

Practical Cyber Security for Cyber Security Practitioners

Prof. Sandeep Kumar Shukla

Department of Computer Science and Engineering

Indian Institute of Technology, Kanpur

Lecture 09

Mapping to ATT & CK from RAW Data

So for the next three classes, I'll be taking the class and we'll be discussing ATT&CK Navigator and how we can extract TTPs, tactics, techniques, and procedures from the raw data. And along with that, how we can store and analyze those extracted TTPs. and how it is being done in the real world, we'll be seeing in the next three classes. Along with that, we'll be also seeing how analysts can make a defensive recommendation to the organization for which they are performing and extracting threat intel data. So in this class, as some of a few of the slides are already covered, I'll be going quickly with those slides to make a revision on how we can attack TTPs from raw data. So, so far we have been working with Intel where the activities were already listed in the threat reports, finished threat reports, which we did in the last class.



Mapping to ATT&CK from Raw Data



- **So far, working from intel where activity has already been analyzed**
- **Analysis of techniques/behaviors directly from source data**
 - Likely more information available at the procedure level
 - Not reinterpreting another analyst's prose
 - Greater knowledge/expertise required to interpret intent/tactic
- **Broad set of possible data can contain behaviors**
 - Shell commands, malware, forensic disk images, packets

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

So how those TTPs were being made by the, like the threat reports must be written by someone, like some engineer or some analyst who must have extracted these TTPs from the raw data, like the Windows events, logs, and the commands and other things as evidence we extract from the attack infrastructure. So, rather than extracting TTPs from the already written and already extracted TTPs, we will be looking into how investigators

and analysts see the raw data in terms of log files or commands and how they process that information and map that information to the TTP in the attack navigator. So this data mostly comes from the various data sources which are present in the victim infrastructure. And analyzing those behaviors and mapping to the TTPs may likely be focused more towards process level.

Like the logs, the place in the logs where something is happening related to the malicious or the malicious behavior. Or if there is something in the commands, if it is performing some action which is something like collecting the data or sending data to somewhere else or communicating on some other IP. So such kinds of behaviors and the procedures we first analyze extract what all such kinds of mini procedures available in the source data and then we process the information. Also, whatever we are doing from the finished report, that was already processed and analyzed by analysts and presented in front of us. But here, in that case, this TTP is kind of biased, like the people who are interpreting the data based on their understanding, their expertise.

Sometimes it adds to the bias. In such a case, we are supposed to do this analysis all together with multiple cyber threat analysts. So till now we were not doing, not interpreting the data from multiple perspectives, rather there was a report where all perspectives clubbed all together as a conclusion and we were extracting data. So in raw data there is a need to collaborate between the analysts in which they will be extracting TTP based on their perspective and the expertise and then they will collaborate and they will discuss and then they can finalize to some set of TTPs which they have observed, commonly observed. Also, to do these things, there is great knowledge required with respect to how we interpret the tactics, the intent.

So one command may represent encryption or the sending data to some other C2 server. So that can be interpreted with the, if there are some people who are from the, other, not exactly the network security guy, and there is a packet capture, and there is a packet behavior where people are, analysts are supposed to extract the TTPs from the packet capture. So a malware guy, a network guy, or a cyber risk guy, they may have a different way to look into it, and there's a high chance of that if there is less expertise, then they can miss some of the key information, which network guy can do. can get the details. So, all these things we need to take care of while mapping TTPs from raw to threat intel.

So, to see where to look into the raw data, there is a broad set of possible data which can contain the behavior. Some of the examples are shell commands, malware and the forensic disk images which analysts usually investigate and the network patterns. So as we discussed in the last class, the steps to map the TTPs from any data, either it's a threat report or raw data, are some six sets of steps. First step is understanding the attack, which

we already did. All these steps we followed for the threat report, and now we'll follow the same way to extract TTPs from raw data.



Process of Mapping to ATT&CK



1. Understand ATT&CK
2. Find the behavior
3. Research the behavior
4. Translate the behavior into a tactic
5. Figure out what technique applies to the behavior
6. Compare your results to other analysts

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

So, we look for the behavior in the raw data where the behavior lies and then we can research about the behavior because the analyst may not have the all broad information. So, one can do research on it and then once we understand what exactly that behavior contained in that behavior do and then we can translate that behavior into the tactics like which tactic is mapped to. And then once we finalize the tactic, within that tactics, which techniques, it is being applied in that behavior. And then followed by we can compare results with our other colleagues or other analysts. So this set of commands, sir, was discussed in the last class.



1. Find the Behavior



`ipconfig /all`

`sc.exe \\In334656-pc create`

`.\recycler.exe a -hpfGzq5yKw C:\$Recycle.Bin\old
C:\$Recycle.Bin\Shockwave_network.vsd`

Commands captured by Sysmon being run interactively via cmd.exe

`10.2.13.44:32123 -> 128.29.32.4:443`

`128.29.32.4:443 -> 10.2.13.44:32123`

Flows from malware in a sandbox

`HKLM\Software\Microsoft\Windows\CurrentVersion\Run`

`HKLM\Software\Microsoft\Netsh`

New reg keys during an incident

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

So it starts with IP config all, first command, where one may try to see the network details, like IP and all. Then the next command was SC command, which is used to create and operate with the services, Windows services. Then there is a unknown, very unknown command, which you may not be much familiar with, what exactly this recycle, since recycle to that BSDX, this long command. This one, so this is by just seeing this, there is a very less sense we are getting what exactly it might be doing. So we'll be investigating these commands to see how and what is being done while running this command.

So these commands were captured from just a small. So this information we'll be using to map to the TTPs. Followed by that, there is a behavior captured while performing malware analysis from the sandbox. Like this is a kind of IP communicating to some different IP and the communication is being returned from the same, on the other side. And following that, there is a new registry key introduction in the run registry and Netsh, which we'll be seeing how we can leverage this set of raw data to the TTP.



ipconfig /all

```
Command Prompt
C:\Users\Sandeep K. Shukla>ipconfig /all

Windows IP Configuration

Host Name . . . . . : Sandeep
Primary Dns Suffix . . . . . :
Node Type . . . . . : Hybrid
IP Routing Enabled. . . . . : No
WINS Proxy Enabled. . . . . : No

Ethernet adapter vEthernet (Default Switch):

Connection-specific DNS Suffix . . :
Description . . . . . : Hyper-V Virtual Ethernet Adapter #2
Physical Address. . . . . : 00-15-5D-78-8E-40
Dhcp Enabled. . . . . : No
Autoconfiguration Enabled . . . . . : Yes
Link-local IPv6 Address . . . . . : fe80::e8c0:71b3:d85c:e00f%32(Preferred)
IPv4 Address. . . . . : 172.29.96.1(Preferred)
Subnet Mask . . . . . : 255.255.240.0
Default Gateway . . . . . :
Dhcpv6 IAID . . . . . : 536876381
Dhcpv6 Client DUID. . . . . : 00-01-00-01-29-02-06-5B-0C-37-96-33-D9-CF
NetBIOS over Tcpip. . . . . : Enabled

Ethernet adapter vEthernet (Wi-Fi):

Connection-specific DNS Suffix . . :
Description . . . . . : Hyper-V Virtual Ethernet Adapter #6
```

So this is just for the demonstration that the ipconfig /all, it leads the details about the IP configuration and the Ethernet details. So one can see the details related to the IP, subnet mask, and other details related to the network. So this is the help page for the SC command, which is used to create the service. So you can see there are different syntax examples where one can enumerate, display, and the other activities one can perform with the services, Windows services. So after seeing that this sc.

Command sc query



```

Command Prompt
this case. If the query command is followed by nothing or one of
the options listed below, the services are enumerated.
type= Type of services to enumerate (driver, service, userservice, all)
      (default = service)
state= State of services to enumerate (inactive, all)
      (default = active)
bufsize= The size (in bytes) of the enumeration buffer
        (default = 4096)
ri= The resume index number at which to begin the enumeration
    (default = 0)
group= Service group to enumerate
      (default = all groups)

SYNTAX EXAMPLES
c query - Enumerates status for active services & drivers
c query eventlog - Displays status for the eventlog service
c queryex eventlog - Displays extended status for the eventlog service
c query type= driver - Enumerates only active drivers
c query type= service - Enumerates only Win32 services
c query state= all - Enumerates all services & drivers
c query bufsize= 50 - Enumerates with a 50 byte buffer
c query ri= 14 - Enumerates with resume index = 14
c queryex group= "" - Enumerates active services not in a group
c query type= interact - Enumerates all interactive services
c query type= driver group= NDIS - Enumerates all NDIS drivers

C:\Users\Sandeep K. Shukla>

```

exe, ln, this -pc create, so this is by seeing this we can understand that there are some services being created with this name, so on this server. So you can see an example of this, so this is kind of not a correct server address, so it faulted as an error, but it says that it is used to create the service in the register service database. So one can do such a type of research on the run command and one can understand exactly what this extracted behavior is doing. So this, researching the behavior, we were doing same in the threat, while mapping to the threat report to the TTP, we were doing the same, like once we understand where that any process, once we understand any sentence or any paragraph or something is being done, is being represented, then we were researching about that, if we are not aware of. So it is similar to what we did previously.

Command sc <server> create



```

Command Prompt
Link-local IPv6 Address . . . . . : fe80::45c7:c497:dfe6:8f71%61
IPv4 Address. . . . . : 172.20.64.1
Subnet Mask . . . . . : 255.255.240.0
Default Gateway . . . . . :

C:\Users\Sandeep K. Shukla>sc \SANDHEEP create
DESCRIPTION:
    Creates a service entry in the registry and Service Database.
USAGE:
    sc <server> create [service name] [binPath= ] <option1> <option2>...
OPTIONS:
NOTE: The option name includes the equal sign.
    A space is required between the equal sign and the value.
type= <own|share|interact|kernel|filesys|rec|userown|usershare>
      (default = own)
start= <boot|system|auto|demand|disabled|delayed-auto>
      (default = demand)
error= <normal|severe|critical|ignore>
      (default = normal)
binPath= <BinaryPathName to the .exe file>
group= <LoadOrderGroup>
tag= <yes|no>
depend= <Dependencies(separated by / (forward slash))>
obj= <AccountName|ObjectName>
      (default = LocalSystem)
DisplayName= <display name>
password= <password>

C:\Users\Sandeep K. Shukla>

```

2. Research the Behavior

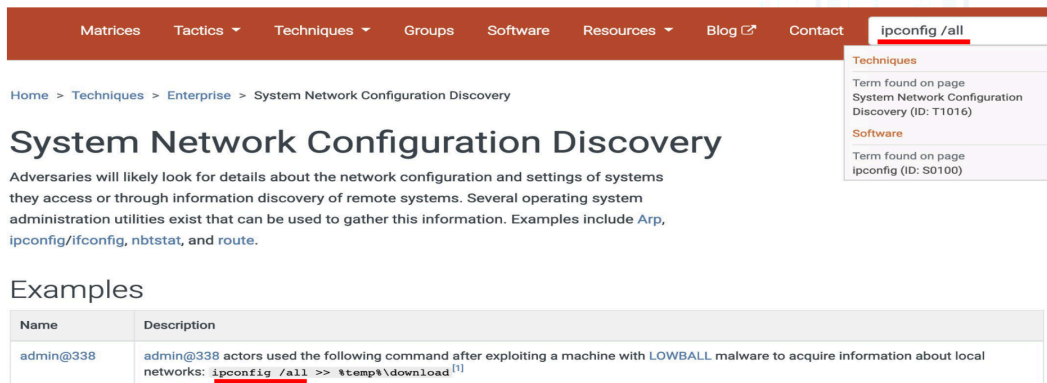
- Can be similar to analysis of finished reporting for raw data
- May require expertise in the specific data type
 - Network, forensics, malware, Windows cmd line, etc
- May require multiple data sources, more context
 - Additional questions to responders/analysts



© 2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

And then one may require to understand the terminology and to help you in researching the behavior. One may require expertise in the specific kind of data types, which can be a network analysis or forensic or malware and other like CMD line tools. Also, one may need to club all data collected from the data sources and to get the more context, like what exactly, how data collected from the multiple sources are correlated so that you can understand the flow and understand how attack flowed during the attack. So to research the behavior, once you have `ipconfig /all`, when we go directly to this MITRE web page, there is a search bar where one can search and see the things, but sometimes these things cannot be as straightforward. So for the first command, this is too easy.

2. Research the Behavior



The screenshot shows the MITRE ATT&CK web interface. The navigation bar includes Matrices, Tactics, Techniques, Groups, Software, Resources, Blog, and Contact. A search bar contains the text "ipconfig /all". Below the search bar, the breadcrumb path is "Home > Techniques > Enterprise > System Network Configuration Discovery". The main heading is "System Network Configuration Discovery". The text below the heading states: "Adversaries will likely look for details about the network configuration and settings of systems they access or through information discovery of remote systems. Several operating system administration utilities exist that can be used to gather this information. Examples include `Arp`, `ipconfig/ifconfig`, `nbtstat`, and `route`." Below this text is a section titled "Examples" with a table:

Name	Description
admin@338	admin@338 actors used the following command after exploiting a machine with LOWBALL malware to acquire information about local networks: <code>ipconfig /all >> %temp%\download^[1]</code>

© 2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

One can go search and there is a technique named system network configuration discovery. And all these techniques have a unique ID. So you can see for this technique, there is a T1016. So, for some of them, you can say, for some of the commands or some of the raw data, you can get straightforward results or mapping to the technique and you can map the technique and go ahead. But for a little of the raw data, one needs to do research and understand, break the whole information collected piecewise to understand what exactly it is doing.



2. Research the Behavior



```
.\recycler.exe a -hpfGzq5yKw C:\$Recycle.Bin\old  
C:\$Recycle.Bin\Shockwave_network.vsd
```

- Can make some educated guesses, but not enough context

File analysis:

When recycler.exe is executed, it gives the following output:

```
C:\recycler.exe  
RAR 3.70 Copyright (c) 1993-2007 Alexander 22 May 2007  
Roshal Shareware versionType RAR -? for  
help
```

- Aha! Based on the analysis we can Google the flags to RAR and determine that it is being used to compress and encrypt the file

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved
for public release. Distribution unlimited 19-01075-15.

So this is more likely the third command which we saw, that recycler one. So after seeing this, the whole command, we can only see that there is something being, there is a two input and output, output is being given to this command recycler.exe, but what exactly it is doing, we have no idea, right now at least. So we have to make some educated guesses, but still we are not getting much context of the behavior. So we can run this recycle.

exe in the sandbox environment to see what exactly it does. If you run this recycle.exe, this is just an assumption case. So if you run, we can see that this output is like this, which represents that this is something related to the RAR, okay? Like compression. So we can guess an idea that there is something compressing or some obfuscation or encryption is being done, which is not clear yet.

But once we have, we can see here, there is one more switch, there is this flag and the A, which might be representing as an argument for this recycler. And then this flag might be representing what exactly we're supposed to do. And this can be output, this can be input. This is mostly a general behavior we give a command. So after seeing this, this -hp in a

flag, it represents the completion or encryption behavior.



2. Research the Behavior



```
.\recycler.exe a -hpfGzq5yKw C:\$Recycle.Bin\old  
C:\$Recycle.Bin\Shockwave_network.vsd
```



vsd



People also ask

What can open a VSDX file?

A **VSDX file** is a drawing saved in the **VSDX file** format introduced with Visio 2013, a program used for making drawings and technical illustrations.

And the file being compressed/encrypted is a Visio diagram, probably exfiltration

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

So one can guess that this Recycler.bin is old, this is the destination path where this file, this VSDX file is being encrypted or compressed and being placed here. So which is, we are not sure, but we just guessed it. But when we go forward, what is this VSDX? Look, we resolve this, there is this. We guess this is the destination and the source and destination path.

But what exactly is this VSDX? So one can do research on this VSDX on the internet, and we got to know that this VSDX is used, it's a Visio diagram, which is being used as an input in this command. So there is a kind of more contextual guess now that there is some Visio diagram file which is being complexed or being encrypted and probably when it is being done, like why the attacker will encrypt or compress the file on the victim system. This is kind of used out. So this is mostly done while collecting the data so that data size can be less and can get evaded in the network phase while communicating to the CNC server. So now we have a little bit more educated guess that this is a kind of stage where an attacker collects the data and does encryption or compresses it to send it to the CNC server.

3. Translate the Behavior into a Tactic



`ipconfig /all`

- Specific procedure only mapped to System Network Configuration Discovery
- System Network Configuration Discovery -> **Discovery** ✓
- Seen being run via Sysmon -> **Execution**

`.\recycler.exe a -hpfGzq5yKw C:\$Recycle.Bin\old
C:\$Recycle.Bin\Shockwave_network.vsdX`

- We figured out researching this that “**vsdx**” is Visio data
- Moderate confidence **Exfiltration**, commands around this could make clearer
- Seen being run via Sysmon -> **Execution**

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved
for public release. Distribution unlimited 19-01075-15.

So now we'll club together whatever we analyzed for all three commands which we saw. So ipconfig all, this is the map directly to the system network configuration discovery. So the tactics which are being used here are discovery and technique. Also, we noticed that in this command, all three commands were extracted from Sysmon. So, this is one that also represents the execution behavior like these commands are being executed in the victim machine.

4. Figure Out What Technique Applies



- **Similar to working with finished reporting we may jump straight here**
 - Procedure may map directly to Technique/Tactic
 - May have enough experience to compress steps

`ipconfig /all`

- Specific procedure in **System Network Configuration Discovery (T1016)**
- Also **Command-Line Interface (T1059)**

`.\recycler.exe a -hpfGzq5yKw C:\$Recycle.Bin\old
C:\$Recycle.Bin\Shockwave_network.vsdX`

- We figured out researching this that “**a-hp**” compresses/encrypts
- Appears to be **Data Compressed (T1002)** and **Data Encrypted (T1022)**
- Also **Command-Line Interface (T1059)**

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved
for public release. Distribution unlimited 19-01075-15.

So, there are two tactics which we discovered from this command: discovery and execution. Further, this recycler command after figuring out this VSDK is being compressed using this recycler, we can connect it with the exfiltration phase, which is a kind of strong evidence that exfiltration might be being done. And also the sysmon as we extracted from the sysmon events, so this is related to the execution tactic, okay? So the fourth stage is now to figure out what techniques apply. For some of them, we already saw that. So this is also similar to the threat reports mapping.

We have to go to the attack database, knowledge base, and once we understand the tactics or the behavior, we have to look into the techniques, which techniques have been used, which is kind of starting again for the ipconfig /all. Plus, all these commands are being run on the command line. So there is one more technique named as command line interface. If things are being run on the command line, using the command line, this is a technique that represents that behavior. So this technique and this technique finally got mapped to this.

And this command line interface comes under the execution tactic. Further, the next, this recycler command, as we figure out that this -hp and command a is representing to send the argument and denote it to perform the compression or encryption. This is appearing to be a data compressed phase. This is a technique present in the collection tactic. And or it can be data encrypted which is not cleared and further this is again aligned with the command line interface T1059 okay.



4. Concurrent Techniques



- **Don't just think of what's happening – think of *how* it's happening**
- **Certain tactics commonly have concurrent techniques:**
 - Execution
 - Defense Evasion
 - Collection
- **Examples:**
 - Data Compressed + Data Encrypted (2x Exfiltration)
 - Spearphishing Attachment + User Execution (Initial Access + Execution)
 - Data from Local System + Email Collection (2x Collection)
 - Process Discovery + Command-Line Interface (Discovery + Execution)

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

So now we understand that we extracted some of the techniques for the first few of the commands which we saw. Now we have to see these techniques. Now how this technique is happening. To understand that we have to, the techniques can occur concurrently. So

it's like we saw that execution and the like if ipconfig /all.

So execution and discovery is happening. Similarly, the compression, if they're performing two techniques together, something like you can see, one can receive the spear phishing attachment once the user executes that or clicks or opens that. So that will come under the two tactics, initial access and execution, and there will be two techniques, spear phishing attachment and the user execution. So the two techniques can occur all together. So once we understand the one, we can look into the other aspect of the behavior, whether we can extract multiple techniques from the single behavior or not. So there can be chances that two multiple techniques can exist in a single behavior.



4. Different Types of Techniques



- **Not all techniques are created equal!**
 - Credit to Red Canary: <https://www.redcanary.com/blog/avoiding-common-attack-pitfalls/>
- **Some are specific**
 - Rundll32
 - Netsh Helper DLL
- **Some are broad**
 - Scripting
 - Obfuscated Files or Information
- **Some capture “how” the behavior occurs**
 - Masquerading
 - Data Transfer Size Limits
 - Automated Collection

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

Similarly, data from the local system can be collected and the email can be collected altogether, which comes into the collection phase. And as we saw in the, no, we didn't see, but discovery and command interface, both of this can occur altogether with the two different tactics, discovery and the execution. Okay, so now in the MITRE knowledge base, we have total approximately if I remember correctly 190 plus techniques divided over all over the tactics which you saw, but all techniques and all TTPs are not equally, they don't have equal weightage in any attack phases. So there are some techniques like the command line interface which is mostly being used by all adversaries. So that is being mostly used TTPs out of all TTPs we have.

Also, the TTPs can be straightforward, like there are some TTPs. If rundll is being used to execute any DLL file, so they are directly mapping to take me to rundll. There's also Netsh, if any DLL is trying to come, trying to do something with the network details, then this netsh Helper DLL is being executed. So this too is an example of

straightforward mapping. If you're seeing anywhere rundll in your behavior, that can be directly mapped to the technique rundll.

And the similar goes for the netsh DLL. But some techniques can have a broader aspect like scripting. Scripting can be, in the scripting technique, one can do anything, like one can use PowerShell, one can use other scripting like Python or to develop the executable or payload. So, scripting is kind of broad, which is again bragged into the sub-techniques, which we'll see in the knowledge-based detail, probably in tomorrow's class or today's class. These techniques are named as obfuscated files or information. So what it is representing is a file or information is being obfuscated, but how and what is being done exactly is not clear.

So these techniques, like one can do compression, one can encrypt, one can hide that in some hidden folder. or as a hidden item in the window machine. So this is still, it is not clear, not exactly specific how obfuscation is being done. So there are some techniques which have specific details and some are very broad. Also, There are some techniques which help us to understand how this behavior occurs.

So one can masquerade the, mostly malicious files masquerade themselves as a benign or legitimate process name. So mostly svchost in the windows. So one can masquerade themselves and this is showing how attackers are masquerading, like behaving. The behavior comes, there is nothing like any procedure is being happening. happening there, then it is just masquerading them with their name.



5. Compare Your Results to Other Analysts



- Same caveats about hedging biases
- May need a broader set of skills/experience to work with types of data

Analyst 1

- Packets
- Malware/Reversing
- Windows command line

Analyst 2

- Windows Events
- Disk forensics
- macOS/Linux

So we need to see also how any behavior, there are some techniques which represent that

not exactly something is being done rather than how it is being done. So this data transfer size limit and the automated collection, all these things come into this category. Now, as we discussed earlier, the way to look into the dataset and translate it, research it and map it to the TTP can have different results for the different analysts having different perspective, different expertise. So, when a threat analyst needs to get collaborated with each other and do collaboration to understand each other's perspective and finalize the set of TTPs or modus operandi observed during the attack. So in one example, an analyst may have more expertise related to analyzing the packets or reverse engineering or be more friendly with the Windows command line than the other analyst may have expertise related to the Windows event analysis or forensic or that analyst may be more comfortable with the Mac OS or Linux.

So in such cases, if you have a machine with the window, high chances that this analyst one can get more contextual and broader information of the behaviors and he may have more understanding of the mapping the behaviors to the TTP rather than if there is something related to forensic, analyst two may have better insights rather than analyst one. But also we cannot go with the one because any attack may have a clubbed information of all these things that needs to be analyzed. So, this is a general way which we follow in the real world where multiple analysts collaborate together to discuss and finalize the TTPs. So, why do we do this? So, basically there is a term of profiling the adversaries. So, this TTP majorly represents the modus operandi of the attacker like how attackers usually behave to achieve some target.

So, assume there is attacker A and B and attacker A is mostly trying to get initial access into the victim machine, attacker B at the same time is trying to get access to the victim machine. So, the behavior of the attacker A is can be if we have profiled in the by analyzing the past attacks. So, it can be seen that the attacker A is trying to send this spear phishing email to the victim and they try to get into the victim's infrastructure. Either user clicks on the send a spear phished email. So why we are doing it, why we are extracting these TTPs from the raw data, how it is going to be helpful in the real world.

We took a case for attacker A and B where attacker A generally follows the spear phishing techniques to phase the user or the victim into the network or system. Where the attacker B is more sophisticated and they develop a zero-day vulnerability or they exploit any public facing applications of the victim and they get into, they invade that and then they get into the network or system. So this can be two different behaviors which are being used by two different adversaries based on their expertise and their knowledge. This helps us to segregate the behavior or the modus operandi of the attack groups or the threat groups where we can see a set of TTPs which one attacker group follows usually and what other attacker group is being followed in the past years.

But obviously this can be changed. This can be changed, one can follow, one can mimic the behavior of others because these TTPs and behavior is all publicly available. One can follow other groups' behavior to mask their identity. This is mainly helping in profiling the threat groups identity or the profiling their behavior. So, let us understand what are the pros and cons of mapping TTPs from the raw data and the finished threat report ok.

Pros/cons of Mapping from the Two Different Sources



Step	Raw	Finished
Find the behavior	Nearly everything may be a behavior (not all ATT&CK)	May be buried amongst prose, IOCs, etc
Research the behavior	May need to look at multiple sources, data types. May also be a known procedure	May have more info/context, may also have lost detail in writing
Translate the behavior into a tactic	Have to map to adversary intent, need domain knowledge/expertise	Often intent has been postulated by report author
Figure out what technique applies to the behavior	May have a procedure that maps straight to technique, or may require deep understanding to understand how accomplished	May be as simple as a text match to description/procedure, or may be too vague to tell
Compare your results to other analysts	May need multiple analysts to cover all data sources	More likely in a form where other analysts needed for coverage/hedge against bias

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

MITRE

So, the first stage was finding the behavior. For raw data, we were nearly exactly to the behavior which is being demonstrated in the attacker victim infrastructure. But in the finished report, it may be buried amongst pros and IOCs. So attackers, in the threat reports, we directly have a list of IOCs which is being listed in the threat reports generally. And things are directly already analyzed and placed in front of us. In the research behavior stage, for the raw data mapping case, one needs to collect the data from all multiple data sources and different data types.

Also, this can be helpful to first understand the procedure before going to the tactics or techniques. Whereas in the finished threat reports, we may have more information and the context has already been analyzed and described after the threat analysis. More information and the contextual information in the threat report rather than the raw data. Okay, but this information whatever is present in the threat report is totally based on the understanding of the analyst who has written the report and performed the analysis. So there are chances that there may be a loss of data while writing the threat report rather than in raw data we will be having all possible behavior and the evidence in front of us and we can perform analysis based on our expertise.

While translating the behavior into the tactics, We have to map to the adversary intent and for that one needs to have a domain knowledge or expertise rather than for finished threat reports, often the intent has been already processed and postulated and by the report author and has been placed directly in front of us. Now while figuring out the techniques, which techniques have been used in the attack, For the raw data, we may have a processor which is directly straightforward as we saw for IP config or one may require a deep understanding of the domain knowledge from where we are mapping the TTPs and understand how it is being accomplished. So for this one needs to have a deep understanding of the domain rather than in the finished report. If you want to map to the TTP, There is a domain with the help of the NLP, Natural Language Processing. One may process the text, understand what exactly is being explained in the sentences of the threat reports or the description, and one can use an NLP model to understand the behavior from the sentences of the threat reports, and one can develop a model where we can map directly from the threat report sentences to the TTP. So, this can also be a kind of complex thing which one needs to understand, which one needs to train a model, an NLP model where one needs to let the model understand the contextual information of the sentences and the paragraphs.

And then this text match also may help in the mapping to the sentences to the TTP. Now, for comparing results with another analyst, one may need multiple analysts to cover all kinds of different data sources present at the victim ends. In the finished threat report case, it is more likely in the form of that other analyst already needed for the coverage to hedge against the bias, means to restrict the bias An analyst are need to having a more kind of diverse coverage and understanding of the domain so that they can understand the sentences and written in the threat report and map it to the attack navigator.



Exercise Working with raw data



- You're going to be examining two tickets from a simulated incident
- Ticket 473822
 - Series of commands interactively executed via cmd.exe on an end system
- Ticket 473845
 - Pieces of a malware analysis of the primary RAT used in the incident
- Both tickets are at <https://attack.mitre.org/training/cti> under Exercise 3

- Use whatever to record your results or download and edit
- Identify as many behaviors as possible
- Annotate the behaviors that are ATT&CK techniques

So now there is an exercise where we will be directly, we'll be seeing in the further slides,

but it is expected that students can go to this link and access the exercise and do it after the class. So there is an exercise at this link under the exercise three, where there are two tickets, 473 at double two and 473 at four five, which has a series of commands and piece of malware analysis, primarily a remote access chosen.

And you have to see the series of commands present in this ticket, and you have to map those commands to the TTPs present in the MITRE ATT&CK knowledge base. So these commands are mostly executed by cmd.exe, which is kind of pre-assumable. And the other piece, ticket 473845, is a piece of malware analysis of a remote access trojan, which has been used in some past attacks. So, we have to go to this website, access the stored tickets and analyze and map the commands to the TTP and the pieces of whatever malware information there is listed in that ticket, map that to the MITRE ATT&CK TTP.

Okay, you have to record your results and download and edit. Then you have to identify as many behaviors as possible from the commands on the malware analysis behavior. And you have to annotate those behaviors on the MITRE ATT&CK techniques on the document. Further, we'll also see once we have this set of TTP, how we are going to store and analyze in the next few slides. While performing this exercise, you have to understand that if you are an analyst, what you are supposed to ask, what has been asked to the incident responder to get that behavior. Whatever behavior is present in the ticket, you have to understand that.



Exercise Questions



- What questions would you have asked of your incident responders?
- What was easier/harder than working with finished reporting?
- What other types of data do you commonly encounter with behaviors?
- Did you notice any behaviors that you couldn't find a technique for?



©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

Then what was easier or harder for working? If you are working with the raw data and you're working with the finished threat reports, you can feel a difference between how easy and how hard it is to map a TTP from raw data and the finished threat reports. Then

you can also look into what type of data you commonly encounter with the behaviors. So once we do this exercise multiple times, we can have a TTP on that fingertip. We can understand which kind of TTP is being used mostly by the attackers or which threat group. Also while exercising, did you notice any behavior that you could not find for the technique form? So there can be some behavior for which there is no tactic listed in the MITRE ATT&CK knowledge base.

So it is kind of interesting to see, can we find some behavior which is not mapped to any TTP? So, this MITRE technology has been adopted continuously since the last few years. So, each year they add different techniques and they club multiple techniques in one category. So, some behavior present in the current situation for which there is no TTP at all. If one can find such TTP or such behavior, MITRE has one channel to report such techniques or behavior to them and they may list out, if they analyze that and if it is found true behavior which they might be missing till now, so they may add that to the ATT&CK navigator or knowledge base by giving credit to you obviously. Okay, so we'll directly see the result of the Ticket 473822, which contains the set of commands.

Going Over the Exercise (Ticket 473822)



```
ipconfig /all System Network Configuration Discovery (T1016)
arp -a System Network Configuration Discovery (T1016)
echo %USERDOMAIN%\%USERNAME% System Owner / User Discovery (T1033)
tasklist /v Process Discovery (T1057)
sc query System Service Discovery (T1007)
systeminfo System Information Discovery (T1082)
net group "Domain Admins" /domain Permission Groups Discovery (T1069)
net user /domain Account Discovery (T1087)
net group "Domain Controllers" /domain Remote System Discovery (T1018)
netsh advfirewall show all System Network Configuration Discovery (T1016)
netstat -ano System Network Connections Discovery (T1049)
All are Execution - Command-Line Interface (T1059)
```

Discovery

© 2019 The MITRE Corporation. ALL RIGHTS RESERVED. Approved for public release distribution on unlimited terms.

So this is the exercise you are supposed to do after the class, but I'll just demonstrate and explain the things to you, having a better understanding of mapping. So as we did earlier, ipconfig all, it is directly having a technique T1016. Next arp -a, so this can be sent to translate the IP and find the system in the network. So this can be a system network configuration discovery.

So then further there is "echo user domain and username". So probably it is trying to print the user domain and username of the machine. So this can map to this system owner or user discovery, which has a technicality T1033, then "tasklist /v", which is mostly

listing the running process in the machine. So this is gonna map to the process discovery T1057. Then there is one more command, "sc query".

This is kind of querying the service state or what exactly they're doing. So this can be listed out, this can come under the class of TTP system service discovery having ID T1007. Victim may run this command, "systeminfo", which has been listed in this ticket. This "systeminfo" gives the output about system information. So that can be mapped directly to information discovery, T1082. Then there is one more command, "net group "Domain Admins" /domain", which can be used to see the group's permission, whatever groups present on the victim machine.

So this can be mapped to the permission group's discovery, which is T1069. Then there is one more command "net user /domain" which can be used to see what all accounts present on the machine that can be mapped directly towards the account discovery. Also there is one command related to "net group "Domain Controllers" which can be used to see all the groups present on the remote domain controller where one can map that to the remote system discovery. Further, "netsh advfirewall show all" is to list out the kinds of firewall related details available implemented on the machine. So, this also again is related to the network configuration and it is being mapped to the system network configuration discovery. Okay, then the final one is "net stat -ano" which lists out all the network connections that are actively communicating to the victim machine.

So this can be listed out, this can be mapped directly to system network connection discovery. So you must have seen a common pattern that we are understanding what exactly this command is doing and based on our knowledge of ATT&CK knowledge base, we are finding where this lies in and we are directly mapping it. And also if you see all these techniques present in these tactics are related to discovery stage or discovery tactics in which an adversary is trying to list and collect and see information related to the victim machine and which is being mapped towards a discovery, performing discovery on the victim machine. Also as this all commands are being, will be either using a PowerShell or the CMD on the victim machine, so that will come under this tactics of execution and mostly it is a command line. Okay, so in the execution, it will get mapped to the T1059, which is used for running the commands line interface or the PowerShell commands.



Going Over Exercise 3 (Ticket 473845)



Command and Control - Data Encoding (T1132)

C2 protocol is base64
30 seconds requesting

Command and Control - Standard Application Layer Protocol (T1071)

UPLOAD file (upload a file server->client)

DOWNLOAD file (download a Command and Control - Remote File Copy (T1105)

SHELL command (runs a command Execution - Command-Line Interface (T1059)

PSHELL command (runs a command via power Execution - Powershell (T1086)

EXEC path (executes a PE at the Execution - Execution through API (T1106)

SLEEP n (skips n beacons)

10.1.1.1:24123 -> 129.83.44.12:443

Command and Control - Commonly Used Port (T1043)

129.83.44.12:443 -> 10.1.1.1:24123

Copy C:\winspool.exe -> C:\Windows\System Defense Evasion - Masquerading (T1036)

HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Run\winspool
REG_SZ "C:\Windows\System32\winspool.exe Persistence - Registry Run Keys (T1060)

for public release. Distribution unlimited 15-02075-13

MITRE

Similarly, the other, the next ticket 473845, for this, we have this set of informal behaviors present in the ticket. So we'll first see this C2 protocol. This C2 protocol is base 64. So here this C2 represents the command control where base 64 is a standard protocol used to encode for data encoding.

So attackers sometimes encode the data or to obfuscate their real information. So in this, we can map that data encoding present in the command and control tactics directly also, as this is the standard application layer protocol, one can use this, one can map this to the T1071. Further, there is information related to uploading the file from server to client like from victim machine to C2 server. So this can come under the remote file copy where T1105 or this can also happen from the attacker server to the victim machine that is the client. So that comes under this downloading the file when the victim advertiser downloads the file or malicious payloads hosted on the attacker's C2 server.

Then there is shell command where some shell command is being run in the environment. So this can be mapped to the command line interface again. Then there is a PSHELL and which can be again, which runs the command via PowerShell. So this PSHELL command can get mapped towards the PowerShell 1086 execution phase. Further, there is EXEC path which represents the path of any PE executable and which is being used using these API. So because of that, this can map execution through API techniques present in the execution tactics.

Now we can see this IP is being communicated to this and there is a reverse communication, but we may not get much detail from this communication rather only we have this port number 443, okay. So this is the standard port for the HTTP

communication and this is kind of non-standard port. So this port number can get mapped. There is one more technique. If the attacker is using some commonly used port number, that presents under the command and control.

If they're using any commonly used port to communicate with the command and control server, there is a technique. So we can directly map T1043 for this kind of behavior, okay? Now there is a copy command where even the spool is being renamed with some other thing which has been hidden here. So aware of this one, sometimes attackers masquerade and they rename their file with something else with some other extension.

So you may not be realizing that the extension with the .pdf file is actually an executable file. So such behavior represents masquerading and there is a technique for that in the defense evasion tactics. So this is T1036. Further, there is a registry information. If you can see, the modification is being done in this registry named as run, where this winspool, which they have renamed just now, added that into the run registry. So the run registry mostly contains the list of items which are supposed to run, which are supposed to get executed or pointed out once the machine is getting, once we are rebooting the machine.

So this represents that they are trying to make a persistent connection. where they are modifying the registry run key and there is a specific technique which is present in the persistent tactics. Okay. So you can see this registry is given a path to see this path which they have just now they renamed it. Okay, so till now we talked about commenting on the reports with the ATT&CK, how we can understand and map the TTPs from the threat report.



From Raw Data to Finished Reporting with ATT&CK



- We've talked about augmenting reports with ATT&CK and analyzing data with ATT&CK, possibly in parallel with analysis for reporting
- If you are creating reporting with ATT&CK techniques, we recommend keeping the techniques with the related procedures for context
 - Allows other analysts to examine the mapping for themselves
 - Allows much easier capture of how a technique was done

Also we discussed till now that one can use the raw data, analyzing data ATT&CK to mapping the TTP. Things can be done in parallel. So this mapping from information to the TTPs, it allows the other analysts to examine the mapping for themselves. Also it allows the easy capturing of how techniques are done, which we have seen while mapping the things. Okay, so in most of the threat reports, if you see, some of the threat reports from the prominent company already list the behavior in the threat report.



Finished Reporting Examples



During operation Tangerine Yellow, the actors used Pineapple RAT to execute 'ipconfig /all' via the Windows command shell².

1. Discovery – System Network Configuration Discovery (T1016)
2. Execution – Command-Line Interface (T1059)

System Network Configuration Discovery (T1016) and Command-Line Interface (T1059) - During operation Tangerine Yellow, the actors used Pineapple RAT to execute 'ipconfig /all' via the Windows command shell.

instead of

Appendix C – ATT&CK Techniques

- System Network Configuration Discovery
- Command-Line Interface
- Hardware Additions

©2019 The MITRE Corporation. ALL RIGHTS RESERVED Approved for public release. Distribution unlimited 19-01075-15.

So rather, they wrote the whole threat report explaining the behavior. In the last, they can give some appendix or list of techniques which has been seen in the whole attack. So this IP, if they have written some sentence like this during operation, Tangerine Yellow, the act is used to execute this IP config via Windows command cell. So directly it is representing these two techniques, which you already discussed.

So instead of writing this set of things, if I want to communicate. with another one. So one can list this set of techniques, like there is a network configuration discovery being done in the attack, there is a command line interface technique is being done, and there is some, if there is a hardware addition they can do. So this enables communication between the analysts. So rather than explaining the whole story in the sentences, one can give a set of TTPs that have been used in the attack, which can summarize the other analyst or other expert to understand how the attack flows. Any doubt? Okay, so this is the end of the module. So now we see a glimpse of the navigator tool which lists out all these techniques and we see how once we have a set of techniques, how we can use that to store an analysis purpose.

As I said, this is used to do profiling, threat profiling. So the set of TTPs which we extracted from the raw data or from the threat reports, one needs to process that to get an

understanding and visualization of the TTPs or the attack scenarios, okay. So we'll see the attack navigator now. So this is just a kind of basic introduction of attack navigator.

We will do the whole analysis of the TTPs in the next class. If anyone has any doubt, you can ask. No doubt? So you can go to attack.mitre.org. There's a metric section. So they have, we sir already discussed this, that we have three kinds of matrices for enterprise, mobile and ICS, and industrial control systems.

So on the enterprise metrics, we are more interested in the IT system. So we'll go with the enterprise metrics. And there is a tab named view attack navigator. This is the navigator, which is openly available. One can use it for their own purpose.

So we have a different section here. We'll discuss all these layers differently in the next class tomorrow. For now, I'll just show you a quick demo. So this is the whole navigator, you can see. These are the tactics listed one by one, all 14 tactics. And these are the techniques within each of the tactics. So, we will see how extracted techniques, we will be mapping here, we will be mapping multiple attacks, tactics all together, TTPs all together and perform analysis to understand what kind of all common techniques has been used and what all kind of different techniques has been used by two threat groups or even a single threat group of two different attacks.

So that will give us an understanding of how attackers are changing their behavior and how they are distinguishing from the other threat groups. Okay, so these all tactics list kind of, reconnaissance has the 10 techniques and the others. Also the sub-techniques, which I saw, this command and script interpreter, we saw this T1059 multiple times. So this has the difference, So, one can use this scripting and the command cell to perform the attack and within this there are multiple ways one can achieve, one can do, one can perform these techniques.

So, these are the kind of sub-techniques, you can see the power cell is T1059.001, similarly this python and the other command cell have 003 sub-techniques. So this is the interface. So we'll wrap up now, and if someone has any doubt, you can ask. Otherwise, in the next class, we'll see this navigator tool, how we can store and analyze, and how we can list out the defensive recommendation for the victim.

So the next class will be crucial because the second assignment, we'll be releasing the second assignment today. And the second assignment has, in the second assignment, you are expected to, you'll be giving a report for each group, and you are expected to extract the TTPs from the reports by reading the report based on your group's understanding. And once you have a list of TTP, you can map that to the navigator. You can give

contextual information, how you came up with that TTP, and then you are supposed to give the defensive recommendation for each of the attack methods which we have seen in the threat report. So, we will discuss how to give this defensive recommendation and how to map and perform analysis on the navigator tool and how to add contextual information in this tool in the next class. Okay.

MITRE ATT&CK Navigator: <https://mitre-attack.github.io/attack-navigator/>