegarrod.com/images/logos.gif?645ed3=65778750 http://brucegarrod.com/images/logos.gif?64647e=59213934 http://brucegarrod.com/images/logos.gif?23dfd3=11755295 http://brucegarrod.com/images/logos.gif?3a7d2=1916560 http://brucegarrod.com/images/logos.gif?3b54a=1944144
2	(45, 47, 45, 47, 45)	(55, 55, 55, 53, 53)

Malware dynamics:
It’s common for many mw categories and different from most of legitimate traffic

Parallel to action recognition:
Each bag (set of mw flows) is an action of mw

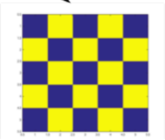
1 image ≈ 1 flow

Action recognition can be solved with self-similarity matrix

Example

	Malicious Bag - Sality v1	Malicious Bag - Sality v2
1	http://sevgikresi.net/logof.gif?8134c8=846765	http://brucegarrod.com/images/logos.gif?645ed3=65778750
	http://sevgikresi.net/logof.gif?25aa74=22216212	http://brucegarrod.com/images/logos.gif?64647e=59213934
	http://sevgikresi.net/logof.gif?4fa0c=1630780	http://brucegarrod.com/images/logos.gif?23dfd3=11755295
	http://sevgikresi.net/logof.gif?a1d1c8=42420000	http://brucegarrod.com/images/logos.gif?3a7d2=1916560
	http://sevgikresi.net/logof.gif?87ddc=1788312	http://brucegarrod.com/images/logos.gif?3b54a=1944144
2	(45, 47, 45, 47, 45)	(55, 55, 55, 53, 53)

3



Self-similarity matrix:

$$S_i^k = \begin{pmatrix} s_{11}^k & s_{12}^k & \dots & s_{1m}^k \\ s_{21}^k & s_{22}^k & \dots & s_{2m}^k \\ & & \vdots & \\ s_{m1}^k & s_{m2}^k & \dots & s_{mm}^k \end{pmatrix} \qquad s_{pq}^k = d(x_{pk}, x_{qk})$$

4

Example

Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

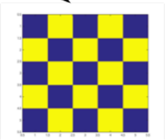
http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

3



Self-similarity matrix:

|45-45| |45-47| …

⋮

Shifting, scaling invariance

4

Example

Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

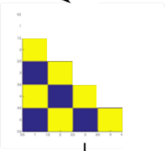
http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

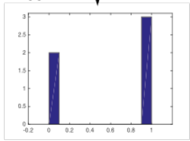
(55, 55, 55, 53, 53)

3



We are not interested in geometrical interpretation

h^S



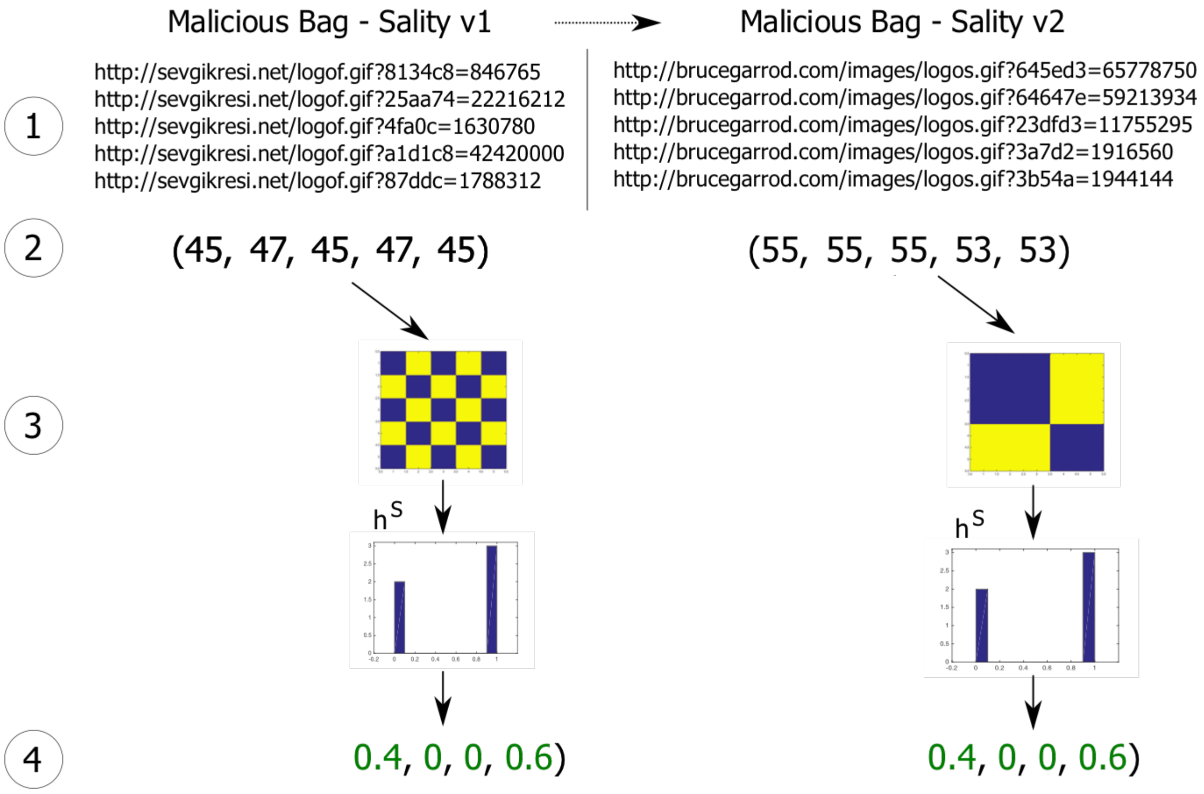
Histogram

4

0.4, 0, 0, 0.6)

Permutation and size invariance

Example



Malicious Bag - Sality v1

Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

3

h^S

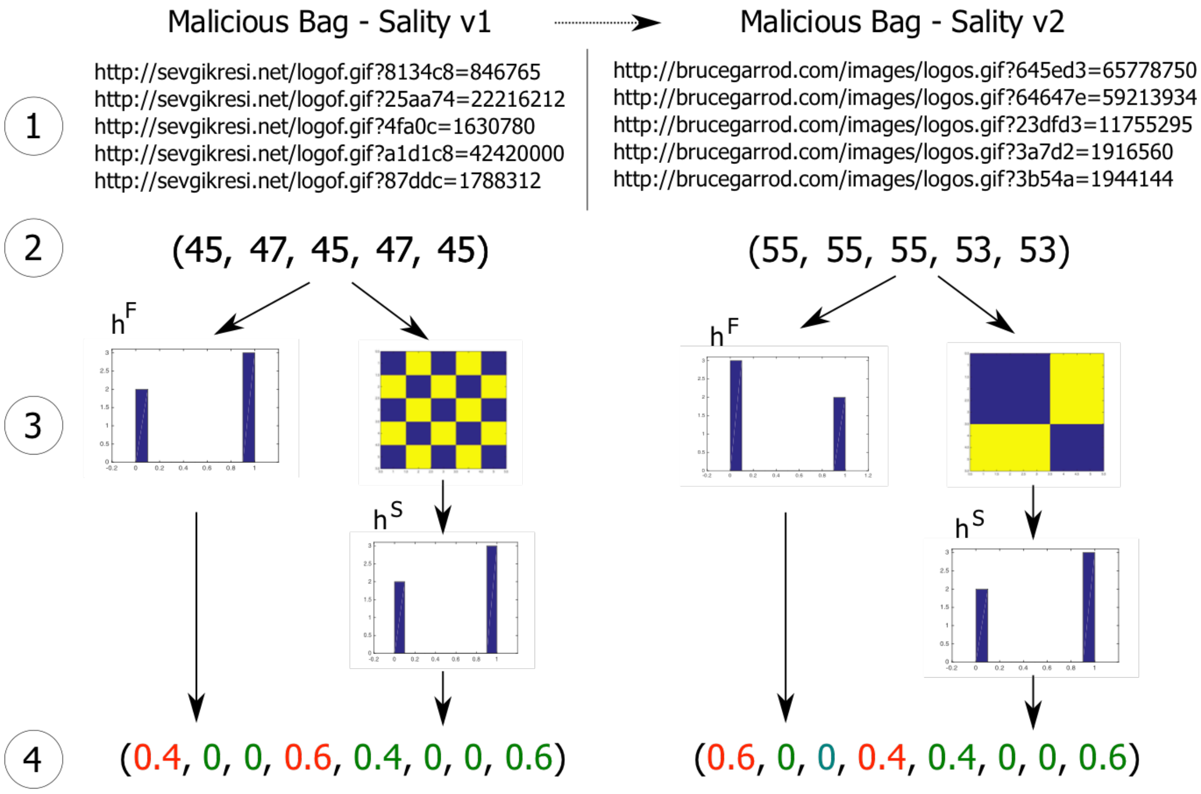
h^S

4

0.4, 0, 0, 0.6)

0.4, 0, 0, 0.6)

Example



Malicious Bag - Sality v1 → Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

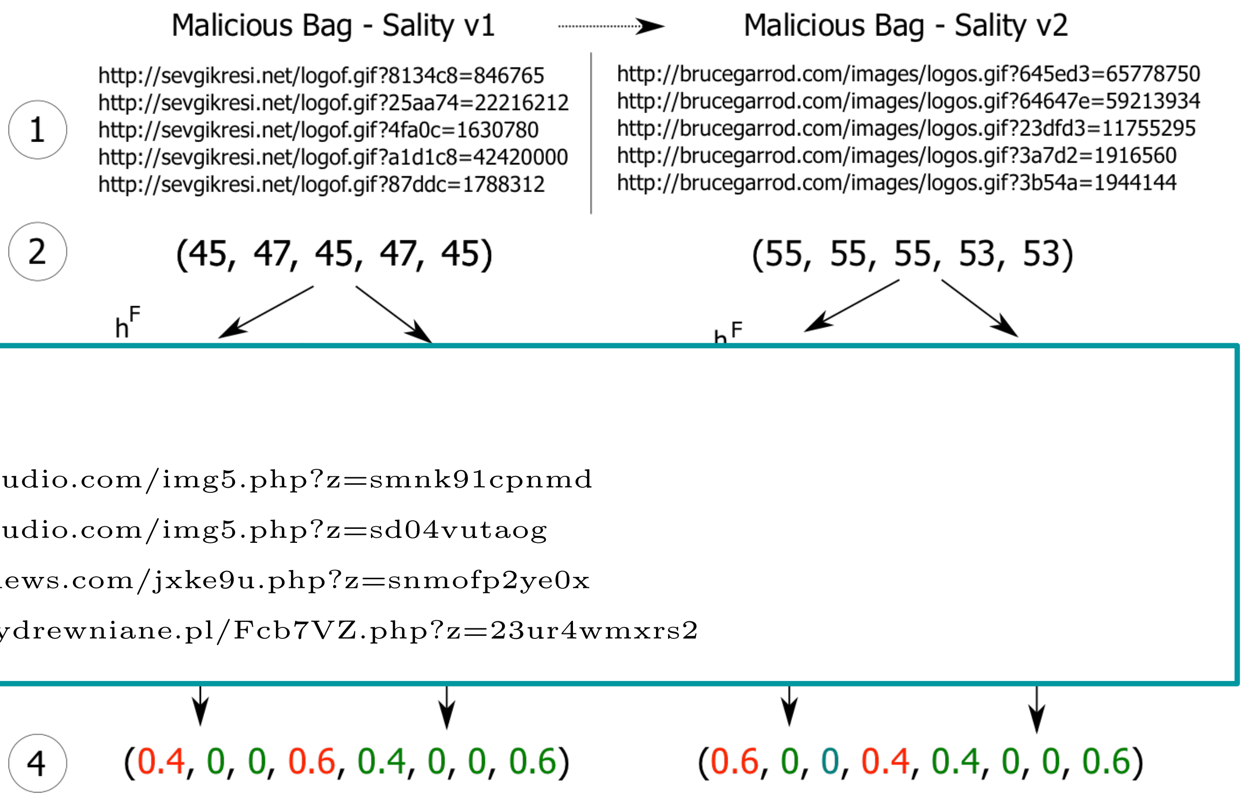
3

4

(0.4, 0, 0, 0.6, 0.4, 0, 0, 0.6)

(0.6, 0, 0, 0.4, 0.4, 0, 0, 0.6)

Example



Malicious Bag - Sality v1

Malicious Bag - Sality v2

1

http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

2

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

4

(0.4, 0, 0, 0.6, 0.4, 0, 0, 0.6)

(0.6, 0, 0, 0.4, 0.4, 0, 0, 0.6)

Cryptowall

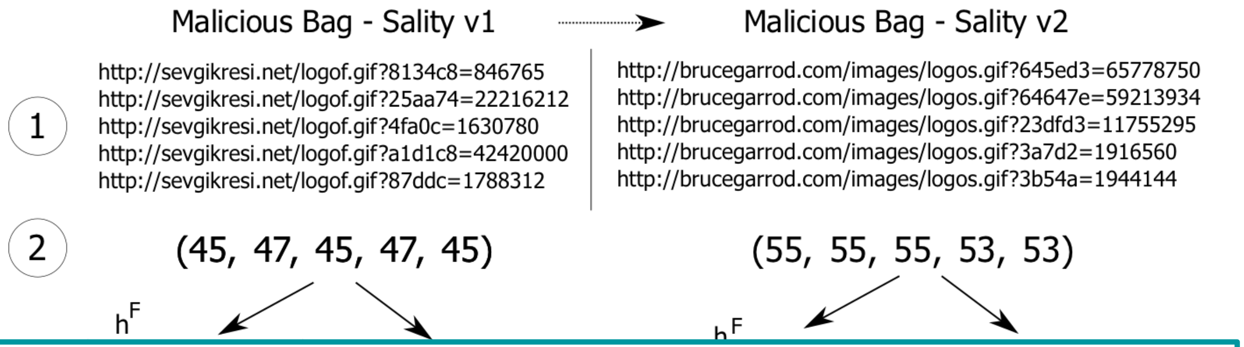
hxxp://zerosumstudio.com/img5.php?z=smnk91cpnmd

hxxp://zerosumstudio.com/img5.php?z=sd04vutaog

hxxp://zemamranews.com/jxke9u.php?z=snmofp2ye0x

hxxp://balustradydrewniane.pl/Fcb7VZ.php?z=23ur4wmxrs2

Example



http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

(45, 47, 45, 47, 45)

(55, 55, 55, 53, 53)

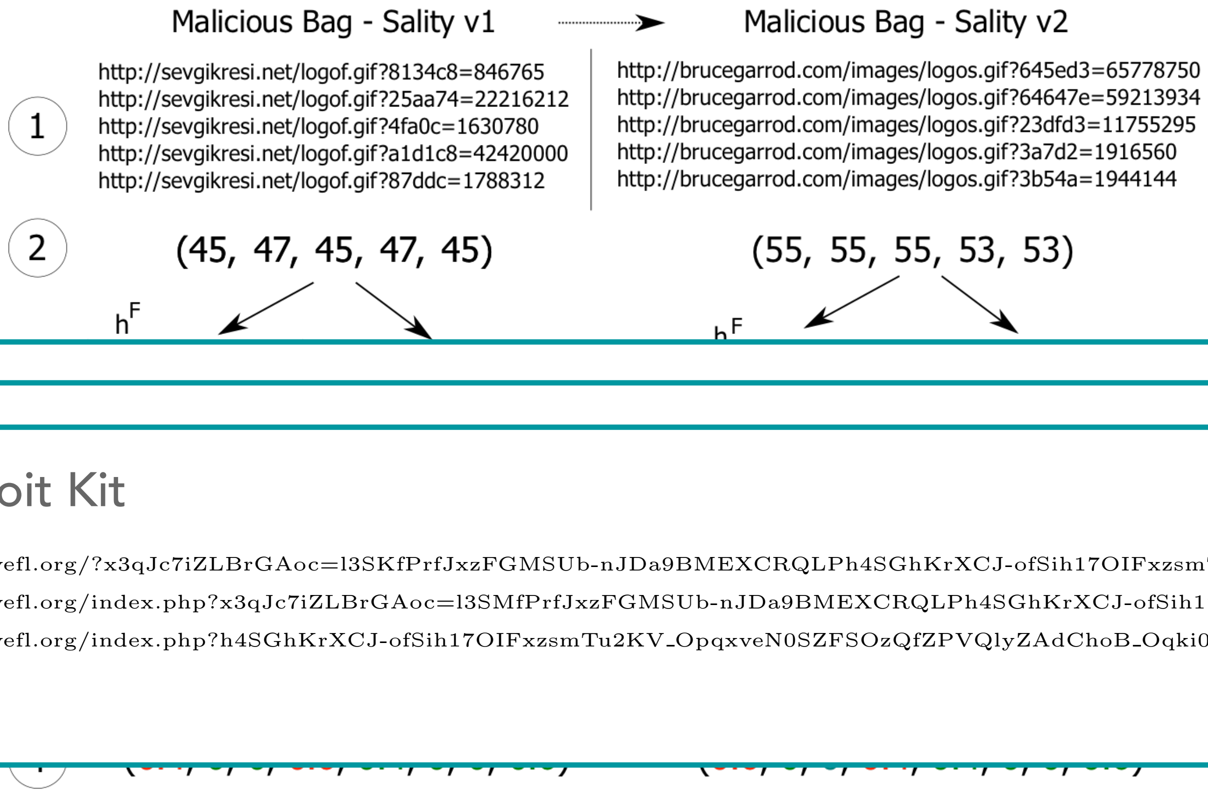
DNS Changer

hxxp://ukhealer.net/u/?a=KELQFAJusqu6Gd33DB0T1zATPwoXsmYFciyO9THSYS7na3zZfVczZ8GzHHydLYn8hVyiy1l0...
hxxp://sethealer.com/u/?a=L4ZTRAn2VVC9F_-BkobTaxsNyaqCKxReHIOOWoVFd–YZxFkES4Y_mBgSCaN_1K1rWdeM...
hxxp://sethealer.net/u/?a=qF1coIn2VVE3OFYDC1NXrm24fgDShSqjFsut7gMXRymFe3zZuFTQPw1lI4X6t2MQIMntv2It...

4 (0.4, 0, 0, 0.6, 0.4, 0, 0, 0.6)

(0.6, 0, 0, 0.4, 0.4, 0, 0, 0.6)

Example



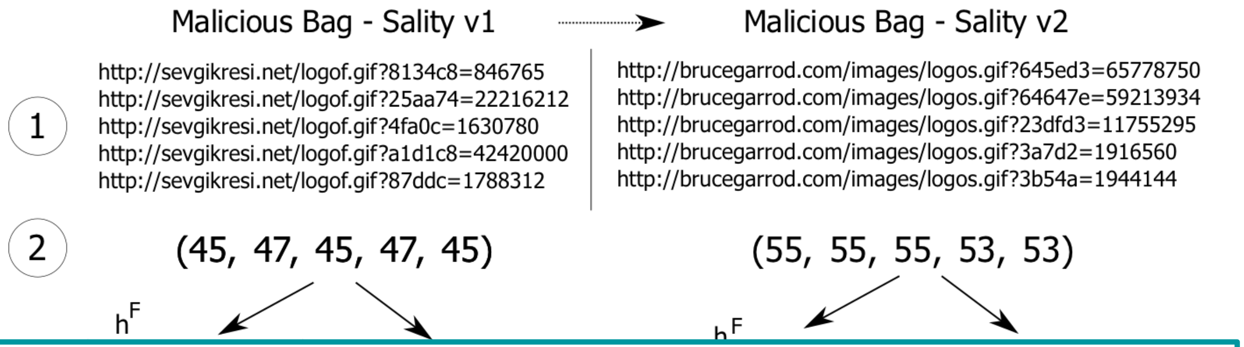
http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312

http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144

Rig Exploit Kit

hxxp://ds.revivefl.org/?x3qJc7iZLBrGAoc=l3SKfPrfJxzFGMSUb-nJDa9BMEXCRQLPh4SGhKrXCJ-ofSih17OIFxzsmT...
hxxp://ds.revivefl.org/index.php?x3qJc7iZLBrGAoc=l3SMfPrfJxzFGMSUb-nJDa9BMEXCRQLPh4SGhKrXCJ-ofSih17...
hxxp://ds.revivefl.org/index.php?h4SGhKrXCJ-ofSih17OIFxzsmTu2KV_OpqxveN0SZFSOzQfZPVQlyZAdChoB_Oqki0v...

Example



Malicious Bag - Sality v1

```
http://sevgikresi.net/logof.gif?8134c8=846765
http://sevgikresi.net/logof.gif?25aa74=22216212
http://sevgikresi.net/logof.gif?4fa0c=1630780
http://sevgikresi.net/logof.gif?a1d1c8=42420000
http://sevgikresi.net/logof.gif?87ddc=1788312
```

(45, 47, 45, 47, 45)

Malicious Bag - Sality v2

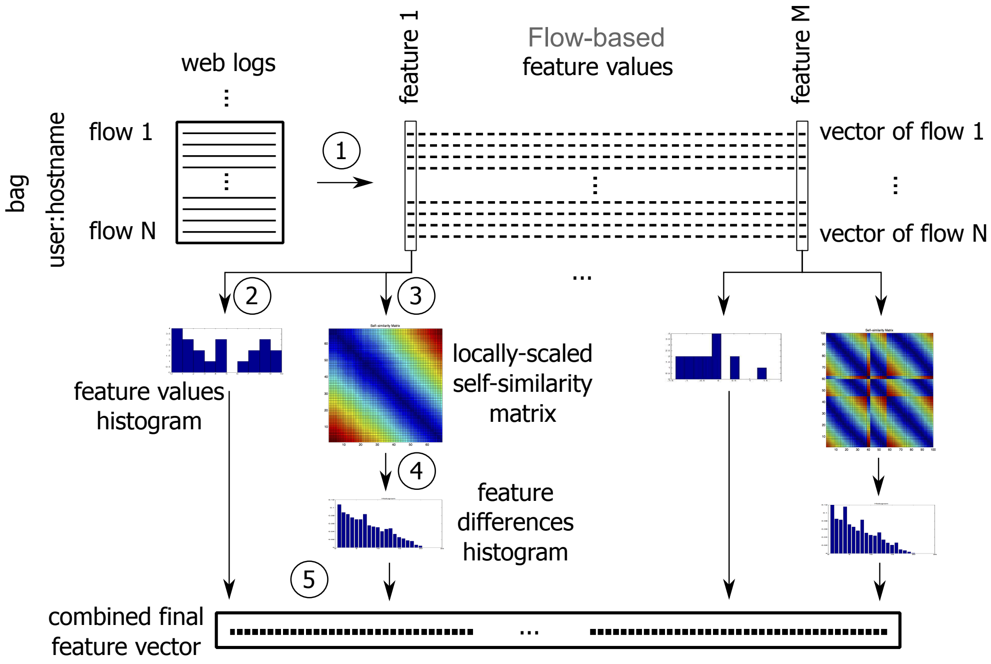
```
http://brucegarrod.com/images/logos.gif?645ed3=65778750
http://brucegarrod.com/images/logos.gif?64647e=59213934
http://brucegarrod.com/images/logos.gif?23dfd3=11755295
http://brucegarrod.com/images/logos.gif?3a7d2=1916560
http://brucegarrod.com/images/logos.gif?3b54a=1944144
```

(55, 55, 55, 53, 53)

Dridex

```
hxxp://27.54.174.181/8qV578&$o@HU6Q6S/gz$J0l=iTTH 28%2CM/we20%3D
hxxp://27.54.174.181/C4GyRx%7E@RY6x /M&N=sq/bW_ra4OTJ
hxxp://27.54.174.181/gPvh+=GO/9RPPfk0%2CzXOYU%20/Vq8Ww/+a_m%7Ez
hxxp://27.54.174.181/qE0my4KIz48Cf3H8wG%7Evpz=iJ%26fqMl%24m/46JoELp=GJww%3D%26Ib+Ar.y3  iu%2D1E/sso
hxxp://27.54.174.181/kv7tig2s1vslfv&i_&/s&no%2Ds83%7E%2B+ns5%2D%3F+%20&1/kjx%26e8x=$.pfilr@s3j66%2D
```

Overview



1 – create bag
+ extract flow-based feature vectors

2 – create feature values histogram

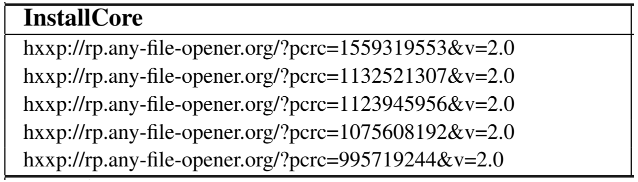
3 – create self-similarity matrix

4 – create feature differences histogram

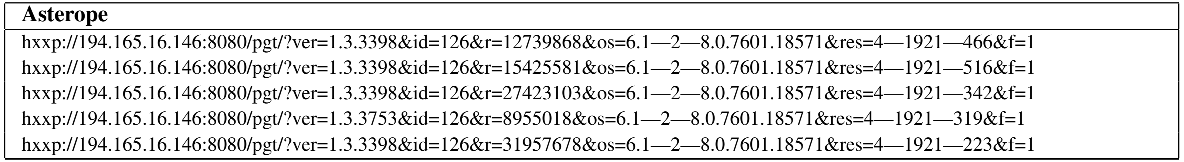
5 – combine into final feature vector

Invariant to the following changes:

- Malicious code, payload, obfuscation
- Server or hostname
- URL path or filename
- Names, values, or number of URL parameters
- Encoded URL content
- Number of flows
- Thinking time
- Ordering of flows
- Size of flows



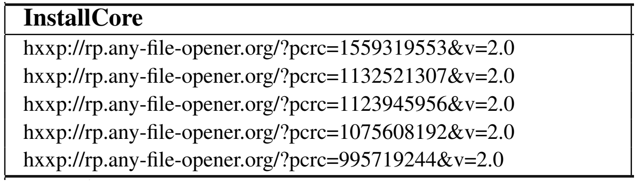
InstallCore
hxxp://rp.any-file-opener.org/?pcrc=1559319553&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=1132521307&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=1123945956&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=1075608192&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=995719244&v=2.0



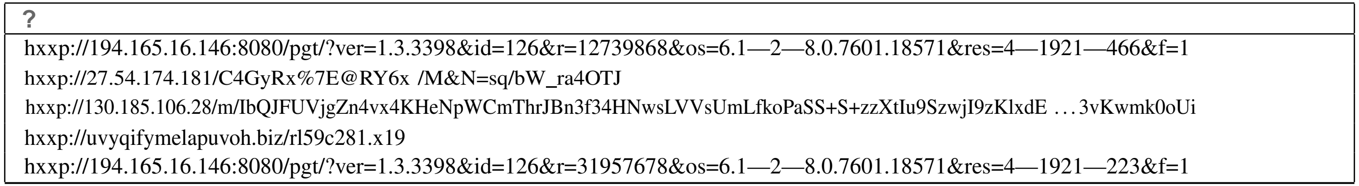
Asterope
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=12739868&os=6.1—2—8.0.7601.18571&res=4—1921—466&f=1
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=15425581&os=6.1—2—8.0.7601.18571&res=4—1921—516&f=1
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=27423103&os=6.1—2—8.0.7601.18571&res=4—1921—342&f=1
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3753&id=126&r=8955018&os=6.1—2—8.0.7601.18571&res=4—1921—319&f=1
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=31957678&os=6.1—2—8.0.7601.18571&res=4—1921—223&f=1

Not invariant to the following changes:

- Static behavior is not considered
- Multiple behaviors in a bag
- Encrypted HTTPS traffic
- Real-time changes and fast evolution



InstallCore
hxxp://rp.any-file-opener.org/?pcrc=1559319553&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=1132521307&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=1123945956&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=1075608192&v=2.0
hxxp://rp.any-file-opener.org/?pcrc=995719244&v=2.0

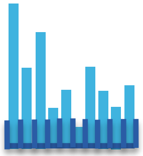


?
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=12739868&os=6.1—2—8.0.7601.18571&res=4—1921—466&f=1
hxxp://27.54.174.181/C4GyRx%7E@RY6x /M&N=sq/bW_ra4OTJ
hxxp://130.185.106.28/m/IbQJFUVjgZn4vx4KHeNpWCmThrJBn3f34HNwsLVVsUmLfkoPaSS+S+zzXtIu9SzwjI9zKlxdE …3vKwmk0oUi
hxxp://uvyqifymelapuvoh.biz/rl59c281.x19
hxxp://194.165.16.146:8080/pgt/?ver=1.3.3398&id=126&r=31957678&os=6.1—2—8.0.7601.18571&res=4—1921—223&f=1

Parameters of the Representation

Number of bins, all of them are equidistant

➔ major impact on the results



How to choose the correct number?

We want to learn the parameters automatically from the data

Proposed Optimization Algorithm

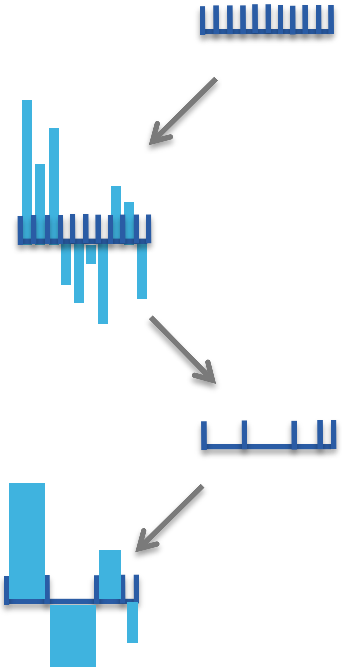
1) Define the initial number of mini-bins (256)

2) Find a set of weights by solving:

$$\min_{\boldsymbol{w}\in\mathbb{R}^{b\cdot p}, w_0\in\mathbb{R}} \left[\underbrace{\gamma \sum_{i=1}^{n}\sum_{j=1}^{b-1} |w_{i,j}-w_{i,j+1}|}_{\text{merging}} + \underbrace{\frac{1}{m}\sum_{i=1}^{m} \max\left\{0, 1-y^i\langle \boldsymbol{\phi}(\boldsymbol{z}^i;\boldsymbol{\theta}), \boldsymbol{w}\rangle\right\}}_{\text{hinge loss}} \right]$$

3) Create new bins (by merging mini-bins)

4) Solve standard SVM with new bins



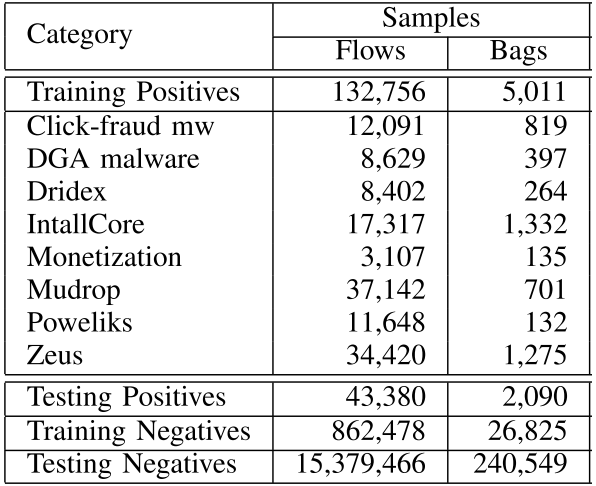
Experiments – Dataset Description

Training data:

- positives: 8 mw categories
- negatives: company A

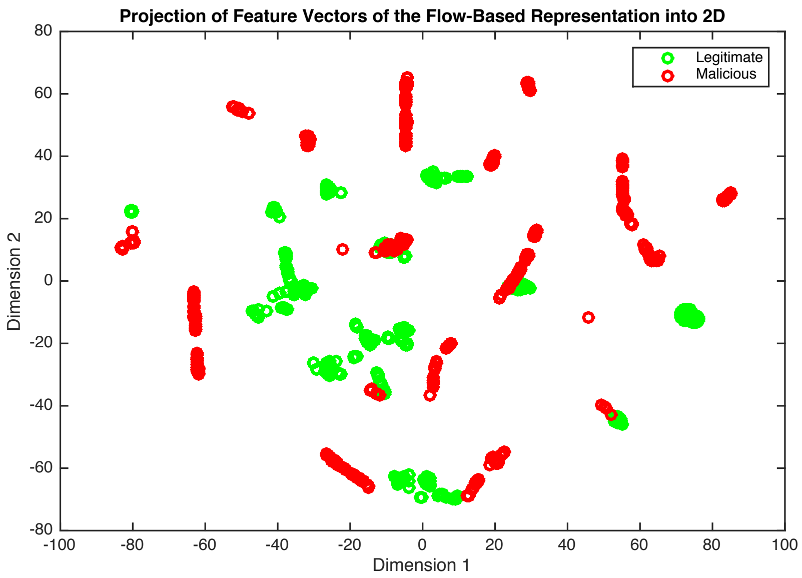
Testing data:

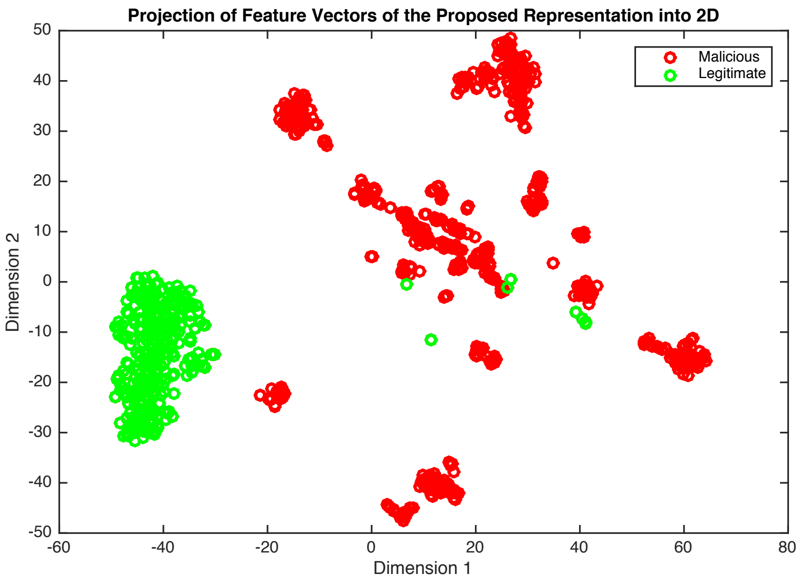
- positives: 24 unseen mw categories
- negatives: company B



Category	Samples	
	Flows	Bags
Training Positives	132,756	5,011
Click-fraud mw	12,091	819
DGA malware	8,629	397
Dridex	8,402	264
IntallCore	17,317	1,332
Monetization	3,107	135
Mudrop	37,142	701
Poweliks	11,648	132
Zeus	34,420	1,275
Testing Positives	43,380	2,090
Training Negatives	862,478	26,825
Testing Negatives	15,379,466	240,549

Experiments – 2D projection (t-SNE)





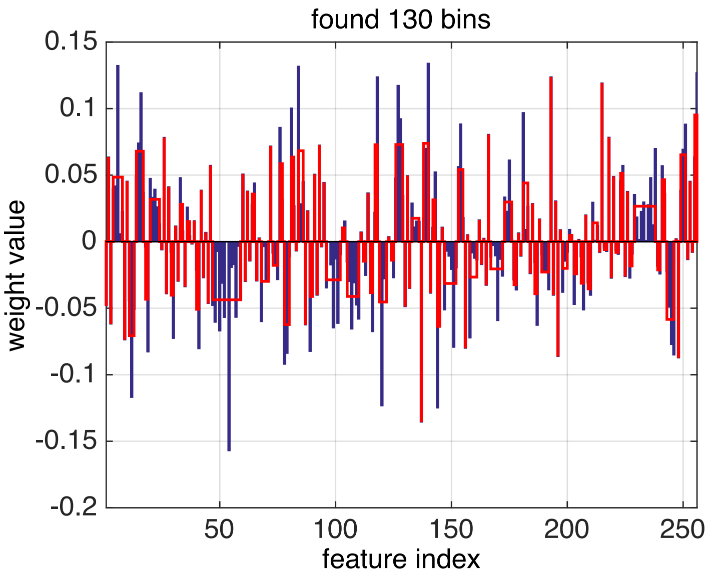
Flow-based features

Good for individual malware families

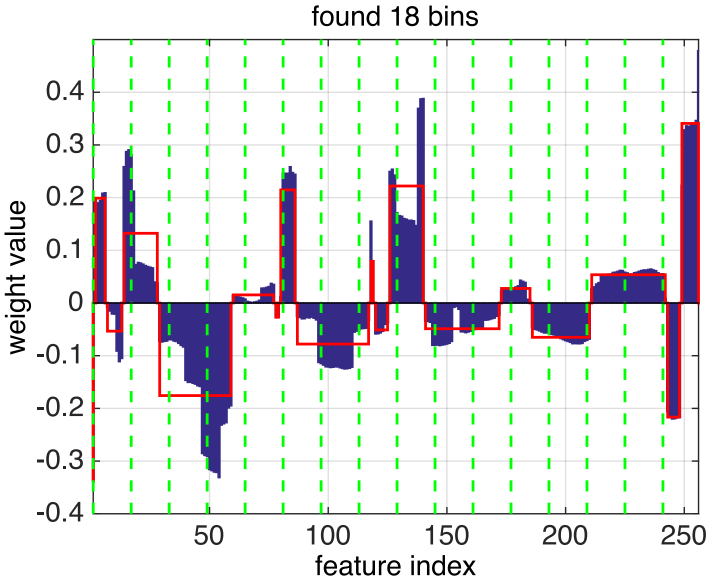
Bag Invariant Features

Good for general malware

Learning bins from 256 equidistant mini-bins



Standard SVM

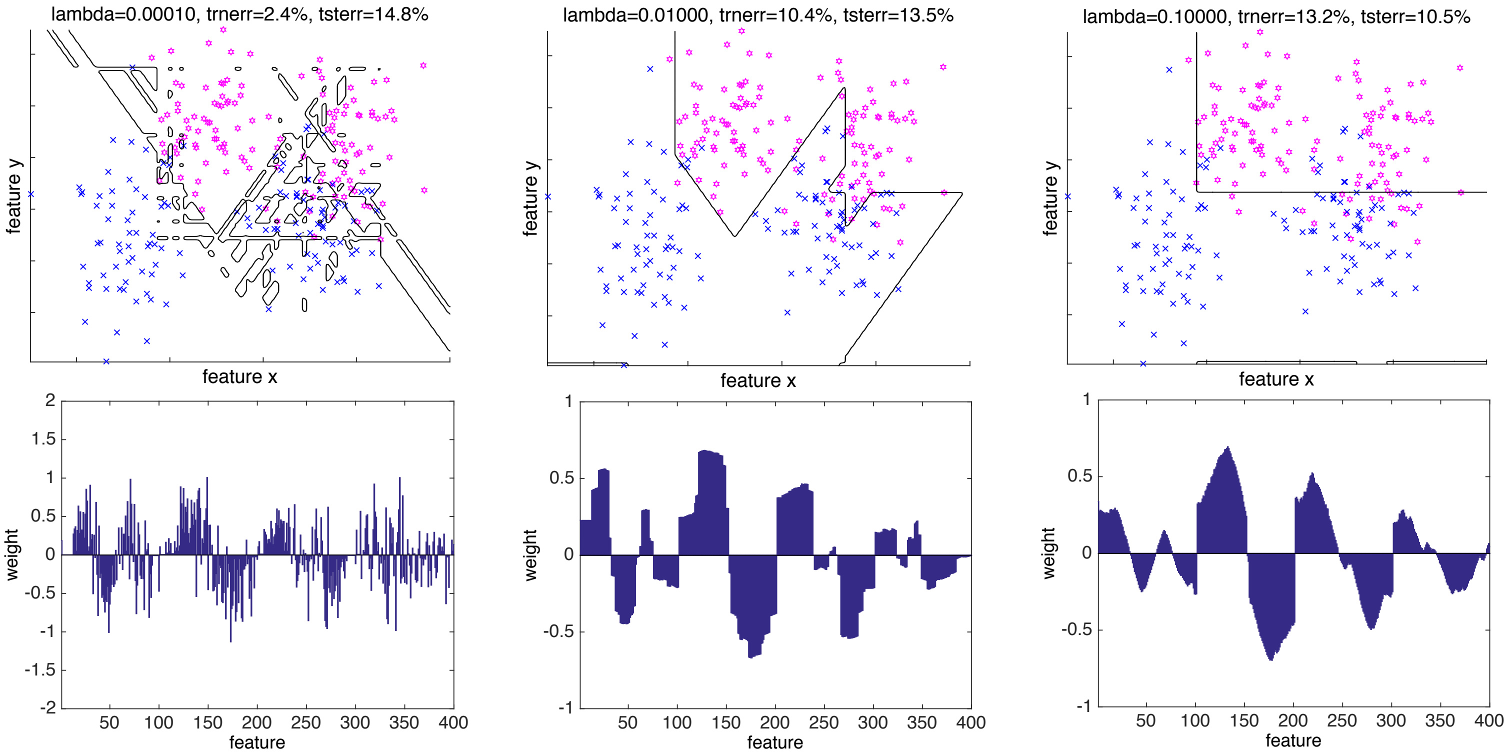


Modified SVM with merging

Blue bars… weights
Red lines… bins

All mini-bins with the same weight sign create new bin.

Optimizing Decision Boundary



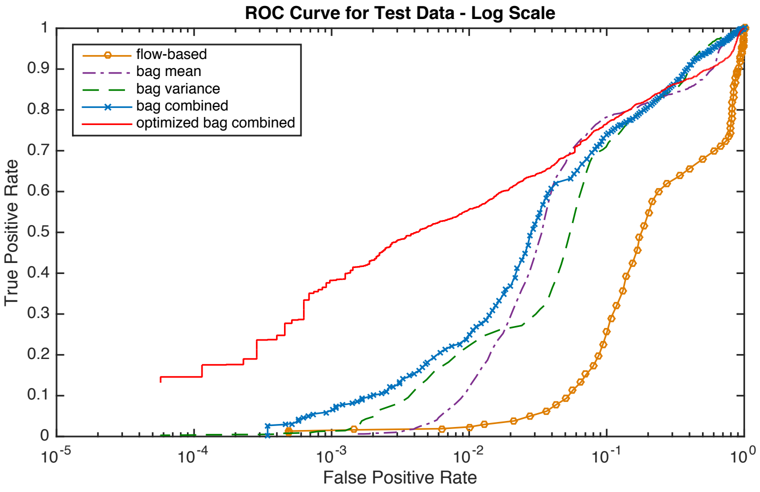
lambda=0.00010, trnerr=2.4%, tsterr=14.8%

lambda=0.01000, trnerr=10.4%, tsterr=13.5%

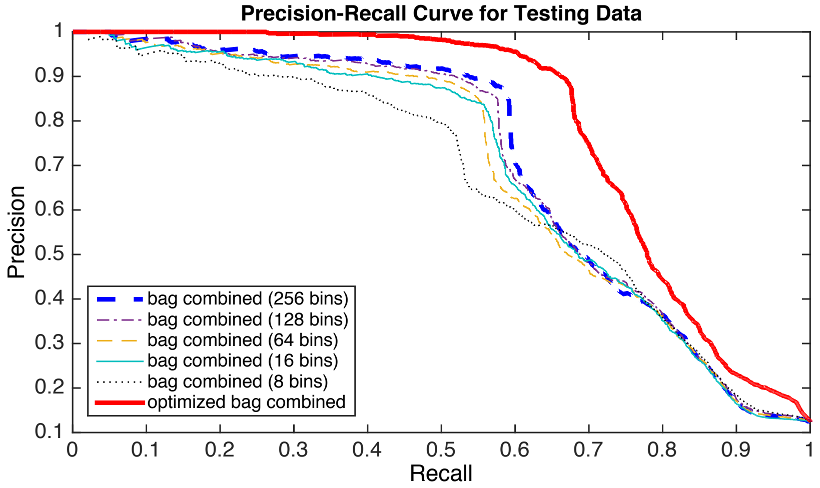
lambda=0.10000, trnerr=13.2%, tsterr=10.5%

Efficacy Results – Unseen Malware

ROC Curve – log scale



Precision – Recall Curve



90% precision, 67% recall

Conclusion and Future Work

- Flaws of flow-based representation

- New representation based on the dynamics of malware bags

- New optimization method that learns the parameters of the **representation automatically from the data**

- In progress:
  - Modified version for HTTPS

Thank you

Q&A ?