INTRODUCTION TO DATA SCIENCE IN INFO-SEC

@HenriettaTante

CATNAT

OVERVIEW

- What is information security and why should we care
- The classical hacker narrative and it’s reality
- What is Feminist Hacking?
- Data Science in info-sec
- Botnet example with jupyter

WHAT IS INFORMATION SECURITY ? – KEYWORDS

Cybersecurity

Malware

Ransom Ware

Cryptography

Hacking

Pentesting

privacy rights

Spying

Secret Services

Script Kiddies

privacy enhancing technology

Virus

Antivirus Software

Attack Vector

Phishing Emails

Info-sec “as a service”

Industrial Control System Hacking

Internet of Things

Security

WHAT IS INFORMATION SECURITY ? – A DEFINITION

Information Security:
The means and processes of keeping entities save from unwanted access.

“Means and processes”:
Firewalls, Encryption, good changing passwords, Security Operation Centers, …

“Entities”:
states, companies, critical infrastructure, and we, the public.

“Unwanted Access ”:
Hacking, collection of personal data, physical access (breaking in)

WHAT IS INFORMATION SECURITY ? – A LITTLE ADD-ON

BUT

Information Security can also mean active hacking:

- Secret services spying on each other or actively destroying infrastructure (eg. Stuxnet)
- Pentesting: When you are paid to break into a entity by electronical or physical means.

WHAT IS INFORMATION SECURITY ? – CHOOSE YOUR CAREER

Three main paths:

- Hacker: yes, hacks stuff for good or evil (pentester, malware architect, intruder, …).
- Info-sec analyst: detects when hackers try to intrude.
- Privacy enhancement expert/activist: makes stuff more secure and harder to be hacked, accessed.

Naturally, those paths are overlap a great deal.

HOW TO BECOME A HACKER/ INFO-SEC PERSON (ACCORDING TO THE HACKER MAINSTREAM)

1. The mindset

A hack is "a clever, benign, and 'ethical' prank or practical joke, which is both challenging for the perpetrators and amusing to the MIT community."
(MIT Hackers, the grandparents of hackers, http://hacks.mit.edu/Hacks/)

HOW TO BECOME A HACKER/ INFO-SEC PERSON (ACCORDING TO THE HACKER MAINSTREAM)

1. The mindset

Hackers are like artists, philosophers, and engineers all rolled up into one. They believe in freedom and mutual responsibility. The world is full of fascinating problems waiting to be solved. Hackers take a special delight in solving problems, sharpening their skills, and exercising their intelligence.
(wikihow https://www.wikihow.com/Become-a-Hacker)

HOW TO BECOME A HACKER/ INFO-SEC PERSON (ACCORDING TO THE HACKER MAINSTREAM)

1. The mindset, according to "THE" hacker manifesto

I am a hacker, enter my world...
Mine is a world that begins with school...
I'm smarter than most of the other kids, this crap they teach us bores me...

Damn underachiever. They're all alike.

Yes, I am a criminal. My crime is that of curiosity. My crime is that of judging people by what they say and think, not what they look like. My crime is that of outsmarting you, something that you will never forgive me for.

I am a hacker, and this is my manifesto. You may stop this individual, but you can't stop us all... after all, we're all alike. http://www.mithral.com/~beberg/manifesto.html

HOW TO BECOME A HACKER/ INFO-SEC PERSON (ACCORDING TO THE HACKER MAINSTREAM)

1. The skillset

Learn to Program in C
Learn to Program in python
Learn Ruby
Learn HTML
Learn JavaScript
Learn UNIX
Learn Networking
Concepts
Learn Cryptography
Participate In Hacking Challenges
Go Next Level: Write Vulnerability
Build a computer
Learn security concepts
Wireshark
Virtualisation
Kali Linux
Database skills
Learn Reverse Engineering

Well, maybe just this (imao) ....

Learn to Program

Learn Networking Concepts

Learn UNIX

Study common attack types

LEARN PSYCHOLOGY

F.E.M.M.I.N.I.S.T HACKING

Key Question: What can feminist hacking look like?

- Hacking mindset has this smell of practical jokes, genius, bullied nerd and underdog hero –> What can a feminist hacking narrative look like?
- What would you personally like to hack/secure/analyse?
- How to get into hacking the feminist way?

HOW TO GET INTO HACKING, THE FEMINIST WAY.

1. The mindset

- What would you personally like to hack/secure/analyse?

My personal list:
* Find out about those who threaten online.
* Understand how insecure critical infrastructure is (like power plants)
* Find out how to secure/enhance democracy through this knowledge
* For the puzzle fun of it: How does malware work, and how can it be detected.

What’s your list?

HOW TO GET INTO HACKING, THE FEMINIST WAY.

2. The skillset (my choice)

Well, to start just this (imao) ....

Learn to Program

Learn Networking Concepts

Learn UNIX

Study common attack types

Learn Psychology

LEARN DATA SCIENCE

INFO-SECURITY DATA

- Internet /Network traffic logs (network intrusion detection)
- Malware binaries (partial hashing, pattern recognition)
- Correlating publicly available vulnerability data.
- Gather and correlate data for social engineering, defence settings
- Adversary machine learning (malware, tricks info-sec ML)

ON THE NATURE OF THE DATA:
HOW DOES INFORMATION TRAVEL ON THE INTERNET

- The internet is a global network of computers.
- Every computer connected to the internet has a unique address : the IP address.
- Doing anything on the internet (visiting a website, writing an email,...) involves exchanging information/messages between computers.
- Every piece of information is disassembled and travels in small packages trough the network. This makes the internet super efficient.
- There are communication protocols to ensure smooth re-assembly and synchronisation.

LEARN NETWORKING CONCEPTS:

PLAY WAR GAMES

Overthewire: http://overthewire.org/wargames/ :
a bunch of “games” where you log onto the overthewire server and have to retrieve the password of the following level to get ahead .

You learn everything about protocols, encryption, unix you need!

ON THE NATURE OF THE DATA:
HOW DOES INFORMATION TRAVEL ON THE INTERNET

Internet property	features	example feature names in the dataset
Ip addresses/ports	Ipv4, Ipv6, Source/ Destination Ip/Ports.	Dst port
Information travels in small packages.	Duration, number of packets, length of packets	Fwd Pkt Len, Bkw_ Pkt Len,
Communication Protocols	TCP/UDP, Syn-Ack-Fin Counts	Protocol, FIN Flag Cnt, SYN Flag Cnt

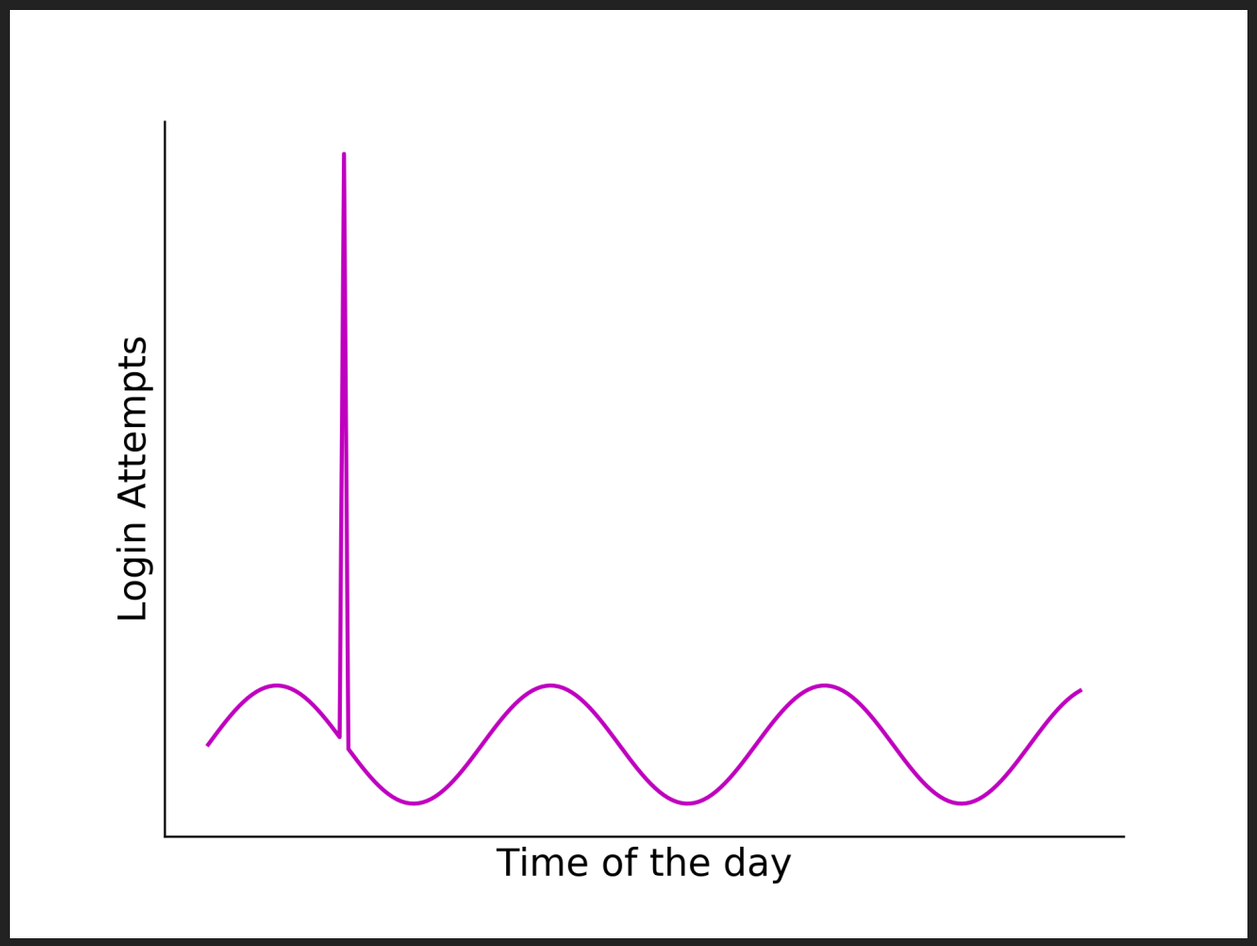
WHAT ARE WE EVEN LOOKING FOR?

Short answer: Anomalies

MOST BASIC ATTACK TYPES

- Brute Force Passwords (guesspwd)
- DOS/DDOS Attacks ((distributed)-Denial of service)
- Bot-nets

THE NIDS FINGERPRINT OF DOS/DDOS ATTACKS // BRUTE FORCE PASSWORDS



Brute force attack:
Somebody is trying to brute force an access.

Fingerprint:
The usual seasonality of daily traffic is broken.

THE NIDS FINGERPRINT OF BOT-NETS

A botnet is a collection of compromised devices. Those devices are controlled trough a ‘CnC’ server and conduct malicious tasks such as spam mails and denial of service attacks.

Fingerprint:

- Often use high ports.
- Localized up/download packet size.up/download happens at the same time every day.
- Tends to beakon out.
- Network jitter (interpacket arrival time) is high because normal desktop computers are used as servers.

OK! LET'S LOOK AT THE DATA!

https://www.unb.ca/cic/datasets/ids-2018.html

pip install awscli

aws s3 sync --no-sign-request --region eu-central-1 “s3://cse-cic-ids2018/“ dest-dir

DATA STRUCTURE OF TRAFFIC LOGS

- Mixture of categorical, binary, ordinal, nominal data.
- The nominal part is often count data.
- Time dependent/correlated data.
- Even the nominal data is usually not normally/malahanobis distributed.
- Communication between hosts can be displayed as graph

ANALYSIS METHODS FOR INFO-SEC DATA

- Anomaly detection/Clustering
- Time series and sequence analysis.
- Seasonality detection
- Embedding of categorical data (e.g. one-hot-encoding)
- Binary classification with balancing schemes.
- Network analysis

LAST WORDS #1

- Info-sec data science is still in its infancy
- High quality labelled training data is really hard to get.
- Info-sec data science is vulnerable to adversary ML attacks.
- The info-sec data is in general meaningful and multifaceted.
- The high false positive rate is a huge problem for businesses.
- The fact that it has become a huge business does not mean the discipline is moving fast.

LAST WORDS #2

- Everybody can do hacking/IT security
- Find you motivation and get hacking.
- Or not :)

THE END.

THANKS!

NETWORKING CONCEPTS: LAYERS, PROTOCOLS, LOGFILES

Jude Milhon, aka St. Jude

Back in the 70s we connected bay area books stores and libraries. We called it the Community memory database, a kind of proto-internet.
Also, i care about privacy enhancing technologies.

SECURITY ENHANCING TECHNOLOGY

Katherine Johnson

I calculated flight trajectories for NASA aircraft, keeping astronauts save.

COMMON ATTACK TYPES AND HOW THEY ARE DETECTED

Maria Mitchell

my computation of the motion of Venus will be soon on github. Also check out out my blog on voting rights for women.

NETWORKING

- How does the internet work (aka how is information sent and received)
  In short:
  DNS
  Network Layers,
  TCP/IP Stack,
  TLS (Transport Layer Security)
  Symmetric and asymmetric cryptography.

- Awesome starting Videos:
  What is the internet: https://www.youtube.com/watch?v=Dxcc6ycZ73M
  Packets, Routing and Reliability: https://www.youtube.com/watch?v=AYdF7b3nMto
  IP Adresses and DNS : https://www.youtube.com/watch?v=5o8CwafCxnU

DETECTION & ANALYSIS

- Static Analysis: analyising malware without execution, most common: hashing (antivirus software, virustotal https://www.virustotal.com/#/home/upload
- Dynamic Analysis and Reverse Engineering: execution in controlled environment (sandbox systems)
- Network Analysis (NIDS): analysing how network reacts, data: log files

  (this is how a raw tcpdump looks like: https://en.wikipedia.org/wiki/Tcpdump#/media/File:Tcpdump.png, and you can analyse it with wireshark https://www.wireshark.org/download.html)

NETWORK ANALYSIS

- Intrusion detection == Anomaly Detection
- For starters: find sample set for example
  https://github.com/FransHBotes/UNSW_NB15
- Use python/scikit learn to perform anomaly detection/outlier detection/
  /isolation forest/…/
  https://scikit-learn.org/stable/modules/outlier_detection.html

WHAT'S NEXT?

- Discussion: What is feminist hacking?
- What would you like to do hands on?

[ ] Hacking
[ ] Learn about security enhancing tech
[ ] Intrusion Detection and Analysis

SECURITY ENHANCING TECHNOLOGY

- Secure Messengers: How does the signal messenger work?
  (https://www.signal.org/docs/)
- Secure Android: Should you go for lineage OS? https://
  www.lineageos.org/
- Secure Python: How can we include security concepts into python?
  >>>

SECURE PYTHON

- Secure.py - secure.py is a lightweight package that adds optional security headers and cookie attributes for Python web frameworks. (https://github.com/TypeError/secure.py)
- Flask-HTTPAuth - Simple extension that provides Basic, Digest and Token HTTP authentication for Flask routes (https://github.com/miguelgrinberg/flask-httpauth/)
- Flask Talisman - Talisman is a small Flask extension that handles setting HTTP headers that can help protect against a few common web application security issues. (https://github.com/GoogleCloudPlatform/flask-talisman)
- Django Session CSRF - CSRF protection for Django without cookies. (https://github.com/mozilla/django-session-csrf)

again thanks to https://github.com/guardrailsio/awesome-python-security

HACKING: OTHER PLAYGROUNDS

- Let's be bad Guys - Shiny, Let's Be Bad Guys: Exploiting and Mitigating the Top 10 Web App Vulnerabilities (https://github.com/mpirnat/lets-be-bad-guys)
- django.nV - django.nV is a purposefully vulnerable Django application provided by nVisium. (https://github.com/nVisium/django.nV)
- DSVW - Damn Small Vulnerable Web (DSVW) is a deliberately vulnerable web application written in under 100 lines of code, created for educational purposes. (https://github.com/stamparm/DSVW)
- DVPWA - Damn Vulnerable Python Web Application was inspired by famous dvwa project and bobby-tables xkcd comics (https://github.com/stamparm/DSVW)

thanks to https://github.com/guardrailsio/awesome-python-security