Investigating co-occurrences of MITRE ATT&CK Techniques

Md Rayhanur Rahman, Laurie Williams

***Abstract*—Cyberattacks use adversarial techniques to bypass system defenses, persist, and eventually breach systems. The MITRE ATT&CK framework catalogs a set of adversarial techniques and maps between adversaries and their used techniques and tactics. Understanding how adversaries deploy techniques in conjunction is pivotal for learning adversary behavior, hunting potential threats, and formulating a proactive defense. *The goal of this research is to aid cybersecurity practitioners and researchers in choosing detection and mitigation strategies through co-occurrence analysis of adversarial techniques reported in MITRE ATT&CK.* We collect the adversarial techniques of 115 cybercrime groups and 484 malware from the MITRE ATT&CK. We apply association rule mining and network analysis to investigate how adversarial techniques co-occur. We identify that adversaries pair *T1059: Command and scripting interface* and *T1105: Ingress tool transfer* techniques with a relatively large number of ATT&CK techniques. We also identify adversaries using the *T1082: System Information Discovery* technique to determine their next course of action. We observe adversaries deploy the highest number of techniques from the *TA0005: Defense evasion* and *TA0007: Discovery* tactics. Based on our findings on co-occurrence, we identify six detection, six mitigation strategies, and twelve adversary behaviors. We urge defenders to prioritize primarily the detection of *TA0007: Discovery* and mitigation of *TA0005: Defense evasion* techniques. Overall, this study approximates how adversaries leverage techniques based on publicly reported documents. We advocate organizations investigate adversarial techniques in their environment and make the findings available for a more precise and actionable understanding.**

I. Introduction

Information technology (IT) systems draw continuous attention from threat actors with financial motives [95] and organized backing [89]. In 2021, corporate businesses suffered 50% more cyberattacks per week compared to 2020 [19]. The incurred cost of cyberattacks also keeps rising. Cybersecurity Ventures forecasts that financial damage by cyberattack will be $6,000B in 2022 and will go high as $10,500B in the following three years [93]. Moreover, current cyberattacks are sophisticated and often consist of combinations of multiple adversarial techniques deployed by cybercrime groups and malware. For example, in Fig 1, we show actual attacks demonstrating how adversaries use multiple adversarial techniques. The figure shows that admin@338 [1], a cybercrime group, first uses the command line interface to collect information about existing users and then sends phishing emails to trick users into executing malicious files. Another cybercrime group, APT-C-36 [3] first obfuscates email attachments to bypass malware checks and then sends phishing emails to trick users into executing code via the terminal. Through code execution, attackers send the collected information back to a remote server via file transfer. Adversaries launching attacks through various adversarial techniques are often hard to detect, and thus organizations face difficulty in defending their systems [13].

Attackers can breach defense mechanisms through multiple attack vectors, and thus defending against attacks requires an understanding of how adversaries deploy techniques in combination to compromise defenses [16]. Hence, cybersecurity analysts identify specific adversarial techniques in cyberattacks and publish press articles and technical reports on the usage. The MITRE ATT&CK framework [9] maintains a catalog of the adversarial techniques used by cybercrime groups and in malware activities documented in these reports. The catalog contains information on what techniques adversaries used in cyberattacks and thus, reflects adversaries breaching systems from the aspect of adversarial techniques used in conjunction. Thus, the ATT&CK catalog enables researchers to analyze various aspects of cyberattacks through atomic adversarial behaviors based on the cataloged techniques. For example, in the literature, researchers have utilized the MITRE ATT&CK framework to perform cyberthreat attribution [82], [96], [108], malware profiling [87], [92], and tracking the provenance of attack indicators along with utilized/impacted resources [91].

Although MITRE ATT&CK enables researchers to correlate malware traces and intrusion alerts to adversarial techniques [84], [101], thwarting cyberattacks requires the understanding of the complex relationship among adversarial techniques. The pyramid of pain [85], a conceptual model of thwarting adversaries by responding to detected indicators, also emphasizes that responding to adversarial techniques is the most effective albeit the toughest way to prevent cyberattacks. Thus, the ATT&CK catalog of adversarial techniques used together in cybercrime groups and malware can aid practitioners in capturing the relationship among techniques. Practitioners can also look for evidence of other potential adversarial techniques in the victim environment. Practitioners can also identify to what extent security enforcement is performing and prioritize what ATT&CK techniques they should mitigate first to facilitate mitigating other ATT&CK techniques. Practitioners can devise proactive detection and mitigation strategies to improve security practices. *The goal of this research is to aid cybersecurity practitioners and researchers in choosing detection and mitigation strategies through co-occurrence analysis of adversarial techniques reported in MITRE ATT&CK.* We investigate the following research questions (RQs):

- **RQ1:** What are the top techniques used by adversaries? What are the top co-occurring techniques reported in adversary activities?
- **RQ2:** Given the occurrence of a technique, how can we predict what other associated techniques adversaries can use? What adversary behavior regarding technique usage do the co-occurrences indicate?

To answer the research questions, we collect adversarial technique usage data by cybercrime groups and malware from

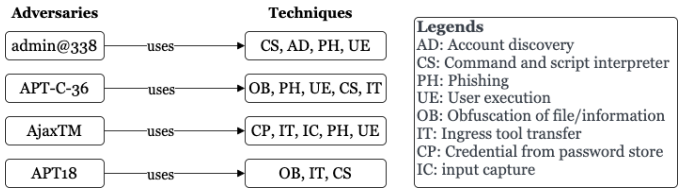

Fig. 1. Example of four real-world cyberattacks consisting of multiple techniques [1], [3], [2], [5]

the MITRE ATT&CK framework. To answer RQ1, we use frequent itemset mining to identify the following used by adversaries: (a) a set of top adversarial techniques; (b) a set of top co-occurring adversarial techniques. To answer RQ2, we first use association rule mining to identify a set of probabilistic rules of co-occurring techniques. We then build a co-occurrence network from the obtained rules and apply three centrality measures to identify adversary behaviors through technique co-occurrences. Overall, we list our contribution:

1) Identification of top adversarial techniques used in conjunction by adversaries, thus benefiting cybersecurity researchers and practitioners in understanding cyberattacks from the perspective of leveraging techniques;
2) Identification of co-occurrence rules among adversarial techniques that can aid practitioners in threat hunting through predicting potential technique(s) in use based upon the evidence of a given technique;
3) Identification of detection, and mitigation strategies, and adversary behaviors from techniques co-occurrences that can aid practitioners in deriving mitigation strategies;
4) Practitioners can use the methodology and findings of the paper as a starting point to investigate adversarial techniques in their environment, derive mitigation strategies, and improve security practices. The dataset and analysis script are made available (see § X) to replicate the study on future ATT&CK versions and any other co-occurrence data on the usage of adversarial techniques.

The rest of the paper is organized below. In § II, we discuss a few key concepts. In § III, we discuss the methodology of our work. In § V, and § VI, we discuss our findings. We § VII, we reflect on observations from the findings and future research paths. In § VIII, we discuss several study limitations. In § IX, we discuss related works and in § XI, we conclude the paper.

II. Key Concepts

In this section, we discuss concepts related to the methodology and findings of this study.

A. Tactics, techniques and procedures (TTPs)

Threat actors utilize a plethora of tactics, techniques, and procedures (TTPs) to compromise the security of target organizations or systems. *Tactic* refers to adversary's tactical goal for performing an action, *Techniques* refers to *how* an adversary achieves a tactical goal by performing an action [9], [105]. *Procedures* refer to the specific method adversaries have used for implementing techniques to achieve a particular tactic [105]. TTPs are used to profile and analyze the lifecycle and behavior of adversaries launching attacks on a targeted system [59]. For example, *privilege escalation* [64] is a tactic for gaining elevated permission on a system. An example technique of privilege escalation is *access token manipulation* [14]. Thus, an attacker can gain elevated privilege in a system by tampering with the access token to bypass the access control mechanism. An example procedure for this tactic and procedure is the FIN6 cybercrime group [7] *manipulating an access token* by *using Metasploit's named-pipe impersonation* [100].

B. MITRE ATT&CK

The MITRE [56] organization introduced the ATT&CK [9] framework in 2013, derived from real-world observations of adversarial TTPs deployed by cybercrime groups and malware. ATT&CK catalogs an enumeration of tactics. Each tactic has an enumeration of corresponding techniques. Each technique has an enumeration of corresponding procedure(s). The procedures are collected from online articles and technical reports describing cyberattack incidents. ATT&CK cites each of the articles and reports their corresponding TTPs. ATT&CK catalogs and annually updates the sets of TTPs observed in *cybercrime groups* [43] (i.e., cyber-criminals or malicious campaigns tracked by a common name in the security community) and the *malware* [73] (i.e., software/tools, scripts, executable used for malicious purposes). We use the enterprise ATT&CK Version 10 in this study, and this version enumerates 14 tactics and 188 adversarial techniques.

C. Frequent Itemset Mining

Frequent Itemset Mining (FIM) refers to the extraction of frequently occurring items, events, and patterns from data [98]. Hence, FIM can extract techniques that are used frequently by adversaries. For example, in Fig. 1, we observe four adversaries using a collection of techniques where the four adversaries use eight techniques in total. Throughout the study, we refer to the collection of techniques used by an adversary as *technique-set*, and hence, Fig. 1 contains four *technique-sets* used by four adversaries.

In the scope of FIM, each of the techniques are *items*, and each of the *technique-sets* are *itemset*. FIM takes one input called *minimum support (minSup)* and returns *frequent itemsets* referring to the *itemsets* appearing in at least *minSup%* of all *itemsets* present in the data. *Frequent itemsets* consist of sets of items (i.e. techniques) and each of the sets can contain a single item (i.e. technique) or multiple items (i.e. collection of techniques). A set containing a single item represents a technique that is used by *minSup%* of adversaries. A set containing multiple items represents co-occurring techniques used by *minSup%* of adversaries in the scope of an attack.

We provide examples from Fig. 1. We assume *minSup* is 0.25 which denotes all *frequent itemsets* returned by FIM must appear in at least 1 out of 4 *itemsets*. One such *frequent itemsets* is $\{OB\}$ as the technique OB with $support = 0.5$ (i.e. OB appears in 2 out of 4 *itemsets*) satisfies the *minSup*. Similarly, another *frequent itemsets* is $\{CS\}$ as the technique CS with $support = 0.75$ (i.e. CS appears in 3 out of 4 *itemsets*) satisfies the *minSup*. Note that these two examples

show that each of the two *frequent itemsets* contains only a single item, and we refer *frequent itemsets* containing only a single item as *frequent individual technique*. Hence, OB or CS are examples of *frequent individual techniques*. However, FIM also returns *frequent itemsets* containing multiple items such as the following two sets: $\{OB, IT\}$, and $\{PH, UE\}$. We see both techniques, OB and IT, are used together by two adversaries; hence, both appear at 2 out of 4 itemsets satisfying *minSup*. Similarly, we also see both techniques: PH and UE are used together by three adversaries, and hence, both techniques appear at 3 out of 4 itemsets satisfying *minSup*. We refer *frequent itemsets* containing multiple items as *frequent co-occurring techniques*. Hence, $\{OB, IT\}$, and $\{PH, UE\}$ are examples of *frequent co-occurring techniques*.

D. Association Rule Mining

Frequent co-occurring techniques obtained from frequent itemset mining indicate potential associations among the techniques. For example, we obtain a *frequent co-occurring techniques*: $\{PH, UE\}$ (see § II-C) which indicates that a potential association (such as correlation, causation, relation) exists between the two techniques. Hence, we can identify an association rule $PH \implies UE$, which indicates that if PH appears in an *itemset*, then UE also appears in the same *itemset*. Association Rule Mining (ARM) refers to the extraction of association rules from *itemsets*. As in this study, we are investigating techniques occurring together in a single cybercrime group or malware activity. We refer to association rules as *co-occurrence rules* among adversarial techniques. Hence, $PH \implies UE$ is an example of *co-occurrence rules*.

A *co-occurrence rule* looks identical to an *if-then* expression. For example, the example *co-occurrence* rule: $PH \implies UE$ denotes that *if* PH occurs, *then* UE also occurs. We refer to the *if* portion of the rule as an *antecedent* (e.g. PH), and the *then* portion of the rule as an *consequent* (e.g. UE). ARM takes two input to extract *co-occurrence* rules: *minimum support (minSup)*, and *minimum confidence (minConf)*. *minSup* denotes that the *co-occurrence rule* materializes in at least *minSup%* of *itemsets*. *minConf* denotes that, in a *co-occurrence rule*, given that *antecedent* appears in an itemset, *consequent* has a *minConf* probability of appearing in the *itemset*.

We provide examples from Fig. 1 where we assume, ARM is run with $minSup = 0.5$, and $minConf = 0.5$. We identify a *co-occurrence rule*: $CS \implies OB$ having $support = 0.5$ and $confidence = 0.66$. We see: (a) both CS and OB appear in 2 out 4 itemsets; (b) CS appears in 3 itemsets and OB appears in 2 out of the 3 itemsets. Note that, in this rule, both *antecedent* and *consequent* contain only one technique each. We refer to such rule as *simple co-occurrence rule*. From Fig. 1, we identify another rule: $PH \wedge UE \implies CS$ having $support = 0.5$ and $confidence = 0.66$. Note that, in this rule, *antecedent* contains two techniques. We refer to *co-occurrence rule* where either *antecedent* or *consequent* contain more than one technique as *compound co-occurrence rule*.

E. Co-occurrence network

A co-occurrence network refers to the graph representing relations between items appearing in the same dataset [103].

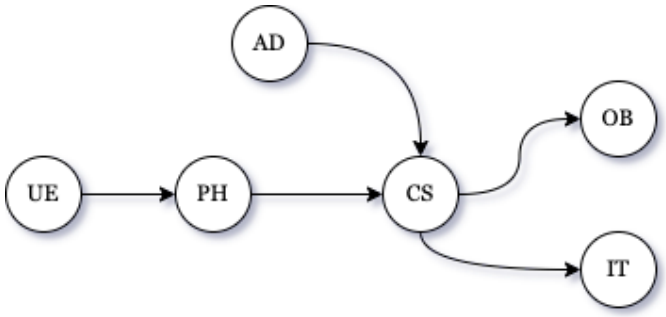

Fig. 2. Example of co-occurrence network derived from Fig 1

In this study, we use co-occurrence network to represent *co-occurrence rules* among the adversarial techniques. The co-occurrence network is a directed graph where each node represents a technique and each directed edge represents a *simple co-occurrence rule*. We demonstrate how we construct a co-occurrence network obtained from *co-occurrence rules* based on Fig. 1. Assume these five following *simple co-occurrence rules* are obtained: $UE \implies PH$, $PH \implies CS$, $AD \implies CS$, $CS \implies IT$, and $CS \implies OB$. Five techniques exist in these five rules. We represent these five techniques by five nodes and then, we represent *co-occurrence rules* by directed edge from *antecedent* techniques to *consequent* technique. Thus, we obtain a co-occurrence network shown in Fig. 2.

Note that, while each edge in the network represents a *simple co-occurrence rule*, a path in the network represents a *chained co-occurrence rule* referring to two or more rules being chained together. For example, a path: (PH, CS, OB) represents a *chained co-occurrence rule*: $PH \implies CS \implies OB$. The path denotes the following: (a) given that PH occurs, CS is likely to occur; (b) given that both PH and CS occur, OB is likely to occur. In the *chained co-occurrence rule*, CS works as an antecedent for OB and consequent for PH. Hence, we refer to the intermediate nodes in a path as (e.g. CS) as *intermediate antecedent* or *intermediate consequent*. We use these two terms interchangeably throughout the paper. Also, note that while PH can directly imply the probability of occurrence of CS. However, PH cannot directly imply the probability of occurrence of OB because, in the given rule, PH can only imply the occurrence of OB, given that CS also occurs. Thus, in a *chained co-occurrence rule*, the probability of a consequent depends on both the probability of its *intermediate antecedent*, and *antecedent*. Hence, the longer a path is in the network, the extent of an *antecedent*, implying the probability of the occurrence of the *consequent* is weaker. A shortest path between an *antecedent* and a *consequent* reflects the strongest implication of occurrence of the *consequent* given that the *antecedent* occurs.

F. Centrality Measures

In the context of graphs or networks, centrality measures refer to the computation of how *important* a node is compared to other nodes [99]. The notion of *importance* depends on the context and structure of the graph or network. Hence, in a co-occurrence network, centrality measures can indicate the importance of adversarial techniques in the context of how

they imply the occurrence of one another. In this study, we apply the three centrality measures. We discuss the measures from the perspective of a hypothetical technique te in Table I.

III. Methodology

We discuss the methodology in this section. An overview of the methodology appears in Fig. 3. The rectangle boxes represent steps in the method. The hexagonal boxes represent information that is the input/output of each step. We also use arrows from steps to RQs to show which step is used to draw observations to answer the RQs.

A. Construct Dataset

MITRE ATT&CK catalogs a collection of cybercrime groups and malware whose activities are reported in publicly-accessible documents. MITRE ATT&CK catalogs the procedures and the corresponding adversarial technique(s) and tactics used by each group or malware during cyberattacks. We obtain the catalog of groups and malware along with the associated techniques from the MITRE ATT&CK website [34], [35]. We next obtain the catalog of tactics and techniques from MITRE ATT&CK website [36], [37]. We then combine the acquired data to build the dataset according to the schema shown in Fig 4. The schema shows: (a) each group or malware uses a collection of techniques where the group or malware implements techniques through procedures; (b) the adversary can use multiple techniques to accomplish a single tactic.

Fig. 5 provides a real example according to the schema. In the example, we show the reported techniques of a cybercrime group named APT12 [4] who targeted numerous government organizations, technology companies, and media outlets. The group has been reported to use five adversarial techniques to gain three tactical goals: (a) use of *T1566: Phishing* [63] technique to achieve the tactical goal *TA0001: Initial access* [50]; (b) use of *T1203: Exploitation for client execution* [41] and *T1204: User execution* [79] technique to achieve the tactical goal *TA0002: Execution* [38]; (c) use of *T1568: Dynamic resolution* [32] and *T1102: Web service* [80] techniques to achieve the tactical goal *TA0011: Command and control* [21]. Thus, the dataset contains a collection of groups and malware. We refer to the groups and malware as adversary entities. Each entity contains a collection of techniques reported to be used by the corresponding adversary entity. We refer to the collection of techniques used by an adversary entity as *technique-set* (§ II-C). We then keep only the *technique-sets* if the *technique-set* contains at least three techniques.

B. Apply frequent itemset mining

We apply frequent itemset mining on technique-sets to identify: (a) frequent individual techniques and (b) frequent co-occurring techniques. We set $minSup = 0.10$, and thus the mining returns frequent individual techniques and frequent co-occurring techniques appearing in at least 10% of all the technique-sets. We obtain the *tactic-set* from each of the corresponding *technique-sets* and then apply frequent itemset mining with $minSup = 0.10$ to find what set of tactics are aimed by adversaries from their corresponding used techniques. We use *mlxtend* [8] package to perform the mining. We answer the *RQ1* based upon our observation on the obtained frequent individual tactics and techniques in this step.

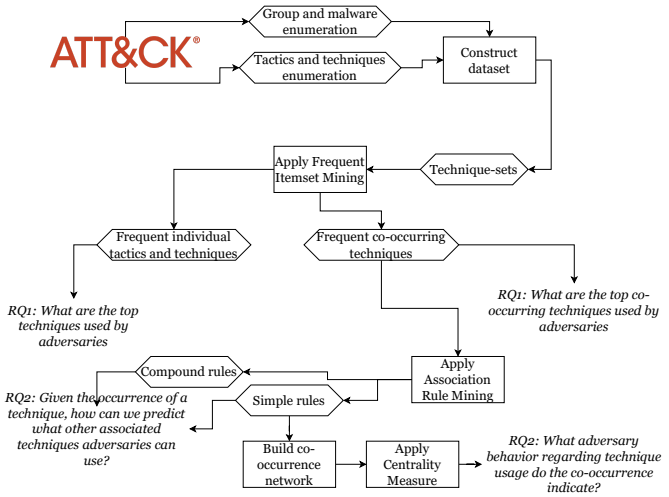

Fig. 3. An Overview of the Methodology

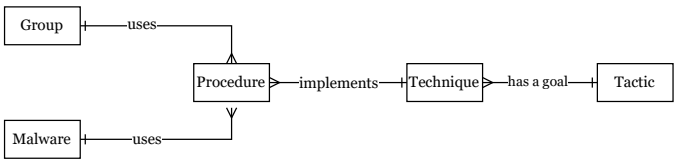

Fig. 4. The Dataset Schema

C. Apply association rule mining

We apply association rule mining to the obtained frequent co-occurring techniques to identify simple and compound co-occurrence rules among the techniques. We set $minSup = 0.10$, and $minConf = 0.15$. Thus the mining returns rules that materialize in at least 10% of all the technique-sets, and given that the antecedent technique occurs, the consequent technique occurs with at least 0.15 probability. We use *mlxtend* package to perform the mining. We answer the *RQ2: Given the occurrence of a technique, how can we predict what other associated techniques adversaries can use?* based upon our observation on obtained rules in this step.

D. Build co-occurrence network

After applying association rule mining, we obtain a set of simple and compound co-occurrence rules(§ II-D). Each simple co-occurrence rule contains antecedent and consequent techniques. We create a co-occurrence network from the set of obtained simple co-occurrence rules where the confidence value of the rule is at least greater than or equal to the median confidence value of all simple co-occurrence rules.

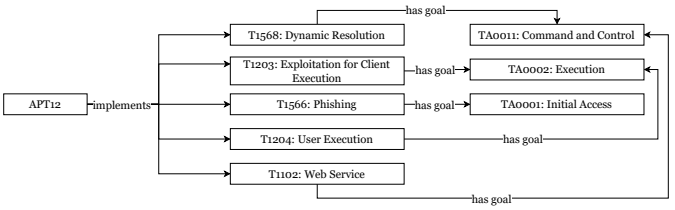

Fig. 5. Dataset Example: APT12 group [4]

TABLE I. CENTRALITY MEASURES FROM THE PERSPECTIVE OF A TECHNIQUE: te COMPARED TO OTHER TECHNIQUES IN THE NETWORK

Centrality	Definition	Explanation	Implication
In-degree centrality [90]	Number of incoming edges	Higher value indicates te is common consequent in relatively higher number of rules. E.g., in Fig. 2, CS has the highest in-degree centrality score of the other four. CS has two incoming edges, while the remaining four have one incoming edge.	A relatively high number of techniques imply the occurrence of te
Out-degree centrality [90]	Number of outgoing edges	Higher value indicates te is common antecedent in relatively higher number of rules. E.g., in Fig. 2, CS has the highest out-degree centrality score of the other four. CS has two outgoing edges while the remaining four has one outgoing edge	te implies the occurrence of a relatively higher number of other techniques
Betweenness centrality [88]	Number of shortest paths between any two techniques in the network where the paths go through te	Higher value indicates te acts as an *intermediate antecedent* or *consequent* in a relatively higher number of the shortest chained rules between any two other technique. E.g. in Fig. 2, CS is an *intermediate antecedent/consequent* of 6 shortest paths between each of: (UE, OB), (UE, IT), (PH, OB), (PH, IT), (AD, OB), and (AD, IT)	Any two techniques' co-occurrence depends on the occurrence of te more than that of any other techniques

TABLE II. SUMMARY OF THE DATASET

Adversary	Count	Technique			Tactic		
		Avg (Med)	Min	Max	Avg (Med)	Min	Max
Groups	115	19.0 (13)	3	64	7.1 (7)	2	14
Malware	484	12.3 (11)	3	56	5.0 (5)	1	11
Total	599	13.6 (11)	3	64	5.5 (5)	1	14

The co-occurrence network is a directed graph where each node represents a technique. We create directed edges between two techniques such that the direction goes from antecedent to consequent techniques. For each node, we record the tactic of the corresponding technique and the number of times the technique appears in the dataset. For each edge, we record the number of times the same adversary entity uses both the antecedent and consequent techniques.

E. Apply centrality measures

The constructed co-occurrence network reflects the probabilistic ties among the co-occurrence of adversarial techniques. Hence, in the co-occurrence network, centrality measures can indicate the importance of adversarial techniques in how their co-occurrence is probabilistic-ally tied to one another. In this study, we apply three centrality measures discussed in Table I. We use *Networkx* [10] package to compute the centrality measures, and we use the default parameters provided by Networkx while computing each measure. We answer the *RQ2: What adversary behavior regarding technique usage do the co-occurrence indicate?* based upon our observation from the computed measures in this step.

IV. THE CONSTRUCTED DATASET

After constructing the dataset according to the schema, we obtain 669 technique-sets used by 125 groups and 544 malware. We then drop the technique-sets where each of the dropped technique-sets contains less than three techniques. Thus, we obtain 599 technique-sets used by 115 groups and 484 malware. We present a summary of the dataset in Table II. Overall, the 599 technique-sets contain 172 of the 188 techniques, and the 172 techniques cover all 14 of the tactics cataloged in MITRE ATT&CK.

V. FINDINGS ON RQ1

In this section, we discuss our findings on frequent individual tactics, frequent individual techniques, and co-occurring techniques to answer RQ1.

TABLE III. TOP TEN INDIVIDUAL TECHNIQUES USED BY ADVERSARIES

Technique	Tactic	Support
T1059: Command and Scripting Interpreter [22]	EX	0.62
T1105: Ingress Tool Transfer [49]	CC	0.56
T1027: Obfuscated Files or Information [61]	DE	0.51
T1071: Application Layer Protocol [17]	CC	0.47
T1082: System Information Discovery [76]	DC	0.47
T1083: File and Directory Discovery [42]	DC	0.40
T1070: Indicator Removal on Hosts [47]	DE	0.39
T1547: Boot or Logon Autostart Execution [18]	PE	0.36
T1057: Process Discovery [65]	DC	0.34
T1016: System Network Configuration Discovery [77]	DC	0.32
Total techniques on $support > 0.1$		44
Mean support of techniques on $support > 0.1$		0.23
Median support of techniques on $support > 0.1$		0.18
Total techniques on $0.05 < support < 0.1$		24
Total techniques on $0.01 < support < 0.05$		63
Total techniques on $support < 0.01$		41
Total techniques		172

EX: TA0002 - Execution [38], CC: TA0011 - Command and control [21], DE: TA0005 - Defense evasion [29], DC: TA0007 - Discovery [31], PS: TA0003-Persistence [62]

A. Frequent individual techniques used by adversaries

We obtain a set of frequent individual techniques from applying frequent itemset mining (§ III-B). In Table III, we show the Top 10 techniques used by the adversaries. *Support* column in the table denotes the percentage of adversary entities (i.e., cybercrime groups and malware) using the technique. For example, the table shows the top technique is *T1059: Command and Scripting Interpreter* which is used by 62% of adversary entities. The table shows, 62%, 56%, and 51% of the adversary entities use the first three techniques, respectively. However, the bottom three techniques have relatively high support with use, respectively, by 36%, 34%, and 32%, of the adversary entities. The total number of techniques used by at least 10% of adversary entities is 44. However, the number of techniques in the dataset is 172. The observation suggests that 26% of the techniques in the dataset are used by at least 10% of adversary entities. We also identify 104 techniques appearing in less than 5% of adversary entities, and these 104 techniques constitute 60% of all techniques in the dataset. While adversaries use these Top 10 techniques for malicious purposes, the techniques are an abuse of legitimate system functions. Thus, detecting these techniques is challenging, and filtering malicious activities from benign system activities

becomes difficult. We manually checked the procedures of the top three techniques from the ATT&CK and identified the following: (a) *T1059* is mainly used for downloading and executing malicious scripts; (b) *T1105* is mainly used for downloading scripts and tools; and (c) *T1082* is mainly used for encrypting/encoding malicious payload and commands. From the observation, we identify the following detection and mitigation strategies (DS and MS, respectively), below.

☞ **DS-1**: Approximately half of the adversaries have three common features: command executing, downloading of files, and encrypted/encoded files or payloads. Thus, organizations can detect a relatively large number of malicious activities by detecting indicators of these three features.
☞ **MS-1**: Enforcing mitigation for all 188 ATT&CK techniques may be infeasible for organizations with limited resources. However, organizations can achieve reasonable protection against one-third of the adversaries by mitigating a small (n=10) number of techniques.

As shown in Table III, the Top 10 techniques are from five tactics. *TA0007: Discovery* has four techniques. Both *TA0011: Command and Control* and *TA0005: Defense Evasion* have two techniques each. *TA0002: Execution* and *TA0003: Persistence* have one technique each. The top three techniques are from *TA0002: Execution*, *TA0011: Command and Control*, and *TA0005: Defense evasion* respectively. The observation provides a bigger picture of generic intrusion activity. Adversaries primarily figure out the victim system by obtaining information, hiding the evidence of intrusion, remaining in the system by attaching malicious programs to periodic system functions, and executing malicious tasks through commands. By definition, techniques from *TA0003: Persistence*, and *TA0005: Defense Evasion* are challenging to detect, and techniques from *TA0002: Execution* indicate a potential breach. Meanwhile, techniques from *TA0007: Discovery* indicate potentially an early phase of a breach where the adversary is figuring out its follow-on course of actions. However, ATT&CK reports that no mitigation exists for the four *TA0007: Discovery* related techniques, and *T1547: Boot or Logon autostart execution* [76], [18], [65], [77], [42].

DS-2: Five of the Top 10 techniques: *T1082, T1547, T1057, T1016, T1083* do not have any mitigation. Hence, detection of these five techniques is of topmost priority to make an early intervention to stop system breaches.

We investigate to what extent adversary entities use techniques from each of the 14 ATT&CK tactics. We report the findings Table IV. We show the following for each tactic: (a) support denoting the percentage of entities using at least one technique; (b) the total number of techniques from the corresponding tactics having at least 0.1 support; (c) minimum, mean, median, standard deviation, and a maximum of support values of the techniques; and (d) the technique having maximum support value. The first row in the table denotes the following: (a) 85% of the entities use at least one technique from *TA0005: Defense Evasion* tactic; (b) 12 techniques from the tactic have minimum support of 0.1 (c) among the support values of 12 techniques, minimum and maximum values are respectively 0.1 and 0.51; (d) *T1027: Obfuscated Files or Information* technique has the highest support value.

We do not find any technique from *TA0043: Reconnaissance*, *TA0042: Resource Development*, *TA0004: Privilege Escalation*, and *TA0040: Impact* tactics with support greater than 0.10. No technique from *TA0043: Reconnaissance*, *TA0042: Resource Development* may indicate, how adversaries initially plan their operation at an early phase of an attack is hard to know for the organizations. No technique from *TA0040: Impact* may indicate that adversaries rather focus on remaining hidden and persistent instead of using a hit-and-run approach. However, no technique from the *TA0004: Privilege escalation* tactic may indicate adversaries circumvent *TA0004: Privilege escalation* through performing techniques from *TA0001: Initial access*, *TA0006: Credential access*, and *TA0009: Collection*. The tactics may aid an adversary in obtaining required information for the escalating privilege by gaining credentials. Another reason could be that privilege escalation attempts are not detected in the same proportions as that of other tactics.

Table IV shows that adversaries primarily use techniques from the top four tactics: *TA0005: Defense evasion, TA0011: Command and control, TA0002: Execution, and TA0007: Discovery*. 75% of the adversary entities use at least one technique from each of the four tactics. We also observe that 50% of the entities use at least one technique each from *TA0003: Persistence*, and *TA0009: Collection*. *TA0005: Defense Evasion*, and *TA0007: Discovery* tactics have the two highest number of techniques: 12 and 9 respectively. The table shows techniques having the highest support from each tactic, and we observe that techniques from the top five tactics in the table are among the Top 10 techniques shown in Table III.

A tactic can have a comparatively lower number of techniques but comparatively higher support, suggesting that adversaries use the techniques from the tactic more often than the techniques from other tactics. For example, the *TA003: Persistence* tactic has only two techniques and has a support value of 0.53. However, the *TA0009: Collection* tactic has five techniques and has a support value of 0.51. The *TA0003: Persistence* tactic also has the highest median and mean value of the support of its corresponding techniques. However, the tactic has only two techniques. In the case of support values of the techniques, we also observe that (a) *TA0009: Collection* has the lowest standard deviation; (b) *TA0001: Initial access* has the highest minimum value. The observation indicates that while adversaries lean toward using *T1566: Phishing* to gain initial access to systems, they deploy multiple techniques of *TA0009: Collection* tactic with similar probabilities. Adversaries use the highest number of techniques for *TA0005: Defense evasion* with 85% of the entities using at least one technique from this tactic. The observation indicates a primary trait of multi-stage attacks that the adversaries always intend to keep their presence hidden, enabling them to conduct their operation persistent.

☞ **MS-2**: We identify the highest number of techniques are used by the adversaries to achieve the goal of *TA0005: Defense evasion*. Organizations can prioritize the mitigation of techniques of the tactic so that adversaries cannot remain hidden for a prolonged period of time.

B. Frequent co-occurring techniques

Our analysis found 188 sets of frequent co-occurring techniques where each set appears in at least 10% of the adversary entities. We identify 33 techniques used in these 188 sets.

TABLE IV. TECHNIQUES USED BY AT LEAST TEN PERCENT OF ADVERSARIES FROM EACH ATT&CK TACTICS

Tactic	Support	Count	Min	Avg	Med	Std.	Max	Most used technique
TA0005: Defense Evasion [29]	0.85	12	0.1	0.22	0.18	0.13	0.51	T1027: Obfuscated Files or Information [61]
TA0011: Command and Control [21]	0.79	6	0.12	0.28	0.2	0.19	0.56	T1105: Ingress Tool Transfer [49]
TA0002: Execution [38]	0.76	5	0.14	0.27	0.21	0.2	0.62	T1059: Command and Scripting Interpreter [22]
TA0007: Discovery [31]	0.75	9	0.11	0.26	0.25	0.13	0.47	T1082: System Information Discovery [76]
TA0003: Persistence [62]	0.53	2	0.2	0.28	0.28	0.11	0.35	T1547: Boot or Logon Autostart Execution [18]
TA0009: Collection [20]	0.51	5	0.13	0.18	0.19	0.03	0.22	T1056: Input Capture [51]
TA0006: Credential Access [24]	0.28	2	0.13	0.13	0.13	-	0.13	T1555: Credentials from Password Stores [25]
TA0001: Initial Access [50]	0.25	1	0.18	0.18	0.18	-	0.18	T1566: Phishing [63]
TA0010: Exfiltration [40]	0.25	1	0.16	0.16	0.16	-	0.16	T1041: Exfiltration Over C2 Channel [39]
TA0008: Lateral Movement [53]	0.16	1	0.12	0.12	0.12	-	0.12	T1021: Remote Services [69]
TA0042: Resource Development [70]	0.10	0	-	-	-	-	-	-
TA0004: Privilege Escalation [64]	0.05	0	-	-	-	-	-	-
TA0043: Reconnaissance [68]	0.03	0	-	-	-	-	-	-
TA0040: Impact [45]	0.02	0	-	-	-	-	-	-

TABLE V. TOP TEN CO-OCCURRING TECHNIQUES USED TOGETHER BY ADVERSARIES

Combination of Techniques	Support
T1059/EX & T1105/CC	0.40
T1059/EX & T1027/DE	0.37
T1105/CC & T1027/DE	0.35
T1059/EX & T1071/CC	0.34
T1105/CC & T1071/CC	0.33
T1059/EX & T1082/DC	0.33
T1105/CC & T1082/DC	0.32
T1027/DE & T1071/CC	0.30
T1059/EX & T1070/DE	0.29
T1027/DE & T1082/DC	0.29
Total sets: 188, Unique Techniques: 33, Mean: 0.16, Median: 0.13	
T1059: Command and Scripting Interpreter [22], T1105: Ingress Tool Transfer, T1027: Obfuscated Files or Information [61], T1071: Application Layer Protocol [17], T1082: System Information Discovery [76], T1070: Indicator Removal on Hosts [47]	
EX: TA0002 - Execution [38], CC: TA0011 - Command and control [21], DE: TA0005 - Defense evasion [29], DC: TA0007 - Discovery [31]	

Discussing all 188 co-occurrences is out of scope. Hence, we report the Top 10 co-occurrences having the highest support value in Table V. The corresponding tactic of a technique is reported followed by a forward slash.

We observe six techniques in the Top 10 sets. We also observe that *T1059: Command and Scripting Interpreter* appears in five sets, *T1105: Ingress Tool Transfer*, and *T1027: Obfuscated Files or Information* appear in 4 sets. The top 2 co-occurrences also suggest that *T1059: Command and Scripting Interpreter* appears most with *T1105: Ingress Tool Transfer* and *T1027: Obfuscated Files or Information*. A potential correlation might exist between co-occurring techniques, such as co-occurring techniques that may aid the adversary in exploiting a single attack vector. For example, using *remsec* [11] malware, adversary can *download* malicious tools through *T1105: Ingress tool transfer* using *T1071: Application layer protocol*. Given two techniques, an adversary may also use the first technique as a requirement to use the second technique. For example, *chimera* [6] group applied *T1082: System information discovery* technique by *executing fsutil* command, which is an use of *T1059: Command and scripting interpreter* technique. Thus, co-occurring techniques may help defenders identify other potential correlated techniques, given the identification of a technique. We also observe that the ten co-occurrence are among the four tactics: *TA0002: Execution*, *TA0005: Defense evasion*, *TA0007: Discovery*, and *TA0011: Command and control*. Among 6 out of the 10 co-occurrences, we find: (a) *TA0002: Execution* and *TA0005: Defense evasion* twice; (b) *TA0005: Defense evasion*, and *TA0011: Command and control* twice; (c) *TA0002: Execution* and *TA0011: Command and control* twice.

☞ **DS-3**: Adversaries' chosen set of techniques primarily achieve three goals: *TA0002: Execution*, *TA0005: Defense evasion*, and *TA0011: Command and control*. Our results indicate that adversaries will likely leverage techniques from these three tactics to breach a system. Hence, detection from one of these tactics can potentially indicate adversaries using techniques from the other two tactics.

VI. FINDINGS ON RQ2

We discuss our findings on co-occurrence rules and the co-occurrence network to answer the RQ2 in this section.

A. Co-occurrence rules among technique

We obtain a set of 376 simple and 4,670 compound co-occurrence rules from applying frequent itemset mining (§ III-B). We report the Top 10 simple rules in Table VI and Top 10 compound rules in Table VII. The corresponding tactic of a technique is reported followed by a forward slash. Both tables show the rules along with the corresponding support and confidence of the rule. The rules are sorted by their corresponding confidence score. For example, in Table VI, the first rule suggests if *T1566: Phishing* technique is used in n adversarial entities, then we can predict *T1204: User Execution* technique is used in $0.95 \times n$. The table indicates that the first rule materializes in the activities of 17% of the entities.

We identify six rules containing *T1059: Command and Scripting Interpreter* technique. In all six rules, the technique is a consequent. The observation suggests the *T1059: Command and Scripting Interpreter* technique may co-occur with a relatively high number of other adversarial techniques. Moreover, the other adversarial techniques implying the occurrence of *T1059* indicate that *T1059* has an overlapping attack vector with other adversarial techniques. For example, *T1204: User execution* and *T1059* both use command execution. These rules may indicate the order of adversarial actions, such as in Rule 7, where adversaries apply deobfuscation to obfuscated files before further usage. The rules can also suggest the usage of related techniques together: such as in Rule 2, where both techniques aid adversary in identifying system information. We also observe mutual interaction between two techniques, such

TABLE VI. TOP 10 SIMPLE CO-OCCURRENCE RULES ALONG WITH CITATIONS TO THE DEFINITIONS OF THE TECHNIQUES IN ATT&CK

SI.	Rules	Supp.	Conf.
1	T1566: Phishing/IA [63] $\implies$ T1204: User Execution/EX [79]	0.17	0.95
2	T1033: System Owner/User Discovery/DC [78] $\implies$ T1082: System Information Discovery/DC [76]	0.21	0.86
3	T1132: Data Encoding/CC [27] $\implies$ T1059: Command and Scripting Interpreter/EX [22]	0.12	0.86
4	T1053: Scheduled Task or Job/EX [71] $\implies$ T1059: Command and Scripting Interpreter/EX	0.17	0.83
5	T1566: Phishing/IA $\implies$ T1059: Command and Scripting Interpreter/EX	0.15	0.82
6	T1041: Exfiltration Over C2 Channel/EF [39] $\implies$ T1059: Command and Scripting Interpreter/EX	0.13	0.82
7	T1140: Deobfuscate/Decode Files or Information/DE [30] $\implies$ T1027: Obfuscated Files or Information/DE [61]	0.24	0.82
8	T1218: System Binary Proxy Execution/EX [75] $\implies$ T1059: Command and Scripting Interpreter/EX	0.15	0.81
9	T1204: User Execution/EX $\implies$ T1059: Command and Scripting Interpreter/EX	0.17	0.80
10	T1204: User Execution/EX $\implies$ T1566: Phishing/IA	0.17	0.80

Total: 376, minimum confidence: 0.16, mean confidence: 0.49, median confidence: 0.51

IA: TA0001 - Initial access [50], EX: TA0002 - Execution [38], DC: TA0007 - Discovery [31], CC: TA0011 - Command and control [38], EF: TA0010 - Exfiltration [40], DE: TA0005 - Defense evasion [29]

TABLE VII. TOP 10 COMPOUND CO-OCCURRENCE RULES

SI.	Rules	Supp.	Conf.
1	T1105: Ingress Tool Transfer/CC [49] $\wedge$ T1566: Phishing/IA [63] $\implies$ T1204: User Execution/EX [79]	0.12	0.99
2	T1105: Ingress Tool Transfer/CC $\wedge$ T1059: Command and Scripting Interpreter/EX [22] $\wedge$ T1566: Phishing/IA $\implies$ T1204: User Execution/EX	0.11	0.98
3	T1105: Ingress Tool Transfer/CC $\wedge$ T1033: System Owner/User Discovery/DC [78] $\wedge$ T1057: Process Discovery/DC [65] $\implies$ T1082: System Information Discovery/DC [76]	0.10	0.95
4	T1027: Obfuscated Files or Information/DE [61] $\wedge$ T1566: Phishing/IA $\implies$ T1204: User Execution/EX	0.13	0.97
5	T1027: Obfuscated Files or Information/DE $\wedge$ T1566: Phishing/IA $\wedge$ T1059: Command and Scripting Interpreter/EX $\implies$ T1204: User Execution/EX	0.11	0.97
6	T1057: Process Discovery/DC [65] $\wedge$ T1033: System Owner/User Discovery/DC $\implies$ T1082: System Information Discovery/DC	0.13	0.96
7	T1059: Command and Scripting Interpreter/EX $\wedge$ T1566: Phishing/IA $\implies$ T1204: User Execution/EX	0.14	0.95
8	T1033: System Owner/User Discovery/DC $\wedge$ T1071: Application Layer Protocol/CC [17] $\wedge$ T1016: System Network Configuration Discovery/DC [77] $\implies$ T1082: System Information Discovery/DC	0.10	0.95
9	T1083: File and Directory Discovery/DC [42] $\wedge$ T1033: System Owner/User Discovery/DC $\implies$ T1082: System Information Discovery/DC	0.12	0.95
10	T1057: Process Discovery/DC [65] $\wedge$ T1105: Ingress Tool Transfer/CC [49] $\wedge$ T1016: System Network Configuration Discovery/DC $\implies$ T1082: System Information Discovery/DC	0.12	0.95

Total: 4670, minimum confidence: 0.16, mean confidence: 0.49, median confidence: 0.47

IA: TA0001 - Initial access [50], EX: TA0002 - Execution [38], DC: TA0007 - Discovery [31], CC: TA0011 - Command and control [38], EF: TA0010 - Exfiltration [40], DE: TA0005 - Defense evasion [29]

as in Rule 1 and Rule 10, where phishing and user execution both implies the occurrence of each other. We also observe the transitive property [107] among the rules. Such as the Rule 1, 9, and 5, where, *T1566* implies *T1059* based on the fact that, *T1566* implies *T1204*, and *T1204* implies *T1059*. Seven of the ten rules have techniques from *TA0002: Execution* tactic, which reflects that adversaries pick techniques that can eventually aid them in getting the opportunity to execute malicious commands in the victim environment.

Table VII shows the Top 10 compound rules. For example, Rule 1 denotes that when an adversary uses both *T1105: Ingress Tool Transfer* and *T1566: Phishing*, then we can predict the adversary is also 99% likely to use the *T1204: User Execution*. In 12% of the adversary entities, this rule materializes. While simple rules provide a one-to-one probability of the co-occurrence of two techniques, compound rules may provide a bigger context of how adversaries use techniques in conjunction. For example, Rule 3 shows that malicious tools used by adversaries are related to discovering system information. Rule 4 captures the context that an adversary has phished a user with an obfuscated malicious attachment which the user would later execute. Discussing all the thousands of simple and compound rules is out of the scope. However, cybersecurity researchers and practitioners can obtain and investigate all the rules from the replication package of this study. Organizations can use these rules as the starting point for the prediction of adversarial techniques. As Milajerdi et al. emphasized, false alarms of intrusion activity is one of the existing challenges in indicator-based threat detection systems [101]. Co-occurrence rules can help in filtering malicious activity from benign ones.

☞ **DS-4**: Organizations can predict potentially associated techniques from prior occurrence of techniques which can aid them detect and forecast intrusion activities with higher precision.
☞ **MS-3**: Organizations can enforce appropriate security controls and improve security practices upon the investigation of the underlying reasons for technique co-occurrences in their environment, such as dependency and similar attack vectors among techniques.

B. Adversary Behavior

1) Co-occurrence network: We obtain 376 simple co-occurrence rules, and the median confidence value of the rules is 0.51, as shown in Table VI. We build a co-occurrence network based on the simple rules having a confidence value of at least 0.51. In Table VIII, we show the co-occurrence network in adjacency matrix format. The network has 33 nodes and 188 edges representing 188 simple co-occurrence rules among 33 techniques. We reflect on our findings from the network upon several following assumptions: (a) a directed edge from $technique_a$ to $technique_b$ implies that an adversary entity is likely to use $technique_b$ by at least 51% whenever

TABLE VIII. ADJACENCY MATRIX OF THE CO-OCCURRENCE NETWORK. IN THE TABLE, * DENOTES A DIRECTED EDGE EXISTS FROM THE TECHNIQUE LOCATED ON ROW TO THE TECHNIQUE LOCATED ON THE COLUMN

		IA	EX			PS		DE								DC					CL				CC			EF
Tactic	**Technique**	**T1566**	**T1047**	**T1053**	**T1059**	**T1106**	**T1204**	**T1543**	**T1547**	**T1027**	**T1036**	**T1055**	**T1070**	**T1112**	**T1140**	**T1218**	**T1562**	**T1564**	**T1016**	**T1033**	**T1057**	**T1082**	**T1083**	**T1518**	**T1005**	**T1056**	**T1113**	**T1560**	**T1071**	**T1090**	**T1105**	**T1132**	**T1573**	**T1041**
IA	T1566: Phishing				*		*			*																				*			
EX	T1047: Windows Management Instrumentation [81]				*																												
EX	T1053: Scheduled Task/Job [71]				*				*	*			*									*							*	*			
EX	T1059: Command and Scripting Interpreter [22]									*												*							*	*			
EX	T1106: Native API [58]				*					*			*									*	*							*	*			
EX	T1204: User Execution [79]	*			*				*	*												*								*	*			
PS	T1543: Create or Modify System Process [23]				*					*			*									*								*	*			
PS	T1547: Boot or Logon Autostart Execution [18]				*					*			*									*								*	*			
DE	T1027: Obfuscated Files or Information [61]				*								*									*								*	*			
DE	T1036: Masquerading [54]				*					*												*								*	*			
DE	T1055: Process Injection [66]				*				*	*			*								*	*	*							*	*			
DE	T1070: Indicator Removal on Host [47]				*					*												*	*							*	*			
DE	T1112: Modify Registry [57]				*					*												*									*			
DE	T1140: Deobfuscate/Decode Files or Information [30]				*					*												*	*							*	*			
DE	T1218: Signed Binary Proxy Execution [75]				*				*	*												*								*	*			
DE	T1562: Impair Defenses [46]				*																													
DE	T1564: Hide Artifacts [44]				*																													
DC:	T1016: System Network Configuration Discovery [77]				*					*			*								*	*	*							*	*			
DC:	T1033: System Owner/User Discovery [78]				*					*			*						*		*	*	*							*	*			
DC:	T1057: Process Discovery [65]				*					*			*						*			*	*							*	*			
DC:	T1082: System Information Discovery [76]				*					*			*						*		*		*							*	*			
DC:	T1083: File and Directory Discovery [42]				*					*			*								*	*								*	*			
DC:	T1518: Software Discovery [74]				*					*											*	*								*	*			
CL:	T1005: Data from Local System [28]				*					*												*	*								*	*			
CL:	T1056: Input Capture [51]				*				*	*			*								*	*	*			*				*	*			
CL:	T1113: Screen Capture [72]				*				*	*			*								*	*	*		*					*	*			
CL:	T1560: Archive Collected Data [15]				*					*			*									*	*								*	*			
CC	T1071: Application Layer Protocol [17]				*					*												*									*			
CC	T1090: Proxy [67]				*																													
CC	T1105: Ingress Tool Transfer [49]				*					*												*							*					
CC	T1132: Data Encoding [27]				*																	*								*	*			
CC	T1573: Encrypted Channel [33]				*					*												*								*	*			
EF	T1041: Exfiltration over C2 Channel [39]				*					*												*	*							*	*			

IA: TA0001 - Initial access [50], EX: TA0002 - Execution [38], PS: TA0003 - Persistence [62], DE: TA0005 - Defense evasion [29], DC: TA0007 - Discovery [31], CL: TA0009 - Collection [20], CC: TA0011 - Command and control [21], EF: TA0010 - Exfiltration [40]

the entity uses $technique_a$; (b) the directed edge, however, cannot imply the order of the usage of technique - we cannot conclude $technique_b$ follows $technique_a$; (c) two techniques co-occurring together happens due to one or multiple reasons as discussed in § VI-A. Investigating the reason(s) for each specific co-occurrence is out of the scope. However, a technique being a common consequent of multiple antecedent techniques may indicate that the antecedents have an interaction dependency on the consequent. A technique being a common antecedent of multiple consequent techniques, may indicate the antecedent technique determines the set of consequent techniques chosen by the adversary entities. We identify the adversary behaviors of techniques used from the observation of the assumptions in the paragraph. Below, we report our observation from the table, and upon the observation, we identify corresponding adversary behaviors (AV).

- Techniques from *DC: TA0007: Discovery* tactics co-occurs with other techniques from the same tactic most (n = 17). The observation indicates **AV-1**: adversaries deploy multiple means to discover system information
- *DE: TA0005: Defense Evasion*, and *CC: TA0011: Command and Control* is the pair of different tactics having the highest co-occurrence among their techniques (n=13). The observation indicates **AV-2**: adversaries utilize techniques of the two tactics together most: *TA0005: Defense Evasion*, and *TA0011: Command and Control*. The behavior may also indicate the adversaries aim to hide their footprint while communicating with adversary-controlled remote devices
- *EX: TA0002: Execution* and *CL: TA0009: Collection* have the highest number of consequent tactics (n = 6). The observation indicates **AV-3**: *TA0002: Execution*, and *TA0009: Collection* tactics mostly determine the follow-on tactics adversaries would perform through techniques
- All the seven tactics reported in the table can imply the occurrence of techniques from *EX: TA0002: Execution*, *DE: TA0005: Defense Evasion*, and *CC: TA0011: Command and Control* tactics. The observation indicates **AV-4**: adversaries require to achieve the goal of *TA0002: Execution*, *TA0005: Defense Evasion*, *TA0011: Command and Control* tactics to perform techniques from all other tactics reported in the table
- Techniques from *DC: TA0007: Discovery* tactic have the highest number of consequent techniques (n = 46). The observation indicates **AV-5**: adversaries use tech-

niques from *TA0007: Discovery* to determine follow-on techniques

- Techniques from *CC: TA0011: Command and Control* have the highest number of antecedent techniques (n = 54). The observation indicates **AV-6**: the techniques from *TA0011: Command and Control* tactic have the most interaction dependency with other techniques
- No techniques from the following six tactics appear in the table: *TA0004: Privilege escalation*, *TA0006: Credential access*, *TA0008: Lateral movement*, *TA0040: Impact*, *TA0043: Reconnaissance*, and *TA0042: Resource development*. Techniques from these tactics did not appear as antecedents or consequents in any rule with 0.51 confidence value. In Table IV, we also observe that only three techniques exists from these six tactics and their support is also the lowest compared to techniques from other tactics. The observations indicate **AV-7**: *TA0004: Privilege escalation*, *TA0006: Credential access*, *TA0008: Lateral movement*, *TA0040: Impact*, *TA0043: Reconnaissance*, and *TA0042: Resource development* tactics are independent of other tactics.

2) Centrality Measures: We compute the three centrality measures (Table I) on the co-occurrence network, and we report the scores for each of the 33 techniques in Table IX. For each of the centrality measures, we bold the top five techniques and any techniques having equal value to the fifth topmost technique. We discuss our observations from the table in the following sections.

3) In-degree (IDC) and out-degree (ODC) centrality: IDC of a technique te indicates how many other techniques can imply the occurrence of te. ODC of a technique te indicates the occurrence of how many techniques te can imply. We observe the following from Table IX, and we identify adversary behaviors upon the observation.

T1059: Command and Scripting Interpreter has the highest in-degree centrality, and the technique has incoming edges from all the other 32 techniques. The four next topmost technique in in-degree centrality are: *T1105: Ingress Tool Transfer* (n=28), *T1027: Obfuscated File/Information* (n=27), *T1082: System Information Discovery* (n=27), and *T1071: Application Layer Protocol* (n=26). The top five techniques have incoming edges from at least 78% of the techniques in the network. However, 19 techniques do not have any incoming edges at all. The observation indicates a relatively high number of techniques depends on the top five techniques in in-degree centrality.

We observe that these techniques are rather an abuse of legitimate system functionality. All operating systems (OS) and platforms ship with built-in command execution capabilities, which adversaries abuse through *T1059: Command and Scripting Interpreter*. OS and platforms also can communicate with remote systems through various application layer protocols such as FTP, HTTPS, and SSH. Adversaries abuse the capability to import malicious files and tools through *T1105: Ingress tool transfer* using *T1071: Application layer protocol*. All OS and platforms also facilitate encryption or encoding, which aids adversaries in obfuscating malicious files or their traces through *T1027: Obfuscated files or information*. Adversaries also identify critical software and hardware information along with potential vulnerabilities and patches of the victim environment through *T1082: System information discovery*. Overall, the observation indicates **AV-8**: adversaries primarily abuse the following legitimate system functionality: command line interface, application layer protocol, file transfer, encoding/encryption, and read system properties. A relatively high number of other malicious techniques depend on abusing these five functionalities.

We observe from Table IX that *T1056: Input Capture* and *T1113: Screen Capture* have the highest out-degree centrality (n=10). The five next topmost technique in out-degree centrality are: *T1033: System Owner/User Discovery* (n=9), *T1055: Process Injection* (n=9), *T1016: System Network Configuration Discovery* (n=8), *T1057: Process Discovery* (n=8), and *T1082: System Information Discovery* (n=8). We also observe, through six out of the top seven techniques in the out-degree centrality, adversaries attempt to understand aspects of the victim system such as architecture, network, file system, or collect the information of interest. The observation suggests **AV-9**: an adversary's follow-on techniques depend on the obtained information of victim systems. We also observe that while many techniques have no incoming edges, all have at least one outgoing edge. The observation indicates **AV-10**: while the majority of the techniques do not act as a precursor for other techniques, all the techniques can imply adversaries' follow-on choice of techniques. Finally, we observe *T1082: System Information Discovery* technique is the only technique among the top five in both in-degree and out-degree centrality. The observation indicates **AV-11**: *T1082: System Information Discovery* acts as a common antecedent and consequent both. The technique aids adversaries by acting as a precursor for implementing other techniques. Moreover, the technique also shapes adversaries' follow-on behavior.

The top techniques in in-degree and out-degree centrality reflect the local aspect of malicious actions. For example, the incoming and outgoing edges from *T1059* reflect how an adversary can use the command line in conjunction with other adversarial techniques. Thus the in-degree and out-degree centrality can aid practitioners in detection and mitigation by the following:

☞ **DS-5**: Organizations can build detection rules around the following techniques: *T1016, T1027, T1033, T1056, T1057, T1059, T1071, T1082, T1105, T1113*, aiding them in detecting many other correlated techniques.
☞ **MS-4**: The following techniques: *T1016, T1027, T1033, T1056, T1057, T1059, T1071, T1082, T1105, T1113* - are associated with many techniques. Mitigation of these can eventually aid in mitigating the associated techniques.
☞ **MS-5** Organizations should investigate why specific techniques are highly associated with others and whether such associations are correlated with the limitations of security solutions and practices. For example, suppose an organization finds *T1059: Command and Scripting Interpreter* is associated with many adversarial techniques. The organization may investigate whether the system configuration allows the non-restricted command execution in the environment.

4) Betweenness Centrality (BC): BC of a technique te indicates to what extent te serves as a bridge between one

TABLE IX. CENTRALITY MEASURES THE TECHNIQUES IN THE NETWORK

Technique	Tactic	IDC	ODC	BC
T1566: Phishing [63]	TA0001: Initial Access [50]	1	4	0
T1047: Windows Management Instrumentation [81]	TA0002: Execution [38]	0	1	0
T1053: Scheduled Task/Job [71]	TA0002: Execution [38]	0	7	0
T1059: Command and Scripting Interpreter [22]	**TA0002: Execution** [38]	**32**	4	**33.7**
T1106: Native API [58]	TA0002: Execution [38]	0	7	0
T1204: User Execution [79]	TA0002: Execution [38]	1	7	2.2
T1543: Create or Modify System Process [23]	TA0003: Persistence [62]	0	6	0
T1547: Boot or Logon Autostart Execution [18]	TA0003: Persistence [62]	6	6	0.67
T1027: Obfuscated Files or Information [61]	**TA0005: Defense Evasion** [29]	**27**	5	**9.65**
T1036: Masquerading [54]	TA0005: Defense Evasion [29]	0	5	0
T1055: Process Injection [66]	TA0005: Defense Evasion [29]	0	**9**	0
T1070: Indicator Removal on Host [47]	TA0005: Defense Evasion [29]	14	6	2.2
T1112: Modify Registry [57]	TA0005: Defense Evasion [29]	0	4	0
T1140: Deobfuscate/Decode Files or Information [30]	TA0005: Defense Evasion [29]	0	6	0
T1218: Signed Binary Proxy Execution [75]	TA0005: Defense Evasion [29]	0	6	0
T1562: Impair Defenses [46]	TA0005: Defense Evasion [29]	0	1	0
T1564: Hide Artifacts [44]	TA0005: Defense Evasion [29]	0	1	0
T1016: System Network Configuration Discovery [77]	**TA0007: Discovery** [31]	3	**8**	0
T1033: System Owner/User Discovery [78]	**TA0007: Discovery** [31]	0	**9**	0
T1057: Process Discovery [65]	**TA0007: Discovery** [31]	8	**8**	**3.33**
T1082: System Information Discovery [76]	**TA0007: Discovery** [31]	**27**	**8**	**72.3**
T1083: File and Directory Discovery [42]	**TA0007: Discovery** [31]	13	7	**4**
T1518: Software Discovery [74]	TA0007: Discovery [31]	0	6	0
T1005: Data from Local System [49]	TA0009: Collection [20]	0	6	0
T1056: Input Capture [51]	**TA0009: Collection** [20]	1	**10**	0
T1113: Screen Capture [72]	**TA0009: Collection** [20]	1	**10**	0
T1560: Archive Collected Data [15]	TA0009: Collection [20]	0	7	0
T1041: Exfiltration Over C2 Channel [39]	TA0010: Exfiltration [40]	0	6	0
T1071: Application Layer Protocol [17]	**TA0011: Command and Control** [21]	**26**	4	0.25
T1090: Proxy [67]	TA0011: Command and Control [21]	0	1	0
T1105: Ingress Tool Transfer [49]	**TA0011: Command and Control** [21]	**28**	4	1.7
T1132: Data Encoding [27]	TA0011: Command and Control [21]	0	4	0
T1573: Encrypted Channel [33]	TA0011: Command and Control [21]	0	5	0

IDC: in-degree centrality, ODC: out-degree centrality, BC: betweenness centrality

group of techniques to another. High betweenness captures the context that the occurrence of te would lead to the occurrence of te's antecedent and consequent technique with greater probability than the case if te does not occur. We observe the following from Table IX: (a) only 10 out of 33 have a non-zero value in this centrality. Among the 10, only *T1082: System Information Discovery* (n=72.3), and *T1059: Command and Scripting Interpreter* (n=33.7) have a relatively high score.

The BC score of these two techniques suggests that *T1082* and *T1059* technique lies on 72% and 33.7% respectively of all the shortest chained rules among all pairs of techniques in the network. Thus, betweenness centrality may reflect how adversaries can use techniques together in a sequence. For example, *T1082* lies on the shortest chained rule between these two techniques: *T1105: Ingress Tool Transfer*, and *T1070: Indicator Removal on Hosts*. The example indicates that the adversary transfers malicious contents into the victim system from a remote and hides its traces. However, the adversary uses *T1082* to figure out the victim environment to determine how to hide the footprints of the malicious content. The observation indicates **AV-12**: The *T1082: System Information Discovery*, and *T1059: Command and Scripting Interpreter* techniques aid an adversary in understanding and broadening the attack surface. These two techniques, thus, may act as common steps in different possible sequences of techniques.

☞ **DS-6**: Practitioners can build detection rules around command line execution and reading system properties to capture potential sequences of adversarial techniques.
☞ **MS-6** By mitigating only the two techniques: *T1082* and *T1059*, organizations can disrupt adversaries' sequence of malicious actions.

VII. DISCUSSION AND FUTURE RESEARCH DIRECTIONS

What adversarial techniques require the most attention? Although MITRE ATT&CK enlists 188 techniques, we observe only 44 of those are prevalent among at least 10% of the adversaries. *T1059: Command and control*, *T1105: Ingress tool transfer*, *T1027: Obfuscated files or information*, *T10721: Application layer protocol*, and *T1082: System Information Discovery* techniques are the predominant techniques reported in 50% of the cybercrime groups and malware. Adversaries pair these five techniques with a large number of ATT&CK techniques. Hence, these techniques work like hubs for other techniques – as adversaries can utilize many other techniques if they can apply these five. Although these five techniques are abuses of system functionality, we observe three means of abuse. Through *T1105: Ingress tool transfer*, *T1027: Obfuscated files or information* techniques, adversaries performs a write/execute operation to gain tactical objectives. However, adversaries exploit the built-in features of an OS through *T1059: Command and control* and *T10721: Application layer protocol* techniques. On the other hand, through *T1082: System information discovery* technique, adversaries read the sys-

tem/network information. As per ATT&CK, *TA0007: Discovery*-related techniques shown in Table III do not have any mitigation at all. Two common mitigation techniques exist for the top four in the table: using anti-malware systems and network intrusion detection. Hence, enforcing these two mitigation techniques should be prioritized by defenders.

Both defense-in-depth and defense-in-breadth are required. Modern information technology infrastructures use a myriad of devices and thus create a heterogeneous environment consisting of network devices, servers, personal computers, mobile devices, and applications. Thus, enforcing security controls on only a subset of devices or applications may leave opportunities for adversaries to exploit elsewhere where security enforcement is insufficient. For example, an adversary can use *T1059: Command and scripting interpreter* on personal computers, routers, and even mobile devices. Thus, an organization may enforce the strongest security measures on execution privileges on personal computers – which is an example of defense in depth. However, an adversary may trick a user into executing a command in an unprotected mobile device – which is an example of an adversary taking advantage due to the lack of defense in breadth. Thus, organizations require a synchronized effort to detect and mitigate techniques across all endpoints so that adversaries cannot circumvent the security enforcement of one facet of an environment by exploiting another unprotected one.

Machine learning: We identify a set of co-occurrence rules among the techniques which can pave the path for building probabilistic models for predicting and detecting techniques, such as naive Bayesian or hidden Markov models. Intrusion detection systems generally suffer from false alert issues [101]. These rules can aid these systems with greater precision in generating alerts.

Detection and mitigation strategies should be adaptive to accommodate the change in the threat landscape. Our study identifies co-occurrence rules of techniques documented by MITRE ATT&CK. However, the threat landscape changes and techniques used by adversaries also depend on the adversaries' expertise and the weakness of the victim environment. Hence, organizations can run co-occurrence analysis on their environment and adapt detection and mitigation strategies that suit organizations' workflow, system architecture, and security enforcement. Organizations can also conduct longitudinal analysis to discover the correlation between adversarial techniques and the limitations of security solutions.

Cyberattack data should be made available for open research. One major challenge in developing a co-coordinated effort to thwart adversaries is that companies are reluctant to openly share and distribute the knowledge on how system compromise happened and what security enforcement failed [97]. However, the lack of fact-finding and sharing eventually deters organizations from capturing a better and actionable understanding of protecting themselves from attacks. Based on our findings, we emphasize that organizations should share pre- and post-compromise facts so that independent researchers and cybersecurity vendors can collaborate to build an actionable knowledge base of cyberattack incidents, indicators, behaviors, and pathways for creating a secure environment.

Deeper investigation is required for technique co-occurrences. Our study is built upon the primary assumption that frequently co-occurring techniques may correlate. However, deeper analysis is required to uncover the underlying reasons for the observed correlation, such as dependency, similar attack vectors, or flow along a sequence among the execution of techniques. Such investigation could be crucial for enforcing and enhancing security measures to prevent adversaries from using the techniques.

VIII. Limitations

We discuss several limitations of the study in this section. The dataset we use reflects the adversarial techniques documented by MITRE ATT&CK. The dataset captures only the techniques used by 115 cybercrime groups and 484 malware. Moreover, the dataset contains only the techniques identified by security experts and then reported after cyberattacks. Consequently, adversaries may have used other techniques that were not detected and remain unreported. Thus, the dataset only reflects a subset of the techniques adversaries used. MITRE ATT&CK mapped the techniques from publicly-reported documents using an automated manner [12] - which may have introduced mapping bias in the dataset. Overall, the study's findings reflect the context of technique co-occurrence data documented by the ATT&CK framework. Nonetheless, the study approximates how cybercrime groups and malware are leveraging techniques on various types of organizations across the globe. Practitioners and defenders can use the methodology and findings of the paper as a starting point to investigate adversarial techniques in their environment, derive mitigation strategies, and improve security practices. Hence, the approach presented in the study requires further data collection and ground truth construction from numerous types of organizations across various parts of the globe. The coordinated effort among organizations and cybersecurity vendors can capture generalized information on how techniques interact with one another and their probability distribution to co-occur.

IX. Related Work

In this section, we discuss our related work. **Prediction**: In [104], the authors proposed a machine learning model to predict a set of probable TTPs based upon the prior probabilities of detected TTPs in an environment. They trained the model with a Bayesian probability network [102] of TTPs obtained from reports describing cyberattacks. They evaluated their model on five datasets obtained from NSL-KDD [60], CICIDS2017 [52], and MITRE ATT&CK, and they obtained a prediction accuracy of 93-97%. **Association among TTPs**: In [83], the authors investigated whether the TTPs demonstrate correlation in cyberattack incidents. They obtained a dataset of 66 APT and 204 software attacks from MITRE ATT&CK. They applied hierarchical clustering to group similar TTPs and identified 37 and 61 clusters for APT and software attacks. **Association among indicators and TTPs** In [82], the authors investigated the relationship between Indicators of Compromise (IoC) [48] and malware from network traffic data. They applied an association rule mining algorithm and identified association rules among malware and corresponding traces, such as the Mirai [55] malware's strong association with the following IP: 212.61.180.100.

Threat model In [86], the authors constructed a dataset of attack sequences on industrial control systems. The authors trained a hidden Markov model reflecting probabilistic transition among MITRE ATT&CK tactics and techniques. In [94], the authors proposed an automatic threat hunting model based upon genetic programming. Using their proposed domain-specific language, the authors transformed cyberattack descriptions of MITRE ATT&CK into TTPs. Then they applied genetic programming to obtain variations of a set of TTPs from known threat descriptions. They stored these variations in a database to facilitate querying and threat hunting for unknown threats. In [84], the authors proposed formalized advanced persistent attacks from the perspective of both attackers and defenders. They construct two graphs for attackers and defenders reflecting adversarial and defensive actions performed on a common set of objects in an environment. The authors then demonstrated how the model could be used to (a) unify the perception of both attackers and defenders; (b) identify traces of attacks on the objects; and (c) improve the threat hunting by measuring the success of defensive actions. The authors evaluated the model on simulated attacks mimicking cyberattacks launched by APT29 actors.

Threat actor attribution In [96], the authors attributed threat actors from the ATT&CK TTPs of mobile malware. The authors used a dataset of 120 mobile malware and 12 threat actors. For each malware, the authors computed the cosine similarity score of the malware based on the observed TTPs and IoCs. The similarity score is used as a distance function for cluster formation, where each cluster represents a responsible actor group. Their proposed attribution model shows 82% and 90% precision and recall scores. In [108], the authors used graph embedding to represent known APT attacks. The authors then attributed malicious behaviors to known APT attacks through graph matching. They evaluated their proposed model on five real-world APT attacks where the model showed a 0.95 AUC score. In [109], the authors proposed GroupTracer to identify the responsible actors of cyberattacks launched on Internet of Things (IoT) devices. GroupTracer constructs TTPs profile of attacks performed on IoT devices by responsible groups. Based upon the profile, GroupTracer performs hierarchical clustering on the TTPs profile of unseen attacks on IoT and finds potentially-accountable groups.

Association among indicators and TTPs In [92], the authors utilized MITRE ATT&CK to identify TTPs and associated resources from the Windows malware execution trace. They proposed a machine learning model named MAMBA to automatically identify the TTPs along with corresponding API and resources. Their model showed over 90% score in precision and recall. In [87], the authors investigated the execution flow of Android malware from the perspective of MITRE ATT&CK. They proposed a graph-based neural network-based learning model to identify specific TTPs from malware control flow graphs. They evaluated the model on 3,250 malware APKs, and the model achieved a 93% F1 score. In [101], the authors mapped the intrusion activities to cyber kill-chain [26] stages and identified correlation among log messages involving different stages and resources of attacks. Based upon the mapping and correlation, the authors constructed attack graphs of various stages of APT attacks, and they evaluated their proposed model on nine real-world APT attack scenarios. In [106], the authors proposed a model for detecting multi-stage attacks in an attack's early phases. Their proposed model can predict a malware's behavior and identify corresponding MITRE ATT&CK techniques. In [91], the authors proposed a tactical provenance graph based on MITRE ATT&CK that represents causal dependencies among threat alerts. The authors showed that their proposed model could help practitioners reduce false alerts by their proposed alert scoring mechanism and preserve the semantics of threat alerts/log data using limited memory.

Overall, in the literature, researchers utilized MITRE ATT&CK for threat modeling, threat attribution, profiling malware actions, and studying the association among TTPs with indicators and intrusion alerts. In this work, we utilize ATT&CK to understand how adversaries leverage multiple ATT&CK techniques and how the understanding can drive proactive detection and mitigation strategies.

X. Dataset and source code

The dataset and analysis scripts are available at: https://github.com/brokenquark/ttps-co-occurrence.

XI. Conclusion

This study aims to understand how attackers use adversarial techniques in conjunction to enable organizations to formulate detection and mitigation strategies. To this end, we collect the publicly-reported MITRE ATT&CK techniques from 599 cybercrime groups and malware. We obtain co-occurrence rules among the techniques through association rule mining and identify adversary behaviors using network analysis. We make our dataset and analysis scripts available for researchers to rerun the analysis on future versions of ATT&CK and different datasets. Organizations can use our approach as a starting point to formulate proactive defensive strategies for protecting their environments. We also advocate researchers and practitioners make synchronized efforts to collect more data to draw a comprehensive picture of how cyberattack happens and how we can make organizations more secure.

Acknowledgment

The authors would like to thank [redacted to protect anonymity.]

References

[1] admin@338, Group G0018 — MITRE ATT&CK. https://attack.mitre.org/groups/G0018/. [Online; accessed 10-Feb-2022].

[2] AJAX TM, Group G0130 — MITRE ATT&CK. https://attack.mitre.org/groups/G0130/. [Online; accessed 10-Feb-2022].

[3] APT-C-36, Group G0009 — MITRE ATT&CK. https://attack.mitre.org/groups/G0099/. [Online; accessed 10-Feb-2022].

[4] APT12, Group G0005 — MITRE ATT&CK. https://attack.mitre.org/groups/G0005/. [Online; accessed 10-Feb-2022].

[5] APT18, Group G0026 — MITRE ATT&CK. https://attack.mitre.org/groups/G0026/. [Online; accessed 10-Feb-2022].

[6] Chimera Group G0114 — MITRE ATT&CK. https://attack.mitre.org/groups/G0114/. [Online; accessed 10-Feb-2022].

[7] FIN6: Group G0037 — MITRE ATT&CK. https://attack.mitre.org/groups/G0037/. [Online; accessed 10-Feb-2022].

[8] Home - mlxtend. http://rasbt.github.io/mlxtend/. [Online; accessed 10-Feb-2022].

[9] MITRE ATT&CK. https://attack.mitre.org. [Online; accessed 10-Feb-2022].

[10] Networkx. https://networkx.org/. [Online; accessed 10-Feb-2022].

[11] Remsec Software S0125 — MITRE ATT&CK. https://attack.mitre.org/software/S0125/. [Online; accessed 10-Feb-2022].

[12] Threat report ATT&CK Mapper - TRAM — CTID. https://ctid.mitre-engenuity.org/our-work/tram/, 2014. [Online; accessed 10-Feb-2022].

[13] Multi-stage attack techniques are making network defense difficult. https://www.helpnetsecurity.com/2019/07/15/multi-stage-attack-techniques/, 2019. [Online; accessed 10-May-2022].

[14] Access Token Manipulation T1134 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1134/, 2022. [Online; accessed 10-Feb-2022].

[15] Achieve collected data T1560 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1560/, 2022. [Online; accessed 10-Feb-2022].

[16] Anatomy of APT attacks. https://www.fireeye.com/current-threats/anatomy-of-a-cyber-attack.html, 2022. [Online; accessed 10-May-2022].

[17] Application layer protocol T1071 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1071/, 2022. [Online; accessed 10-Feb-2022].

[18] Boot or Logon autostart execution T1547 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1547/, 2022. [Online; accessed 10-Feb-2022].

[19] Check Point Research: Cyber Attacks Increased 50% Year over Year. https://blog.checkpoint.com/2022/01/10/check-point-research-cyber-attacks-increased-50-year-over-year/, 2022. [Online; accessed 10-May-2022].

[20] Collection Tactic TA0009 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0009/, 2022. [Online; accessed 10-Feb-2022].

[21] Command and Control Tactic TA0011 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0011/, 2022. [Online; accessed 10-Feb-2022].

[22] Command and scripting interpreter T1059 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1059/, 2022. [Online; accessed 10-Feb-2022].

[23] Create or modify system process T1543 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1543/, 2022. [Online; accessed 10-Feb-2022].

[24] Credential Access Tactic TA0006 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0006/, 2022. [Online; accessed 10-Feb-2022].

[25] Credentials from password stores T1555 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1555/, 2022. [Online; accessed 10-Feb-2022].

[26] Cyber kill chain - lockheed martin. https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html, 2022. [Online; Accessed 10-Feb-2022].

[27] Data encoding T1132 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1132/, 2022. [Online; accessed 10-Feb-2022].

[28] Data from Local Systems T1005 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1005/, 2022. [Online; accessed 10-Feb-2022].

[29] Defense Evasion Tactic TA0005 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0005/, 2022. [Online; accessed 10-Feb-2022].

[30] Deobfuscate/decode files or information T1140 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1140/, 2022. [Online; accessed 10-Feb-2022].

[31] Discovery Tactic TA0007 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0007/, 2022. [Online; accessed 10-Feb-2022].

[32] Dynamic resolution T1568 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1568/, 2022. [Online; accessed 10-Feb-2022].

[33] Encrypted channel T1573 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1573/, 2022. [Online; accessed 10-Feb-2022].

[34] Enterprise att&ck groups. https://attack.mitre.org/docs/enterprise-attack-v10.1/enterprise-attack-v10.1-groups.xlsx, 2022. [Online; Accessed 10-Feb-2022].

[35] Enterprise att&ck software. https://attack.mitre.org/docs/enterprise-attack-v10.1/enterprise-attack-v10.1-software.xlsx, 2022. [Online; Accessed 10-Feb-2022].

[36] Enterprise att&ck tactics. https://attack.mitre.org/docs/enterprise-attack-v10.1/enterprise-attack-v10.1-tactics.xlsx, 2022. [Online; Accessed 10-Feb-2022].

[37] Enterprise att&ck techniques. https://attack.mitre.org/docs/enterprise-attack-v10.1/enterprise-attack-v10.1-techniques.xlsx, 2022. [Online; Accessed 10-Feb-2022].

[38] Execution Tactic TA0002 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0002/, 2022. [Online; accessed 10-Feb-2022].

[39] Exfiltration over C2 channel T1041 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1041/, 2022. [Online; accessed 10-Feb-2022].

[40] Exfiltration Tactic TA0010 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0010/, 2022. [Online; accessed 10-Feb-2022].

[41] Exploitation for client execution T1203 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1203/, 2022. [Online; accessed 10-Feb-2022].

[42] File and directory discovery T1083 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1083/, 2022. [Online; accessed 10-Feb-2022].

[43] Groups - Enterprise — MITRE ATT&CK. https://attack.mitre.org/groups/, 2022. [Online; accessed 10-Feb-2022].

[44] Hide artifacts T1564 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1564/, 2022. [Online; accessed 10-Feb-2022].

[45] Impact Tactic TA0040 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0040/, 2022. [Online; accessed 10-Feb-2022].

[46] Impair defenses T1562 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1562/, 2022. [Online; accessed 10-Feb-2022].

[47] Indicator removal on host T1070 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1070/, 2022. [Online; accessed 10-Feb-2022].

[48] Indicators of compromise. https://www.trendmicro.com/vinfo/us/security/definition/indicators-of-compromise, 2022. [Online; Accessed 10-Feb-2022].

[49] Ingress tool transfer T1105 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1105/, 2022. [Online; accessed 10-Feb-2022].

[50] Initial Access Tactic TA0001 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0001/, 2022. [Online; accessed 10-Feb-2022].

[51] Input capture T1056 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1056/, 2022. [Online; accessed 10-Feb-2022].

[52] Intrusion detection evaluation dataset. https://www.unb.ca/cic/datasets/ids-2017.html, 2022. [Online; Accessed 10-Feb-2022].

[53] Lateral Movement Tactic TA0008 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0008/, 2022. [Online; accessed 10-Feb-2022].

[54] Masquerading T1036 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1036/, 2022. [Online; accessed 10-Feb-2022].

[55] The mirai botnet - threats and mitigations. https://www.cisecurity.org/insights/blog/the-mirai-botnet-threats-and-mitigations, 2022. [Online; Accessed 10-Feb-2022].

[56] Mitre corporation. https://www.mitre.org/, 2022. [Online; Accessed 10-Feb-2022].

[57] Modify registry T1112 - Enterprise — MITRE ATT&CK. https://

attack.mitre.org/techniques/T1112/, 2022. [Online; accessed 10-Feb-2022].

[58] Native API T1106 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1106/, 2022. [Online; accessed 10-Feb-2022].

[59] NIST Glossary. https://csrc.nist.gov/glossary/term/Tactics{_}Techniques{_}and{_}Procedures, 2022. [Online; accessed 10-Feb-2022].

[60] Nsk-kdd dataset. https://www.unb.ca/cic/datasets/nsl.html, 2022. [Online; Accessed 10-Feb-2022].

[61] Obfuscated files or information T1027 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1027/, 2022. [Online; accessed 10-Feb-2022].

[62] Persistence Tactic TA0003 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0003/, 2022. [Online; accessed 10-Feb-2022].

[63] Phishing T1566 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1566/, 2022. [Online; accessed 10-Feb-2022].

[64] Privilege Escalation Tactic TA0004 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0004/, 2022. [Online; accessed 10-Feb-2022].

[65] Process discovery T1057 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1057/, 2022. [Online; accessed 10-Feb-2022].

[66] Process Injection T1055 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1055/, 2022. [Online; accessed 10-Feb-2022].

[67] Proxy T1090 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1090/, 2022. [Online; accessed 10-Feb-2022].

[68] Reconnaissance Tactic TA0043 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0043/, 2022. [Online; accessed 10-Feb-2022].

[69] Remote services T1021 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1021/, 2022. [Online; accessed 10-Feb-2022].

[70] Resource Development Tactic TA0042 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/tactics/TA0042/, 2022. [Online; accessed 10-Feb-2022].

[71] Scheduled Task/Job T1053 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1053/, 2022. [Online; accessed 10-Feb-2022].

[72] Screen capture T1113 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1113/, 2022. [Online; accessed 10-Feb-2022].

[73] Software - Enterprise — MITRE ATT&CK. https://attack.mitre.org/software/, 2022. [Online; accessed 10-Feb-2022].

[74] Software discovery T1518 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1518/, 2022. [Online; accessed 10-Feb-2022].

[75] System binary proxy execution T1218 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1218/, 2022. [Online; accessed 10-Feb-2022].

[76] System Information Discovery T1082 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1082/, 2022. [Online; accessed 10-Feb-2022].

[77] System Network Configuration Discovery T1016 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1016/, 2022. [Online; accessed 10-Feb-2022].

[78] System Owner/User discovery T1033 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1033/, 2022. [Online; accessed 10-Feb-2022].

[79] User Execution T1204 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1204/, 2022. [Online; accessed 10-Feb-2022].

[80] Web service T1102 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1102/, 2022. [Online; accessed 10-Feb-2022].

[81] Windows Management Instrumentation T1047 - Enterprise — MITRE ATT&CK. https://attack.mitre.org/techniques/T1047/, 2022. [Online; accessed 10-Feb-2022].

[82] Md Sahrom Abu, Siti Rahayu, Robiah Yusof, and Aswami Ariffin. An Attribution of Cyberattack using Association Rule Mining (ARM). *International Journal of Advanced Computer Science and Applications*, 11(2), 2020.

[83] Rawan Al-Shaer, Jonathan M. Spring, and Eliana Christou. Learning the Associations of MITRE ATT & CK Adversarial Techniques. In *2020 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9, Avignon, France, June 2020. IEEE.

[84] Aimad Berady, Mathieu Jaume, Valerie Viet Triem Tong, and Gilles Guette. From TTP to IoC: Advanced Persistent Graphs for Threat Hunting. *IEEE Transactions on Network and Service Management*, 18(2):1321–1333, June 2021.

[85] David Biancho. The Pyramid of Pain. http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html, 2014. [Online; accessed 10-Feb-2022].

[86] Seungoh Choi, Jeong-Han Yun, and Byung-Gil Min. Probabilistic Attack Sequence Generation and Execution Based on MITRE ATT&CK for ICS Datasets. In *Cyber Security Experimentation and Test Workshop*, pages 41–48, Virtual CA USA, August 2021. ACM.

[87] Jeffrey Fairbanks, Andres Orbe, Christine Patterson, Janet Layne, Edoardo Serra, and Marion Scheepers. Identifying ATT&CK Tactics in Android Malware Control Flow Graph Through Graph Representation Learning and Interpretability. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 5602–5608, Orlando, FL, USA, December 2021. IEEE.

[88] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[89] Yali Gao, Xiaoyong Li, Hao Peng, Binxing Fang, and Philip S. Yu. HinCTI: A Cyber Threat Intelligence Modeling and Identification System Based on Heterogeneous Information Network. *IEEE Transactions on Knowledge and Data Engineering*, 34(2):708–722, February 2022.

[90] Jennifer Golbeck. Chapter 3 - Network Structure and Measures. In Jennifer Golbeck, editor, *Analyzing the Social Web*, pages 25–44. Morgan Kaufmann, Boston, 2013.

[91] Wajih Ul Hassan, Adam Bates, and Daniel Marino. Tactical Provenance Analysis for Endpoint Detection and Response Systems. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 1172–1189, San Francisco, CA, USA, May 2020. IEEE.

[92] Yi-Ting Huang, Chi Yu Lin, Ying-REN Guo, Kai-Chieh Lo, Yeali S. Sun, and Meng Chang Chen. Open Source Intelligence for Malicious Behavior Discovery and Interpretation. *IEEE Transactions on Dependable and Secure Computing*, pages 1–1, 2021.

[93] Sophie Ireland. Revealed: The True Cost of Rising Cyber Attacks. https://ceoworld.biz/2022/02/02/revealed-the-true-cost-of-rising-cyber-attacks/, 2022. [Online; accessed 10-May-2022].

[94] Prakruthi Karuna, Erik Hemberg, Una-May O'Reilly, and Nick Rutar. Automating Cyber Threat Hunting Using NLP, Automated Query Generation, and Genetic Perturbation, April 2021.

[95] Swati Khandelwal. New Group of Hackers Targeting Businesses with Financially Motivated Cyber Attacks. https://thehackernews.com/2019/11/financial-cyberattacks.html, 2019. [Online; accessed 10-Feb-2022].

[96] Kyoungmin Kim, Youngsup Shin, Justin Lee, and Kyungho Lee. Automatically Attributing Mobile Threat Actors by Vectorized ATT&CK Matrix and Paired Indicator. *Sensors*, 21(19):6522, September 2021.

[97] Robert Knake, Adam Shostack, and Tarah Wheeler. Learning from cyber incidents: Adapting aviation safety models to cybersecurity. Technical report, Belfer Center for Science and International Affairs, Harvard Kennedy School, 2021.

[98] José María Luna, Philippe Fournier-Viger, and Sebastián Ventura. Frequent itemset mining: A 25 years review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6):e1329, 2019.

[99] Peter V. Marsden. Network Analysis. In Kimberly Kempf-Leonard, editor, *Encyclopedia of Social Measurement*, pages 819–825. Elsevier, New York, 2005.

[100] Bredan Mckeague. Pick-Six: Intercepting a FIN6 Intrusion, an Actor Recently Tied to Ryuk and LockerGoga Ransomware. https://www.gao.gov/products/gao-22-105103, 2019. [Online; accessed 10-Feb-2022].

[101] Sadegh M. Milajerdi, Rigel Gjomemo, Birhanu Eshete, R. Sekar, and V.N. Venkatakrishnan. HOLMES: Real-Time APT Detection through Correlation of Suspicious Information Flows. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1137–1152, San Francisco, CA, USA, May 2019. IEEE.

[102] Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

[103] Elad Segev. *Semantic Network Analysis in Social Sciences*. Routledge, 2021.

[104] Ayan Sentuna, Abeer Alsadoon, P. W. C. Prasad, Maha Saadeh, and Omar Hisham Alsadoon. A Novel Enhanced Naïve Bayes Posterior Probability (ENBPP) Using Machine Learning: Cyber Threat Analysis. *Neural Processing Letters*, 53(1):177–209, February 2021.

[105] Blake Strom, Andy Applebaum, Doug Miller, Kathryn Nickels, Adam Pennington, and Cody Thomas. Mitre att&ck: Design and philosophy. Technical report, MITRE, 2020.

[106] Yuvraj Sanjayrao Takey, Sai Gopal Tatikayala, Satyanadha Sarma Samavedam, P R Lakshmi Eswari, and Mahesh Uttam Patil. Real Time early Multi Stage Attack Detection. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 283–290, Coimbatore, India, March 2021. IEEE.

[107] James S Tanton. *Encyclopedia of mathematics*. Infobase Publishing, 2005.

[108] Renzheng Wei, Lijun Cai, Lixin Zhao, Aimin Yu, and Dan Meng. DeepHunter: A Graph Neural Network Based Approach for Robust Cyber Threat Hunting. In Joaquin Garcia-Alfaro, Shujun Li, Radha Poovendran, Hervé Debar, and Moti Yung, editors, *Security and Privacy in Communication Networks*, volume 398, pages 3–24. Springer International Publishing, Cham, 2021.

[109] Yixin Wu, Cheng Huang, Xing Zhang, and Hongyi Zhou. GroupTracer: Automatic Attacker TTP Profile Extraction and Group Cluster in Internet of Things. *Security and Communication Networks*, 2020:1–14, December 2020.