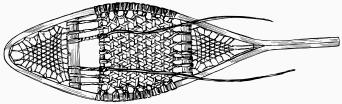


Combating Snowshoe Spam with Fire

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

November 13, 2018

University of Twente, Design and Analysis of Communication Systems

ICT OPEN 2018

Overview

Introduction

Methodology

Results

Conclusions

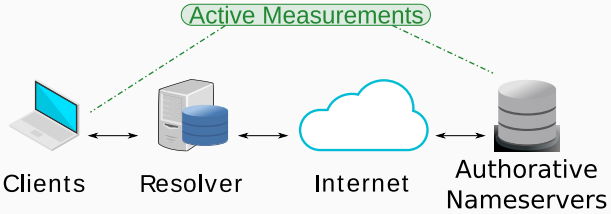
Introduction

Background Info

- Active DNS Measurements

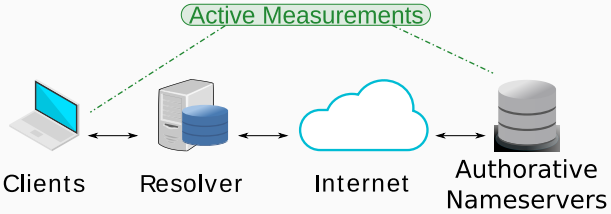
Background Info

- Active DNS Measurements



Background Info

- Active DNS Measurements
- Snowshoe Spam



Starting Point

- Snowshoe spam is hard to detect

Starting Point

- Snowshoe spam is hard to detect
- Sender Policy Framework (SPF)

Starting Point

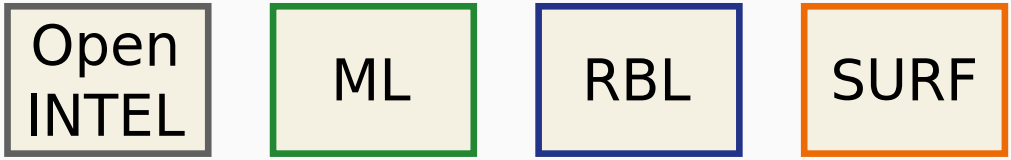
- Snowshoe spam is hard to detect
- Sender Policy Framework (SPF)
- DNS domain

Research Question

How can we detect snowshoe spam through active DNS measurements?

Methodology

Methodology



- Active DNS Measurement Platform

OpenINTEL

- Active DNS Measurement Platform
- Queries more than 60% of registered domain names

Datasets & Features

- Two types of datasets
  - Labeled
  - Unlabeled

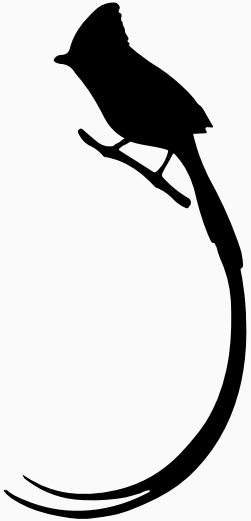
Datasets & Features

- Two types of datasets
  - Labeled
  - Unlabeled
- 37 Features

Datasets & Features

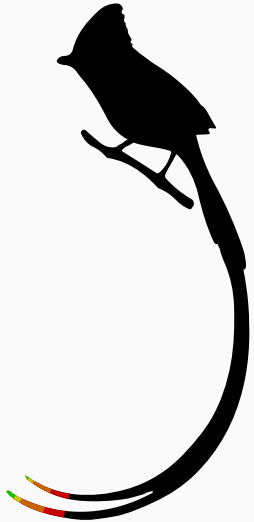
- Two types of datasets
  - Labeled
  - Unlabeled
- 37 Features
- Long Tail Analysis

Long Tail Analysis



The long tail of the DNS

Long Tail Analysis



The **long** tail of the DNS

Machine Learning

- Trained and evaluated many classifier algorithms

Machine Learning

- Trained and evaluated many classifier algorithms
- Ranked performance based on ‘precision’ metric

Precision

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

Machine Learning

- Trained and evaluated many classifier algorithms
- Ranked performance based on ‘precision’ metric
- Selected AdaBoost Classifier as classifier of choice (110 false positives out of 10851 ham domains)

Realtime Blackhole List

- DNS based way of hosting a blacklist

Realtime Blackhole List

- DNS based way of hosting a blacklist
- Daily detections

Realtime Blackhole List

- DNS based way of hosting a blacklist
- Daily detections
- Compared to other blacklists

SURF

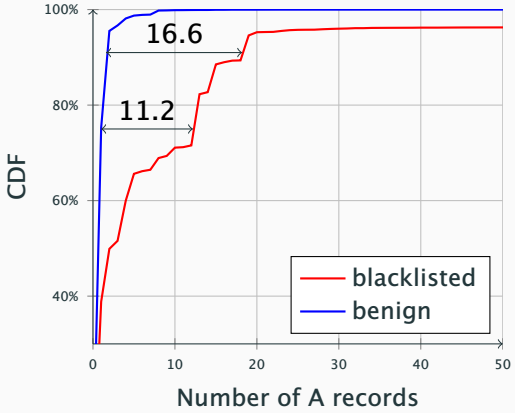
- SURFmailfilter

SURF

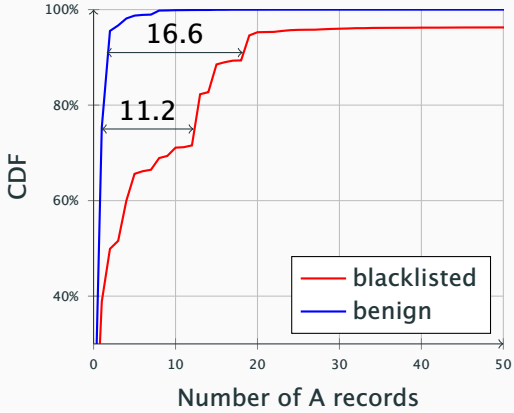
- SURFmailfilter
- Initially in evaluation mode

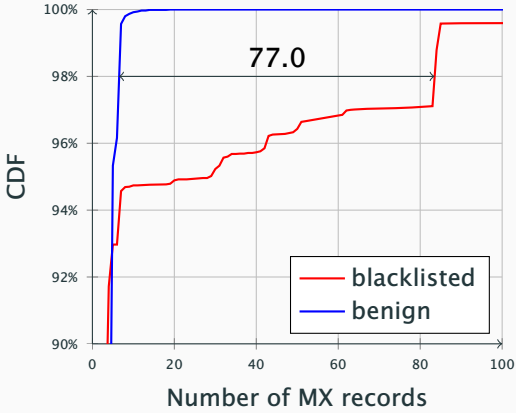
Results

Comparison training data

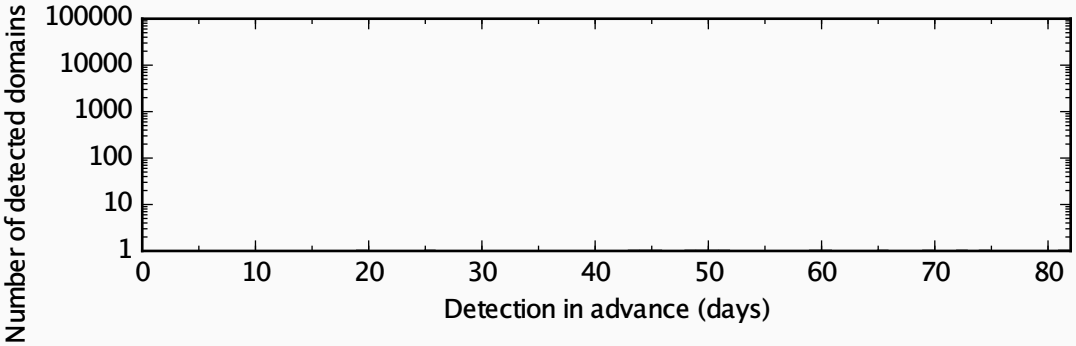


Comparison training data

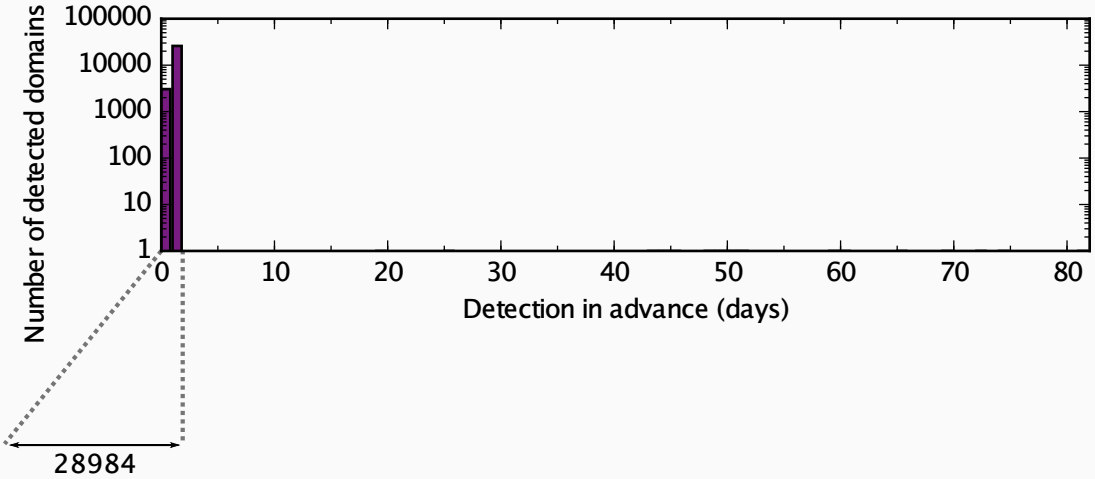




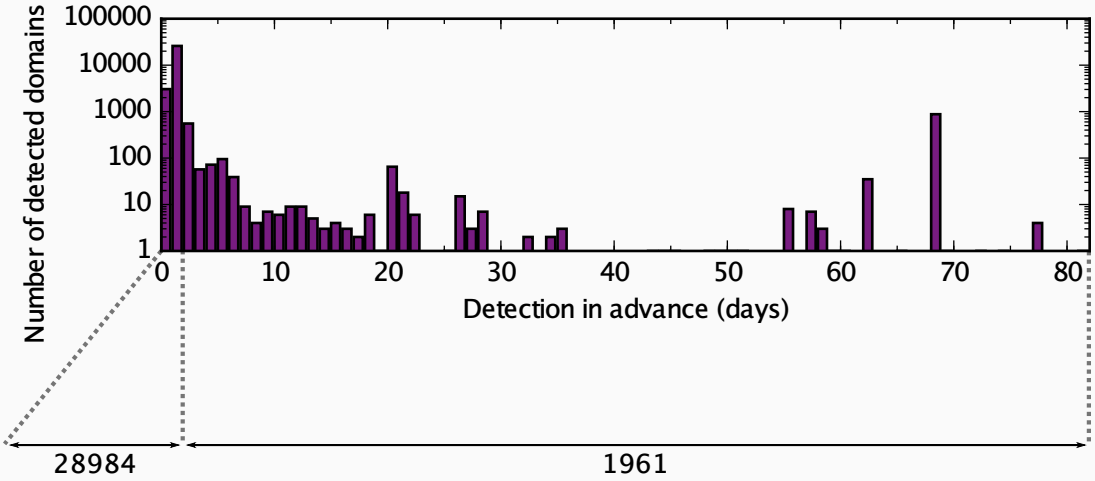
Early Detection



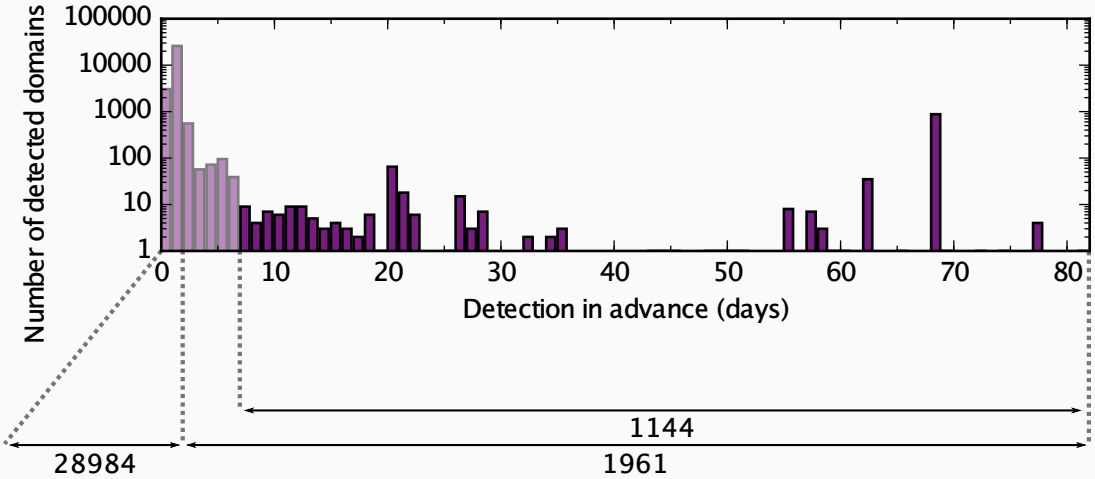
Early Detection



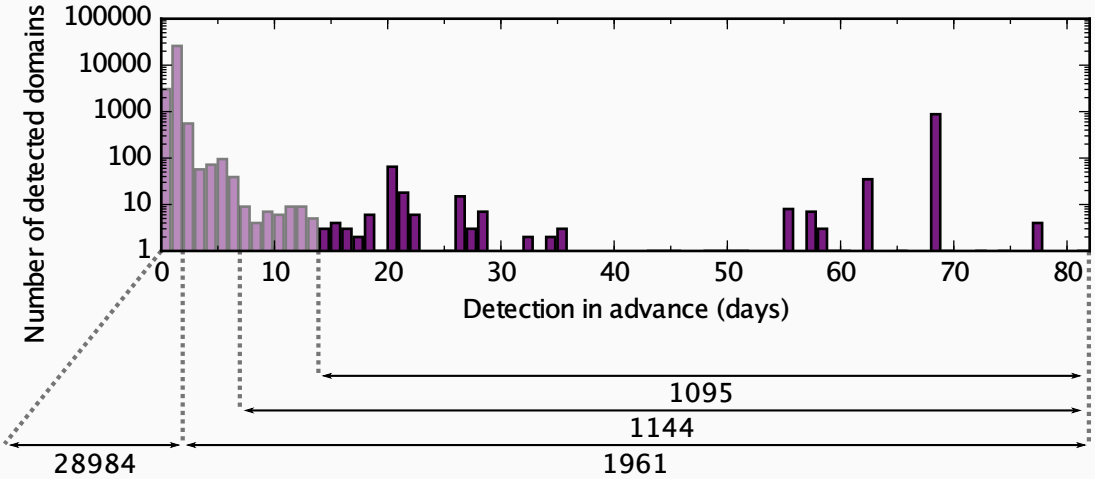
Early Detection



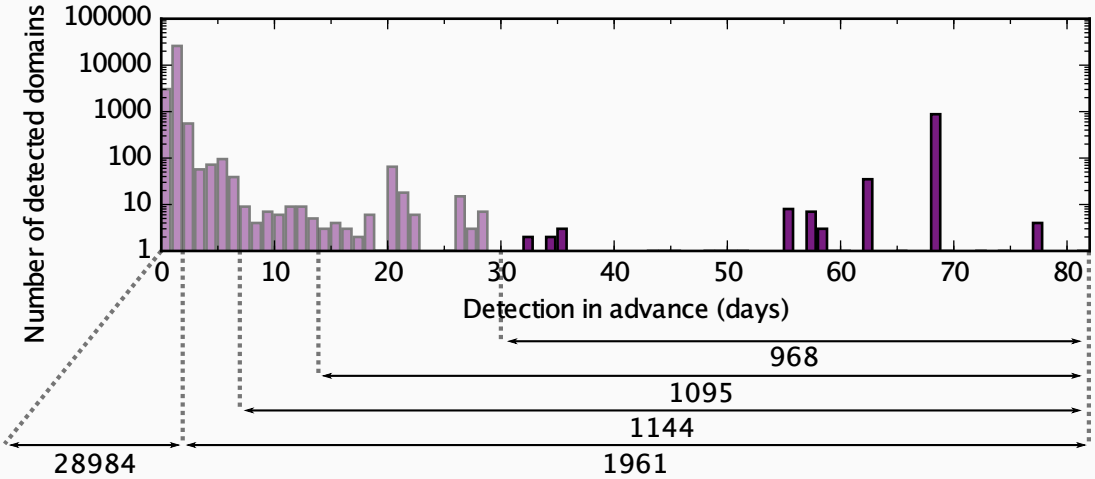
Early Detection



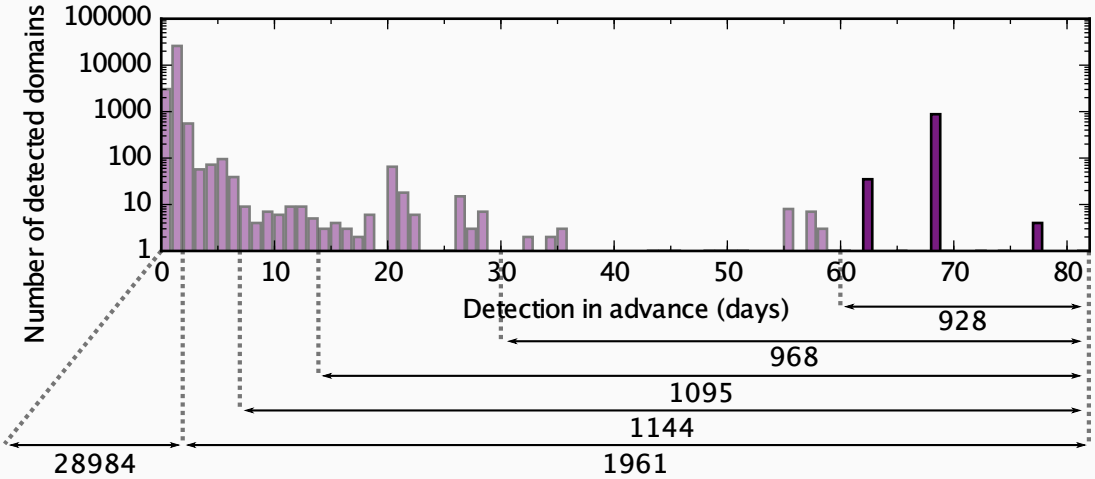
Early Detection



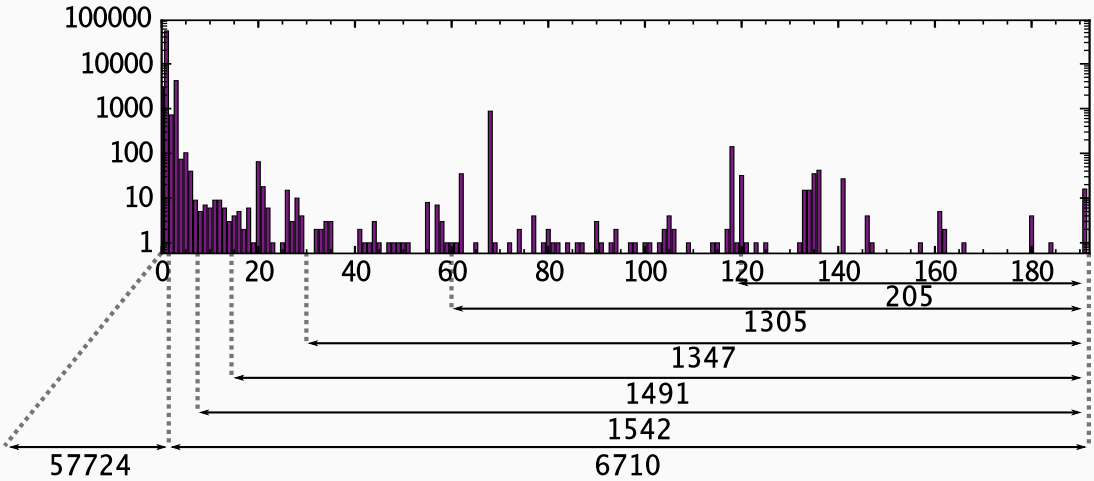
Early Detection



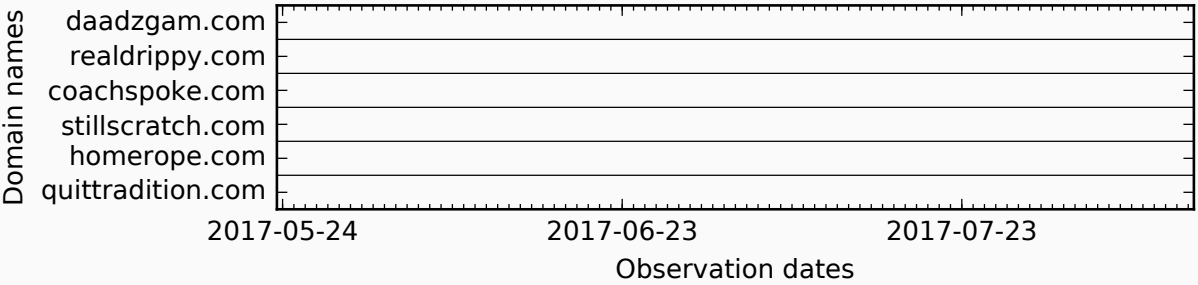
Early Detection



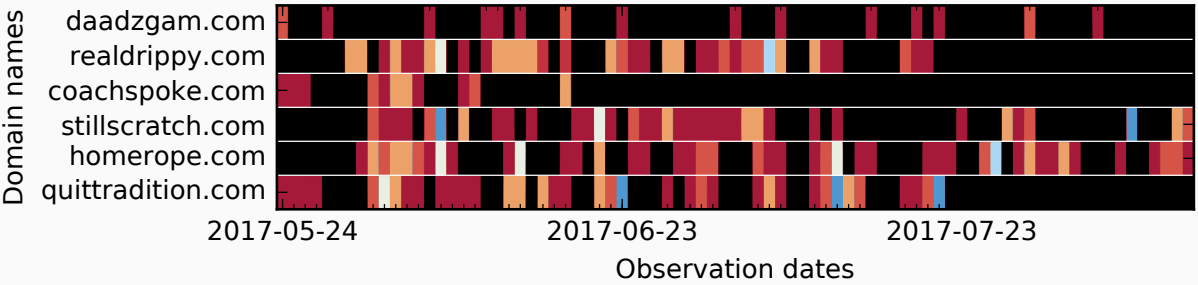
Early Detection (update)



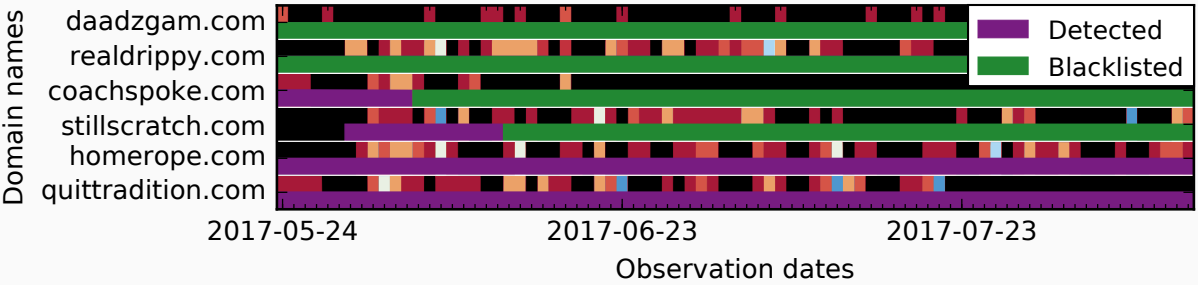
SURF Results



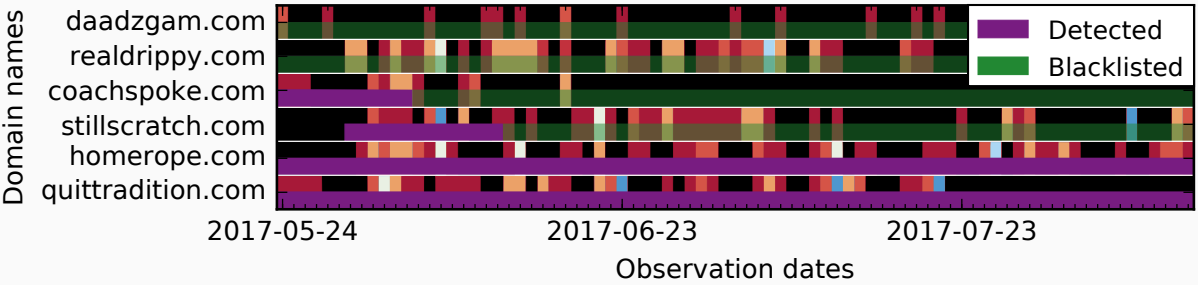
SURF Results



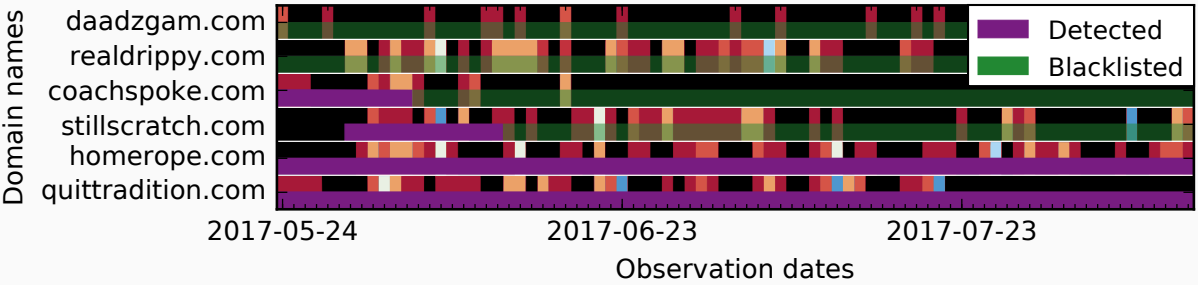
SURF Results



SURF Results

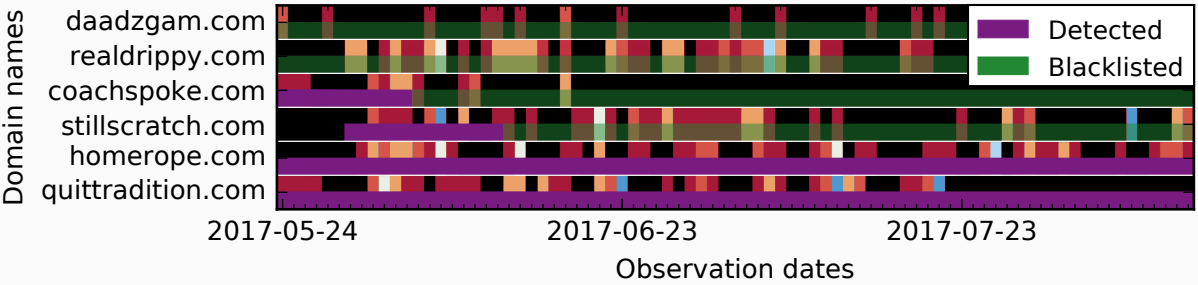


SURF Results



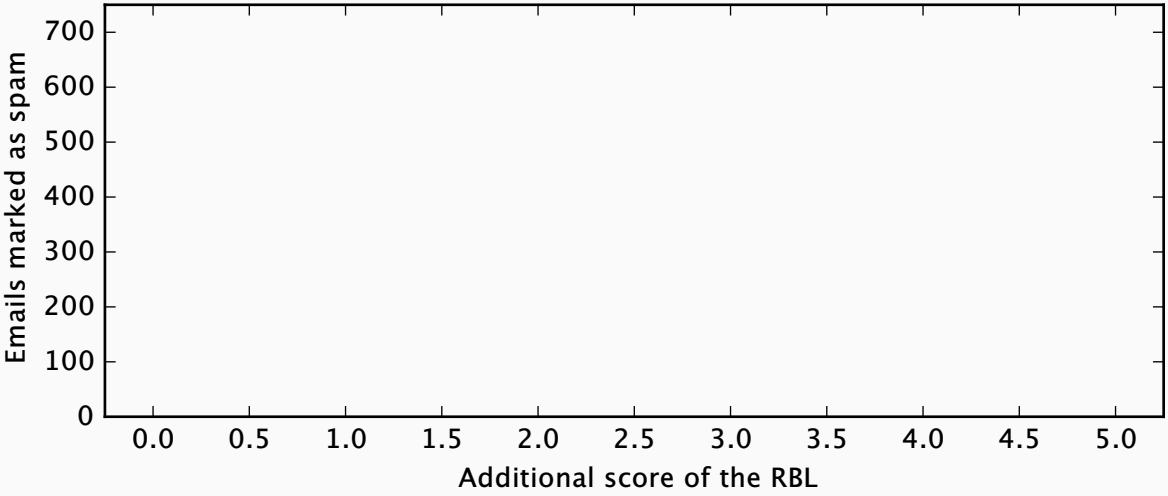
- 1080 emails
- 447 (41.39%) emails with a score of five or higher

SURF Results

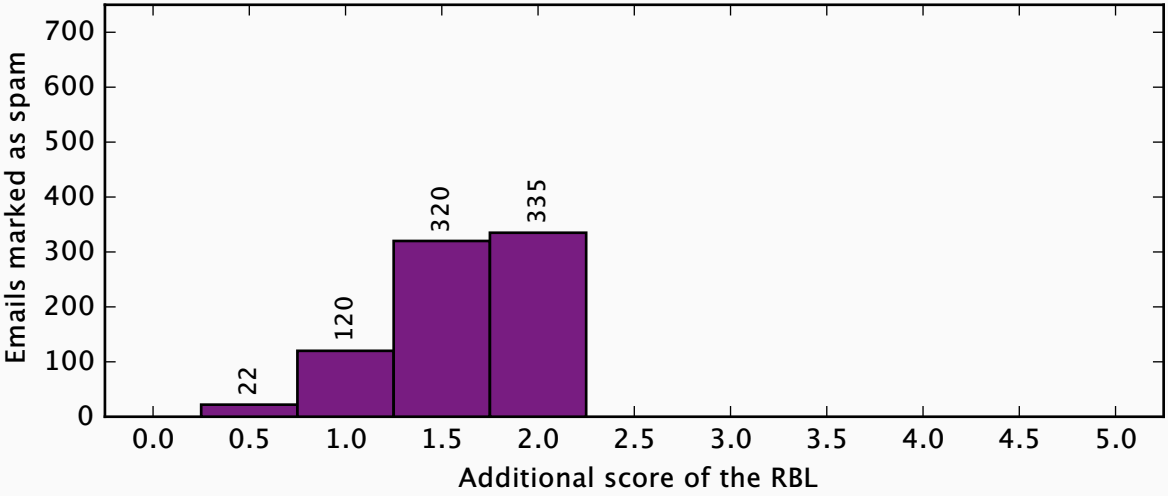


- 633 (58.61%) emails have a score below five
- 52 unique domains in the body
- of which 13 domains have never appeared in an email classified as spam
- these 13 domains appeared in 31 emails (2.87%)

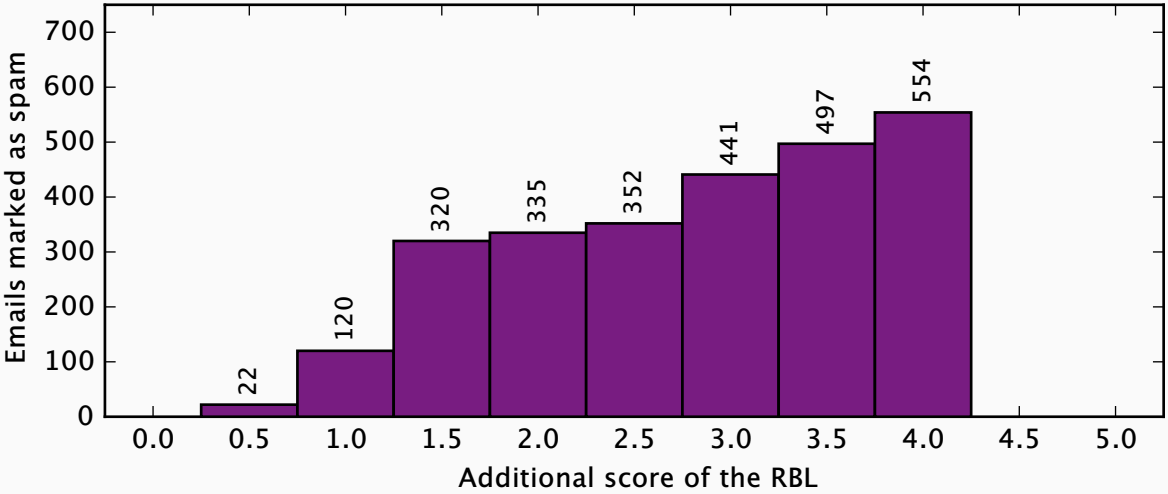
Additional email blocked



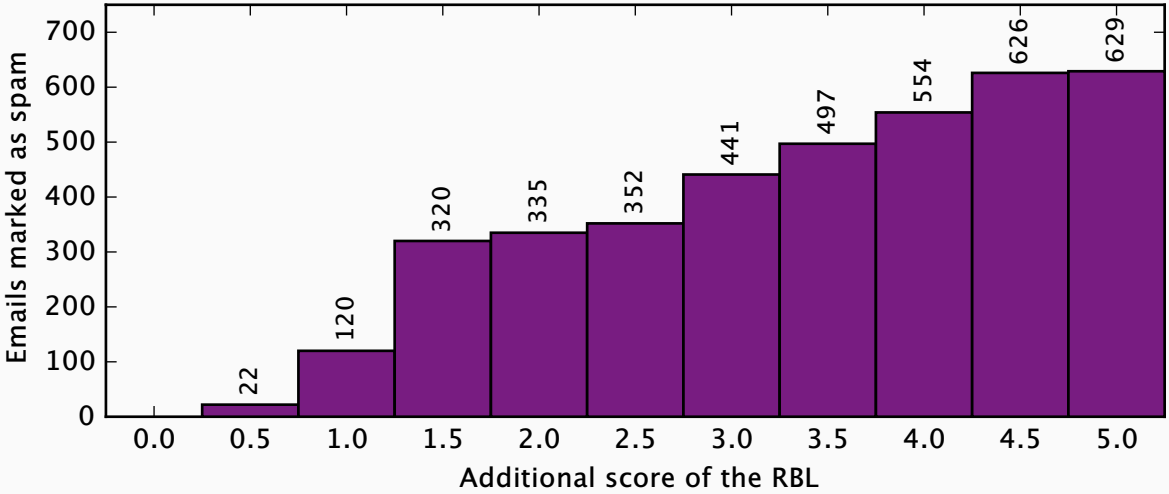
Additional email blocked



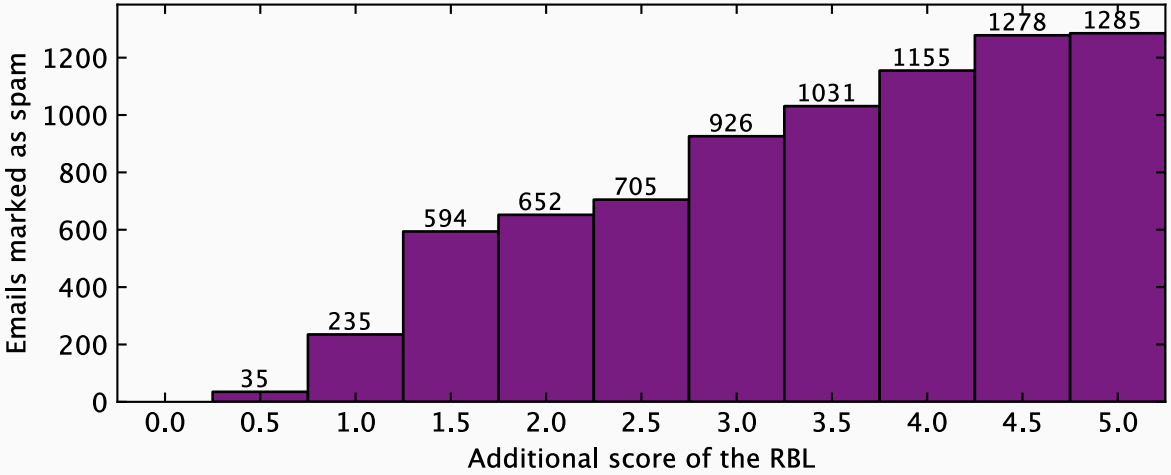
Additional email blocked



Additional email blocked



Additional email blocked (update)



Conclusions

Conclusions

- Hard to detect spam is detectable

Conclusions

- Hard to detect spam is detectable
- Early detection

Conclusions

- Hard to detect spam is detectable
- Early detection
- Additional spam blocked

Questions