Johnny Zhong
CS373 – DADA
Aug 16 2019

Email Security

Security Spaces commonly involve databases.

Phishing Quiz

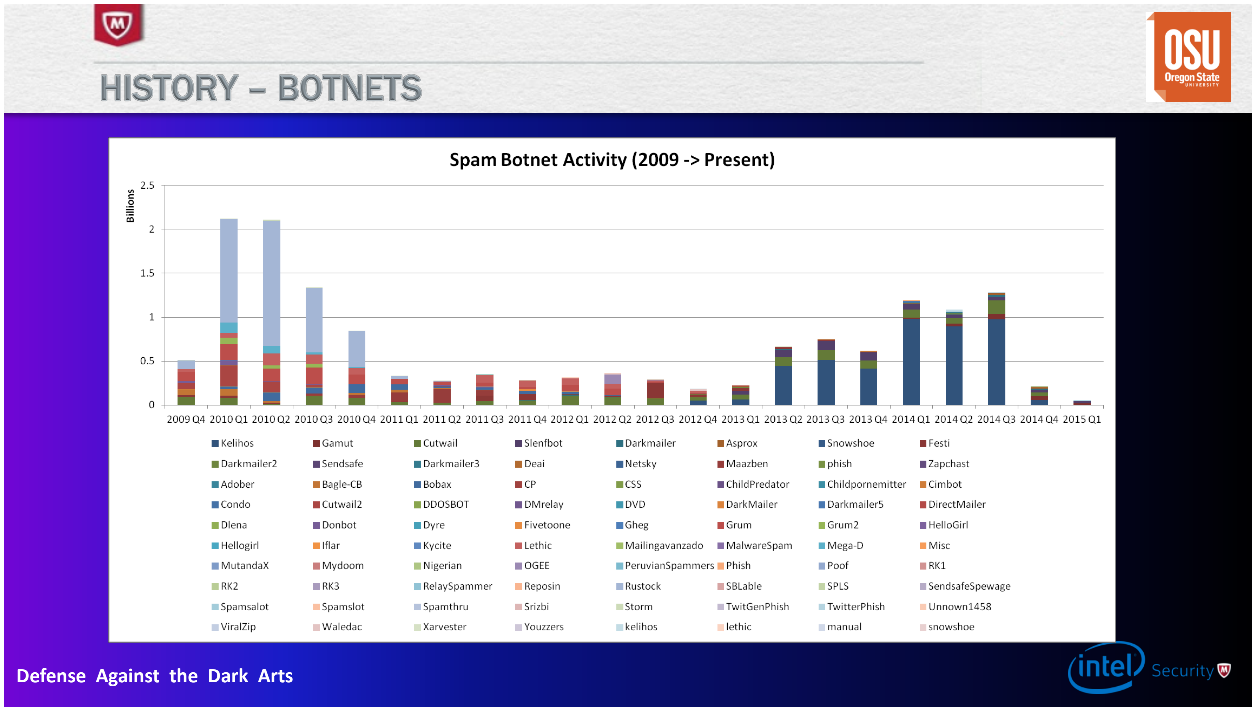
Education – significant portion of security.
The following portion was devoted to running the phishing quiz with the class.
The class identified basic criteria for determination of phishing emails vs legitimate emails.
Some notable criteria:

- Origin email address
- Email address location
- Coupon codes
- Ads
- Request for personal information

Terminology:

Spam/Ham: illegitimate vs legitimate
Spamtrap/Honeypot: unprotected attractive machine which is used for collecting malware (and in this case, spam). An email that should, under normal circumstances, not receive emails. This would be used to unequivocally determine what is spam.
Snowshoe spam: use of many machines to send spam slowly to avoid reputation classifiers.
RBL: real time blackhole list – IP reputation: anything on the list is blocked and considered spam. Vendors have to subscribe to the RBL.
Heuristics: process for determination of spam (or other malware). Commonly using regexes, machine learning, etc.
Bayesian Filtering: finding common features that would indicate the presence of spam using probability. Modified dynamically based on filtering.

Classics
419 phishing – "African Prince" phishing (promise of return on money if initial investment is provided). Preys on people's need for a quick fix/quick method for money.
Canadian Pharmacy – embedded span tags to avoid heuristic filters
Pump n' Dump – artificial inflation of stock by spamming people to buy a stock, then the spammer dumps the stock.

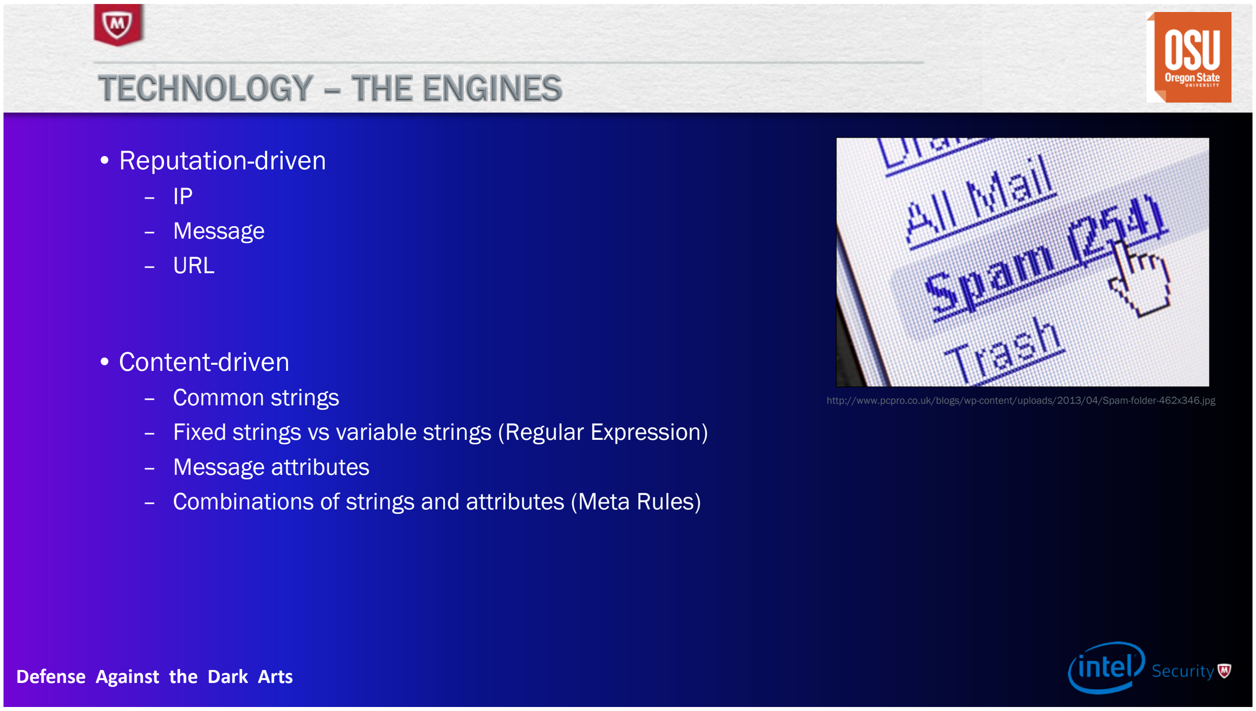


HISTORY – BOTNETS

Defense Against the Dark Arts

Spam Botnet

Rustock – original botnet – making a lot of money at the beginning of emails.
Spam numbers decreased after the initial push of botnets/spam bots



TECHNOLOGY – THE ENGINES

- Reputation-driven
  - IP
  - Message
  - URL
- Content-driven
  - Common strings
  - Fixed strings vs variable strings (Regular Expression)
  - Message attributes
  - Combinations of strings and attributes (Meta Rules)

All Mail
Spam (254)
Trash

http://www.pcpro.co.uk/blogs/wp-content/uploads/2013/04/Spam-folder-462x346.jpg

Defense Against the Dark Arts

Tools

Reputation vs Content Driven
Content Driven explored in the class

Content Driven Tools
DIG aka Domain Information Groper – DNS querying, looks up MX records

WHOIS – IP and Domain Reg
grep/sed/awk – string parsing
databases - postgres
regex coach – string parsing, regex

Regex Techniques were discussed in the lecture to be able to intelligently filter for advertisements for Viagra.

Research Techniques - Considerations

Extraction of Metadata:
Accuracy needs to increase as the number of samples increases.
Value aggregation is necessary to filter through all the samples.
COUNT DISTINCT is commonly used.

Scalability:
Can't have random and can't have individuals writing random regexes.
Need to have an intelligent and strategic method to scale.
Via human training, machine learning?

Classification Methods:
Probability Scoring vs Additive Scoring
Different methods, so methods don't work.

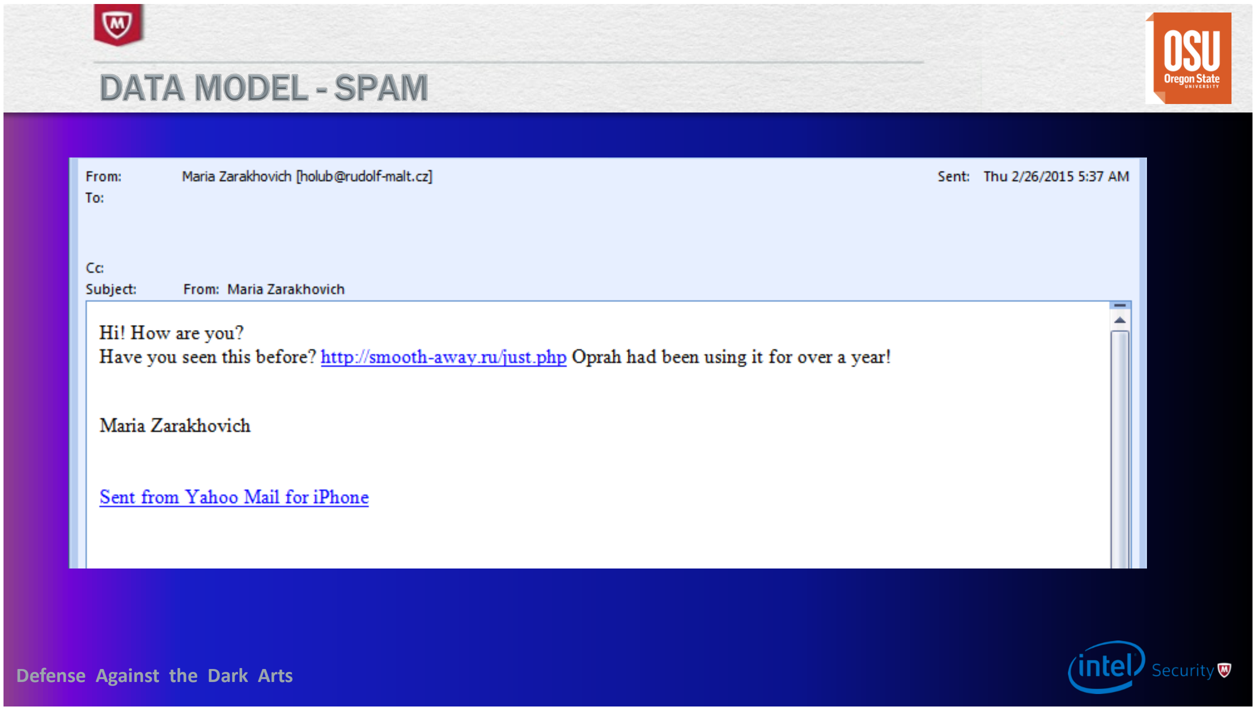
Message Security

Constructing an SMTP Conversation:

1. Port 25
2. Get MX server location from domain
3. Provide recipient address
4. Write in data. Subject line, sender address, body.
5. Send message
6. Quit

SMTP protocol was not designed for security.
Response code 554 – denied – complete rejection

From my observations of the next section, we're trying to build a matrix of features to distinguish spam and ham. This looks like we going to use supervised learning on a collection of spam and ham with the features established in both.



DATA MODEL - SPAM

From: Maria Zarakhovich [holub@rudolf-malt.cz] Sent: Thu 2/26/2015 5:37 AM
To:

Cc:
Subject: From: Maria Zarakhovich

Hi! How are you?
Have you seen this before? http://smooth-away.ru/just.php Oprah had been using it for over a year!

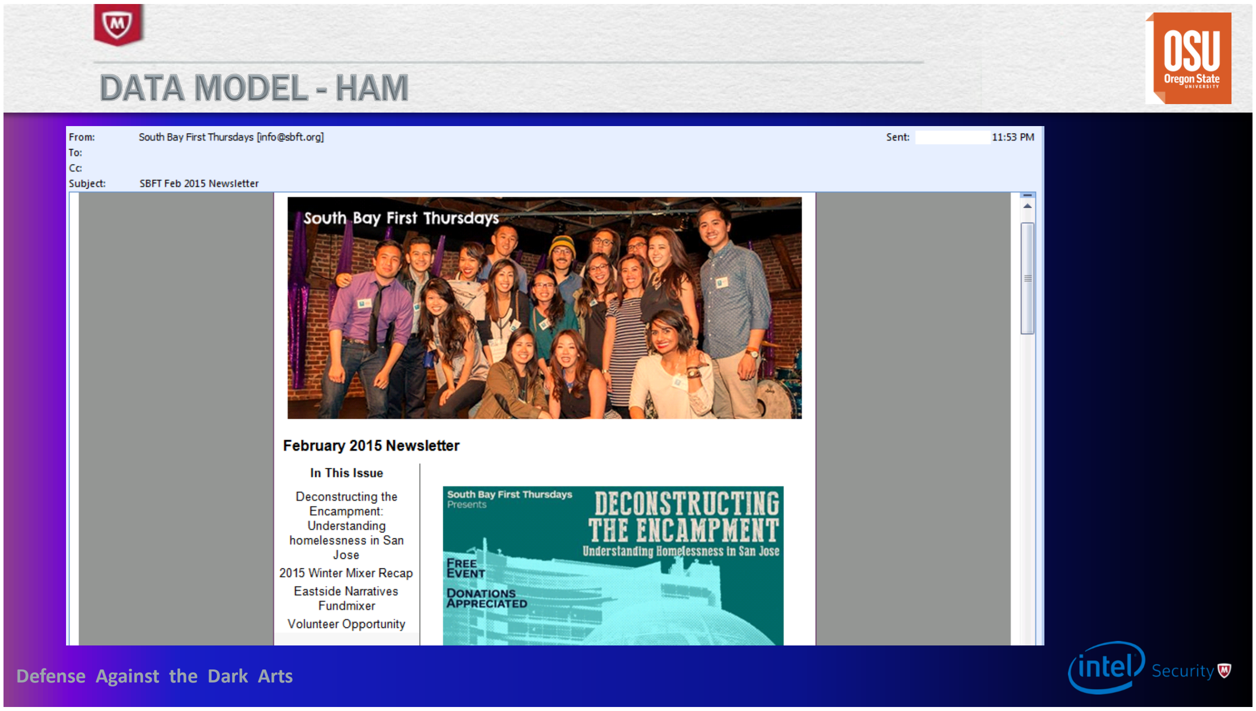
Maria Zarakhovich

Sent from Yahoo Mail for iPhone

Defense Against the Dark Arts

Some Indicators of a Spam Email:

To Subject line empty
Russian domain that points to Personal Home Page
Coming from the .cz
“Yahoo Mail”? No exclamations.
No attachments
Invoke Oprah/Major media figure?
No recipient in the message body
Didn’t use a period in any part of the message body
URL ends in a PHP file



DATA MODEL - HAM

From: South Bay First Thursdays [info@sbft.org]
To:
Cc:
Subject: SBFT Feb 2015 Newsletter
Sent: 11:53 PM

South Bay First Thursdays

February 2015 Newsletter

In This Issue

Deconstructing the Encampment: Understanding homelessness in San Jose
2015 Winter Mixer Recap
Eastside Narratives Fundmixer
Volunteer Opportunity

South Bay First Thursdays
Presents
DECONSTRUCTING THE ENCAMPMENT
Understanding Homelessness in San Jose
FREE EVENT
DONATIONS APPRECIATED

Defense Against the Dark Arts

<u>Some Indicators of a Ham Email:</u>
Heavily formatted
List Mailing Service
Unsubscribe Button
From a .org
No greeting
Message size (length and bytes)

A star plot is created to represent n-dimensional vectors that functions as a representation for spam and ham. Clustering is used to determine likeness of features.