Week 8 Notes

This week’s focus is on email security

Regex Coach – window’s tool and let’s you put in the pattern that your trying to match against

\*\*\* FIND OUT MORE ABOUT THIS \*\*\*\*

Find out more about YARA rules \*\*\*\*\*\*

w/ a pile of data, how do I get something meaningful out of it.

You don’t want to have false positives because it could hinder legitimate business and the business line

Spear phish vs phishing

Spear phishing is a very targeted attack where the attackers know exactly who they want to affect and compromise vs phishing is just trying to find anyone who will take the bait

80% is the standard for HR person in the industry – the most phished role and HR is the most vulnerable and the most susceptible

Most successful from a phish is the HR team

Terminology

Spamtrap/Honeypot – is a domain or an address that receives only spam messages. Its setup so that no legitmate email will ever go there.

What the big IT companies and ISP providers are doing is one a person deletes an email account, they are keeping it alive for additional 6 months and responding w/ a 500 code, and if after those 6 months if those authored emails continue to send then the ISP’s and big IT companies start to flag it as spam

\*\*\*\*\* THIS IS INTERESTING, MAY WANT TO FIND OUT HOW BIG IT COMPANIES ARE COMBATING SPAM W/ TECHNIQUES LIKE THIS ONE \*\*\*\*\*\*\*\*\*

They retire an address and turn it into a spam trapper

Botnet – series of machines that have been taken over that the end user doesn’t support the end goal (could be spamming, bitcoin mining, etc)

Most common botnet is a spam botnet.

One example of a botnet is Zeus.

Snowshoe spam – technique of spamming. Distributes the weight is what a snowshoe does. So the actor is snowshoe spamming the loads across various computers to avoid putting machines on anti-spamming machine radars.

* To get around IP address reputation filters/machines

RBL = real time blackhole list – it’s an IP reputation, if your on the list then you are known as a spammer. No middle ground. RBL’s are owned by various commercial companies.

* It’s very binary

PBL – policy block list

Heuristics – finding spam through rules such as strings matches (“Buy Viagra Now!”) and search by this string and anything w/ this string will be flagged as spam.

* Spam rules
* There are also meta rules – look at features of email via an aggregate and if the email meet’s all these criteria’s it’s then flagged as spam

Another way is through Bayesian (statistical) analysis – through token (mostly through strings) found in emails and the more string tokens you find across multiple emails the more likely that it’s probably the email is spam

419 Phishing

Nigerian scam – African prince scams, etc.

Pump N Dump Spams

Spam that says they have insider information on a particular stock to drive the stock’s value up and then dump it at it’s peak

\*\*\*\*\*FIND OUT MORE ABOUT BOTNET’s for SPAMMING \*\*\*\*\*\*

The 2 best tools for detecting spam are:

1. Reputation driven
2. Content driven

Reputation Driven

* IP
* Message
* URL

Content-driven

* Common strings
* Fixed strings vs variable strings (regular expressions)
* Message attributes
* Combinations of strings and attributes (meta rules)

Tools for Research Purposes

* Linux tools
  + DIG – Domain information groper – investigation of DNS records
  + WHOIS – searching for IP/Domain registration information
  + Grep, SED, AWK – data parsing and manipulation
* Open source databases
  + PostgreSQL – “ the worlds most advanced open source database”
  + MySQL – the worlds most popular open source database
* The Regex Coach
  + Regular Expresssion syntax/functionality learning aid
* Trustedsource.org
  + Historical and current reputations based on McAfee data
* Spamhaus.org
  + Widely accepted as an authoritative source of reputation data (especially in NA)

MX record is a DNS record for a mail exchange server

Research Techniques – Samples or Metadata

* Parsing
  + Extraction of key meta data (Source\_ip, Subject, Mail\_From, HELO, etc.)
* Grouping
  + By timestamp, source\_ip or cidr, subject, URL, other common elements
* Aggregation values over time
  + Values/time
    - Count distinct \*
    - Etc.
  + Exposure of relevant research paths
* Identification of outliers
  + When A, which is highly similar to B, is classified as spam but B is not
    - Why?
    - Why not?

Research techniques – considerations

* Huma input required?
  + How much over time?
  + What quality assurances can be expected?
  + Value of human input vs automation?
* Fully automated?
  + Initial human resource cost?
  + Fault tolerance/resiliency?
* Combination of automation and human input?
* Probability scoring vs additive scoring

\*\*\*\* LOOK AT SCORING MODELS FOR DETECTING SPAM \*\*\*\*\* PROBABILITY SCORE / WEIGHTS ETC.

* Additive scoring vs other scoring models

Day 2 Lecture

Email header is usually read from the bottom up

Authentication or security protocols on top of SMTP

* How you would change/spoof the sender email in an email

***Find out more about SPF, DKIM, DMARC authentication methods***