Threat Identification Using Active DNS Measurements

AIMS 2018: Ph.D. track

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

November 13, 2018

University of Twente, Design and Analysis of Communication Systems

Who am I?

Who am I?

Who am I?

HELLO
My name is
Olivier

- Bachelor Electrical Engineering @University of Twente

Who am I?

- Bachelor Electrical Engineering @University of Twente
    - A First Look at HTTP(S) Intrusion Detection Using NetFlow/IPFIX (IM 2015)

Who am I?

- Bachelor Electrical Engineering @University of Twente
    - A First Look at HTTP(S) Intrusion Detection Using NetFlow/IPFIX (IM 2015)
- Master Electrical Engineering @University of Twente

Who am I?

- Bachelor Electrical Engineering @University of Twente
  - A First Look at HTTP(S) Intrusion Detection Using NetFlow/IPFIX (IM 2015)
- Master Electrical Engineering @University of Twente
  - Melting The Snow: Using Active DNS Measurements to Detect Snowshoe Spam Domains (NOMS 2018)

Who am I?

- Bachelor Electrical Engineering @University of Twente
  - A First Look at HTTP(S) Intrusion Detection Using NetFlow/IPFIX (IM 2015)
- Master Electrical Engineering @University of Twente
  - Melting The Snow: Using Active DNS Measurements to Detect Snowshoe Spam Domains (NOMS 2018)

- Voluntary System Administrator at:

Who am I?

- Bachelor Electrical Engineering @University of Twente
    - A First Look at HTTP(S) Intrusion Detection Using NetFlow/IPFIX (IM 2015)
- Master Electrical Engineering @University of Twente
    - Melting The Snow: Using Active DNS Measurements to Detect Snowshoe Spam Domains (NOMS 2018)

- Voluntary System Administrator at:
    - Scintilla (study association of Electrical Engineering)

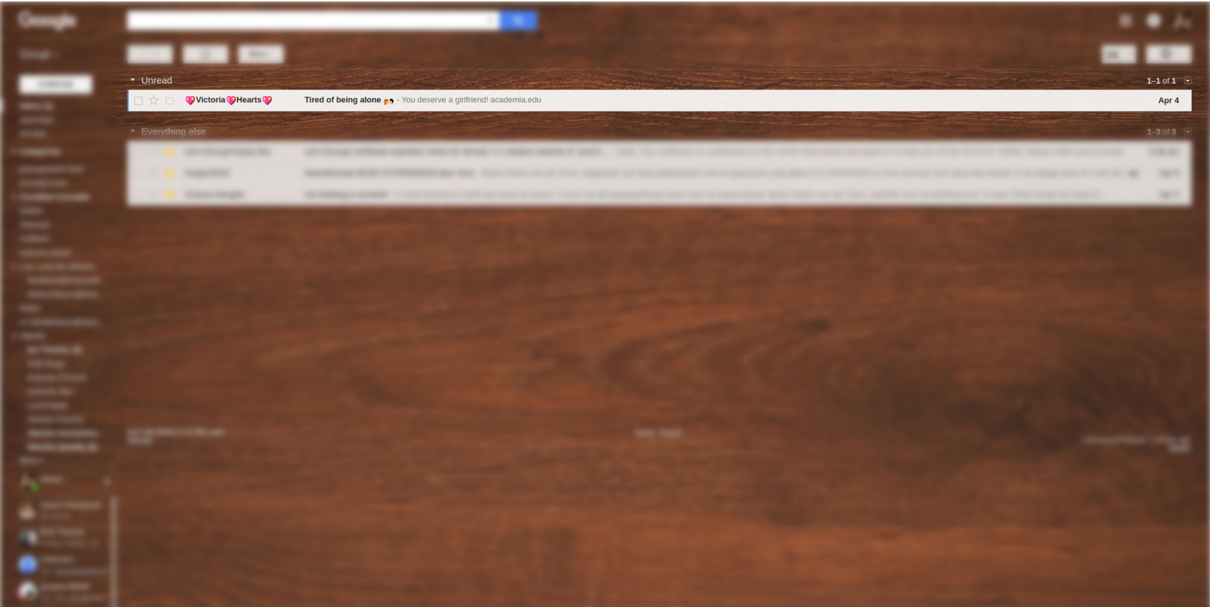
Who am I?

- Bachelor Electrical Engineering @University of Twente
    - A First Look at HTTP(S) Intrusion Detection Using NetFlow/IPFIX (IM 2015)
- Master Electrical Engineering @University of Twente
    - Melting The Snow: Using Active DNS Measurements to Detect Snowshoe Spam Domains (NOMS 2018)

- Voluntary System Administrator at:
    - Scintilla (study association of Electrical Engineering)
    - Student Network Twente (SNT)

Introduction

Introduction

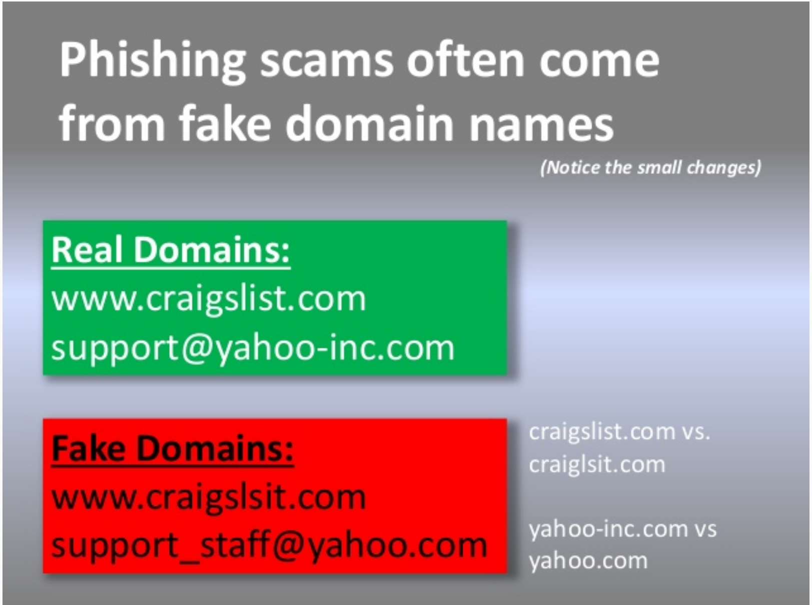


Unread 1–1 of 1

□ ☆ 💖Victoria💖Hearts💖 **Tired of being alone** 👫 - You deserve a girlfriend! academia.edu **Apr 4**

Everything else 1–3 of 3

Introduction



Phishing scams often come from fake domain names

(Notice the small changes)

Real Domains:

www.craigslist.com

support@yahoo-inc.com

Fake Domains:

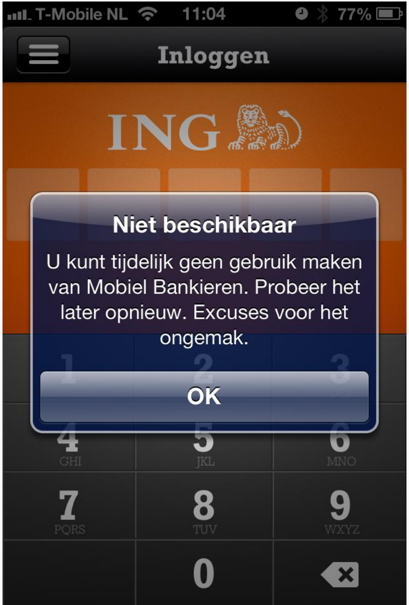
www.craigslsit.com

support_staff@yahoo.com

craigslist.com vs.
craiglsit.com

yahoo-inc.com vs
yahoo.com

Introduction



Niet beschikbaar

U kunt tijdelijk geen gebruik maken van Mobiel Bankieren. Probeer het later opnieuw. Excuses voor het ongemak.

Introduction

What do DDoS, phishing and spam attacks have in common?

Introduction

What do DDoS, phishing and spam attacks have in common?

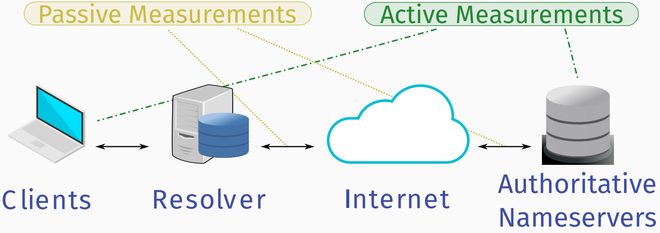
They leave traces...

Introduction

What do DDoS, phishing and spam attacks have in common?

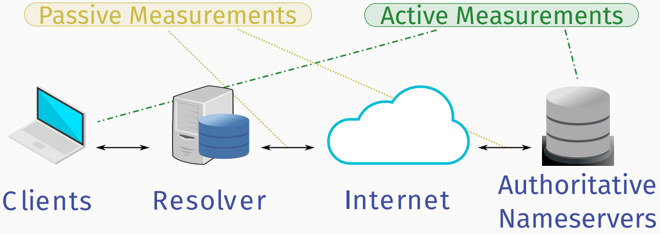
They leave traces... in the DNS!

Introduction



Passive DNS measurements

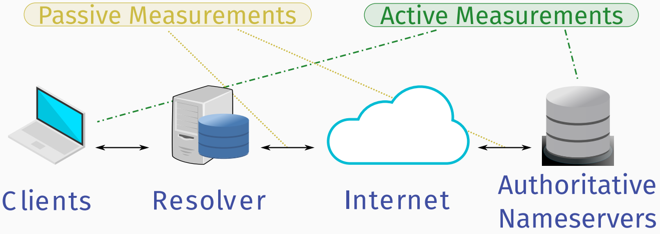
Introduction



Passive DNS measurements

- Detailed DNS usage
- Usage biases
- Time series are difficult

Introduction

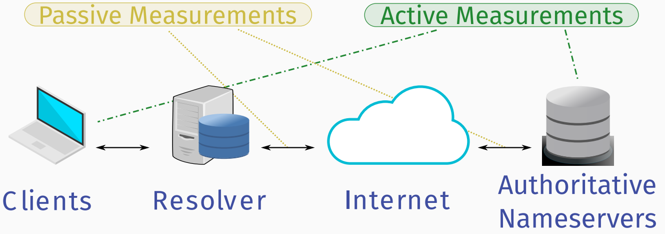


Passive DNS measurements

- Detailed DNS usage
- Usage biases
- Time series are difficult

Active DNS measurements

Introduction



Passive DNS measurements

- Detailed DNS usage
- Usage biases
- Time series are difficult

Active DNS measurements

- Greater overview
- Possibility of a time-advantage
- Less detailed than passive measurements

Introduction

Our proposal:

Introduction

Our proposal:

Pro-active threat identification of malicious domains through active DNS measurements.

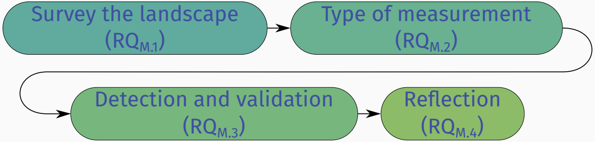
Research questions

Research questions

RQ_M: How can we use active DNS measurements to pro-actively identify malicious domains, and what are the benefits of such an approach?

Research questions

RQ_M: How can we use active DNS measurements to pro-actively identify malicious domains, and what are the benefits of such an approach?



Survey the landscape ($RQ_{M.1}$)
Type of measurement ($RQ_{M.2}$)
Detection and validation ($RQ_{M.3}$)
Reflection ($RQ_{M.4}$)

Sub-research question 1

We would like to detect all attacks. However, not all attacks make use of the DNS.

Sub-research question 1

We would like to detect all attacks. However, not all attacks make use of the DNS.

$\text{RQ}_{\text{M.1}}$: Which attacks make use of DNS and how do they use it?

Sub-research question 1

We would like to detect all attacks. However, not all attacks make use of the DNS.

$\mathrm{RQ}_{M.1}$: Which attacks make use of DNS and how do they use it?

- Survey literature

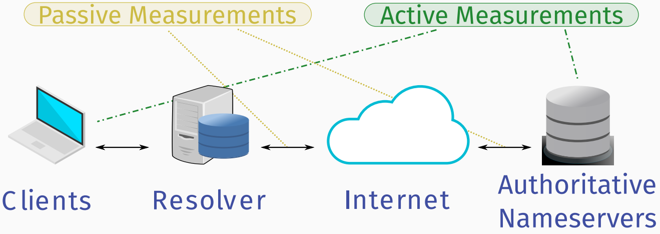
Sub-research question 1

We would like to detect all attacks. However, not all attacks make use of the DNS.

$\text{RQ}_{\text{M.1}}$: Which attacks make use of DNS and how do they use it?

- Survey literature
- Interview experts in the field

Sub-research question 2



Answering **RQ$_{M.1}$** gives a list of attacks which makes use of the DNS in some way.

There are (roughly) two types of DNS measurements, active and passive.

Sub-research question 2

$RQ_{M.2}$: What are the strengths and weaknesses of both types of DNS measurements with respect to the attacks?

Sub-research question 2

$\text{RQ}_{\text{M.2}}$: What are the strengths and weaknesses of both types of DNS measurements with respect to the attacks?

As a starting point we want to use Entrada and OpenINTEL to analyse how well both approaches fare in the detection of the surveyed attacks.

Sub-research question 3

We want to be able to perform detections on the entire DNS namespace.

Attacks are dynamic, therefore our detection method needs to be dynamic too.

Sub-research question 3

We want to be able to perform detections on the entire DNS namespace.

Attacks are dynamic, therefore our detection method needs to be dynamic too.

$\mathrm{RQ}_{\mathrm{M.3}}$: How can we perform efficient, large-scale, detections using Machine Learning and how do we validate these detections?

Sub-research question 3

We want to be able to perform detections on the entire DNS namespace.

Attacks are dynamic, therefore our detection method needs to be dynamic too.

$RQ_{M.3}$: How can we perform efficient, large-scale, detections using Machine Learning and how do we validate these detections?

- Evaluate different classifier algorithms

Sub-research question 3

We want to be able to perform detections on the entire DNS namespace.

Attacks are dynamic, therefore our detection method needs to be dynamic too.

$RQ_{M.3}$: How can we perform efficient, large-scale, detections using Machine Learning and how do we validate these detections?

- Evaluate different classifier algorithms
- Compare results with established blacklists

Sub-research question 4

At this point we have a list of bad domains.

Are we able to infer identifiable information about the parties behind the domains by clustering domains together?

Sub-research question 4

At this point we have a list of bad domains.

Are we able to infer identifiable information about the parties behind the domains by clustering domains together?

$\text{RQ}_{\text{M.4}}$: What additional information can be obtained by clustering similar domain-configurations together?

Sub-research question 4

At this point we have a list of bad domains.

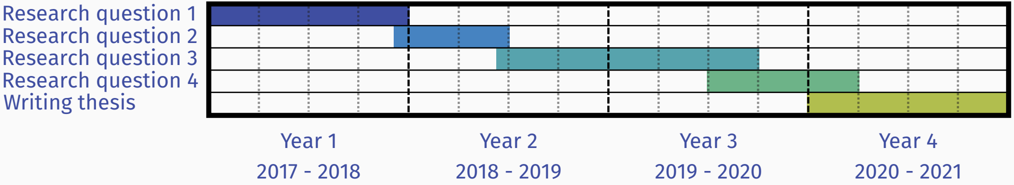
Are we able to infer identifiable information about the parties behind the domains by clustering domains together?

$\text{RQ}_{\text{M.4}}$: What additional information can be obtained by clustering similar domain-configurations together?

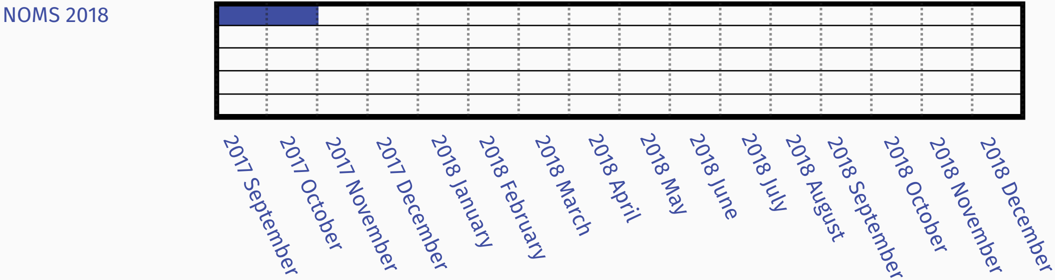
Cluster domains with similar configurations together, then analyse the similarities and differences.

Planning

Planning: Global



Planning: NOMS 2018

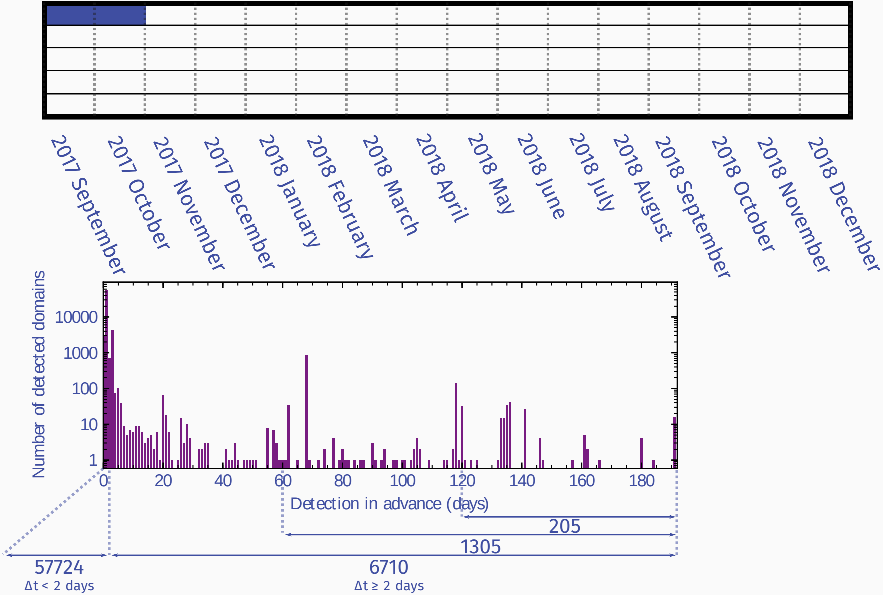


NOMS 2018 – September & October 2017 (Published)

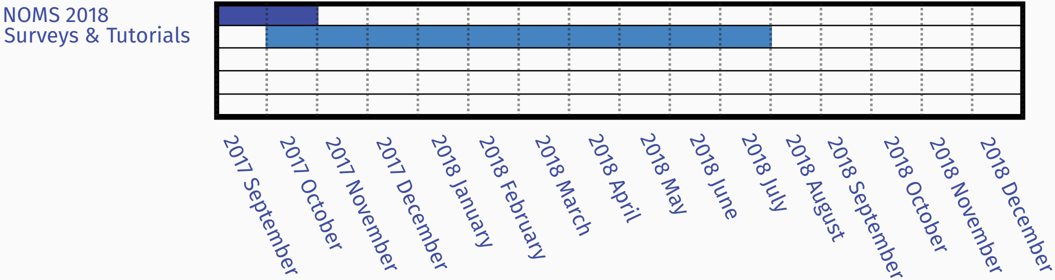
Detection of Snowshoe Spam by using active DNS measurements. Snowshoe Spam domains using SPF typically feature many records.

Planning: NOMS 2018

NOMS 2018



Planning: Surveys & Tutorials

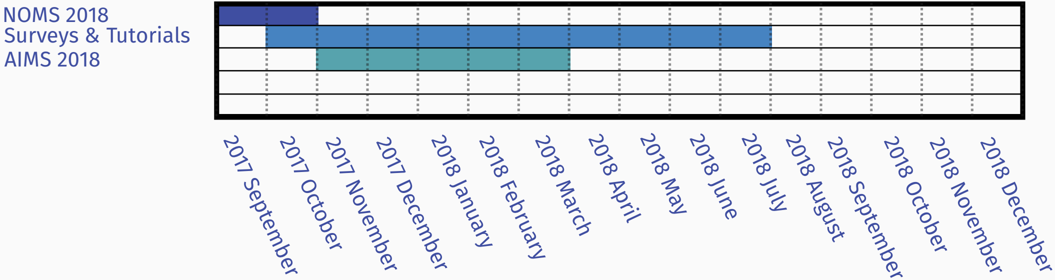


Surveys & Tutorials – October 2017, July 2018 (WIP)

Survey paper looking at state-of-the-art attack detection using either active, or passive, DNS measurements.

Based on research question **$\text{RQ}_{\text{M.1}}$**.

Planning: AIMS 2018

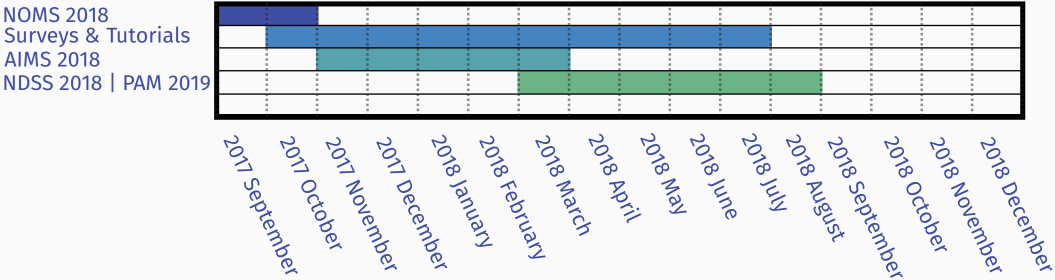


AIMS 2018 – November 2017, March 2018 (Accepted)

Ph.D. project proposal for the TIDE project.

Formal definition of the research questions.

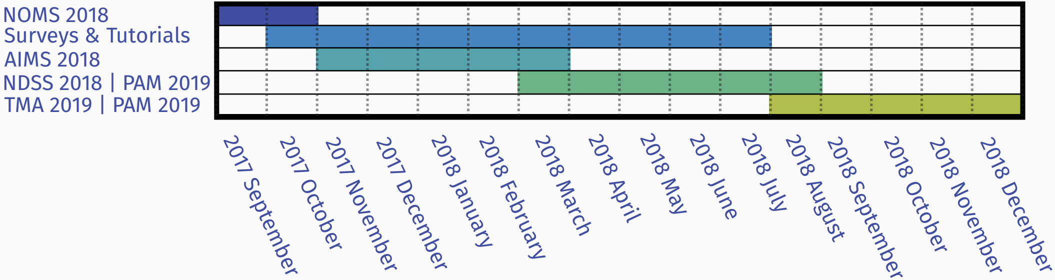
Planning: NDSS 2018 | PAM 2019



NDSS 2018 | PAM 2019 – March 2018, August 2018 (Planned)

Paper about the detection of malware code in DNS TXT resource records.

Planning: TMA 2019



TMA 2019 | PAM 2019 – August 2018, December 2018 (Planned)

Measurement paper. Here we want to put the theories learned from the survey into practice.

Based on research question **RQ$_{M.2}$**.

Conclusion

Conclusion

We want to make the Internet a safer place by pro-actively identifying malicious DNS domains.

Discussion?

Discussion?

- What attacks can’t we miss in our research?

Discussion?

- What attacks can’t we miss in our research?
- Do you have recommendations on how to validate a predictive blacklist?

Discussion?

- What attacks can’t we miss in our research?
- Do you have recommendations on how to validate a predictive blacklist?
- How would you use such a predictive blacklist?

Extra slides

Why active DNS measurements?

Why study active DNS measurements? And not go with passive DNS measurements like everyone else?

Why active DNS measurements?

Why study active DNS measurements? And not go with passive DNS measurements like everyone else?

- Approach for primer domains.

Why active DNS measurements?

Why study active DNS measurements? And not go with passive DNS measurements like everyone else?

- Approach for primer domains.
- We believe that the configuration of a domain can give a real insight into the purpose of a domain.