IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 November 2023, accepted 13 December 2023, date of publication 18 December 2023,
date of current version 4 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3344680

RESEARCH ARTICLE

Analysis and Characterization of Cyber Threats Leveraging the MITRE ATT&CK Database

BADER AL-SADA*, ALIREZA SADIGHIAN*, AND GABRIELE OLIGERI, (Member, IEEE)
Division of Information and Computing Technology (ICT), College of Science and Engineering (CSE), Hamad Bin Khalifa University (HBKU), Doha, Qatar

Corresponding author: Gabriele Oligeri (goligeri@hbku.edu.qa)

This work was supported by the Qatar National Research Fund (a member of Qatar Foundation) under Grant GRSA7-1-0516-20061. This work was supported by the Open Access funding provided by the Qatar National Library.

*Bader Al-Sada and Alireza Sadighian contributed equally to this work.

ABSTRACT MITRE ATT&CK is a comprehensive knowledge-base of adversary tactics, techniques, and procedures (TTP) based on real-world attack scenarios. It has been used in different sectors, such as government, academia, and industry, as a foundation for threat modeling, risk assessment, and defensive strategies. There are valuable insights within MITRE ATT&CK knowledge-base that can be applied to various fields and applications, such as risk assessment, threat characterization, and attack modeling. No previous work has been devoted to the comprehensive collection and investigation of statistical insights of the MITRE ATT&CK dataset. Hence, this work aims to extract, analyze, and represent MITRE ATT&CK statistical insights providing valuable recommendations to improve the security aspects of Enterprise, Industrial Control Systems (ICS), and mobile digital infrastructures. For this purpose, we conduct a hierarchical analysis starting from MITRE ATT&CK threat profiles toward the list of techniques in the MITRE ATT&CK database. Finally, we summarize our key findings while providing recommendations that will pave the way for future research in the area.

INDEX TERMS Cyber security, MITRE ATT&CK, advanced persistent threat, cyber threat analysis, cyber threat intelligence.

I. INTRODUCTION

With continuous digitization, interconnectivity, and smartness of various environments and infrastructures, such as homes, organizations, Industrial Control Systems (ICS), etc., the importance of modeling existing and emerging cyber-threats has significantly escalated in today's world. Adversaries increasingly employ sophisticated strategies to exploit vulnerabilities and weaknesses in the digital environment, making it very difficult for cyber security detection and analysis solutions to identify and prevent their malicious activities. To address this challenge, one major solution is access to a comprehensive knowledge-base of adversary tactics, techniques, and procedures (TTP) continuously updated based on emerging cyber threats.

MITRE ATT&CK is a publicly available knowledge-base of adversary TTP based on real-world observations that are classified as groups or Advanced Persistent Threats (APTs) [1]. It provides various information regarding APTs' motivations, interests, targeted regions, and used tactics and techniques. Each TTP in MITRE ATT&CK comes with a mitigation strategy that provides solutions to protect critical infrastructures. Therefore, it can help reliably predict and mitigate various types of threat that attackers pursue after an initial intrusion. MITRE ATT&CK has been widely used for threat modeling and risk assessment by numerous cyber-security products and service providers both in the private and public sectors. By studying MITRE ATT&CK knowledge-base and extracting its statistical insights, significantly valuable recommendations and defensive strategies can be issued that cyber security experts can apply to digital products, solutions, and services.

To enrich MITRE ATT&CK knowledge-base, on the one hand, individuals and organizations continuously share various threat intelligence data and information such that cyber security professionals integrate this information into the knowledge-base as threat profiles, tactics, techniques, etc. On the other hand, cyber security researchers and

The associate editor coordinating the review of this manuscript and approving it for publication was Mouquan Shen.

professionals have attempted to correlate MITRE ATT&CK knowledge-base with other information sources, such as MITRE Common Vulnerability and Exposures (CVE)[§] [2], [3], MITRE Common Attack Pattern Enumeration and Classification (CAPEC)[§] [4], [5], National Institute of Standards and Technology (NIST)[§] [6], system Logs/Audits [7], [8], [9], security reports [10], [11], [12], etc.

MITRE ATT&CK knowledge-base covers TTP related to various environments and technologies, including enterprises, Industrial Control Systems (ICS), smartphones, and Internet of Things (IoT). Hence, it is considered a highly reliable source of threat information that can be utilized in the modeling and study of various cyber threats within a wide range of sectors. Often, the MITRE ATT&CK TTP are applied to associate malware with a specific family of attacks (e.g., spyware and Trojan), and this mapping to TTP helps characterize zero-day malware.

Given its empirical nature, MITRE ATT&CK cannot be described by omni-comprehensive mathematical models. Indeed, the ATT&CK matrix is a database of elementary malicious actions extracted from previous real-world attacks. Although MITRE ATT&CK allows the break-down of (old and new) cyber security attacks into building blocks, no previous contributions have been made towards mathematical modeling.

A major research gap in the literature is to perform a comprehensive statistical analysis of the MITRE ATT&CK dataset and represent its hidden insights to be applied by experts in various contexts. To address this shortcoming, in this research, we conduct an analytical and statistical study of MITRE ATT&CK dataset to extract and represent hidden insights related to threat profiles, domains, motivations, target platforms, target sectors, tactics and techniques that help to improve public and expert knowledge about recent and emerging threats.

Contributions. The major contributions of this paper are as following:

- We provide the results of an analytical study on MITRE ATT&CK knowledge-base entities to show how they are correlated and the type of relationships thus providing insights about threat profiles, tactics, techniques, and groups.
- For each key entity in MITRE ATT&CK knowledge-base, we statistically represent its relationship with other entities and list potential insights.
- We provide several recommendations in the discussion section to be utilized for detection, prediction, prevention, and mitigation purposes.

Roadmap. The remainder of this paper is organized as follows. Section II is dedicated to background knowledge about the MITRE ATT&CK framework. In Section III, we discuss the related work. Section IV describes the under analysis dataset and corresponding terminology in this research. In Section V, we present our research methodology. Section VI describes in detail our statistical analysis and the list of insights extracted. Section VII summarizes the key findings of our research. Finally, Section VIII draws some concluding remarks.

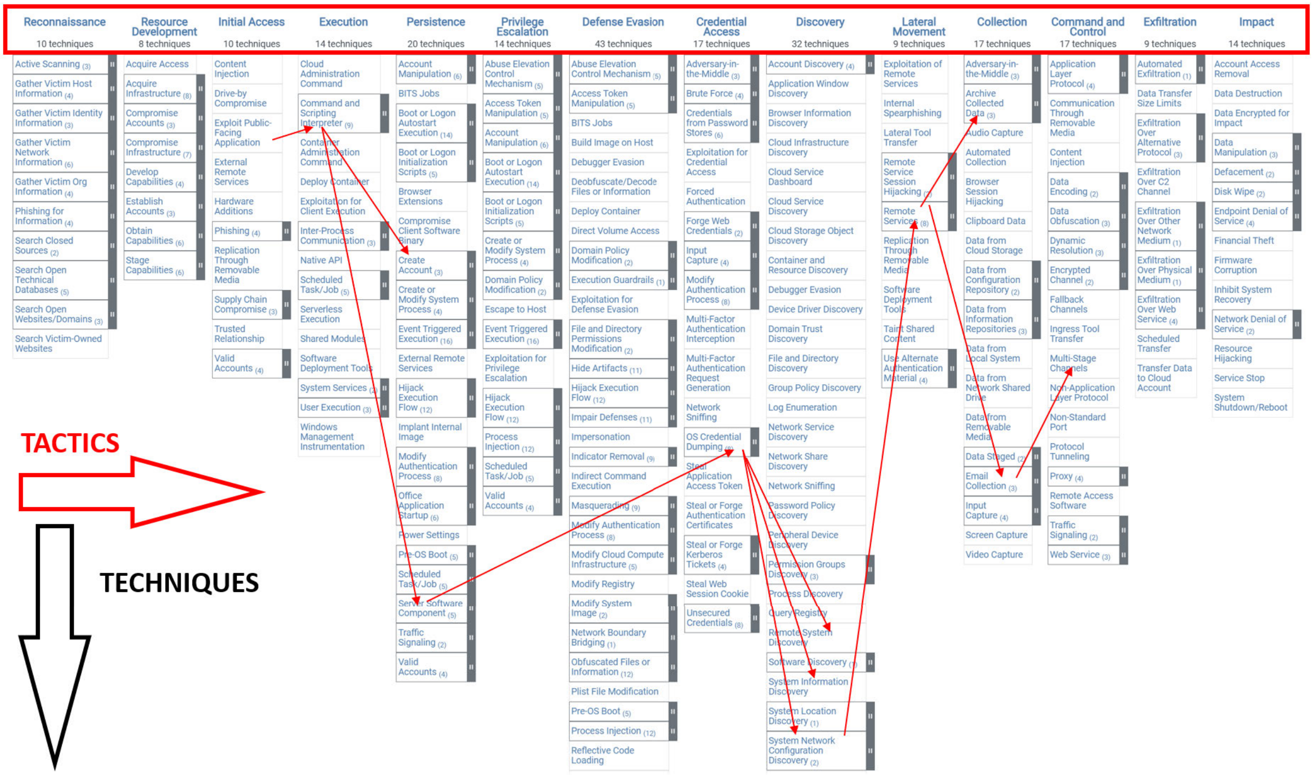
II. BACKGROUND

MITRE ATT&CK refers to MITRE Adversarial Tactics, Techniques, and Common Knowledge. It was created in 2013 after the Fort Meade Experiment (FMX), in which a group of researchers simulated a scenario that involved both attackers and defenders. The objective of the experiment was to enhance the forensic analysis of attacks by studying the behavior patterns of the adversaries. The MITRE ATT&CK framework serves as a repository of malicious activities in the real world, classified into groups or Advanced Persistent Threats (APTs). An APT is characterized by a sequence of malicious activities that represent the adversary's behavior. Taking into account the bottom-up perspective, these malicious activities are executed using specific procedures providing low-level and detailed descriptions of techniques, which are the methods employed by adversaries to achieve their goals. Techniques are further classified into tactics, which represent the adversary's highest-level description of the behavior. Therefore, an APT achieves its objective by implementing a sequence of malicious activities that can be modeled as Tactics, Techniques, and Procedures (TTP). Fig. 1 shows the typical arrangement of the ATT&CK matrix, with tactics organized in columns and techniques organized in rows. Tactics are logically ordered from left to right, highlighting the possible phases that occur during an attack.

The MITRE ATT&CK framework encompasses three technology domains: *Enterprise*, *Mobile*, and *ICS*. Each domain is associated with different tactics and techniques, although some minor overlaps may exist. These domains are fundamentally different from each other, leading to variations in adversarial behavior. The attack surfaces and the adversaries' objectives also differ among these domains, necessitating customized tactics, techniques, and procedures.

The *Enterprise* domain is the most extensive, consisting of 191 techniques and 385 sub-techniques. It covers multiple platforms such as Windows, Linux, macOS, and others. This domain is rich in attacks as a result of its broad technological scope. It encompasses platforms like Windows, macOS, Linux, Cloud, Network, and Containers. Additionally, it is the only domain that includes specific techniques (within 2 tactics) for analyzing preparatory adversarial activities (PRE-ATT&CK). In particular, entry points in the ATT&CK matrix such as Search Open Websites/Domains and Search Victim-Owned Websites (Reconnaissance) are indicative of social engineering tactics commonly employed in this domain. Typical adversary objectives in the enterprise domain include Data Destruction, Disk Wipe, and Service Stop (Impact), which are techniques frequently utilized by adversaries in enterprise scenarios.

[§] https://cve.mitre.org/
[§] https://capec.mitre.org/
[§] https://www.nist.gov/



Reconnaissance
10 techniques

- Active Scanning (3)
- Gather Victim Host Information (4)
- Gather Victim Identity Information (3)
- Gather Victim Network Information (6)
- Gather Victim Org Information (4)
- Phishing for Information (4)
- Search Closed Sources (2)
- Search Open Technical Databases (5)
- Search Open Websites/Domains (3)
- Search Victim-Owned Websites

Resource Development
8 techniques

- Acquire Access
- Acquire Infrastructure (8)
- Compromise Accounts (3)
- Compromise Infrastructure (7)
- Develop Capabilities (4)
- Establish Accounts (3)
- Obtain Capabilities (6)
- Stage Capabilities (6)

Initial Access
10 techniques

- Content Injection
- Drive-by Compromise
- Exploit Public-Facing Application
- External Remote Services
- Hardware Additions
- Phishing (4)
- Replication Through Removable Media
- Supply Chain Compromise (3)
- Trusted Relationship
- Valid Accounts (4)

Execution
14 techniques

- Cloud Administration Command
- Command and Scripting Interpreter (9)
- Container Administration Command
- Deploy Container
- Exploitation for Client Execution
- Inter-Process Communication (3)
- Native API
- Scheduled Task/Job (5)
- Serverless Execution
- Shared Modules
- Software Deployment Tools
- System Services (2)
- User Execution (3)
- Windows Management Instrumentation

Persistence
20 techniques

- Account Manipulation (6)
- BITS Jobs
- Boot or Logon Autostart Execution (14)
- Boot or Logon Initialization Scripts (5)
- Browser Extensions
- Compromise Client Software Binary
- Create Account (3)
- Create or Modify System Process (4)
- Event Triggered Execution (16)
- External Remote Services
- Hijack Execution Flow (12)
- Implant Internal Image
- Modify Authentication Process (8)
- Office Application Startup (6)
- Power Settings
- Pre-OS Boot (5)
- Scheduled Task/Job (5)
- Server Software Component (5)
- Traffic Signaling (2)
- Valid Accounts (4)

Privilege Escalation
14 techniques

- Abuse Elevation Control Mechanism (5)
- Access Token Manipulation (5)
- Account Manipulation (6)
- Boot or Logon Autostart Execution (14)
- Boot or Logon Initialization Scripts (5)
- Create or Modify System Process (4)
- Domain Policy Modification (2)
- Escape to Host
- Event Triggered Execution (16)
- Exploitation for Privilege Escalation
- Hijack Execution Flow (12)
- Process Injection (12)
- Scheduled Task/Job (5)
- Valid Accounts (4)

Defense Evasion
43 techniques

- Abuse Elevation Control Mechanism (5)
- Access Token Manipulation (5)
- BITS Jobs
- Build Image on Host
- Debugger Evasion
- Deobfuscate/Decode Files or Information
- Deploy Container
- Direct Volume Access
- Domain Policy Modification (2)
- Execution Guardrails (1)
- Exploitation for Defense Evasion
- File and Directory Permissions Modification (2)
- Hide Artifacts (11)
- Hijack Execution Flow (12)
- Impair Defenses (11)
- Impersonation
- Indicator Removal (9)
- Indirect Command Execution
- Masquerading (9)
- Modify Authentication Process (8)
- Modify Cloud Compute Infrastructure (5)
- Modify Registry
- Modify System Image (2)
- Network Boundary Bridging (1)
- Obfuscated Files or Information (12)
- Plist File Modification
- Pre-OS Boot (5)
- Process Injection (12)
- Reflective Code Loading

Credential Access
17 techniques

- Adversary-in-the-Middle (3)
- Brute Force (4)
- Credentials from Password Stores (6)
- Exploitation for Credential Access
- Forced Authentication
- Forge Web Credentials (2)
- Input Capture (4)
- Modify Authentication Process (8)
- Multi-Factor Authentication Interception
- Multi-Factor Authentication Request Generation
- Network Sniffing
- OS Credential Dumping (8)
- Steal Application Access Token
- Steal or Forge Authentication Certificates
- Steal or Forge Kerberos Tickets (4)
- Steal Web Session Cookie
- Unsecured Credentials (8)

Discovery
32 techniques

- Account Discovery (4)
- Application Window Discovery
- Browser Information Discovery
- Cloud Infrastructure Discovery
- Cloud Service Dashboard
- Cloud Service Discovery
- Cloud Storage Object Discovery
- Container and Resource Discovery
- Debugger Evasion
- Device Driver Discovery
- Domain Trust Discovery
- File and Directory Discovery
- Group Policy Discovery
- Log Enumeration
- Network Service Discovery
- Network Share Discovery
- Network Sniffing
- Password Policy Discovery
- Peripheral Device Discovery
- Permission Groups Discovery (3)
- Process Discovery
- Query Registry
- Remote System Discovery
- Software Discovery (1)
- System Information Discovery
- System Location Discovery (1)
- System Network Configuration Discovery (2)

Lateral Movement
9 techniques

- Exploitation of Remote Services
- Internal Spearphishing
- Lateral Tool Transfer
- Remote Service Session Hijacking (2)
- Remote Services (8)
- Replication Through Removable Media
- Software Deployment Tools
- Taint Shared Content
- Use Alternate Authentication Material (4)

Collection
17 techniques

- Adversary-in-the-Middle (3)
- Archive Collected Data (3)
- Audio Capture
- Automated Collection
- Browser Session Hijacking
- Clipboard Data
- Data from Cloud Storage
- Data from Configuration Repository (2)
- Data from Information Repositories (3)
- Data from Local System
- Data from Network Shared Drive
- Data from Removable Media
- Data Staged (2)
- Email Collection (3)
- Input Capture (4)
- Screen Capture
- Video Capture

Command and Control
17 techniques

- Application Layer Protocol (4)
- Communication Through Removable Media
- Content Injection
- Data Encoding (2)
- Data Obfuscation (3)
- Dynamic Resolution (3)
- Encrypted Channel (2)
- Fallback Channels
- Ingress Tool Transfer
- Multi-Stage Channels
- Non-Application Layer Protocol
- Non-Standard Port
- Protocol Tunneling
- Proxy (4)
- Remote Access Software
- Traffic Signaling (2)
- Web Service (3)

Exfiltration
9 techniques

- Automated Exfiltration (1)
- Data Transfer Size Limits
- Exfiltration Over Alternative Protocol (3)
- Exfiltration Over C2 Channel
- Exfiltration Over Other Network Medium (1)
- Exfiltration Over Physical Medium (1)
- Exfiltration Over Web Service (4)
- Scheduled Transfer
- Transfer Data to Cloud Account

Impact
14 techniques

- Account Access Removal
- Data Destruction
- Data Encrypted for Impact
- Data Manipulation (3)
- Defacement (2)
- Disk Wipe (2)
- Endpoint Denial of Service (4)
- Financial Theft
- Firmware Corruption
- Inhibit System Recovery
- Network Denial of Service (2)
- Resource Hijacking
- Service Stop
- System Shutdown/Reboot

FIGURE 1. MITRE ATT&CK matrix lay-out for Enterprise domain: tactics are organized by columns while techniques by rows.

The *Mobile* domain encompasses a total of 66 techniques and 41 sub-techniques, with its primary objective being the behavioral analysis of attacks targeting smartphones and related devices. The widespread use of smartphones to access and share personal and corporate data, such as in the Bring Your Own Device (BYOD) model, has significantly expanded the attack surface and increased associated threats. Consequently, a multitude of attacks have been developed specifically for smartphones, with a growing focus on targeting users' organizations due to the vulnerabilities present in these devices. To address this scenario, the mobile ATT&CK matrix specifically considers attack behavior that exploits common entry points in smartphones, such as Lockscreen Bypass and Supply Chain Compromise (Initial Access). Furthermore, it aims to identify targets that go beyond the user's data, such as Input Injection and Data Encrypted for Impact (Impact).

Industrial Control Systems (ICSs) refer to a domain comprising 78 techniques without any sub-techniques. In the context of modern cyber-physical systems, the convergence of Information Technology (IT) and Operational Technology (OT) is occurring through emerging models of fog and edge computing. This convergence involves connecting the OT infrastructure to the network with processing capabilities, allowing the utilization of AI techniques in certain cases. Consequently, it enables seamless integration of data-centric computing (IT) with monitoring and sensing of the physical world (OT). Although the integration of OT and IT improves the effectiveness and efficiency of industrial processes, it also presents an unprecedented opportunity for adversaries. They can now transition between these two worlds, thereby increasing their potential to disrupt the physical environment by launching cyber attacks. The ICS ATT&CK matrix addresses this IT-OT relationship by encompassing attack behaviors that originate in the IT world, such as Spearphishing attachment, Wireless Compromise, and Internet Accessible Devices (Initial access). These attacks culminate in physical impacts within the OT world, such as Damage to Property, Denial of Control, and Loss of Safety (Impact).

The ATT&CK matrix undergoes updates approximately every 6 months, with significant changes occurring since its initial launch as a wiki in 2015. At the time of writing this manuscript, the matrix has reached version 12. One notable major change was the introduction of PRE-ATT&CK in 2017, which encompasses tactics and techniques for describing adversarial behavior before the actual deployment of an attack. Subsequently, PRE-ATT&CK was integrated into the ATT&CK framework as the *Reconnaissance* and *Resource Development* columns. Furthermore, in July 2017, Linux and macOS were incorporated into the framework. In October 2019, ATT&CK for cloud was added, specifically addressing cloud-based Infrastructure as a Service (IaaS). Finally, in 2020, ATT&CK for Industrial Control Systems (ICSs) was introduced.

III. RELATED WORK

During recent years, MITRE ATT&CK framework and dataset have shown the potential to be used for various applications, such as modeling cyber threats, risk assessments, and mainly for cyber attack detection, prediction, prevention, and mitigation. In this section, we briefly review MITRE ATT&CK framework and its dataset usage within recent research work.

Researchers have widely used MITRE ATT&CK in their projects. Wang et al. in [13] introduce a three-layer model, including the goal, behavior, and capability layers, to analyze the statistical characteristics of the APT groups. For this purpose, they use the MITRE ATT&CK dataset to construct a knowledge-base at a behavioral level to calculate similarities among APT groups in order to discover potential communities. This helps security experts to quickly analyze cyber attack behaviors by constructing the whole picture of the attacks. Nisioti et al. in [14] use Bayesian games based on the game theory framework to study the interaction between a cyber forensic investigator and a cyber security attacker. Accordingly, a game-theoretic decision support framework is introduced. They use MITRE ATT&CK Structured Threat Information Expression (STIX) [15] repository as one of the sources of threat reports to design case studies to evaluate their proposed model. In [16], the same authors introduce a data-driven decision support framework, called DISCLOSE, for cyber forensic investigation optimization. DISCLOSE is evaluated using MITRE ATT&CK knowledge-base and the Common Vulnerability Scoring System (CVSS). In [17] leverage the machine-readable representation of the STIX 2.1 model of MITRE ATT&CK Groups knowledge-base in order to query various types of contextual and threat intelligence information, such as group motivations, countries of origin, and targeted sectors and countries. They used this information to study the behaviors and activities of the groups.

Some research works have correlated MITRE ATT&CK with other information sources. Wu et al. in [18] propose a framework, called GroupTracer, to automatically observe and predict complex IoT attacks based on ATT&CK matrix. The proposed framework identifies attack activities using IoT honeypots and retrieves complementary information from the corresponding logs. Next, it maps the attack behaviors to the ATT&CK matrix to perform automatic extraction of the TTP profiles. An Industrial Control Systems Threat Hunting Framework (ICS-THF) for the early detection of cyber threats of ICS is proposed in [10]. In the proposed framework, Cyber Threat Intelligence (CTI) is extracted from the actions of identified adversaries. Straub in [19] introduce the Blackboard Architecture Cyber Command Entity Attack Route (BACCER) which is a generalized approach to model and analyze framework/paradigm-based attacks. In this work, the authors propose to combine rules and facts that demonstrate attack types and the corresponding decision-making logic with actions. Ampel et al. in [3] propose a self-distillation model called the CVE Transformer (CVET) that aims to label CVEs based on information from the MITRE ATT&CK knowledge-base. CVET includes a fine-tuned distillation model that is the result of using a pre-trained language model called RoBERTa [20]. Huang et al. in [8] propose a MITRE ATT&CK-based malicious behavior analysis system (MAMBA) incorporating ATT&CK knowledge-base into neural network models for MS Windows malware detection. This work exploits Open Source Intelligence (OSINT) for the analysis of several malware. The goal is to discover malicious activities and corresponding TTP by analyzing the execution trace associated with the malware under the Windows operating system. In [21], the authors propose a fingerprint for mobile-sensor APT detection framework (FORMAP) based on the correlation between the MITRE ATT&CK knowledge-base, the mobile sensors, and the attack trees. The goal of the proposed framework, namely FORMAP, is to improve security awareness to detect APT attacks on smartphones. Chierzi and Mercês in [22] propose an approach to keep track of the evolution path of the techniques and capabilities employed by IoT Linux malware. For this purpose, they leverage MITRE ATT&CK matrix to discover and characterize current threats (also covering pre- and post-exploitation aspects) and provide useful insights. Fairbanks et al. in [23] propose an approach to automate the identification and location of TTP in a Control Flow Graph (CFG) by using Graph Machine Learning techniques on Android Malware. The authors believe that identifying sub-graphs in the malware CFG provides insights about malware behavior and corresponding mitigation strategies. Georgiadou et al. in [24] propose to correlate a well-defined set of characteristics that includes organizational and individual characteristics with the MITRE ATT&CK framework TTP and vulnerability databases. In this work, a multidisciplinary security culture framework covering attack patterns of the MITRE ATT&CK matrix, designed for critical infrastructures is being evaluated. Kwon et al. in [6] propose an attack-defense mapped framework called Cyber Threat Dictionary (CTD), linking MITRE ATT&CK matrix to the NIST framework. The two major components of CTD are its search engine, which provides details and solutions to deal with an attack, and its suggestion component, which issues appropriate solutions. Legoy et al. in [25] propose an automated cyber threat report analysis tool called rcATT, to extract valuable security information from Cyber Threat Reports (CTRs). In order to automate information extraction from CTRs, the authors investigate multilabel text classification techniques and map them to the MITRE ATT&CK tactics and techniques. Mavroeidis and Bromander in [26] use MITRE ATT&CK knowledge-base to propose a CTI model enabling security analysts to analyze cyber threats. The authors discuss a significant need for an ontology for CTI, and list existing difficulties toward proposing such an ontology, such as lack of standard data representation, comprehensive terminology list, and their layered hierarchical relationships. Parmar and Domingo in [27] describe NATO Allied Command Transformation (ACT) and NATO Communication and Information Agency (NCI Agency) experiments that focus on using deception techniques, such as honeypots, to capture and analyze attackers' activities data. In particular, the MITRE ATT&CK knowledge-base has been utilized during the experiments to identify known adversary patterns and to be correlated with the collected traffic in order to provide the required information for responsible commanders. Kim et al. in [11] propose an automated methodology to classify mobile malware using MITRE ATT&CK knowledge-base. For this purpose, they utilize mathematically vectorized ATT&CK matrix for each TPP to calculate similarities and compare Indicators of Compromise (IoC). Consequently, they provide complementary insights and reduce false positive phenomena.

MITRE ATT&CK knowledge-base has been employed in risk assessment and threat modeling research [28], [29], [30], [31], [32], [33]. Ahmed et al. in [34] propose a hybrid cyber risk assessment methodology based on (i) the vulnerability-oriented approach, (ii) asset/impact-oriented approach, (iii) threat-oriented approach. This work combines the MITRE ATT&CK repository with attack graphs to calculate the probability of attacks and their potential success rate. Xiong et al. in [35] uses the MITRE Enterprise ATT&CK matrix to propose a threat modeling language, called enterpriseLang, for enterprise security. The proposed threat modeling language, which is a Domain Specific Language (DSL), is designed based on the Meta Attack Language (MAL) framework [36] and concentrates on representing system assets, asset associations, attack phases, and defense strategies. Choi et al. in [37] propose an automatic attack sequence generation approach consistent with MITRE ATT&CK tactics and techniques to define and deploy an attack dataset. In this research, the authors consider three aspects of a qualified attack sequence generation: (i) reproducibility, (ii) diversity, and (iii) reality. Accordingly, they propose an attack sequence generation approach that uses Hidden Markov Models (HMM) as its core component. The authors of [38] leverage the MITRE ATT&CK knowledge-base to propose an Attack Specific Language (ASL) that provides a unified representation for all cyber-threat scenarios. ASL provides information and knowledge about attack techniques in a brief format that streamlines and automates the malicious activities of threat and challenge execution. Pell et al. in [39] proposes an approach to extend MITRE ATT&CK knowledge-base by adding a list of 5G related cyber-threat techniques. This research investigates the knowledge sharing between early 5G network threat analysis and adversarial TTP of MITRE ATT&CK knowledge-base, and it proposes solutions to bridge these gaps. It analyzes the knowledge and theoretical shortcomings of current 5G technology enablers (e.g., Software-defined networking (SDN) and network functions virtualization (NFV)). A formal method that models cyber security concepts from both the attacker and the defender points of view is proposed in [40]. For this purpose, two persistent graphs are generated: (i) the execution of the attacker's procedures mapped to the MITRE ATT&CK framework, (ii) the exposed resources during the execution of the attacker's procedures. The graphs help to reduce false positive alarm rates for defenders.

IV. DATASET AND TERMINOLOGY

In this section, we introduce the terminology that we will adopt in the remainder of the paper while describing the details of our dataset. Our dataset has two main tables: (i) Threat Profiles and (ii) Threat Profiles Relationships. The dataset generated for the analysis proposed in this work can be requested by email to the authors. First, we describe the dataset preparation procedure. Next, we define the columns of each table.

A. DATASET GENERATION

We considered the MITRE ATT&CK dataset Version 12.1 for *Enterprise*, *ICS* and *Mobile* as Microsoft Excel files. Then, we applied the following steps to prepare our datasets.

For the Threat Profiles dataset:

1) Starting from the relationship sheet, we filtered *Source Type* column and selected *Campaign*, *Group* and *Software* values. Next, we filtered *Mapping* column and selected *Uses* value. Finally, we filtered *Target Type* column and selected *Technique* value.
2) We copied the following columns into our Threat Profiles dataset: *Source ID* as *Threat Profile ID*, *Source Name* as *Threat Profile Name*, *Source Type*, and *Target ID* as *Technique ID*. Next, we removed the duplicates, add a new column *Domain*, and populated it using values from the input dataset.
3) We created *Tactic* and *Platform* columns and filled their values from the input dataset by mapping them to techniques.
4) By analyzing the description of the threat groups, we extracted and created *Threat Actor Country/Region*, *Targeted Country/Region*, *Targeted Sector* and *Motivation~~s~~* columns. For *Targeted Sector* and *Motivation*, we normalized and created a list of possible sectors and common motivations.
5) We created two new columns *Threat Actor Region* and *Targeted Region* and mapped them to each country/region from selected naming conventions of regions. We removed *Threat Actor Country/Region* and *Targeted Country/Region* because they will not be used for our analysis.

Next, for our Threat Profile Relationships dataset preparation, we performed the following steps:

1) From MITRE ATT&CK datasets, we extracted the following columns. From the Relationship sheet, we filtered the *Source Type* column and selected *Campaign*, *Group* and *Software* values. Next, we filtered the *Mapping* column and selected the *Uses* value. Finally,

we filtered the *Target Type* column and selected the *Software* and *Group* values.

2) We copied *Source Name* as *Threat Profile Name*, *Source Type* as *Threat Profile Type*, *Target ID* as *Target Profile ID*, and *Target Name* as *Target Profile Name*.
3) We created the *Domain* column and populated it with the corresponding values: *Enterprise*, *ICS*, or *Mobile*.

B. THREAT PROFILES

An example of the Threat Profiles table is shown in Table 1. Each row represents a threat profile (9 in the example), while the columns represent the attributes:

- **Threat Profile ID**: the ID that MITRE assigns to each threat profile. The naming convention of ID is the type of threat profile (*G* as *Group*, *S* as *Software*, and *C* as *Campaign*) and a four-space numerical value. For example, *C* for *Campaign* and 0001 form the first threat profile ID: *C0001*.
- **Threat Profile Name**: the name that MITRE has assigned or taken from other sources. There may be other names or aliases for the same profile, especially the threat *groups/actors* where multiple vendors/researchers are attributing by time to the groups.
- **Threat Profile Type**: it contains the threat profile type that MITRE has categorized: (i) *Group* also known as Threat Actor, (ii) *Software* which indicates tools such as malware, (iii) *Campaign* to describe any collection of intrusion activity carried out during a defined time frame targeting the same objectives and entities.
- **Technique ID**: this column represents techniques used in a threat profile. The naming convention is *T* for *Technique* and four numbers as an incremental value (e.g. *T1119*).
- **Domain**: this column represents the domains from MITRE ATT&CK main datasets. There are three domains: (i) *Enterprise*, (ii) *Mobile* and (iii) *ICS*.
- **Tactic**: the value representing the main goal each technique tries to achieve. To represent techniques spread over multiple tactics, we have generated multiple rows in the table.
- **Platform**: it represents the operating system or the application that a technique utilizes (e.g. *Cloud Administration Command* is only applicable in Azure AD, and IaaS).
- **Threat Group Region**: this column lists countries suspected of hosting threat groups. Based on the MITRE descriptions, we have mapped countries to regions (e.g., Japan is mapped to region East Asia). In particular, we extracted the columns *Threat Group Region*, *Targeted Region*, *Targeted Sector*, and *Motivations* from only threat groups (neither for campaigns nor software) as conductors of cyber attacks.
- **Targeted Region**: this column lists the regions targeted by threat groups.
- **Targeted Sector**: this column lists the targeted sector/industry by Threat Groups. We extracted, normalized, and mapped more than 110 unique mentions of industries/sectors to 34 general ones (e.g., telecom, telecommunication, and information technology are mapped to the general sector of *Information and Communication Technology*).
- **Motivation**: the column represents the motivations behind the threats described by the MITRE researchers in the description of threat groups. We have generalized and mapped them into 5 main categories: (i) *Espionage*, (ii) *Financial*, (iii) *Cyber Warfare / Destruction*, (iv) *Political or Ideological*, (v) *Not specified*.

C. THREAT PROFILES RELATIONSHIPS

This table concentrates on the relationship between different types of threat profiles to provide more context to our analysis. Table 2 shows this table and its header data.

- **Threat Profile Name**: the name of the source threat profile.
- **Threat Profile Type**: as mentioned above, it contains the threat profile type that MITRE has categorized: (i) *Group*, (ii) *Software*, (iii) *Campaign*.
- **Related/Target Threat Profile ID**: the ID of related/target threat profile that is in relation to the main *Threat Profile*.
- **Related Threat Profile Name**: the name of the related/target threat profile.
- **Domain**: as mentioned above, this column represents the domains from MITRE ATT&CK main datasets, namely: *Enterprise*, *Mobile* and *ICS*.

V. METHODOLOGY

In this section, we describe the methodology we adopted to analyze the MITRE ATT&CK dataset. Fig. 2 shows the conceptual architecture of the MITRE ATT&CK dataset. There are *Many-to-Many* relationships among the components of the framework. Each threat group can be related to multiple software profiles and campaigns and vice-versa. Campaigns are led by threat actors (groups or individuals). Groups and campaigns utilize one or more software to implement their threat scenarios.

Only one campaign (*C0024: SolarWinds Compromise*) is in relation to the group threat profiles. 13 campaigns are related to software threat profiles, and one campaign does not have a relationship with other threat profiles. The domain layer defines the type of infrastructure that an APT applies to, that is, *Enterprise*, *ICS* or *Mobile*, while the *Platform* layer mainly specifies the type of operating system targeted by an APT.

MITRE ATT&CK includes 22 platforms, such as different Operating Systems (*Windows*, *Linux*, *macOS*), network operating system, Infrastructure as a Service (IaaS), Software-as-a-Service (SaaS), *ICS* environments, *Mobile* environments, etc. that have been targeted by threat profiles. In terms of the number of tactics and techniques, MITRE ATT&CK Version 12.1 constitutes 16 tactics and 309 unique techniques

TABLE 1. Threat profiles table head rows.

Threat Profile ID	Threat Profile Name	Threat Profile Type	Technique ID	Domain	Tactic	Platform	Threat Actor Region	Targeted Region	Targeted Sector	Motivation
C0001	Frankenstein	campaign	T1119	Enterprise	Collection	IaaS	N/A	N/A	N/A	N/A
S0001	Trojan.Mebromi	software	T1542	Enterprise	Persistence	macOS	N/A	N/A	N/A	N/A
S0002	Mimikatz	software	T1555	Enterprise	Credential Access	Linux	N/A	N/A	N/A	N/A
G0018	admin@338	group	T1566	Enterprise	Initial Access	Windows	East Asia	Not specified	Financial	Espionage
G0018	admin@338	group	T1566	Enterprise	Initial Access	macOS	East Asia	Not specified	Financial	Espionage
G0130	Ajax Security Team	group	T1056	Enterprise	Collection	Linux	Middle East	North America	Security and Defense	Not specified
S0089	BlackEnergy	software	T0859	ICS	Persistence	DataHistorian	N/A	N/A	N/A	N/A
G0112	Windshift	group	T1521	Mobile	Command and Control	iOS	Not specified	Middle East North Africa	Critical Infrastructure	Espionage
G0026	APT18	group	T1078	Enterprise	Defense Evasion	AzureAD	Not specified	Not specified	Information and Communication Technology	Not specified

TABLE 2. Threat profiles relationships table head rows.

Threat Profile Name	Threat Profile Type	Target Profile ID	Target Profile Name	Domain
Wizard Spider	group	S0552	AdFind	Enterprise
Wizard Spider	group	S0534	Bazar	Enterprise
Wizard Spider	group	S0521	BloodHound	Enterprise
Wizard Spider	group	S0154	Cobalt Strike	Enterprise
Wizard Spider	group	S0575	Conti	Enterprise
Wizard Spider	group	S0024	Dyre	Enterprise
Wizard Spider	group	S0367	Emotet	Enterprise
Wizard Spider	group	S0363	Empire	Enterprise
Wizard Spider	group	S0632	GrimAgent	Enterprise

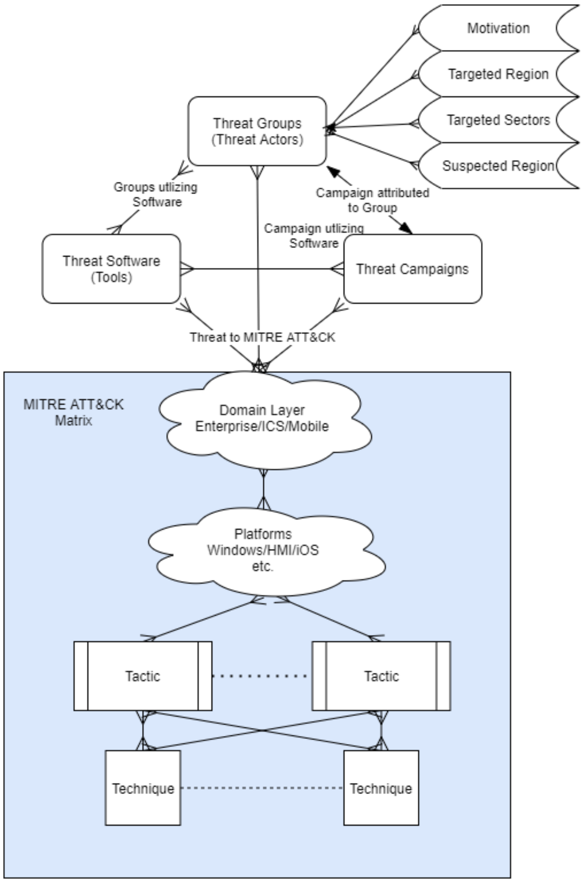


FIGURE 2. MITRE ATT&CK conceptual architecture.

observed over 860 Threat Profiles. 360 techniques are observed in the entire spectrum of tactics.

In the following section, we investigate and analyze the relationships between the components of this dataset and describe our findings.

VI. MITRE ATT&CK ANALYSIS

This section describes our statistical analysis of the MITRE ATT&CK dataset and extracted insights. Each subsection is devoted to one specific analytical perspective or dimension. In the following, we explain each dimension in hierarchical order.

A. THREAT PROFILES

MITRE ATT&CK is characterized by three threat profile types: (i) Group, (ii) Campaign, (iii) Software. Software collects the largest number of threat profiles, while Campaign represents the smallest. Fig. 3 illustrates the count of threat profiles as a function of their types. The total number of 865 threat profiles comprises 717 Software, 134 Groups, and 14 Campaigns.

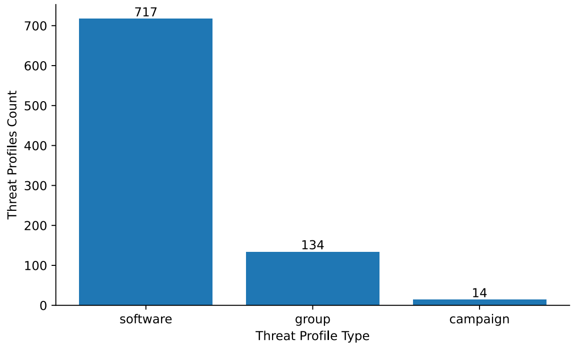


FIGURE 3. Number of threat profiles as a function of their types.

Groups and campaigns may deploy software to implement various types of threats. Fig. 4 shows the top 20 most used Software threat types in terms of their adoption in threats campaigns and groups.

- *Mimikatz* has been used by 42 threat profiles to extract credentials in *Windows* platform, which indicates the objective of the threat profiles towards persistence or/and lateral movement.
- *PsExec* and *Net* are popular software among threat profiles. They are pre-built in *Windows OS* with various capabilities leading to achieve various objectives, such as *Lateral Movement*, *Persistence*, *Defense Evasion*, etc. It indicates that threat groups prefer those tools that do not require any malicious download in order to avoid being detected by the underlying network and instead utilize valid tools to achieve the same goals.
- Although there are very low numbers of campaigns in the dataset, they are present among the top 20 software, and affect the most used popular software. For example, *Systeminfo* is utilized by 3 campaigns, whereas without

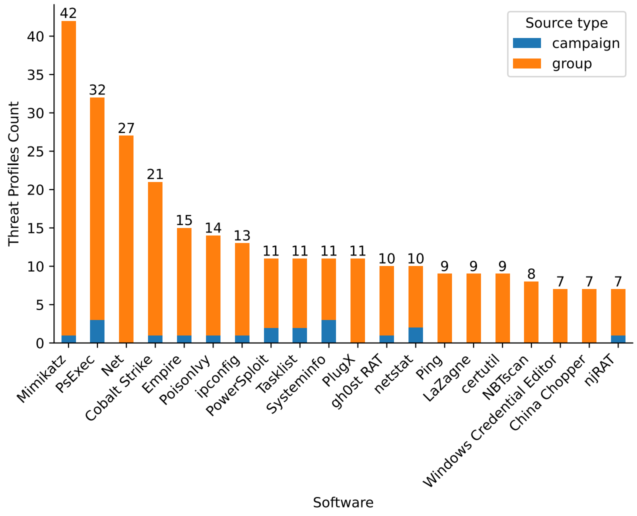


FIGURE 4. Top 20 software as a function of their adoption in campaigns and groups.

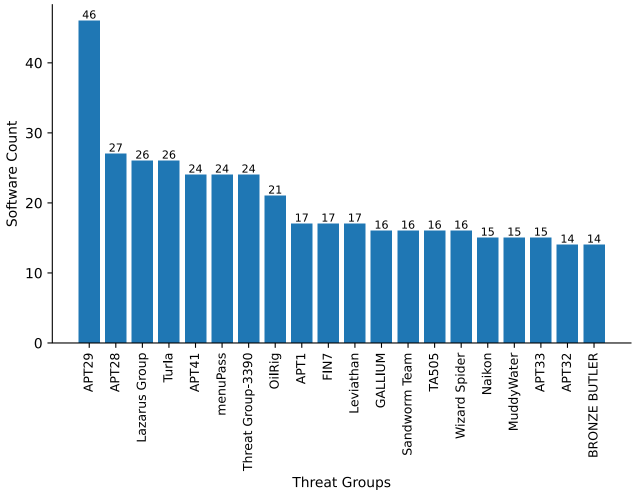


FIGURE 5. MITRE ATT&CK top 20 groups based on software usage.

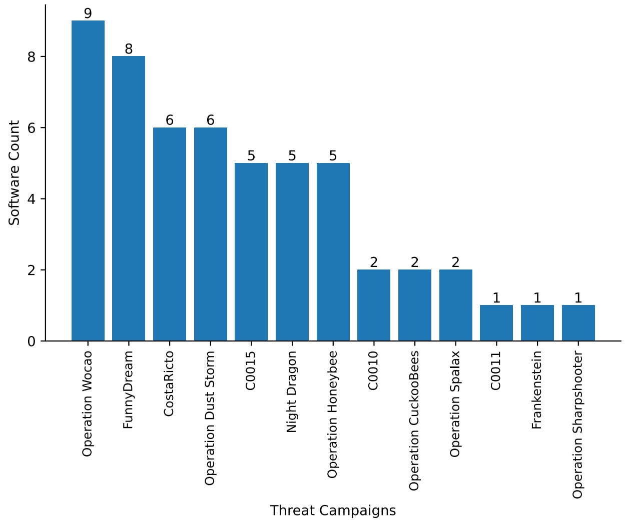


FIGURE 6. MITRE ATT&CK campaigns and software count distribution.

these observations the position of *Systeminfo* would have been degraded from position 10 to 17.

- It is worth mentioning that of the top 20 software, there are 6 malware shared among multiple threat profiles, namely: (i) *China Chopper*, (ii) *Cobalt Strike*, (iii) *PlugX*, (iv) *PoisonIvy*, (v) *gh0st RAT*, (vi) *njRAT*. *China Chopper* is a web-shell malware, and the rest are Remote Access Tool (RAT). Hence, persistence in systems to execute remote commands is the core function of popular malware among threat groups and campaigns.

From a different dimension, Fig. 5 illustrates MITRE ATT&CK top 20 groups based on the number of software usage. Here are the extracted insights:

- The software count shows threat groups' capabilities utilizing or employing various software to achieve their objectives. This indicates their ability to understand historical data and adapt to new procedures in order to achieve their goals.
- APT29 with a significant difference uses the highest number of software (46) compared to the second group (27).
- The top 3 groups (*APT29*, *APT28* and *Lazarus Group*) have utilized 96 unique software, of which only 4 are in common. This reveals their different strategies for conducting malicious activities.

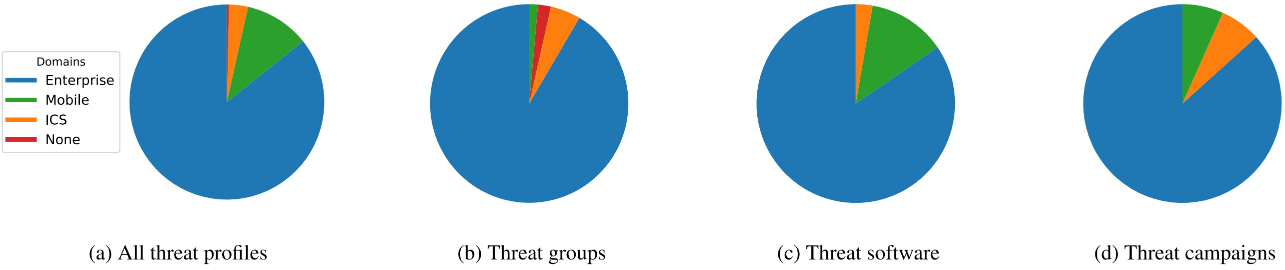
Similarly, Fig. 6 illustrates MITRE ATT&CK campaigns and software count relationships. Due to the low number of campaigns in the dataset (only 14 campaigns), the extracted insights might not be appropriately generalizable. Here are the extracted insights:

- The average software usage count is around 4, while the highest usage count is 9 and the lowest usage count is 0 since there is one campaign that is not in a relationship with *Software* threat profiles. This indicates that the total number of software used by threat campaigns is distributed linearly.
- *Operation Wocao* and *FunnyDream* utilize the highest number of software among all campaigns whereas *C0011*, *Frankenstein* and *Operation Sharpshooter* utilize the lowest number of software among all campaigns. Among the top 3 threat campaigns, there are 4 software in common, none of which is identified as malware.

B. THREAT PROFILES AND DOMAINS DISTRIBUTION

In terms of domain, MITRE ATT&CK has three major domains that represent the type of targeted environments. These domains are: (i) *Enterprise*, (ii) *ICS*, (iii) *Mobile*. Fig. 7 illustrates the domains related to MITRE ATT&CK threat profiles. Our extracted insights are as follows:

- Fig. 7.a represents domains targeted by all threat profiles, *Enterprise* domain is more dominant (85.65%)



(a) All threat profiles (b) Threat groups (c) Threat software (d) Threat campaigns

FIGURE 7. Threat profiles and domain distribution.

compared to *ICS* (3.16%) and *Mobile* (10.85%) domains. This indicates that threat profiles mainly target the *enterprise* domain. The main reason may be the wider platform coverage (e.g., cloud-based, networking, and OS) and its openness to more end-users causing a higher variety of cyber attacks. As of these points, the *Enterprise* domain has been more attractive for threat profiles.

- Fig. 7.a shows 862 threat profiles that have targeted at least one domain. Among them, 25 threat profiles have targeted two domains. It is also worth mentioning that there are 3 threat profiles that targeted the zero domain. This may indicate that the probability for threat profiles to target multiple domains is limited.
- Fig. 7.b represents domains targeted by *Group* threat profiles, 94% of *Group* profiles targeted *Enterprise* domain which is higher than *Software* and *Campaign* threat profiles. This indicates that threat groups have been identified and attributed to relying heavily on targeting enterprises.
- Fig. 7.c represents domains targeted by *Software* threat profiles. Proportionally, *Software* threat profiles have targeted the *Mobile* domain more than *Group* and *Campaign* threat profiles.
- Fig. 7.d shows the domains targeted by *Campaign* threat profiles. In turn, *Campaign* threat profiles have targeted *ICS* domain more than other threat profiles. This indicates that designing attacks towards the *ICS* domain is more complicated and requires more expert knowledge.

C. THREAT PROFILES AND PLATFORMS DISTRIBUTION

MITRE ATT&CK contains information for platforms, such as Windows, macOS, Android, PRE, Azure AD, DataHistorian, Office 365, IaaS, Network, Containers, etc. Figure 8 illustrates MITRE ATT&CK targeted platforms by threat profiles. In the following, we list our insights:

- *Windows*, *Linux* and *macOS* operating systems are the most considered platforms, each targeted by 88% of the threat profiles. *Network* is targeted by 78% of threat profiles. Cloud solutions have been targeted by more than 50%, and *Mobile* operating systems have been targeted by 11% of threat profiles.
- *Groups* mainly targeted enterprise operating systems (95%), cloud solutions (up to 87%), and networking (85%). *ICS* and *Mobile* platforms have not been highly attributed to *Group* threat profiles.
- *Software* threat profiles target enterprise operating system platforms (86%), networking platform (77%), and cloud solution platforms (up to 62%). *ICS* related platforms have rarely been targeted by software threat profiles. Compared to other threat profiles, *Software* threat profiles have targeted mobile operating systems with a higher percentage.
- Similar to *Group* threat profiles and unlike *Software* threat profiles, *Campaign* threat profiles have rarely targeted *ICS* and *Mobile* platforms.
- *PRE* platform also defined as the PRE-attack phase, is mostly utilized by groups and campaigns. Its main goals are collecting information about target environments and developing the necessary capabilities to carry out various cyber attacks.

D. THREAT PROFILES AND TACTICS DISTRIBUTION

As described, MITRE ATT&CK framework has 16 tactics: *Reconnaissance*, *Resource Development*, *Initial Access*, *Execution*, *Persistence*, *Privilege Escalation*, *Defense Evasion*, *Credential Access*, *Discovery*, *Lateral Movement*, *Collection*, *Command and Control*, *Exfiltration*, *Impair Process Control Inhibit Response Function* and *Impact*. Fig. 9 illustrates the threat profiles and tactics distribution observed in MITRE ATT&CK dataset as a function of the domains. We observe that:

- The most used tactics (over 70%) by threat profiles are: *Defense Evasion*, *Discovery*, *Command&Control*. *Defense Evasion* and *Discovery* tactics include the highest number of techniques compared to other tactics. *Command&Control* is one of the main tactics employed in most of the attack scenarios.
- *Reconnaissance* and *Resource-Development* (pre-attack phases) are among the least used tactics due to the wide variety of assumptions. For both of them, the *Enterprise* domain constitutes all the threat profiles.
- *Impair-Process-Control* and *Inhabit-Response-Function* are the two tactics dedicated to *ICS* domain because they are used mainly for operational technologies. The

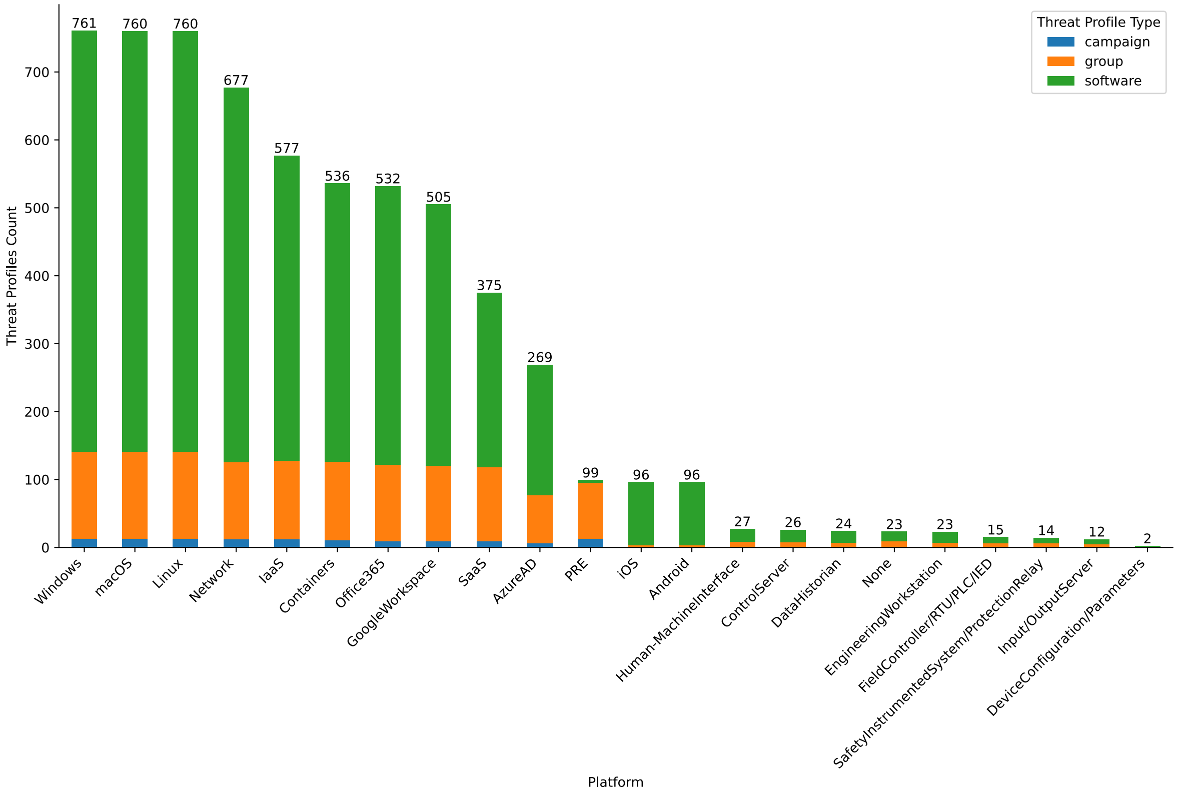


FIGURE 8. Number of threat profiles as a function of the type of platform.

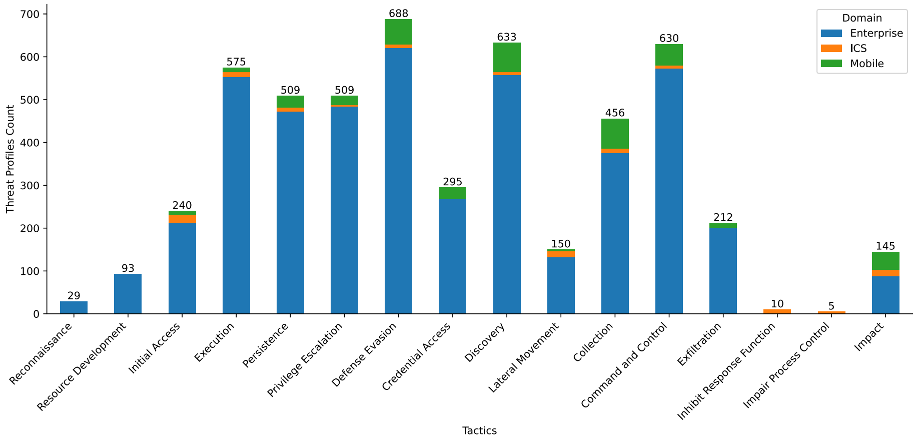


FIGURE 9. All threat profiles and tactics distribution by domains.

ICS domain is not present in *Credential-Access* or *Exfiltration* tactics because these two tactics are not applicable to *ICS* domain.

- Except for two tactics that are mainly dedicated to *ICS* domain, For most of the tactics, *Enterprise* domain constitutes the highest percentage of threat profiles.
- The top 3 tactics used by the *Mobile* domain are: *Defense-Evasion*, *Discovery* and *Collection*. The main reason behind the *Collection* tactic is that it can be used for all types of users and social categories that show potential for malicious activities, such as information collection.

Figure 10 illustrates the threat profile and tactics distribution observed in MITRE ATT&CK dataset representing threat profile types too.

- The most used tactics from *Group* threat profiles with more than 80% are *Defense Evasion*, *Execution* and *Initial Access*.
- *Exfiltration* and *Impact* are among the least used tactics although they are considered essential phases of cyber

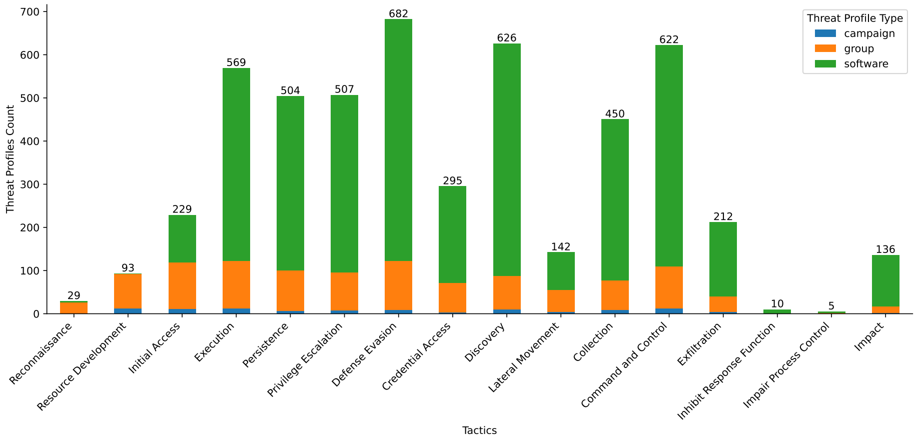


FIGURE 10. All threat profiles and tactics distribution by threat types.

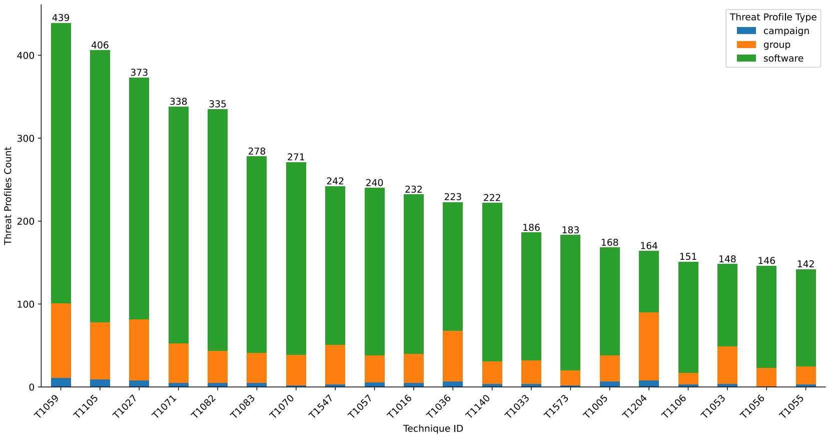


FIGURE 11. All threat profiles and techniques distribution by domains.

attacks. One potential reason may be the resistance to information sharing as a result of reputational side effects, particularly for enterprises.

- Tactics related to the pre-attack phase along with tactics that are unique to *ICS* domain have not been highly utilized by *software*. In addition, *Initial-Access* and *Lateral-Movement* are among the least employed tactics by *software* threat profiles although they are still popular among the threat groups that use these software, such as *Mimikatz*, *PsExec*, *Cobalt Strike*.
- *Defense-Evasion*, *Discovery* and *Command&Control* are the tactics used most frequently by more than 500 software. This indicates the amount of effort and unique tools required to achieve such tactics.
- Top utilized tactics by *Campaigns* are *Resources-Development*, *Execution* and *Command&Control*. Furthermore, what may be interesting is that *Resources-Development* is the highest (percentage-wise) detected in *Campaigns* compared to other types of threat profiles. This is due to the nature of campaigns that take into account the time period and common objectives/targets.

E. THREAT PROFILES AND TECHNIQUES DISTRIBUTION

Techniques represent *how* an adversary achieves a tactical goal by performing an action. For example, an adversary may dump credentials to achieve credential access. Fig. 11 illustrates MITRE ATT&CK threat profiles and the distributions of the top 20 techniques. Table 3 lists these top 20 techniques.

- The top 5 techniques in order are: (i) *Command and Scripting Interpreter*, (ii) *Ingress Tool Transfer*, (iii) *Obfuscated Files or Information*, (iv) *Application Layer Protocol*, (v) *System Information Discovery*. These techniques have been employed by more than 38% of the threat profiles and are assigned unique tactics.
- 8 tactics are the corresponding tactics of top 20 techniques: (i) *Defense-Evasion* (5 times), (ii) *Discovery* (5 times), (iii) *Execution* (4 times), (iv) *C&C* (3 times), (v) *Privilege-Escalation* (3 times), (vi) *Persistence* (2 times), (vii) *Collection* (2 times), (viii) *Credential-Access* (1 time).
- Two examples of techniques that played a key role for threat groups, in particular, are: (i) T1204

TABLE 3. Top 20 techniques.

Rank	Technique ID	Count of Threat Profiles
1	T1059	439
2	T1105	406
3	T1027	373
4	T1071	338
5	T1082	335
6	T1083	278
7	T1070	271
8	T1547	242
9	T1057	240
10	T1016	232
11	T1036	223
12	T1140	222
13	T1033	186
14	T1573	183
15	T1005	168
16	T1204	164
17	T1106	151
18	T1053	148
19	T1056	146
20	T1055	142

(*User-Execution*) moved from the 16th place to the 2nd place to represent the importance of achieving the execution tactic for threat groups, (ii) T1566 (*Phishing*) moved from 23rd place to the 3rd place which indicates the preference of threat groups towards using this technique. It should be mentioned that techniques related to *Resource Development*, *Initial access* and *Lateral Movement* tactics are of great importance to threat groups.

- For Software threat profiles, two new techniques escalated into the top 20s are: T1113 (*Screen Capture*) and T1543 (*Create-or-Modify-System-Process*). This indicates the importance of tactics, such as *Collection*, *Persistence* and *Privilege Escalation* for Software threat profiles.
- Similar to group threat profiles, Campaign threat profiles prefer utilizing T1566 (*Phishing*) technique to achieve *Initial access*. Unlike other threat profiles, T1583 (*Acquire-Infrastructure*) and T1588 (*Obtain Capabilities*) emerge within the top 20 techniques only for Campaign threat profiles showing the importance of *Resource Development* tactic for them.

As mentioned above, threat profiles utilize various techniques to conduct their malicious activities. For example, Fig. 12 illustrates the number of unique techniques used by all threat profiles. As we see, 6 threat profiles use more than 60 unique techniques (5 Groups and 1 Software) indicating the complexity of their malicious scenarios and their significant capabilities. Around 147 threat profiles use only three or fewer unique techniques, which represent their level of simplicity and selective tactic set to achieve.

Figure 13 illustrates the number of occurrences of MITRE ATT&CK techniques within all threat profiles. The highest occurrences (e.g., more than 300 occurrences) are very sparse and limited compared to the lowest occurrences part of the graph, which is highly concentrated. This provides information for cyber security teams to build and enhance their security controls against high-risk techniques.

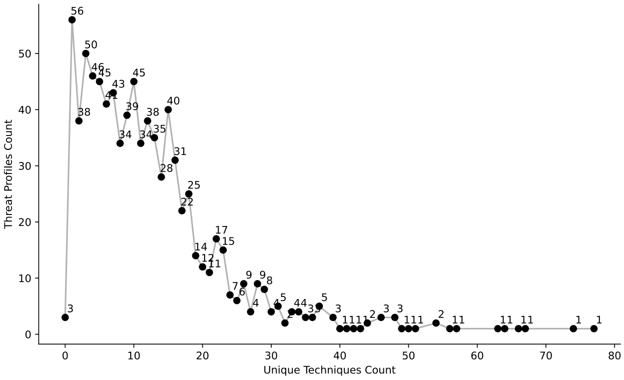


FIGURE 12. Threat profiles count over the number of unique techniques.

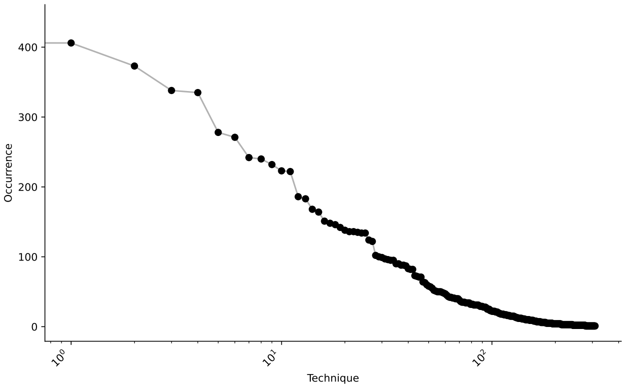


FIGURE 13. Frequency of MITRE ATT&CK techniques over threat profiles.

F. TACTICS AND TARGETED PLATFORMS DISTRIBUTION

Each attack scenario requires specific platforms and selected tactics in order to carry out its activities. This section describes the tactics and corresponding platforms within the MITRE ATT&CK dataset. Fig. 14 illustrates MITRE ATT&CK targeted platforms over tactics. In the following, we list our extracted insights:

- *Initial-Access* tactic has been utilized for 20 platforms (out of 22) as the top tactic. There are 2 platforms that *Initial-Access* tactic is not applicable to: (i) *DeviceConfiguration/Parameters* representing system settings/configurations, (ii) *PRE* that represents the pre-attack phase.
- *Persistence* and *Lateral Movement* tactics are the 2nd and 3rd highest tactics, respectively. These tactics are highly related to *Initial-Access* as shown in Fig. 15. Other tactics that overlap with their compatible platforms are *Defense-Evasion* and *Discovery*.
- *Exfiltration* tactic was observed on 6 platforms and *Impact* tactic was observed in 16 platforms. Both tactics are considered important objectives in most cyber attacks.

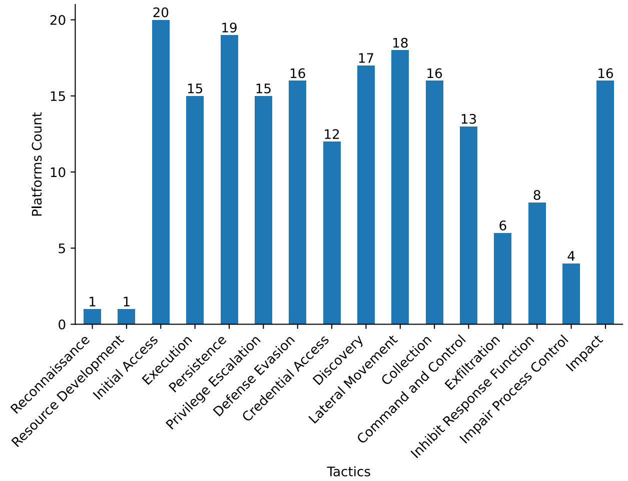


FIGURE 14. Tactics and platforms distribution.

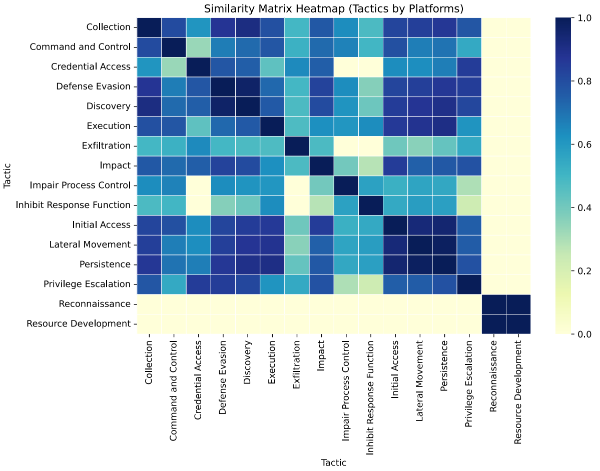


FIGURE 15. MITRE ATT&CK tactics measured by platform types heatmap.

G. TECHNIQUES AND TARGETED PLATFORMS DISTRIBUTION

As another analytical phase, we analyze techniques and the corresponding targeted platforms. Figure 16 illustrates MITRE ATT&CK techniques frequency and targeted platforms distribution, and Fig. 17 illustrates the corresponding heatmap representation. In the following, we list our extracted insights.

- Unique techniques that use *Windows* Operating System (OS) are 50% while for *Linux* and *macOS* are around 43%. As expected, popular OSs have been targeted by most of the techniques. As 17 illustrates, these platforms are similar not only in terms of frequency but also in terms of the type of techniques.
- The second category is mobile operating systems, which have been targeted by only 10% of threat profiles. This category has more diverse techniques than other types of platforms, such as ICS and Cloud platforms.

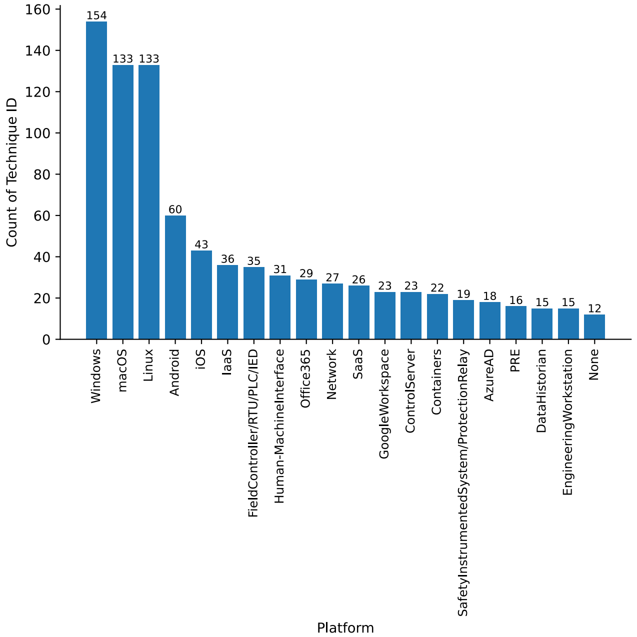


FIGURE 16. The count of techniques and targeted platforms distribution.

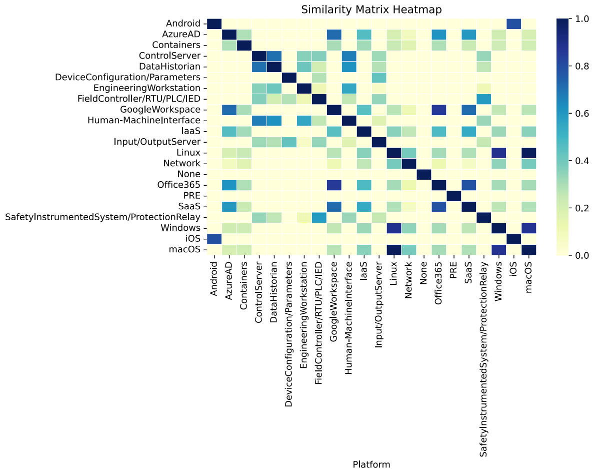


FIGURE 17. The targeted platforms measured by types of techniques heatmap.

- Figure 17 illustrates the similarity between other platforms as well. For instance, *Office365* and *GoogleWorkspace* have higher similarity due to the types of services. The same description is applied to *GoogleWorkspace* and *AzureAD*.

H. TACTICS-TECHNIQUES DISTRIBUTION

In MITRE ATT&CK framework, each tactic represents a particular objective of a group of techniques. Figure 18 illustrates the distribution of MITRE ATT&CK unique number of techniques for each tactic. Here are the extracted insights:

- In terms of the number of techniques, the *Defense-Evasion* tactic with 56 techniques includes the highest number of techniques. This tactic, as demonstrated in the previous sections, is the most utilized tactic by threat profiles. *Discovery* and *Collection* tactics with more than 38 techniques are among the most diverse tactics.

- Although *Impact* tactic is one of the most diverse tactics that includes various techniques, it has not been widely used by threat profiles, unlike other diverse tactics.
- We see the lowest number of techniques for *Pre*-attack phase (i.e. *Reconnaissance* and *Resource Development*) and ICS (i.e. *Inhibit-Response-Function* and *Impair-Process-Control*) tactics. Among the lowest, we observed *Exfiltration* tactic despite being considered one of the main outcomes and the last phase of most cyber attacks.

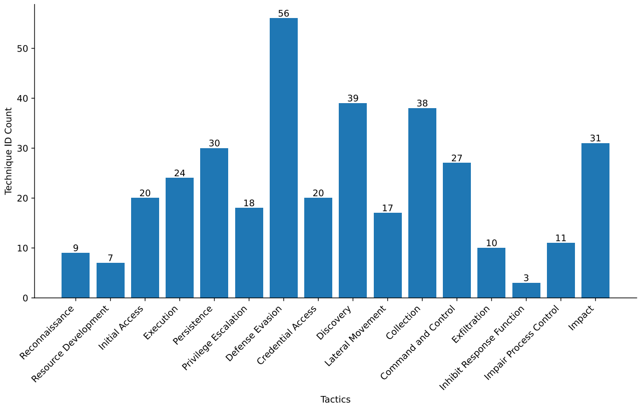


FIGURE 18. **Tactics and techniques distribution.**

Figure 19 illustrates MITRE ATT&CK techniques occurrence over multiple tactics. In the following, we list our insights.

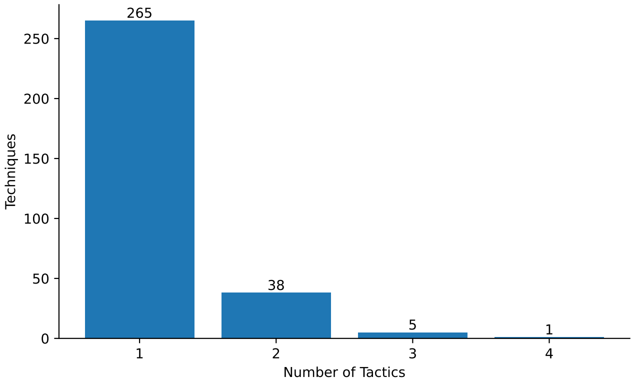


FIGURE 19. **Techniques occurrence over multiple tactics.**

- In general, there are 338 techniques in the MITRE ATT&CK framework. Observed techniques are 309 (*Enterprise* = 180, *Mobile* = 61, *ICS* = 68) and not yet observed are 29 techniques (*Enterprise* = 13, *Mobile* = 5, *ICS* = 11).
- To understand techniques and their overlaps in different tactics, the likelihood that an observed technique achieving multiple objectives is 14%. This knowledge can be valuable to security experts while analyzing intrusions in order to design appropriate countermeasures and defense strategies.

I. THREAT GROUPS MOTIVATIONS

In this section, we describe MITRE ATT&CK dataset threat groups' motivations for conducting various threats. Fig. 20 illustrates MITRE ATT&CK threat groups motivations. Here are the extracted insights:

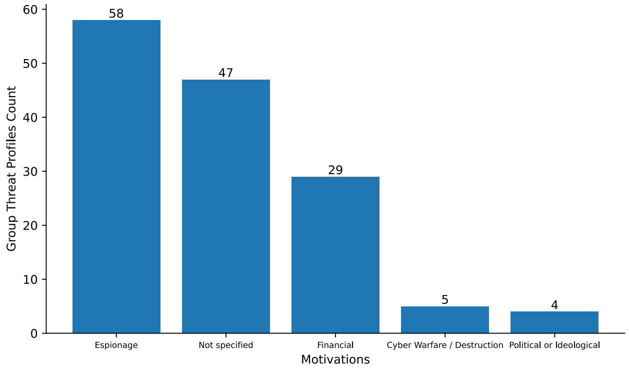


FIGURE 20. **Threat groups motivations.**

- 43% of threat groups had *Espionage* as one of their main motivations to carry out various cyber crimes. However, 70% of the threat groups that have *Espionage* as their motivation do not include information about using *Exfiltration* tactic. This tactic is an eventual phase of *Espionage* cyber attacks.
- There are 47 threat groups whose motivations have not been clearly identified. Consequently, *Financial* motivation is considered as the second motivation with 29 threat groups. Compared to the other motivations that targeted more than 20 platform types, unexpectedly, *Financial* motivation only targeted around 60% of platforms not including any *ICS* and *Mobile* platforms.
- Overall, the main concentration of threat profiling scope for MITRE researchers is focused on the type of threat groups. As illustrated, *Financial* motivations count significantly less than *Espionage* motivations.
- *Cyber Warfare/Destruction* motivation is associated with only 5 threat groups. Subsequently, *Political or Ideological* motivation is associated with 4 threat groups. Both are the only motivations that used all the tactics and are only dedicated to ICS related tactics (*Inhibit-Response-Function* and *Impair-Process-Control*).
- In particular, *Mobile* platforms have been targeted only by threat groups that followed *Espionage* motivations.
- There are 64% threat profiles that have at least one motivation. This is an essential factor in understanding the main objective of any cyber offense.

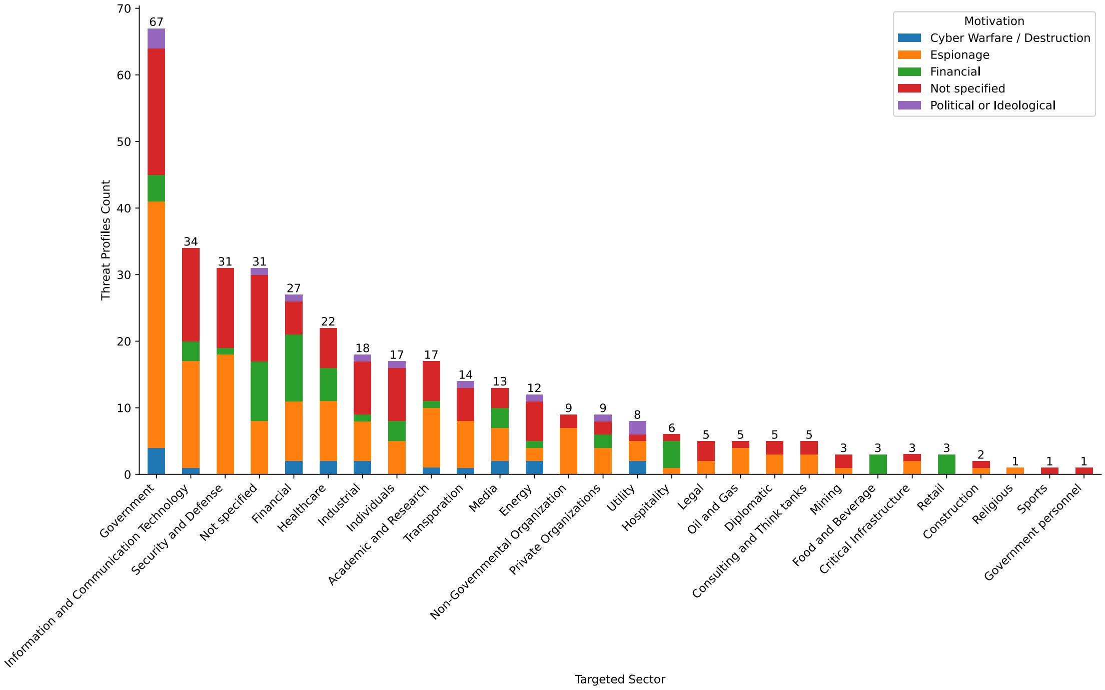


FIGURE 21. Targeted sectors by threat groups.

J. TARGETED SECTORS BY THREAT GROUPS

In this section, we describe the sectors targeted by threat groups. Having information related to targeted sectors provides valuable road-maps for security experts. Fig. 21 illustrates MITRE ATT&CK threat groups targeted sectors while highlighting motivations. Here are the extracted insights:

- *Government* sector is highly targeted with 50% of threat groups committing various cyber offenses against this sector. Most *Espionage* related threat groups target critical/sensitive information sources of *Government* and *Security-and-Defense* sectors despite spreading over 23 sectors.
- Noticeably, the second sector is *Information and Communication Technology* with almost 50% of the threat groups numbers that targeted the *Government* sector. This indicates the importance of this sector for attackers.
- *Financial* motivation is spread over more than 15 sectors. Obviously, the highest occurrence of this motivation is over the *Financial* sector. Among all threat groups with *Financial* motivation, 17% targeted at least one of the highly critical sectors, such as *Healthcare*, *Energy* or *Industrial*.
- 11 sectors have been targeted by *Cyber-Warfare-and-Destruction* motivation including critical sectors, such as *Healthcare* and *Utility*. The corresponding threat groups reveal their potential capabilities targeting various critical sectors.
- 9 sectors have been targeted with *Political-or-Ideological* motivation that is highly interesting to hacktivists. 50% of the corresponding threat groups have targeted *Cyber-Warfare-and-Destruction* motivation too.

K. TARGETED REGIONS BY THREAT GROUPS

From a different perspective, in this section, we analyze the MITRE ATT&CK dataset regarding targeted regions by threat groups. Fig. 22 illustrates MITRE ATT&CK targeted regions by threat groups. Moreover, Fig. 23 and Fig. 24 illustrate MITRE ATT&CK threat groups attributed to suspected regions. Here are the extracted insights:

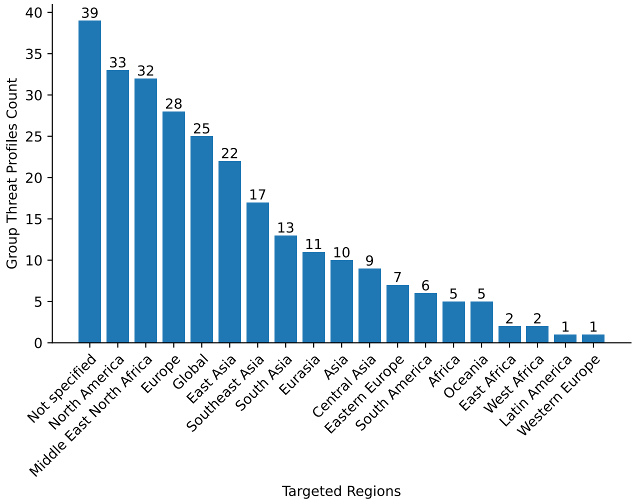


FIGURE 22. Targeted regions by threat groups.

- There are more than 70% of threat groups attributed to a specific region. Among all regions, North-America and Middle East-North Africa regions have been targeted by 25% and 24% of threat groups respectively. Overall, there is a linear distribution of the targeted regions by the number of groups.
- Figure 23 shows that more than 60% of threat groups have been attributed to a suspected region.
- Based on our analysis, we have observed biases while analyzing the suspected origin of threat groups: (i) As Fig. 24 shows, there are 7 suspected regions that are mostly located in the eastern side of the world compared to targeted regions that are distributed across the globe, (ii) National Cyber Power Index 2022 [41] have evaluated cyber offense capabilities by countries wherein 4 of the top 10 countries have not been attributed to any threat groups.
- It is worth mentioning that the corresponding threat groups of suspected regions have mainly targeted their own regions, indicating their geopolitical interests.

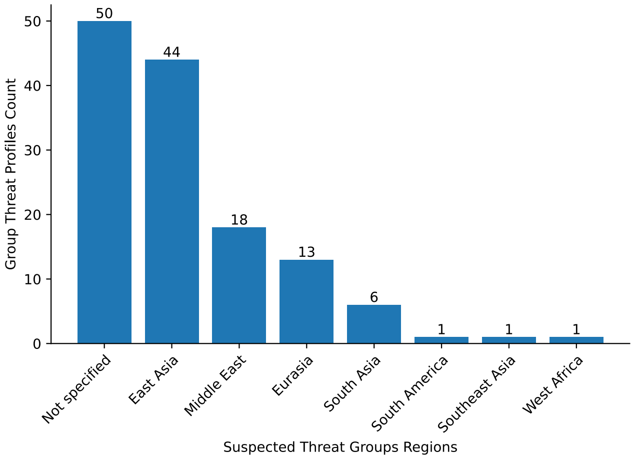


FIGURE 23. Suspected threat group regions.

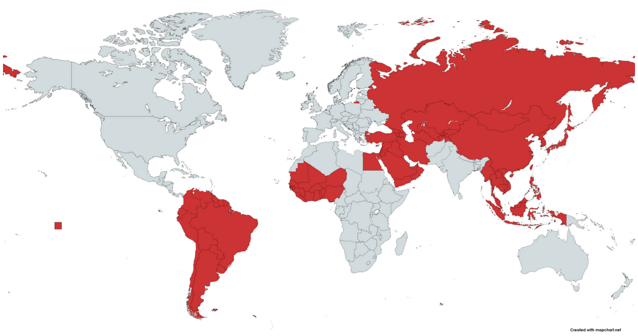


FIGURE 24. Suspected threat group regions over map.

VII. DISCUSSION

We conducted a hierarchical and analytical study of the MITRE ATT&CK dataset to extract its hidden insights. This section discusses the major findings of our analysis from different perspectives, such as the application scenario, targeted domains, targeted sectors, motivation behind threat groups, etc.

The insights in the previous section provide a deep understanding of MITRE ATT&CK framework and dataset including details, such as threat profiles, threat profile relationships, corresponding hierarchical structures, and mappings. We described the applications of these insights during this analytical study to be utilized within various contexts and environments by cyber security analysts, incident responders, vulnerability assessors, etc. In particular, our insights can be useful for highly important purposes, such as cyber attack mitigation, prediction, and prevention. In the following, we describe these purposes in detail.

Software profiles constitute a large number of threat profiles compared to Groups and Campaigns indicating that Groups and Campaigns have various choices of software tools to conduct their malicious activities. Taking into account the underlying domain, Groups and Campaigns can utilize appropriate software to perform their cyber activities. On the other side, organizations depending on their domains should identify problematic software lists having weaknesses/vulnerabilities that might be exploited by tools that threat groups/campaigns utilize, and define appropriate countermeasures.

Threat profiles' preferences in domains, platforms, tactics, and techniques provide valuable insights about strategies and road-maps attackers follow and invest in. Accordingly, appropriate mitigation decisions can be issued to intervene and break the cyber attack offense cycle. As an example, according to our analytical study, *Defense Evasion* is the most common tactic employed by threat profiles to carry out their cyber attacks. This insight helps organizations and cyber security experts develop detection strategies or implement countermeasures against these types of tactics and their techniques to mitigate the corresponding attacks.

The insights provided on the motivations of threat groups can help organizations and nations identify common global cyber attacks and the corresponding defense strategies. Similarly, extracting information from targeted sectors helps identify the most important motivations for each sector (e.g., *Espionage* is the most important motivation for attacks targeting the government sector).

From application scenarios perspectives, we can apply our analysis to various applications, such as (i) Operational Technology, (ii) Defense, (iii) Commercial, (iv) Telecom, (v) cyber security. Each one has its own sub-categories as listed in Table 4.

In order to complement the extracted insights, we can correlate them with other information sources, such as Common Vulnerabilities and Exposures (CVE),[§] National Institute of Standards and Technology (NIST),[§] Common Attack Pattern

[§] https://cve.mitre.org/

[§] https://www.nist.gov/

TABLE 4. Extracted insights application scenarios.

Application Scenarios	
Operational Technology	Power Grid
	Maritime
	Healthcare
	ICS
	IoT
	Water System
Commercial	Finance
	E-Commerce
Telecom	5G
	Cellphone
Cyber Security	Malware Detection

Enumeration and Classification (CAPEC),[§] cyber security reports, etc. The results provide complementary information about threat groups and the corresponding countermeasures that can help model various attacks and defenses.

The insights provided can be used for tactics and techniques associations. For this purpose, they can be structured as a Cyber Threat Intelligence (CTI) ontology and knowledge-base to be employed within several cyber security analysis systems within various contexts, such as analyzing security aspects of different industrial control systems, Internet of Things (IoT), self-driving vehicles, 5G Core Networks (5GCN), etc. Such an ontology or knowledge-base helps to find similar tactics and techniques exchangeable within cyber attack scenarios and extend attack detection scopes. Furthermore, the insights provided can be used to label threat groups and train machine learning and deep learning models for detection and prediction purposes. The results can be used for automated and real-time cyber attack detection and malware investigation purposes.

This study and its provided insights can be used for attacker characterization. Threat groups and campaigns information can be aggregated to define special characteristics (utilized tactics and techniques, platforms, motivations, targeted sectors, etc.) for each threat group and campaign so that unique profiles can be defined accordingly. Using these profiles, appropriate mitigation and prevention strategies can be issued by experts. Additionally, studying these profiles can help for risk analysis and assessment purposes.

VIII. CONCLUSION

This paper performed a statistical analysis of the MITRE ATT&CK dataset and presented insights to improve the security aspects of various digital infrastructures. For this purpose, we have proposed a hierarchical architecture representing the involved entities, such as threat actors, software, campaigns, domains, platforms, tactics, and techniques, in the MITRE ATT&CK framework and their relationships within the dataset. Finally, we have analyzed entities' relationships while providing high-level statistics to describe the current landscape of cyber security attacks.

[§] https://capec.mitre.org/

ACKNOWLEDGMENT

The contents herein are solely the responsibility of the author[s].

REFERENCES

[1] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, "MITRE ATT&CK: Design and philosophy," MITRE Corporation, McLean, VA, USA, Tech. Rep. MP180360R1, 2018.

[2] A. Kuppa, L. Aouad, and N.-A. Le-Khac, "Linking CVE's to MITRE ATT&CK techniques," in *Proc. 16th Int. Conf. Availability, Rel. Secur.*, 2021, pp. 1–12.

[3] B. Ampel, S. Samtani, S. Ullman, and H. Chen, "Linking common vulnerabilities and exposures to the MITRE ATT&CK framework: A self-distillation approach," 2021, *arXiv:2108.01696*.

[4] H. M. Farooq and N. M. Otaibi, "Optimal machine learning algorithms for cyber threat detection," in *Proc. UKSim-AMSS 20th Int. Conf. Comput. Modeling Simulation (UKSim)*, Mar. 2018, pp. 32–37.

[5] O. Mendsaikhan, H. Hasegawa, Y. Yamaguchi, and H. Shimada, "Automatic mapping of vulnerability information to adversary techniques," in *Proc. 14th Int. Conf. Emerg. Secur. Inf., Syst. Technol. (SECUREWARE)*, 2020, pp. 1–7.

[6] R. Kwon, T. Ashley, J. Castleberry, P. Mckenzie, and S. N. G. Gourisetti, "Cyber threat dictionary using MITRE ATT&CK matrix and NIST cybersecurity framework mapping," in *Proc. Resilience Week (RWS)*, Oct. 2020, pp. 106–112.

[7] F. S. Toker, K. O. Akpinar, and I. ÖZçelik, "MITRE ICS attack simulation and detection on EtherCAT based drinking water system," in *Proc. 9th Int. Symp. Digit. Forensics Secur. (ISDFS)*, Jun. 2021, pp. 1–6.

[8] Y.-T. Huang, C. Y. Lin, Y.-R. Guo, K.-C. Lo, Y. S. Sun, and M. C. Chen, "Open source intelligence for malicious behavior discovery and interpretation," *IEEE Trans. Depend. Sec. Comput.*, vol. 19, no. 2, pp. 776–789, Mar. 2022.

[9] Y. S. Takey, S. G. Tatikayala, S. S. Samavedam, P. R. L. Eswari, and M. U. Patil, "Real time early multi stage attack detection," in *Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2021, pp. 283–290.

[10] Z. Jadidi and Y. Lu, "A threat hunting framework for industrial control systems," *IEEE Access*, vol. 9, pp. 164118–164130, 2021.

[11] K. Kim, Y. Shin, J. Lee, and K. Lee, "Automatically attributing mobile threat actors by vectorized ATT&CK matrix and paired indicator," *Sensors*, vol. 21, no. 19, p. 6522, Sep. 2021.

[12] P. Karuna, E. Hemberg, U.-M. O'Reilly, and N. Rutar, "Automating cyber threat hunting using NLP, automated query generation, and genetic perturbation," 2021, *arXiv:2104.11576*.

[13] W. Wang, B. Tang, C. Zhu, B. Liu, A. Li, and Z. Ding, "Clustering using a similarity measure approach based on semantic analysis of adversary behaviors," in *Proc. IEEE 5th Int. Conf. Data Sci. Cyberspace (DSC)*, Jul. 2020, pp. 1–7.

[14] A. Nisioti, G. Loukas, S. Rass, and E. Panaousis, "Game-theoretic decision support for cyber forensic investigations," *Sensors*, vol. 21, no. 16, p. 5300, Aug. 2021.

[15] S. Barnum, "Standardizing cyber threat intelligence information with the structured threat information expression (STIX)," *Mitre Corp.*, vol. 11, pp. 1–22, Jan. 2012.

[16] A. Nisioti, G. Loukas, A. Laszka, and E. Panaousis, "Data-driven decision support for optimizing cyber forensic investigations," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2397–2412, 2021.

[17] M. Zych and V. Mavroeidis, "Enhancing the STIX representation of MITRE ATT&CK for group filtering and technique prioritization," in *Proc. ECCWS 21st Eur. Conf. Cyber Warfare Secur.*, 2022, pp. 1–7.

[18] Y. Wu, C. Huang, X. Zhang, and H. Zhou, "GroupTracer: Automatic attacker TTP profile extraction and group cluster in Internet of Things," *Secur. Commun. Netw.*, vol. 2020, pp. 1–14, Dec. 2020.

[19] J. Straub, "Modeling attack, defense and threat trees and the cyber kill chain, ATT&CK and STRIDE frameworks as blackboard architecture networks," in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*, Nov. 2020, pp. 148–153.

[20] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.

[21] A. A. Al-Kadhimi, M. M. Singh, and T. Jabar, "Fingerprint for mobile-sensor apt detection framework (FORMAP) based on tactics techniques and procedures (TTP) and MITRE," in *Proc. 8th Int. Conf. Comput. Sci. Technol. (ICCST)*, Aug. 2022, pp. 515–533.

[22] V. Chierzi and F. Merces, "Evolution of IoT Linux malware: A MITRE ATT&CK TTP based approach," in *Proc. APWG Symp. Electron. Crime Res. (eCrime)*, Dec. 2021, pp. 1–11.

[23] J. Fairbanks, A. Orbe, C. Patterson, J. Layne, E. Serra, and M. Scheepers, "Identifying ATT&CK tactics in Android malware control flow graph through graph representation learning and interpretability," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 5602–5608.

[24] A. Georgiadou, S. Mouzakitis, and D. Askounis, "Assessing MITRE ATT&CK risk using a cyber-security culture framework," *Sensors*, vol. 21, no. 9, p. 3267, May 2021.

[25] V. Legoy, M. Caselli, C. Seifert, and A. Peter, "Automated retrieval of ATT&CK tactics and techniques for cyber threat reports," 2020, *arXiv:2004.14322*.

[26] V. Mavroeidis and S. Bromander, "Cyber threat intelligence model: An evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence," in *Proc. Eur. Intell. Secur. Informat. Conf. (EISIC)*, Sep. 2017, pp. 91–98.

[27] M. Parmar and A. Domingo, "On the use of cyber threat intelligence (CTI) in support of developing the commander's understanding of the adversary," in *Proc. MILCOM IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2019, pp. 1–6.

[28] P. Zilberman, R. Puzis, S. Bruskin, S. Shwarz, and Y. Elovici, "SoK: A survey of open-source threat emulators," 2020, *arXiv:2003.01518*.

[29] S. Choi, J. Choi, J.-H. Yun, B.-G. Min, and H. Kim, "Expansion of $*ICS*$ testbed for security validation based on $*MITRE*$$*ATT&CK*$ techniques," in *Proc. 13th USENIX Workshop Cyber Secur. Experimentation Test (CSET)*, 2020, pp. 1–9.

[30] F. J. Stech, K. E. Heckman, and B. E. Strom, "Integrating cyber-D&D into adversary modeling for active cyber defense," *Cyber Deception: Building the Scientific Foundation*. Springer, 2016, pp. 1–22.

[31] F. Maymí, R. Bixler, R. Jones, and S. Lathrop, "Towards a definition of cyberspace tactics, techniques and procedures," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 4674–4679.

[32] H. Manocha, A. Srivastava, C. Verma, R. Gupta, and B. Bansal, "Security assessment rating framework for enterprises using MITRE ATT&CK matrix," 2021, *arXiv:2108.06559*.

[33] Y. Jo, O. Choi, J. You, Y. Cha, and D. H. Lee, "Cyberattack models for ship equipment based on the MITRE ATT&CK framework," *Sensors*, vol. 22, no. 5, p. 1860, Feb. 2022.

[34] M. Ahmed, S. Panda, C. Xenakis, and E. Panaousis, "MITRE ATT&CK-driven cyber risk assessment," in *Proc. 17th Int. Conf. Availability, Rel. Secur.*, Aug. 2022, pp. 1–10.

[35] W. Xiong, E. Legrand, O. Åberg, and R. Lagerström, "Cyber security threat modeling based on the MITRE enterprise ATT&CK matrix," *Softw. Syst. Model.*, vol. 21, no. 1, pp. 157–177, Feb. 2022.

[36] P. Johnson, R. Lagerström, and M. Ekstedt, "A meta language for threat modeling and attack simulations," in *Proc. 13th Int. Conf. Availability, Rel. Secur.*, Aug. 2018, pp. 1–8.

[37] S. Choi, J.-H. Yun, and B.-G. Min, "Probabilistic attack sequence generation and execution based on MITRE ATT&CK for ICS datasets," in *Proc. Cyber Secur. Experimentation Test Workshop*, Aug. 2021, pp. 41–48.

[38] S. Arshad, M. Alam, S. Al-Kuwari, and M. H. A. Khan, "Attack specification language: Domain specific language for dynamic training in cyber range," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2021, pp. 873–879.

[39] R. Pell, S. Moschoyiannis, E. Panaousis, and R. Heartfield, "Towards dynamic threat modelling in 5G core networks based on MITRE ATT&CK," 2021, *arXiv:2108.11206*.

[40] A. Berady, M. Jaume, V. V. T. Tong, and G. Guette, "From TTP to IoC: Advanced persistent graphs for threat hunting," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 2, pp. 1321–1333, Jun. 2021.

[41] D. C. J. Voo and I. Hemani, "National cyber power index 2022," Harvard Kennedy School, Belfer Center Sci. Int. Affairs, Tech. Rep., 2022. [Online]. Available: https://www.belfercenter.org/publication/national-cyber-power-index-2022

BADER AL-SADA received the B.Sc. degree (Hons.) in computer networks from the University of Portsmouth, U.K., and the M.Sc. degree in cyber security from Hamad Bin Khalifa University, Qatar, where he is currently pursuing the Ph.D. degree with the College of Science and Engineering.

ALIREZA SADIGHIAN received the Ph.D. degree in computer software engineering from the University of Montreal and École Polytechnique. He is currently a Postdoctoral Fellow with the College of Science and Engineering, Hamad Bin Khalifa University, Qatar. His research interest includes data-driven cyber security.

GABRIELE OLIGERI (Member, IEEE) received the Ph.D. degree in computer engineering from the University of Pisa. He is currently an Associate Professor with the College of Science and Engineering, Hamad Bin Khalifa University, Qatar. His research interest includes signals intelligence.

Open Access funding provided by 'Qatar National Library' within the CRUI-CARE Agreement