CS 3546

Introduction to Security Analytics

Topics: Threat Hunting

Instructor: Mir Mehedi A. Pritom

Department of Computer Science

Spring 2023

What is Threat?

- In cybersecurity context, any activities or artifacts that can cause damage to any system/network/organization is a **threat**
- Traditional threat Defn: **Capability** + **Intent**
- New Threat Paradigm: **Capability** + **Intent** + **knowledge**
  - Capability: includes tools and ability to access
  - Intent: the motivation
  - Knowledge: is specific, sophisticated ability to operate within a system/network after getting access!

Attacker's Advantage

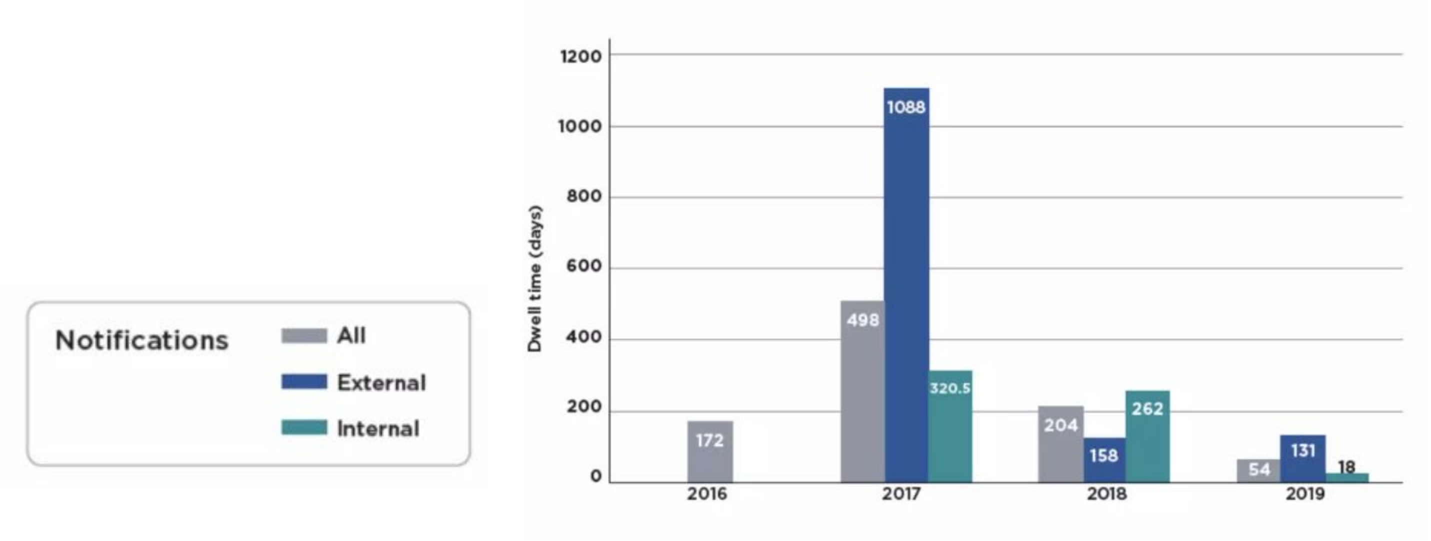
- Attacker’s need to be successful only once!
- Can be persistent
- Custom malware, 0days, social engineering
- Often funded adversaries!

Defender's Disadvantages

- Often times Underfunded/ Understaffed
- Increase complexity of IT infrastructure:
  - Moving to Cloud
  - Bring your Own device
  - Work from home
  - Virtualization
- Prevention control fails to block all threats
- Hundreds of System Vulnerabilities to patch

Dwell Time

- Dwell time: The number of days the attack artifacts is present in a victim network before it is detected and reported.
- Goal: Is to reduce the dwell time



Existing Issues

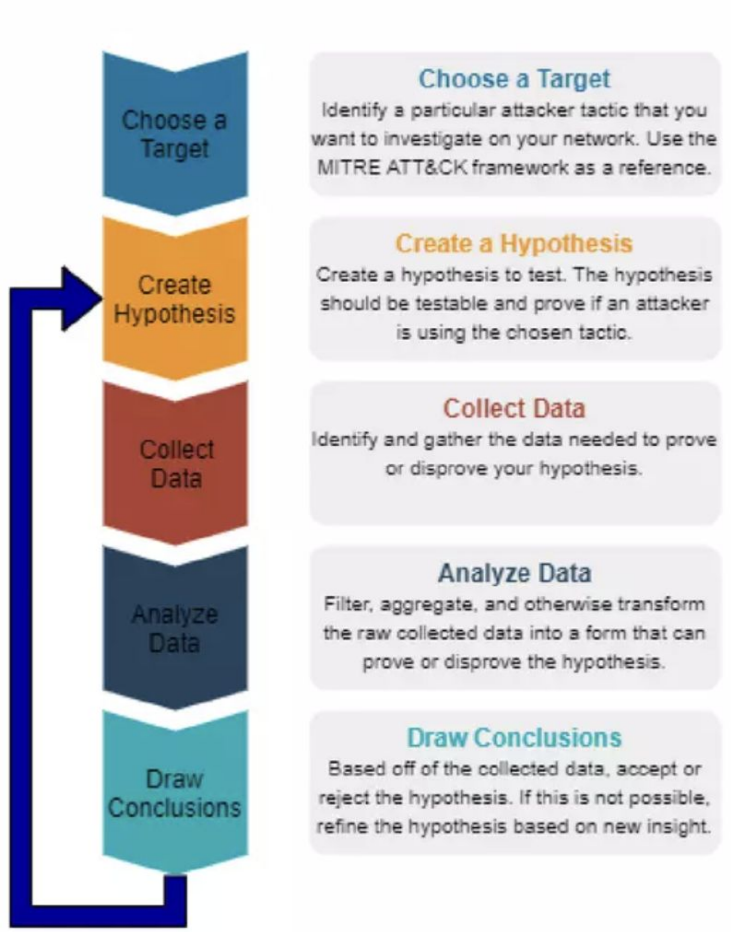
- Endpoint devices, OS, and networks, all have a certain level of vulnerabilities
- Organizations fail to prevent adversaries from getting in.
- Advanced Adversaries (APTs) manage to hide their tracks for days, month, or even years.

Threat Hunting- The Proactive Cyber defense Approach

- Hunting processes: Perform proactive and iterative discovery through network, endpoints, and other infrastructures to detect and respond to cyber threats those evaded existing protection
- Hunting is different from traditional defense approaches: IDS, Firewall, malware sandbox, or SIEM
  - Investigation of evidence-based data after there is a warning of potential threat
- No warning is generated in hunting
  - Based on the intuition/hunches of the Hunter (human intelligence)

Threat Hunting- Principle

- **Presumption of compromise:** Your prevention mechanism will fail (or already failed), thus assume breach mentality will increase awareness for potential compromise of assets



Choose a Target
Identify a particular attacker tactic that you want to investigate on your network. Use the MITRE ATT&CK framework as a reference.

Create a Hypothesis
Create a hypothesis to test. The hypothesis should be testable and prove if an attacker is using the chosen tactic.

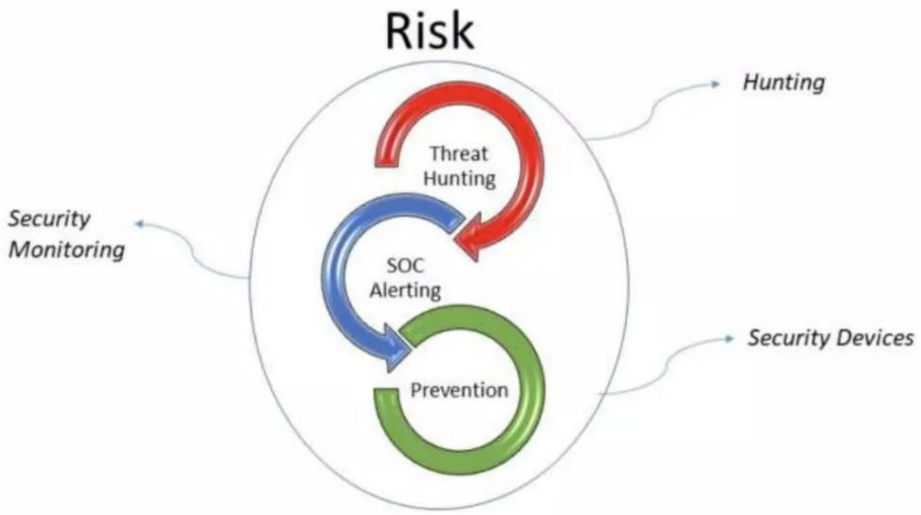
Collect Data
Identify and gather the data needed to prove or disprove your hypothesis.

Analyze Data
Filter, aggregate, and otherwise transform the raw collected data into a form that can prove or disprove the hypothesis.

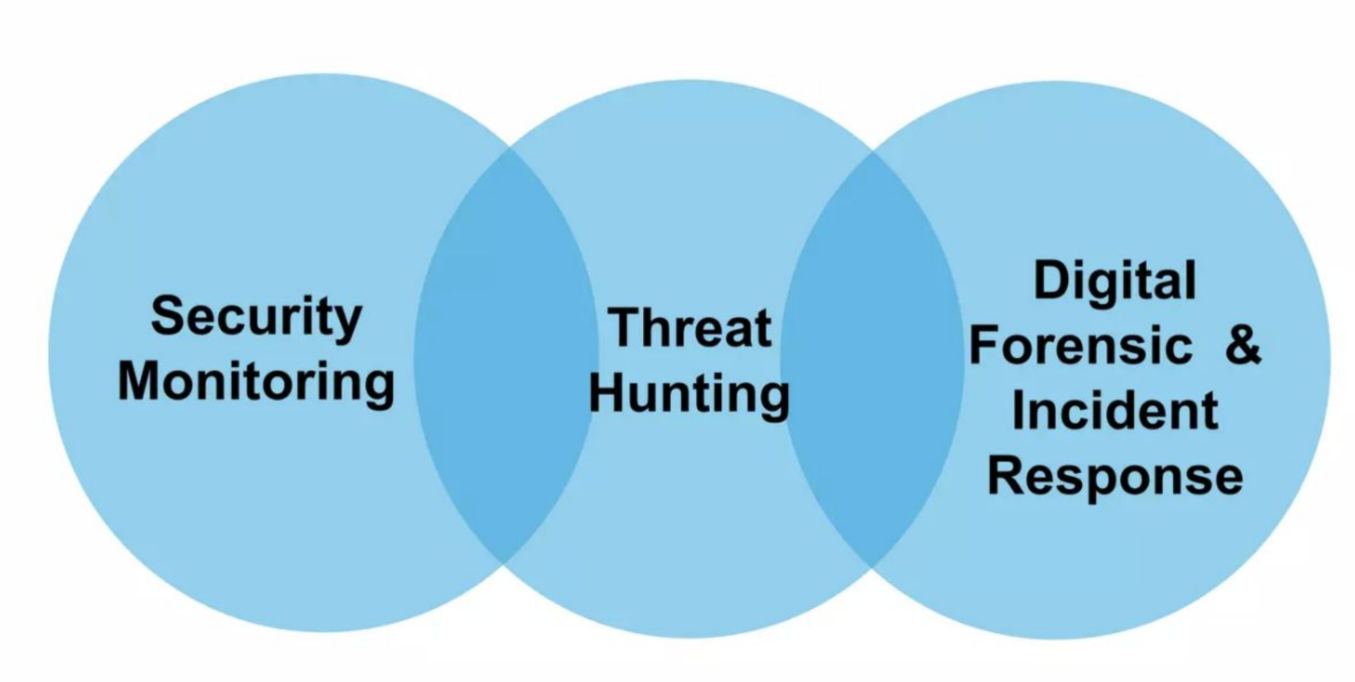
Draw Conclusions
Based off of the collected data, accept or reject the hypothesis. If this is not possible, refine the hypothesis based on new insight.

Why threat hunting?

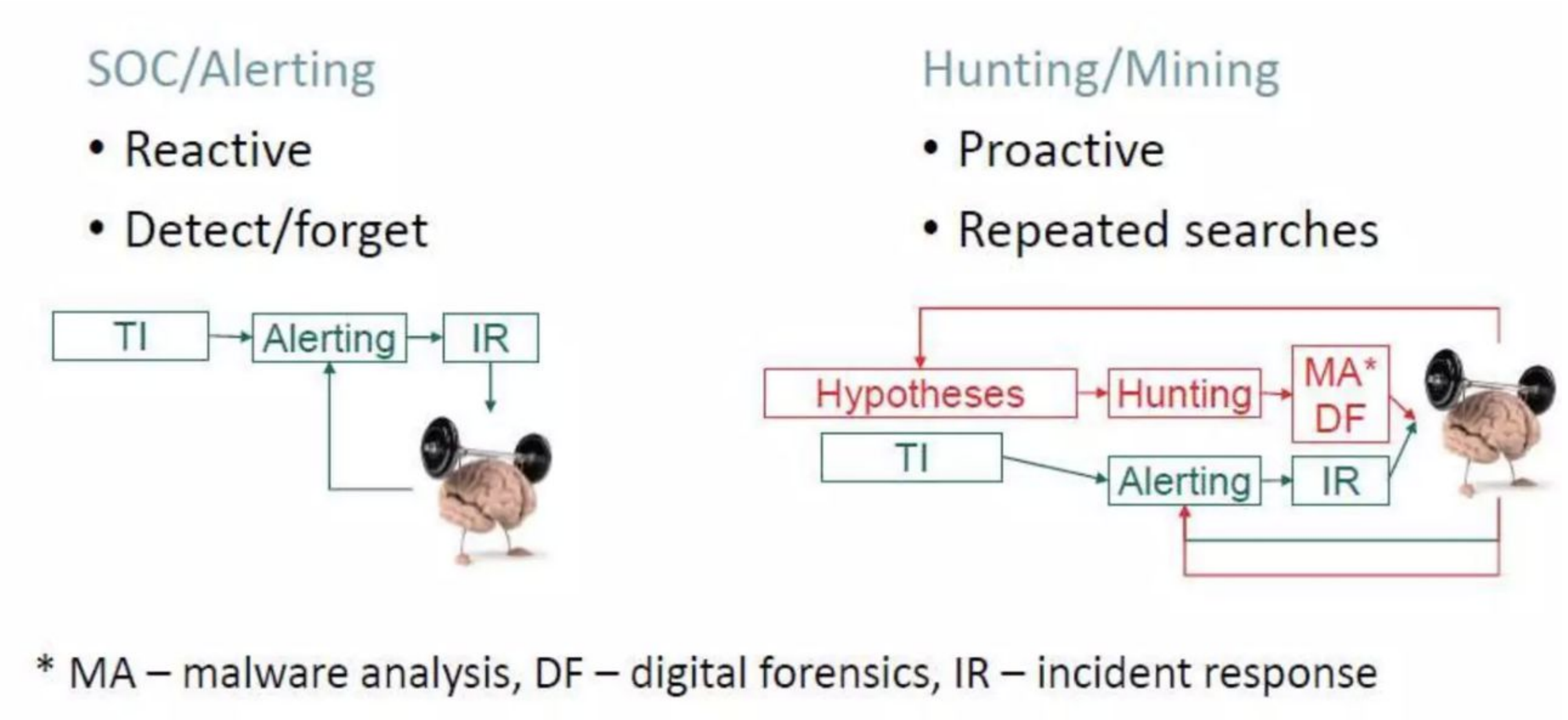
- Business perspective:
  - Minimize Risks
  - Minimize Dwell time
- Technical perspective:
  - Advanced/new attack detection
  - TTPs based detection
  - Non-malware attack detection



Why threat hunting?



threat hunting Vs Alert-Based Investigation



SOC/Alerting

- Reactive
- Detect/forget

TI → Alerting → IR

Hunting/Mining

- Proactive
- Repeated searches

Hypotheses → Hunting → MA* DF

TI → Alerting → IR

* MA – malware analysis, DF – digital forensics, IR – incident response

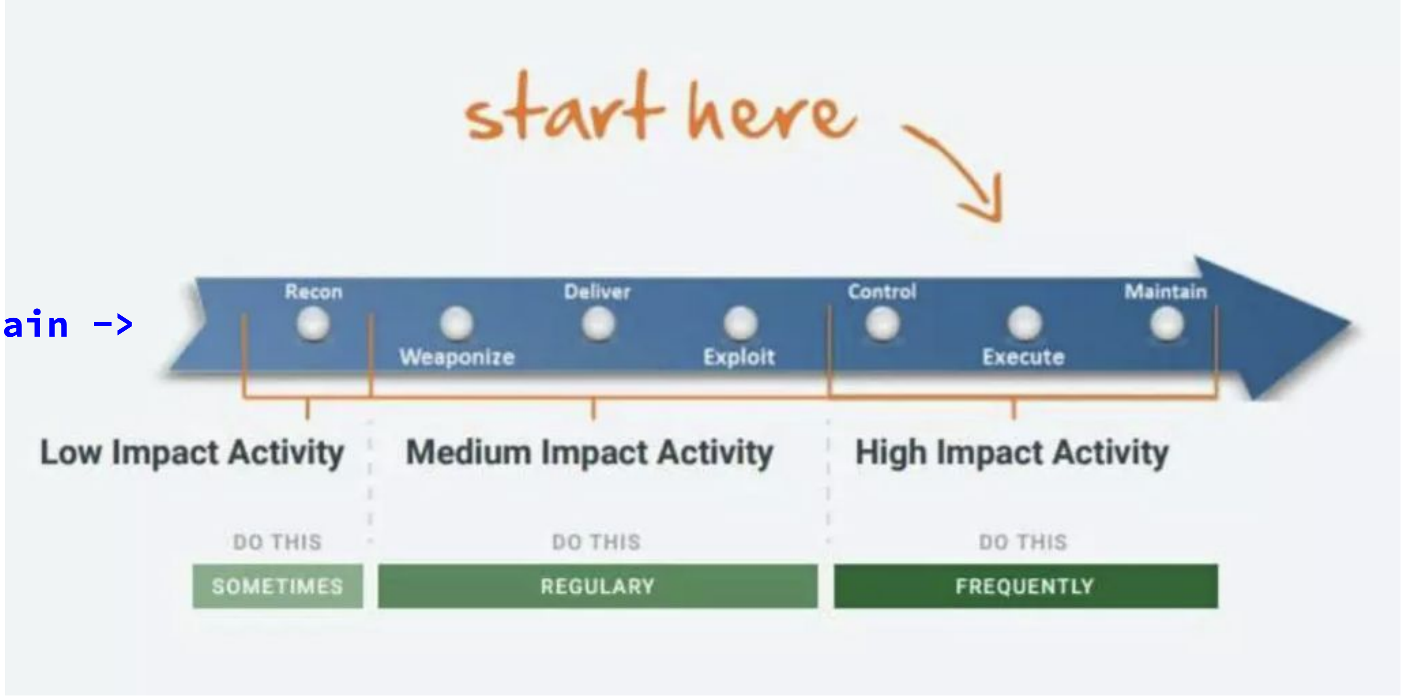
THREAT HUNTING Vs DIGITAL FORENSICS (COMPROMISE ASSESSMENT)

There are differences in 3 perspectives:

- Situation & Condition: TH-> assume compromise will happen **[Proactive]**; CA-> compromise already happened **[Reactive]**
- Location & Object: TH-> Anywhere in the system/network; CA-> selected system/network component/suspected area
- Actor (who performed the activities?): TH->Part of SOC team; CA-> usually DFIR (digital forensic incident response) team

Threat Hunting Activity

Kill chain ->



Human Hunters- Skillsets

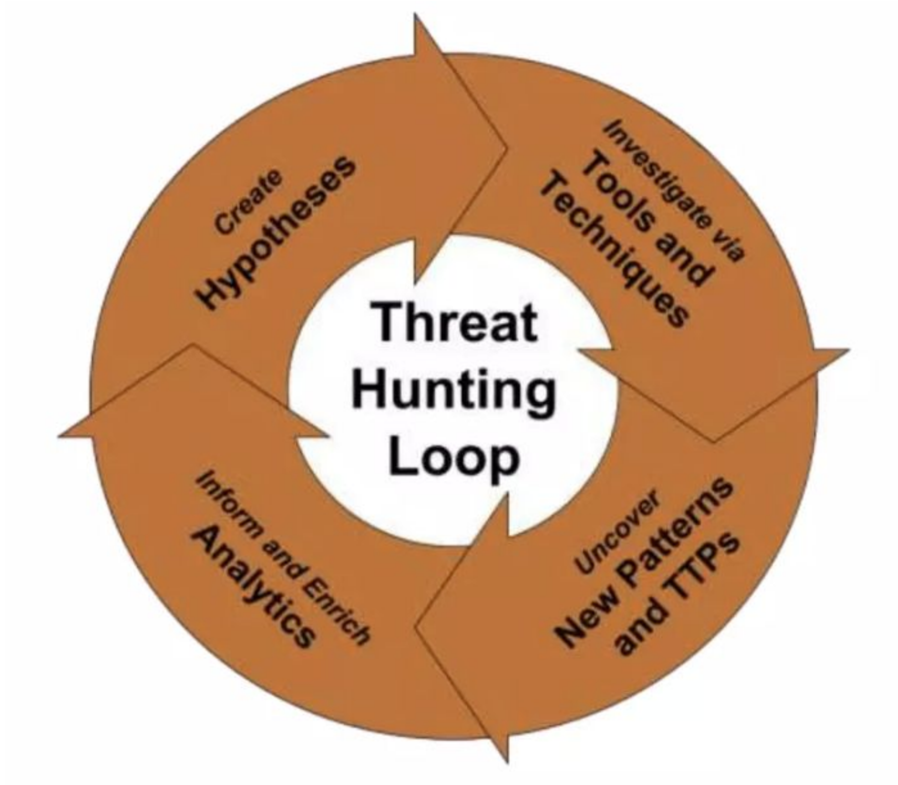
- Analytical mindset: Having a mindset of curiosity, generate and investigate hypothesis, try to create RQs and answer them through your data.
- Network Architecture: generic computer network knowledge, OSI layers, TCP/IP, Basics of protocols (DNS, DHCP, HTTP, FTP, …)
- OS: Basic knowledge of OS and system security (for system/kernel level analysis)
- Attack methods/TTPs/Attack life cycle: know about about attack vectors (entry points/targets), what tools and techniques attackers may use at different attack stages.

Human Hunters- Skillsets

- Log analysis: handle logs and data analysis skills, try to see the big picture from logs/data
- Cyber Threat intelligence: Threat intel can give you lead on new attacks to unearth them in your system/network
- Malware analysis: To determine the purpose and origin by analyzing an instance of malware
- Tools for hunting: Python Pandas, ELK Platform, Open PCAP (wireshark), Collect and manage network/OS logs

Threat Hunting Processes

-



Create Hypotheses
Investigate via Tools and Techniques
Uncover New Patterns and TTPs
Inform and Enrich Analytics
Threat Hunting Loop

Threat Hunting Processes

- **Step 1: Create Hypothesis:**

  threat hunting begins with questions “How can attacker infiltrate our infrastructure?”

- These questions need to be broken down into specific measurable hypotheses:
  - What is our prime assets?
  - What threats can be present in the network?
  - How can we identify threat actors?
- **Hypothesis can not be generated by tools!**
  - It is defined by threat hunter mindset and knowledge

Threat Hunting Processes

- **Step 2- Tools and Techniques:**

  Once the hypothesis is set, then we need to leverage all the available tools and techniques to implement detection engineering tasks.

Threat Hunting Processes

- **Step 3- Uncover New Patterns/TTPs:**

  Prove if the hypothesis can be proven or not. Even if the hypothesis result is not proven, that does not necessarily mean that no malicious activity is present or the hunters create a wrong hypothesis. It might be the case that current infrastructure is insufficient to prove the hypothesis, which can be reveal new TTPs later on.

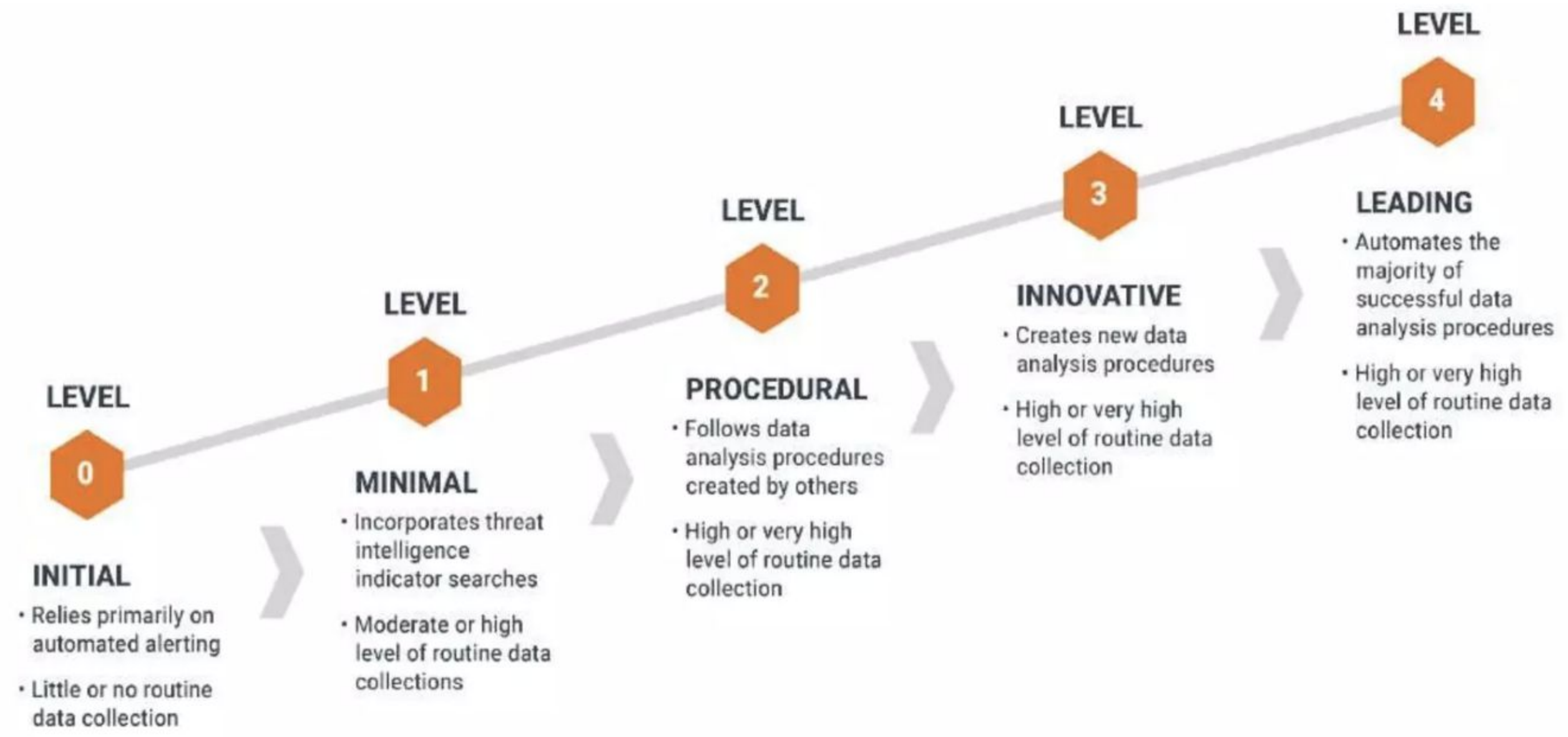
Threat Hunting Processes

- **Step 4-Inform & Enrich Analytics:**

  Once new knowledge is gained, the team should have a mindset to automate the process to save time for continuous repetitive process. Can be done in different ways-

  - Scheduling a saved search
  - Develop a new analytic/visualization within the existing tool
  - Providing feedback to a supervised ML model

Threat Hunting Maturity Model



LEVEL 0

INITIAL

- Relies primarily on automated alerting
- Little or no routine data collection

LEVEL 1

MINIMAL

- Incorporates threat intelligence indicator searches
- Moderate or high level of routine data collections

LEVEL 2

PROCEDURAL

- Follows data analysis procedures created by others
- High or very high level of routine data collection

LEVEL 3

INNOVATIVE

- Creates new data analysis procedures
- High or very high level of routine data collection

LEVEL 4

LEADING

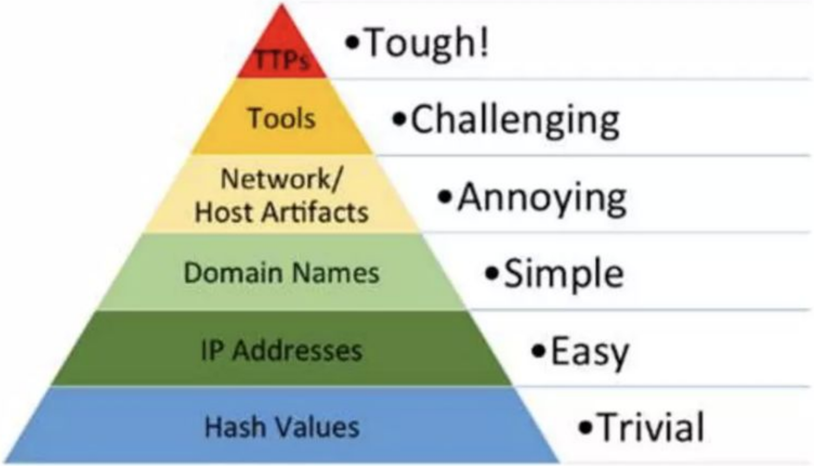
- Automates the majority of successful data analysis procedures
- High or very high level of routine data collection

Threat Hunting Framework

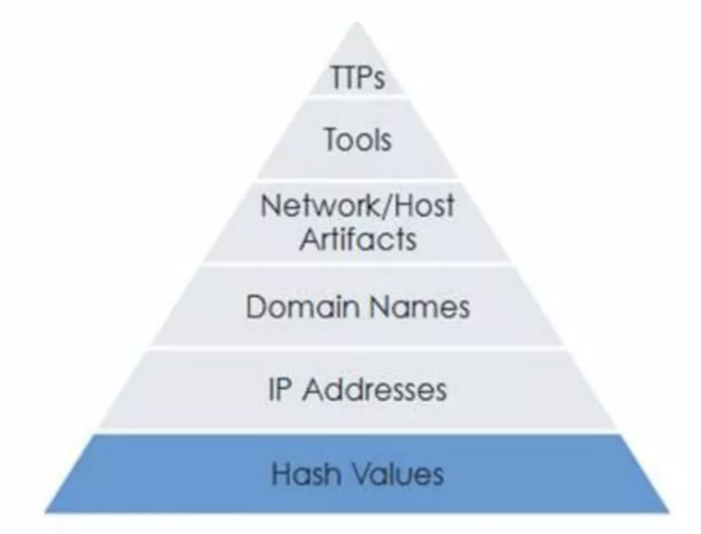
- The common Cybersecurity framework used by threat hunters to start their processes:
  - Pyramid of pain
  - Cyber Kill Chain
  - ~~MITRE ATT&CK~~ [Not covered in this class]

Pyramid of Pain

- Pyramid of pain represents the usefulness of your intelligence
- The higher the stack the more the adversaries have to expend for resources
- It also indicates to gather artifacts/threat intelligence from adversaries.



Pyramid of Pain: Hashes

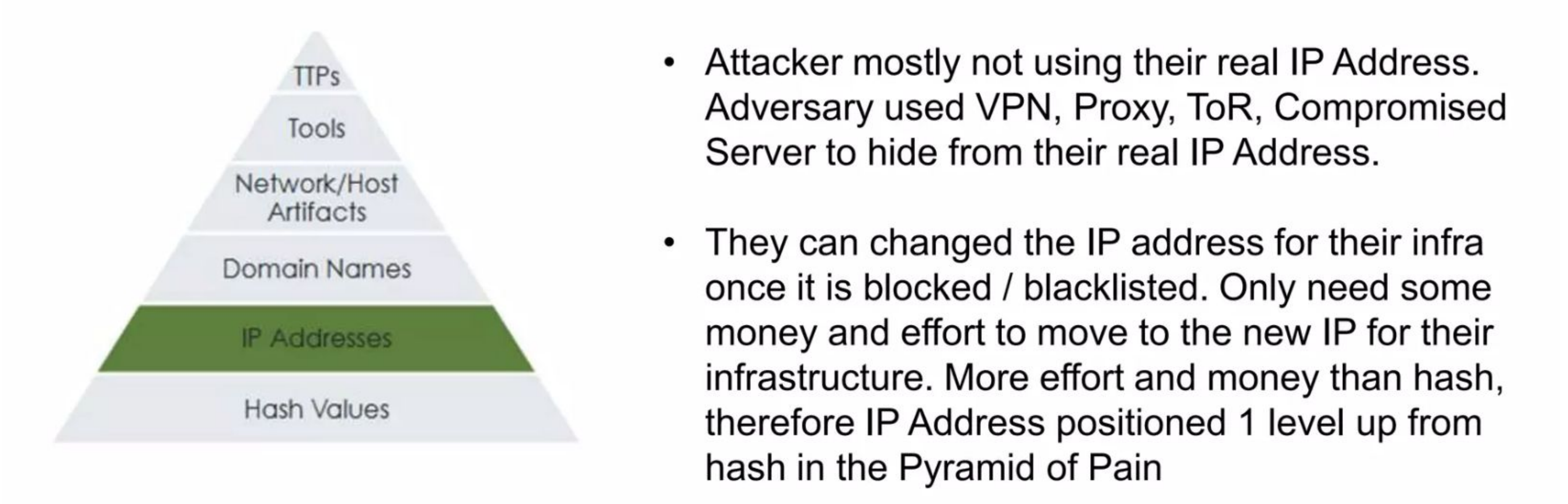


- Hash is so far the **highest confidence level** from artifacts collected or gathered from intel resources

- But, hash is **very easy to change**. Adversaries only need a lil bit effort to modify and create a new hash for their tradecraft

- It is maybe the reason why hash positioned **in the bottom of the pyramid** stack

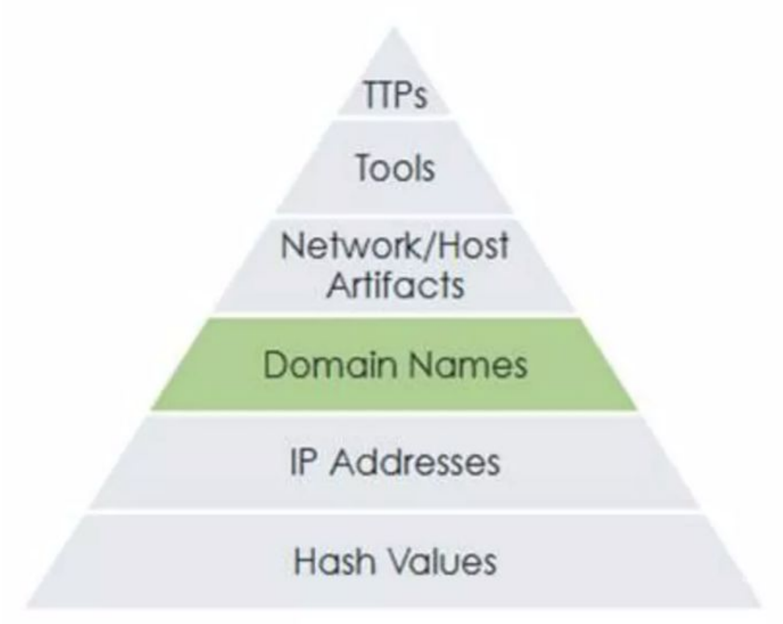
Pyramid of Pain: IP Address



- Attacker mostly not using their real IP Address. Adversary used VPN, Proxy, ToR, Compromised Server to hide from their real IP Address.

- They can changed the IP address for their infra once it is blocked / blacklisted. Only need some money and effort to move to the new IP for their infrastructure. More effort and money than hash, therefore IP Address positioned 1 level up from hash in the Pyramid of Pain

Pyramid of Pain: Domain Name

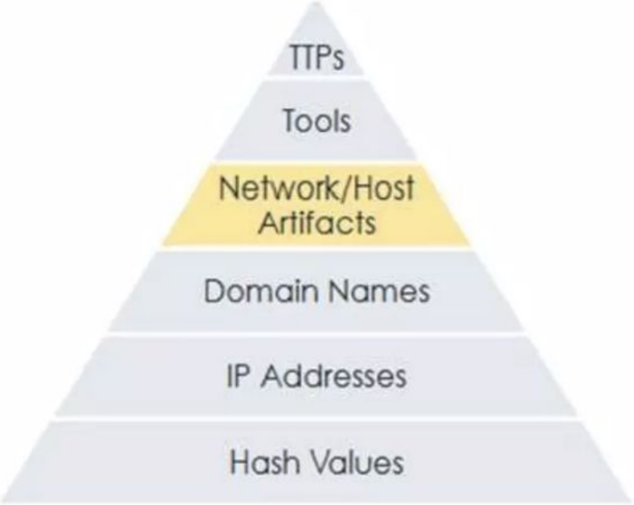


- Almost easy as IP Address to change the domain name. But need more time (Domain propagation in DNS)

- Need some registration, and for some reason they mostly hide the whois for domain privacy offered by domain registrars. Need more money for this services.

- Need to define the domain name. And it is not easy. Sometimes adversaries made bot to automatically create a new domain using certain algorithm (DGA)

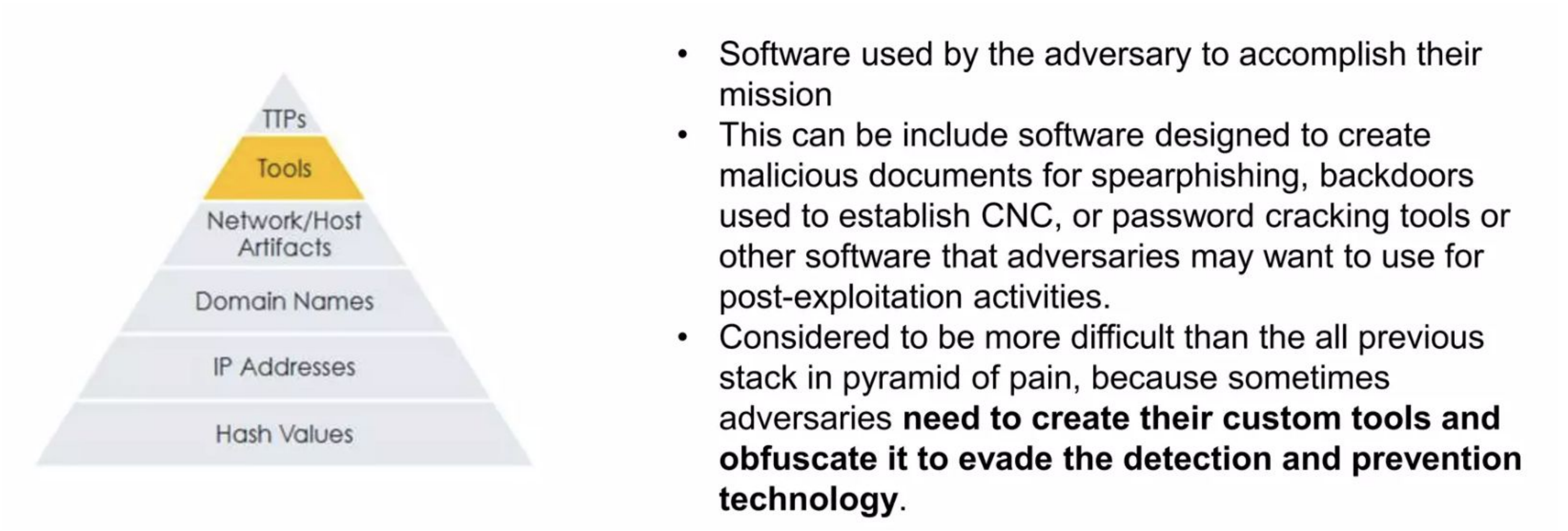
Pyramid of Pain: Network/Host Artifacts



- Network Artifacts : indicators of activities performed by the adversaries on the network. Anything communicated over the network by the adversary can be referred to as network artifact, which includes URI patterns, SMTP mailer values, HTTP user agent, and the like.

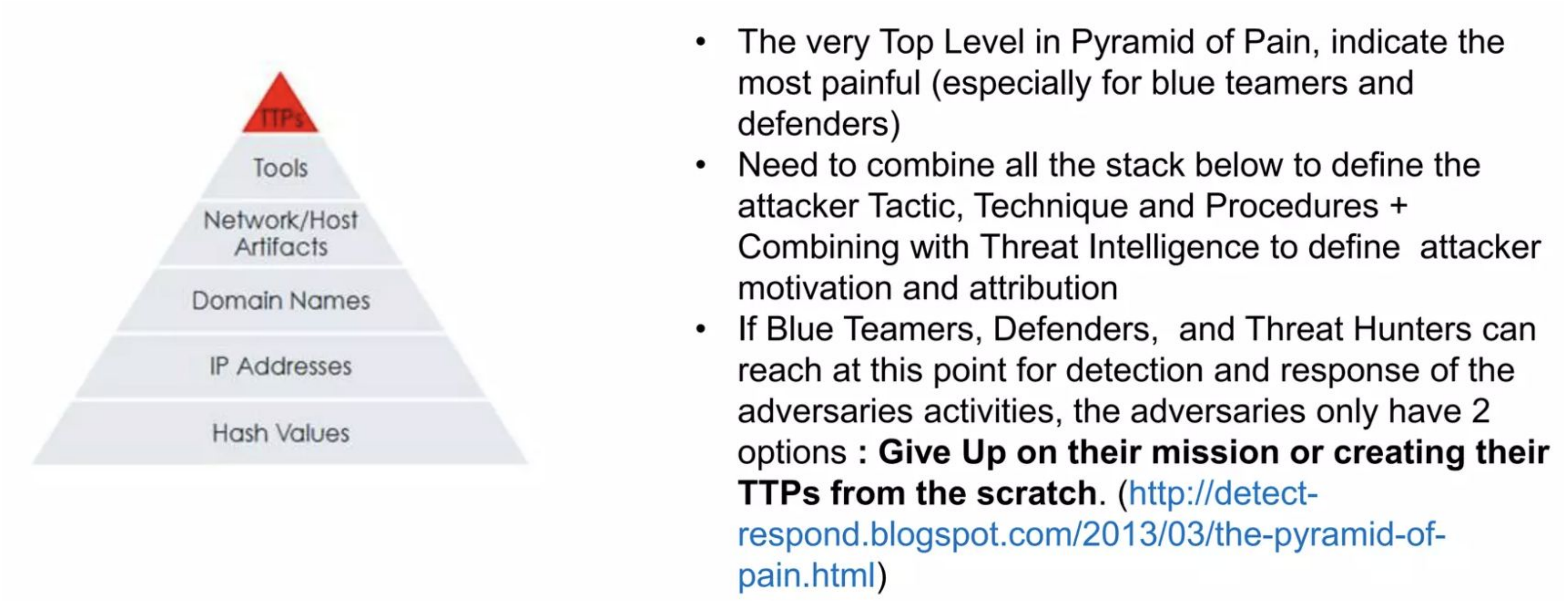
- Host Artifact : Indicators of activities performed by the adversaries on the hosts. Artifacts like registry keys or values created by malware. Files or directories injected in specific locations, and the like are considered as host artifacts.

Pyramid of Pain: Tools



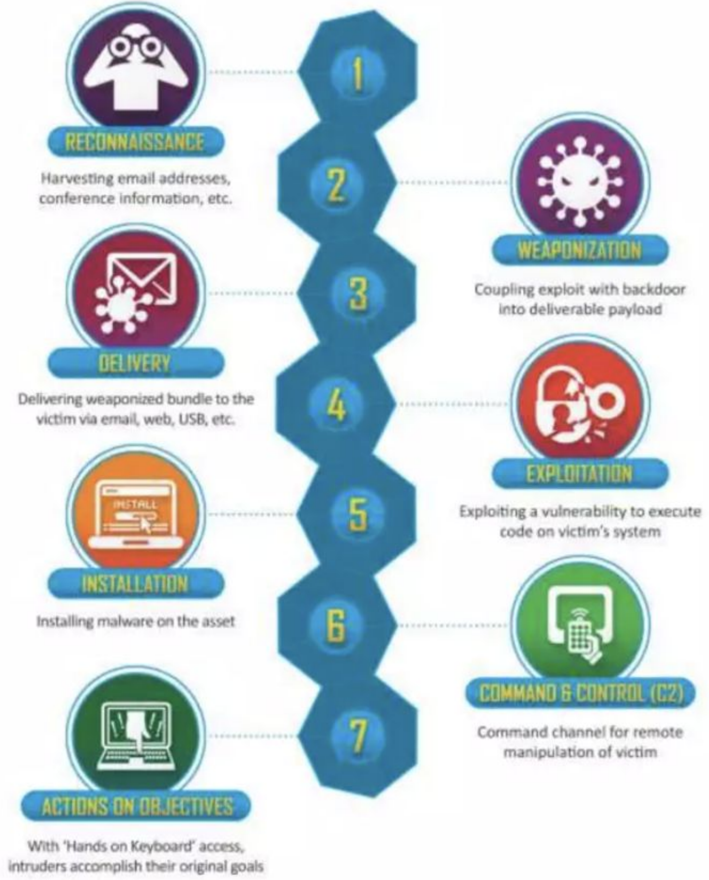
- Software used by the adversary to accomplish their mission
- This can be include software designed to create malicious documents for spearphishing, backdoors used to establish CNC, or password cracking tools or other software that adversaries may want to use for post-exploitation activities.
- Considered to be more difficult than the all previous stack in pyramid of pain, because sometimes adversaries **need to create their custom tools and obfuscate it to evade the detection and prevention technology**.

Pyramid of Pain: TTPs



- The very Top Level in Pyramid of Pain, indicate the most painful (especially for blue teamers and defenders)
- Need to combine all the stack below to define the attacker Tactic, Technique and Procedures + Combining with Threat Intelligence to define attacker motivation and attribution
- If Blue Teamers, Defenders, and Threat Hunters can reach at this point for detection and response of the adversaries activities, the adversaries only have 2 options : **Give Up on their mission or creating their TTPs from the scratch**. (http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html)

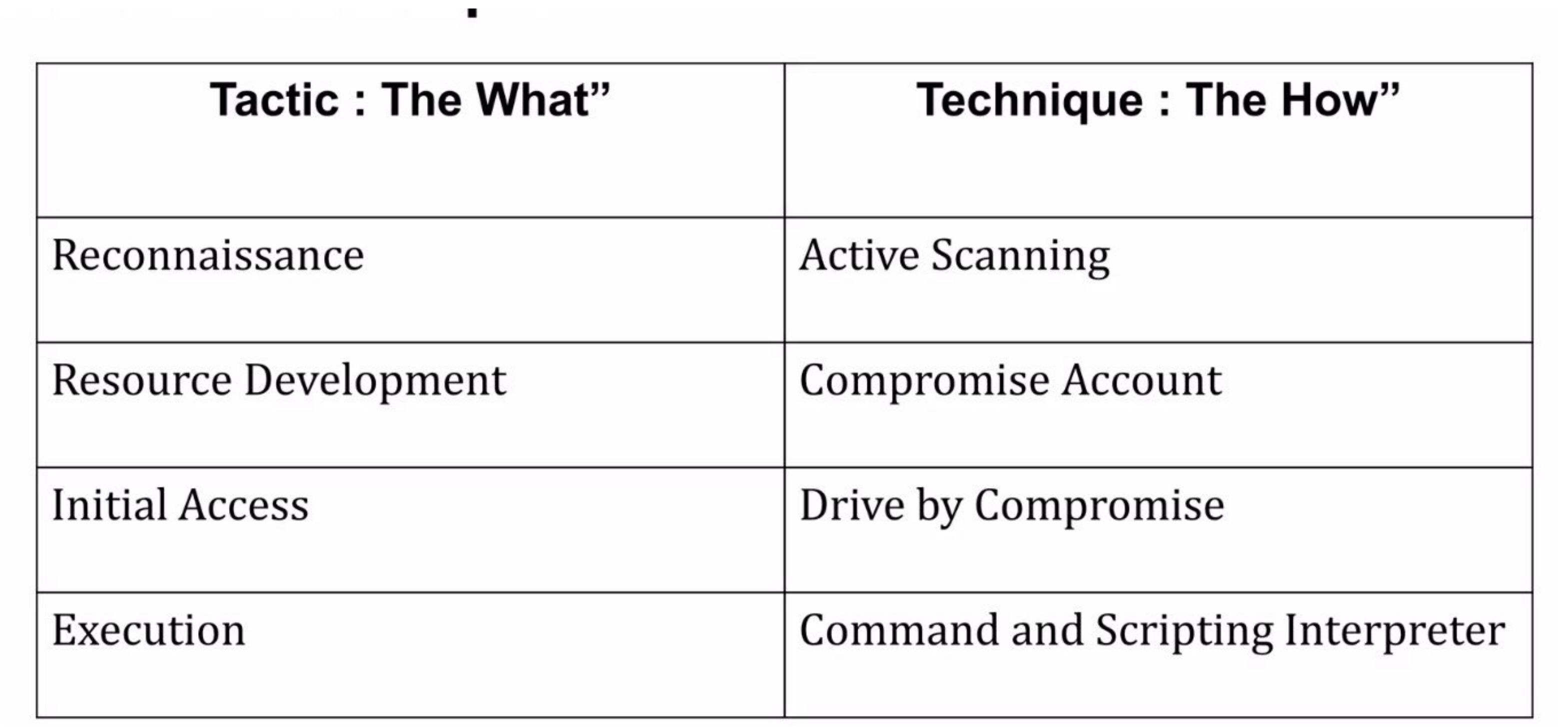
Cyber Kill Chain



1
RECONNAISSANCE
Harvesting email addresses, conference information, etc.
2
WEAPONIZATION
Coupling exploit with backdoor into deliverable payload
3
DELIVERY
Delivering weaponized bundle to the victim via email, web, USB, etc.
4
EXPLOITATION
Exploiting a vulnerability to execute code on victim's system
5
INSTALL
INSTALLATION
Installing malware on the asset
6
COMMAND & CONTROL (C2)
Command channel for remote manipulation of victim
7
ACTIONS ON OBJECTIVES
With 'Hands on Keyboard' access, intruders accomplish their original goals

https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

Technique Vs Tactic



Tactic : The What”	Technique : The How”
Reconnaissance	Active Scanning
Resource Development	Compromise Account
Initial Access	Drive by Compromise
Execution	Command and Scripting Interpreter

Types of Threat Hunting

1. IoC (Indicator of Compromise) based hunting
2. Hypothesis based hunting
3. Baseline based hunting
4. Anomaly Based hunting

IOC based hunting

- Hunting based on IOC collected from Threat Intel sources
- More like Compromise Assessment (Forensic job)
- Cross check if IOC is present in environment
- Checking on specific threat actor and threat intel report

Hypothesis based hunting

- Creating a Hypothesis for certain TTPs on specific objects
- Leverage specific framework for creating hypothesis
- Defining assets for hunting

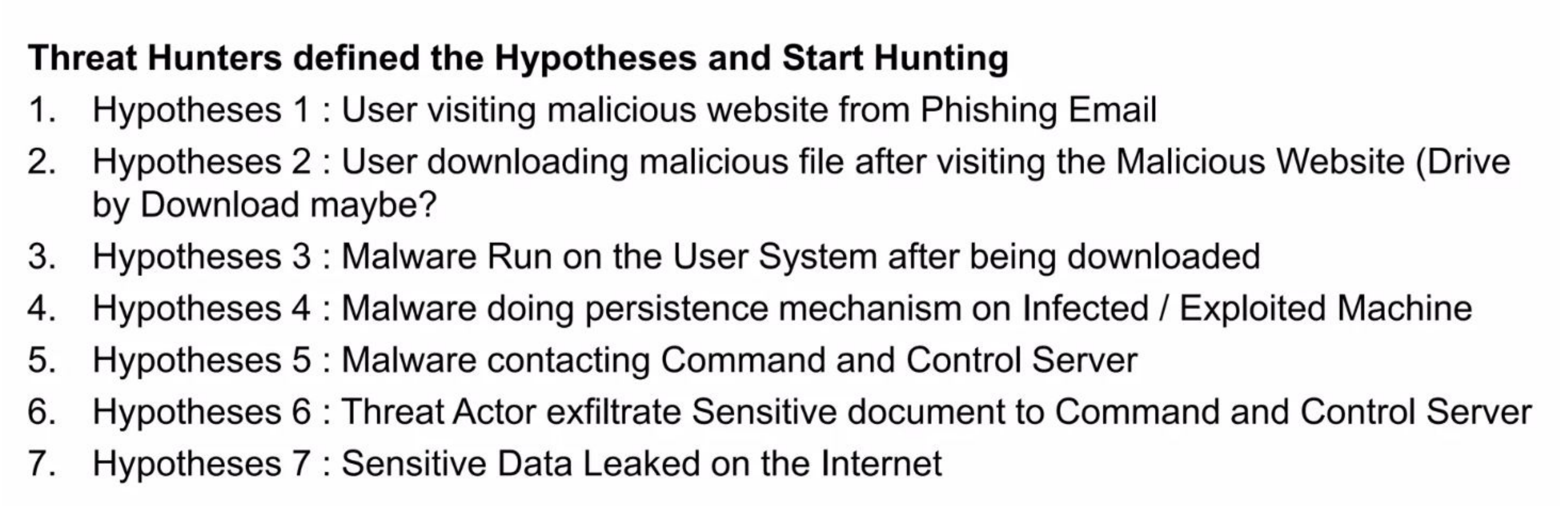
Baseline based hunting

- Detect something haven’t seen before based on baseline data
- Needs larger set of data for creating the baseline
- Triggers lot of false positives
- Quite effective to spot changed in your environment

Anomaly based hunting

- Going through the log data available to spot irregularities that might be malicious
- Applying patterns on the data
- Quite useful in fraud detection

Examples of Hypothesis for Threat Hunting



Threat Hunters defined the Hypotheses and Start Hunting

1. Hypotheses 1 : User visiting malicious website from Phishing Email
2. Hypotheses 2 : User downloading malicious file after visiting the Malicious Website (Drive by Download maybe?
3. Hypotheses 3 : Malware Run on the User System after being downloaded
4. Hypotheses 4 : Malware doing persistence mechanism on Infected / Exploited Machine
5. Hypotheses 5 : Malware contacting Command and Control Server
6. Hypotheses 6 : Threat Actor exfiltrate Sensitive document to Command and Control Server
7. Hypotheses 7 : Sensitive Data Leaked on the Internet

Hypothesis 1: USer visiting malicious website from phishing email

- Data source for Hunting:
  - Passive DNS log, DNS server logs, email logs, mail security gateway log
- Platform for hunting:
  - SIEM
  - ELK Stack
  - Data Science tools
- Analysis/Data Enrichment:
  - DNSTwist, Phishing Domain blacklist (Phishtank), TI Feeds, URL/Domain sandbox analysis

Thank you

Ref: https://www.slideshare.net/digitoktavianto/cyber-threat-hunting-workshop-239366830