# A Decade-long Landscape of Advanced Persistent Threats: Longitudinal Analysis and Global Trends

Shakhzod Yuldoshkhujaev
Sungkyunkwan University
Suwon, Republic of Korea
shakzod02@g.skku.edu

Mijin Jeon
Sungkyunkwan University
Suwon, Republic of Korea
jinijxxn@g.skku.edu

Doowon Kim
University of Tennessee
Knoxville, TN, United States
doowon@utk.edu

Nick Nikiforakis
Stony Brook University
Stony Brook, NY, United States
nick@cs.stonybrook.edu

Hyungjoon Koo*
Sungkyunkwan University
Suwon, Republic of Korea
kevin.koo@skku.edu

## Abstract

An advanced persistent threat (APT) refers to a covert and long-term cyberattack, typically conducted by state-sponsored actors, targeting critical sectors and often remaining undetected for long periods. In response, collective intelligence from around the globe collaborates to identify and trace surreptitious activities, generating substantial documentation on APT campaigns publicly available on the web. While a multitude of prior works predominantly focus on specific aspects of APT cases, such as detection, evaluation, cyber threat intelligence, and dataset creation, limited attention has been devoted to revisiting and investigating these scattered dossiers in a longitudinal manner.

The objective of our study lies in filling the gap by offering a macro perspective, connecting key insights and global trends in the past APT attacks. We systematically analyze six reliable sources—three focused on technical reports and another three on threat actors— examining 1,509 APT dossiers (*i.e.,* totaling 24,215 pages) spanning from 2014 to 2023 (a decade), and identifying 603 unique APT groups in the world. To efficiently unearth relevant information, we employ a hybrid methodology that combines rule-based information retrieval with large-language-model-based search techniques. Our longitudinal analysis reveals shifts in threat actor activities, global attack vectors, changes in targeted sectors, and the relationships between cyberattacks and significant events, such as elections or wars, which provides insights into historical patterns in APT evolution. Over the past decade, 154 countries have been affected, primarily using malicious documents and spear phishing as the dominant initial infiltration vectors, and a noticeable decline in zero-day exploitation since 2016. Furthermore, we present our findings through interactive visualization tools, such as an APT map or a flow diagram, to facilitate intuitive understanding of the global patterns and trends in APT activities.

*Corresponding author.


*CCS '25, Taipei, Taiwan.*

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3765085

## CCS Concepts

• **General and reference** → **Measurement**; • **Security and privacy** → *Social aspects of security and privacy*.

## Keywords

Advanced Persistent Threats; Longitudinal Analysis; Global Trends

**ACM Reference Format:**
Shakhzod Yuldoshkhujaev, Mijin Jeon, Doowon Kim, Nick Nikiforakis, and Hyungjoon Koo. 2025. A Decade-long Landscape of Advanced Persistent Threats: Longitudinal Analysis and Global Trends. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25), October 13–17, 2025, Taipei, Taiwan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3719027.3765085

## 1 Introduction

Advanced persistent threats (APTs) are covert and sophisticated cyberattacks, typically orchestrated by state actors. APT campaigns attempt to gain unauthorized access to remote machines and stay undetected for extended periods, enabling targeted campaigns against governments and financial institutions [6]. Their primary objectives are to steal sensitive data, to disrupt critical operations, and to undermine national security or economic stability [2].

Given the severity of risks and threats posed by APTs, both industry and academia put in constant effort to monitor and understand APT-involving incidents. The security industry (*e.g.,* companies, experts, practitioners) investigates each APT incident, producing invaluable technical reports and articles. These *scattered* dossiers from varying sources provide detailed information on attack strategies, threat actors, and exploitation techniques on the web. Moreover, security practitioners track technical reports [9, 22, 30] and threat actors [20, 75, 101], organizing them in (public) repositories. The security industry also compiles Cyber Threat Intelligence (CTI) [120] databases to detect, analyze, and mitigate threats, including ① Indicators of Compromise (IoCs) [33] for collecting forensic evidence, ② Common Vulnerabilities and Exposures (CVEs) [18] for maintaining publicly disclosed security flaws, ③ MITRE ATT&CK Techniques [19] for identifying adversarial tactics, and ④ YARA rules [111] for detecting specific malware patterns. For the intuitive understanding of APTs, visual representations with a map [20, 51, 56] have been implemented on the web. Orthogonally to the aforementioned work by the cyber-security industry, academics also invest resources in better understanding

past APTs. The majority of recent academic studies have focused on studying and analyzing specific attributes related to APT incidents, such as detection and evaluation [1, 27, 48, 72, 74, 98, 108, 113–115], CTI [8, 53, 62, 67, 76, 99], and APT dataset [8, 13, 53, 62, 99].

In this paper, we observe that few studies have systematically investigated the landscape of APTs over an extended period of time. To help draw a complete picture of APT activity, we study the landscape of APT incidents over a period of 10 years. We uncover longitudinal changes through an in-depth analysis of fragmented, publicly available APT documentation. Our analysis can assist in identifying broader trends and patterns in surreptitious APT activities, offering valuable insights into the evolution of APT targets, malware samples, and sophisticated attack techniques.

More specifically, this study attempts to identify global trends from a macro viewpoint, including vulnerability exploitations, threat actor behaviors, and target changes through a large-scale investigation of 1,509 unique APT dossiers and 603 APT groups *over the past decade*. Due to the large volume of publicly available APT reports (*i.e.,* 24,215 pages of technical reports), we adopt a hybrid information-retrieval approach by leveraging the inference capabilities of Large Language Models (LLMs) combined with a rule-based extraction tool. To boost the LLM's accuracy, we carefully design the questions and prompts, evaluating multiple models to identify the most effective retrieval for our goals.

With a comprehensive collection of APT dossiers [9, 22, 30], our study aims to analyze ① the evolution of APTs over the past decade in terms of victim countries, threat actors, target sectors, initial attack vectors, and zero-day vulnerabilities; ② CTI records in APT cases (*e.g.,* CVEs, MITRE ATT&CK, YARA rules); ③ common traits of APT campaigns that demonstrate concealment (*e.g.,* attack duration) and aggressiveness; and ④ external factors affecting APT campaigns, such as political events, international conflicts, global pandemics, or economic instabilities.

Our analysis yielded the following findings: ① Over the past decade, APT campaigns have impacted 154 countries (80% of all nations), with the United States, India, and South Korea among the most frequently targeted. While 446 unique threat actors have been identified in our dataset, a small set of actors is responsible for a significant share of attacks, including Lazarus [123], APT28 [122], and APT29 [119]. In terms of targets, the government and corporate sector attract the majority of APT attention, with malicious documents and spear phishing serving as the dominant initial infiltration vectors. ② Vulnerability-wise, while the exploited CVEs are highly severe (average score of 8.5), our findings indicate that many of the attacks do not need to rely on zero-day vulnerabilities to be successful, which peaked between 2014 and 2016 but has declined thereafter. ③ In terms of the lifetime of the recorded incidents, APT campaign duration varies widely, from a single day to nearly five years (137 days on average). ④ Finally, an important finding is that APT activity frequently coincided with political events, international conflicts, global crises like COVID-19, indicating that attackers had already performed target reconnaissance and were waiting for an opportune time to act upon their findings.

To facilitate the exploration of APT campaign data by reviewers and eventually the general public, we designed an interactive map [1] that incorporates decade-long historical data, including threat actor(s), CVEs, attack vector(s), malware, target sector(s), and estimated duration, with support for *dynamic updates* using LLMs to retrieve content from technical reports. Additionally, we provide a flow diagram [2] illustrating the relationships between selected threat actors and target countries.

**Table 1: Summary of prior work on APTs. We classify them into five categories where the majority of those studies focus on specific domains, covering limited periods. Our work aims to offer insights into the evolution of APTs over the past decade from macro perspective (Section 2).**

| Category | Topic Focus |
|---|---|
| **Survey** | APT Survey [2, 6] |
| **Detection and evaluation** | Detection techniques [1, 27, 48, 72, 74, 108, 114, 115]<br>Evaluating APT detection systems [108]<br>APT reconstruction [98, 113] |
| **Cyber Threat Intelligence** | Information retrieval [8, 67, 76, 99]<br>Information recognition [53, 62] |
| **APT Dataset** | Dataset creation [8, 53, 62, 99]<br>Dataset evaluation [13] |
| **Technical articles** | APT trends [38, 106]<br>Special reports [16, 47] |

In summary, this paper makes the following original contributions:

- We conduct a longitudinal measurement study of APT campaigns over the last decade (2014 – 2023), organizing 1,509 unique technical reports, 603 threat actors, and 177 news articles.
- To unveil longitudinal APT campaigns, we retrieve and refine responses to 10 identified questions using (context-aware) LLMs.
- We carefully define four research questions centered on the evolution, CTI records, common traits, and external dynamics of APT campaigns over a decade. Our findings reveal global trends and key insights, including that the campaigns have affected 80% of countries worldwide; a small number of actors are responsible for a disproportionate share of attacks; and the exploitation of both zero-day and one-day vulnerabilities is prevalent.
- We publicly release our curated dataset [3] and an interactive APT-campaign map to foster future research in the field of APT studies.

Due to space constraints, supplementary materials, such as the LLM prompts, questions, and the snapshots from visualization tools, are included in the extended version of this paper [4].

## 2 Background

This section provides some background on advanced persistent threats, CTI, and the focus of prior work on APT campaigns.

**Advanced Persistent Threats.** APTs refer to a critical and insidious category of cyberattacks characterized by their sustained, targeted, and highly sophisticated nature. Typical attacks include individual actors attacking any and all systems for financial gain, activism, or to demonstrate technological proficiency. However, APTs are often executed by well-organized, resource-rich groups

[1] https://lngt-apt-study-map.vercel.app/
[2] https://public.tableau.com/views/TopMentionedCountries/Top30Countries
[3] https://zenodo.org/records/16869733
[4] https://arxiv.org/abs/2509.07457

with strategic objectives [2]. These entities persist in their efforts to infiltrate and maintain a presence within a network over extended periods, often evading detection. Evidenced by exploiting zero-day vulnerabilities, deploying custom-developed malware, and adopting advanced social-engineering techniques, APTs' resource-intensive operations indicate clear underlying motives - political [25, 47, 127], economic [60, 86, 103], and military [80, 90, 128]. APT campaigns predominately aim at espionage, disruption, or sabotage, posing serious threats to national security and critical infrastructure.

**Cyber Threat Intelligence.** Cyber Threat Intelligence [120] is an essential pillar of modern cybersecurity for threat detection, prevention, and response. CTI enables organizations to collect and analyze information on cyber threats, anticipate emerging threats, and mitigate potential risks beforehand, empowering them to take proactive measures against cyberattacks. Indicators of Compromises [33] are critical elements of CTI, representing specific pieces of evidence that signal malicious activities within a system or network. Integrating such artifacts with CTI serves as actionable data for detecting and preventing future threats. IoCs are often aligned with standardized CTI frameworks to maximize their effectiveness. For instance, the MITRE ATT&CK [19] framework categorizes adversarial behaviors into distinct techniques, offering a structured approach to understanding how IoCs align within the attack lifecycle. In a similar vein, CVEs [18] maintain up-to-date (publicly known) vulnerabilities exploited by attackers, while YARA rules [111] facilitate the detection of malicious artifacts through signature-based matching. As such, various tools have been developed to automate the extraction of IoCs from technical reports. One notable example is IoCParser [105], which specializes in retrieving diverse IoCs from URLs and texts. We leverage IoCParser to extract three primary IoCs – CVEs, MITRE ATT&CK technique IDs, and YARA rules.

**Previous APT Studies.** Table 1 summarizes prior research on APTs. While extensive surveys on APTs have been conducted [2, 6], the majority of these studies are dedicated to specific domains or cover limited periods. Beyond those surveys, we classify existing work into four distinct areas. First, a significant portion of the literature concentrates on detection and evaluation by exploring detection techniques [1, 27, 48, 72, 74, 108, 114, 115], evaluating existing detection systems [108], or reconstructing APT scenarios [98, 113]. Second, several studies focus on information retrieval [8, 67, 76, 99] and information recognition [53, 62] contributing to a deeper understanding of Cyber Threat Intelligence. Third, another direction explores APT datasets in terms of their creation [8, 53, 62] and evaluation [13]. Lastly, the security industry has analyzed APT trends [38, 106] and issued special reports on particular incidents [16, 47]. However, our work differs from prior work by offering a longitudinal analysis based on a multitude of scattered APT dossiers, providing valuable insights into the evolution of APTs over time.

**Threat Maps.** Geo-location maps are often used to visualize APT incidents. Kaspersky [56], for example, offers a cyber attack map that depicts real-time worldwide cyber assaults. However, this tool needs to include historical data for APT activities. Both APT threat actor map [51] and APTMAP [20] aim to map APT-related data. However, both focus on APT groups' physical locations, which are missing a longitudinal study. Although these works introduce an APT map, they focus on the distribution of victim nations by year alone. In this work, we devise a new APT map that represents combined information associated with a certain threat actor or victim country.

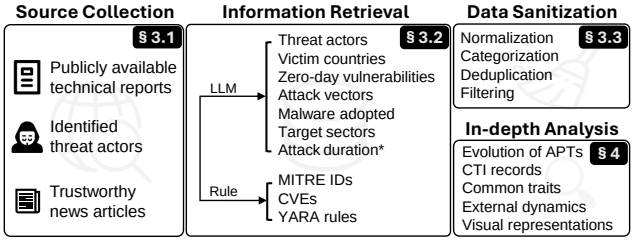



**Figure 1: Overview of our methodology for longitudinal APT analysis. We collect technical reports, threat actors, and news articles across the web (Section 3.1). Then, we probe valuable information from technical reports based on rules and LLMs (Section 3.2). Note that we manually inspect attack duration (*) for precise analysis. Next, we refine raw information via normalization, categorization, de-duplication, and filtering (Section 3.3). Lastly, we conduct in-depth analyses to answer our research questions (Section 4).**

## 3 Methodology

This section sketches our methodology to reveal the landscape of APT campaigns over the past 10 years, including details on source collection, information retrieval, and raw data refinement.

**Overview.** Figure 1 illustrates the overview of our methodology to analyze decade-long APT cases spanning from 2014 to 2023. We explore open APT technical reports that provide details of an individual attack, threat actors, and varying news articles related to APT attacks (Section 3.1). Among those sources, we choose the dossiers that are ① publicly accessible, ② offering up-to-date information, and ③ written by a trustworthy entity. Additionally, we gather past articles from an authoritative outlet specializing in security-focused news. Considering the volume of technical reports (*i.e.,* 24,215 pages), we retrieve useful information (*e.g.,* victim countries, target sectors, attack vectors) by leveraging an LLM that helps contextual inference. Note that we also utilize a rule-based (*i.e.,* regular expressions) tool that performs better than the LLM probe when looking for specific information, such as, MITRE IDs, CVEs, and YARA rules. (Section 3.2). Subsequently, we refine the raw data using normalization, categorization, deduplication, and filtering (Section 3.3).

### 3.1 APT Source Collection

As a primary source, we use a series of open datasets that collect technical reports and threat actors. We acknowledge the collective intelligence of security experts and practitioners who collaborate to uncover covert threat activities around the globe. Of all, we carefully select reliable sources, which focus on three collections of technical reports [9, 22, 30] (Table 2) and another three collections of threat actors [20, 75, 101] (Table 3). While the former provide in-depth analyses of each APT incident, the latter offer individual

**Table 2: Statistics on our collection of technical reports (TRs) and news articles on APTs. Out of 2,563 TRs, we analyze 1,509 unique TRs (after removing 1,003 duplicates and 51 APT trend dossiers), along with 177 news articles. The numbers in parentheses indicate the sum of all TRs before refinement. We discuss the credibility of TR sources in Section 3.1.**

| Year | TR#1 [22] | TR#2 [9] | TR#3 [30] | All TRs | News Articles |
|---|---|---|---|---|---|
| **2014** | 128 | 104 | 16 | 124 (248) | 2 |
| **2015** | 150 | 87 | 28 | 135 (265) | 8 |
| **2016** | 171 | 104 | 13 | 168 (288) | 14 |
| **2017** | 124 | 87 | 26 | 142 (237) | 15 |
| **2018** | 169 | 24 | 46 | 160 (239) | 18 |
| **2019** | 222 | 27 | 57 | 201 (306) | 15 |
| **2020** | 226 | 10 | 105 | 207 (341) | 21 |
| **2021** | 156 | 13 | 94 | 162 (263) | 19 |
| **2022** | 51 | 77 | 117 | 136 (245) | 30 |
| **2023** | 21 | 42 | 68 | 74 (131) | 35 |
| **Total** | 1,418 | 575 | 570 | **1,509 (2,563)** | **177** |

**Table 3: Statistics on our collection of threat actors (TAs). Out of 1,684 TAs, we organize 603 unique TAs after removing 800 duplicates and 281 entries containing no information beyond their names. Note that the information of APT groups has been maintained by reliable sources (Section 3.1).**

| Collection | Source of Threat Actors' Information | # of TAs |
|---|---|---|
| **TA#1** [75] | MISP Project [23] | 562 |
| **TA#2** [101] | Palo Alto [77], IBM X-Force[95, 96], Malpedia [24], Kaspersky [57], Crowdstrike [21], Mandiant [65], Secureworks [94], Dragos [26], Venafi [109], CERT-UA [15], Microsoft [70] | 692 |
| **TA#3** [20] | MITRE ATT&CK [19], ETDA [28], VX-underground [112] | 430 |
| **Total** | – | **603 (1,684)** |

APT group information. We include both sources because technical reports often lack detailed information about APT groups.

**Collection of Technical Reports.** We thoroughly examined the available web sources to obtain the most reputable technical reports (TRs), which we refer to as dossiers. We adopt three TRs as reliable sources in this work. First, TR#1 [22] is a GitHub repository containing an extensive collection of 1,418 TRs. This repository is continually updated as new dossiers on APT cases become available. Similarly, TR#2 [9] is another GitHub repository that hosts 575 TRs documenting various APT campaigns. Lastly, TR#3 [30] is a curated list of 570 TRs included as part of the comprehensive dataset maintained by Malpedia [24]. As shown in Table 2, we aggregate these three sources as *the collection of* 1,509 *technical reports* between 2014 and 2023 for our longitudinal analysis of APT cases. Although a small number of old TRs are available, we define our collection from the year of 2014 that shows a significant rise in the volume of TRs. Similarly, we exclude 2024 due to the small number of available TRs. Finally, we convert all web page reports into the PDF format for consistency.

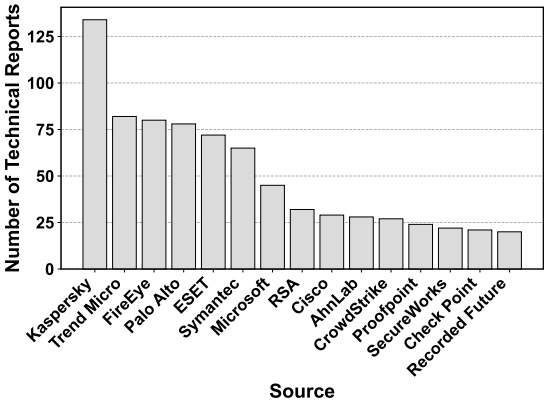



**Figure 2: Top 15 sources from the collection of technical reports. Most reports come from reputable sources such as Kaspersky [55] and Trend Micro [69]. We confirmed that 1,412 (93.6%) TRs are highly credible (Section 3.1).**

**Collection of Threat Actors (APT Groups).** We separately collect threat actors (TAs) since the collection of TRs does not contain that information. First, TA#1 [75] includes information about the threat actors obtained from the MISP project [23], one of the well-known Open Source Threat Intelligence Platforms. Second, TA#2 [101] is maintained by diverse security companies and non-profit organizations which includes a wide range of APT groups. Third, TA#3 [20] updates APT group information independently from the above two sources. Similar to the collection of TRs, we combine all the above three sources, obtaining *the collection of* 603 *unique threat actors*. Table 3 provides the statistics of our collection, which encompasses the information on TAs' official names and their alias(es), country of origin, motives, first-seen year, activity patterns, and sponsors.

**Collection of News Articles.** Oftentimes, APTs occur in the context of cyber warfare. To identify plausible connections between APTs and external factors (including national conflicts, geopolitical events, and global crises) we gathered security news articles and media reports about APT incidents for the last decade. Note that our collection includes 177 news articles (Table 2) through the Google News search engine [78] using the `APT` and `attack` keywords.

**Source Credibility.** Characterizing the sources is crucial for further analysis since unreliable source(s) could severely distort our collected data, compromising the validity of the findings. First, we explicitly extract the origin of a report (*e.g.,* organization name) from the collection of TRs. We confirmed that 1,412 (93.6%) of our TR collection come from highly reliable sources, including reputable security companies like Kaspersky [55], RSA [97], FireEye [61], and Microsoft [71], as well as governmental agencies like NATO [84], US-CERT [17], FBI [79], or trustworthy news sources. Figure 2 highlights the top 15 sources, collectively representing approximately half of TRs. The remaining 97 TRs (6.4%) come from individual web blogs and security experts, which may be relatively less reputable or credible. Second, for the collection of TAs, Table 3 shows that the original data comes from highly trusted companies and projects that deal with CTI, such as MITRE ATT&CK [19], MISP [23], ETDA [28], and CrowdStrike [21].

**Table 4: Performance comparison of precision (P), recall (R), and F1 scores across three different LLMs. To evaluate LLM models, we obtain the ground truth by manually inspecting around 120 technical reports. Due to its superior performance, we choose the GPT-4-Turbo model (Section 3.2.1).**

| Language Model | Gemini Flash [7] | | | GPT-4-Turbo [81] | | | GPT-4o [44] | | |
|---|---|---|---|---|---|---|---|---|---|
| Search Item | P | R | F1 | P | R | F1 | P | R | F1 |
| **Threat Actor** | **0.98** | 0.78 | 0.87 | **0.98** | **0.80** | **0.88** | **0.98** | 0.77 | 0.86 |
| **Victim Country** | 0.84 | 0.77 | 0.80 | **0.88** | **0.86** | **0.86** | 0.82 | 0.77 | 0.79 |
| **Zero-day** | **0.96** | 0.65 | 0.77 | 0.95 | **0.95** | **0.95** | 0.89 | 0.74 | 0.81 |
| **Average** | 0.93 | 0.73 | 0.81 | **0.94** | **0.87** | **0.90** | 0.90 | 0.76 | 0.82 |

**Table 5: Comparison of precision (P), recall (R), and F1 scores between a signature-based (*e.g.,* IoCParser) and an LLM-based approach (*e.g.,* GPT-4-Turbo). Note that IoCParser is capable of seeking CVE, MITRE ID, and YARA rules alone. (*) represents the items that we adopt GPT Turbo's results that demonstrate the best LLM performance (Section 3.2.2).**

| Tool | IoCParser [105] | | | GPT-4-Turbo [81] | | |
|---|---|---|---|---|---|---|
| Search Item | P | R | F1 | P | R | F1 |
| **CVE** | **0.98** | **0.95** | **0.97** | 0.97 | 0.84 | 0.90 |
| **MITRE ID** | 0.97 | **0.96** | **0.97** | **0.99** | 0.93 | 0.96 |
| **YARA** | **1.00** | **0.96** | **0.98** | 0.94 | 0.86 | 0.90 |
| **Attack vector*** | – | – | – | 0.89 | 0.77 | 0.83 |
| **Malware*** | – | – | – | 0.74 | 0.70 | 0.72 |
| **Target sector*** | – | – | – | 0.82 | 0.89 | 0.85 |
| **Average** | **0.98** | **0.96** | **0.97** | **0.89** | **0.83** | **0.86** |

## 3.2 APT Information Retrieval

We carefully define ten items that have the potential to reveal longitudinal APT changes and global trends, including MITRE IDs [19], CVEs [18], YARA rules [111], threat actors (APT groups), victim country, zero-day vulnerabilities, initial attack vectors, associated malware, target sectors, and attack durations.

*3.2.1 LLM-based Retrieval.* In this work, we leverage the language model's contextual inference capabilities to accurately extract relevant information from our collected technical reports.

**LLM Prompt and Questions Design.** We design our prompt and questions to enhance the accuracy of responses during the retrieval process of an LLM. We systematically evaluate varying question formulations and methodological techniques to identify the most effective strategies. Following the recommendations of Kumarasinghe *et al.* [53], we incorporate both the "Role Play" and "Specificity and Precision" means for prompt generation. The former approach enables the LLM to adopt a defined perspective, thereby generating more contextually appropriate and relevant responses, while the latter approach reduces the likelihood of irrelevant or inaccurate outputs. Likewise, for query construction, we employ "Specificity and Precision" to ensure simplicity and straightforwardness, which minimizes ambiguity and improves overall quality of the retrieved responses.

**LLM Model Choice and Evaluation.** To obtain the desirable answers as reliably as possible, we sample three items for retrieval (*e.g.,* threat actors, victim countries, zero-day vulnerabilities if any). To this end, we evaluate three popular LLMs: Gemini-1.5-Flash [7], GPT-4o [44], and GPT-4-Turbo [81] with the same prompt and questions. However, it is well-known that LLM models may generate inaccurate responses (*i.e.,* hallucinations). This means we cannot completely trust the responses from an LLM. In response, we randomly picked around 120 articles for manual inspection: *i.e.,* a human compares the LLM responses with the ground truth in a technical report. Note that we use precision ($P = \frac{TP}{TP+FP}$), recall ($R = \frac{TP}{TP+FN}$), and F1 ($F1 = \frac{2PR}{P+R}$) for evaluation metrics where TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively. Notably, FPs refer to cases where the LLM incorrectly identifies attributes that are not present in the report, while FNs arise when the model fails to detect attributes that are present. Table 4 presents a performance comparison of three LLM models. GPT-4-Turbo achieves the highest scores in all Precision (0.94), Recall (0.87), and F1 score (0.90) over other LLM models.

**Table 6: We investigate 10 items from each technical report on a specific APT. We adopt rule-based (*e.g.,* IoCParser [105]) and LLM-based approaches. The number (ratio) of TRs denotes retrieved items out of 1,509 TRs because not every piece of information was available (Section 3.2.4).**

| Search Item | Retrieval Approach | # of TRs | Ratio |
|---|---|---|---|
| **CVE** | Rule | 416 | 27.6% |
| **MITRE ID** | Rule | 175 | 11.6% |
| **YARA** | Rule | 131 | 8.7% |
| **Threat actor** | LLM | 1,089 | 72.2% |
| **Victim country** | LLM | 886 | 58.7% |
| **Zero-day** | LLM | 839 | 55.6% |
| **Attack vector** | LLM | 1,186 | 78.6% |
| **Malware** | LLM | 1,287 | 85.3% |
| **Target sector** | LLM | 1,228 | 81.4% |
| **Attack duration** | LLM, Manual inspection | 235 | 15.6% |

*3.2.2 Rule-based Retrieval.* The IoCParser [105] tool is designed for processing IoCs from various data sources (*e.g.,* CTI reports, security logs, other security-related texts). As the parser extracts three specific types of information, including CVEs, MITRE ATT&CK Technique IDs, and YARA rules, we compare it with GPT-4-Turbo. Table 5 demonstrates that IoCParser surpasses GPT-4-Turbo in extracting CVEs, MITRE IDs, and YARA rules, achieving F1 scores of 0.97, 0.97, and 0.98, respectively. Hence, we decided to include IoCParser's results as a complementary tool. Our further examination of IoCParser's retrieval failure reveals that ① some TRs do not follow the standardized CVE format (*e.g.,* `year-vul_id` instead of `CVE-year-vul_id`), which reduces the recall; ② some TRs extracted from web pages have the parser retrieve irrelevant IoCs due to a noise in the extracted content, decreasing its precision. As a final note, IoCParser does not have features to retrieve other items.

*3.2.3 Manual Retrieval.* We manually verify the accuracy of attack durations retrieved by the LLM to estimate both the lifecycle of individual APT campaigns and the time required to patch associated vulnerabilities. However, determining the precise lifecycle of an APT incident is inherently challenging due to its persistence and

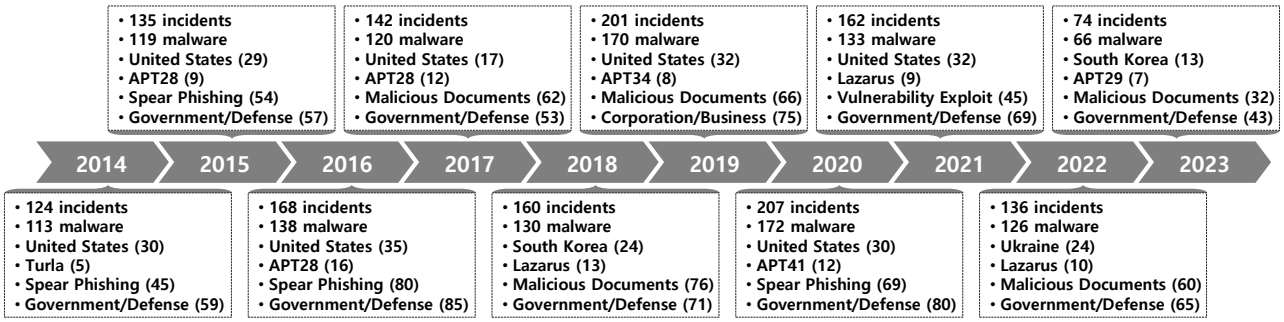



**Figure 3: Summary of global APT trends over the past decade. Each box represents six key points for a given year: the number of APT campaigns, the number of associated malware samples, the most frequently attacked country, the primary threat actor, the most commonly used initial attack vector, and the most targeted sector. For the last 10 years, the most frequently targeted country, the most active APT group, the most predominant attack method, and most commonly targeted sector have been the United States, Lazarus group, malicious documents (and spear phishing almost equally contributed), and the government/defense sector, respectively (Section 4.1). Note that the numbers in parentheses represent the occurrences per year.**

stealthy nature. Such campaigns may persist for weeks, months, or even years, depending on the adversary's ultimate objective (*e.g.,* disrupting target operations vs. stealing sensitive data). Hence, we define the start and end of an APT case based on the earliest and latest known discoveries of related activity, using day-level granularity. When exact dates are unavailable, we approximate using the midpoint of the month (*i.e.,* the 15th day). Note that we exclude cases where the year-based (rough) information is solely available because such imprecision could lead to inaccurate estimations.

*3.2.4 Availability of Retrieved Information.* As one might imagine, not every piece of information was available in our collection of TRs. Indeed, we discover that only one TR contained all 10 items of interest. Table 6 presents the number of TRs (out of 1,509) that include each retrievable item. YARA rules and MITRE ID are sparsely available, appearing solely 8.7% and 11.6% of the TRs, respectively. In contrast, malware information and target sectors are prevalent, in approximately 8 out of 10 TRs (85.3% and 81.4%, respectively). Given that our analysis relies on the records in the collected TRs as a best-effort approach, we find it encouraging that a non-negligible number of items was successfully retrieved. We hypothesize that the covert nature of APT campaigns contributes to the limited availability of malware samples and lack of precise information regarding the duration of attacks.

*3.2.5 Information Retrieval from Technical Reports.* We convert web-based technical reports into PDF format using `pdfkit` [34] in conjunction with `wkhtmltopdf` [107]. Then, we utilize LangChain [58] that offers a structured framework for data processing and retrieval. To enable context-aware question answering, we employ a Retrieval-Augmented Generation [89] (RAG) pipeline by vectorizing the technical reports into a vector database. For each query, relevant passages are retrieved and incorporated into the prompt to provide contextually rich information to the LLM. In summary, our approach consists of the following phases: ① extracting texts from each PDF document using PyPDFLoader [59], ② vectorizing those texts with the OpenAI's [83] embedding model, ③ storing vectorized embeddings in the FAISS library [50], and ④ generating responses by constructing prompt templates and chaining them with user queries to facilitate effective question answering. For performance comparisons (Table 4), we access LLMs via LangChain's API integrations, including GPT-4-Turbo [81], GPT-4o [44], and gemini-1.5-flash-latest [7]. It is worth noting that achieving full accuracy and reliability remains challenging due to complications from PDF processing (Section 5).

## 3.3 Refining LLM-generated Responses

This section describes the refinement process of LLM-generated responses using our prompt and questions.

**Normalization and Categorization.** For our analysis, it is essential to interpret and normalize the responses generated by the LLM, particularly for the victim countries, attack vectors, and target sectors. To ensure consistency, victim countries were standardized by converting them into their corresponding two-letter country codes [32]. For the categorization of attack vectors, we referred to the work of Sharma *et al.* [6], which identifies the most common attack vectors. This process results in classifying the attack vectors into the following 12 distinct categories: Spear Phishing, Phishing, Watering Hole, Credential Reuse, Social Engineering, Vulnerability Exploitation, Malicious Documents, Covert Channels, Drive-by Download, Removable Media, Website Equipping and Meta Data Monitoring. Meanwhile, for the categorization of target sectors, we adopt the taxonomy proposed in [93], which includes nine categories: Government and Defense Agencies, Corporations and Businesses, Education and Research Institutions, Critical Infrastructure (*e.g.,* transportation, water supply), Financial Institutions, Individuals, Media and Entertainment Companies, Non-Governmental Organizations (NGOs) and Nonprofits, and Manufacturing. In line with [35], we treat Energy and Utilities and Healthcare as distinct categories from critical infrastructure. Additionally, we include Cloud/IoT Services as a category to capture its relevance in recent attacks, such as those involving supply chains. This process results in classifying the target sectors into 12 distinct classes.

**De-duplication and Filtering.** After consolidating all sources from the collection of TRs and TAs, we identified a substantial number of duplicate entries. To address this, we extracted the full text from each PDF and computed cosine similarity using OpenAI's embedding [82]. We apply an empirically-chosen similarity threshold of 0.85. This reduces the number of TRs to 1,560 unique dossiers after removing 1,003 duplicates. Furthermore, we found that 51 TRs focus on APT trends involving multiple campaigns rather than analyzing individual APT instances. We filter out these reports, resulting in a final dataset of 1,509 TRs. Similarly, the number of TAs was reduced to 884 unique APT groups after trimming 800 duplicates. Next, we further filtered out the threat actors ($n$ = 281; 31.8%) that hold little information beyond their names, such as aliases and country of origin. Finally, we use 603 identifiable APT groups for further analysis.

**APT Name Aliases.** The same APT group may operate under multiple aliases. For instance, the notorious APT group known as APT28 [122] is frequently referred to as FANCY BEAR, Pawn Storm, or Sofacy depending on the security vendor. To minimize complications, we de-duplicate APT groups with the following two-phase process. First, we use the identifiers as primary names from each threat actor source, along with their known aliases. Then, we merge alias entries across different sources based on these primary names. It is possible that despite these steps, due to the inconsistent naming conventions of APT groups across vendors and the absence of ground truth, some inaccuracies may remain in our final set.

## 3.4 Visual Representations

We develop the APT map and its corresponding timeline chart using the React framework [68] and the amCharts library [5] to enable interactive user experiences. The front-end has been integrated with a Flask-based backend [87], deployed on the Heroku platform [91]. Additionally, we use Tableau [92], a visual analytics platform, to construct a flow diagram illustrating the relationships between years, threat actors, and victim countries by incorporating the `Sankey Viz` extension [102].

# 4 Decadal Landscape of APT Campaigns

To longitudinally understand the landscape of APTs, we define the following research questions (RQs) under four themes.

- **RQ1: Evolution of APTs over a decade.** How have APT campaigns evolved over the past 10 years regarding victim countries, threat actors, target sectors, initial attack vectors, and zero-day vulnerabilities?
- **RQ2: Cyber Threat Intelligence records for APTs.** How is APT-related information captured across common threat intelligence sources, including vulnerability databases, attack frameworks, and indicators of compromise?
- **RQ3: Common traits of APTs.** To what extent do APT campaigns exhibit concealment and aggressiveness?
- **RQ4: External dynamics affecting APTs.** How do external factors, such as political events, international conflicts, global pandemics, or economic instabilities, affect APT activity?

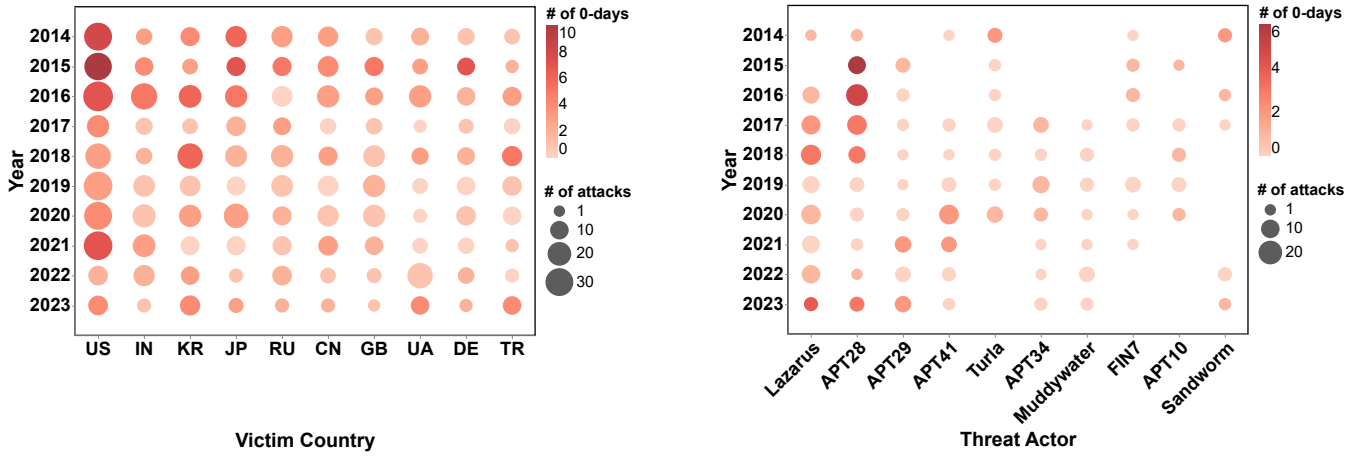
## 4.1 Evolution of APT Campaigns Over a Decade

This section explores the temporal changes in APT campaigns. We use the end year of the attack identified by the LLM, since the start dates of APT operations are often unknown or partially known (around 27%) due to the covert nature of APT operations. Otherwise, we base our statistics on the year in which the corresponding technical reports were published. Accordingly, each APT is counted only once, even if it spans multiple years. However, we note that a single APT case may involve multiple target sectors or attack vectors, each of which is counted separately, as illustrated in Figure 5.

**Comprehensive Overview.** Figure 3 presents a comprehensive overview of APT trends over the past decade, highlighting key attributes for each year, including the most frequently targeted countries and sectors, the most active threat actors, and the most widely employed initial attack vectors. The United States (US) consistently appears as the primary target, while the Lazarus group [123] emerges as the most active threat actor during this period. Malicious documents constitute the most popular initial attack vector. Furthermore, the frequency of APT attacks involving malware closely aligns with the overall trend in APT activity, indicating that malware remains a core component of most campaigns.

**Victim Countries.** A total number of 154 countries were identified as victims across 1,509 APT campaigns, representing *around 80% of all nations worldwide*. We analyze the 10 most victimized countries that account for 43.1% (650) of all incidents. Figure 4a presents the trends in APT attacks against these countries from 2014 to 2023. US remains the primary target throughout the decade with the exceptions of South Korea (in 2018, 2023) and Ukraine (in 2022). Despite minor fluctuations, our findings reveal that the following countries have also been heavily targeted: India (IN), South Korea (KR), Japan (JP), Russia (RU), China (CN), Great Britain (GB), Ukraine (UA), Germany (DE), and Türkiye (TR). A notable spike in attacks across the 10 most countries occurred in 2016 with 174 recorded cases. There has been a steady decline in APT activity since 2021, with cases dropping from 157 in 2020 to 121 in 2021, and to 67 in 2023.

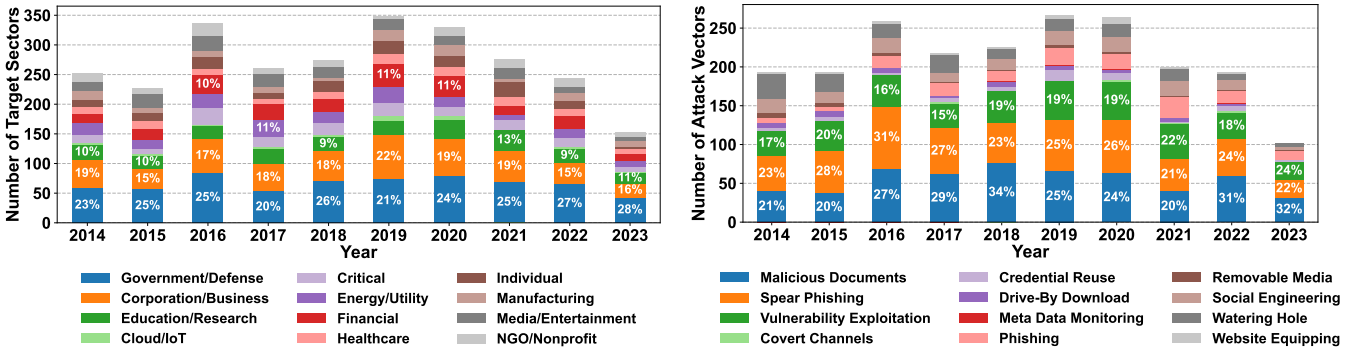
**Threat Actors.** A total of 446 unique threat actors have been identified in our APT dataset, 263 of which are known to be state-sponsored (while others unknown). The 10 most active APT groups are responsible for 326 attack incidents, representing 21.6% of all campaigns. Figure 4b shows the number of attacks and the frequency of associated zero-day vulnerabilities by TAs. Among them, Lazarus [123], APT28 [122], and APT29 [119] stand out as the most prolific threat actors over the past decade, with 76, 64, and 35 campaigns, respectively. APT28, in particular, exhibits a strong inclination toward exploiting zero-day vulnerabilities, with 22 instances accounting for 34.4% of its operations. Interestingly, the activity levels of these threat actors are not uniformly distributed over time; rather, their campaigns tend to cluster within specific periods, likely reflecting shifting strategic objectives. For instance, the Sandworm [125] group conducted the majority of its operations within a concentrated five-year span. The peak year for activity among these top actors was 2017, during which 48 attacks were recorded. Since 2021, however, a gradual decline in activity has been observed, with the number of attacks falling from 47 in 2020 to 27 in 2021, and further to 26 in 2023.




**(a) Number of APT attacks against the 10 most victimized countries over the past decade. The United States has experienced the highest number of attacks, followed by India, South Korea, Japan, and Russia.**

**(b) Number of APT campaigns conducted by the 10 most active threat actors over the past decade. The Lazarus group has been identified as the most active APT group, while APT28 is notable for its frequent exploitation of zero-day vulnerabilities.**

**Figure 4: Decadal trends in APT activity by victim countries (left) and threat actors (right). A circle size reflects the frequency of APT incidents, while color gradation represents the number of zero-day vulnerabilities associated with each entity as a concrete value (*i.e.,* lighter red indicates fewer occurrences) (Section 4.1).**




**(a) APT campaigns consistently target the government and business sectors, as well as increasing attacks on the education and financial sectors in recent years.**

**(b) Malicious documents and spear phishing remain the most common initial attack vectors consistently harnessed by APTs, followed by vulnerability exploitation.**

**Figure 5: Decadal APT trends in 12 target sectors (left) and 12 initial attack vectors (right). We follow the categories of sectors from the guides [35, 93] and the attack vectors that Sharma *et al.* [6] proposed. The figures illustrate the distributions of each sector/vector over the last 10 years. The percentages within a stacked bar chart indicate the three most common target sectors and attack vectors for each year, along with their respective proportions. Note that a single APT case may entail multiple target sectors or attack vectors, which we count individually (Section 4.1).**

**Zero-Day Vulnerabilities.** One of the common beliefs is that APT campaigns frequently exploit zero-day vulnerabilities due to their capability to bypass defenses and maintain stealth. We identify 204 APT incidents (13.5%) that incorporate zero-day vulnerabilities into their campaigns. Figure 4a and Figure 4b depict the annual use of zero-day vulnerabilities over the past decade from the perspectives of the top 10 victim countries and threat actors. Our analysis reveals notable peaks in zero-day exploitation between 2014 and 2016, with an average of 39.7 cases affecting victim countries and 7.7 cases linked to threat actors. The highest recorded number of zero-day incidents targeting countries occurred in 2015, reaching 50 cases. In contrast, 2019 saw a sharp decline, with only nine cases affecting victim countries and just one attributed to threat actors. Since 2018, a general downward trend in zero-day vulnerabilities has been observed, possibly due to the increasing prevalence of 1-day vulnerabilities (*i.e.,* non-patched systems) or rising costs and the complexity of developing zero-day vulnerabilities.

**Target Sectors.** Figure 5a depicts the distribution of 12 sectors targeted by APT campaigns for 10 years, demonstrating notable shifts in attack patterns over time. The Government and Defense sector consistently accounts for a substantial share of APT activities, with an average of 65.6 incidents per year. The second most targeted is the Corporations and Businesses sector, which has emerged as a prominent focus, averaging 48.5 incidents annually. Notably, the third most frequently targeted sector varies across the years. The Education and Research sector ranks third in 6 out of 10 years, with a peak of 35 cases in 2021 (possibly due to the COVID-19 pandemic). Meanwhile, the Critical Infrastructure sector has seen fluctuations, peaking 29 in 2016 and declining since 2019.

**Initial Attack Vectors.** Figure 5b illustrates the distribution of 12 initial attack vectors in APT attacks over the past decade. According to our investigation, malicious documents appear the most preferred initial access in APT campaigns, averaging 54.6 incidents per year. The second most common attack vector is spear phishing, with an average of 53.6 cases annually. This sophisticated phishing often entails sending fraudulent emails or deceptive messages to carefully selected individuals or organizations, rendering the attack more convincing. Malicious documents are often combined with spear phishing to gain the initial access to target system. Furthermore, across the whole decade, vulnerability exploitation consistently ranks as the third most frequently employed attack vector, with an average of 38.9 incidents per year. This technique involves leveraging software vulnerabilities to gain unauthorized access to the system. Interestingly, watering hole attacks [5], ranked as one of most utilized attack vectors in 2014, with 32 instances, exhibited a gradual decline afterwards, with only 4 cases recorded in 2023.

**Key Takeaway 1:** Over the past decade, APT campaigns have impacted 154 countries worldwide where US, IN, and KR remain frequently targeted. Although 446 unique threat actors have been identified, a small subset (Lazarus, APT28, APT29) accounts for a significant portion of attacks. Zero-day usage peaked between 2014 and 2016 but has declined in recent years. Government and corporate sectors stay the most targeted. Malicious documents and spear phishing dominate as initial penetration vectors.

## 4.2 CTI Records in APT Campaigns

This section explores how APT campaigns can be described across common threat intelligence sources, including a common vulnerability database, attack frameworks, and indicator of compromise.

**MITRE IDs.** We extract a total of 2,582 MITRE ATT&CK techniques [19] from 175 TRs available in our collection (11.6% in Table 6), among which 263 technique IDs are unique. Considering the whole 359 distinct identifiers in the MITRE ATT&CK framework, this reveals that APT campaigns harness a wide spectrum of techniques to facilitate intrusions. Table 7 demonstrates the 10 most frequently observed techniques (22.5% of all instances), along with their descriptions, tactics, occurrences, and proportional representation. As expected, the most commonly observed tactics in APT campaigns include ① execution that involves the running code on local or remote systems; ② defense evasion that aims to bypass detection mechanisms; and ③ discovery that focuses on reconnoitering intelligence about internal systems or networks. Additional frequently observed tactics are consistent with the advanced and stealthy characteristics of APT campaigns, such as command and control, exfiltration, persistence, and privilege escalation.

[5]A watering hole attack targets a specific group by compromising websites they frequently visit with malicious code. Upon successful exploitation, attackers can gain access target systems and exfiltrate information while maintaining long-term control.

**Table 7: Top 10 most frequently observed MITRE ATT&CK techniques [19] in APT campaigns over the past decade. Notably, Execution, Defense Evasion, and Discovery are the most common tactic categories, using Command and Scripting Interpreter and Application Layer Protocol being among the most prevalent. These tactics are well aligned with the nature of APTs (Section 4.2).**

| MITRE ID | Description | Tactic | Count | Ratio |
|---|---|---|---|---|
| **T1059** | Command/scripting interpreter | Execution | 77 | 3.0% |
| **T1071** | Application layer protocol | Command and control | 76 | 2.9% |
| **T1082** | System information discovery | Discovery | 65 | 2.5% |
| **T1027** | Obfuscated files or information | Defense evasion | 60 | 2.3% |
| **T1140** | Deobfuscate/decode files or information | Defense evasion | 56 | 2.2% |
| **T1041** | Exfiltration over C2 channel | Exfiltration | 54 | 2.1% |
| **T1204** | User execution | Execution | 51 | 2.0% |
| **T1053** | Scheduled task/job | Execution, persistence, privilege escalation | 49 | 1.9% |
| **T1083** | File/directory discovery | Discovery | 47 | 1.8% |
| **T1036** | Masquerading | Defense evasion | 45 | 1.7% |

**CVEs.** We extract a total of 1,088 CVEs from 416 TRs available in our collection (27.6% in Table 6), among which 431 CVEs are unique. It is noted that multiple CVEs may be leveraged in a single APT campaign to achieve full-chain exploitation. Notably, the vulnerabilities exhibit high severity levels, with an average score of 8.5. Table 8 presents the top 10 most frequently observed CVEs (22.8% of all instances), along with their severity scores, vulnerability types, the number of affected software, occurrences, and proportions. Our further analysis shows that Microsoft Windows and Office are the most commonly targeted platforms/software, relating to 90% of the identified CVEs. Unsurprisingly, the most prevalent vulnerability types include remote code execution (RCE), memory corruption, use-after-free (UAF), stack overflow and privilege escalation (PE). Notably, the CVE with the highest severity score (9.3), CVE-2015-5119, is associated with the exploitation of Adobe Flash Player, which was officially discontinued back in 2020.

**YARA Rules.** We extract a total of 419 YARA rules from 131 TRs available in our collection (8.7% in Table 6), among which 419 rules are unique. We hypothesize that the limited availability of YARA rules may be attributed to the sensitive nature of APT campaigns. Contributing factors include: ① ongoing private investigations, ② disclosure restrictions imposed by non-disclosure agreements or internal policies, ③ the risk of rule evasion or misuse by threat actors, and ④ high variability and obfuscation in malware samples, which complicates the creation of generalizable detection rules.

**Table 8: Top 10 most frequently exploited vulnerabilities in APT campaigns. The APT groups show a strong preference for remote code execution (RCE) and memory corruption vulnerabilities, with CVE-2012-0158 being the most widely exploited (5.4%). High severity vulnerabilities such as privilege escalation (PE) and use-after-free (UAF) are also commonly leveraged. (Section 4.2). Note that we follow vulnerability naming from the National Vulnerability Database [11].**

| CVE | Severity | Vuln | Affected S/W | Count | Ratio |
|---|---|---|---|---|---|
| **CVE-2012-0158** | 8.8 (High) | RCE | 19 | 59 | 5.4% |
| **CVE-2017-11882** | 7.8 (High) | Memory Corruption | 4 | 44 | 4.0% |
| **CVE-2017-0199** | 7.8 (High) | RCE | 8 | 33 | 3.0% |
| **CVE-2018-0802** | 7.8 (High) | Memory Corruption | 4 | 20 | 1.8% |
| **CVE-2015-5119** | 9.8 (Critical) | UAF | 7 | 18 | 1.7% |
| **CVE-2015-1641** | 7.8 (High) | Memory Corruption | 11 | 16 | 1.5% |
| **CVE-2010-3333** | 7.8 (High) | Stack Overflow | 8 | 15 | 1.4% |
| **CVE-2014-6332** | 9.3 (High) | RCE | 11 | 15 | 1.4% |
| **CVE-2015-1701** | 7.8 (High) | PE | 3 | 15 | 1.4% |
| **CVE-2014-4114** | 7.8 (High) | RCE | 10 | 13 | 1.2% |

**Key Takeaway 2:** APT campaigns leverage a diverse set of intrusion techniques, with 2,582 MITRE ATT&CK instances, highlighting frequent use of tactics like execution, defense evasion, and discovery. Analysis of 1,088 CVEs reveals that Windows operating systems are the most targeted platforms, with remote code execution being the most common vulnerability type. The CVEs are highly severe, averaging the score of 8.5. On the other hand, YARA rules are scarce in public APT reports, likely due to confidentiality concerns, evasion risks, and the technical challenges posed by malware variability.

## 4.3 Common Traits of APT Campaigns

This section explores the underlying characteristics of APTs, mostly focusing on concealment and aggressiveness.

**APT Duration.** Figure 6 presents the distribution of attack durations and their cumulative distribution function (CDF). Our findings indicate that approximately half of APT campaigns have lasted five months or less (137 days), while the remaining half have extended beyond that duration. We observe a significant variation in attack durations, ranging from a single day to nearly five years. The longest recorded campaign spans from June 2011 to April 2016 (1,766 days), which is associated with Project Sauron [36]. Attributed to a suspected US-based threat actor, this campaign primarily targets governmental and research institutions in Russia, Iran, Rwanda, and Italy. The second longest campaign lasts 1,706 days, and was linked to the Iranian APT group Ajax Security Team as part of Operation Saffron Rose [64]. This operation was first detected on July 12, 2009, and remained active until March 2014. At the other end of the spectrum, the shortest attack lasts only a single day, which targeted France's TV5Monde broadcasting network [4]. This incident, attributed to Cyber Caliphate that is reportedly linked to the Islamic State of Iraq and Syria, results in a shutdown on April 8, 2015. The second shortest attack, spanned just two days, involves the distribution of malware via Hangul document files [3].

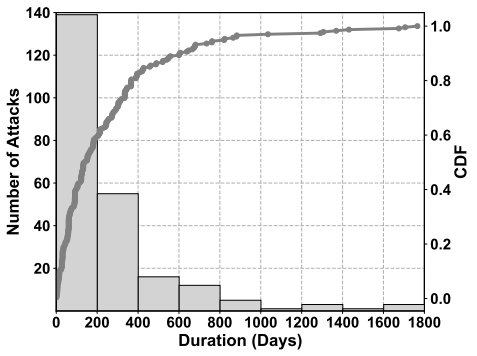



**Figure 6: CDF and histogram of APT campaign durations from 235 cases (Table 6). The plot shows that around half of APT incidents lasted 137 days or fewer, while the remaining half extended beyond this duration. Notably, the longest recorded campaign persisted for 1,766 days, whereas the shortest lasted only a single day (Section 4.3).**

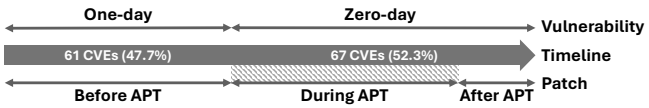



**Figure 7: CVEs exploited in APT campaigns based on patch availability relative to the attack timeline. We consider a zero-day vulnerability when CVE (with a patch) becomes available during (shadowed region) or after APT campaigns. 61 one-day vulnerabilities (47.7%) were disclosed and patched prior to the APT campaign, whereas 67 zero-day vulnerabilities (52.3%) were exploited before a patch was available (Section 4.3).**

The campaign was observed between December 9 and 11, 2014, primarily targeting South Korea's power infrastructure.

**Vulnerabilities and Patches.** We further investigate 62 TRs that include both CVE identifiers and corresponding attack durations, recognizing 128 distinct CVEs. We assume that the CVE release date corresponds to the availability of a patch. Figure 7 displays the relationship between the attack timelines and patch availability. Our findings reveal that around half of those CVEs (67 or 52.3%) are exploited as zero-day vulnerabilities, while the remaining half CVEs (61 or 47.7%) had been patched prior to the associated APT campaigns. On average, the time required to develop and release a patch for a zero-day vulnerability is approximately 200 days.

**Two-sided Nature as Both Attacker and Victim.** APT campaigns reveal the dual role of nations, where a country can simultaneously act as both an aggressor and a victim. The APT campaigns in our dataset reveal that 23 *countries have been involved as attackers at least once*, while 154 *countries have been targeted as victims in one or more incidents*. We analyze the top 20 most attacking and victimized countries over the past decade, which account for 44.7% of all APT cases. Notably, Figure 8 uncovers a significant asymmetry between attacker and victim nations. For instance, the CN-US attack ratio is 31:1, indicating that China has conducted 31 times more attacks against the United States than the reverse. Similarly, both

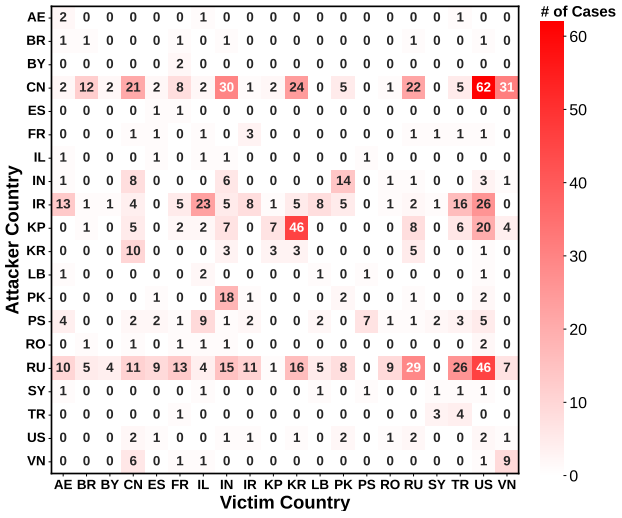



**Figure 8: Heatmap that depicts the distribution of APT campaigns based on the 20 origins (attacker country) and destinations (victim country), accounting for approximately 45% of the whole incidents over the past decade. The darker shades in each cell indicate the higher frequencies of APT cases. Notable patterns include a high volume of attacks originating from China, Russia, and North Korea, primarily targeting the United States, South Korea, and other regions. Notably, some countries are observed to engage in self-directed attacks (Section 4.3).**

the KP-KR and RU-US attack ratios stand at around 15:1, reflecting substantial imbalances in cyber offensive activity. Russia, China, and Iran emerge as the most active attacking country, responsible for 229, 232, and 125 APT cases (based on horizontal occurrences). Meanwhile, the United States, South Korea, and India stay the most frequently targeted, with 174, 95 and 89 incidents. (based on vertical occurrences). Figure 8 also highlights the three most prominent bilateral attack relationships: namely, China targeting the US (62 cases), North Korea targeting South Korea (46 cases), and Russia targeting the US (46 cases). It is noteworthy to mention that not all the origin countries for APT groups are known.

**Self-directed APT Attacks.** A closer examination of Figure 8 unveils that, in some APT campaigns, the origin and target countries are identical (*i.e.,* main diagonal values of the matrix). For instance, Russia has been observed targeting its own systems in 29 cases, while China appears to have done so in 21 cases. Our further investigation identifies several contributing factors to these self-directed attacks. First, APT campaigns may target individuals within the same country, often focusing on political dissidents or human rights activists [45, 63, 66, 85]. These campaigns are typically conducted by domestic threat actors aligned with government interests. Second, such campaigns may also be directed at foreign companies operating within the country, particularly in sectors such as banking and payment systems [14, 37]. Third, geopolitical and territorial disputes may drive this behavior, especially in politically tense regions such as the Russia–Ukraine border [29, 40, 73, 104]. Finally, a few self-targeted incidents appear to result from unusual circumstances. For instance, the Longhorn threat actor –linked to the United States– compromised a US system, only to deploy an uninstaller within hours, suggesting the attack was likely unintentional [49]. Another notable case involves APT17 [31], a China-associated group that reportedly targeted Chinese entities suspected of leaking sensitive information domestically [46].

**Key Takeaway 3:** APT campaigns exhibit a wide range of durations, from a single day to nearly five years. About half of the 128 analyzed vulnerabilities are exploited as zero-days, while the rest are one-days. Notably, 23 countries have acted as attackers and victims in APT operations, revealing the dual-role nature of many nations and significant asymmetries (*e.g.,* China launching 31 times more attacks on the United States than the opposite). Some campaigns were self-directed, with countries like Russia and China targeting domestic entities, often due to political repression or foreign company surveillance.

## 4.4 External Dynamics of APT Campaigns

Although APT operations are typically goal-driven and backed by sponsors, uncovering behind these attacks or external factors inherently remains a significant challenge. This section examines our curated collection of news articles and media reports from the web, discussing any known relationships (*e.g.,* triggers, motives, or impacts) between global affairs and APT activities. We classify such external dynamics into four categories: political events, international conflicts, global pandemics, and economic gains.

**Political Events.** APT groups frequently exploit politically sensitive events to further their objectives. A notable example was the Russia-affiliated group Fancy Bear (APT28) [122], which launched a spear phishing campaign targeting the Democratic National Committee (DNC) during the 2016 US presidential election [52]. As part of this operation, APT28 successfully breached systems linked to the presidential campaign, resulting in the theft of sensitive data. In a similar vein, Fancy Bear targeted Macron's presidential campaign during the 2017 French election. The group infiltrated campaign systems, stealing credentials and sensitive data [43]. Fancy Bear is also suspected of compromising the infrastructure of a German political party. By infiltrating the network, the group harvested critical information for political advantage [41].

**International Conflicts.** APTs linked to international conflicts often involve the intersection of geolocational tensions and cyber warfare. For instance, during the Russo-Ukrainian War [124], the Russian-sponsored APT group known as Sandworm [125] hacked Ukrainian energy infrastructure [117] back in 2015, causing power outages that affected 230K consumers. Sandworm also attempted to disrupt the news agency in Ukraine [110] afterwards. Similarly, around 2023, the notorious Russian cyberespionage group, named Fancy Bear, launched an attack on a critical energy facility [100].

**Global Pandemics.** Beginning in late 2019, the COVID-19 pandemic brought the unprecedented impacts across various sectors worldwide [118]. Figure 5a shows a noticeable rise in APT attacks targeting the healthcare sector, with 14 cases (4.3%), and the education and research sector, with 33 cases (10.0%) in 2020, coinciding

with the global spread of the virus. During this period, the China-sponsored APT41 [121] group exploited vulnerabilities in remote desktop services to attack healthcare organizations [42]. Similarly, the Lazarus [123] group attempted to steal COVID-19-relevant intelligence [88] by targeting a pharmaceutical company. Additionally, CozyBear (APT29) [119] has been suspected of attempting to steal the COVID-19 vaccines information [12]. The World Health Organization (WHO) was not spared, as the APT group called DarkHotel launched a password-stealing attack against WHO staff [126].

**Economic Gains.** APT activities are associated with the financial sector, often driven by economic gain. For instance, the Carbanak group (FIN7) [116] gained notoriety for its cyberattacks on Russian banking institutions in 2016 [54], coinciding with Russia's gradual recovery from a prolonged economic recession [39]. With the rise of cryptocurrencies, many threat actors have shifted their focus to these digital assets as primary targets. Notably, the Lazarus group, reportedly sponsored by North Korea, was implicated to a major cryptocurrency theft in 2022 [10]. Both FIN7 and Lazarus are recognized for their persistent targeting of financial institutions, underscoring the significant economic motivations behind their operations.

**Key Takeaway 4:** APT groups frequently exploit politically sensitive events, such as presidential elections, to infiltrate voting systems and influence public perception. Geopolitical conflicts often coincide with cyber operations, with Russian APTs launching attacks during international tensions. During the COVID-19 pandemic, healthcare and research sectors experienced a spike in APT campaigns (*e.g.,* APT41, Lazarus, APT29), targeting pandemic-related organizations. Economic motives also drive APT activity (*e.g.,* FIN7, Lazarus), conducting cyberattacks on financial institutions and cryptocurrency platforms reflecting a persistent focus on financial gain.

## 4.5 Visual Representations

We design an interactive map to visualize APT campaigns, enabling users to explore detailed information by selecting either an attacking or victimized country. The map integrates decade-long historical data by year, presenting key attributes such as associated threat actors, source(s), CVE identifier(s), initial attack vector(s), related malware, targeted sector(s), and estimated attack duration (when such information is available in the corresponding technical report). More importantly, the map maintains *up-to-date information by dynamically retrieving content* from technical reports using an LLM. Note that the source has been currently linked to TR#1 [22]. In addition, we incorporate a timeline chart that links APT campaigns to relevant news articles for contextual reference. Finally, we display an additional interactive diagram that visualizes the relationships between the top 10 threat actors and the 30 most frequently targeted countries over the past decade.

# 5 Discussion and Limitations

This section discusses the limitations of our work.

**Representativeness of APT Campaigns.** Although our dataset covers a wide spectrum of APT campaigns, recent incidents likely remain unreported or undocumented. For instance, publicly available technical reports in recent years appear less comprehensive, as shown by the decline observed in 2023 (Table 2). Besides, due to the covert nature of APT operations, capturing every case is inherently infeasible. While we chose three independent repositories to include as many APT instances as possible, community-aggregated datasets may still introduce selection biases and coverage gaps. Nonetheless, we believe that the collective intelligence drawn from both official and unofficial sources worldwide provides a representative sample, which would be sufficient to approximate a meaningful reflection of the broader ground truth.

**Limited Responses from an LLM.** While extracting targeted information using an LLM can be highly effective, it comes with several limitations related to accuracy, reliability, and completeness. First, PDF files are unstructured, leading to extract misordered or misaligned content during extraction. Complex layouts including tables and figures could further result in broken or fragmented text. Second, long documents must be segmented into smaller chunks due to the limited context window of LLMs, which may disrupt contextual coherence. Third, LLMs are susceptible to hallucination, potentially creating inaccurate or fabricated information when encountering incomplete or ambiguous input. Lastly, extracted text may contain extraneous or irrelevant content such as page numbers, headers, footers, legal disclaimers, or noises. With the possibly above reasons, our empirical assessments with sampled TRs (Table 4) demonstrate that the GPT-4-Turbo model achieves around 90% F1 scores. Nevertheless, we acknowledge that our study represents a best-effort approach, as the retrieval process is inherently dependent on the interpretive capabilities of the LLM.

**Attack Duration.** We understand that estimating the attack duration of APT campaigns is inherently challenging due to the covert and persistent nature of these operations. First, the reports may have inconsistent timelines across multiple reports by different organizations. Second, the reported dates may be distorted due to the unwillingness of a victim or national security reasons. Third, reliance on public sources may miss internal or classified timelines. Lastly, inferring the end date may be open because detecting the last known activity is difficult to determine. This work attempts to reconstruct durations based on the (known) records.

**CVE and Patch Timing.** In many cases, a patch is available at the time of CVE publication. However, the timing of a CVE's release does not always guarantee patch availability, as it relies on responsible disclosure practices. Note that our study aims to approximate the timeline required to develop patches for zero-day vulnerabilities exploited in APT campaigns.

**Future Work.** While we carefully examine attack durations (Section 4.3) and attacker motivations (Section 4.4), there remain additional opportunities to gain deeper insights. One promising direction is to further investigate the relationship between attack duration and attacker objectives. Another is to uncover the evolution of persistent techniques and remediation trends over time.

# 6 Related Work

We categorize prior APT works into three main areas, including the detection and evaluation, Cyber Threat Intelligence, dataset regarding APT campaigns.

**APT Detection and Evaluation.** A substantial body of research has focused on the detection and evaluation of APT campaigns, including the development of frameworks such as HOLMES [72], ProvG-Searcher [27], Zimba *et al.* [1], Marchetti *et al.* [74], and Hassan *et al.* [113]. Notably, MAGIC [48] leverages graph neural networks and self-supervised learning to construct behavior graphs from system logs, applying a masked graph learning strategy to capture relationships between entities and events. On the other hand, CAPTAIN [114] introduces a rule-based intrusion detection system, which enhances traditional provenance-based approaches by learning fine-grained detection rules through gradient descent optimization. Wang *et al.* [115] propose a provenance graph-based detection framework by reconstructing attack chains and countering adversarial strategies through empirical evaluation. Meanwhile, Shen *et al.* [98] further contribute to the field by systematically evaluating the effectiveness of modern Endpoint Detection and Response (EDR) systems in detecting and mitigating adversarial tactics commonly used by APT groups in realistic attack scenarios. Malik *et al.* [108] present a multi-layered mitigation framework, evaluating network-based, host-based, and AI-driven detection methods. Our work differs from previous approaches, with a focus on trends and insights over the last decade from a macroscopic perspective (*e.g.,* temporal and global analysis).

**Cyber Threat Intelligence.** Several studies have concentrated on the collection and evaluation of CTI. With an in-depth analysis of 22 APT reports, Ussath *et al.* [67] identify common tactics, tools across different stages of APT campaigns. Similarly, Bahrami *et al.* [8] propose a taxonomy based on the cyber kill chain model to systematically categorize the tactics, techniques, and procedures (*i.e.,* TTPs) by threat actors. Kumarasinghe *et al.* [53] present a multi-stage ranking framework designed to identify and prioritize the most relevant MITRE ATT&CK techniques by leveraging a combination of pre-trained and fine-tuned language models. Meanwhile, TTPHunter [76] introduce an automated system that applies machine-learning-based sentence classification to extract MITRE ATT&CK-aligned TTPs from unstructured APT reports. Note that our work uses an LLM to retrieve pre-defined questions from TRs.

**APT Dataset.** TREC [62] collects a dataset by capturing APT behaviors at the kernel level, making it publicly available. Similarly, Kumarasinghe *et al.* [53] introduce an open benchmark dataset for training and validating varying techniques used in APT campaigns, which contains threat behavior descriptions from real-world reports and MITRE ATT&CK technique identifiers. Siracusano *et al.* [99] provide a collection of publicly available technical reports and corresponding CTI information. Moreover, Stojanović *et al.* [13] provide a comprehensive review of existing APT datasets and the frameworks employed to model attacks for the development of automated detection techniques. Their analysis highlights the limited availability of publicly accessible datasets and discusses the considerable challenges involved in collecting realistic, high-quality attack data. In this work, we compile several (large) collections of prior technical reports (*e.g.,* 2,563 before dataset refinement), threat actor profiles, and related news articles.

## 7 Conclusion

This study presents a comprehensive, decade-long (from 2014 to 2023) analysis of APT campaigns, offering a macroscopic perspective of how these threats have evolved across countries, sectors, and attack techniques. By leveraging a hybrid information retrieval approach that combines LLM inference and rule-based extraction, we systematically process and analyze over 1,500 technical reports, revealing key trends in threat actor behavior, vulnerability exploitation, and campaign duration. Our key findings highlight the global reach, persistent nature, and strategic targeting patterns of APT groups, while also uncovering the contextual influence of geopolitical events, global crises, and economic motivations. Although most observations align with our expectations, our findings also discover interesting (but less known) facts such as a wide range of attack durations, self-directed attacks, notable prevalence of one-day CVEs. To support ongoing research and awareness in the area of APTs, we release an interactive visualization platform, which promotes deeper engagement with the evolving threat landscape.

## Acknowledgments

We thank the anonymous reviewers for their constructive feedback. This work was partially supported by the grants from Institute of Information & communications Technology Planning & Evaluation (IITP), funded by the Korean government (MSIT; Ministry of Science and ICT): No. RS-2022-II221199, No. RS-2024-00437306, No. RS-2024-00337414, No. RS-2025-25457342, and No. RS-2025-25394739. Additional support was provided by the National Science Foundation under awards CNS-2126654, CNS-2440819, and DGE-2335798. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.

## References

[1] Zhaoshun Wang Aaron Zimba, Hongsong Chen and Mumbi Chishimba. 2020. Modeling and Detection of the Multi-Stages of Advanced Persistent Threats Attacks Based on Semi-Supervised Learning and Complex Networks Characteristics. *Future Generation Computer Systems* (2020).

[2] Ankur Chowdhary Adel Alshamrani, Sowmya Myneni and Dijiang Huang. 2019. A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities. *IEEE Communications Surveys & Tutorials* (2019).

[3] AhnLab. 2014. Vulnerability, Malicious Code Appeared in the MBR Destruction Function Using Hangul File. https://asec.ahnlab.com/ko/1015/.

[4] AhnLab. 2015. Targeted Attack on France TV5Monde. https://orkl.eu/libraryEntry/01871652-6001-4c34-b246-181978941024.

[5] amCharts. 2025. amCharts 4 Documentation. https://www.amcharts.com/docs/v4/.

[6] Awadhesh Kumar Singh Amit Sharma, Brij B. Gupta and V. K. Saraswat. 2023. Advanced Persistent Threats (APT): Evolution, Anatomy, Attribution and Countermeasures. *Journal of Ambient Intelligence and Humanized Computing* (2023).

[7] Gemini Team Google: Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, and David Silver et al. 2024. Gemini: A Family of Highly Capable Multimodal Models. *arXiv preprint arXiv:2312.11805* (2024).

[8] Pooneh Nikkhah Bahrami, Ali Dehghantanha, Tooska Dargahi, Reza M. Parizi, Kim-Kwang Raymond Choo, and Hamid H. S. Javadi. 2019. Cyber Kill Chain-Based Taxonomy of Advanced Persistent Threat Actors: Analogy of Tactics, Techniques, and Procedures. *Journal of Information Processing Systems* (2019).

[9] Kiran Bandla and Santiago Castro. 2025. Data. https://github.com/aptnotes/data.

[10] BBC. 2022. North Korean Hackers Target Gamers in $615m Crypto Heist - US. https://www.bbc.com/news/world-asia-61036733.

[11] Harold Booth, Doug Rike, and Gregory A. Witte. 2013. The National Vulnerability Database (NVD): Overview. *ITL Bulletin, National Institute of Standards and Technology* (2013).

[12] Becky Bracken. 2020. Pfizer COVID-19 Vaccine Targeted in EU Cyberattack. https://threatpost.com/pfizer-covid-19-vaccine-cyberattack/162170/.

[13] Katharina Hofer-Schmitz Branka Stojanović and Ulrike Kleb. 2020. APT Datasets and Attack Modeling for Automated Detection Methods: A Review. *Computers & Security* (2020).

[14] Nick Carr. 2017. Cyber Espionage Is Alive and Well: APT32 and the Threat to Global Corporations. https://cloud.google.com/blog/topics/threat-intelligence/cyber-espionage-apt32/.

[15] CERT-UA. 2025. CERT-UA Threat actors. https://cert.gov.ua/search/UAC.

[16] CISA. 2020. Iranian Advanced Persistent Threat Actors Threaten Election-Related Systems. https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-296b.

[17] Cybersecurity & Infrastructure Security Agency (CISA). 2025. US-CERT. https://www.cisa.gov/.

[18] The Mitre Corporation. 2024. Common Vulnerabilities and Exposures (CVE). https://cve.mitre.org/.

[19] The Mitre Corporation. 2024. MITRE ATT&CK. https://attack.mitre.org/.

[20] Andrea Cristaldi. 2025. APTmap. https://github.com/andreacristaldi/APTmap/.

[21] CrowdStrike. 2025. CrowdStrike Adversaries. https://www.crowdstrike.com/adversaries/.

[22] CyberMonitor. 2024. APT & Cybercriminals Campaign Collection. https://github.com/CyberMonitor/APT_CyberCriminal_Campagin_Collections.

[23] Gérard Wagener Cynthia Wagner, Alexandre Dulaunoy and Andras Iklody. 2016. MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security (WISCS '16)*.

[24] Steffen Enders Daniel Plohmann, Martin Clauß and Elmar Padilla. 2017. Malpedia: A Collaborative Effort to Inventorize the Malware Landscape. *The Journal on Cybercrime and Digital Investigations* (2017).

[25] Lizzie Dearden. 2017. Emmanuel Macron Campaign Hack: French Presidential Candidate Targeted by Cyber Attack Similar to DNC Leak. https://www.independent.co.uk/news/world/europe/emmanuel-macron-leaks-hack-en-marche-cyber-attack-russia-dnc-marine-le-pen-election-france-latest-a7721796.html.

[26] Dragos. 2025. Dragos Threat Groups. https://www.dragos.com/threat-groups/.

[27] Hüsrev Taha Sencar Enes Altinisik, Fatih Deniz. 2023. ProvG-Searcher: A Graph Representation Learning Approach for Efficient Provenance Graph Search. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23)*.

[28] Electronic Transactions Development Agency (ETDA). 2019. Electronic Transactions Development Agency. https://apt.etda.or.th/cgi-bin/aptgroups.cgi.

[29] Robert Falcone and Jen Miller-Osborn. 2016. Scarlet Mimic: Years-Long Espionage Campaign Targets Minority Activists. https://unit42.paloaltonetworks.com/scarlet-mimic-years-long-espionage-targets-minority-activists/.

[30] Fraunhofer FKIE. 2024. Malpedia Inventory. https://malpedia.caad.fkie.fraunhofer.de/library.

[31] Fraunhofer FKIE. 2025. APT17. https://malpedia.caad.fkie.fraunhofer.de/actor/apt17.

[32] International Organization for Standardization. 2025. ISO 3166 Country Codes. https://www.iso.org/iso-3166-country-codes.html.

[33] Fortinet. 2025. Indicators Of Compromise (IoCs). https://www.fortinet.com/resources/cyberglossary/indicators-of-compromise.

[34] Stanislav Golovanov. 2021. pdfkit 1.0.0. https://pypi.org/project/pdfkit/.

[35] Google. 2025. Advanced persistent threats (APTs). https://cloud.google.com/security/resources/insights/apt-groups?hl=en.

[36] GREAT. 2016. ProjectSauron: Top Level Cyber-Espionage Platform Covertly Extracts Encrypted Government Comms. https://securelist.com/faq-the-projectsauron-apt/75533/.

[37] GREAT. 2019. Recent Cloud Atlas Activity. https://securelist.com/recent-cloud-atlas-activity/92016/.

[38] GREAT. 2024. APT Trends Report Q2 2024. https://securelist.com/apt-trends-report-q2-2024/113275/.

[39] World Bank Group. 2016. Russian Economy Inches Forward, Says World Bank. https://www.worldbank.org/en/news/press-release/2016/11/09/russian-economy-inches-forward-says-world-bank.

[40] Josh Grunzweig. 2017. DragonOK Updates Toolset and Targets Multiple Geographic Regions. https://unit42.paloaltonetworks.com/unit42-dragonok-updates-toolset-targets-multiple-geographic-regions/.

[41] Claudio Guarnieri. 2015. Digital Attack on German Parliament: Investigative Report on the Hack of the Left Party Infrastructure in Bundestag. https://netzpolitik.org/2015/digital-attack-on-german-parliament-investigative-report-on-the-hack-of-the-left-party-infrastructure-in-bundestag/.

[42] Health Sector Cybersecurity Coordination Center (HC3). 2020. APT41 Citrix and Zoho Attacks on Healthcare. https://www.hhs.gov/sites/default/files/apt41-citrix-and-zoho-attacks-on-healthcare.pdf.

[43] Alex Hern. 2017. Macron Hackers Linked to Russian-Affiliated Group Behind US Attack. https://www.theguardian.com/world/2017/may/08/macron-hackers-linked-to-russian-affiliated-group-behind-us-attack/.

[44] OpenAI: Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, and Alec Radford et al. 2024. GPT-4o System Card. *arXiv preprint arXiv:2410.21276* (2024).

[45] Amnesty International. 2021. Click and Bait: Vietnamese Human Rights Defenders Targeted with Spyware Attacks. https://www.amnesty.org/en/latest/research/2021/02/click-and-bait-vietnamese-human-rights-defenders-targeted-with-spyware-attacks/.

[46] Intrusiontruth. 2019. Encore! APT17 Hacked Chinese Targets and Offered the Data for Sale. https://intrusiontruth.wordpress.com/2019/07/25/encore-apt17-hacked-chinese-targets-and-offered-the-data-for-sale/.

[47] Luke Jenkins and Dan Black. 2024. APT29 Uses WINELOADER to Target German Political Parties. https://cloud.google.com/blog/topics/threat-intelligence/apt29-wineloader-german-political-parties.

[48] Zian Jia, Yun Xiong, Yuhong Nan, Yao Zhang, Jinjing Zhao, and Mi Wen. 2024. MAGIC: Detecting Advanced Persistent Threats via Masked Graph Representation Learning. In *Proceedings of the 33rd USENIX Security Symposium (Security '24)*.

[49] A.L. Johnson. 2017. Longhorn: Tools Used by Cyberespionage Group Linked to Vault 7. https://community.broadcom.com/symantecenterprise/communities/community-home/librarydocuments/viewdocument?DocumentKey=7ca2e331-2209-46a8-9e60-4cb83f9602de.

[50] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data* (2019).

[51] Nils Kuhnert. 2025. APTMAP. https://aptmap.netlify.app/.

[52] Mohit Kumar. 2016. Hillary Clinton's Presidential Campaign Also Hacked in Attack on Democratic Party. https://thehackernews.com/2016/07/hillary-clinton-hacked.html.

[53] Udesh Kumarasingh, Ahmed Lekssay, Husrev Taha Senca, Sabri Boughorbe, Charitha Elvitigala, and Preslav Nakov. 2024. Semantic Ranking for Automated Adversarial Technique Annotation in Security Text. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security (ASIACCS '24)*.

[54] Kaspersky Lab. 2016. Carbanak and Beyond: Banks Face New Attacks. https://www.kaspersky.es/about/press-releases/carbanak-and-beyond-banks-face-new-attacks?srsltid=AfmBOooc537SIDl8XO0gOiV-ILgYN7dcprjGliqXabXLdW6MgegfKZv1.

[55] Kaspersky Lab. 2025. Kaspersky Lab. https://www.kaspersky.com/.

[56] Kaspersky Lab. 2025. MAP | Kaspersky Cyberthreat Live Map. https://cybermap.kaspersky.com/.

[57] Kaspersky Lab. 2025. Targeted Cyberattacks Logbook. https://apt.securelist.com/.

[58] LangChain. 2022. LangChain. https://www.langchain.com/.

[59] LangChain. 2025. PyPDFLoader. https://python.langchain.com/docs/integrations/document_loaders/pypdfloader/.

[60] Denis Legezo. 2016. InPage Zero-Day Exploit Used to Attack Financial Institutions in Asia. https://securelist.com/inpage-zero-day-exploit-used-to-attack-financial-institutions-in-asia/76717/.

[61] Musarubra US LLC. 2025. Trellix. https://www.trellix.com/.

[62] Mingqi Lv, HongZhe Gao, Xuebo Qiu, Tieming Chen, Tiantian Zhu, Jinyin Chen, and Shouling Ji. 2024. TREC: APT Tactic / Technique Recognition via Few-Shot Provenance Subgraph Learning. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24)*.

[63] Asheer Malhotra and Kendall McKay. 2021. Transparent Tribe APT Expands Its Windows Malware Arsenal. https://blog.talosintelligence.com/transparent-tribe-infra-and-targeting/.

[64] Mandiant. 2014. Operation Saffron Rose: Iranian Threat Actors Conduct Cyber Espionage Against U.S. Targets. https://www.infopoint-security.de/medien/fireeye-operation-saffron-rose.pdf.

[65] Mandiant. 2025. Mandiant Threat Intelligence. https://www.mandiant.com.

[66] Bill Marczak, Jakub Dalek, Sarah McKune, Adam Senft, John Scott-Railton, and Ron Deibert. 2018. BAD TRAFFIC: Sandvine's PacketLogic Devices Used to Deploy Government Spyware in Turkey and Redirect Egyptian Users to Affiliate Ads? https://citizenlab.ca/2018/03/bad-traffic-sandvines-packetlogic-devices-deploy-government-spyware-turkey-syria/.

[67] Feng Cheng Martin Ussath, David Jaeger and Christoph Meinel. 2016. Advanced Persistent Threats: Behind the Scenes. In *Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS '16)*.

[68] Meta. 2025. React. https://react.dev/.

[69] Trend Micro. 2014. Trend Micro. https://www.trendmicro.com/en_us/business.html.

[70] Microsoft. 2025. How Microsoft Names Threat Actors. https://learn.microsoft.com/en-us/defender-xdr/microsoft-threat-actor-naming?view=o365-worldwide.

[71] Microsoft. 2025. Microsoft. https://www.microsoft.com/.

[72] Sadegh M. Milajerdi, Rigel Gjomemo, Birhanu Eshete, R. Sekar, and V.N. Venkatakrishnan. 2019. HOLMES: Real-Time APT Detection through Correlation of Suspicious Information Flows. In *Proceedings of the 2019 IEEE Symposium on Security and Privacy (S&P '19)*.

[73] Aleksandar Milenkoski. 2023. Gaza Cybergang | Unified Front Targeting Hamas Opposition. https://www.sentinelone.com/labs/gaza-cybergang-unified-front-targeting-hamas-opposition/.

[74] Michele Colajanni Mirco Marchetti, Fabio Pierazzi and Alessandro Guido. 2016. Analysis of High Volumes of Network Traffic for Advanced Persistent Threat Detection. *Computer Networks* (2016).
[75] MISP. 2025. Misp-Galaxy. https://github.com/MISP/misp-galaxy.
[76] Vikas Maurya Nanda Rani, Bikash Saha and Sandeep Kumar Shukla. 2023. TTPHunter: Automated Extraction of Actionable Intelligence as TTPs from Narrative Threat Reports. In *Proceedings of the 2023 Australasian Computer Science Week (ACSW '23)*.
[77] Palo Alto Networks. 2025. Palo Alto Networks. https://unit42.paloaltonetworks.com/.
[78] Google News. 2024. Google News. https://news.google.com/.
[79] Federal Bureau of Investigation (FBI). 2025. FBI. https://www.fbi.gov/services.
[80] U.S. Department of the Treasury. 2024. Treasury Sanctions China-Linked Hackers for Targeting U.S. Critical Infrastructure. https://home.treasury.gov/news/press-releases/jy2205.
[81] OpenAI. 2024. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* (2024).
[82] OpenAI. 2024. New Embedding Models and API Updates. https://openai.com/index/new-embedding-models-and-api-updates/.
[83] OpenAI. 2025. OpenAI Models. https://platform.openai.com/docs/models.
[84] The North Atlantic Treaty Organization. 2025. NATO. https://www.nato.int/.
[85] Ovi. 2023. RE:archive | Reverse Engineering APT37's GOLDBACKDOOR Dropper. https://www.0x0v1.com/rearchive-goldbackdoor/.
[86] Pierluigi Paganini. 2020. Japanese Kawasaki Heavy Industries Discloses Security Breach. https://securityaffairs.com/112765/data-breach/kawasaki-heavy-industries-cyber-attack.html.
[87] Pallets. 2024. Flask Documentation. https://flask.palletsprojects.com/en/stable/.
[88] Seongsu Parkk. 2020. Lazarus Covets COVID-19-Related Intelligence. https://securelist.com/lazarus-covets-covid-19-related-intelligence/99906/.
[89] Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20)*.
[90] Alexander Rogan. 2024. The Unseen Siege: China's Persistent Cyber Offensive Against U.S. Critical Infrastructure. https://energycentral.com/c/pip/unseen-siege-chinas-persistent-cyber-offensive-against-us-critical-infrastructure.
[91] Salesforce. 2025. Heroku. https://www.heroku.com/.
[92] Salesforce. 2025. Tableau. https://www.tableau.com/.
[93] Sectrio. 2024. Complete Guide to Advanced Persistent Threat (APT) Security. https://sectrio.com/blog/complete-guide-to-apt-security/.
[94] Secureworks. 2025. Secureworks Threat Profiles. https://www.secureworks.com/research/threat-profiles.
[95] IBM Security. 2025. IBM X-Force Hive. https://exchange.xforce.ibmcloud.com/search/hive.
[96] IBM Security. 2025. IBM X-Force ITG. https://exchange.xforce.ibmcloud.com/search/ITG.
[97] RSA Security. 2025. RSA Security. https://www.rsa.com/.
[98] Xiangmin Shen, Zhenyuan Li, Graham Burleigh, Lingzhi Wang, and Yan Chen. 2024. Decoding the MITRE Engenuity ATT&CK Enterprise Evaluation: An Analysis of EDR Performance in Real-World Environments. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security (ASIACCS '24)*.
[99] Giuseppe Siracusano, Davide Sanvitom, Roberto González, Manikantan Srinivasan, Sivakaman Kamatchi, Wataru Takahashi, Masaru Kawakita, Takahiro Kakumaru, and Roberto Bifulco. 2023. Time for aCTIon: Automated Analysis of Cyber Threat Intelligence in the Wild. *arXiv preprint arXiv:2307.10214* (2023).
[100] Dark Reading Staff. 2023. Russia's 'Fancy Bear' APT Targets Ukrainian Energy Facility. https://www.darkreading.com/cyberattacks-data-breaches/russia-fancy-bear-apt-ukrainian-energy-facility.
[101] StrangerealIntel. 2023. EternalLiberty. https://github.com/StrangerealIntel/EternalLiberty.
[102] Tableau. 2024. Sankey. https://exchange.tableau.com/products/932.
[103] Threat Hunter Team. 2024. Carderbee: APT Group Use Legit Software in Supply Chain Attack Targeting Orgs in Hong Kong. https://www.security.com/threat-intelligence/carderbee-software-supply-chain-certificate-abuse.
[104] Check Point Software Technologies. 2022. Cloud Atlas Targets Entities in Russia and Belarus Amid the Ongoing War in Ukraine. https://research.checkpoint.com/2022/cloud-atlas-targets-entities-in-russia-and-belarus-amid-the-ongoing-war-in-ukraine/.
[105] Tines. 2025. IoCParser. https://iocparser.com/.
[106] Orlaith Traynor. 2024. Top Threat Actors on the Dark Web | 2023 Recap. https://cybelangel.com/top-threat-actors-on-the-dark-web-recap/.
[107] Jakob Truelsen and Ashish Kulkarni. 2022. wkhtmltopdf. https://wkhtmltopdf.org/.
[108] Nandan Sharma Vaibhav Malik, Anirudh Khanna and Suryaprakash Nalluri. 2024. Advanced Persistent Threats (APTs): Detection Techniques and Mitigation Strategies. *International Journal of Global Innovations and Solutions* (2024).
[109] Venafi. 2025. Venafi. https://venafi.com/.
[110] Jai Vijayan. 2023. Russia's Sandworm APT Launches Swarm of Wiper Attacks in Ukraine. https://www.darkreading.com/cyberattacks-data-breaches/russia-sandworm-apt-swarm-wiper-attacks-ukraine.
[111] VirusTotal. 2022. YARA Documentation. https://yara.readthedocs.io/en/stable/index.html.
[112] Vx-Underground. 2025. Vx-Underground. https://vx-underground.org/.
[113] Adam Bates Wajih Ul Hassan and Daniel Marino. 2020. Tactical Provenance Analysis for Endpoint Detection and Response Systems. In *Proceedings of the 2020 IEEE Symposium on Security and Privacy (S&P '20)*.
[114] Lingzhi Wang, Xiangmin Shen, Weijian Li, Zhenyuan Li, R. Sekar, Han Liu, and Yan Chen. 2025. Incorporating Gradients to Rules: Towards Lightweight, Adaptive Provenance-based Intrusion Detection. In *Proceedings of the 32nd Annual Network and Distributed System Security Symposium (NDSS '25)*.
[115] Yuntao Wang, Han Liu, Zhendong Li, Zhou Su, and Jiliang Li. 2024. Combating Advanced Persistent Threats: Challenges and Solutions. *IEEE Network* (2024).
[116] Wikipedia. 2024. FIN7. https://en.wikipedia.org/wiki/FIN7.
[117] Wikipedia. 2025. 2015 Ukraine Power Grid Hack. https://en.wikipedia.org/wiki/2015_Ukraine_power_grid_hack.
[118] Wikipedia. 2025. COVID-19 Pandemic. https://en.wikipedia.org/wiki/COVID-19_pandemic.
[119] Wikipedia. 2025. Cozy Bear. https://en.wikipedia.org/wiki/Cozy_Bear.
[120] Wikipedia. 2025. Cyber Threat Intelligence. https://en.wikipedia.org/wiki/Cyber_threat_intelligence.
[121] Wikipedia. 2025. Double Dragon (Hacking Group). https://en.wikipedia.org/wiki/Double_Dragon_(hacking_group).
[122] Wikipedia. 2025. Fancy Bear. https://en.wikipedia.org/wiki/Fancy_Bear.
[123] Wikipedia. 2025. Lazarus Group. https://en.wikipedia.org/wiki/Lazarus_Group.
[124] Wikipedia. 2025. Russo-Ukrainian War. https://en.wikipedia.org/wiki/Russo-Ukrainian_War.
[125] Wikipedia. 2025. Sandworm (Hacker Group). https://en.wikipedia.org/wiki/Sandworm_(hacker_group).
[126] Davey Winder. 2020. 'Elite Hackers' Thought Behind Cyber Attack on World Health Organization. https://www.forbes.com/sites/daveywinder/2020/03/25/hackers-target-world-health-organization-as-cyber-attacks-double-during-covid-19-pandemic/.
[127] Jonathan Yerushalmy. 2024. China Cyber-Attacks Explained: Who Is Behind the Hacking Operation Against the US and UK? https://www.theguardian.com/technology/2024/mar/26/china-cyber-attack-uk-us-explained-hack-apt-31.
[128] Zawya. 2024. Positive Technologies: Cyberattackers Targeting Telecommunications and the Military-Industrial Complex in the Middle East. https://www.zawya.com/en/press-release/research-and-studies/positive-technologies-cyberattackers-targeting-telecommunications-and-the-military-industrial-complex-in-the-middle-east-f1aazssc.