Writeup 8
CS 373
James Tyler Ball

Messaging Security 1 (Eric Paterson)

As with all of these lectures, educating the end user is one of the most effective things you can do to protect your system.

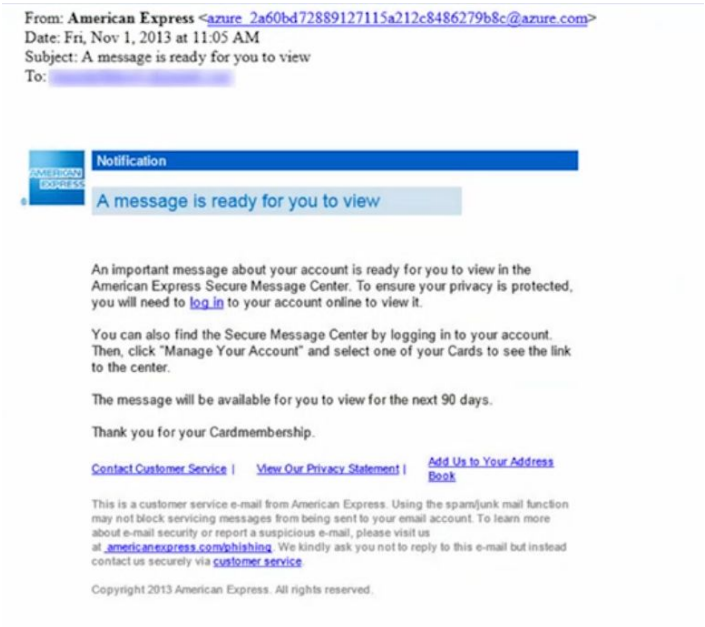
What's the difference between phishing emails and ad emails? Malicious behavior or someone attempting to take information makes it a phishing attempt.

First thing was to look at an email and determine whether or not it's spam. The students noticed that:

- linkedin was an odd source for a coupon
- no logo
- an activation key
- testable code

Eric mentioned that blocking a legitimate email is worse than letting spam go through. A false positive is worse than a false negative. You do not want to impede someone's business.

This was agreed to be phishing:



From: **American Express** <azure_2a60bd72889127115a212c8486279b8c@azure.com>
Date: Fri, Nov 1, 2013 at 11:05 AM
Subject: A message is ready for you to view
To:

Notification

A message is ready for you to view

An important message about your account is ready for you to view in the American Express Secure Message Center. To ensure your privacy is protected, you will need to log in to your account online to view it.

You can also find the Secure Message Center by logging in to your account. Then, click "Manage Your Account" and select one of your Cards to see the link to the center.

The message will be available for you to view for the next 90 days.

Thank you for your Cardmembership.

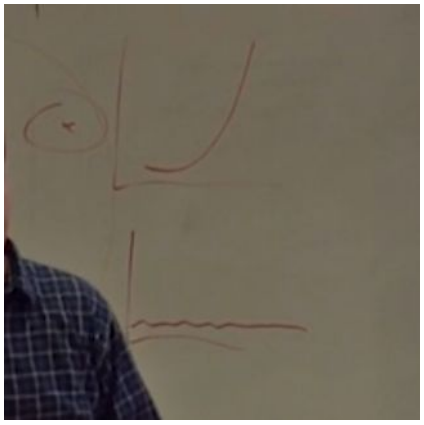
Contact Customer Service | View Our Privacy Statement | Add Us to Your Address Book

This is a customer service e-mail from American Express. Using the spam/junk mail function may not block servicing messages from being sent to your email account. To learn more about e-mail security or report a suspicious e-mail, please visit us at americanexpress.com/phishing. We kindly ask you not to reply to this e-mail but instead contact us securely via customer service.

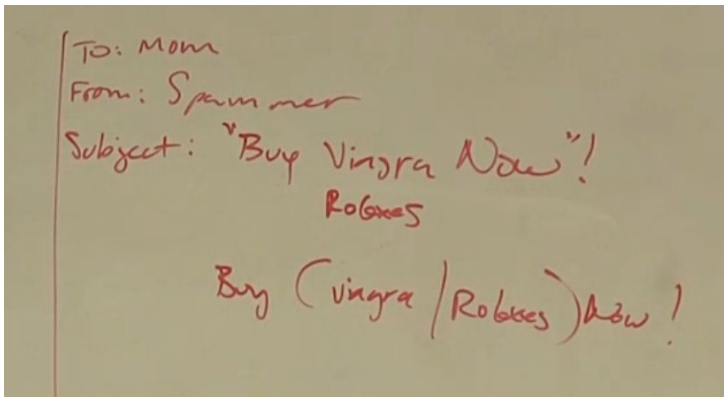
Copyright 2013 American Express. All rights reserved.

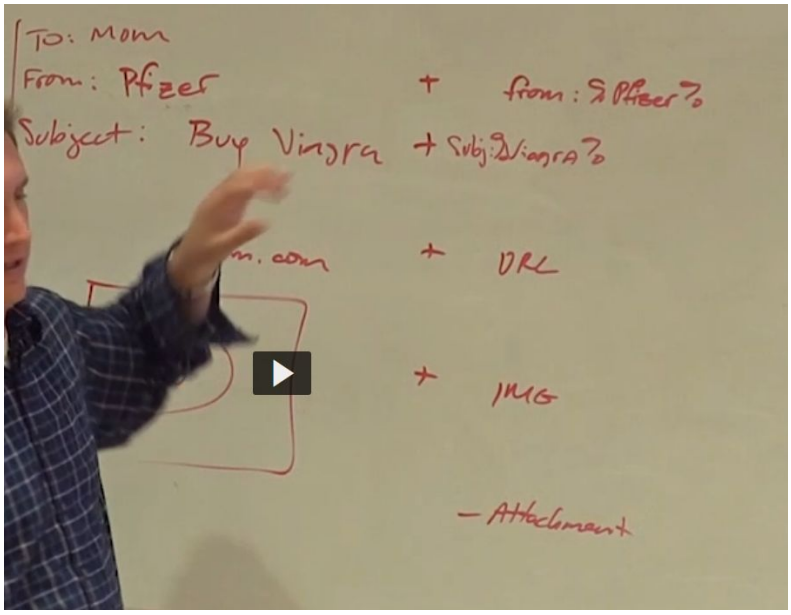
Another was also classified as phishing and the biggest thing that jumped out to me was an odd email account as the sender. Another tool used was hovering over a link and seeing if it actually goes where it claims to. Something to note is that even as a group, the class did not score 100% when trying to classify emails. While there are telling signs, phishing emails can be tricky.

- SPAM - unwanted attempts at private information
- spamtrap/honeypot - unprotected machine used to gather spam emails for analyzation, set up so no legitimate mail should go there
- botnet - series of machines that have been taken over where the end user does not control the end goal
- snowshoe spam - different than phishing, refers to the technique. used to be that a machine would send out a whole lot of spam as soon as it was able. snowshoe refers to spreading out the frequency of phishing attempts to avoid a spike

  
  above - w/o snowshoe, below - with snowshoe

  this is one of the hardest things for to defend against for email vendors
- phishing vs spear phishing - data breaches are huge news at time of lecture (still the case now). abusing email is very profitable and one effective method is to insert yourself into a thread and continue it
- RBL - very effective source of information, but requires buy in by many users
- Heuristics - methodology for analyzing the aspects of the in aggregate instead of looking at simple strings

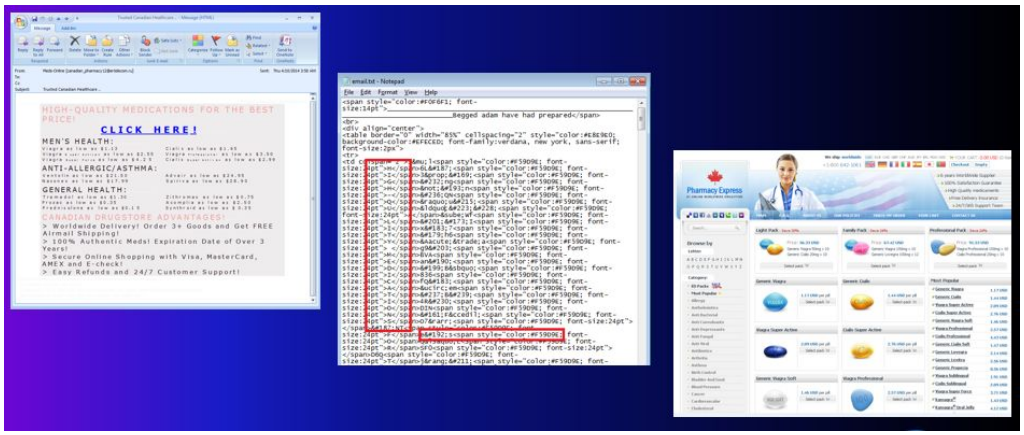
  



- bayesian - for a given set of samples, take a corpus of known HAMs and SPAMS. Gather tokens and observe/weight trends

419 is the code referring to “I have a dead relative that will give you money if you give me this information”

Attackers will try to bypass heuristic filters by embedding the text within other tags:



Spammers will send out emails to try and inflate stock values: “pump and dump”.

Types of spam, much like malware in week 1, cycle in popularity. There are spiking instances of known spammers trying to infect machines and sending out spam, and not necessarily at the same time. Spammers obviously don’t want to be detected.

TOOLS

- DIG is for investigation of DNS records
- WHOIS is more focused on IP and Domain registration info
- Grep, SED, AWK are great for analyzing data
- PostgreSQL is Eric's preference for databases

"write a regular expression that matches all variations"

v|agra, \/iagra, v|4agra, but not Viagra

I had trouble with this, but some classmates claim that

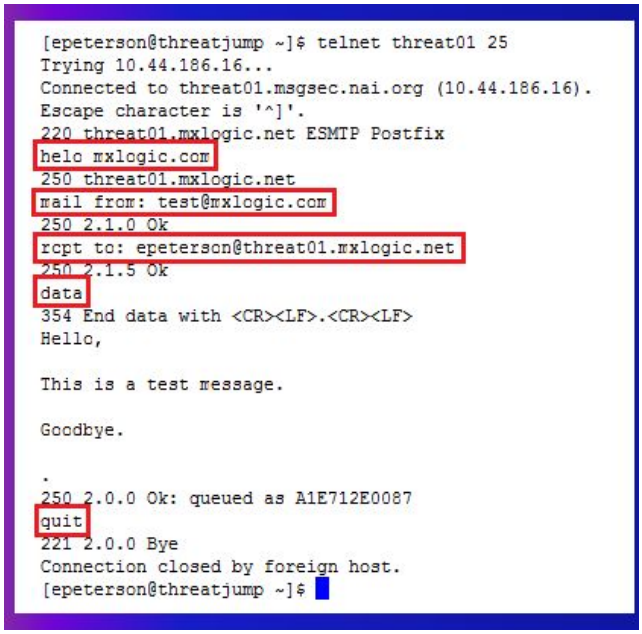
(\\\/i|V\||v[ ]{0,1}\|[ ]{0,1})(a[ ]{0,1}|4)(g[ ]{0,1}|G)r[ ]{0,1}(a|A) and
(?!viagra)((v|\\)\s?(i|\/|)\s?(\||4|)\s?4?\s?g?\s?r?\s?a?) work.

Good question was asked: Do large companies that sort through millions of pieces of data use Regex or something else? Answer was yes, Regex is used on millions of entries.

Eric said peak 2014through hosted platform was 700 million messages, 1 billion caught at the firewall. We established early on that classification is hard, but even 99% detection isn't good enough.

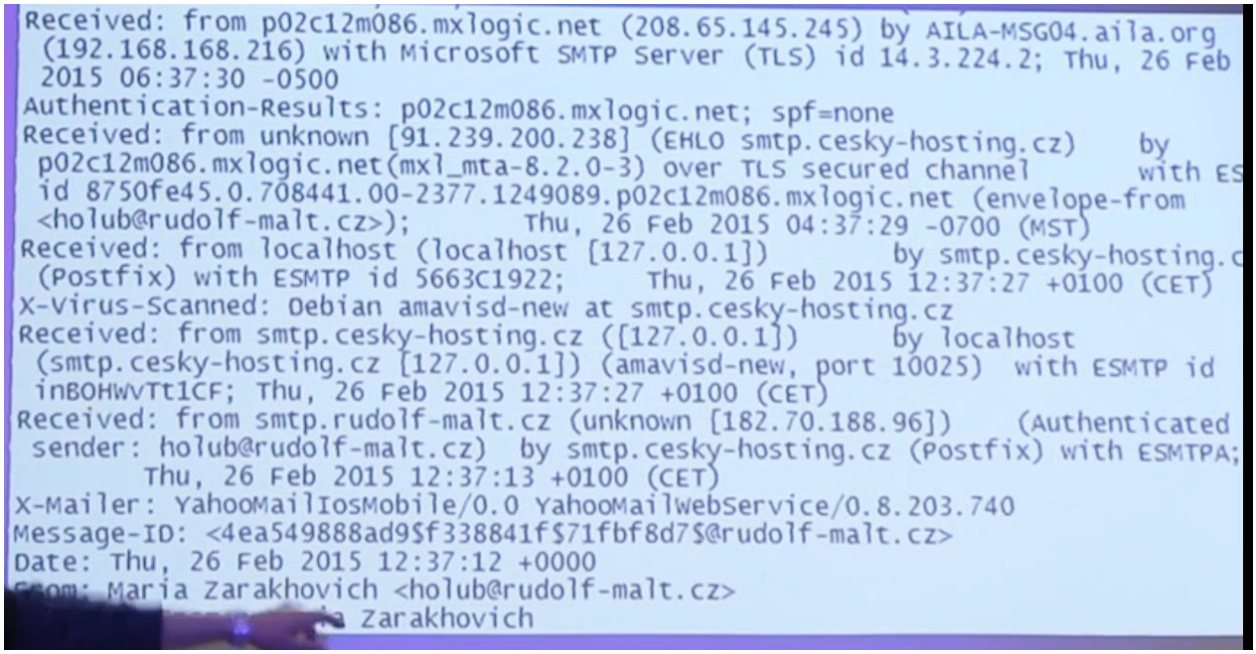
- parsing - extract key data, common elements. how do get something meaningful out of 100000 elements?
- aggregate - highlight pieces of data over time. how to block SPAM and allow HAM?
- outliers - why in certain instances are some things similar to SPAM not classified as SPAM?

Messaging Security 2 (Eric Paterson)



First thing discussed was how easy it was to spoof an email address. The fundamental problem with SMTP is that it wasn't designed to be secure.

Read the following email header bottom to top.



I liked this next exercise. Observe the following SPAM email

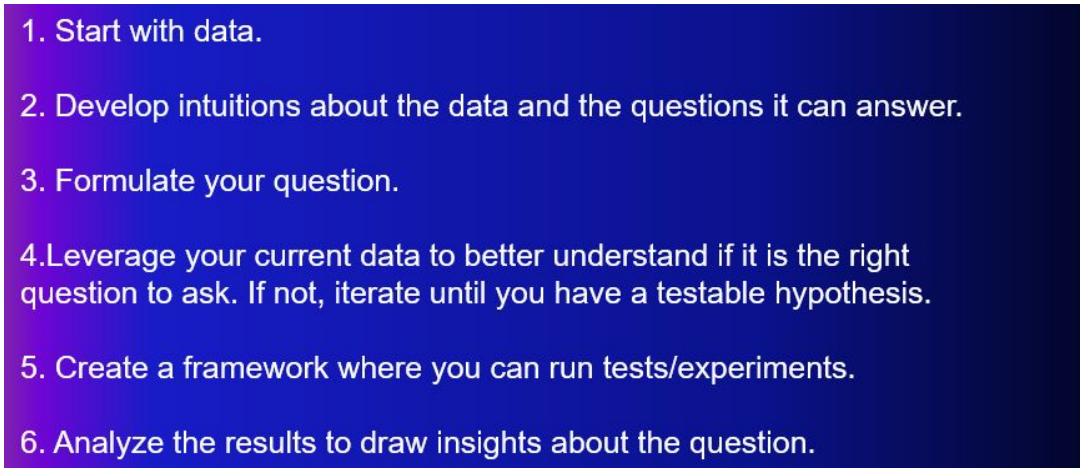
- to line is empty
- russian domain (php page)
- claims to be from yahoo mail (no exclamation)
- .cz
- no attachment
- invoke Oprah
- nothing about the product listed
- no recipient listed in the body (but this is actually common)
- odd punctuation style
- weird subject line
- from a mobile device
- sent at 5:37am
- very short email

With simple observations, we can identify several suspect aspects of this email.

Next we're looking at a HAM email:

- embedded images/URLs
- heavily formatted
- large email
- no visible URLs
- unsubscribe button
- .org
- subject line is relevant
- visible people
- no personalized greeting

Following the comparison between HAM and SPAM examples, Eric brought up how the scientific method factors into the approach:



There are millions of SPAM messages every day, and the better someone is equipped to analyze massive amounts of data, the more effective the analysis will be. We've used basic clues to look at examples, and a similar approach can help an analyst glean what sort of info is in a large sample of data and what questions should be asked. Classification is an iterative process. Databases are a great place to analyze the data.

Next Eric introduced us to the classification lab. A bit of a time-skip later and some students presented their results. The stock pick spam was extremely common in the data set used.

Eric reminds us that SPAM isn't just in email, and learning to recognize it is important to keep our personal information safe.