OREGON STATE UNIVERSITY

CS 373

WINTER 2019

Week 8

Author:
Thomas Noelcke

Instructor:
D. Kevin McGrath

I. Overview

This week we are diving deeper into message security. This also includes Email security. This week we will be talking about various aspects of security in regards to email security. In this write up will talk about topics that I found interesting during the lecture. I will also list out different terms that I learned through out the lectures.

II. Interesting Lecture Topics

A. tools

We will be using databases in this section of the course. When we talk about security we will be using a database. The database it self might differ but in this class we will mostly be using postgres. We will also be using regex expressions. To do this we will use regex coach. This tool will let us play with regular expressions and play around with the regular expressions and get real time results.

B. 419 Phishing

This is type of spam where some one claims to have a lot of money that they would like to give you but you need to send money and your contact info in order to get the million dollars they promise you. As it turns out this is not true. They are not sending you money and are taking your money or contact information. This is so successful and command that it has its own criminal code.

C. Tricking Filters

Another popular way around the Heuristics filters to place every thing in html elements. An example is that you can place each letter in span tags. This gets around the filters because the filtering system was not designed to look for the words in span tags. Often these emails won't look all that great but the web page that it links you to may look acceptable.

D. Pump and Dump

This is an attack that uses spam to push up the value of stock and then dump the stock. The idea is that the spammers send out emails claiming to have an insider tip that every one should buy this stock now. Once they have convinced some people to buy the stock the stock price will go up. Then the spammers dump the stock and the people who believed them are stuck with a bad investment. Moral of the story, Don't take investment advice over email from some one you don't know.

E. Spam Bot Nets

We started looking into spam data in 2009. Initially, There were lots of Canadian pharmacy spam that was very successful. The purps made over 100g per week push fake Viagra. However, After they were shut down because they were very visible, there was a massive serge in virus based spam. This was so that they could rebuild their bot net and come back. We saw a serge some number of years later that reflects this.

F. Defense against spam

Below is a list of ways that we can attempt to defend against spam:

- Reputation Driven: IP, Message, URL
- Common Strings
- Fixed strings vs variable stings (regular expression)
- message attributes
- combos of strings and attributes (meta rules)

On top of these methods there are also a number of tools that we might use to investigate and prevent spam. A list of these tools is shown below.

Linux:

- Dig - Investigation of DNS records
- WHOIS - Finds information about an IP or Domain registration
- Grep, SED, AWK - data parsing and manipulation

Opensource Databases

- PostgreSQL - Best open source database
- MySQL - Most popular open source database

Other Tools:

- Regex coach - regular expression tool
- Trstedsource.org - current reputations according to McAfee.
- Spamhause.org - Authoritative source of reputation data.

G. Research Techniques

This week we also talked about many different research techniques for looking into different spam. We can research spam by looking at the actual samples or sometime just by looking at the metadata. One common technique is parsing. This is done by collecting key metadata from a sample. This can include source ip, subject, and who it is from. Grouping is another common method by which research is done. Items can be grouped by time stamp source ip, subject, url and many other common values. We can also use an aggregate to look into different samples. This technique takes advantage of counting different occurrences of different words and using this to correlate the likely hood that it is spam vs ham.

H. SMTP conversation

SMTP is often done using telnet over port 25. This can be done over different ports but often its port 25. Using telnet we can send a message to an IP address. First we have a hand shake and set up the connection. Once the hand shake has been completed we send the data. Once we are done sending the data we will terminate the connection and the message will either be sent or will be queued. It's also important to note that the mail from is just data that can be manipulated. Just because the email is in the from section doesn't mean that email is actually who they say they are. This is one of the key problems with SMTP.

I. Reading Email Headers

When reading email headers you should start from the bottom up. The pertinent information that you are looking for will be best to read from the bottom up. It will start with a subject and sender. It will then include the actual sender information and a time stamp that shows who actually sent it. This header will also show all of the hops that the email took to get to you. This information can be really useful.

III. Vocabulary

- Spam - This is email that is unwanted email that is not legitimate.

- Ham - This is legitimate email that is from the sources that it claims to be from.

- Spamtrap/Honeypot - An unprotected computer that has no protection on it designed to be attacked to collect samples. Some times this is done with a new domain but some times this is done using a retired email address.

- Botnet - This is a network of connected devices that are infected or running a bot of some kind. These can be used to steal data or to preform DOS attacks.

- Snowshoe spam - This is a technique of spamming that spreads out the load of sending spam to other machines.

- Phishing - Using an email that looks legitimate but actually contains malicious content.

- Spear Phishing - Targeted phishing attack towards a specific person in an organization. This is often an executive or some one of power.

- RBL - Reputation Block List.

- Heuristics - This a method of defining rules that detect suspicious emails or traffic.

- Bayesian (Statistical) - This is another strategy for looking at spam and ham and use tokens to identify mail that is malicious. It then weights the tokens and uses these weights to determine if the message is spam or ham.

- Fingerprinting/Hashing - This is the process of taking a hash of know malicious emails and using that hash to detect future spam.