LI OPSEC, Targeting Analysis & Countermeasures

Targeted User Analytics,
Defensive Persona Development
and OSINT

whoami

Richard Wickersham @RichWickersham

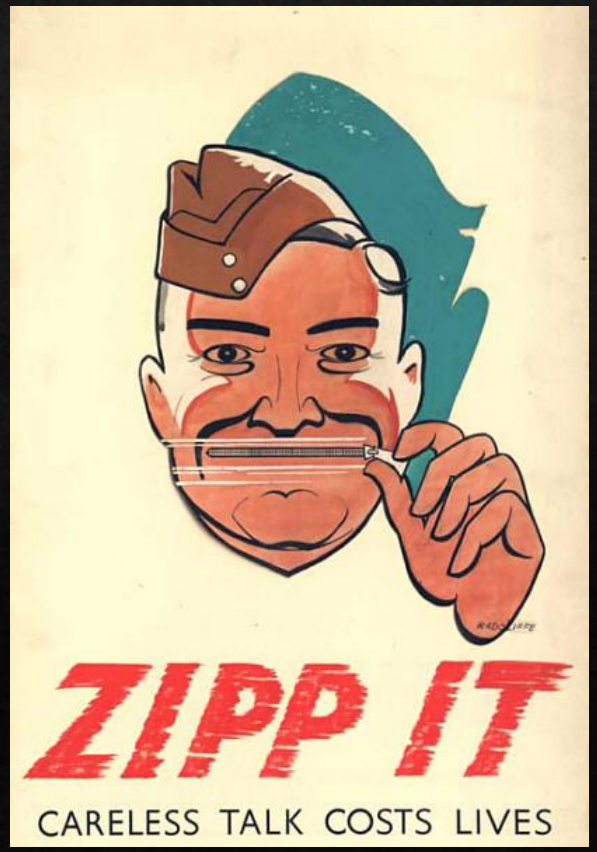
- 21+ years in the Security field
- Conducting OSINT and Security Research since the late 1990's
- Background in Incident Response & Incident Recovery, Cloud Security, Network Security, Threat Intelligence, Security Architecture, Security Strategy and Leadership
- Strong advocate of Threat Modeling as a foundation of Architecture design
- Recognized Social Media Platform Researcher(DEFCON, Voice of America, BSIDES, Podcasts, Threat Briefings)
- Volunteer for BSIDES Nova

Employer Statement

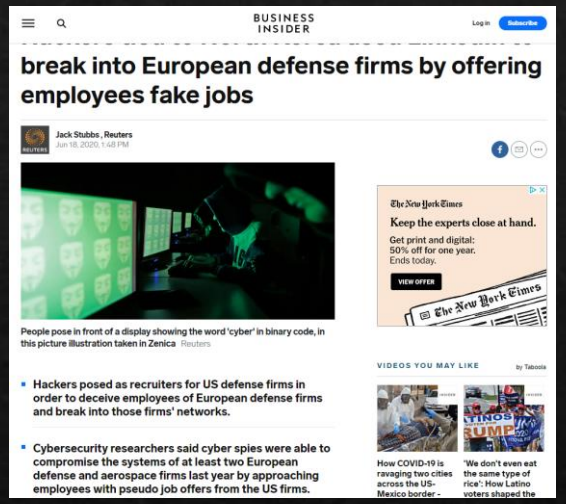
- My views, research and the data expressed in this deck **are my own research** and do not reflect the views of my **current** or previous employers

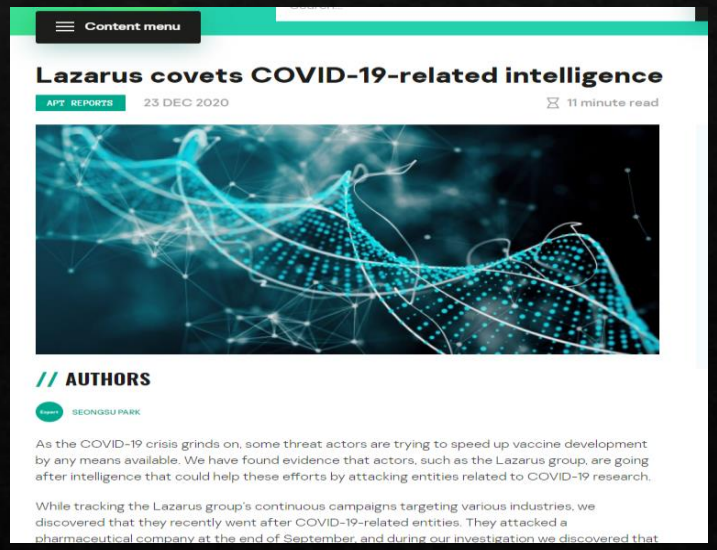
Rich Wickersham

Historical OpSec Parallels



The Threat = Targeting against LinkedIn





Was the focus in the right place?

Definitions and Conditions

- OSINT = Passive Adversary Reconnaissance through publicly available methods to allow for adversary Collection, Analysis and Targeting
- OPSEC = Understand what data is available to an adversary, how to protect the data and the responsibilities in protecting data
  - Platform Responsibility
  - Corporate Responsibility
  - Individual Responsibility
- TUA = Targeted User Analytics, Combines Red and Blue team capabilities to identify and reduce attack surface related to users
- ARM = Attract Retain and Monetize(you are the product)
- HIBP = Have I Been Pwned?

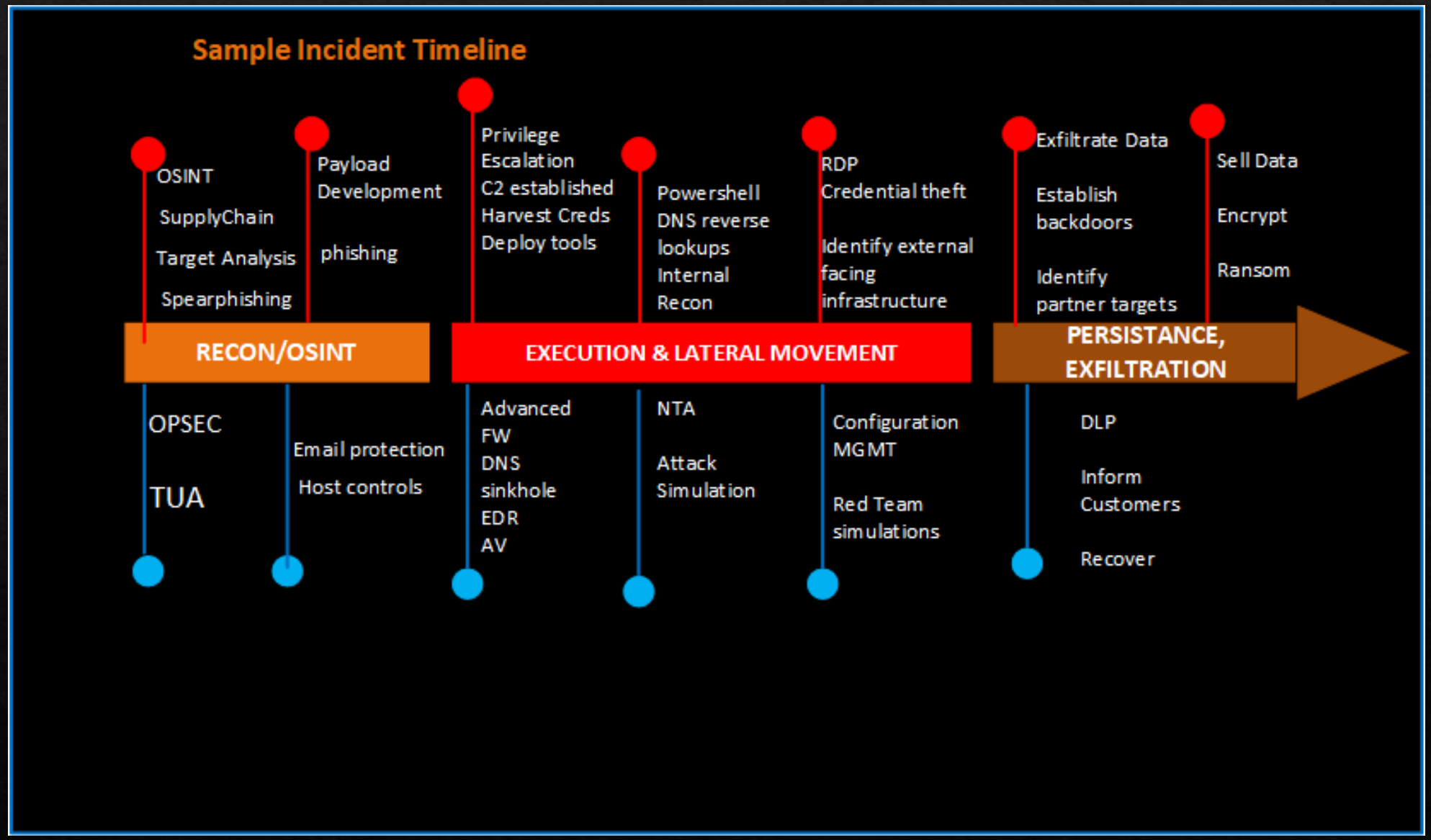
Hypothesis

- Our adversaries are using LinkedIn to harvest and target users with access to targeted data or systems
- Data Driven Targeting models can be built to simulate adversary behavior using openly available data gathered from Social Media platforms
- Targets can be externally enriched using OSINT data, Breach Data and through other methods
- Internal enrichment data **is available only to Defenders** and is extremely valuable when correlated with targeting data models
- Countermeasures can be employed using the same data driven approach our adversaries use to target us

Targeted User Analytics, TUA

- Who has Access?
- What is the probability that someone has access?
- What can go wrong from the Adversarial perspective?
- What story does the available data tell us?
  - Target Acquisition
  - Target Enrichment
- What countermeasures can we introduce?
  - Harden users
  - Harden Our environment
  - Deception

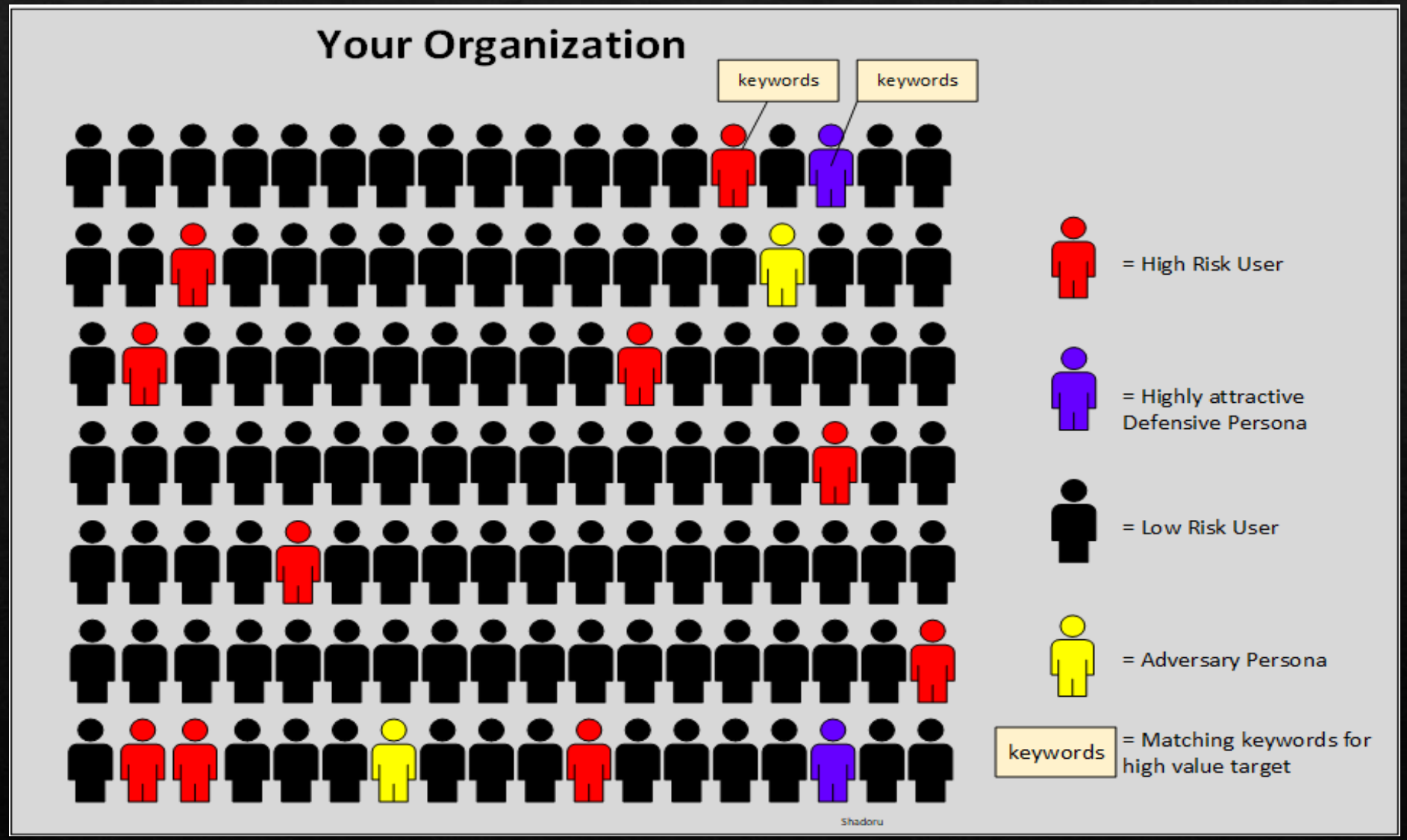
Typical Targeting scenario



TUA Process

- Target Acquisition
  - Keywords
  - Harvest
  - Acquisition Score
- Enrichment- External
  - OSINT
  - Breach Data Footprint
  - Enrichment Score
- Enrichment-Internal
  - Defender only Data Sets
  - Targeting score
- Countermeasures
  - User Hardening- OPSEC
  - Personas
  - Human Sensors

TUA


Your Organization
keywords
keywords
= High Risk User
= Highly attractive Defensive Persona
= Low Risk User
= Adversary Persona
keywords
= Matching keywords for high value target
Shadoru

Responsible Research Practices & Pre Conditions

- Is it legal to use a search Engine to perform **unauthenticated search** against LinkedIn to pull in user generated **public data**?
  - **Yes**, see LI vs HiQ ruling!
- Can we use public data sources for enrichment?
  - **Yes**
- Can researchers aggregate and pull in breach data to build a complete targeting model for the targets identified in the model?
  - **No**, so for this model I leveraged Have I been Pwned, HIBP just to generate a 1 OR 0 based on interesting data fields that are a match
  - Unfortunately we know the bad guys are using breach data sets
- Can we create Defensive Personas in Social Media platforms?
  - **Sometimes…**

LI Adversary Platform Targeting Methods

- Unauthenticated- All of my research follows this method
  - Google Fu
  - Web scraping with Python
- Authenticated
  - In Network (personas)
  - Recruiters
- Data Products
  - 3rd party aggregator data products (DataSift, HiQ)

TUA LI potential USE CASES

Verticals

- Financial Vertical
  - SWIFT
  - W2 Spear phishing
- Critical Infrastructure
  - Elections
  - Energy (several subcategories)
- Intellectual Property
  - TESLA
  - Biomedical
  - Pharmaceutical
- Government Contracting

Use Case = Biopharmaceutical

- Who has access to COVID19 Vaccine Research Data?
- What could go wrong if people with access to Vaccine Research data are collected from LI?
- Methods to enumerate targets using TUA based on LI data
- Modeling the attack surface
- Countermeasures

BioPharma Boolean strings & Seed files

- Sample string 1: syntax = SARS, SARS-CoV-2, engineer “target company” site: linkedin.com
- Sample string: 2 syntax = SARS, SARS-CoV-2, virologist “target company” site: linkedin.com
- SEED FILE: Vaccine Research, Development, clinical trials, virologist, pathogenesis, Clinical viro-Immunology, Experimental Immunology, SARS, SARS-CoV-2, MERS, Severe Acute Respiratory Syndrome, Coronavirus
  - Syntax= keyword + AND/OR/NOT
- MAPS & Models: incomplete

LinkedIn Harvest

- Keyword matching to identify high value targets
  - Seed files are grown based on common skills amongst harvested targets
  - If three or more targets share a skill that is not in the seed file, then add skill to seed file
- Geo Location where the people that have access to the target are located
  - Work locations derived from data posted on LI
  - Seed files can be refined to deanonymize targets that have removed company name based on geography, skill match OR unique job title
- Identified Targets reveal peers in industry
  - Targets leak this data through profile data that they post and connections
  - Knowing who the collaboration partners are in research exposes attack surface
- Org Chart Mapping of Relationships between targets
  - Targets have titles which are assigned a value (e.g. Manager, Employee)
  - Phishing will likely be based on this
  - We can use this data set to simulate the actual event and test our users(simulate phishing)
- Identified targets will be targeted repeatedly
  - Targeting models are valuable for years…

The Harvest

BioPharma Unauthenticated Harvest

Company	Harvest-Approximate	Average Seed match	Geo data leaked	Unique Names	External Enrichment
A	100	5	yes	31	TBD
B	100	4	yes	43	TBD
C	100	7	yes	33	TBD
D	100	4	yes	20	TBD

Note:

- *Anonymized Harvest Data is based on one data pull with 4 unique target org keyword strings and approximately 100 targets*
- *Some targets have a higher keyword match than the average.*

Target Enrichment for Adversaries

What OSINT data is available about the harvested targets:

- Simple Google search of name if the name of the target is unique
  - Geo Location data
  - Pictures, with exif data
  - Phone numbers, email addresses, physical address and other data points

What Breach Data is available about the harvested targets:

- Find External target Email addresses @yahoo, @hotmail, @gmail, etc..
- Leverage *HIBP for pointer to relevant breach data sets and fields
  - Assign **value of 1/Yes** for breach data available, **0/No for no breach data available** and add to conditional probability model
  - 1 or 0 value will be used because “you cant legally collect/store this data” even if you could obtain it
  - Determine whether breach data is available and what fields may have been exposed

Target Enrichment for Adversaries

OSINT & Breach Data Fields:

- User IDs, EMAIL address(@yahoo, @hotmail, @gmail)
- Breach footprint revealed via HIBP, provide a pointer to plaintext passwords or other data fields widely available breach data?
- Security questions-list questions and answers if available?
- Exif Data
- Target Mailing address
- Phone numbers-for vishing, or SIM swap
- Friends and professional contacts- for phishing
- GitHub
- Other Social Media profiles(e.g. Twitter)

Operationalizing Target Data

Adversary Execution Example BIO- Pharma

Data map to Role

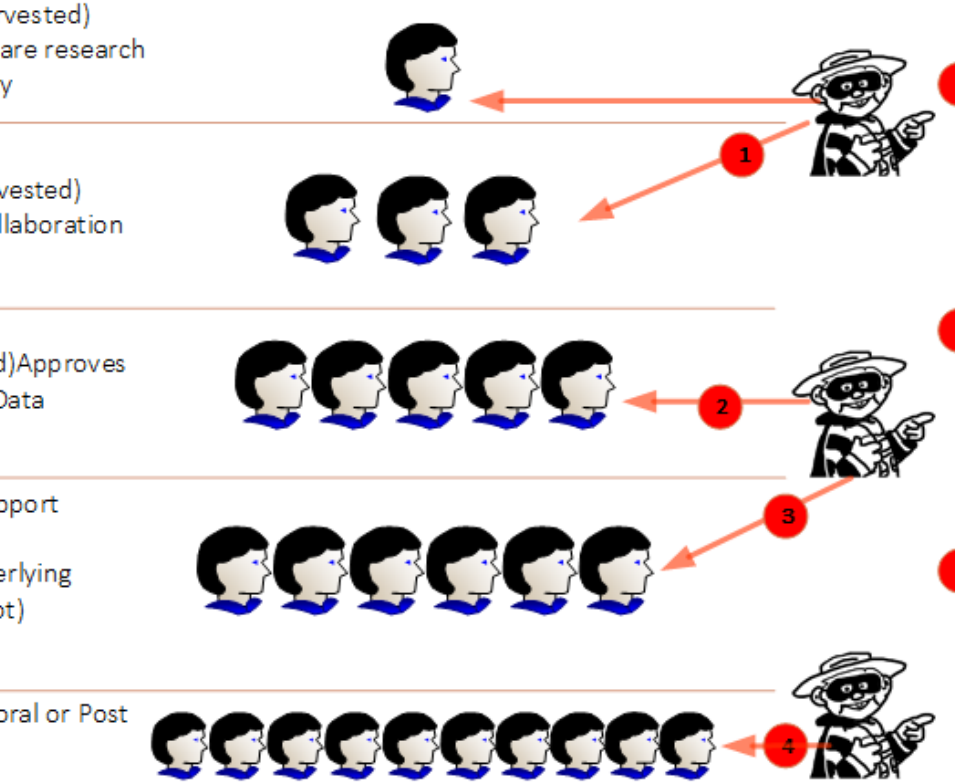
#=Senior Scientist (X harvested)
Requests Subordinates to share research data to third party

##=Sr. Virologist (# harvested)
Approves new research collaboration request

##=Team Lead (# harvested)Approves Access to research Data

##=Infrastructure Support (# harvested)
Has access to the underlying infrastructure/Root)

###=Research fellow, Doctoral or Post Doctoral
(## harvested)
(Presents further collaboration attack surface)



Threat #

1. "Adversary" **Spear Phishes** Senior Scientist(Spoofs VP R&D) with request to have subordinates externally share data
2. "Adversary" **is already inside your network** and enumerates and compromises accounts/ devices with access approval role
3. "Adversary" **is already in your network** and **exploits vulnerability in underlying infrastructure** or **compromises admin account** with Root access"
4. "Adversary" **is already in your network and enumerates and compromises user** or **workstation** with COVID-19 research data or system used to access research data

Vaccine research is expensive. Why would an adversary pay or wait for it when they could "potentially" steal it!

Countermeasures

Countermeasures = Harden Users

- OPSEC training
  - Train users on the impact of what they post
  - Train users on the value of breach data and how it can be used against them
  - Provide Corporate Security Awareness training, show actual data metrics about your users
- Create your own TUA
  - Learn the keywords, build and maintain a seed file
  - Search your user base continuously for high risk data matches
- Share knowledge
  - Partner with your employees, communicate the threats, OPSEC practices and the countermeasures

Countermeasures = Human Sensors

- Utilize Existing High Risk users as sensors
  - Turn a weakness into a strength
  - Add to High Risk User Monitoring groups
  - Trend data(phishing metrics) and enhanced monitoring
  - Build Internal Deception based on these users
  - Build Canary creds based on these users
- Utilize Former Employees as sensors
  - Stale profiles (former employees) appear active to adversaries
  - Deception infrastructure and creds based on real users
- Volunteer an employee as a sensor
  - This can work too but make sure it is believable

Countermeasures = LI Persona Development

- Create Personas that fit the target profile
  - Utilize TUA keywords to establish attractive targets
- Create LI relationships with other potential Targets or controlled personas
  - Short-term approach
  - Long-term approach
  - Establish relationships with known adversary personas
- Success Methods
  - Build Deception profiles over time
  - Build External artifacts(multi-platform personas)for target validation outside of LinkedIn
  - Build Internal Corporate Deception that aligns to external artifacts

The ARM Model

LinkedIn Issues

- User Responsibility vs LinkedIn Responsibility?
  - ARM model
- 2019 Metrics
  - 21,000,000 personas removed
  - Why prior year metrics are unavailable and my theory
- Questions for LinkedIn
  - Have shell companies bought LI data products?
  - How many persona recruiters have been detected prior to 2019?
  - Does LI plan to Inform users if they were contacted/exposed to a known Persona?
  - What capital investment has been made into security as a percentage of LI profit per year to prevent nation state threats?

My proposal to LI

- My proposal
  - LI should protect US interests by offering platform protections to “special” users
    - This is a responsibility that they must undertake.
    - MSFT should be pressured(by regulators) to force this.
  - LI should force a User Security Review(following the Facebook example)
  - LI should consider selling or providing **free** Deception products
  - LI should offer a pay for product that offers an encryption scheme that only shares our data within our trusted Networks(opt in model)

OPSEC

What not to post

- Keywords that indicate you have access to sensitive business data, systems or processes
- Keywords that indicate you have access to Classified data, Facilities, or Programs
  - If you work in these areas(e.g. DoD) you can count on job specific training in this area
- Pictures with metadata that could reveal or connect you or a coworker to a sensitive location
- Keywords with specific tools you use to protect your enterprise

What to be aware of

- Your social media footprint
- Breach data that can be used against you

What to do

- Assume Exposure
- Assume Breach
- Don’t be the reason that your company is breached!

Summary & Actions

- LI data is actively used by our adversaries to target our interests
- Targeting models built on this data can be reused outside of the platform for many years
- Blue teams should use LI data to model high risk users
- Red teams should use LI data to simulate adversary targeting
- Personas(of all types) are prevalent in the LI platform and you should consider building them as sensors!
- OPSEC is a User, Platform and Corporate shared responsibility…

Questions

Backup

- Approaches to handling Adversary Personas
- Tools list
- LABS
  - Labs Exif enrichment and analysis
- Rick Astley

Handling Adversary Personas

Handling Adversary Personas

“I have identified a persona/persona recruiter that is targeting my Corporate Users…”

- Inform LinkedIn
  - Process can take weeks or longer
  - Capture detail about the persona before it is gone
  - Look for other indicators of execution against targets
- Or Leave in place
  - Use as a sensor
  - Who are they connected to?
  - What do they promote?
  - When did the activity start and stop?
  - When to reach out for assistance?

Basic Tools

- Google Fu/Google Dorking!!
- HIBP have I been pwned (and pwned Passwords API)
- Pastebin
- Spiderfoot(or similar OSINT tools)
- Exiftool Exchangeable image file format data analysis tool
- Python
- Reverse image lookup tools(Google)
- Out of the box thinking!

Labs

Lab-Enrichment Exif/GEO Location Data

Exif = EXchangeable Image Format = Metadata in photos

- GeoTagging, location data, camera, altered images
- Used in my model for target location validation
- Most Social media platforms started stripping this data out
  - Source jpg vs Processed jpgs
- This data can be altered!
- If you are trying to obfuscate your location you should remove or alter location data before sharing any photo and be aware of which platforms process an image and which ones don't…
- Class should play “where’s Waldo” with my Defcon photo and find and analyze the image to establish a target location.

Enrichment Exif/GEO Location Data


```
Select C:\Users\Rich\Desktop\exiftool(-k).exe
Media Black Point                : 0 0 0
Red Matrix Column                : 0.43604 0.22249 0.01392
Green Matrix Column              : 0.38512 0.7169 0.09706
Blue Matrix Column               : 0.14305 0.06061 0.71391
Red Tone Reproduction Curve      : (Binary data 32 bytes, use -b option to extract)
Chromatic Adaptation             : 1.04788 0.02292 -0.05019 0.02959 0.99048 -0.01704
 -0.00922 0.01508 0.75168
Blue Tone Reproduction Curve     : (Binary data 32 bytes, use -b option to extract)
Green Tone Reproduction Curve    : (Binary data 32 bytes, use -b option to extract)
Image Width                      : 2448
Image Height                     : 3264
Encoding Process                 : Baseline DCT, Huffman coding
Bits Per Sample                  : 8
Color Components                 : 3
Y Cb Cr Sub Sampling             : YCbCr4:2:0 (2 2)
Aperture                         : 2.4
Image Size                       : 2448x3264
Megapixels                       : 8.0
Scale Factor To 35 mm Equivalent: 7.4
Shutter Speed                    : 1/15
Create Date                      : 2019:08:09 12:22:21.543080
Date/Time Original               : 2019:08:09 12:22:21.543080
Modify Date                      : 2019:08:09 12:22:21.543080
Thumbnail Image                  : (Binary data 7171 bytes, use -b option to extract
)
GPS Altitude                     : 622 m Above Sea Level
GPS Date/Time                    : 2019:08:09 19:22:18Z
GPS Latitude                     : 36 deg 6' 39.06" N
GPS Longitude                    : 115 deg 10' 13.42" W
Circle Of Confusion              : 0.004 mm
Field Of View                    : 71.5 deg
Focal Length                     : 3.4 mm (35 mm equivalent: 25.0 mm)
GPS Position                     : 36 deg 6' 39.06" N, 115 deg 10' 13.42" W
Hyperfocal Distance              : 1.17 m
Light Value                      : 4.3
-- press ENTER --
```

Source data removed from Twitter download

C:\Users\Rich\Desktop\exiftool(-k).exe

```
ExifTool Version Number         : 12.19
File Name                       : wick defcon.jfif
Directory                       : C:/Users/Rich/Desktop
File Size                       : 10 KiB
File Modification Date/Time     : 2021:05:09 12:03:04-04:00
File Access Date/Time           : 2021:05:09 12:03:04-04:00
File Creation Date/Time         : 2021:05:09 12:03:03-04:00
File Permissions                : -rw-rw-rw-
File Type                       : JPEG
File Type Extension             : jpg
MIME Type                       : image/jpeg
JFIF Version                    : 1.01
Resolution Unit                 : None
X Resolution                    : 1
Y Resolution                    : 1
Image Width                     : 194
Image Height                    : 259
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
Image Size                      : 194x259
Megapixels                      : 0.050
-- press ENTER --
```

Altered Exif Data to maintain OPSEC or throw off an adversary

Note: If you found the OSINT object, I was not in Red Square(**55.754093, 37.620407**) during Defcon 27...

Fun Facts

- Rick Astley has a large presence on LinkedIn
  - He has such good OPSEC that he has created over **268 personas**
  - https://www.linkedin.com › pub › dir › rick › Astley
  - Defensive personas work!