

Automating reasoning with ATT&CK?

Jonathan M Spring, Rawan Al-Shaer

FloCon, Jan 8, 2020

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

Copyright 2019 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon®, CERT Coordination Center® and FloCon® are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM19-1296

Introduction

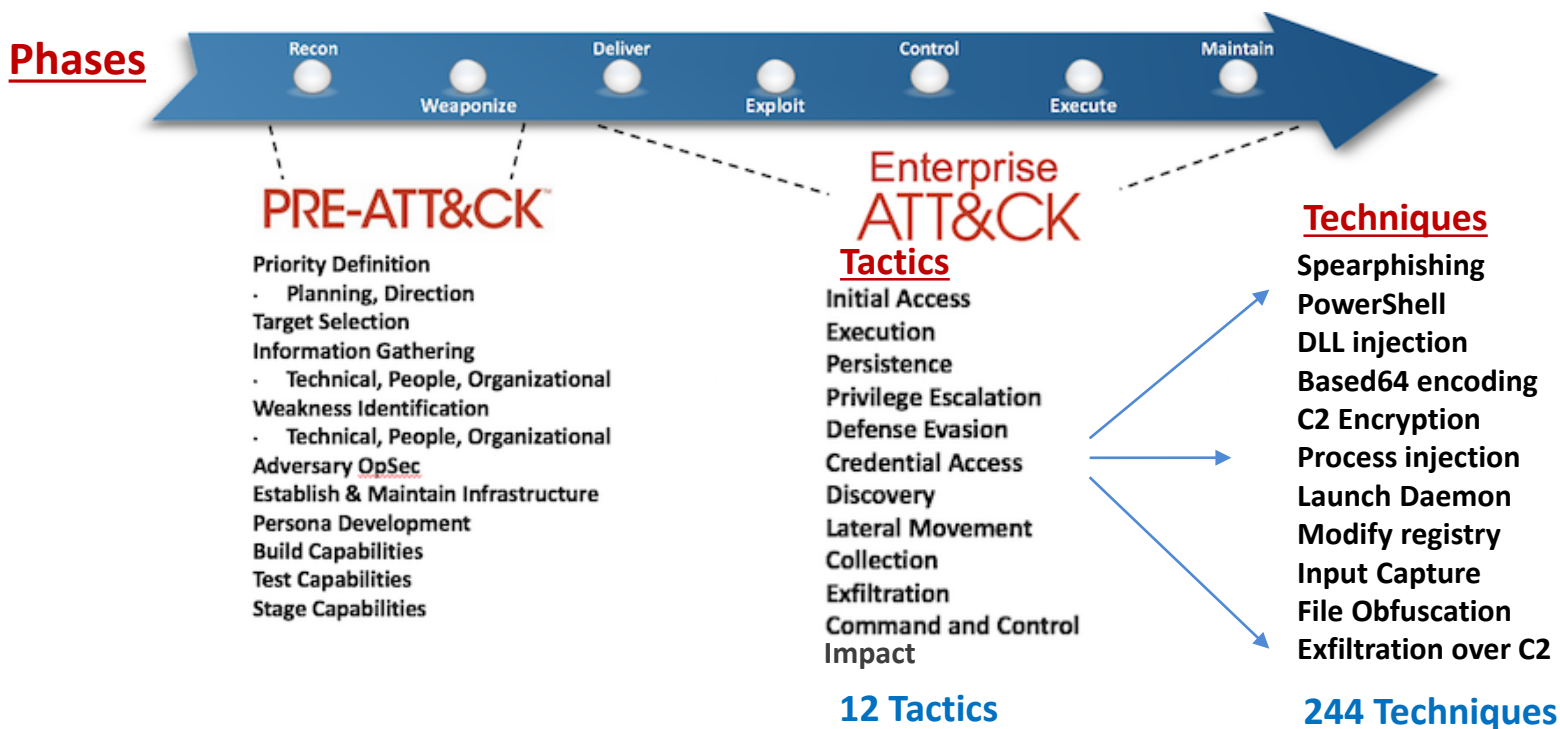
MITRE ATT&CK is made up of **TTPs** (Tactics, Techniques, Procedures)

- Tactics are general goals (e.g., initial access, exfiltration)
- Techniques are descriptions of adversarial actions that achieve tactical goals (e.g., Spearphishing Attachment, Modify Registry, Input Capture)
- The community is interested in using ATT&CK for detection, prediction, forensics, and threat hunting because it provides behavioral observables for detecting attacks

Our goal:

- Characterize ATT&CK's structure and usefulness for automated detection, etc., by analyzing their APT dataset

TTPs in MITRE ATT&CK Framework



Challenges for automated reasoning

MITRE ATT&CK TTPs are not *correlated* at the technique level

MITRE ATT&CK techniques are not *ordered temporally*

- A kill-chain ordered set of techniques would be, for example:

1. Account Discovery
2. [weaponization]
3. Spearphishing Attachment
4. User Execution
5. Bypass User Account Control
6. Automated Collection, Data Compressed
7. Exfiltration over C2 Channel

What do you mean uncorrelated?

Searched for meaningful correlations among techniques using:

- Partitioned Clustering
 - Finding the optimal K clusters
 - K means clustering
 - PAM clustering
 - Fuzzy Analysis clustering
 - Cluster Validation
- Hierarchical Clustering
 - Finding the optimal K clusters
 - Agglomerative clustering
 - Divisive clustering

(Paper on this due out soon)

Only the Agglomerative clustering produces some results

For intuition behind this result, consider clustering coefficient of APT data set

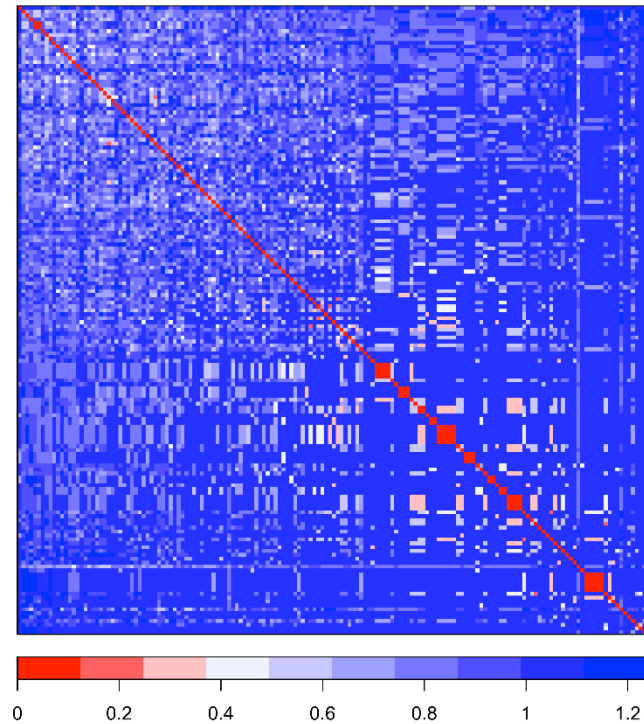
Hopkins Statistic: assess the clustering tendency of a dataset by measuring the probability that a given dataset is generated by a uniform distribution – tests the spatial randomness of the data

Interpretability:

- **$H = 0.5$** : The data set contains no meaningful clusters
- **$H \cong 1$** : The data set contains meaningful clusters
- **$H \cong 0$** : The data set is regularly spaced (neither clustered nor random)

Using Phi Coefficient : **$H = 0.6$**

Dissimilarity Plot for Techniques using Phi Coefficient



Attacks and Campaigns

ATT&CK merges the concept of attack with that of campaign

This is true even though it uses the kill chain as a semi-organizational concept

In the kill chain, an attack is a single exploitation attempt

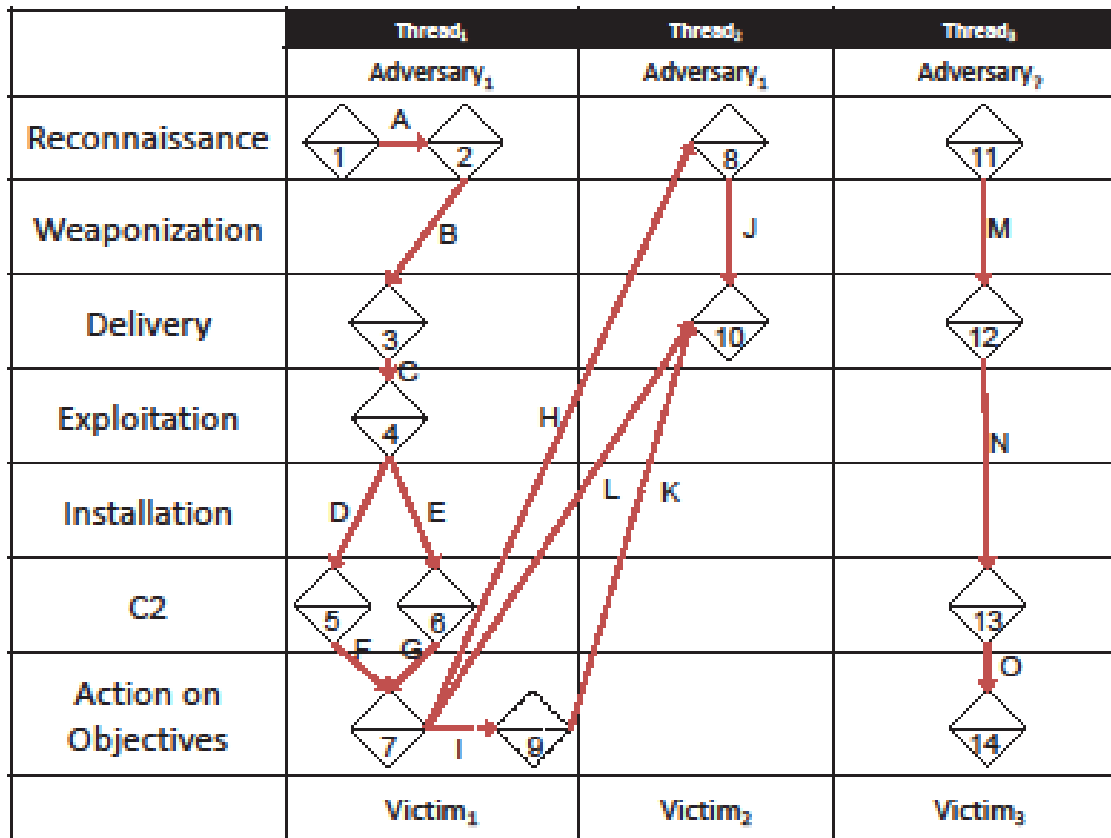
Campaigns are a series of planned or interrelated attacks

The kill chain and diamond model are not perfect, but they are useful mental models to organize general knowledge in security

- See Spring JM, Illari P. Building general knowledge of mechanisms in information security. *Philosophy & Technology*. 2018 Sep 17:1-33.

Diamond model and Campaigns

The diamond model for campaigns includes some things that ATT&CK does not



Sergio Caltagirone, Andrew Pendergast, Christopher Betz. The Diamond Model of Intrusion Analysis. 2014

Does it matter that ATT&CK is missing this temporal structure?

For automated reasoning, yes.

- (It would matter for other things, like real-time analysis, too)

At least, if you want to understand the relationship between MITRE's APT data sets and the techniques they use, temporal kill-chain structure helps

Sequential Pattern Mining

Sequential pattern mining looks for technique rules based on which techniques often showed a related temporal order

Sequential Pattern Discovery using Equivalence Classes:

1. Find the most frequent single length sequence
2. Observe the two-type temporal sequences (A occurs before B) and two-element item groupings (A and B appear together)
3. Based on the most frequent length-two outputs, find three-element sequences and three element item groupings
4. Continues until no longer finds frequent outputs

Confidence: likelihood that the sequence rule $A \rightarrow B$ occurs among transactions containing item set A, where item set A is before B

- Extracted **19** technique rules with confidence of 0.5 or higher
- * Beware base rate issues

Why should we care about related techniques?

To automate reasoning in incident analysis, it would help to know what adversary actions are most likely to look for, given what the analyst has seen already

ATT&CK could provide this, we're working on how to suggest such temporal structure (basically reintegrate the diamond model / kill chain)

If we had it clear and formalized, we could use it in formal, automated reasoning

- See Spring JM, Pym D. Towards Scientific Incident Response. International Conference on Decision and Game Theory for Security 2018 Oct 29 (pp. 398-417). Springer.

Automated reasoning with ATT&CK?

Not yet

Spring and Pym (2018) proposes a system for automated incident analysis, but to make it work it needs a corpus of what is likely given what has been seen

- ATT&CK is an obvious place to look for this
- Therefore we could make progress on automating incident and campaign analysis with some careful improvements

The target for automation is probably improving automated evidence collection and data discovery in a SIEM, so that an analyst can review **incidents**, and not *alerts*

One other barrier to automated reasoning with ATT&CK

Some techniques are subsets of others. For example:

- Scripting
- Powershell
- Bash scripting

There do not seem to be guidelines on when an analyst tags an intrusion with the more general or more specific option

The unified cyberspace ontology (UCO) tries to be a bit more rigorous about these relationships, but it does not have the same level of input from practitioners

Summary

The case studies captured in ATT&CK are valuable information for incident analysis

The ATT&CK structure is not currently amenable to automated reasoning

Two most important things to make it so:

- Restore the temporal relationships between the techniques (as in the Diamond Model)
- Make hierarchy or subset relationships between techniques explicit



Thanks! Questions?

[jspring __ cert.org](http://jspring__cert.org)