Melting the Snow: Detecting Snowshoe Spam Domains Using Active DNS Measurements

Olivier van der Toorn <o.i.vandertoorn@utwente.nl>

November 13, 2018

University of Twente, Design and Analysis of Communication Systems

NOMS 2018

Introduction

Re: Your Abandoned Package For Delivery

Spam

Mr.Joseph Scott <Scot@world.ocn.ne.jp>
to

Apr 5 (1 day ago)

Why is this message in Spam? It's similar to messages that were detected by our spam filters. Learn more

Re: Your Abandoned Package For Delivery

I have very vital information to give to you, but first I must have your trust before I review it to you because it may cause me my job, so I need somebody that I can trust for me to be able to review the secret to you.

I am Mr.Joseph Scott, head of luggage/baggage storage facilities here at the McCarran International Airport,Nevada USA. During my recent withheld package routine check at the Airport Storage Vault, I discovered an abandoned shipment from a Diplomat from London and when scanned it revealed an undisclosed sum of money in a metal trunk box. The consignment was abandoned because the Contents of the consignment was not properly declared by the consignee as "MONEY" rather it was declared as personal effect to avoid interrogation and also the inability of the diplomat to pay for the United States Non Inspection Charges which is $3,700USD. On my assumption the consignment is still left in our Storage House here at the International Airport Nevada till date. The details of the consignment including your name, your email address and the official documents from the United Nations office in Geneva are tagged on the Trunk box.

However, to enable me confirm if you are the actual recipient of this consignment as the assistant director of the Inspection Unit, I will advise you provide your current Phone Number and Full Address, to enable me cross check if it corresponds with the address on the official documents including the name of nearest Airport around your city. Please note that this consignment is supposed to have been returned to the United States Treasury Department as unclaimed delivery due to the delays in concluding the clearance processes so as a result of this, I will not be able to receive your details on my official email account. So in order words to enable me cross check your details, I will advise you send the required details to my private email address for quick processing and response. Once I confirm you as the actual recipient of the trunk box, I can get everything concluded within 48hours upon your acceptance and proceed to your address for delivery.

KINDLY FILL IN THE FOLLOWING INFORMATION:

Your Full Name:
Your Country:
Your Direct Telephone Numbers:
Your Direct Office Phone Numbers:
Mobile Phone Number:
Your Current Home Address:
Your Office Address:
Your age:

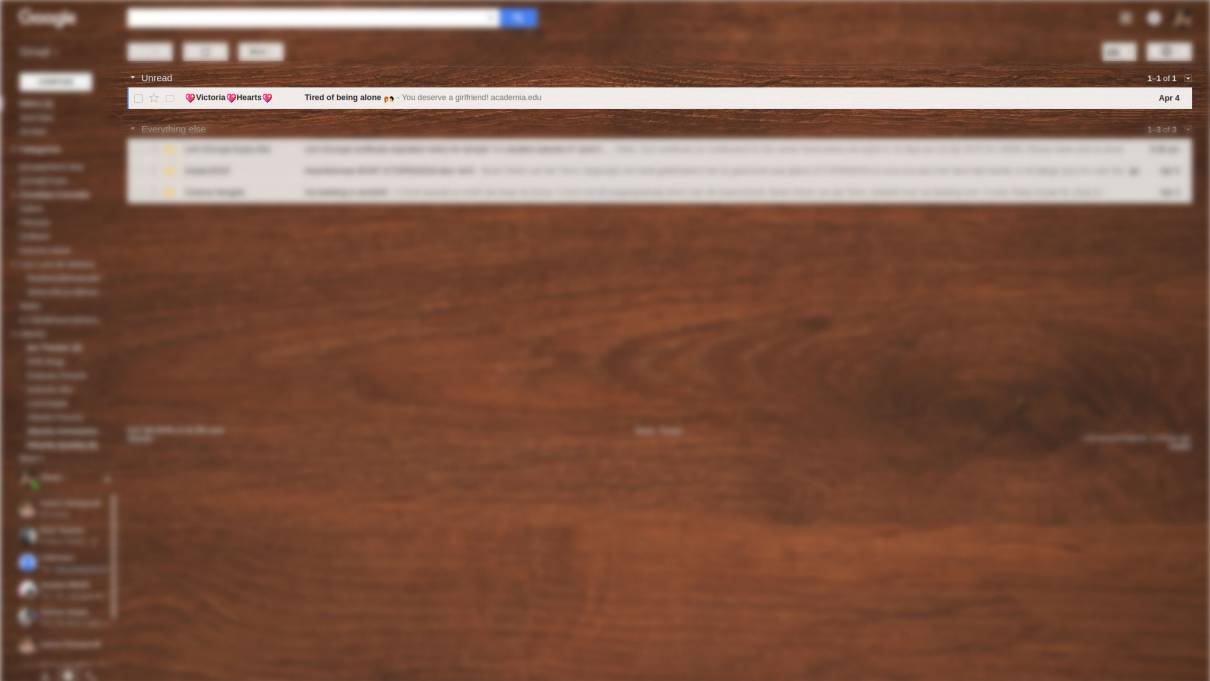
Lastly, be informed that the reason I have taken it upon myself to contact you personally about this abandoned consignment is because I want us to transact this business and share the money 80% for you and 20% for me since the consignment has not yet been returned to the United States Treasury Department after being abandoned by the diplomat so immediately the confirmation is made, I will go ahead and pay for the United States Non Inspection Fee of $3,700 dollars and arrange for the box to be delivered to your doorstep Or I can bring it by myself to avoid any more trouble but you have to assure me of my 20% share.

I wait to hear from you urgently if you are still alive and I will appreciate if we can keep this deal confidential. for further directives Email: (josephscott870@gmail.com) You can call me on my telephone number: +1 702 763 7639 for verbal conversation. Please if am not available to accept your call, please and drop a message or send me a text.

Thank you.

Delete all spam messages now (messages that have been in Spam more than 30 days will be automatically deleted)

Mr.Joseph Scott	Re: Your Abandoned Package For Delivery	- Re: Your Abandoned Package For Delivery I have very vital information to give to you, but first I must have your trust before I review it to you because it may cause me my job, so I neec	Apr 5
UN OFFICE FILE	Contact ORA Bank for your ATM Card	- New Secretary-General Antonio Guterres (UNITED NATION COMPENSATION COMMITTEE PROGRAMME) UNITED STATES REGIONAL HEADQUARTER IN AFFILIATION WITH THE	Apr 4
Evelyn Lai	Business proposal	- Hello, I have a business proposal of great benefit for the both of us. contact me for details. Email: evelynlai1979@gmail.com With Kind Regards. Ms. Evelyn Lai	Apr 2
Pedro SANCHEZ (2)	BATCH NUMMER:LTO/17/2017-5RD/EU	- INTERNATIONALE LOTTO/EUROMILLION PROGRAMME INTERNATIONALE PROMOTION-GEWINNZUTEILUNG REFERENZ NUMMER:PRCH/1097/SD2/ESP BATCH NUMMER:LT	Mar 31
Cpt.Stephen Laurence	God bless you and America!!!	- Hello Dear, I apologize if the content of my email is contrary to your moral ethics but I find it pleasurable to offer you my partnership in business. I am Cpt.Stephen Laurence, an officer in the US	Mar 31
Blood Sugar Blueprint	Urgent news about Metformin	- 1 trick that lets you eat sweets without spiking blood-sugar 🍬	Mar 30
(unknown sender)	Attention: Beneficiary.	- Attention: Beneficiary, We received a payment instruction from the Debt Reconciliation Committee of African Union in conjunction with the International Monetary Fund (IMF) to credit your account with	Mar 29
dutch.glory@yahoo.com	Ubuntu/Kubuntu-dev crash with installing 18.04 on enabeling format	- crash with installing 18.04 on enabeling partition "format" flag... Traceback (most recent call last): File "/usr/lib/ubiquity/ubiquity/frontend/kde_components/Pa	Mar 27
FULTON BANK	FULTON BANK	- Attention Dear Customer I am Dr. Mark McCollom, I work with a reputable financial institution in United State of American,Bank Address 111 Ballanard Ave,Pitman New Jersey 08071,FULTON BANK OF NEW	Mar 26
Robin Walker	(no subject)	- Ich bin Robin Walker, ein älterer Bürger von St. Alberta, Kanada. Ich habe einen Jackpot von 60 Millionen Dollar gewonnen, was der größte Gewinn der Lotterie in Alberta ist. Im Namen von mir und	Mar 26
QR-IAA 2018	CFP - Qualitative Research and Integrating Academic Activities	- Dear Olivier Van Der Toorn, We would like to invite you to contribute to the special three-track event on "Qualitative Research and Integrating Academic Activitie	Mar 25
Keith B. Bolcar	OFFICE OF HOMELAND SECURITY ..	- OFFICE OF HOMELAND SECURITY Immigration and Customs Enforcement US Department of Homeland Security, Street SW Suite 322 Atlanta Georgia 30303 2160 Park-lake Drive Nortl	Mar 23
Mr.Abdullahi Ibrahim	gat back to me as soon as you got my message	- Attn:Sir, We were mandated by the Federal Government of Nigeria last week so effect the release the sum of US$250000.00 to each affected (SCAM) victim around the globe u	Mar 23
United Nations Office	Quickly contact diplomat George Porter here at Bradley International Airport Connecticu...	- We hereby inform you that the Scotland Yard Police, Interpol, Federal Bureau of Investigation, (FBI) United States of America, the I	Mar 22
Mr Ivan Ronald	Re: Investment Proposal/For Your Consideration	- Hello Good morning Do you have any project or an on going project which requires funding? I got your contact from a Business Directory in my search for a trustworthy and reli	Mar 22
Patrick Ozougo Esq.	Contact Him.	- Dear. I want to inform you about the success in transferring the funds.I left $3.8 Million compensation for your efforts. Do contact my secretary Mr.Benjamin Mark via his email's address (dadam07@	Mar 21
Mr.Adebayo Adelabu	Kindly get back to Me..	- Dear Beneficiary, My name is Mr.Adebayo Adelabu,the deputy governor CBN.This is to bring the good news to you that I have been mandated by the President Federal Republic in conjunction with the Fe	Mar 19
Obrist Nicolas	funding	- Are you looking for a business loan, personal loans, mortgages, car loans, student loans, debt consolidation loans, unsecured loans, risk capital, at a low interest rate and affordable Interest rate of	Mar 17
Amo Riselli	Mailen Sie mir, wenn Sie das bekommen	- Meine Namen sind Amo Riselli, der Sieger der Nationalen Lotterie von vierundzwanzig Millionen, fünfhunderteintausend, zweihundertundachtzig Pfund. (£ 24, 501.283.) Am 27. Dezemt	Mar 17
MONEYGRAM FOUNDATION A	Attn ~	- CLAIM YOUR MONEYGRAM PREPAID VISA CARD CREDITED WITH $850000.00 USD - Reply To Claim !!! (null)	Mar 16
Delivery (2)	Contact diplomat John Malvin now at John.F.Kennedy international airport New York Usa, ...	- Attention, I'm Diplomat John Malvin, I have been trying to reach you on your telephone about an hour now just to inform you abou	Mar 14
Kamernet	Vind je nieuwe huurder of huisgenoot eenvoudig bij Kamernet	- Plaats gratis de advertentie en ontvang vandaag nog reacties! kamernet Plaats advertentie advertentie Plaats gratis je advertentie De beschikbare kamer of won	Mar 13
CRYPT0_Currency_Institute	Easy_L00PH0LE~T0~Turn~$1O~~T0~$1OO,OOO?	- Renegade--Russian---Analyst---Reveals---How---To---Turn--$1O To $1OOK	Mar 8


Unread
1–1 of 1
Victoria Hearts
Tired of being alone
You deserve a girlfriend! academia.edu
Apr 4
Everything else
1–3 of 3

Snowshoe Spam

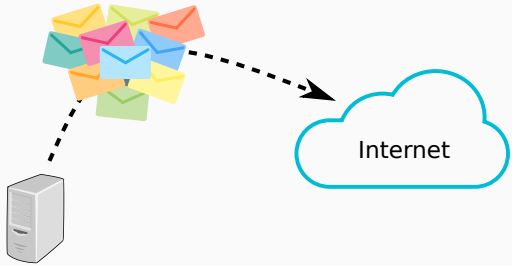
Snowshoe Spam

Internet

Spam

- few hosts

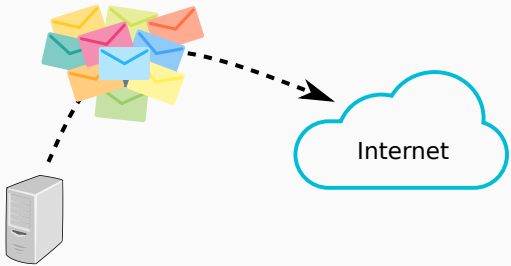
Snowshoe Spam



Spam

- few hosts
- many messages per host

Snowshoe Spam



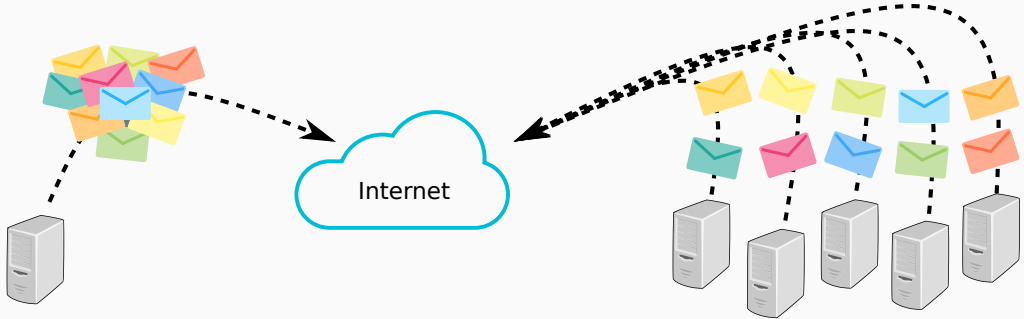
Spam

- few hosts
- many messages per host

Snowshoe Spam

- many hosts

Snowshoe Spam



Spam

- few hosts
- many messages per host

Snowshoe Spam

- many hosts
- few messages per host

Assumption

Original Message

Message ID	<820093659.47491888.1520538809384.JavaMail.root@basil.ocn.ne.jp>
Created at:	Thu, Mar 8, 2018 at 8:53 PM (Delivered after 402 seconds)
From:	MARX MARC <ktaguchi@basil.ocn.ne.jp>
To:	
Subject:	loan offer listings
SPF:	PASS with IP 153.149.236.10 Learn more

Assumption

Original Message

Message ID	<820093659.47491888.1520538809384.JavaMail.root@basil.ocn.ne.jp>
Created at:	Thu, Mar 8, 2018 at 8:53 PM (Delivered after 402 seconds)
From:	MARX MARC <ktaguchi@basil.ocn.ne.jp>
To:	
Subject:	loan offer listings
SPF:	PASS with IP 153.149.236.10 Learn more

Assumption: Background

SPF record from ‘consultant.com’

v=spf1

Assumption: Background

SPF record from ‘consultant.com’

v=spf1 ip4:213.165.64.0/23 ip4:74.208.5.64/26 ip4:74.208.122.0/26 ip4:212.227.126.128/25 ip4:212.227.15.0/24 ip4:212.227.17.0/27 ip4:74.208.4.192/26 ip4:82.165.159.0/24 ip4:217.72.207.0/27

Assumption: Background

SPF record from ‘consultant.com’

v=spf1 ip4:213.165.64.0/23 ip4:74.208.5.64/26 ip4:74.208.122.0/26 ip4:212.227.126.128/25 ip4:212.227.15.0/24 ip4:212.227.17.0/27 ip4:74.208.4.192/26 ip4:82.165.159.0/24 ip4:217.72.207.0/27 -all

Assumption: Background

SPF record from ‘consultant.com’

v=spf1 ip4:213.165.64.0/23 ip4:74.208.5.64/26 ip4:74.208.122.0/26 ip4:212.227.126.128/25 ip4:212.227.15.0/24 ip4:212.227.17.0/27 ip4:74.208.4.192/26 ip4:82.165.159.0/24 ip4:217.72.207.0/27 -all

Email from ‘paypoint_sanchez@consultant.com’

Received-SPF: fail (google.com: domain of paypoint_sanchez@consultant.com does not designate 103.10.4.139 as permitted sender) client-ip=103.10.4.139;

Assumption: Background

SPF record from ‘consultant.com’

v=spf1 ip4:213.165.64.0/23 ip4:74.208.5.64/26 ip4:74.208.122.0/26 ip4:212.227.126.128/25 ip4:212.227.15.0/24 ip4:212.227.17.0/27 ip4:74.208.4.192/26 ip4:82.165.159.0/24 ip4:217.72.207.0/27 -all

Email from ‘paypoint_sanchez@consultant.com’

Received-SPF: fail (google.com: domain of paypoint_sanchez@consultant.com does not designate 103.10.4.139 as permitted sender) client-ip=103.10.4.139;

Typical usage of SPF

v=spf1 a mx ip4:167.160.22.0/24 -all

Hypothesis

While snowshoe spammers are hard to detect, but still leave a trace in the DNS.

Hypothesis

While snowshoe spammers are hard to detect, but still leave a trace in the DNS.

Snowshoe spam + SPF

Hypothesis

While snowshoe spammers are hard to detect, but still leave a trace in the DNS.

Snowshoe spam + SPF
Many hosts + a DNS record for each host or a long SPF record

Hypothesis

While snowshoe spammers are hard to detect, but still leave a trace in the DNS.

Snowshoe spam + SPF
Many hosts + a DNS record for each host or a long SPF record
Domain with many records or long SPF records

Hypothesis

While snowshoe spammers are hard to detect, but still leave a trace in the DNS.

Snowshoe spam + SPF
Many hosts + a DNS record for each host or a long SPF record
Domain with many records or long SPF records

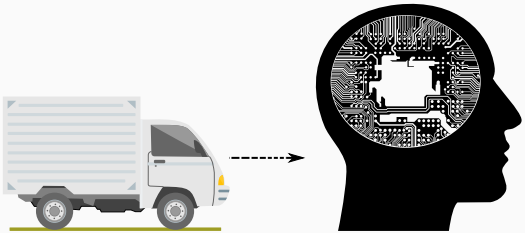
Active DNS measurements are a good way to detect snowshoe spam domains.

Methodology

Overview

OpenINTEL
(DNS data source)

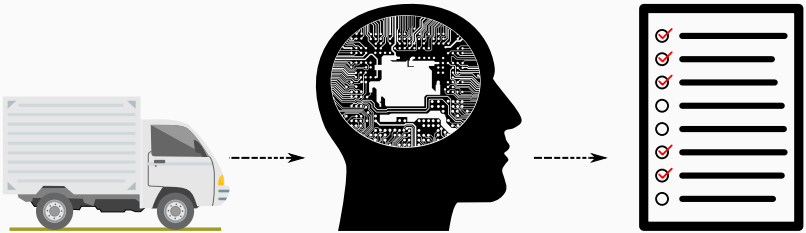
Overview



OpenINTEL
(DNS data source)

Machine Learning
(processing)

Overview

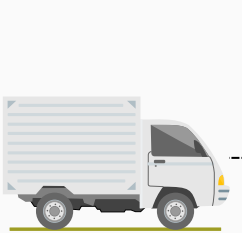


OpenINTEL
(DNS data source)

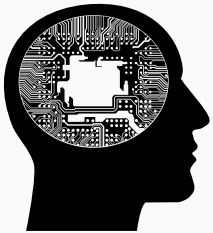
Machine Learning
(processing)

Realtime Blackhole List
(storage)

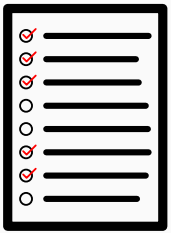
Overview



OpenINTEL
(DNS data source)



Machine Learning
(processing)



Realtime Blackhole List
(storage)

SURFnet
(validation)

OpenINTEL: Background

- Active DNS measurement platform

OpenINTEL: Background

- Active DNS measurement platform
- Queries more than 60% of registered domain names (in total more than 206 million)

OpenINTEL: Background

- Active DNS measurement platform
- Queries more than 60% of registered domain names (in total more than 206 million)
  - A
  - AAAA
  - MX
  - NS
  - ...

OpenINTEL: Background

- Active DNS measurement platform
- Queries more than 60% of registered domain names (in total more than 206 million)
  - A
  - AAAA
  - MX
  - NS
  - ...
- Every 24 hours a measurement is started

OpenINTEL: Datasets & Features

37 features

OpenINTEL: Datasets & Features

37 features

- Simple: number of MX addresses

OpenINTEL: Datasets & Features

37 features

- Simple: number of MX addresses
- Complex: number of IP addresses inside an SPF record

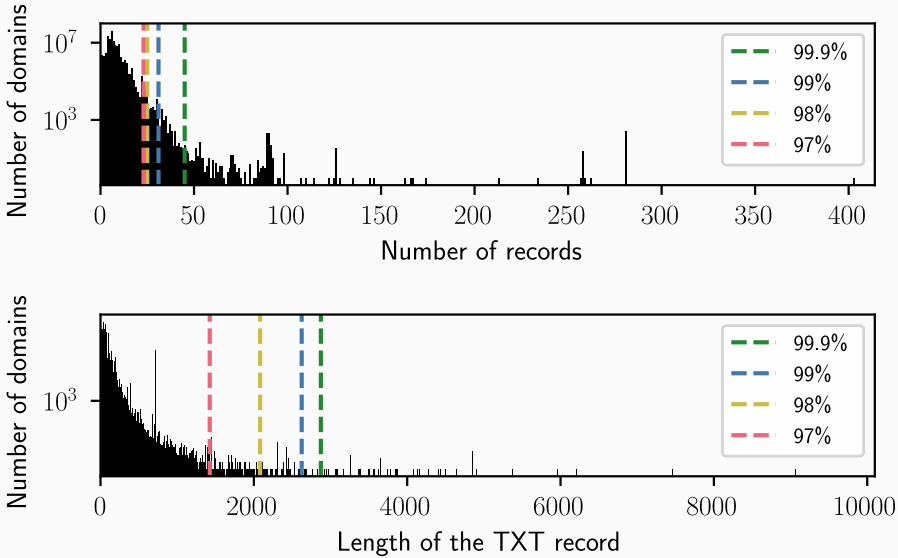
OpenINTEL: Datasets & Features

37 features

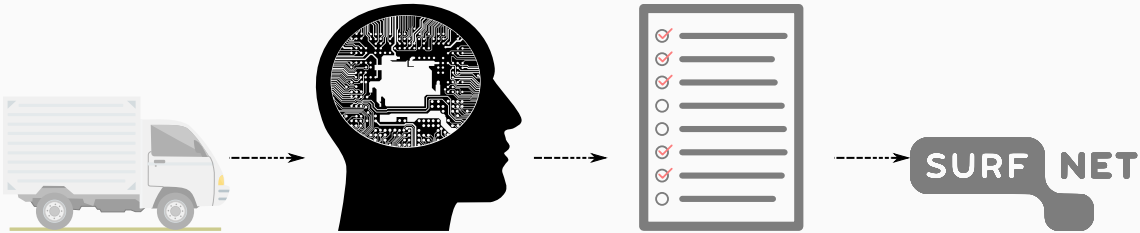
- Simple: number of MX addresses
- Complex: number of IP addresses inside an SPF record

These features are not computed for every domain in OpenINTEL.

OpenINTEL: Long Tail Analysis



Machine Learning



OpenINTEL
(DNS data source)

Machine Learning
(processing)

Realtime Blackhole List
(storage)

SURFnet
(validation)

Machine Learning: 12 algorithms

We have trained and evaluated 12 Machine Learning algorithms.

- Training dataset from domains on the long tail which appear in known blacklists.

Machine Learning: 12 algorithms

We have trained and evaluated 12 Machine Learning algorithms.

- Training dataset from domains on the long tail which appear in known blacklists.

The performance of each classifier is compared based on the precision metric.

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

Machine Learning: 12 algorithms

We have trained and evaluated 12 Machine Learning algorithms.

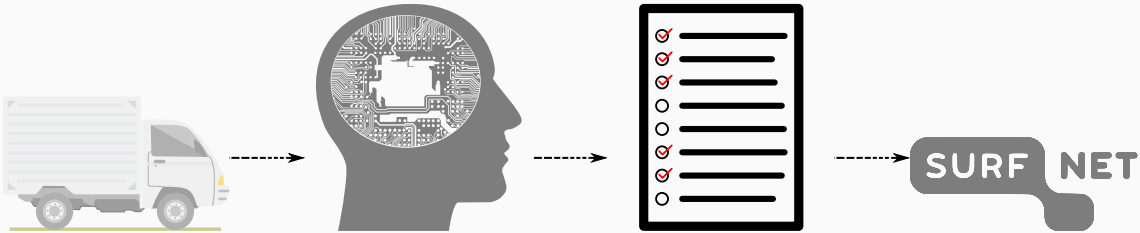
- Training dataset from domains on the long tail which appear in known blacklists.

The performance of each classifier is compared based on the precision metric.

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

Selected the 'AdaBoost' classifier as our classifier of choice, since it had the highest precision (98% with a FPR of 1%).

Realtime Blackhole List (RBL)



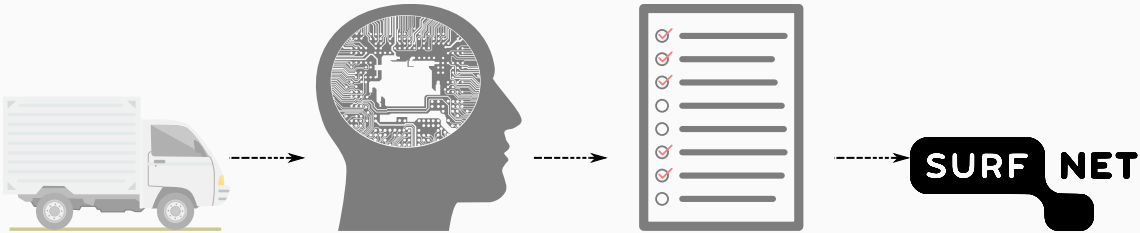
OpenINTEL
(DNS data source)

Machine Learning
(processing)

Realtime Blackhole List
(storage)

SURFnet
(validation)

SURFnet



OpenINTEL
(DNS data source)

Machine Learning
(processing)

Realtime Blackhole List
(storage)

SURFnet
(validation)

Methodology: Recap

Active DNS measurements of more than 60% of registered domain names forms the source of our data. We filter out large domains via the Long Tail Analysis.

Methodology: Recap

Active DNS measurements of more than 60% of registered domain names forms the source of our data. We filter out large domains via the Long Tail Analysis.

We have selected the AdaBoost classifier as our classifier of choice, since it had the highest precision metric.

Methodology: Recap

Active DNS measurements of more than 60% of registered domain names forms the source of our data. We filter out large domains via the Long Tail Analysis.

We have selected the AdaBoost classifier as our classifier of choice, since it had the highest precision metric.

The results of our daily detections are stored in an RBL.

Methodology: Recap

Active DNS measurements of more than 60% of registered domain names forms the source of our data. We filter out large domains via the Long Tail Analysis.

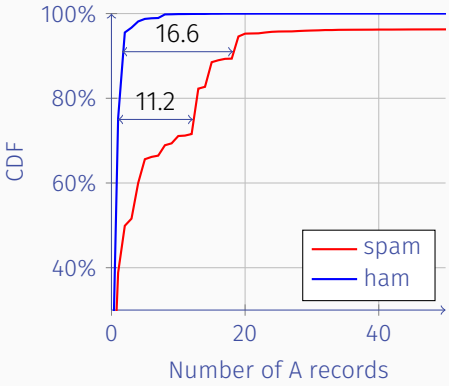
We have selected the AdaBoost classifier as our classifier of choice, since it had the highest precision metric.

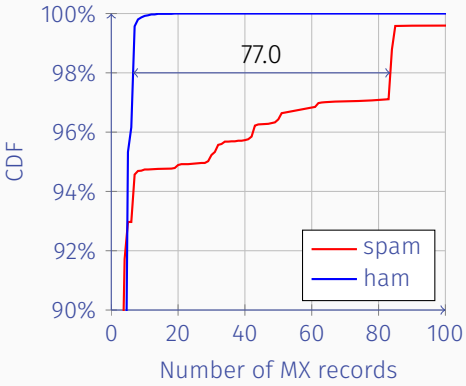
The results of our daily detections are stored in an RBL.

We have evaluated the RBL in SURFmailfilter.

Results

Distinction between two types





Example

	Domain	A records	MX records
(ham)	google.com	1	5

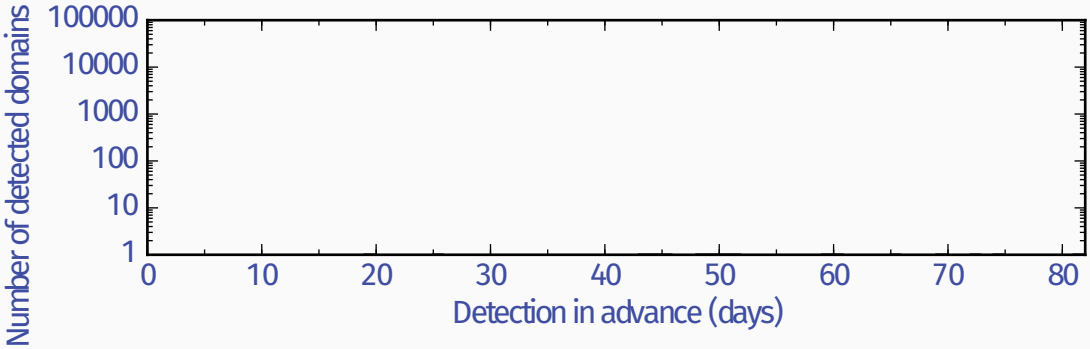
Example

	Domain	A records	MX records
(ham)	google.com	1	5
(spam)	giftiedan.com	61	1

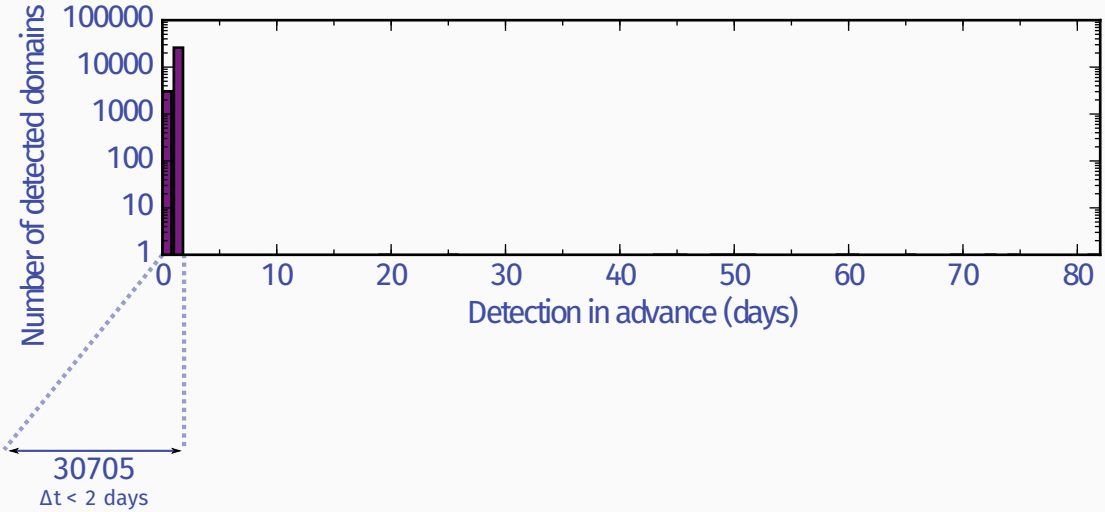
Example

	Domain	A records	MX records
(ham)	google.com	1	5
(spam)	giftiedan.com	61	1
(spam)	twirlmore.com	1	253

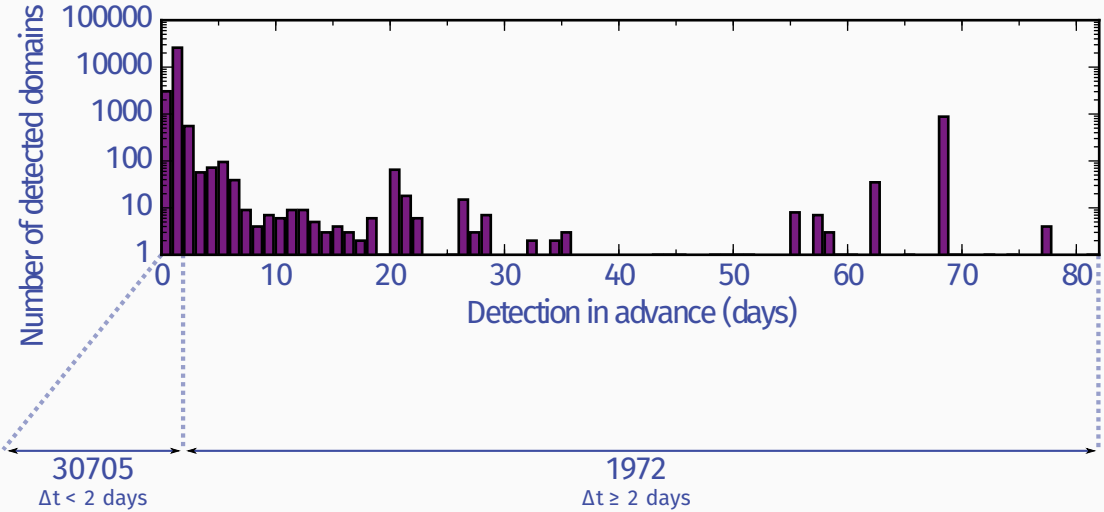
RBL comparison (2 month period)



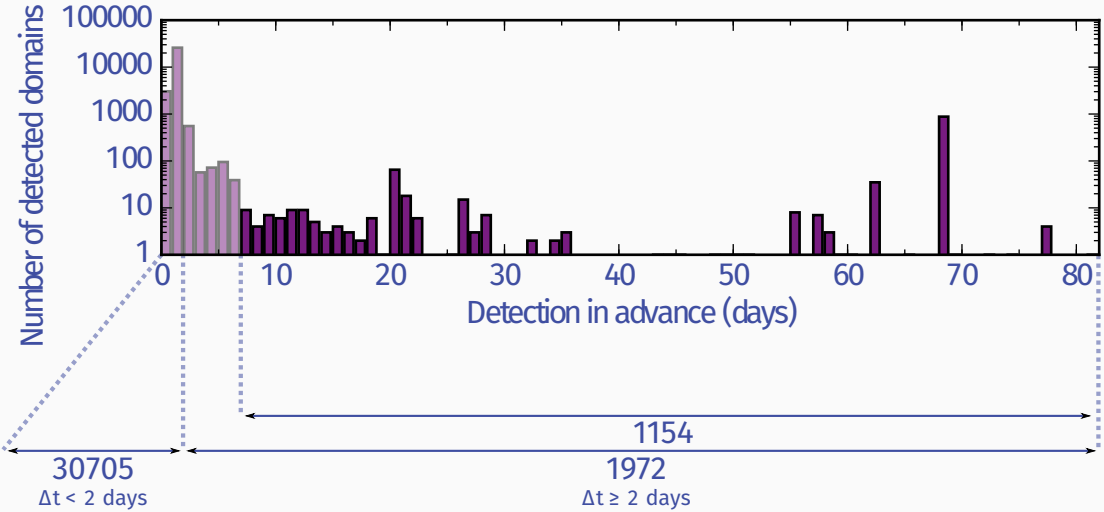
RBL comparison (2 month period)



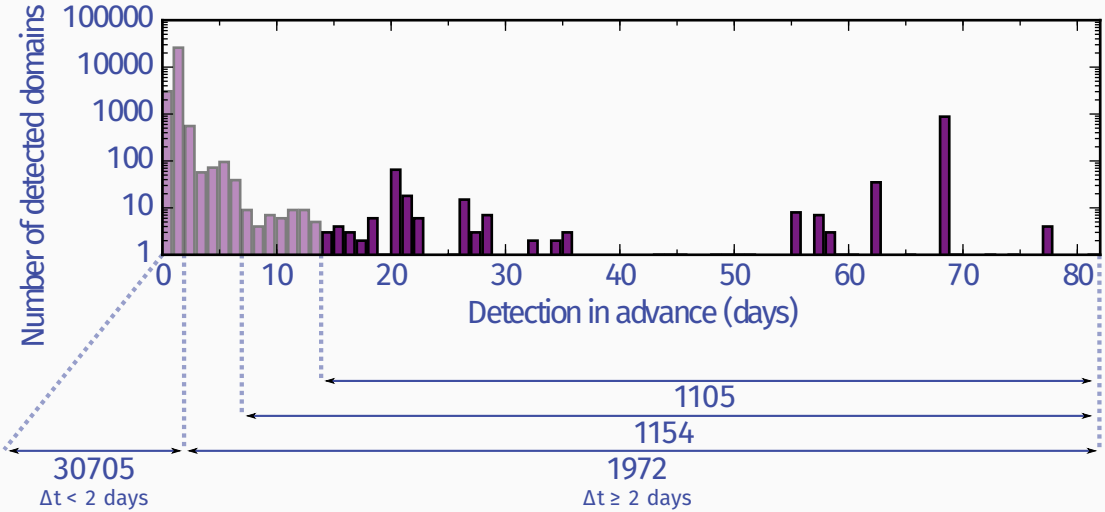
RBL comparison (2 month period)



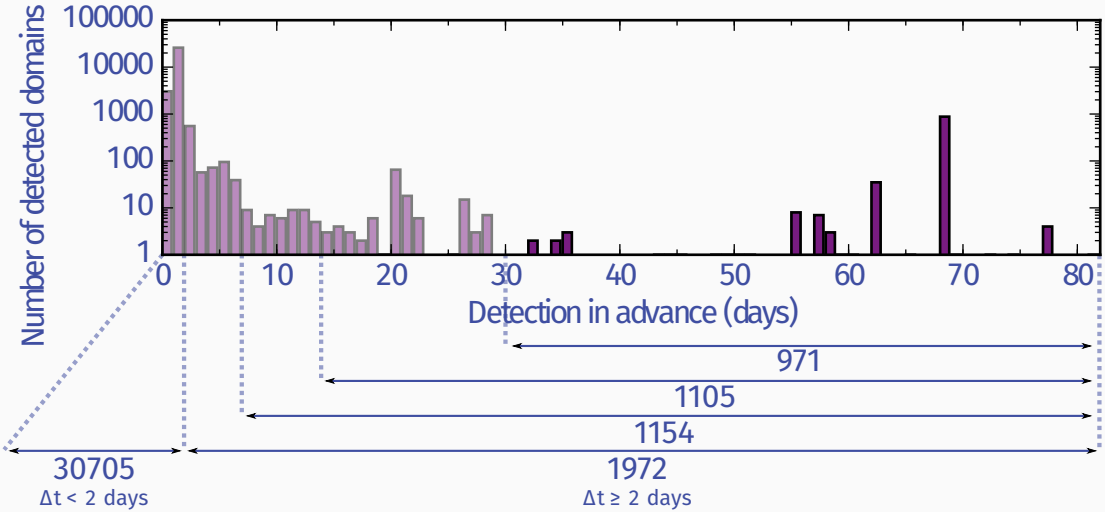
RBL comparison (2 month period)



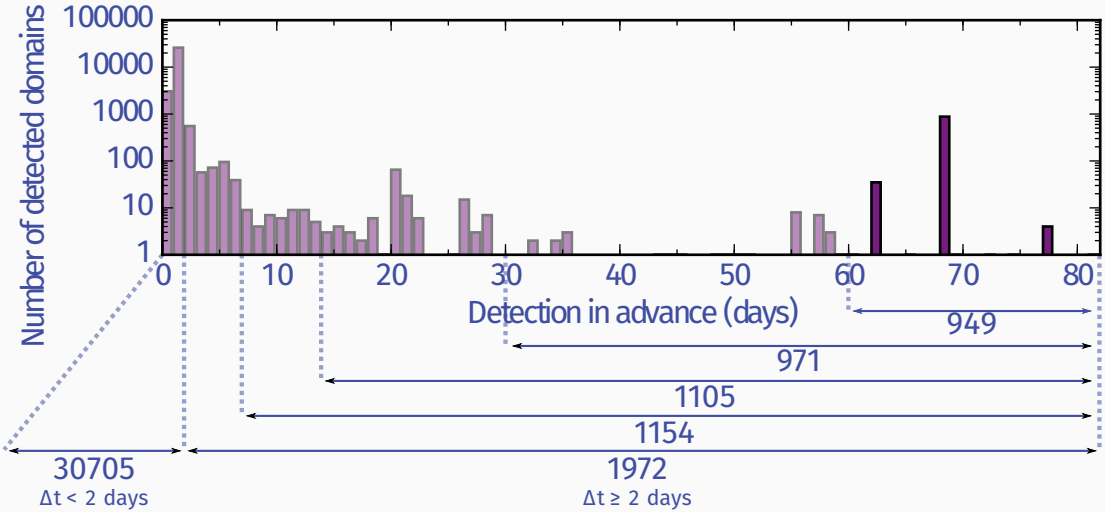
RBL comparison (2 month period)



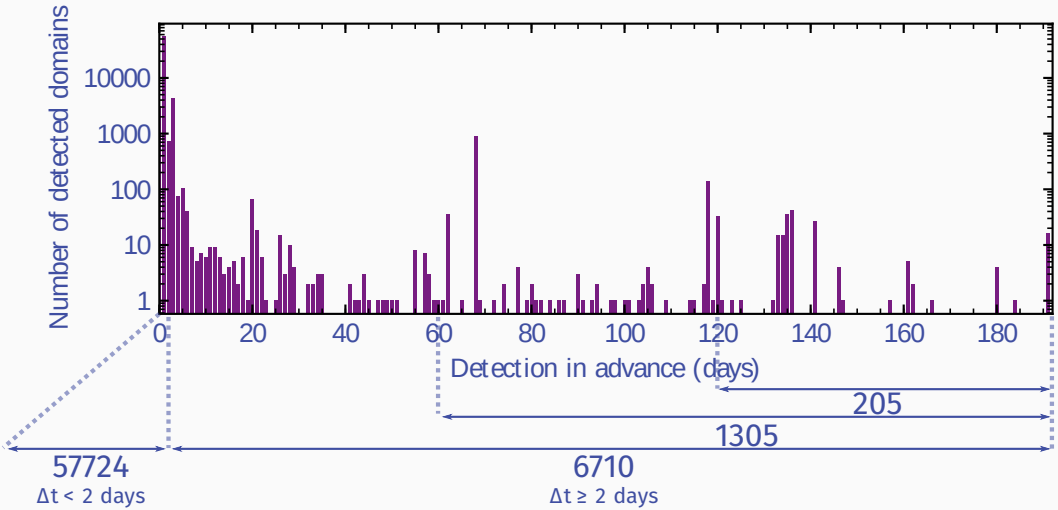
RBL comparison (2 month period)



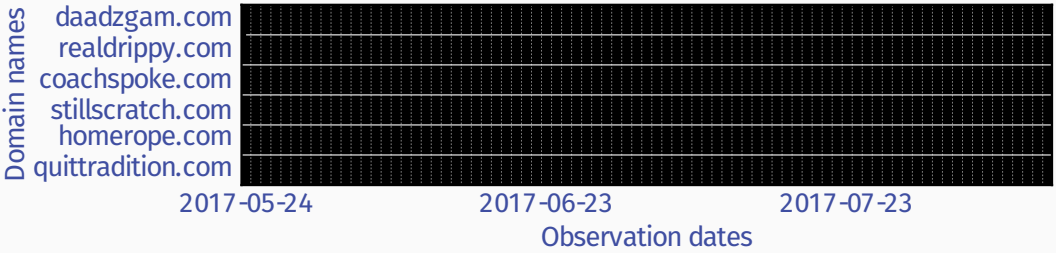
RBL comparison (2 month period)



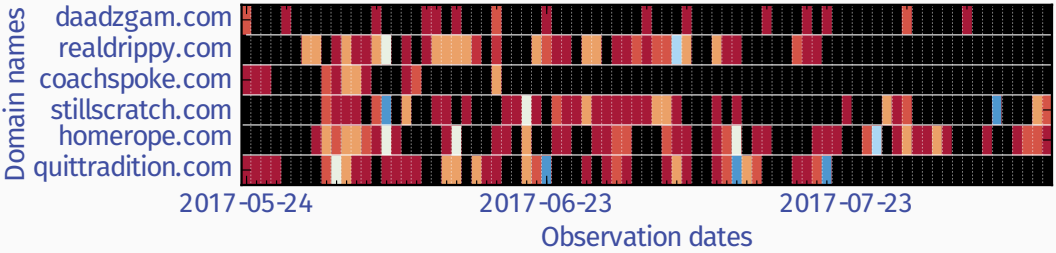
RBL comparison (9 month period)



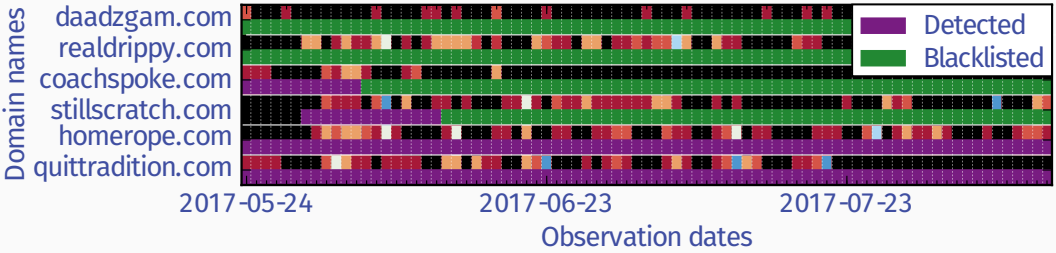
SURFnet evaluation



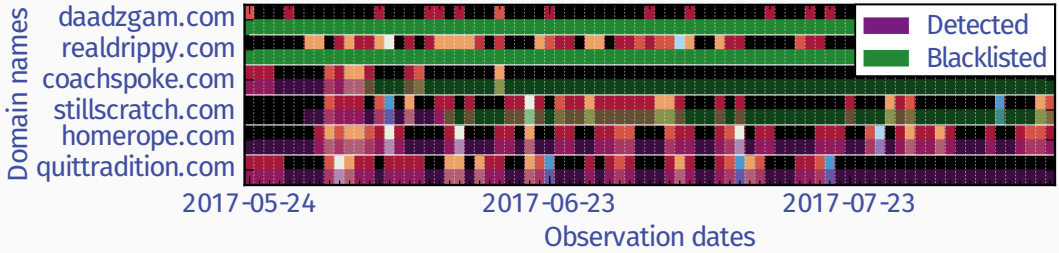
SURFnet evaluation

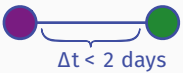


SURFnet evaluation



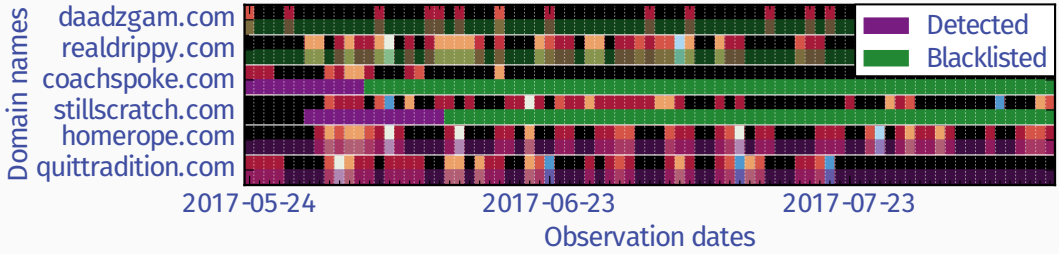
SURFnet evaluation

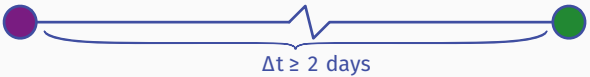




- 45% of received emails fall in this category
- 18% of observed domains fall in this category

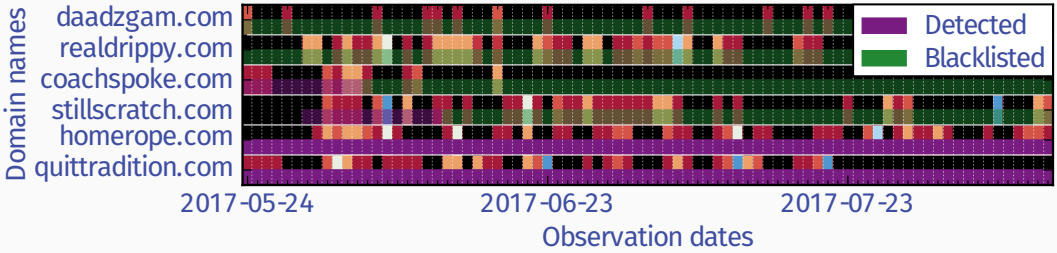
SURFnet evaluation





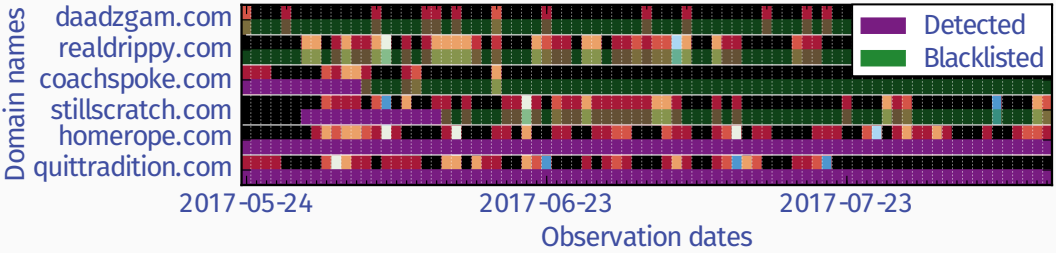
- 17% of received emails fall in this category
- 26% of observed domains fall in this category

SURFnet evaluation



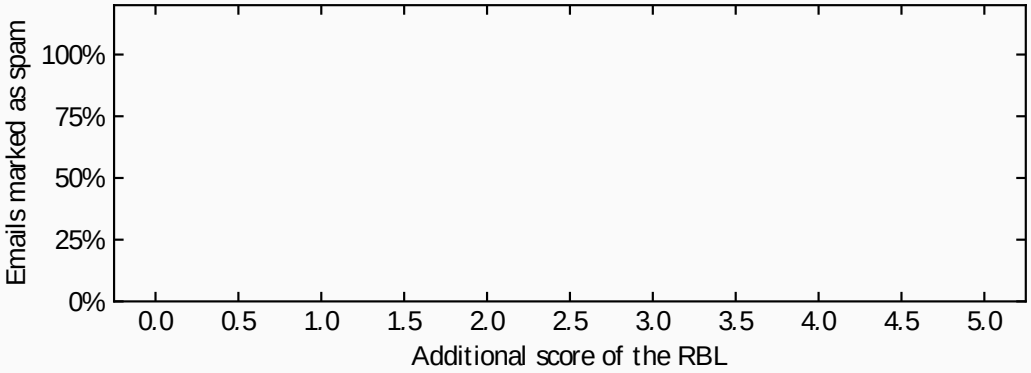
- 38% of received emails fall in this category
- 57% of observed domains fall in this category

SURFnet evaluation

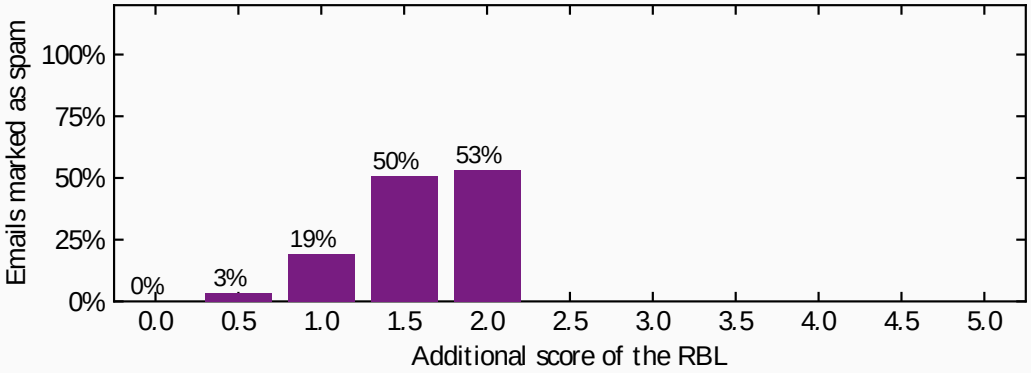


- 41% of emails were received in the purple areas
  - 59% of these emails have not been marked as spam

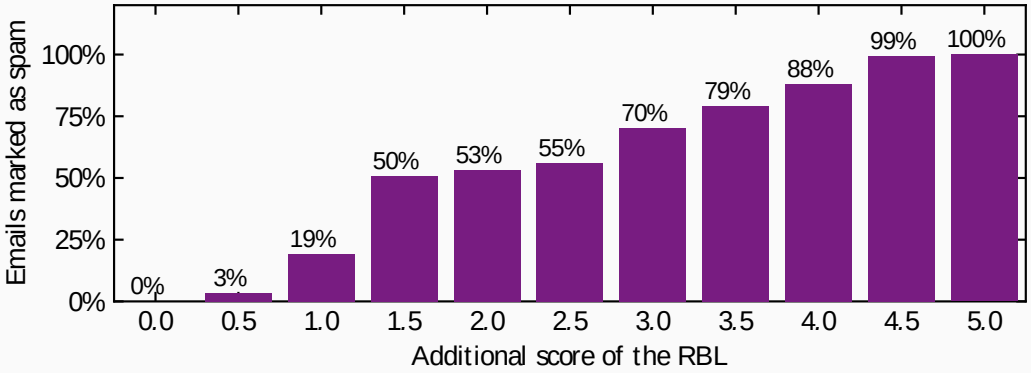
SURFnet evaluation



SURFnet evaluation



SURFnet evaluation



Conclusion

Conclusion

Using active DNS and by applying machine learning we are able to detect snowshoe spam domains.

Conclusion

Using active DNS and by applying machine learning we are able to detect snowshoe spam domains.

We are able to detect domains from 2 to 180 days in advance when compared to other blacklists.

Conclusion

Using active DNS and by applying machine learning we are able to detect snowshoe spam domains.

We are able to detect domains from 2 to 180 days in advance when compared to other blacklists.

This time advantage translates into additional email being marked as spam.

Emails which would otherwise bypass the email filter.

Conclusion

Using active DNS and by applying machine learning we are able to detect snowshoe spam domains.

We are able to detect domains from 2 to 180 days in advance when compared to other blacklists.

This time advantage translates into additional email being marked as spam.

Emails which would otherwise bypass the email filter.

After the evaluation period, SURFnet has deployed our RBL in production.

Thank you & questions

Thank you for listening[1].

Are there any questions?

[1] Images are from Pixabay and Wikimedia