TIDE: Proactive threat detection

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

2019-06-20

University of Twente, Design and Analysis of Communication Systems

Introduction

Who am I

- Ph.D. student from the University of Twente
- System administrator @SNT (ftp.nl.debian.org/ftp.snt.utwente.nl?)
- First FIRST conference

Contact details

tide-project.nl

o.i.vandertoorn@utwente.nl

Introduction

Is there a better way?

Is there a better way?

- Typically reactive detection approaches, or as it happens...
  - Based on passive measurement
  - Proof of suspicious activity is required

Proactive threat detection!

Proactive threat detection!

- Transition towards proactive security
- Use active measurement to pick up on clues of upcoming attacks
- Proactive threat detection gives an early warning

Why do we propose this?

Why do we propose this?

We want to improve attack detection:

Why do we propose this?

We want to improve attack detection:

- Proactive threat detection gives us, the defenders, a better chance against attacks

Why do we propose this?

We want to improve attack detection:

- Proactive threat detection gives us, the defenders, a better chance against attacks
- In the field of DNS this approach works, more on this later

Why do we propose this?

We want to improve attack detection:

- Proactive threat detection gives us, the defenders, a better chance against attacks
- In the field of DNS this approach works, more on this later

The advantages of an proactive approach are:

Why do we propose this?

We want to improve attack detection:

- Proactive threat detection gives us, the defenders, a better chance against attacks
- In the field of DNS this approach works, more on this later

The advantages of an proactive approach are:

- Unbiased towards your own network (depends on the underlaying measurement)

Why do we propose this?

We want to improve attack detection:

- Proactive threat detection gives us, the defenders, a better chance against attacks
- In the field of DNS this approach works, more on this later

The advantages of an proactive approach are:

- Unbiased towards your own network (depends on the underlaying measurement)
- Possible time advantage (alert before the attack happens)

What do we need to do proactive threat detection?

Three components:

What do we need to do proactive threat detection?

Three components:

- Data from active measurements (DNS, ICMP, etc.)

What do we need to do proactive threat detection?

Three components:

- Data from active measurements (DNS, ICMP, etc.)
- Knowledge about what you are measuring (what sets the abnormal apart from the normal?)

What do we need to do proactive threat detection?

Three components:

- Data from active measurements (DNS, ICMP, etc.)
- Knowledge about what you are measuring (what sets the abnormal apart from the normal?)
- Ability to use the detection results

Use cases

Use cases

Use case	Does proactive security work?

Use cases

Use case	Does proactive security work?
Snowshoe spam domains	Yes!

Use cases

Use case	Does proactive security work?
Snowshoe spam domains	Yes!
DDoS domains	Maybe

Use cases

Use case	Does proactive security work?
Snowshoe spam domains	Yes!
DDoS domains	Maybe
DNS TXT records	Maybe

Use cases

Use case	Does proactive security work?
Snowshoe spam domains	Yes!
DDoS domains	Maybe
DNS TXT records	Maybe
Combo-squat domains	No

OpenINTEL: How we measure

- OpenINTEL performs an active measurement, sending a fixed set of queries for all covered domains once every 24 hours
- We do this at scale, covering over 216 million domains per day:
  - gTLDs:
    .com, .net, .org, .info, .mobi, .aero, .asia, .name, .biz, .gov
    + almost 1200 “new” gTLDs (.xxx, .xyz, .amsterdam, .berlin, ...)
  - ccTLDs:
    .nl, .se, .nu, .ca, .fi, .at, .dk, .ru, .рф, .us, <your ccTLD here?>

Use case: Snowshoe spam

Snowshoe Spam

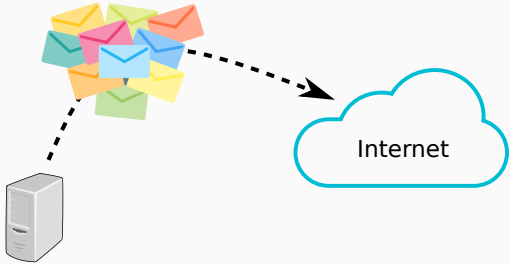
Snowshoe Spam

Internet

Spam

- few hosts

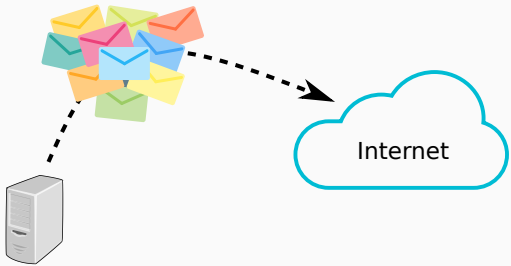
Snowshoe Spam



Spam

- few hosts
- many messages per host

Snowshoe Spam



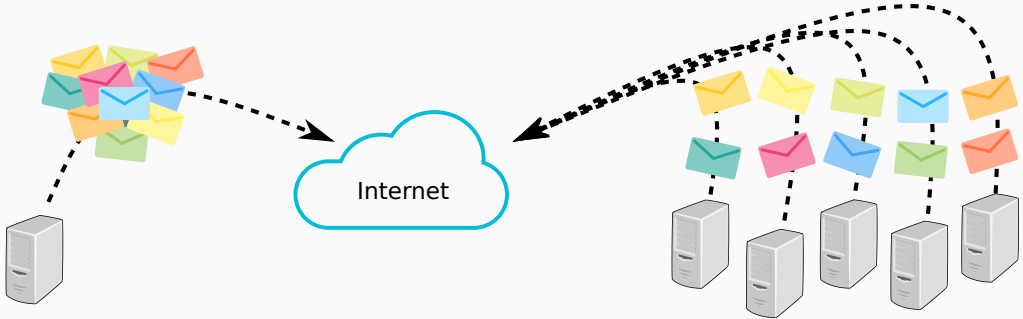
Spam

- 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

Snowshoe spam: Hypothesis

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

Snowshoe spam: Hypothesis

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

Snowshoe spam + SPF

Snowshoe spam: 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

Snowshoe spam: 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

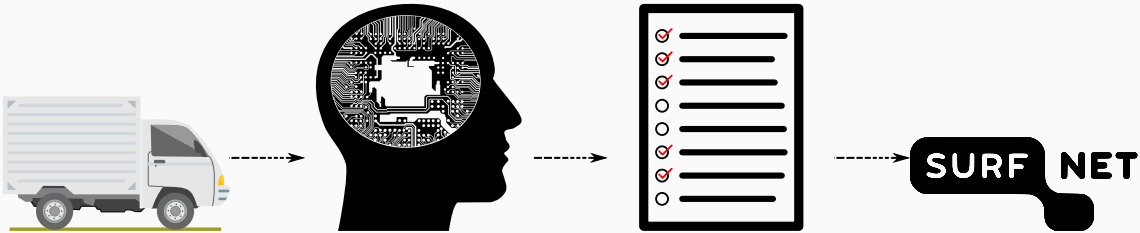
Snowshoe spam: Hypothesis

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.

Snowshoe spam: Methodology



OpenINTEL
(DNS data source)

Machine Learning
(processing)

Realtime Blackhole List
(storage)

SURFnet
(validation)

Snowshoe spam: Datasets & Features

37 features

Snowshoe spam: Datasets & Features

37 features

- Simple: number of MX addresses

Snowshoe spam: Datasets & Features

37 features

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

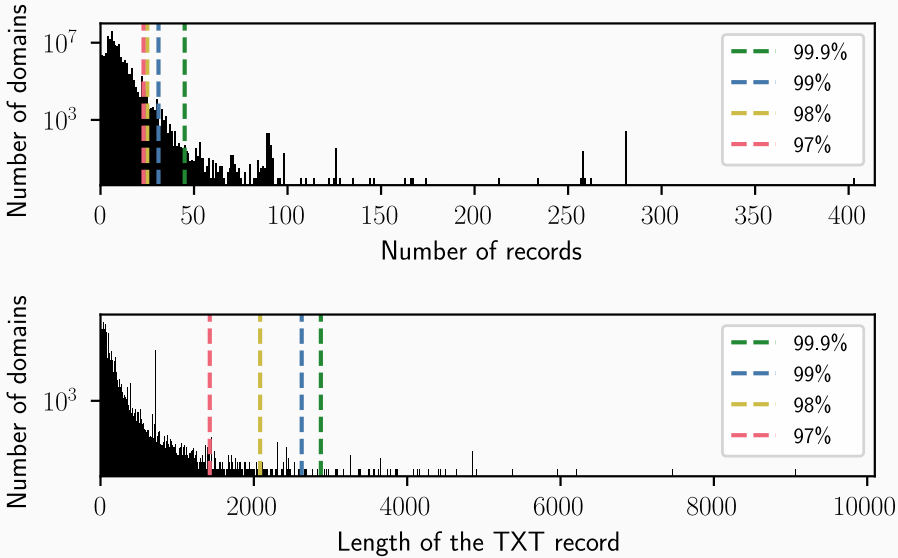
Snowshoe spam: Datasets & Features

37 features

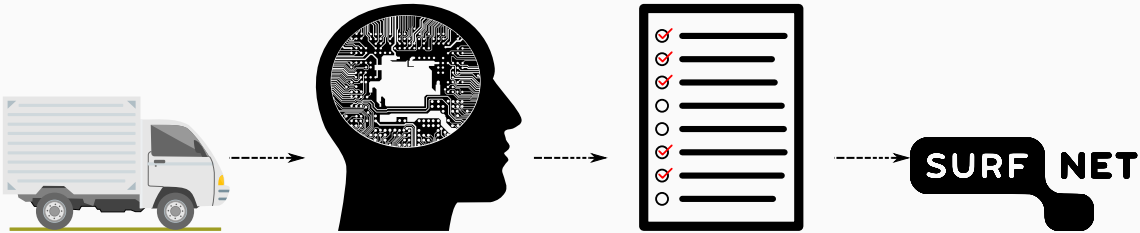
- Simple: number of MX addresses
- Complex: number of IP addresses inside an SPF record

These features are not computed for every domain in OpenINTEL.

Snowshoe spam: Long Tail Analysis



Snowshoe spam: Methodology



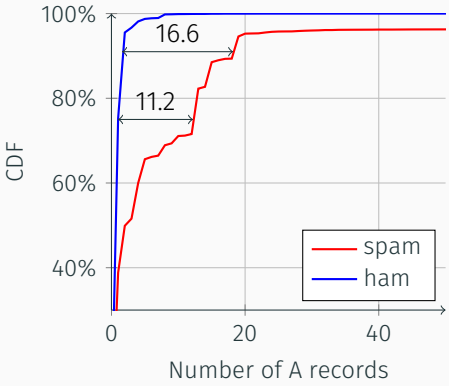
OpenINTEL
(DNS data source)

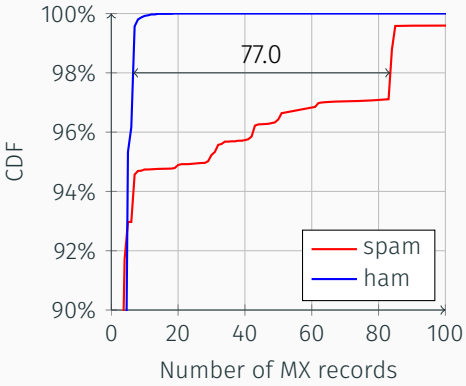
Machine Learning
(processing)

Realtime Blackhole List
(storage)

SURFnet
(validation)

Snowshoe spam: Results





Snowshoe spam: Example

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

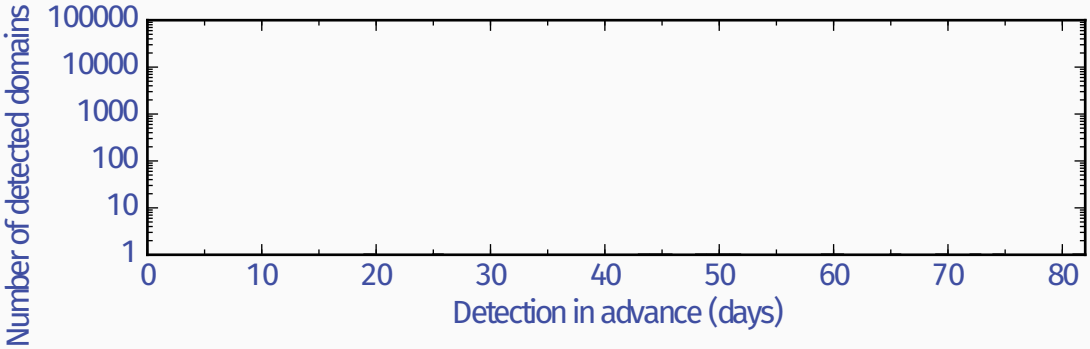
Snowshoe spam: Example

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

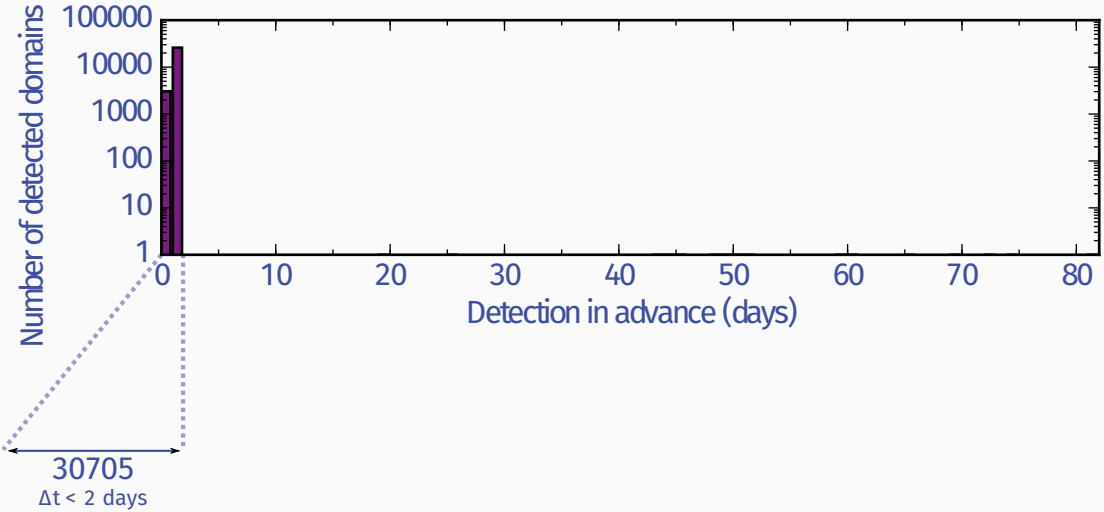
Snowshoe spam: Example

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

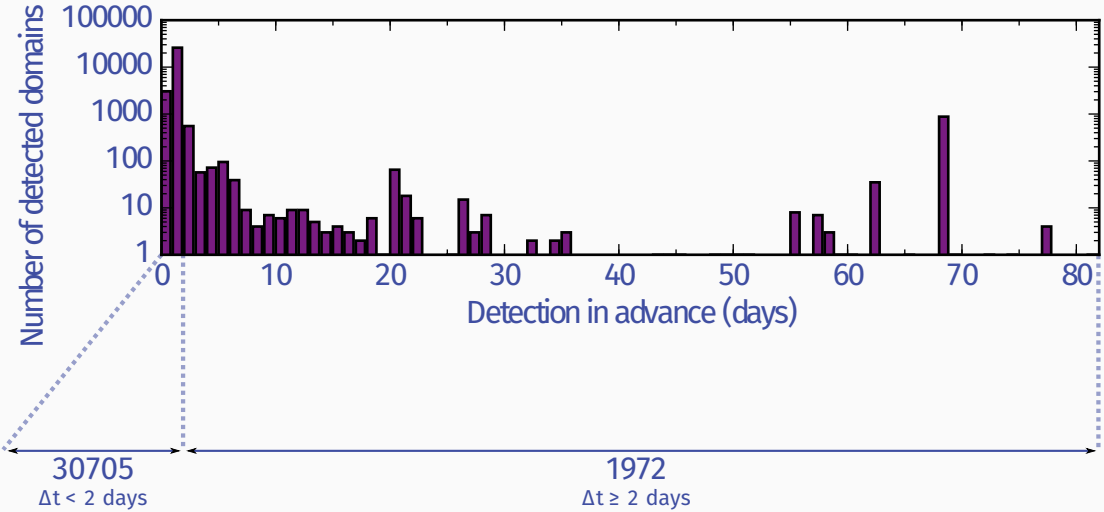
RBL comparison (2 month period)



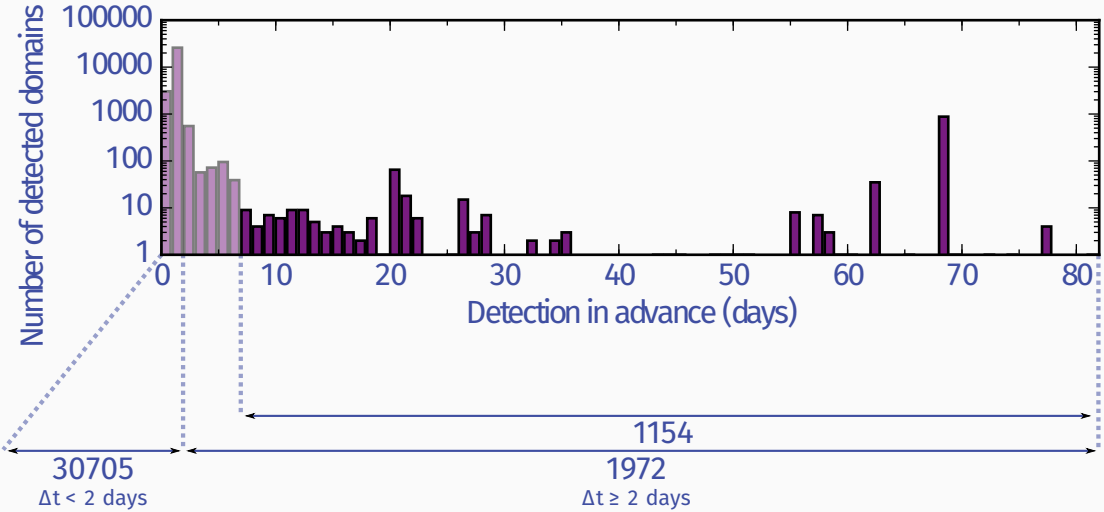
RBL comparison (2 month period)



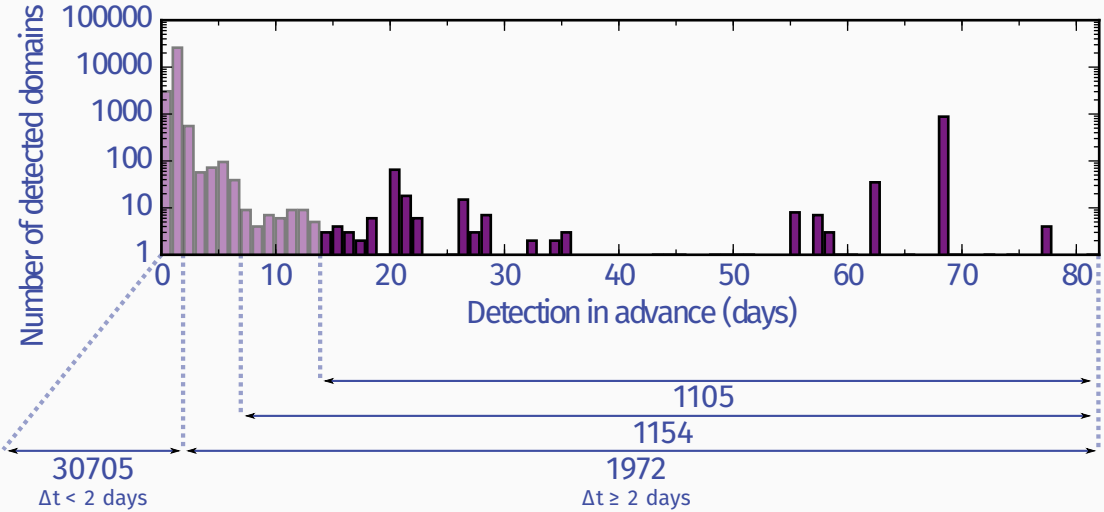
RBL comparison (2 month period)



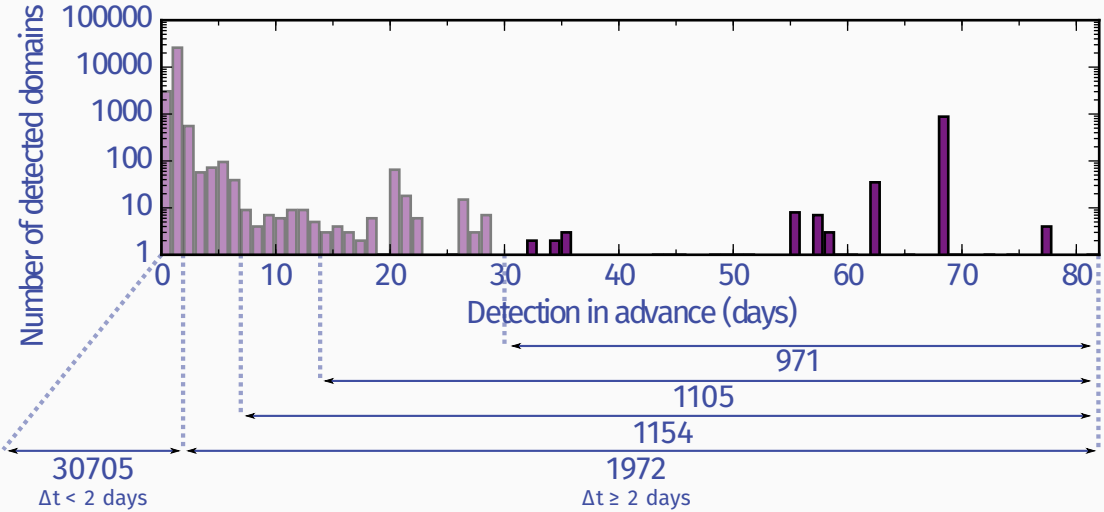
RBL comparison (2 month period)



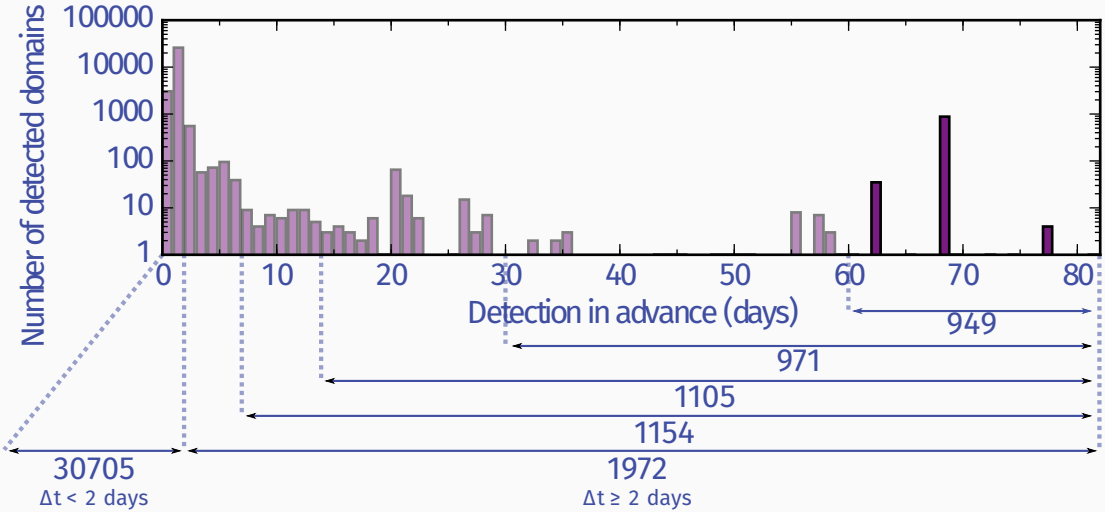
RBL comparison (2 month period)



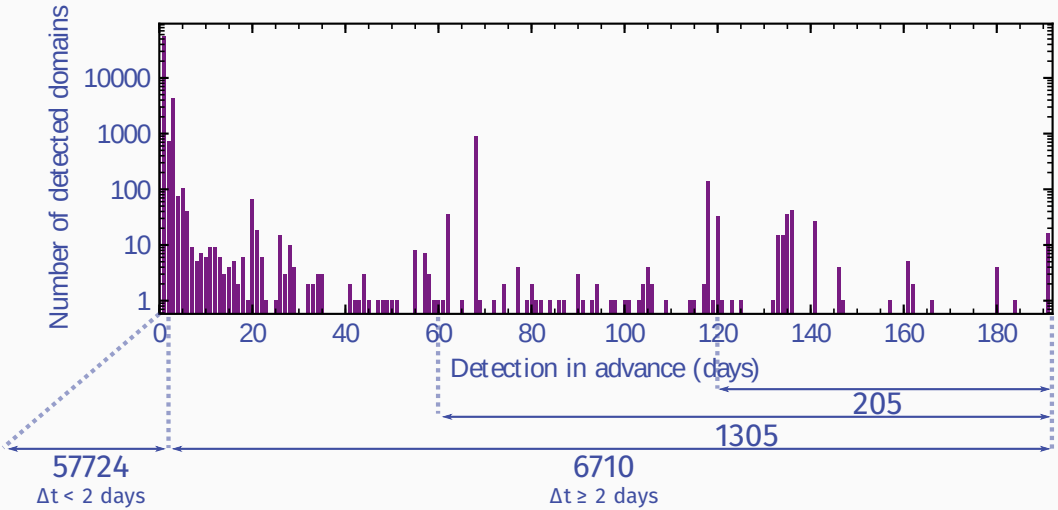
RBL comparison (2 month period)



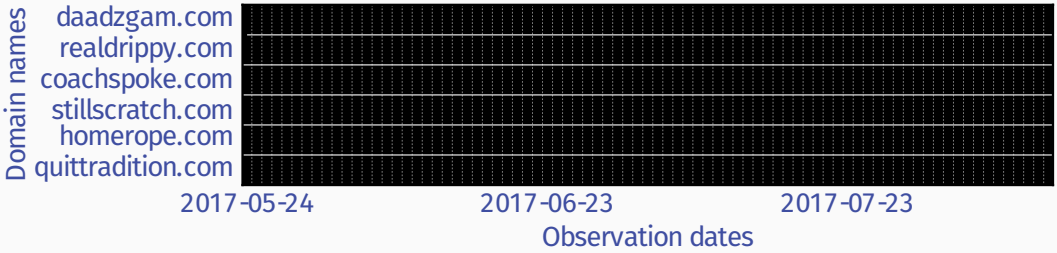
RBL comparison (2 month period)



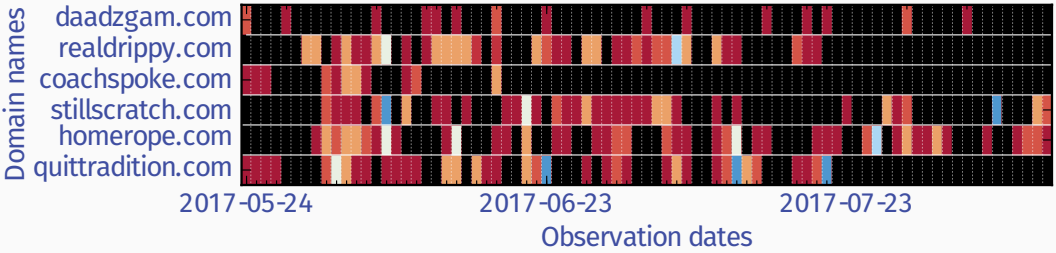
RBL comparison (9 month period)



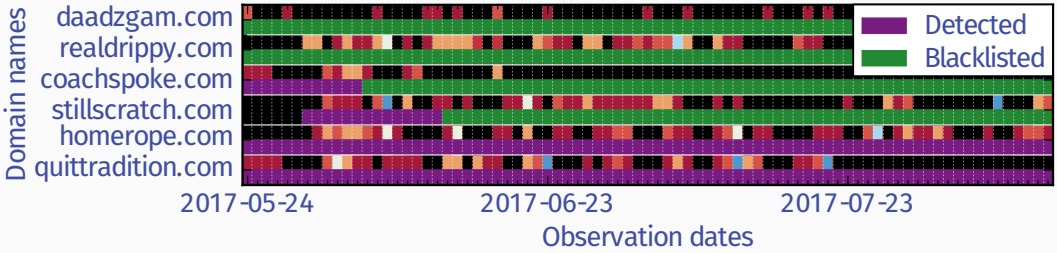
SURFnet evaluation



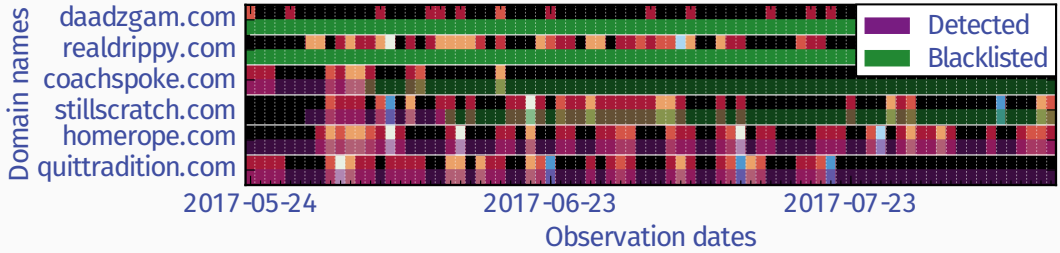
SURFnet evaluation

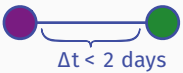


SURFnet evaluation



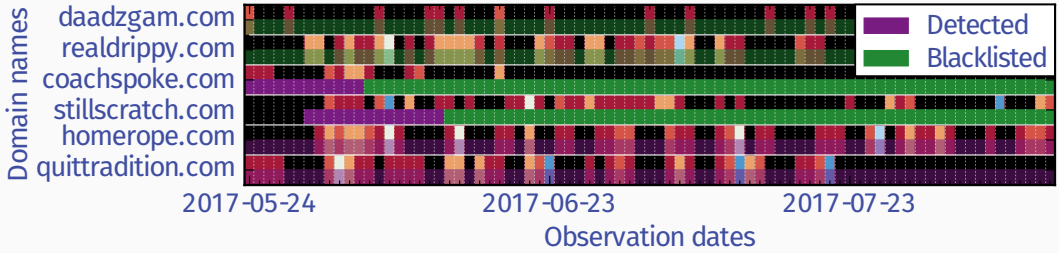
SURFnet evaluation

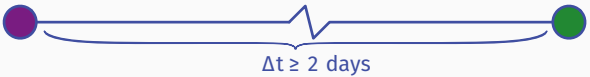




- 45% of received emails fall in this category
- 18% of observed domains fall in this category

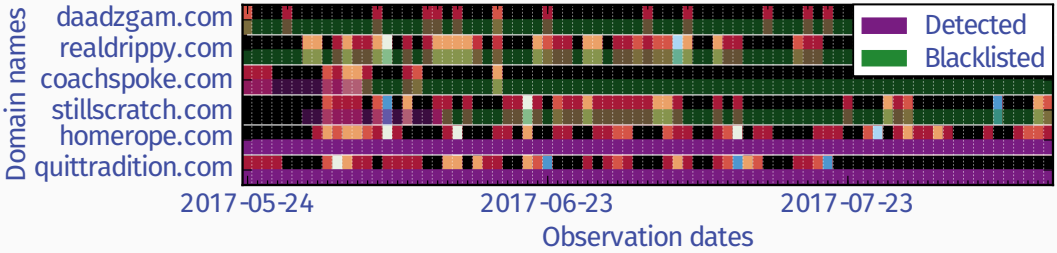
SURFnet evaluation





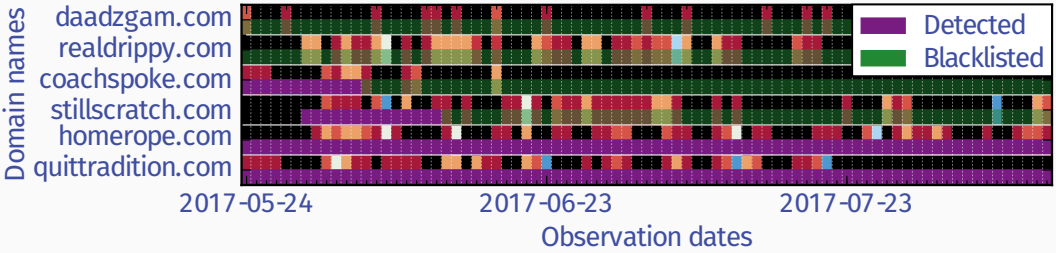
- 17% of received emails fall in this category
- 26% of observed domains fall in this category

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

Use case: DDoS domains

DDoS domains

In DDoS attacks the amplification factor is important.

Domains crafted for DDoS attacks typically have:

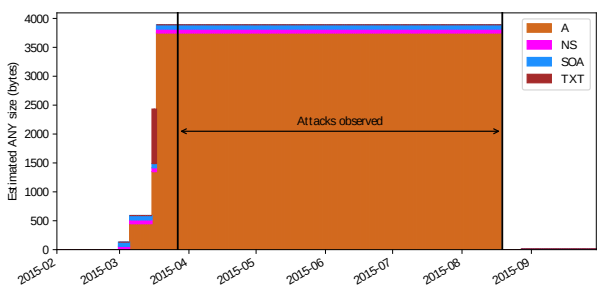
DDoS domains

In DDoS attacks the amplification factor is important.

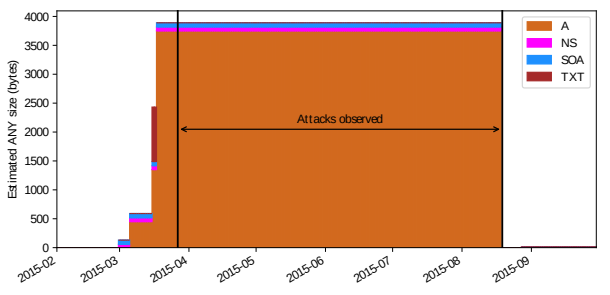
Domains crafted for DDoS attacks typically have:

- Many records
- Long (TXT) records

Lifetime of a DDoS domain

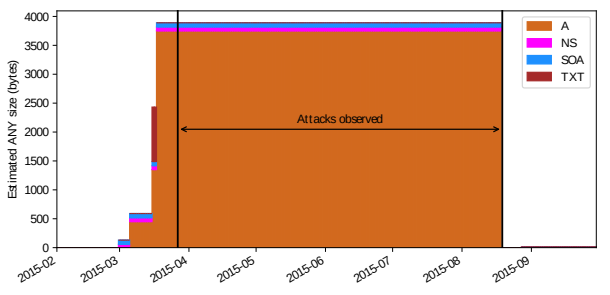


Lifetime of a DDoS domain



Possible methodology could be:

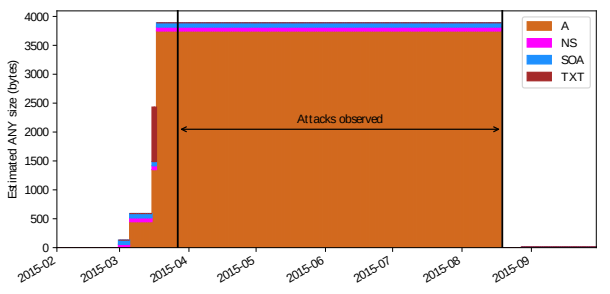
Lifetime of a DDoS domain



Possible methodology could be:

1. Filter domains with more than average number of records, or longer than average TXT record

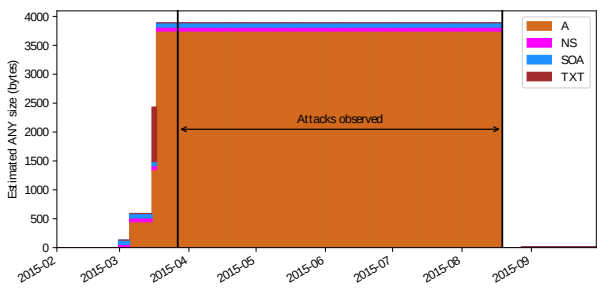
Lifetime of a DDoS domain



Possible methodology could be:

1. Filter domains with more than average number of records, or longer than average TXT record
2. Gather the records for the past X days

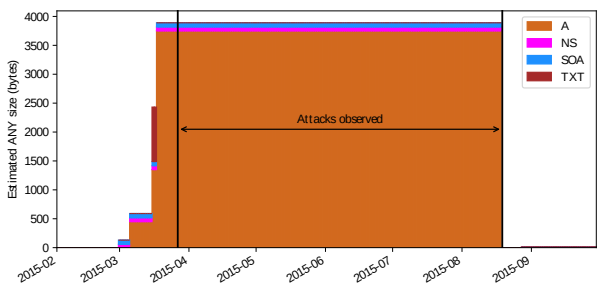
Lifetime of a DDoS domain



Possible methodology could be:

1. Filter domains with more than average number of records, or longer than average TXT record
2. Gather the records for the past X days
3. Determine trend lines

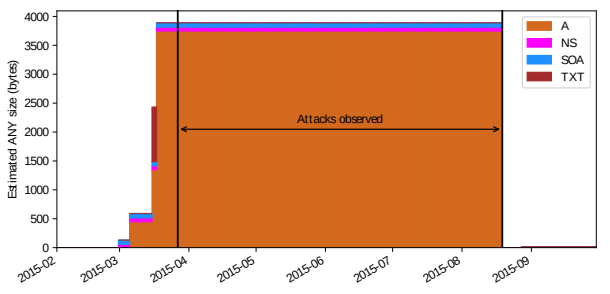
Lifetime of a DDoS domain



Possible methodology could be:

1. Filter domains with more than average number of records, or longer than average TXT record
2. Gather the records for the past X days
3. Determine trend lines
4. Predict the size of the domain, say, ten days from now

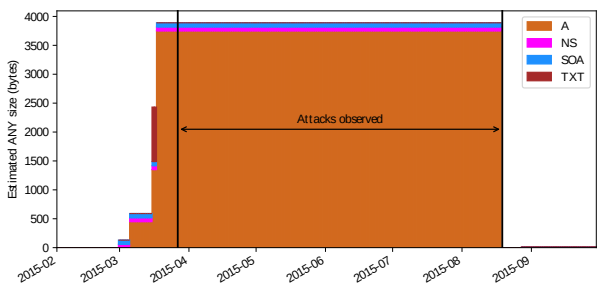
Lifetime of a DDoS domain



Possible methodology could be:

1. Filter domains with more than average number of records, or longer than average TXT record
2. Gather the records for the past X days
3. Determine trend lines
4. Predict the size of the domain, say, ten days from now
5. Flag the domain if the predicted size is above a certain threshold

Lifetime of a DDoS domain

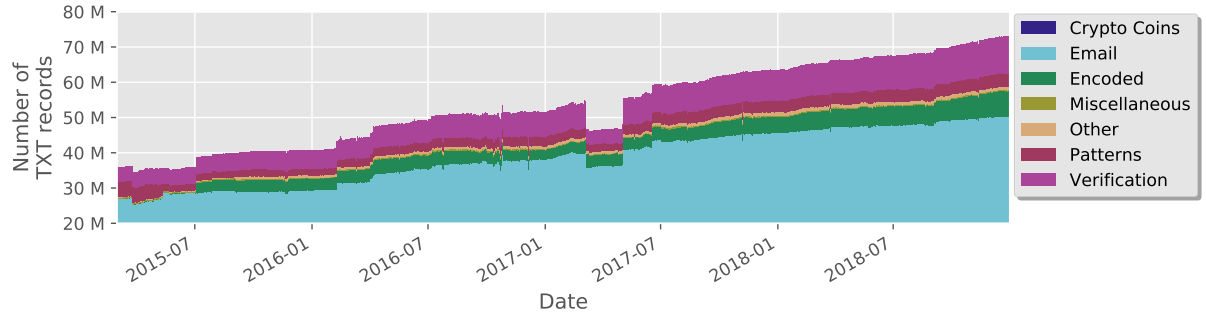


Possible methodology could be:

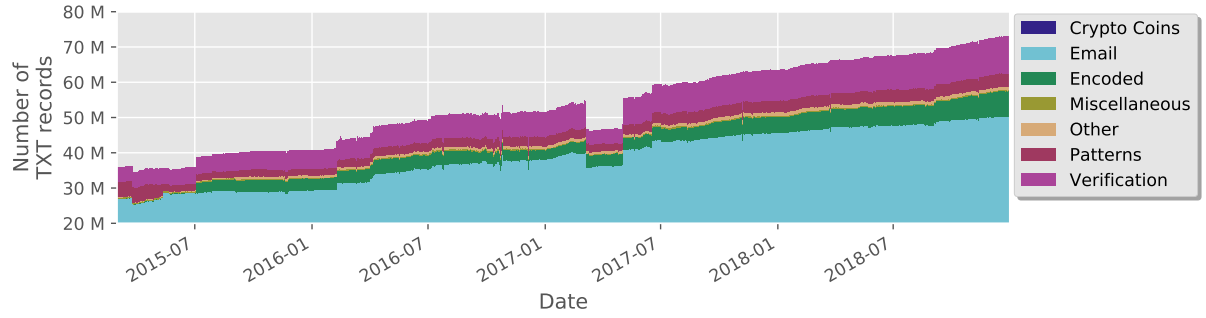
1. Filter domains with more than average number of records, or longer than average TXT record
2. Gather the records for the past X days
3. Determine trend lines
4. Predict the size of the domain, say, ten days from now
5. Flag the domain if the predicted size is above a certain threshold

Use case: DNS TXT records

DNS TXT records

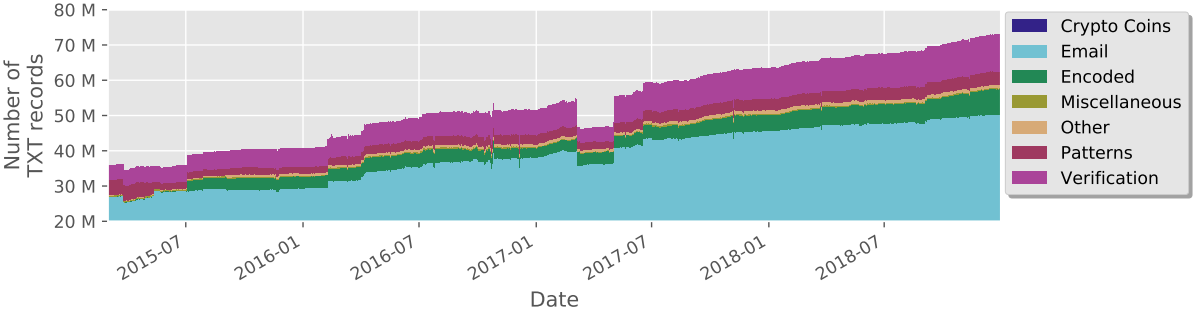


DNS TXT records



- Majority of TXT records are related to email (~70%)

DNS TXT records



- Majority of TXT records are related to email (~70%)
- 1.2% falls in the ‘other’ category

DNS TXT records

One of the hightlights of this ‘other’ category is single character records.

DNS TXT records

One of the hightlights of this ‘other’ category is single character records.

- More than 278K TXT records consisting of a single charcter

DNS TXT records

One of the hightlights of this ‘other’ category is single character records.

- More than 278K TXT records consisting of a single charcter
- Majority contains a ~

DNS TXT records

One of the hightlights of this ‘other’ category is single character records.

- More than 278K TXT records consisting of a single charcter
- Majority contains a ~
- Almost all of these domains are hosted in the same AS

MalwareHunterTeam

@malwrhunterteam

Official MHT Twitter account. CyberTracker is closed. Check out ID Ransomware (created by @demonslay335). Want to talk with us? DM @0x7fff9 anytime.

id-ransomware.malwarehunterteam.com

Joined October 2014

JAMESWT @JAMESWT_MHT · 17 May 2017

😎😂

208.91.197.46

204.11.56.48

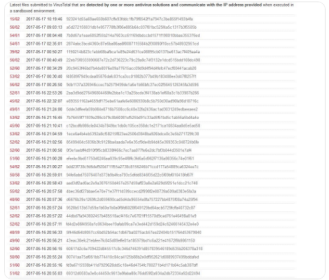
141.8.225.244

virustotal.com/en/ip-address/…

virustotal.com/en/ip-address/…

virustotal.com/en/ip-address/…

@malwrhunterteam



2 5 8

MalwareHunterTeam

@malwrhunterteam

Follow

Replying to @JAMESWT_MHT

AS for all: 40034 (Confluence Networks Inc) - nothing more to say.

3:56 AM - 17 May 2017

1 Retweet **3** Likes

1 3

DNS TXT records

Are these records useful for threat detection?

- Generally, no
- The ‘~’; case could be an identifier for domains from a specific AS

Use case: Combo-squat domains

Combo-squat: What is a combo-squat domain?

Many types of squatting domains:

Type	Example (target: utwente.nl)
Typosquatting	utwent.nl
Combosquatting	utwente-login.nl
Bitsquatting	utwenpe.nl
Homograph-Based squatting	utvvente.nl

Combo-squat: A general approach?

We started out by developing a general machine-learning based detection model.

Combo-squat: A general approach?

We started out by developing a general machine-learning based detection model.
Feeding the detection model a list of trademarks worked a lot better!

Combo-squat: A general approach?

We started out by developing a general machine-learning based detection model. Feeding the detection model a list of trademarks worked a lot better!

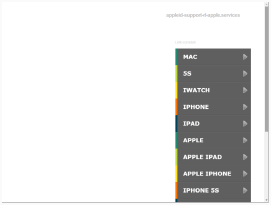
Trademark	Number of domains
Apple	8751
Paypal	1241
Microsoft	711

Combo-squat: The problems with a generic approach

However, a larger problem is the life time of a combosquat domain.

Combo-squat: The problems with a generic approach

However, a larger problem is the life time of a combosquat domain.



(a) Newly registered

(b) Under construction

(c) Up & running

(d) Inactive

Use cases

Where it works:

- Snowshoe spam domains

Use cases

Where it works:

- Snowshoe spam domains

Where it might work:

- DDoS Domains
- Malicious TXT records

Use cases

Where it works:

- Snowshoe spam domains

Where it might work:

- DDoS Domains
- Malicious TXT records

Where it doesn’t work:

- Combo-squat domains

Reflection

Reflection

What have we learned from these use cases?

Reflection

What have we learned from these use cases?

- The data needs to contain hints

Reflection

What have we learned from these use cases?

- The data needs to contain hints
- This approach works for relatively long setup times (in our case >1d)

Improvement?

We realize that our solution is not perfect.

Improvement?

We realize that our solution is not perfect.

We think the “ultimate” solution is to combine passive and active measurements.

Improvement?

We think the “ultimate” solution is to combine passive and active measurements.

Use proactive threat detection to prime passive approaches.

Conclusion

Conclusion

We should move towards proactive threat detection.

Conclusion

We should move towards proactive threat detection.

- Pick up on clues of an upcoming attack
- Look beyond your own network

Conclusion

We should move towards proactive threat detection.

- Pick up on clues of an upcoming attack
- Look beyond your own network

Use the early warning from these methods to feed passive detection approaches.

- Combine the high level of detail of passive measurements with the time advantage from active measurements

Future work

Future work

- Research other areas of attack:
  - DDoS domains
  - C&C domains
  - etc.
- Collaborate with pDNS @ CERT.at
  - Are there more benefits of combining passive and active (DNS) measurements?

Thank you

Thank you for listening!

Any questions?

Contact details


tide-project.nl


o.i.vandertoorn@utwente.nl