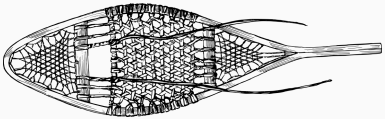
Melting the Snow

Using Active DNS Measurements to Detect Snowshoe Spam Domains

Olivier van der Toorn

November 13, 2018



University of Twente, Design and Analysis of Communication Systems

Introduction

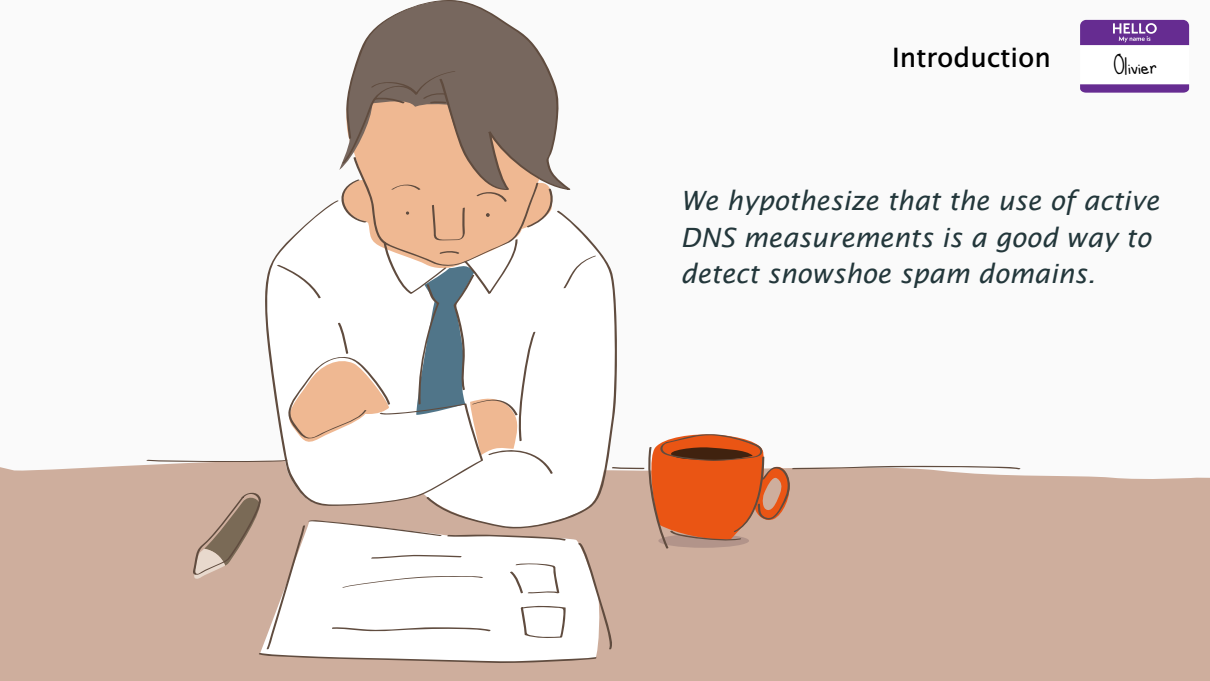
matrix @lordievader:corellian.student.utwente.nl
o.i.vandertoorn@utwente.nl
https://www.tide-project.nl/

Introduction

HELLO
My name is
Olivier

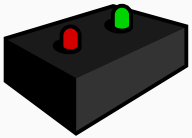
Introduction

Introduction

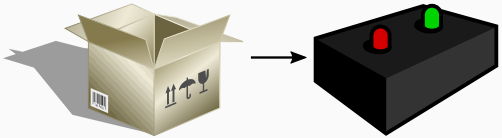


We hypothesize that the use of active DNS measurements is a good way to detect snowshoe spam domains.

Overview

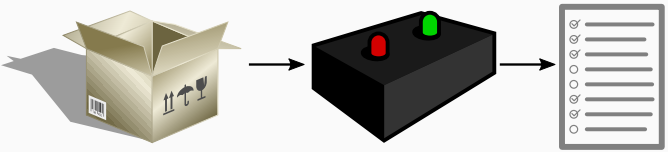


Black box



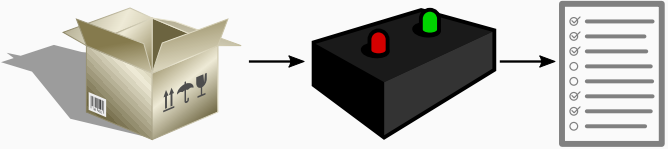
Box of domains

Black box

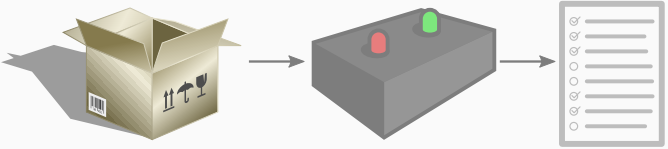


Box of domains Black box Notepad

A Closer Look



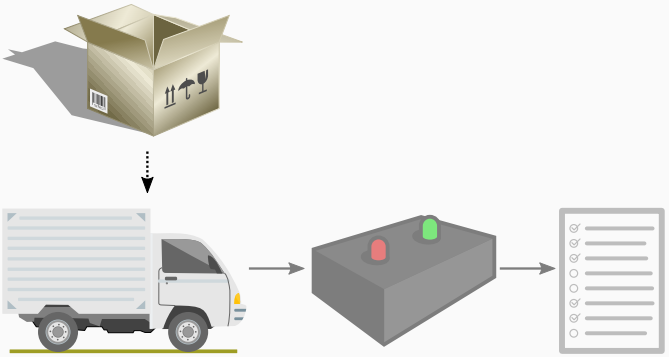
Box of domains Black box Notepad



Box of domains

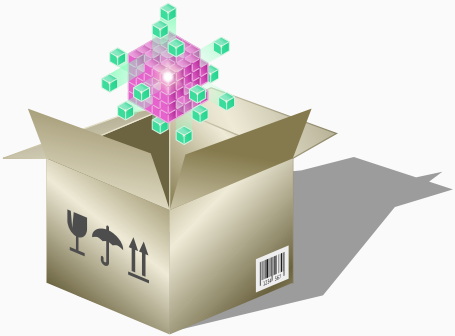
Black box

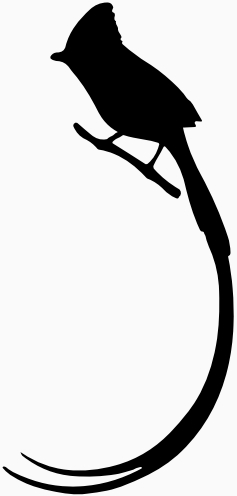
Notepad



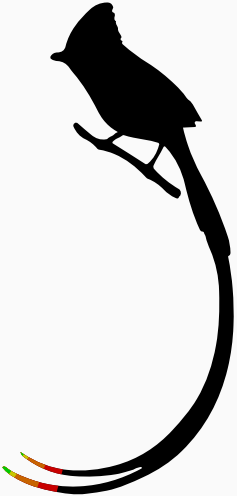
OpenINTEL Black box Notepad

OpenINTEL

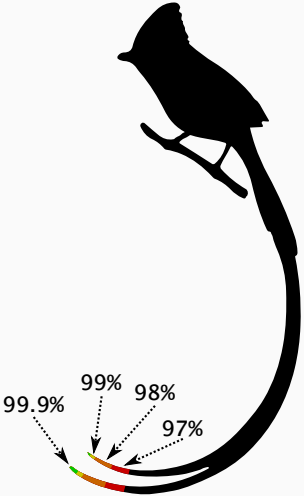




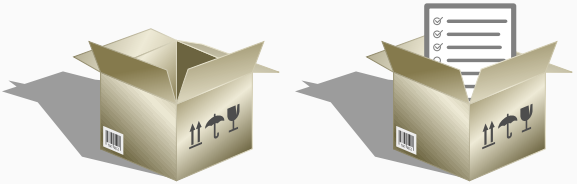
The long tail of the DNS



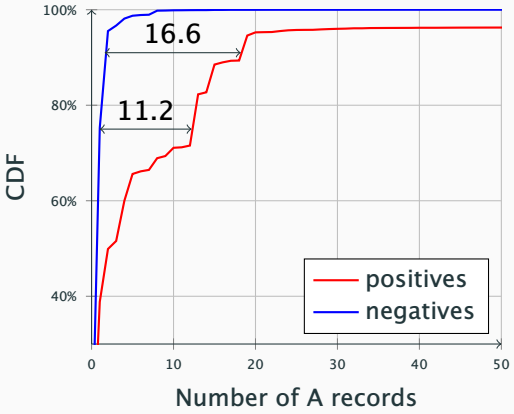
The **long** tail of the DNS

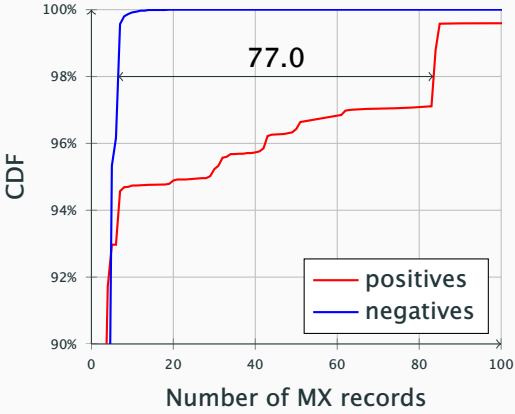


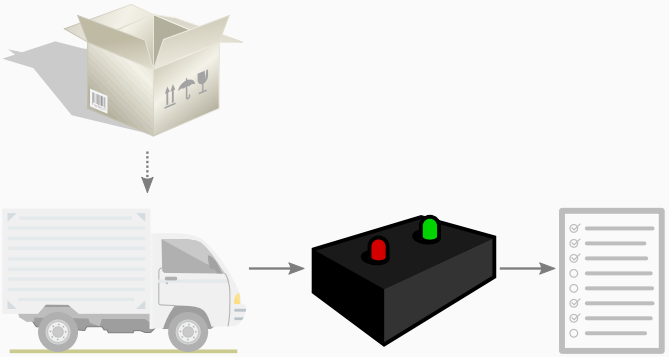
The **long** tail of the DNS



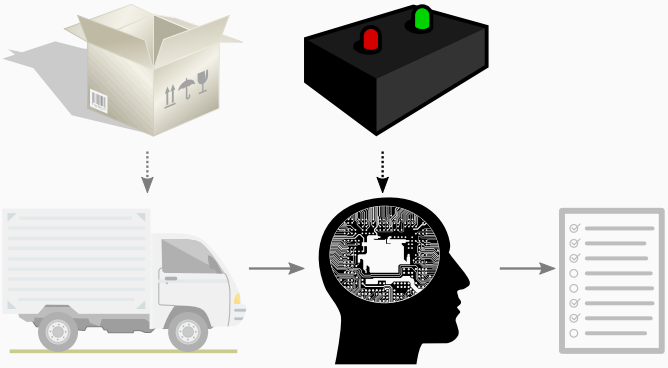
OpenINTEL


100%
80%
60%
40%
CDF
16.6
11.2
positives
negatives
0
10
20
30
40
50
Number of A records


100%
98%
96%
94%
92%
90%
CDF
77.0
positives
negatives
0
20
40
60
80
100
Number of MX records



OpenINTEL Black box Notepad

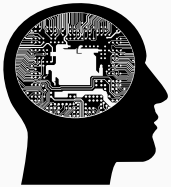


OpenINTEL

Machine Learning

Notepad

Machine
Learning



Machine
Learning

	Spam		Ham	
Type	**TP**	**FN**	**FP**	**TN**

	Spam		Ham	
Type	**TP**	**FN**	**FP**	**TN**
BernoulliNB	12995	1535	2507	8344
GradientBoostingClassifier	12645	1885	9605	1246
MultinomialNB	12179	2351	1397	9454
RandomForestClassifier	11156	3374	1488	9363
MLPClassifier	7273	7257	707	10144
DecisionTreeClassifier	6279	8251	695	10156
AdaBoostClassifier	5971	8559	164	10687
KNeighborsClassifier	4562	9968	676	10175
SGDClassifier	3599	10931	674	10177

	Spam		Ham	
Type	**TP**	**FN**	**FP**	**TN**
RadiusNeighborsClassifier	13318	1212	2367	8484
BernoulliNB	12995	1535	2507	8344
GradientBoostingClassifier	12645	1885	9605	1246
MultinomialNB	12179	2351	1397	9454
RandomForestClassifier	11156	3374	1488	9363
MLPClassifier	7273	7257	707	10144
DecisionTreeClassifier	6279	8251	695	10156
AdaBoostClassifier	5971	8559	164	10687
KNeighborsClassifier	4562	9968	676	10175
SGDClassifier	3599	10931	674	10177

	Spam		Ham	
Type	**TP**	**FN**	**FP**	**TN**
GaussianNB	13330	1200	2075	8776
RadiusNeighborsClassifier	13318	1212	2367	8484
BernoulliNB	12995	1535	2507	8344
GradientBoostingClassifier	12645	1885	9605	1246
MultinomialNB	12179	2351	1397	9454
RandomForestClassifier	11156	3374	1488	9363
MLPClassifier	7273	7257	707	10144
DecisionTreeClassifier	6279	8251	695	10156
AdaBoostClassifier	5971	8559	164	10687
KNeighborsClassifier	4562	9968	676	10175
SGDClassifier	3599	10931	674	10177

	Spam		Ham	
Type	**TP**	**FN**	**FP**	**TN**
SVC	13449	1081	2339	8512
GaussianNB	13330	1200	2075	8776
RadiusNeighborsClassifier	13318	1212	2367	8484
BernoulliNB	12995	1535	2507	8344
GradientBoostingClassifier	12645	1885	9605	1246
MultinomialNB	12179	2351	1397	9454
RandomForestClassifier	11156	3374	1488	9363
MLPClassifier	7273	7257	707	10144
DecisionTreeClassifier	6279	8251	695	10156
AdaBoostClassifier	5971	8559	164	10687
KNeighborsClassifier	4562	9968	676	10175
SGDClassifier	3599	10931	674	10177

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

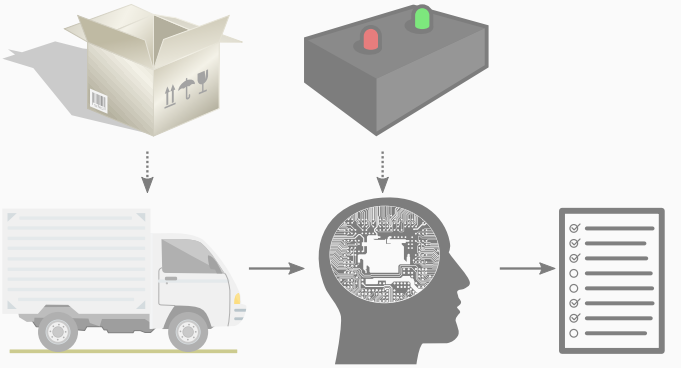
Machine Learning

	Spam		Ham		
Type	**TP**	**FN**	**FP**	**TN**	**Precision**
SVC	13449	1081	2339	8512	85.18%
GaussianNB	13330	1200	2075	8776	86.53%
RadiusNeighborsClassifier	13318	1212	2367	8484	84.90%
BernoulliNB	12995	1535	2507	8344	83.82%
GradientBoostingClassifier	12645	1885	9605	1246	56.83%
MultinomialNB	12179	2351	1397	9454	89.70%
RandomForestClassifier	11156	3374	1488	9363	88.23%
MLPClassifier	7273	7257	707	10144	91.14%
DecisionTreeClassifier	6279	8251	695	10156	90.03%
AdaBoostClassifier	5971	8559	164	10687	97.32%
KNeighborsClassifier	4562	9968	676	10175	87.09%
SGDClassifier	3599	10931	674	10177	84.22%

	Spam		Ham		
Type	**TP**	**FN**	**FP**	**TN**	**Precision**
AdaBoostClassifier	5971	8559	164	10687	97.32%
MLPClassifier	7273	7257	707	10144	91.14%
DecisionTreeClassifier	6279	8251	695	10156	90.03%
MultinomialNB	12179	2351	1397	9454	89.70%
RandomForestClassifier	11156	3374	1488	9363	88.23%
KNeighborsClassifier	4562	9968	676	10175	87.09%
GaussianNB	13330	1200	2075	8776	86.53%
SVC	13449	1081	2339	8512	85.18%
RadiusNeighborsClassifier	13318	1212	2367	8484	84.90%
SGDClassifier	3599	10931	674	10177	84.22%
BernoulliNB	12995	1535	2507	8344	83.82%
GradientBoostingClassifier	12645	1885	9605	1246	56.83%

	Spam		Ham		
Type	**TP**	**FN**	**FP**	**TN**	**Precision**
AdaBoostClassifier Improved	6688	7842	110	10741	98.38%
AdaBoostClassifier	5971	8559	164	10687	97.32%
MLPClassifier	7273	7257	707	10144	91.14%
DecisionTreeClassifier	6279	8251	695	10156	90.03%
MultinomialNB	12179	2351	1397	9454	89.70%
RandomForestClassifier	11156	3374	1488	9363	88.23%
KNeighborsClassifier	4562	9968	676	10175	87.09%
GaussianNB	13330	1200	2075	8776	86.53%
SVC	13449	1081	2339	8512	85.18%
RadiusNeighborsClassifier	13318	1212	2367	8484	84.90%
SGDClassifier	3599	10931	674	10177	84.22%
BernoulliNB	12995	1535	2507	8344	83.82%
GradientBoostingClassifier	12645	1885	9605	1246	56.83%

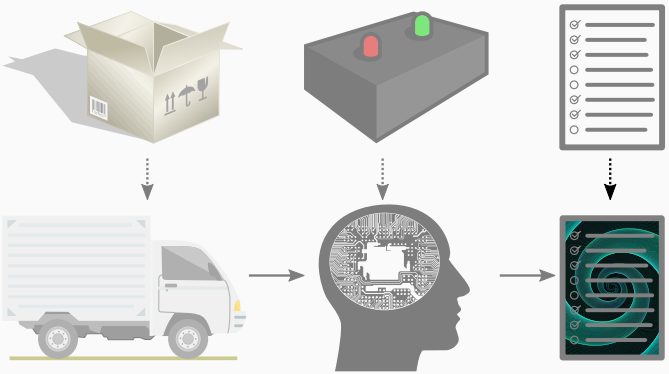
	Spam		Ham		
Type	TP	FN	FP	TN	Precision
AdaBoostClassifier Improved	6688	7842	110	10741	98.38%
AdaBoostClassifier	5971	8559	164	10687	97.32%
MLPClassifier	7273	7257	707	10144	91.14%
DecisionTreeClassifier	6279	8251	695	10156	90.03%
MultinomialNB	12179	2351	1397	9454	89.70%
RandomForestClassifier	11156	3374	1488	9363	88.23%
KNeighborsClassifier	4562	9968	676	10175	87.09%
GaussianNB	13330	1200	2075	8776	86.53%
SVC	13449	1081	2339	8512	85.18%
RadiusNeighborsClassifier	13318	1212	2367	8484	84.90%
SGDClassifier	3599	10931	674	10177	84.22%
BernoulliNB	12995	1535	2507	8344	83.82%
GradientBoostingClassifier	12645	1885	9605	1246	56.83%



OpenINTEL

Machine Learning

Notepad



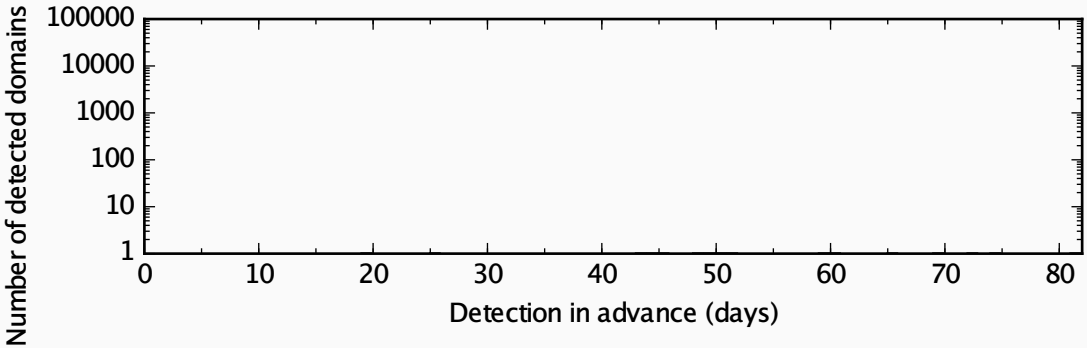
OpenINTEL

Machine Learning

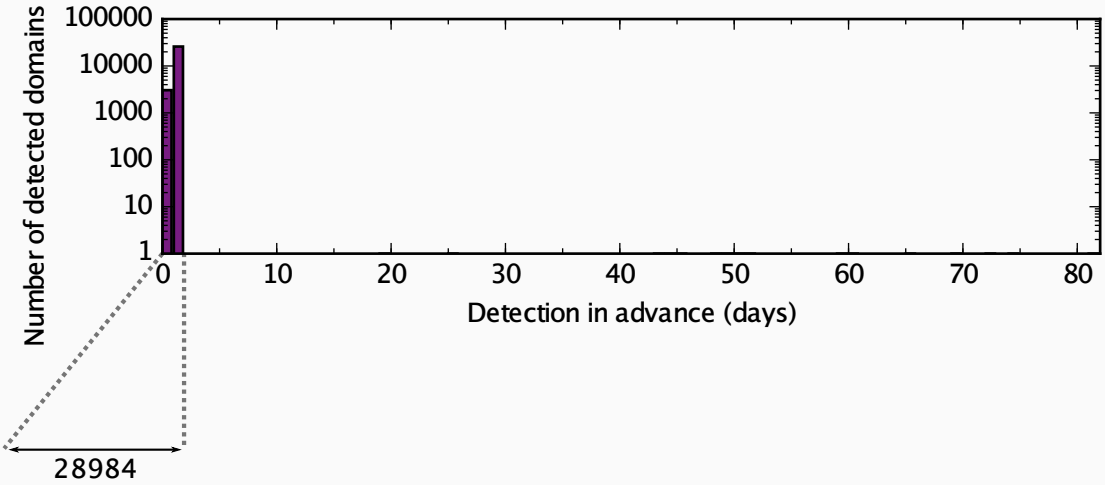
Realtime Blackhole List (RBL)

Realtime Blackhole List (RBL)

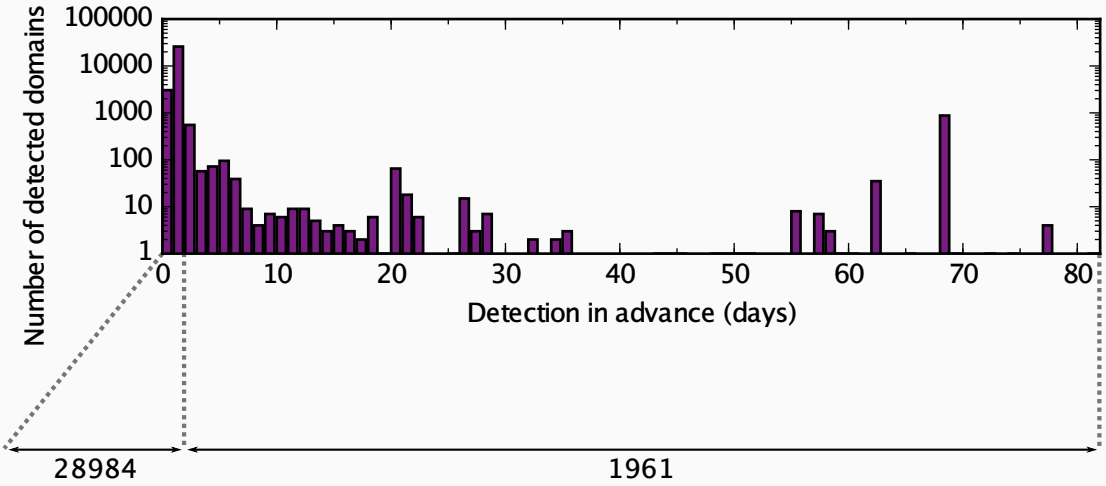
Realtime Blackhole List (RBL)



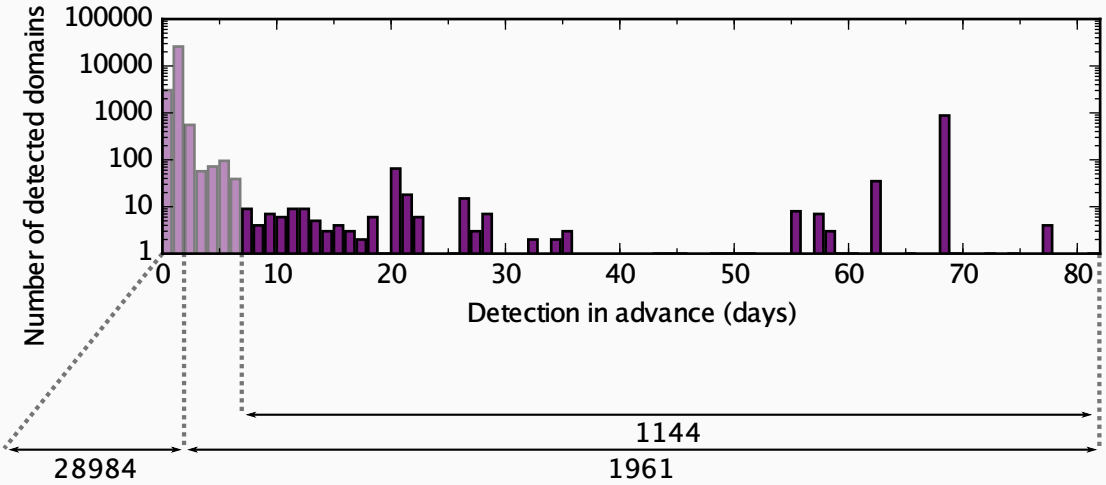
Realtime Blackhole List (RBL)



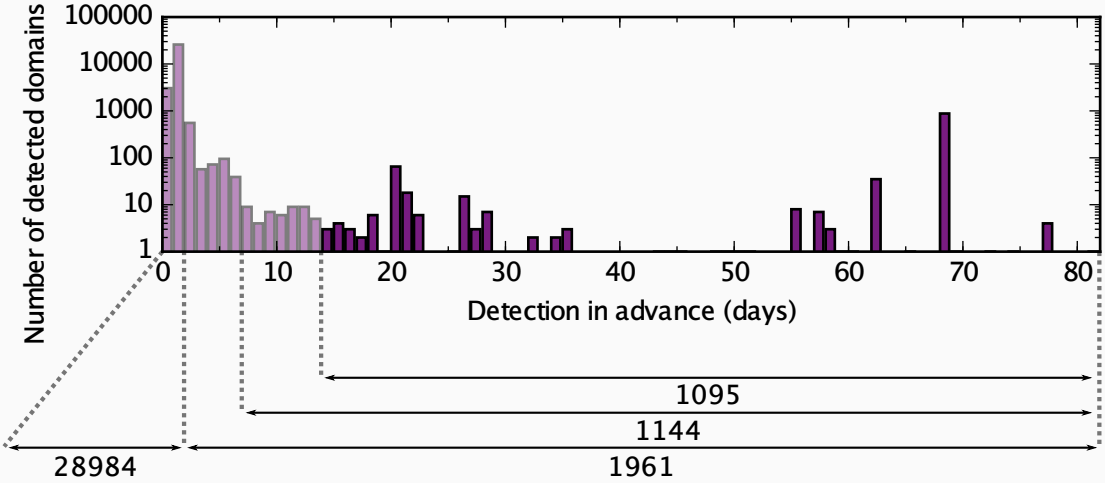
Realtime Blackhole List (RBL)



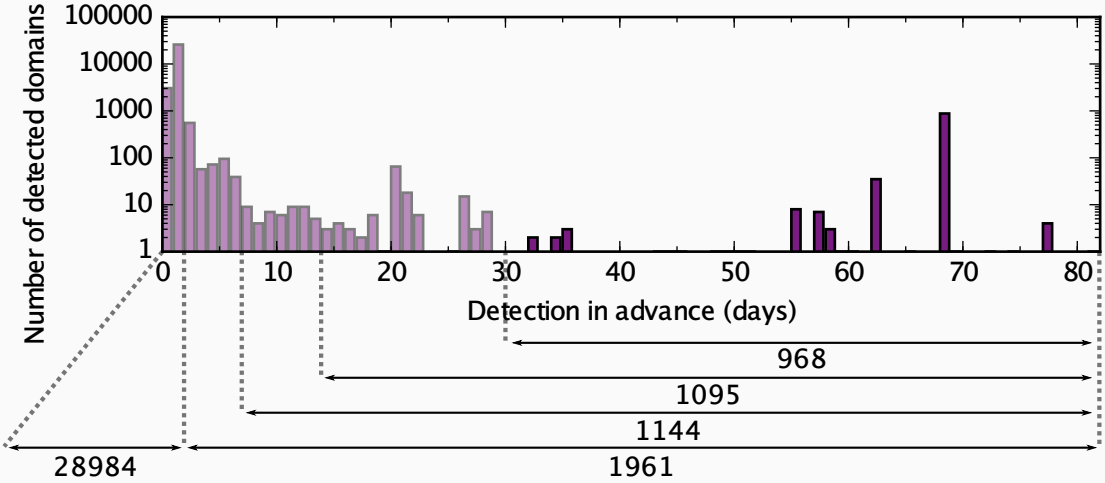
Realtime Blackhole List (RBL)



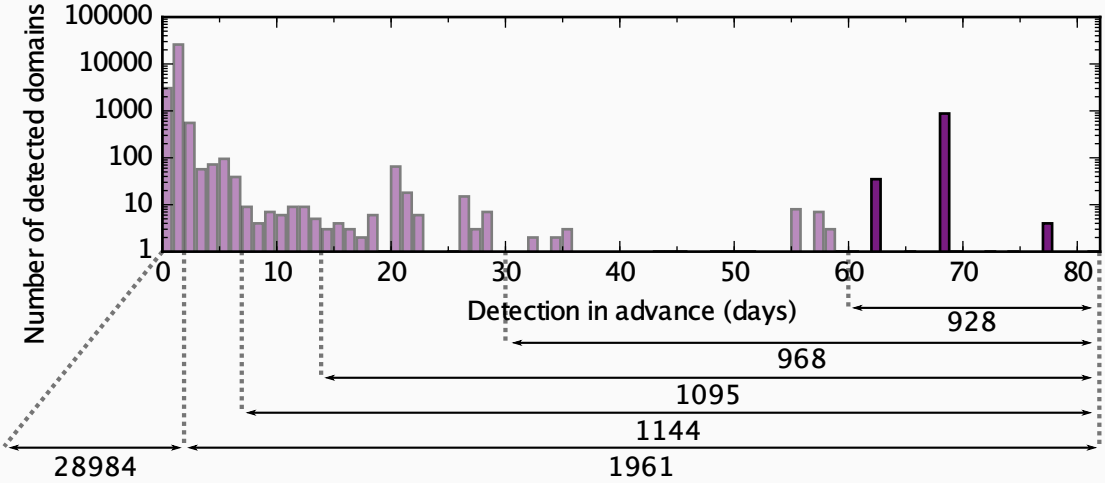
Realtime Blackhole List (RBL)



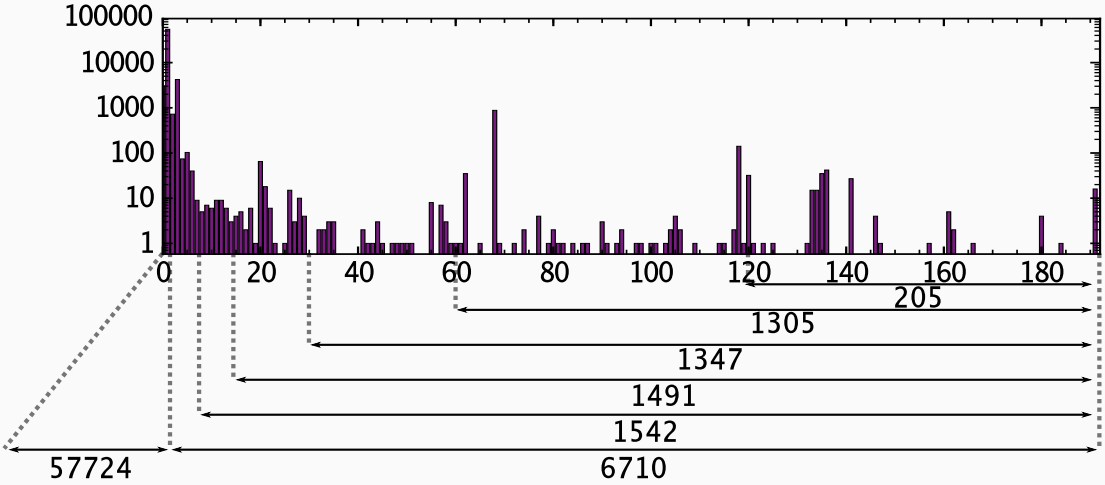
Realtime Blackhole List (RBL)

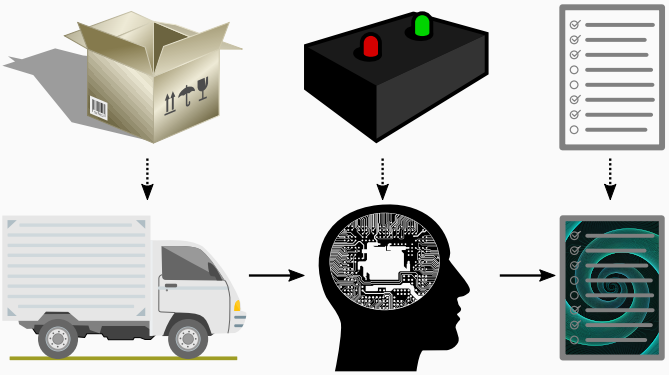


Realtime Blackhole List (RBL)



Realtime Blackhole List (RBL)

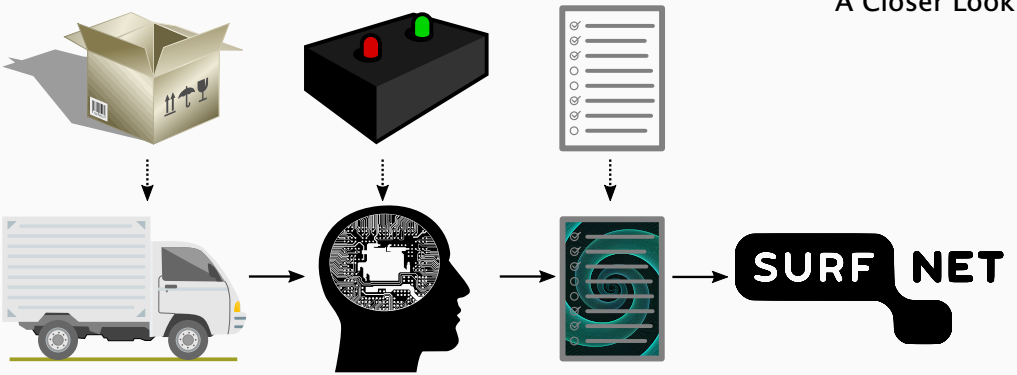




OpenINTEL

Machine Learning

Realtime Blackhole List (RBL)



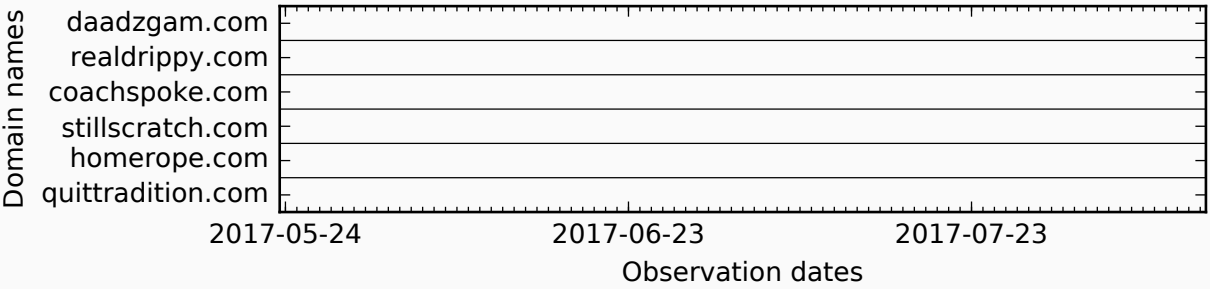
OpenINTEL

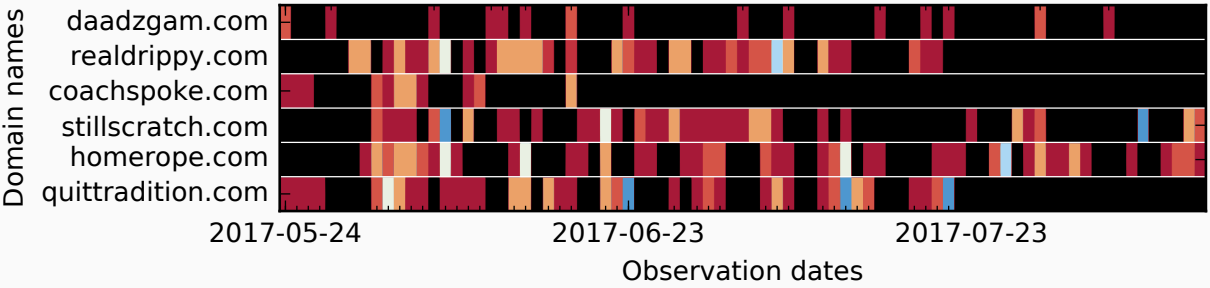
Machine Learning

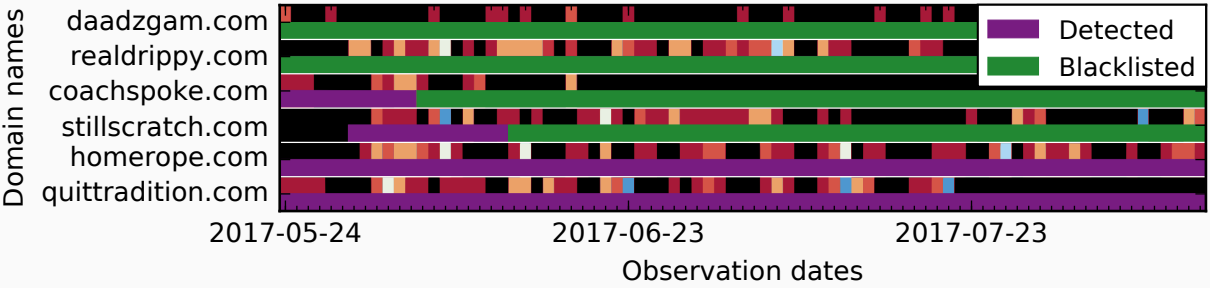
Realtime Blackhole List (RBL)

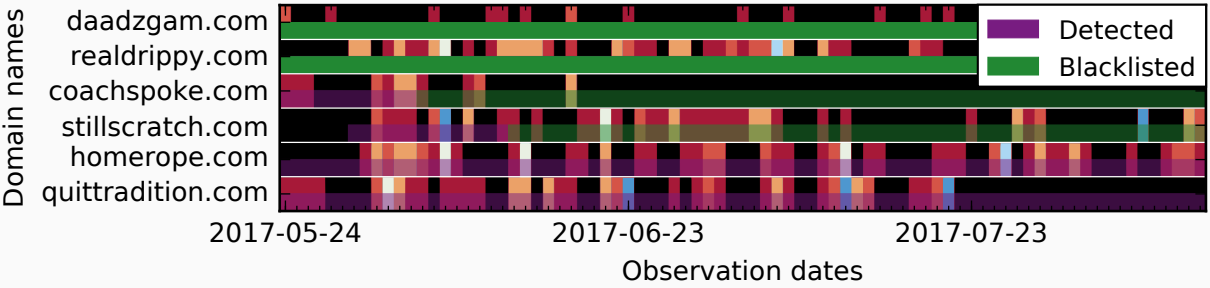
SURFmailfilter

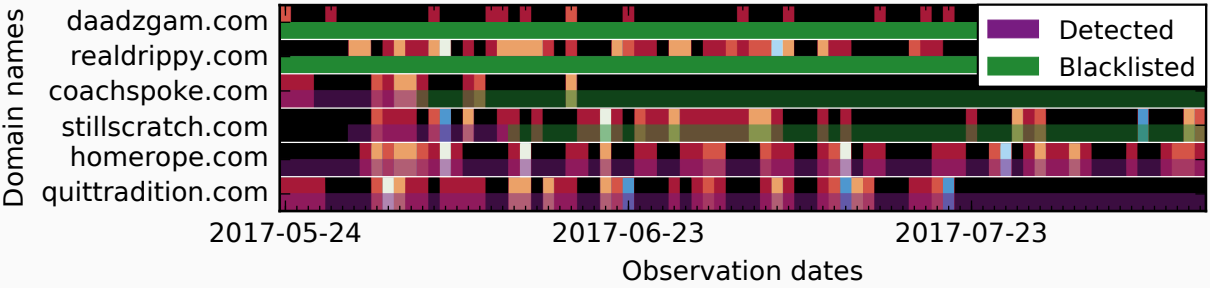
SURFmailfilter


Domain names
daadzgam.com
realdrippy.com
coachspoke.com
stillscratch.com
homerope.com
quittradition.com
2017-05-24
2017-06-23
2017-07-23
Observation dates

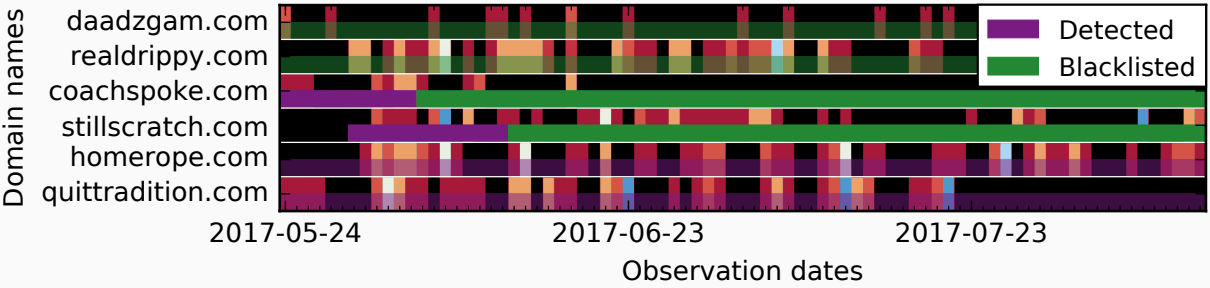

Domain names
daadzgam.com
realdrippy.com
coachspoke.com
stillscratch.com
homerope.com
quittradition.com
2017-05-24
2017-06-23
2017-07-23
Observation dates

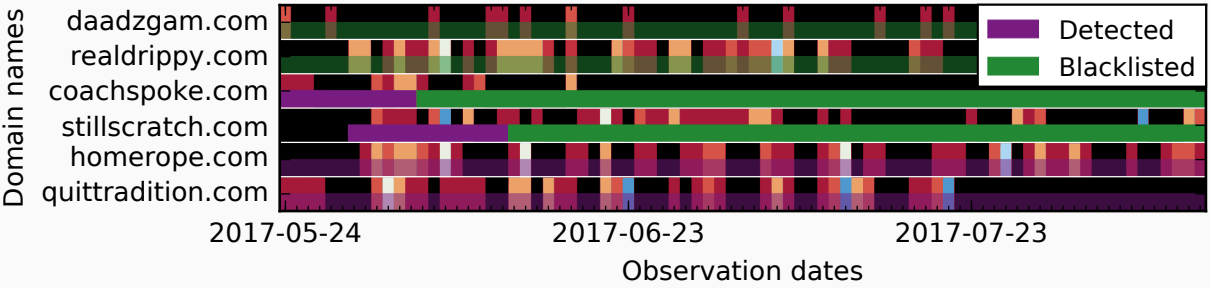

Domain names
daadzgam.com
realdrippy.com
coachspoke.com
stillscratch.com
homerope.com
quittradition.com
Detected
Blacklisted
2017-05-24
2017-06-23
2017-07-23
Observation dates


Domain names
daadzgam.com
realdrippy.com
coachspoke.com
stillscratch.com
homerope.com
quittradition.com
Detected
Blacklisted
2017-05-24
2017-06-23
2017-07-23
Observation dates

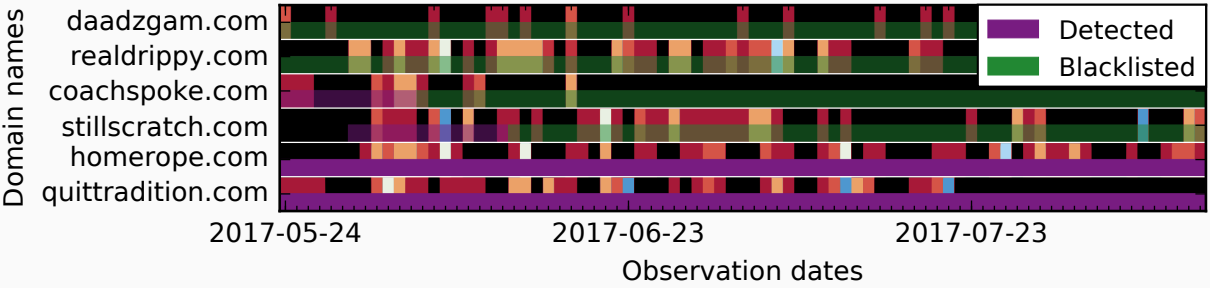


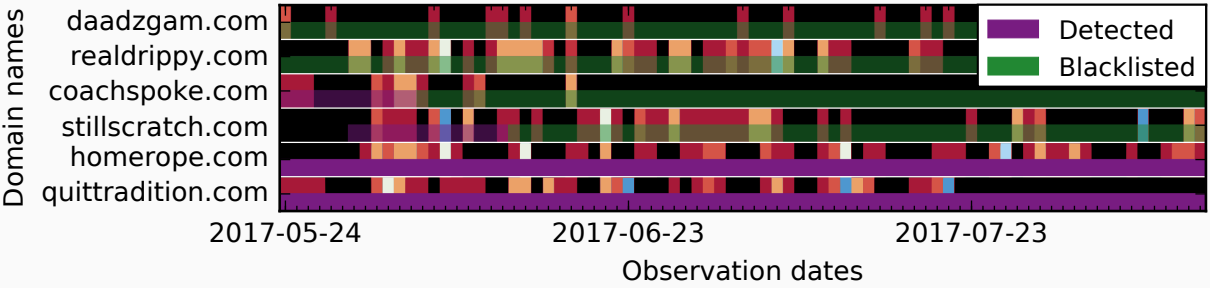
- 1188 emails
- 20 domains unique domains in the body


Domain names
daadzgam.com
realdrippy.com
coachspoke.com
stillscratch.com
homerope.com
quittradition.com
Detected
Blacklisted
2017-05-24
2017-06-23
2017-07-23
Observation dates



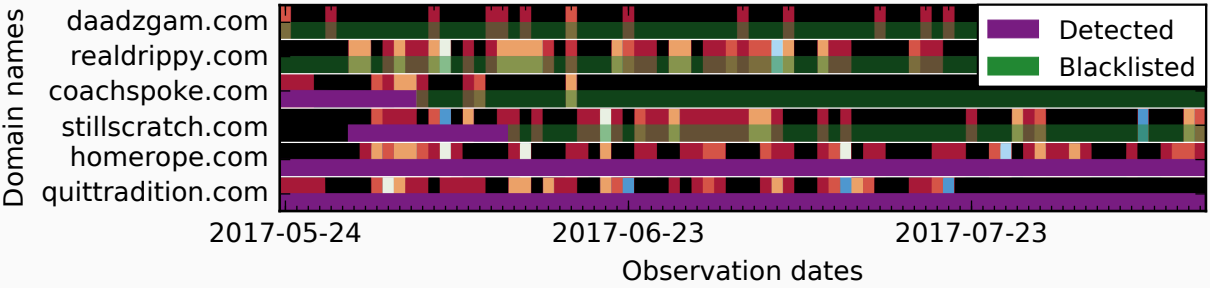
- 448 emails
- 29 unique domains in the body

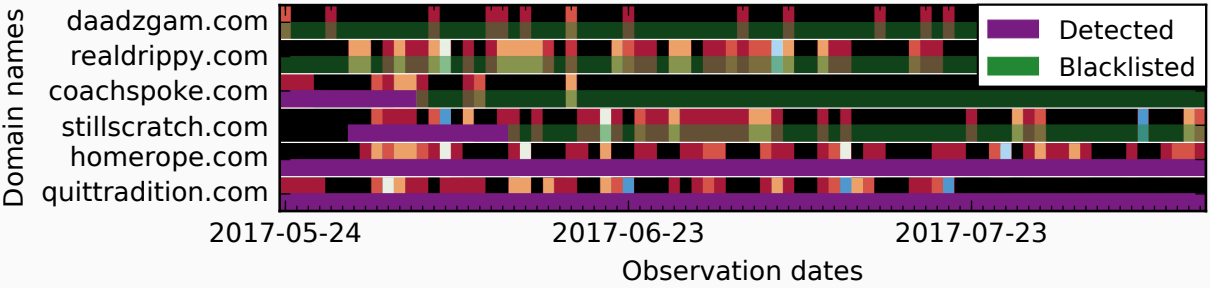

Domain names
daadzgam.com
realdrippy.com
coachspoke.com
stillscratch.com
homerope.com
quittradition.com
Detected
Blacklisted
2017-05-24
2017-06-23
2017-07-23
Observation dates



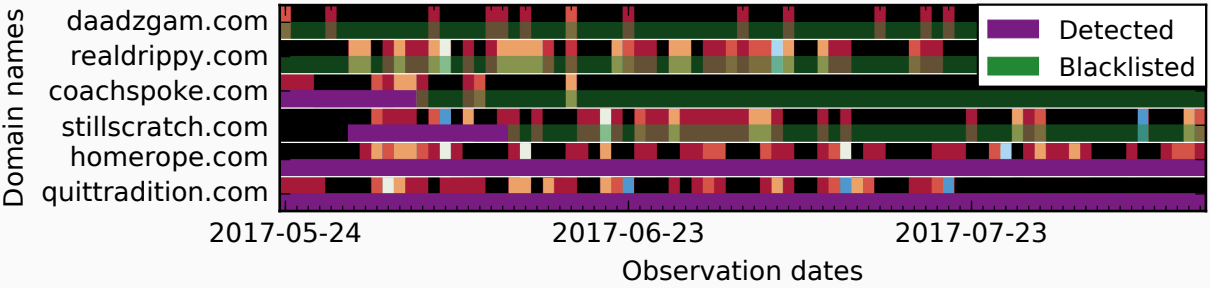
- 1006 emails
- 64 unique domains in the body

SURFmailfilter

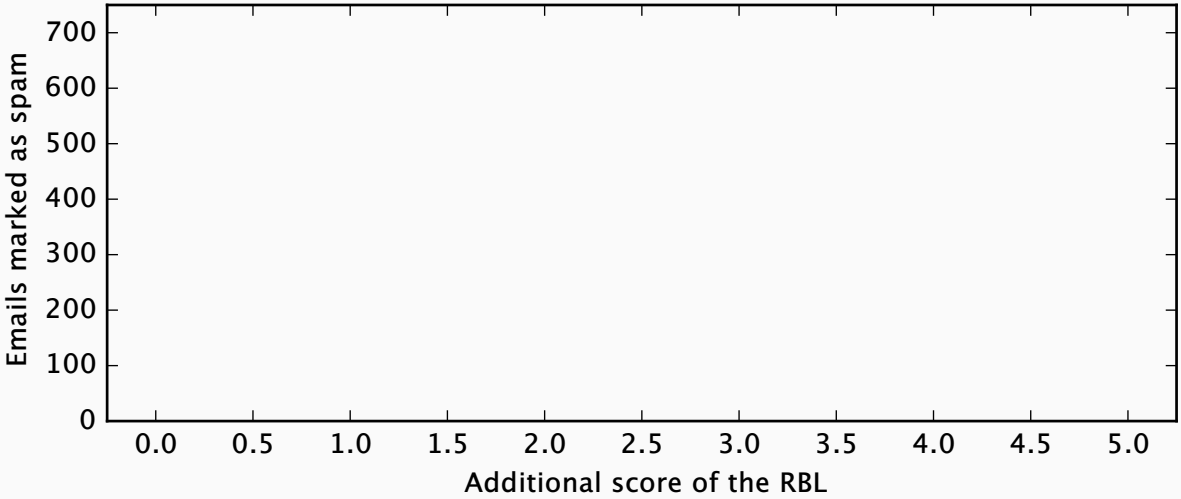


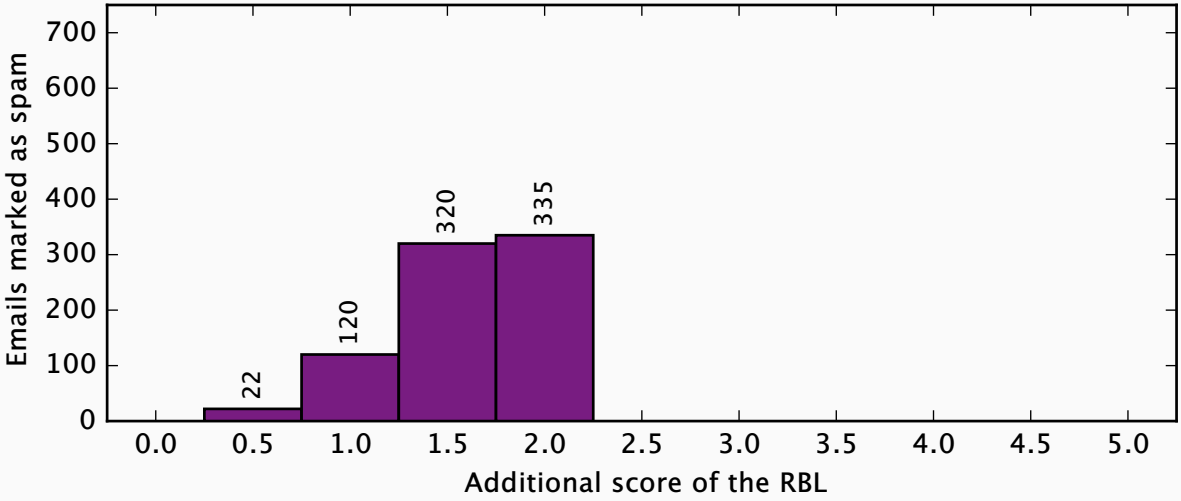


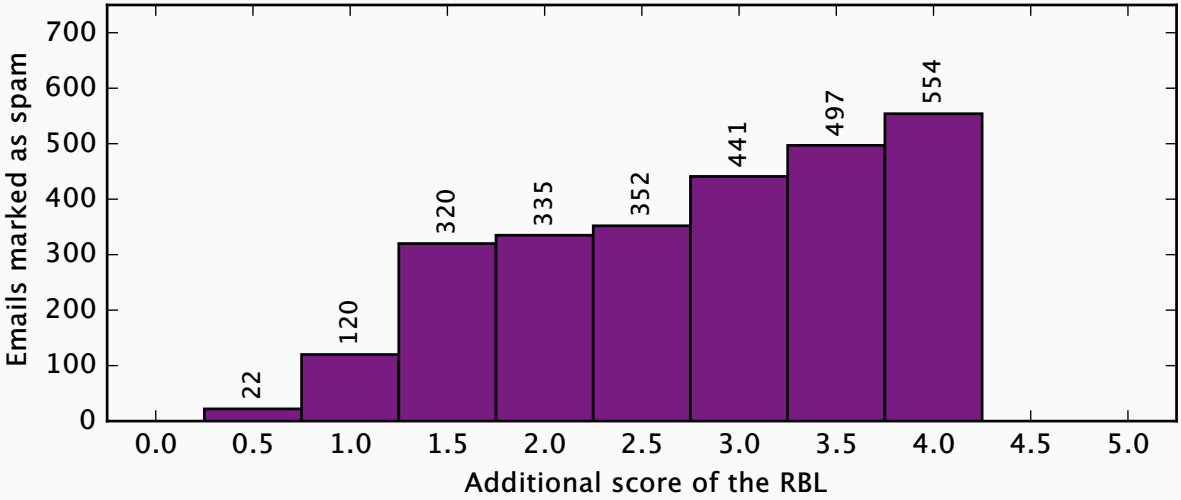
- 1080 emails
- 447 (41.39%) emails have a score of five or higher

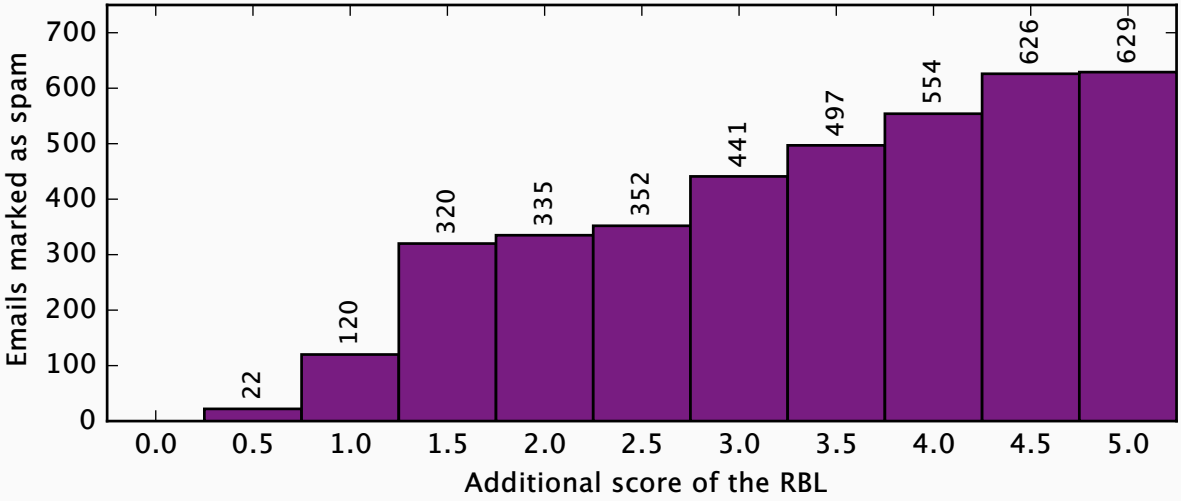


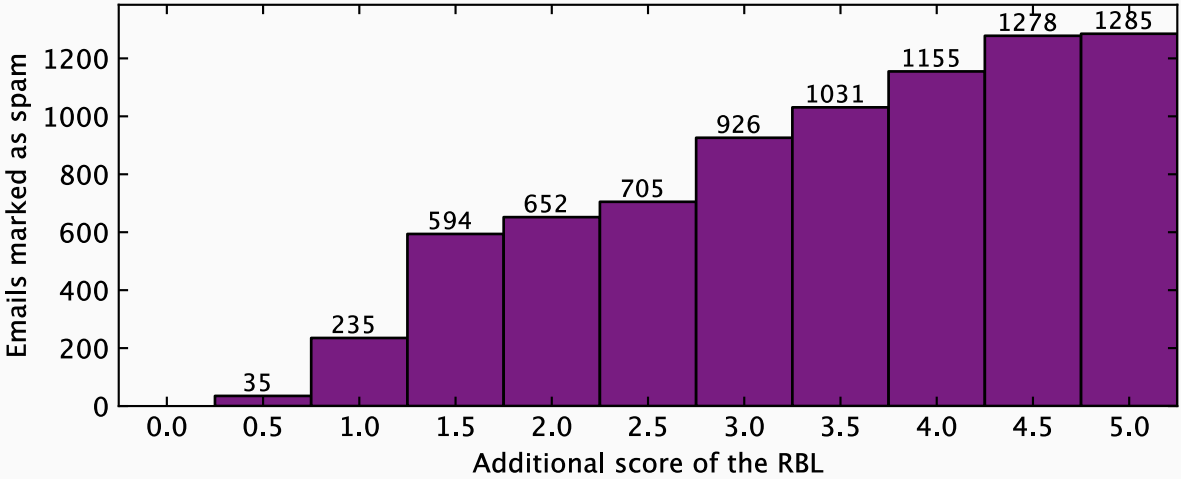
- 633 (58.61%) emails have a score below five
- 52 unique domains in the body
- of which 13 domains never appear in an email classified as spam
- these 13 domains appear in 31 emails (2.87%)


Emails marked as spam
700
600
500
400
300
200
100
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Additional score of the RBL

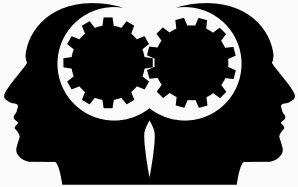

Emails marked as spam
700
600
500
400
300
200
100
0
22
120
320
335
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Additional score of the RBL


Emails marked as spam
700
600
500
400
300
200
100
0
22
120
320
335
352
441
497
554
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Additional score of the RBL


Emails marked as spam
700
600
500
400
300
200
100
0
22
120
320
335
352
441
497
554
626
629
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Additional score of the RBL


Emails marked as spam
0
200
400
600
800
1000
1200
35
235
594
652
705
926
1031
1155
1278
1285
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Additional score of the RBL

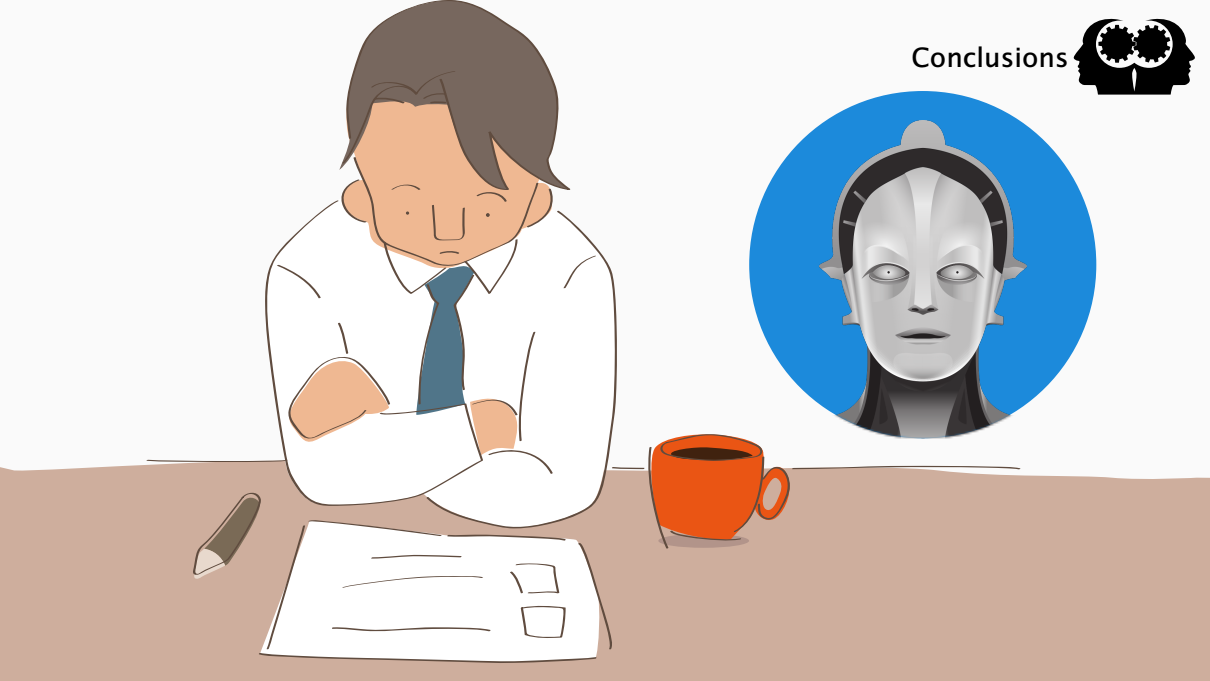
Conclusions



Conclusions

What is the advantage of proactive snowshoe spam domain detection using DNS data?

Conclusions



Conclusions

o.i.vandertoorn@utwente.nl