From Concealment to Exposure: Understanding the Lifecycle and Infrastructure of APT Domains

Athanasios Avgetidis[1], Aaron Faulkenberry[1], Vinny Adjibi[1], Tillson Galloway[1], Panagiotis Kintis[1], Omar Alrawi[1], Zane Ma[3], Fabian Monrose[1], Angelos D. Keromytis[1], Roberto Perdisci[2], and Manos Antonakakis[1]

[1]Georgia Institute of Technology
[2]University of Georgia
[3]Oregon State University

Abstract—**Advanced Persistent Threats (APTs) are sophisticated and long-lived attacks that are often backed by nation-states. Despite the security community's efforts to design and deploy specialized systems to combat them, APTs have remained prevalent while persisting undetected for significantly more time than commodity cyber threats. In this paper, we measure this difference by conducting the first longitudinal analysis of APT infrastructure by shedding light on the lifecycle of their domain names. To enable this study, we build Atropos, a novel measurement methodology that automatically and accurately labels DNS records of APT domain names, enabling us to understand their lifecycle and gain a more comprehensive and contextualized infrastructure picture than the one that is shared in public reports. Using the comprehensive infrastructure view that Atropos provides, we study 405 APT actors over a period spanning a decade and unveil several novel findings regarding their utilization of network infrastructure that have practical implications.**

We find that APT actors provision their IPs to their domain names 317 days on average before an attack is publicly reported. Furthermore, 73.6% of the APT IPs that are part of the attack infrastructure no longer point to their domains at the time of first public disclosure, highlighting that researchers and security practitioners need to consider historic DNS data in order to get a more comprehensive and accurate picture when training network detection, investigation, or attribution systems. Organizations that are more sensitive to APT attacks will need to retain network logs for at least 19 to 25 months in order to have higher probabilities of discovering whether they have been a target of an APT attack. Finally, we provide evidence that APT actors re-use hosting providers, deploy APT network infrastructure close to their intended attack targets, and increasingly utilize more cloud-fronting. These findings are important because they can guide future threat detection and attribution works.

Index Terms—**Cybersecurity, APT, DNS, machine learning**

I. Introduction

Advanced Persistent Threats (APTs) are attacks conducted by well-organized, well-funded, and technically sophisticated actors [2]. The term APT, likely coined in 2006 by analysts of the United States Air Force [14], is used to differentiate commodity and low-sophistication operations (e.g., script kiddies) from those that are more complex and often backed by nation-states and sophisticated crime syndicates. The sophisticated and unique *modus operandi* of these actors—as captured by MITRE's cyber kill chain [74]—has led to specialized mechanisms for APT detection and investigation [53], [32], [6], [39], [67]. Despite active APT research, recent attacks have continued to cause widespread damage, such as the SolarWinds supply chain attack that forced more than 18,000 customers (including the US government) to install malicious code [27] or the 2025 *Bybit hack* [68] that stole $1.5 billion worth of digital tokens.

Prior work on APTs has been mainly focused on detection and investigation systems [53], [32], [6], [39], [67], [35], [47], either aiming to identify APT attacks in real-time, or to support forensic investigations. Measurement studies have focused on understanding the attack surfaces of organizations targeted by APT actors [80], the vulnerabilities they exploit [22], the tactics, techniques, and procedures (TTPs) they employ [65], or sophisticated attacks against specific targets [44] and regions [51]. Despite the prior work to understand and combat APT attacks, APT investigations still remain a highly manual effort done by experts [70]. Among the top challenges expert APT analysts currently face is that the "lack of automation and validation in data ingestion impacts the use of historical threat data [70]." While these challenges are evident across different signals of APT investigations, such as TTPs and malware, they also pose a major problem in the utilization of Indicators of Compromise (IoCs), such as domain names and IPs, which remain the primary signals for APT attribution [70]. Aside from aiding expert APT analysts in investigation and attribution efforts, characterizing and contextualizing the network infrastructure (i.e., domains and IPs) of APTs, which is lacking from public reports and threat intelligence [79], [29], can help us answer and quantify research questions that are still largely unanswered. For instance, the network infrastructure comprehensiveness of public threat reports, the longevity of APT infrastructure before disclosure, and the infrastructure utilization trends and similarities of APT groups over the years are still open research questions. Answering these questions can help the community devise more comprehensive defensive strategies, develop more effective attribution systems by utilizing network attributes, and understand how long organizations need to keep network logs in order to detect whether they have been a victim of an APT attack, considering that APTs are particularly persistent compared to commodity threats, thus requiring higher log retention windows.

One of the main challenges in trying to answer the aforementioned research questions is the fact that the relationship

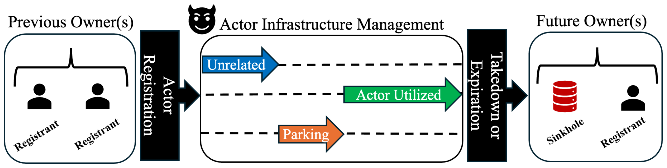

Fig. 1: Lifecycle of an actor-controlled domain name. Multiple owners and infrastructure types complicate forensics.

between the network infrastructure used to orchestrate an attack and the APT actors is transitory, as is evident in their domain names [45], [1] and as shown in Figure 1. For instance, an APT actor can register a previously expired domain name, park it at parking infrastructure, point it to their attack infrastructure for a few days, and then let it expire or be taken down. Another challenge is that APT attacks can persist for years, and the actors can dynamically change the IP addresses utilized by their domain names. Thus, to comprehensively and accurately identify the network infrastructure associated with an APT domain and its lifetime, a forensic analyst needs access to a dataset that is capable of witnessing the historical IP changes, has to filter out unrelated and noisy infrastructure (e.g., parking and sinkhole, etc.), and finally pinpoint the infrastructure and period of time in which each domain name was ”active”. These challenges diminish the usefulness of ”as-is” network IoCs extracted from threat reports, requiring analysts to invest manual effort in enriching, contextualizing, and validating them, which is time-consuming [3], and is typically conducted on a per-incident basis [24].

In this work, we reduce the knowledge gap in the network infrastructure of APT attacks by performing the first longitudinal study of APT infrastructure used by 405 APT actors over a period spanning a decade. We focus on measuring and expanding the comprehensiveness of the publicly known IP infrastructure of APT attacks by enriching known, high-confidence APT domain names appearing across 2,188 APT reports with historical DNS data. To this end, and considering the measurement challenges we discussed, we utilize two historical DNS datasets [78], [42] that witness changes to over 1,100 generic top-level domains (gTLDs) daily, and a novel measurement methodology that automatically and accurately characterizes historic APT infrastructure. Our novel measurement methodology, called Atropos, filters and labels domain-to-IP mappings – Resource Records – related to known domains of APT actors, while discarding IP addresses that are unrelated to APT attacks (i.e. parking, sinkholes, etc.), providing needed automation that expands, validates and contextualizes historical threat data, which has been recently characterized as a major challenge by APT experts [70]. Our contributions are as follows:

- A novel measurement methodology that expands and contextualizes the network infrastructure of known APT domain names and offers three times the IP visibility and domain contextualization than that of public threat reports. The source code of Atropos can be found at: https://github.com/Astrolavos/Atropos/.
- The largest and most comprehensive APT infrastructure analysis to date, spanning over a decade and 405 APT actors.
- We quantify the time window during which organizations need to keep network logs to identify the vast majority of the infrastructure of an APT attack. Our results show that the network logs should be preserved for at least 19 to 25 months.
- We find that while APT actors utilize a plethora of different hosting providers, they only re-use a small portion of them, while, over the years, the use of cloud-fronting has increased significantly. These findings verify expert knowledge [70] and highlight the difficulty of network forensics and attribution.

II. Challenges in Domain Lifecycle Analysis

APT network forensic investigations are often conducted as more of an art than a science. Among the many challenges that network forensic analysts must address, identifying the period of time in which an APT attack was active has traditionally been a highly manual process. In this section, we outline the main challenges investigators face in temporally bounding the active period of the APT attack (Section II-A), then put these challenges into perspective using the SolarWinds attack as a case study (Section II-B), and finally outline the scope and requirements we need to measure the lifecycle of APT domains (Section II-C).

A. Forensic Challenges

Identifying the attack-related (i.e., actor-utilized) IP infrastructure that an APT domain pointed to during a cyberattack and its likely active time window is challenging. The first major challenge comes from the fact that an APT domain name can historically have multiple previous or future owners other than the APT actors [45], [5]. Before an APT actor registers or gains control of a domain name, the domain name can be associated with previous owners whose IP infrastructure is unrelated to the attack. After detection or disclosure, an APT domain can be taken down, sinkholed, or left to expire until it is re-registered by some other legitimate or malicious entity [5]. Such infrastructure and time windows have to be identified by the forensic analysts as unrelated to the attack.

The second major challenge comes from the fact that even during the time window that the domain name is managed by the actors, not all of the IP infrastructure that it points to is related to the attack. For example, after its registration by the APT actors, the domain name could point to its registrar's default parking infrastructure for a period of hours, days, or months [82]. The APT actors also may choose to park the domain at a benign IP outside their control (e.g., an IP with a positive Internet reputation) for “aging” reasons and to establish network reputation, since newly registered domains with no network history are often more suspicious than long-lived ones [28]. Other actors may choose to point and periodically move the domain to arbitrary infrastructure in order to inject deliberate noise into passive and active DNS repositories. However, when a domain name is effectively used in an attack (i.e., to deliver exploits, as a social engineering domain, command and control, or exfiltration point), it must point to the actor-utilized infrastructure.

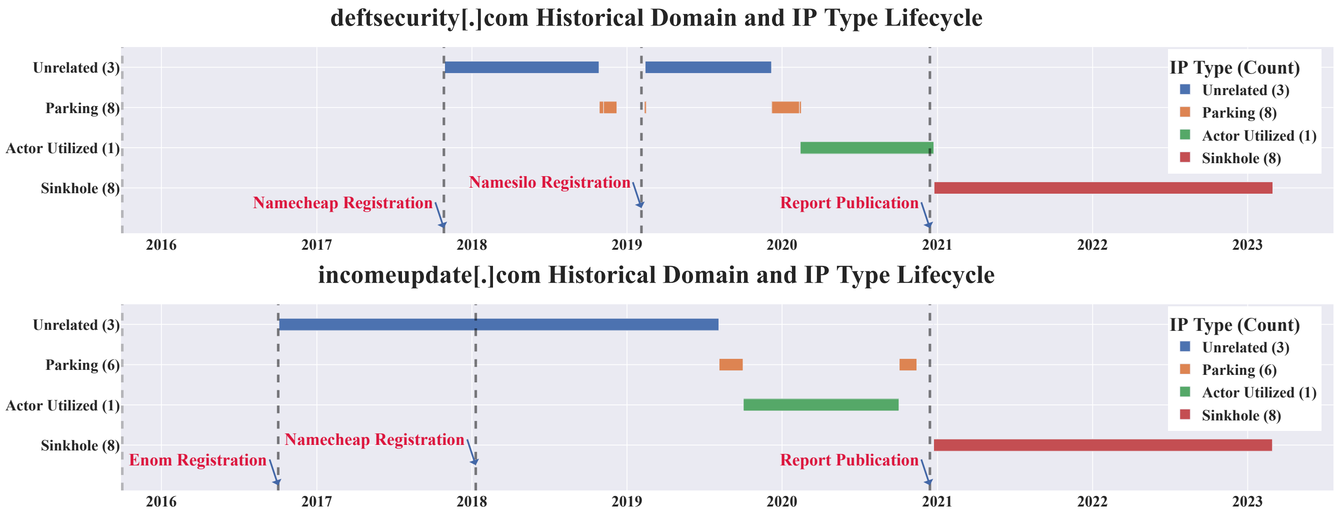

Fig. 2: Domain and IP lifecycle of deftsecurity[.]com and incomeupdate[.]com sunburst domain names initially reported in [49]. In this work, we seek to automatically identify the actor-utilized IPs (colored in green). The numbers inside the parentheses reflect the number of unique IPs of each category.

While some, or all of these events may occur, there is no definitive lifecycle of an APT domain. As a forensic investigator tries to piece together the timeline of the APT attack, we can assume that they will uncover a combination of the actor-utilized hosting infrastructure used in the attack, infrastructure belonging to previous owners of the domain, parking infrastructure, sinkholes, and even deliberate noise added by the actors.

Considering the aforementioned challenges, it is clear that extracting the actor-utilized hosting infrastructure of a domain name is no easy task. To make matters worse, it is also particularly hard to collect a clean and complete picture of the infrastructure to conduct further analyses, as APT domain names are usually utilized by nation-states and high-profile adversaries that do not — obviously — share any information about their operations.

Next, we summarize four main challenges that network forensic investigators face as they characterize the lifecycle and infrastructure of APT domain names. Across all these challenges, we use the term “domain lifetime” to reflect all observable time periods in which a domain name existed.

Non-Actor Ownership. This is the period in the domain lifetime in which an actor does not own a domain due to it belonging to a different owner. As illustrated in Figure 1, this can occur either before the actor registers the domain, or after the domain expires or is taken down.

Sinkholing. This is the period in which the domain points to sinkhole infrastructure. Sinkholing occurs after domain detection and results in a malicious domain pointing to infrastructure controlled by security companies and professionals, registrars, and law enforcement agencies [5].

Various Forms of Parking. This is the period in which the domain points to various forms of parking infrastructure. This infrastructure may be placeholder registrar-controlled parking infrastructure shortly after the domain name is registered, or parking infrastructure where the actors can point their domain names to age them until they use them for their operation.

Deliberate Noise Injection. Actors can point their domain to infrastructure that is not under their control to gain a positive (benign) reputation before they use the domain in an attack. Such an action could easily inject noise into passive and active DNS repositories, effectively making the network forensic investigation of the APT attack significantly harder.

B. *Placing The Challenges In Context: The SolarWinds Attack*

Next, we put the four major classes of challenges into perspective by using the attack against SolarWinds as an example. Figure 2 showcases the historical lifecycles of two domains used in the *SolarWinds* attack, deftsecurity[.]com and incomeupdate[.]com. It also depicts the corresponding IP infrastructure that can be discovered from publicly available DNS datasets [42]. For this example, these IP addresses were labeled taking into account the public reporting of the attack [49], [23] and manual threat analysis from multiple analysts. Although the two domains were used in the same attack, their lifecycles differ in registrar, IP infrastructure, and pre-registration activity. These differences alone make the analysis of the two APT domains used in the same attack and controlled by the same actor non-trivial.

Starting with the domain registration patterns, APT actors re-registered domains from two different registrars that had completely different hosting histories (possibly because of their previous domain owners). By just utilizing WHOIS data and manually trying to pinpoint the most likely window of actor registration, a forensic analyst would not be able to identify which IPs were the ones utilized by the adversaries and used in the attack. That is because there are multiple parking and other unknown IPs in the historical DNS data — even in the period where the actors likely re-registered these two domains. Thus, to identify the actor-utilized IPs, an analyst would need more than just a temporal window of interest.

One solution to filter out the non-actor-utilized IPs would be to use publicly available lists of parking IPs and DNS nameserver infrastructure [48]. By doing so, utilizing the IP and DNS nameserver data from [82], manually inspecting the DNS nameservers, and identifying various parking infrastructure, we could only additionally remove a subset of the publicly known parking infrastructure (colored in orange).

While this methodology has reduced the amount of infrastructure to inspect, it is still not sufficient, as the domain names have been pointing to cloud infrastructure (Amazon and Unified Layer, colored in blue after their latest registration in Figure 2) which has not been attributed to the SolarWinds attack due to its large temporal distance (many months before the attack took place). An analyst, knowing the timeline of the attack and manually inspecting the properties of this unknown cloud infrastructure, would filter out these IPs as likely parking and inactive infrastructure and yield only the actor-utilized IPs as they have been publicly reported [25], [49]. The practice of registering domain names years before their utilization and strategically aging them on infrastructure other than the attack infrastructure has been documented in prior reports [36], [56]. Evidently, filtering out all of these non-attack-related IPs is a non-trivial and labor-heavy process, often left to expert analysts. In this work, we seek to automatically identify the actor-utilized IPs of historical APT domain names in a transparent way and with a low false positive rate.

C. Observations and Takeaways

By taking into consideration the challenges and the lessons learned from the SolarWinds campaign, we arrive at the following three observations: first, APT domains feature unique lifecycles that can differ even within the same campaign, second, these lifecycles can last multiple years, and domain registrations, and third, APT domains can be associated with a diverse set of infrastructure (e.g., parking, sinkhole, etc.) that is often not associated with the actor-utilized IPs. Thus, measuring the lifecycle of APT domains requires:

- A historical dataset that observes and logs the infrastructure changes in APT domains across the years.
- A methodology that filters and labels the IP infrastructure associated with the APT domains and considers the diverse infrastructure types we discussed.
- A methodology that is applicable on a per-domain basis.

To satisfy the aforementioned requirements we take the following steps: first, we utilize two historical DNS datasets that span over a decade and capture the changes in DNS resolutions of 405 APT actors and second, we develop a novel system that filters and labels these historical DNS resolutions taking into account the diverse infrastructure we encountered on our case study and operates on a per domain basis with high accuracy. Next, we discuss the datasets and measurement methodology in more detail.

III. Datasets and Methodology

This section introduces the OSINT datasets (Section III-A) we use to study 405 APT groups as outlined by our visibility in Table III. Then we proceed by diving deep into Atropos (Section III-C), its modules, and how these modules enable Atropos to reliably and accurately identify actor-utilized infrastructure.

A. OSINT Datasets

Threat Actor Information. We utilize the threat actor information from the MISP Galaxy project [54]. This dataset consists of a set of known APT actors, their attributed country code, and a list of all their known aliases. The MISP Galaxy threat actors dataset is also used by the popular threat encyclopedia Malpedia [63] and is more comprehensive than that of MITRE [55]. In our study, we only consider IoCs that have been attributed to these known threat actors.

Threat Report IoCs. To build a set of known APT domain names and IPs, we utilize publicly available threat reports. Threat reports have been highly utilized in prior works to gather IoC datasets related to APT threats [66] and are considered a quality data source as the IoCs shared in them are published by reputable security vendors. We extract threat report IoCs from two well-known data sources. The first data source is AlienVault Open Threat Exchange (OTX) [4]. AlienVault OTX is a large, open threat intelligence community that has released more than 19 million IoCs to date. In our study, we only consider IoCs that map to 1,859 threat reports published on the Alienvault's user account between 2014 and 2025. The second threat report dataset to extract APT IoCs is CyberMonitor [20]. CyberMonitor is an aggregation of popular APT threat reports and datasets such as APTnotes and others, that have been heavily used in former works [50], [9], [66]. We manually parse a subset of 329 threat reports from this data source that were published between April 2013 and June 2019, with the intent of extracting four attributes: APT domains, APT IPs, publication date of the report, and name of the APT actor that is associated with the domain names and IPs.

By combining the two datasets and looking at only threat reports attributed to known threat actors, according to the **Threat Actor Information**, and filtering out reports that mention multiple actors, the final threat report dataset consists of 2,188 APT reports, which is larger than previous APT studies [3], [80]. Table I shows the top 10 publishers in terms of the IoCs we utilize in this study. It is important to note that most of our indicators come from reputable security vendors, and we do not consider IoCs that come without a published report in order to minimize potential noise in our APT datasets from unreliable sources such as random users in the AlienVault community [15]. Table II shows the distribution of APT IoCs for each of the two datasets that together sum up to 31,398 domains and 7,533 IPs.

DNS Resource Records. To populate the Historic DNS Database of Atropos in Figure 3, we use two historical DNS resource records datasets. The first dataset is the **Historical Active DNS dataset**, from the ActiveDNS [42] project, which scans daily millions of domain names from over 1,100 gTLDs and has been utilized in many prior measurement works [7], [76], [52]. The historical DNS records span from January 2016 to January 2025 and include A, AAAA, NS, NX, MX, and SOA query responses. To complement the coverage of ActiveDNS, we use a premium VirusTotal API access to gather historical DNS resource records for all of the 31,398 domain

TABLE I: Coverage of IOCs for the top 10 publishers in terms of reports. Overall, we utilize a total of 2,188 APT reports.

	No. of	No. of	No. of IOCs		
Publisher	**Reports**	**APTs**	**Domain**	**e2LD**	**IP**
Palo Alto Networks	133	81	3024	2738	706
Kaspersky Lab	126	81	2574	1803	392
Trend Micro	90	63	1386	960	441
ESET	80	51	679	607	459
FireEye	65	50	1373	1228	151
Symantec	63	50	972	903	241
Proofpoint	57	45	1091	622	105
Talos	52	38	1906	1680	211
SentinelOne	40	32	940	835	159
Tencent	37	25	308	262	41

names from our Threat Report IoC dataset. This dataset, also referred to as the **Historical VirusTotal DNS**, amounts to 480,093 DNS resource records and spans back to April 2013.

VirusTotal API (VT). To generate needed features for Atropos' feature extraction module, which we detail in Section III-C, we query the VirusTotal API. Features $f_4 - f_7$ for the VirusTotal DNS records and features $f_{12} - f_{22}$ for both Active DNS and VirusTotal DNS records are being gathered by querying this data source.

Parking and Sinkholes. We utilize parking IPs and DNS nameservers from both an academic publication and Maltrail [82], [73], as well as manually labeling the DNS nameservers of APT records to identify parking ones. Additionally, we utilize sinkhole IPs and DNS nameservers from an academic publication and a public list [5], [73], as well as manually labeling the DNS nameservers of APT records to identify sinkholes.

Compromised Domains. To filter out compromised domain names, we remove the APT domains that were mentioned to be compromised in the reports they were published in from the CyberMonitor [20] source. Additionally, we also filter out compromised domain names based on an aggregation of compromised domain list [83], which includes various reputable sources such as abuse.ch and SANS.

B. DNS Datasets and Threat Reports IP Visibility

Since we are mainly interested in identifying actor-utilized IPs to study the infrastructure they utilize, we can just gather the high-confidence domain names and IPs that appear in our 2,188 APT reports and utilize our DNS datasets to match them. This way, we will only utilize known domains and IPs that threat analysts in reputable reports have identified. Table III presents the visibility of our DNS data sources on the report Fully Qualified Domain Names (FQDNs), effective Second Level Domains (E2lds) [26] — i.e., the registrable portion of a domain name —, APT actors, and resource records (RRs) after removing NX records and bogon IPs [21] (e.g., unroutable, private, loopback networks). As we can see, both these DNS sources together can provide at least one IP for 90.84% of the APT FQDNs and 98.06% of the APT actors, showcasing that our DNS datasets have significant IP coverage for the APT domain names.

With this DNS visibility, we can now match the APT domains and IPs that get shared on APT reports and see what percentage of domain names threat reports can characterize with an IP. When we do so, we can see that only 23.52% of the FQDNs and 67.31% of the APT actors can be characterized as demonstrated in Table III. Clearly, if we just utilize the APT report domain and IPs, we would only characterize less than a quarter of the historical APT domains, even with DNS data sources that have over 90% APT domain coverage. Evidently, there are legitimate reasons why APT report authors may not have IP-level visibility of the domain names they have identified or may choose not to share all the IPs they have identified. For example, the IPs that APT actors use can belong to virtual hosting or cloudfronting providers and serve both benign and malicious domains at the same time. Thus, the report authors may omit such IPs from the reports to avoid readers blocklisting them and causing harm to benign services. Additionally, report authors may lack the historical DNS datasets to identify the actor-utilized IPs. We find that the percentage of APT reports that share both domains and IPs is only 44.22% of all reports that share at least one domain. The size imbalance has also been demonstrated in prior work [17]. Thus, conducting a comprehensive APT network infrastructure study cannot be done just by utilizing threat report information.

Takeaways: Simply matching known APT domains to known APT IPs from APT reports using popular DNS data can only characterize **23.52%** of the APT domains. We find that only **44.22%** of the APT reports sharing domains also share IPs, which further substantiates the coverage concerns of threat intelligence that prior works have raised [16], [79].

C. Measurement Methodology

Considering the lack of comprehensive OSINT visibility in domain-to-IP mappings and infrastructure coverage, we need to develop a measurement methodology to expand the APT infrastructure coverage and conduct a representative measurement study. However, as we have described in Section II, identifying the actor-utilized IPs of an APT domain is challenging. Previous works have tried to address similar problems [1], [48], [45], but they are largely not applicable to address all the challenges and satisfy all the requirements we have set. To address these shortcomings and characterize more domains than those that APT reports alone can, we develop a simple supervised model that we call Atropos, which automatically filters and identifies actor-utilized IP addresses of known APT domains. More specifically, Atropos ingests domain-to-IP mappings (i.e., DNS Resource Records — RRs) from DNS data, only for domains that appear in APT reports, and identifies which RRs correspond to infrastructure likely used by the APT actors. Atropos uses a combination of different OSINT datasets and three inline analytical modules which are described below.

*1) **Enrichment and Filtering Module**:* **(Enrichment)** Atropos first ingests the set of APT domain names from the **Threat Report IoCs** (i.e, 31,398 domains in Step 1a) in its Enrichment and Filtering Module. To get historical infrastructure visibility for these APT domains, it utilizes (Step 1b) the Historical DNS Database and gets all the DNS records that

TABLE II: Major datasets utilized in the study.

Type	Source	Dataset	Time Span	Number of Records
Threat Actor Information	MISP Galaxy [54]	Threat Actor Information	2025-03-04 to 2025-03-04	750
Threat Report IoCs	AlienVault OTX[4]	APT Domains	2014-12-02 to 2025-03-01	27,709
Threat Report IoCs	AlienVault OTX[4]	APT IPs	2014-12-02 to 2025-03-01	5,171
Threat Report IoCs	Cybermonitor[20]	APT Domains	2013-04-13 to 2019-06-26	6,621
Threat Report IoCs	Cybermonitor[20]	APT IPs	2013-04-13 to 2019-06-26	2,616
DNS Resource Records	Active DNS Project [42]	Historical Active DNS	2016-01-01 to 2025-01-31	119,959,784
DNS Resource Records	VirusTotal Resolutions [77]	Historical VirusTotal DNS	2013-04-01 to 2025-03-06	480,093
Compromised Domains	Zonefiles.io [83]	Compromised Domains	2013-03-20 to 2025-04-08	132,210
Parking and Sinkholes	Prior Work & Maltrail [82], [5], [73]	Parking and Sinkholes	2007-07-18 to 2024-03-13	85,509

TABLE III: A and AAAA resource record visibility after enriching the known APT domain names with our DNS data sources. APT IPs appearing on threat reports can only characterize 23.52% of APT FQDNs in popular DNS datasets.

Visibility Metrics	Threat Reports	Active DNS and VT	Report IPs Matched on Active DNS and VT
Timespan	2013-04 2025-03	2013-04 2025-03	2013-04 2025-03
FQDNs	31,398	28,524 (**90.84**%)	7,386 (**23.52**%)
E2lds	22,691	20,975 (92.43%)	5,392 (23.76%)
APT Actors	413	405 (98.06%)	278 (67.31%)
RRs	N/A	1,004,614	51,891

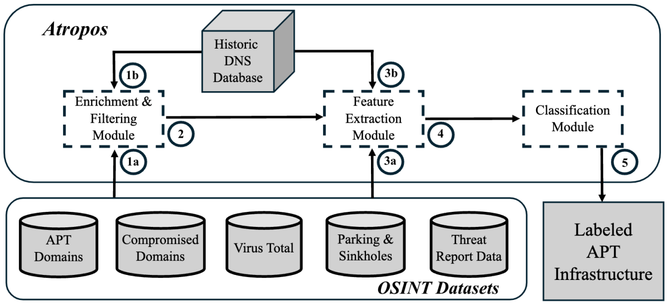

Fig. 3: An overview of Atropos. Atropos utilizes OSINT datasets and historical DNS data to label and filter APT infrastructure in a 3-step process.

were ever associated with the APT domains provided in Step 1a. The Historical DNS Database consists of all the historical DNS Resource Records (RRs) from Active DNS and VirusTotal and which are described in Table II. The output of the enrichment is a set of all the resource records that exist in the Historic DNS Database for the APT domains fed in Step 1a.

(Filtering) Atropos will then filter out all known compromised domains from the **Compromised Domains** dataset II, as well as bogon IPs that exist in the Cymru bogon list [21], non-existing domains (NXDOMAIN), and empty responses. Non-existent domains and empty responses are filtered by removing resource records with an RCODE number equal to 3 and resource records with the RDATA field equal to null. This filtering is necessary as such resource records will not be related to infrastructure provisioned by the APT actors for an attack campaign. Additionally, Atropos filters all domain names related to DNS fast-fluxing. Fast-fluxing is the process that involves the frequent change of the RRs of a domain name to many different IPs that can span hundreds or even thousands [34]. We consider such domain names out-of-scope of Atropos, as during development we found that the tactics and techniques of some fast-fluxing actors make them feature different lifecycles than those of typical APT domains, discussed in Section II, and thus require dedicated models. For example, the *gamaredon group* has been demonstrated to keep utilizing the same detected and reported domain names, long after the reporting of its attacks, thus having malicious activity after its detection [58], [57]. We leave the development of dedicated systems for such lifecycles to future work. To remove such domains from our dataset, inspired by Holz et. al. [34], we count the number of distinct IPs per domain and filter out the top 5% of the domain names in our dataset. This methodology filters out 1,497 domains with 544 IPs on average per domain, with 82.69% of these domain names belonging to gamaredon group, which, as we have described, is known for fast fluxing activities [58], [57]. Since only 5% of the domains have been filtered, we do not consider the impact of this filtering significant for the generalization of our methodology. Table III illustrates Atropos' visibility in the FQDNs, E2LDs, number of APT actors, and RRs after enrichment and before filtering. After all filtering, our DNS visibility spans over a decade, with at least one resource record for 26,615 (84.76%) of the FQDNs in all the reports published between April 2013 and March 2025. The total amount of resource records after the filtering is 195,051.

*2) **Feature Extraction Module**:* The next step of our methodology is to extract the features needed in order to train our models. Table IV illustrates the features of Atropos. Atropos generates features for each of the resource records that come out of the Enrichment and Filtering Module of Step 2. The features are generated by utilizing data extracted from **VirusTotal API** calls, the **Parking and Sinkholes** dataset, as well as the **DNS Resource Records**. We utilize a total of 22 features from four classes, namely temporal, infrastructure, OSINT, and domain name features. We pick our features based on historical forensic experience and argue about their utility. The features are extracted by a set of Python functions shared as part of our artifact. Next, we discuss the main classes of our features. An even more detailed description of the features can be found in the Appendix A.

Temporal Class (3 features): These features are meant to capture the proximity of an IP to the detection of an APT domain name and also their lifetime. Inspired by the *SolarWinds* case 2, the actor-utilized IPs were used close to, and shortly before the detection event, and had a

TABLE IV: The features of Atropos. Atropos utilizes 22 features from four distinct classes.

#f	Feature	Class	#f	Feature	Class
f_1	Detection and IP Fseen Delta	Temporal	f_{12}	IP Reputation	OSINT
f_2	Detection and IP Lseen Delta	Temporal	f_{13}	Number of Malicious Votes	OSINT
f_3	IP Lifetime	Temporal	f_{14}	Number of Harmless Votes	OSINT
f_4	Number of Historic Domains on IP	Infra.	f_{15}	Number of Malicious Analyses	OSINT
f_5	Mean Concurrent Domains on IP	Infra.	f_{16}	Number of Suspicious Analyses	OSINT
f_6	Median Concurrent Domains on IP	Infra.	f_{17}	Number of Undetected Analyses	OSINT
f_7	Number of IP Communicating Files	Infra.	f_{18}	Number of Harmless Analyses	OSINT
f_8	IP is Known Parking	OSINT	f_{19}	Num. of Domain Communicating Files	Domain
f_9	Nameserver is Known Parking	OSINT	f_{20}	Num. of Files Downloaded From Domain	Domain
f_{10}	IP is Known Sinkhole	OSINT	f_{21}	Number of Domain Subdomains	Domain
f_{11}	Nameserver is Known Sinkhole	OSINT	f_{22}	Number of Domain Certificates	Domain

lifetime of multiple months. On the contrary, the sinkhole IPs appeared after detection, and the domains first pointed to unrelated or previous owners' IPs long before their detection. We build the temporal features around these observations. We calculate (f_1) and (f_2) by computing the date difference between the domain's detection and the first and last seen of an IP on a domain name respectively, while (f_3) is the total number of days the IP was seen pointing to that domain.

Infrastructure Class (4 features): These features aim to characterize the IP infrastructure that a domain points to, by looking at other domains that point to the same IP. In our example (Section 2), the actor-utilized IPs had historically only one domain name pointing to them, while the parking IPs had a median of 9,014,949 domains pointing to them, and the sinkhole IPs a median of 664. We calculate (f_4) by counting the total number of domains ever pointed to this IP according to our DNS data sources and (f_5) and (f_6) by computing the mean and median number of domains pointed at that IP at the same time as the APT domain name.

OSINT Class (11 features): This class of features integrates OSINT knowledge around the IPs. Features (f_8) and (f_9) are binary features that report whether the IP or the nameserver of the domain is in known OSINT parking lists. Similarly we compute features (f_{10}) and (f_{11}) with known OSINT sinkhole lists. Features (f_{12}-f_{18}) are computed by querying the VirusTotal API regarding reputation, OSINT community votes, and OSINT analysis scans of each IP.

Domain Name Class (4 features): Domain name features capture differences in the utilization of a domain name by an APT actor on the domain name level that can aid the classification. We extract these features by querying the VirusTotal API and calculating the number of communicating and downloaded files (f_{19} and f_{20}) and number of subdomains and certificates (f_{21} and f_{22}) that each APT domain has.

*3) **Classification Module**:* The final step in our methodology is to feed the feature vectors generated at the Feature Extraction module to the classification Module. The classification module consists of a binary classifier that ingests the 22 features we have described and classifies each resource record as actor-utilized (True) or non-actor-utilized (False). During the development of Atropos, we experimented with various machine learning methods, including heuristics, Decision Trees, Support Vector Machines, Random Forests, XGBOOST [18], and Multi-Layer Perceptrons. In our experimental analysis, while other models had great performance, we found the Random Forest classifier to offer the best ROC AUC performance across datasets, while offering decision interpretability; thus, we picked this model over the rest. During its development, we trained and fine-tuned the hyperparameters only using our training dataset – to prevent data snooping [11] – and optimizing for ROC AUC with grid search. The optimal hyperparameters feature a depth of 10, 100 estimators, and the optimal number of features to consider at each split equal to their squared number. To showcase generalization, we tested Atropos on two out-of-distribution datasets. Finally, to demonstrate transferability across different DNS datasets, we train and test Atropos utilizing different models on each DNS dataset (ActiveDNS and VirusTotal) and show that accuracy is similar.

IV. Evaluation

In this section, we discuss the training and performance evaluation of Atropos. Atropos is trained and fine-tuned on a training dataset based on the public knowledge of public threat reports, which we call the Public Reports Dataset (**PR**). After Atropos is trained and fine-tuned, it is tested on two different test datasets that were not considered during development, with the aim of evaluating our methodology against potential sampling bias and overfitting. Atropos achieves 10-fold cross-validation accuracy scores of 98.16% and 98.90% on Active DNS and VirusTotal DNS datasets, respectively, demonstrating transferability, and accuracy scores of 91.39% and 95.38% when evaluated on the test datasets (**EA**) and (**FR**), respectively, demonstrating generalization.

A. Training and Evaluation Datasets

Collecting ground truth regarding the infrastructure of APT actors is very challenging. Two of the main reasons that contribute to this are that APT actors will not share their attack playbooks with the public and the fact that APT attacks are, by definition, sophisticated. Thus, to create our training and evaluation datasets, we take two steps. First, we utilize the public knowledge of domains and IPs existing in public threat reports, and second, we utilize three analysts for manual labeling. These analysts consist of two PhD students with seven and four years of experience in APT network forensics (*JA1* and *JA2* respectively) and one senior APT network analyst with over 20 years of experience (*SA*). The instructions given to the analysts were the following:

- You are given DNS resource records (RRs) of historical APT domains.
- Your task is to label these RRs as actor-utilized (True) or non-actor-utilized (False).
- A RR is actor-utilized when the IP corresponding to the domain is the infrastructure utilized in the APT operation.
- You can utilize any tool at your disposal to do so.
- Deliver a file with every RR you can confidently label.

Aside from the RRs, the analysts are also provided with open Internet access along with all the features generated, and they are allowed to perform any tasks to validate the correctness of their decision (e.g., reverse IP lookups, searching IPs in IP intelligence and other reports, etc.). The analysts were able to distinguish likely actor-utilized from non-actor-utilized IPs by considering a plethora of factors such as: a) the first and last observation of an IP to their APT domain name relative to the first public disclosure (e.g., actor-utilized IPs are more likely to be first observed before detection), b) the existence of an IP or a nameserver of the domain to known parking and sinkhole lists, c) the number and profile of other domains pointed to the same IP at the same time or historically (e.g., an IP that has hundreds of thousands other domains pointed to it concurrently and a random sample of them are displaying a parking page is a parking IP), and d) the time window of previous WHOIS registrations (e.g., IPs pointed to by the APT domain multiple registrations before the disclosure are likely previous owners of the domain). Next, we provide more details regarding each labeled dataset.

*1) (Training) Public Reports Dataset (**PR**):* This set incorporates the public knowledge from APT reports. As APT actors will not share their infrastructure with the public, the next most accurate set that can be utilized is that of report authors who have manually labeled the infrastructure and openly shared it in threat reports. For this dataset, we utilize all the APT domain-to-IP mappings (RRs) that have been publicly mentioned in the APT reports (i.e. **Threat Report IoCs** dataset) and have been matched together by using the Active DNS dataset, same as in Section III-B. However, these records only represent the positive class (i.e., actor-utilized) of the ground truth. To generate the negative class (i.e., not actor-utilized), and avoid class imbalance [11], we pick an equal amount of other random resource records from Active DNS, for the same domains that have a positive class record, and give all these records for manual labeling to analyst *JA1*. Analyst *JA1* confidently labels 1,915 out of 2,027 RRs and marks 1,065 RRs as actor-utilized and 851 RRs as non-actor-utilized. While the class distribution is not equal, the final dataset does not suffer from class imbalance [11], with 55.61% actor-utilized RRs and 44.43% non-actor-utilized RRs. Overall, this dataset consists of 1,915 resource records from 938 domains associated with 94 APT actors, from threat reports spanning from 2014-02-11 to 2023-04-13. To further evaluate *JA1* records for label inaccuracies [11], after *JA1* has completed the manual labeling, we give the same set of records and instructions to another junior analyst *JA2* from the same organization as *JA1* for labeling. After their inspection, we quantify the level of agreement between the two analysts by computing the Cohen's kappa [75] for the records they both successfully labeled. We find a Cohen's kappa score of 0.9820, suggesting almost perfect agreement, thus giving us confidence that the *PR* dataset has a very high level of agreement among analysts. Only 17 records had different labels. This labeling disagreement was resolved by keeping the labels of the most senior analyst among the two (i.e. *JA1*). Thus, the maximum potential error rate in this dataset, assuming that *JA1* is wrong in all of the 17 assessments, is 0.8%.

*2) (Evaluation) Senior Expert Analyst Dataset (**EA**):* Despite the **PR** dataset incorporating the public reports' APT infrastructure labels and the high confidence agreement between the two analysts in manually labeling, sampling bias could still be apparent [11]. To better understand the potential sampling bias of the *PR* dataset that will be used for training, we ask an expert analyst with over 20 years of experience, from a separate organization of *JA1* and *JA2*, to manually label a second completely disjoint ground truth from that of *PR*. This set consists of all the RRs found in Active DNS for one random domain name per APT actor, totaling 2,293 RRs. *SA* was able to confidently label 831 from the 2,293 RRs and marked 155 RRs as actor-utilized and 683 RRs as non-actor-utilized. The dataset *SA* labeled is not as balanced as *PR*, since *SA* was given all the historical RRs for each domain name and not a balanced set of RRs, in contrast to *JA1*. We do utilize this dataset — since the *PR* dataset is balanced — to evaluate Atropos in a scenario without base rate fallacy [11]. Overall, this dataset consists of 831 RRs from 191 domain names of 191 different APT actors.

*3) (Evaluation) Future Records Dataset (**FR**):* The second test set is created after the system is completed with the intent to evaluate its performance against future distributions of RRs that were not seen during training. To do that, we pick a random sample of 100 RRs from reports spanning from 2023-05-03 to 2025-01-29, which were published after all of the reports from our training dataset. These 100 RRs correspond to 65 domains, 98 IPs, and 33 APT actors. Given the same instructions and data that were outlined in Section IV-A, analysts *JA1* and *JA2* label these 100 RRs and resolve their disagreements to arrive at a single dataset. The two analysts achieved a very high Cohen's Kappa agreement level of 0.86, suggesting an almost perfect agreement, with only six records in disagreement. Post-labeling, the two analysts discussed their labels and resolved all six disagreements. The class distribution of this set is 73 non-actor-utilized and 27 actor-utilized RRs.

TABLE V: Average 10-fold cross-validation performance of Atropos on the PR dataset. Atropos achieves at best a 99.86 ROC AUC score when utilizing VirusTotal DNS data and training on the *PR* dataset utilizing a Random Forest Model.

DNS Dataset	**ML Model**	**Average 10-fold X Validation Scores**				
		ROC AUC	**F1-Macro**	**Accuracy**	**Precision**	**Recall**
Active DNS	Random Forest	**99.82**%	98.14%	98.16%	98.03%	98.60%
Active DNS	Decision Tree	97.67%	97.71%	97.72%	97.74%	98.11%
Active DNS	XGBOOST	99.52%	98.36%	98.37%	98.07%	99.00%
Active DNS	SVM	97.86%	88.20%	88.70%	82.88%	100.0%
Active DNS	MLP	96.83%	94.33%	94.37%	96.50%	93.05%
VirusTotal	Random Forest	**99.86**%	98.86%	98.90%	98.37%	99.77%
VirusTotal	Decision Tree	98.14%	98.39%	98.44%	97.62%	99.77%
VirusTotal	XGBOOST	99.86%	98.66%	98.70%	98.35%	99.44%
VirusTotal	SVM	97.40%	81.37%	83.44%	78.38%	99.33%
VirusTotal	MLP	97.00%	95.54%	95.56%	96.30%	96.31%

B. Experimental Results

*1) **Classification Results**:* Table V shows the average 10-fold cross-validation performance of Atropos on the *PR* training dataset. Atropos achieves significant ROC AUC scores across all utilized machine learning models and the two DNS datasets. The best-performing model in terms of ROC AUC score is the Random Forest with a score of 99.82% and 99.86% on Active DNS and VirusTotal datasets, respectively. This showcases that Atropos has high performance across models and can have high levels of transferability across different DNS datasets. Since Random Forest has the highest performing scores, we pick this model as best for our next test, out-of-distribution experiments.

Our second experiment evaluates Atropos against two test sets (**EA** and **FR**) that consist of records that were not considered during training with the intent to test Atropos performance against out-of-distribution(OOD) datasets and observe its generalization and robustness against sampling bias that has been identified as a major problem in the security field [11]. Table VII demonstrates Atropos' performance against these two test sets and across the two DNS datasets.

TABLE VI: Number of network IoCs for the top 10 actors that are reported in threat reports and identified by Atropos. Atropos provides three times the IP visibility of threat reports and contextualizes three times more domain names than threat reports.

Actor	IP addresses		BGP prefixes		ASN		Domain Coverage (%)	
	Reports	Atropos	Reports	Atropos	Reports	Atropos	Reports	Atropos
Lazarus Group	1,047	776	569	504	371	241	20.25%	76.12%
Gamaredon	361	1,873	130	623	25	253	20.90%	45.19%
Fin7	218	341	132	222	56	126	23.48%	60.75%
Unc1878	208	379	73	134	48	47	63.45%	96.49%
APT28	204	723	155	381	97	202	29.63%	71.89%
Muddywater	173	301	92	206	36	98	06.38%	43.20%
Winnti Group	158	206	85	126	47	74	23.64%	21.28%
APT29	157	144	123	130	93	91	28.35%	65.67%
Sandworm	132	53	93	20	70	13	21.42%	71.42%
CharmingKitten	128	554	62	188	62	81	46.84%	63.19%
Total	**7,553**	**25,049**	**3,530**	**6,115**	**1,291**	**1,762**	**20.20%**	**61.07%**

We observe that in all of the tests, Atropos remains highly accurate with accuracy equal to and higher than 91.00%. We also notice that the precision of Atropos drops especially in the **EA** dataset. We investigate these records and find out that the largest class of false positives comes from Cloudflare and Namecheap web-hosting IPs (35.71%), while the rest are distributed among different ASes. After debriefing the **SA** analyst, they mentioned that they do not consider any cloud-fronting and shared-hosting IP addresses (e.g., Cloudflare, Namecheap shared-hosting) as likely actor-utilized, as they do not provide any basis for pivoting to other infrastructure or evidence that the actors owned the IPs since they can belong to multiple users. Despite that **EA** analyst is correct and these IPs are not useful for pivoting and should not be considered for blacklisting, this comes in contrast with our instructions in which we outlined we wanted to identify the IP corresponding to the domain is the infrastructure utilized in the APT operation, regardless of whether they are cloud-fronting or virtual hosting. Despite that, the overall performance of Atropos across all tests remains very high, and this experiment showcased that its results are generalizable in (OOD) datasets. In the appendix, we demonstrate how Atropos can be adjusted to generalize in similar scenarios of labeling as those considered by the **EA** analyst.

TABLE VII: Out-of-distribution test set evaluation of Atropos. Atropos achieves an over 91.00% accuracy across the two evaluation datasets, demonstrating generalization.

DNS Dataset	Test Set	ROC AUC	F1-Macro	Accuracy	Precision	Recall
Active DNS	FR	95.47%	95.08%	95.38%	92.00%	95.38%
VirusTotal	FR	95.47%	95.08%	95.38%	92.00%	95.38%
Active DNS	EA	87.13%	85.56%	91.00%	73.23%	91.00%
VirusTotal	EA	88.47%	87.20%	91.39%	76.53%	91.39%

*2) **Feature Importance**:* By calculating the Mean Decrease of Impurity (MDI) score on an 80-20% split utilizing the *PR* dataset and Active DNS data, we rank the features that Atropos has used to find out their importance, thus offering model interpretability. We observe that the top five features by MDI include the number of Historic Domains on IP (0.187), the IP first seen Delta (0.177), the number of Communicating Files on IP (0.158), and the Mean and Median Concurrent Domains on IP (0.125 and 0.149), thus highlighting importance across all feature types but specifically in infrastructure and temporal features. This fact aligns with our observations from the SolarWinds case study in Section II-B, where both the temporal (i.e., when an IP was pointed to the domain name compared to detection) and infrastructural (i.e., what kind of infrastructure that IP is) features are necessary to distinguish the actor-utilized from the non-actor-utilized infrastructure. Considering this, we are confident that Atropos makes decisions that follow the principles that a human analyst would also have used. We further list full details around the individual feature importance in Appendix B.

C. Infrastructure Expansion and Lifetime Characterization

Table VI presents the number of IPs, BGP Prefixes, Autonomous System Numbers (ASNs), and domain coverage comparison between what is provided in threat reports and what is identified by Atropos. The table presents the coverage of the top 10 APT actors along with the total number of all the actors. Overall, Atropos provides 3.062 times more high-confidence IPs than OSINT APT reports. The added benefit for BGP prefixes and autonomous systems (ASes) is smaller as they represent bigger groupings of Internet infrastructure, but is still significant. Furthermore, Atropos provides actor-utilized IP mappings for 61.07% of domains provided in APT reports, which is significantly larger than that of just matching the IPs that exist or reports with their domain names. To characterize the lifetime of IP infrastructure, we utilize the historical information provided by the Active DNS dataset [42], enabling us to build lifetimes with a daily granularity for all of the high-confidence IPs Atropos has identified as associated with the APT domains in our dataset.

Takeaways: Our measurement methodology accurately identifies **3.06** times more historically utilized IP infrastructure, compared to that published in threat reports, and characterizes **61.07%** of the APT domains appearing in them, thus enabling a more comprehensive measurement study than just utilizing OSINT threat report data.

V. Infrastructure Analysis

In this section, we conduct the largest APT and most comprehensive APT infrastructure analysis to date. To do so, we

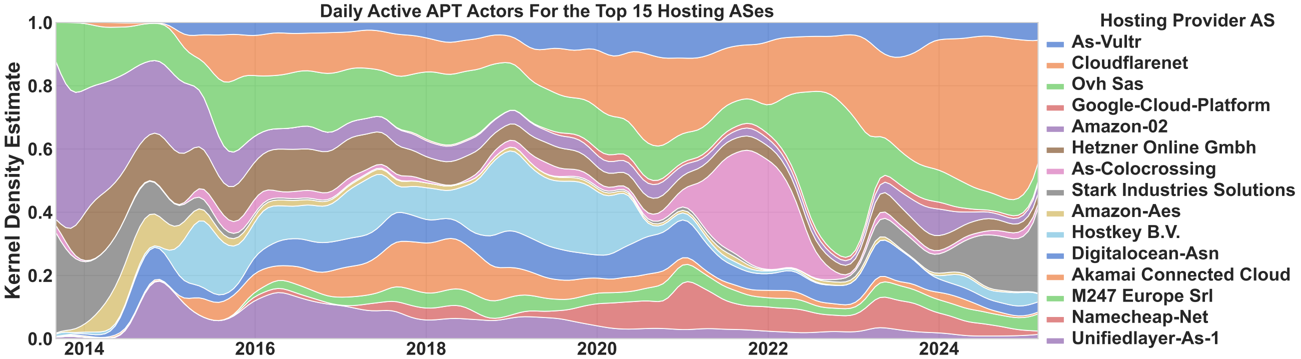

Fig. 4: Kernel density estimate distribution of the daily active APT actors for the top 15 most utilized provider ASes in the last decade. Cloudflare utilization for domain name hosting has increased significantly over the years, making forensic analysis and attribution of IP infrastructure harder due to the overlap of benign and malicious infrastructure on the same IP addresses.

utilize all of the APT IoCs from our OSINT data sources described in Section III-A, as well as the new infrastructure Atropos has identified, utilizing both the ActiveDNS and VirusTotal DNS datasets. For more conservative estimations, we remove RRs on which ActiveDNS and VirusTotal models disagree. The number of these records is only 1.34% of the overall records, and thus it does not bias our measurement results. We structure our analysis around the following research questions:

- Where do APT actors provision their infrastructure, and do they re-use the same hosting providers over the years? (Section V-A)
- What is the lifecycle of the different infrastructure types associated with APT domains, and how does that affect forensic analysis? (Section V-B1)
- How long before the public reporting of an attack are actor-utilized IPs provisioned to the domains, and what is the time window of their observability? (Section V-B2)

A. *Infrastructure Utilization*

*1) **Hosting Provider Utilization**:* Figure 4 demonstrates the density of the daily active APT actors that utilize any of the top 15 hosting providers in our dataset. We observe that these hosting provider ASes that APT actors utilize consist of a mix of cloud-fronting, CDN, and proxying providers (e.g., Cloudflare, Akamai, AWS, Google Cloud), virtual and shared hosting providers (e.g., Vultr, DigitalOcean, OVH, Namecheap, UnifiedLayer), dedicated hosting (i.e., Hetzner, OVH, Hostkey), and providers that are more tolerant to abuse (i.e., Colocrossing, Stark Industries, M247). Thus, APT actors utilize a plethora of different types of hosting providers for their domain name hosting and do not primarily choose a specific category of providers. Temporally, we observe that after 2023, CloudFlare has drastically increased in popularity among actors, with 74 different actors hosting at least one domain in their network. This increased popularity of Cloudflare over the years is justified as this provider offers very lucrative technologies that enhance the stealthiness of APT infrastructure, such as origin IP masking and blending with benign domain traffic behind the same virtual hosting IPs. This trend complicates network threat hunting and forensics as it diminishes the value of IP addresses for such domains, a fact that has been anecdotally verified by APT experts [70]. Another recent rising trend is the increased utilization of the bulletproof hosting provider ”Stark Industries Solution” after 2023. Stark Industries Solution is a new bulletproof hosting provider that was launched in February of 2022 [60] (although its IP space was used in previous attacks under different management). While we observe 23 different groups to have utilized this hosting provider since 2023, the two groups with the highest number of domains are Fin7 and MuddyWater [64], [37]. Lastly, despite the aforementioned rising trends, several hosting providers (e.g., Hetzner, VULTR, M247) have featured a steady utilization by APT groups across the years.

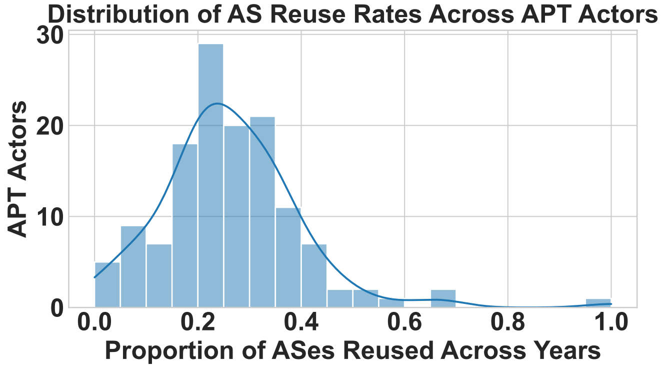

Fig. 5: AS re-utilization among APT actors. The average re-use rate of ASes among actors equals 26.20%.

*2) **Infrastructure Reuse**:* In order to measure the re-use of ASes among different APTs, we identify for each domain name and actor the first time that domain was provisioned to each AS. Then we measure the proportion of ASes that get re-used for more than one year per APT actor, and for statistical relevance, we remove actors with less than 20 domain names, thus focusing this experiment on 135 actors. Figure 5 demonstrates the proportion of ASes that these actors re-use across the years. We notice that most APT actors re-use a small portion of all the ASes they have provisioned their

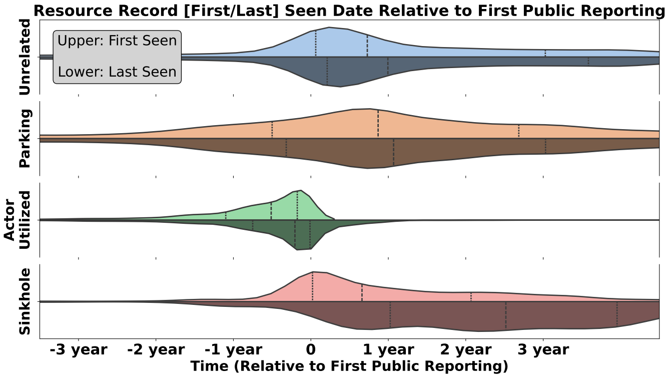

Fig. 6: The unique lifecycle of the infrastructure associated with APT domains compared to the first public report date.

domain names historically, with the average re-use rate for all these actors being 26.20%. This means that most actors do not choose to host their domain names on the exact same set of ASes over the years; however, they do re-use a smaller portion of the same hosting ASes. When we look at the percentage of the APT groups that do re-use at least one AS for over one year, we see that it is 97.03%. Thus, APT actors do re-use network infrastructure in the same hosting providers; however, this re-use only accounts for a small portion of all the infrastructure they have used historically. Threat hunters and attribution experts need to be careful when identifying and attributing new campaigns to existing actors, mainly by network infrastructure signals, and focus only on the infrastructure that is consistently being re-used when doing so.

Takeaways: APT actors provision their infrastructure to a plethora of different types of hosting providers, ranging from cloud-fronting, virtual hosting, dedicated hosting, and bulletproof hosting providers, with cloud-fronting rising significantly in popularity. While the majority of APT actors re-use infrastructure, this only occurs for a small portion of their overall infrastructure, with the average AS re-use rate equaling 26.20%.

B. Infrastructure Lifecycle

Utilizing the labeled RRs that Atropos provides alongside their first and last seen days in the daily DNS records of the Active DNS dataset, we measure the lifecycle of all the different categories of infrastructure related to APT domains and derive key takeaways for forensic analysts and defenders. Subsection V-B1 characterizes and compares the unique lifecycles of all the categories of infrastructure, while subsection V-B2 focuses specifically on actor-utilized infrastructure and its period of observability.

*1) **Infrastructure Type Analysis**:* As we demonstrated in Figure 2, four categories of infrastructure are frequently associated with APT domain names, namely, actor-utilized, parking, sinhkole, and unrelated to the attack. Figure 6 shows the lifecycle of all these types of infrastructure relative to the first public reporting of each of the domains they point at. Actor-utilized infrastructure features a narrower lifecycle than the rest of the categories, with the median IP first and last observed in DNS data 251 and 103 days before detection, respectively. More interestingly, **73.6%** of the actor-utilized IPs no longer point to their domains at the first public disclosure date. This finding has practical applications for analysts and systems detecting and investigating APT infrastructure during and after the disclosure of an attack, considering the lack of comprehensive coverage that threat reports provide. Analysts and systems that do not utilize historical DNS datasets and do not consider the lifetime of their IP infrastructure risk, at best, to incomprehensively discover attacker-utilized infrastructure or, at worst, to misclassify parking, sinkhole, or other unrelated infrastructure that appears after detection as attacker-utilized.

Sinkhole infrastructure features a more long-lived lifecycle than that of actor-utilized IPs, starting very close to the disclosure date and spanning long after detection. However, 17.5% of the sinkhole IPs are pointed to their domain before the day of their public disclosure, and thus, analysts and detection systems utilizing DNS records even at the time before disclosure have to be very careful not to associate sinkholes with actor-utilized IPs.

Parking infrastructure is more uniformly distributed across the lifecycle of APT domains, and the median IP is first pointed 297 days after public disclosure. However, similarly to the sinkhole infrastructure, 35.9% of known parking IPs appear to be first pointed to the domains before public disclosure of their domain, and as we showcased in the SolarWinds case study, such IPs have to be identified and filtered out by forensic analysts looking to uncover actor-utilized infrastructure. Surprisingly, **31%** and **42%** of the parking and sinkhole IPs, respectively, have one or more malicious detections on VirusTotal, and **18%** of the sinkhole IPs have five or more. This can be explained due to the large amount of APT and other malicious domains that end up being pointed at them, which makes some vendors flag them as malicious by association. Nevertheless, this fact highlights that researchers have to be careful and not blindly trust vendors' detections but consider the type of infrastructure when doing forensic analysis or building intrusion detection or investigation systems, especially considering that five or fewer VirusTotal malicious detections are common in malicious IP labeling [10], [30], [31], [81], [69].

The rest of the IP infrastructure that neither Atropos labels as actor-utilized nor is in known parking and sinkhole lists is grouped in the ”Unrelated” category. Such infrastructure is primarily first pointed to by their domains after public disclosure for 75.9% of the IPs, and is mainly associated with infrastructure unrelated to the actors, such as future owners, parking, and sinkhole IPs unknown to the public. The top two IPs of this class are: ”35[.]205[.]61[.]67”, an unknown to our sinkhole list sinkhole [72], and ”54[.]65[.]172[.]3”, an Amazon shared hosting IP that had 995,067 domains historically pointed at it. Almost a quarter of this infrastructure (i.e., 24.1%) is first pointed before public disclosure of their domains, and thus forensic analysts have to be careful of not mis-associating such infrastructure with likely actor-utilized IPs.

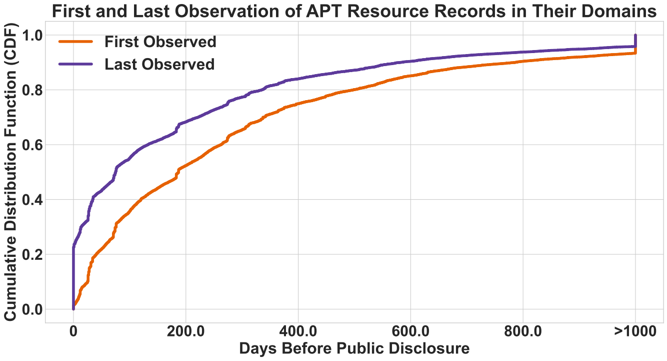

Fig. 7: Number of days that actor-utilized IPs were first and last observed before their domain name public disclosure.

Takeaways: The IP infrastructure of APT domains features unique lifecycles. Actor-utilized IPs are mostly visible before public disclosure, with **73.6%** of the IPs no longer pointing to their domains after their disclosure, highlighting the importance of historical data for comprehensive infrastructure tracking. Parking IPs are more uniformly distributed across time, while sinkholes and other infrastructure unrelated to the actors primarily first appear after disclosure. A significant portion of known parking and sinkhole IPs are mistakenly being flagged by security vendors as malicious, and thus forensic analysts have to be very careful not to mistake them with infrastructure that was likely utilized by the APT actors.

2) ***Actor-Utilized IP Activity:*** Figure 7 demonstrates the first and last seen of the APT utilized resource records among all actors as observed in ActiveDNS compared to their first public disclosure. We observe that there is a wide variation among initial provisioning delta compared to the first public threat reporting. The mean and median first IP provisioning are 317 and 187 days before the first public disclosure, respectively, which indicates that many actor-utilized IPs remain well under the radar for months. This fact reinforces the common knowledge that APT attacks are stealthy, and it takes a significant amount of time to detect them in contrast to other cyber attacks like phishing or password stealers, which feature significantly shorter detection lifecycles of 21 hours and 11 days, respectively [59], [12]. It is important to note that this is a higher bound estimate as it includes the time for a report to be written and published; however, the difference is still significant relative to commodity threats, and the reliance of expert APT analysts on public reports has been recently verified [70]. The long delta between first infrastructure provisioning and public reporting of the attacks can also be explained by the fact that advanced actors have been reported to strategically age their domain names [36], [56]. Looking at the last time the actor-utilized IPs were resolved by their domain names relative to public disclosure, we observe a mean and median time of 173 and 75 days before disclosure, respectively. More importantly, to observe 90% of the actor-utilized resource records, an analyst would need to go back to at least 19 months in time before the first public disclosure of their domains. This fact reiterates the need for historical data to comprehensively track APT infrastructure, and calls for great caution among forensic analysts in order not to mistake parking, sinkhole, and previous owners' infrastructure, which has a significant presence before the public disclosure, as actor-utilized.

These observations can aid network detection systems that are heavily dependent on features related to the short lifespan of malicious domain names, which have been proven not to be adversarially robust [28]. Furthermore, they have practical implications for organizations and government entities that need to forensically investigate APT attacks against them. Our results demonstrate that 90% of the actor-utilized resource records would be observable by keeping logs in a time window between 19 and 25 months before their public reporting. Thus, organizations that are sensitive to APT threats and forensic analysts will need to keep at least 19 months' worth of historical network records to comprehensively evaluate whether they have been a target of a prior APT threat and to thoroughly investigate the network infrastructure of APT actors, respectively.

Takeaways: APT actors first provision infrastructure on their domain names **317** days on average before the APT attack is publicly reported. This number alone provides ample time for actors to successfully conduct their operations while negatively impacting detection systems that assign a positive reputation to longer-lived domains. Organizations need to keep their network logs for a time window of at least **19** to **25** months to be able to identify 90% of the APT infrastructure from a DNS perspective.

VI. Discussion and Limitations

A. *Operationally Relevant Takeaways*

Our study revealed that **73.6%** of the actor utilized IPs no longer pointed to their domains at the time of their first disclosure. This has important implications for properly training APT detection systems that utilize network infrastructure features, as researchers would need to identify the actor-utilized IPs of APT domains by looking back in time so that they do not mistakenly associate them with other types of infrastructure, such as parking, sinkholes, and future owners that frequently appears close or after public disclosure as illustrated in Figure 6. Future works can utilize tools and methodologies like Atropos to do so.

The recent increase in the utilization of cloud-fronting services, such as CloudFlare, among APT actors makes network forensics and attribution harder since such infrastructure is frequently associated with multiple other domain names and owners at the same time. DNS-based detection and attribution systems should be adjusted to work well beyond the infrastructure-level features of prominent works [8], [13] and emphasize lexical, registration, and temporal characteristics of domain names to extract signal that could characterize the APT actors since their infrastructure blends in with normal, benign traffic. Furthermore, researchers can use historical DNS records and methodologies like Atropos and look for the period before the actors enabled cloud-fronting services to find out whether the actors had pointed their domains to other likely utilized infrastructure. As we demonstrate in the Appendix C, in such scenarios, Atropos can be tuned to ignore cloud-fronting and shared hosting infrastructure and focus on dedicated infrastructure that the actors are more likely to own.

Our findings in Section V-B2 revealed that APT actors first provision their IP infrastructure to their domains 317 days on average before disclosure, reaching 25 months for the 90th percentile. Defenders will need to extend their data retention policy for DNS queries, firewall logs, network flows, and endpoint telemetry to at least 25 months. With such policies, retrospective scans of their network data, the moment new IoCs appear in public reports and threat feeds, will yield more comprehensive coverage in terms of identifying likely infected victims in their networks.

B. Limitations

Despite the increased infrastructure visibility that our measurement methodology provides compared to APT threat reports, it cannot identify all actor-utilized IPs for all domains, as illustrated in Table VI. Some of the APT domain names belong to ccTLDs and other TLDs that do not share their zone files, so it is difficult for DNS scanners to pick them up before their detection. APT actors may also set their name servers to respond with a valid command and control IP only to specific target networks (i.e., victims) and with invalid IPs to others, including projects like Active DNS. Additionally, some APT actors may utilize a subdomain that hasn't been observed by a DNS scanner (e.g., 3LD or 4LD) for their command and control server and park their e2LD to known parking locations, which Atropos will filter out. Despite all this, the infrastructure expansion compared to public reports for our measurement study is still significant.

As illustrated in Section IV-B, Atropos performs very well in both evaluation datasets; however, its performance can vary by actor depending on how differently actors utilize the network infrastructure. APT actors can perform mimicry attacks or utilize fast flux [58] to induce false positives and perform *label shift* [11], [46]. Future work can build dedicated models for individual groups and their strategies to address such issues. We did not explore this avenue as the existing high-confidence ground truth for these threats is insufficient to effectively represent each APT actor in a machine-learning model without big class imbalances [11].

VII. Related Work

The lifecycle of domain names has been the subject of prior works in the security and measurement communities. Lever et al.[45] offered an alternative to WHOIS and tried to identify domain ownership changes using Alembic, a lightweight algorithm that utilized passive DNS data. However, their methodology was aimed at identifying changes of ownership and not actor-utilized infrastructure of malicious domains. Affinito et al. [1] studied the lifecycle of domains and malicious domains in blocklists utilizing zone file data, and similarly to Lever, developed a methodology to bound the lifecycles of domain names, but not to label their infrastructure. Lloyd et. al. [48] developed a methodology to classify domain names as "active", "no-IP", or "inactive", with an aim to find domain names serving content under the registrant's control. However, this methodology is not applicable to historic domain names and mainly relies on parking infrastructure lists that we have demonstrated are not sufficient for our scope. Sebastian et al. developed an automated approach to attribute domain names to their most likely ownership [71]. However, our work is different as our goal is to identify the lifecycle of the infrastructure of domain names.

In network-based detection systems, network traffic data and domain lifecycle analysis have also been used as the means of APT detection. Alageel et. al., [3] proposed Hawk-Eye, an APT command and control domain detection system that utilizes PCAP data. Oprea et al [61] propose a framework for early-stage APT detection, by modeling the network communications of the internal hosts of an enterprise with outside hosts and utilizing belief propagation. Lamprakis et al. [43] suggest a system capable of detecting APT commands and controlling traffic in an unsupervised fashion utilizing host weblogs. Chiba et al., proposed a detection system that is based on domain name lifecycle analysis [19]. Other studies suggested techniques for the detection of lateral movement that are applicable to APTs [16], [33], [41], [40], while a large amount of work has focused on provenance detection and investigation systems [53], [32], [6], [39], [67], [35], [47]. Such studies are orthogonal to our scope as they are aimed at the detection of APT domains rather than the investigation of their network infrastructure over the years.

Several measurement studies have analysed APT actors and sophisticated attacks over the years. Marczak et al. [51] were among the first to empirically measure and characterize the modus operandi of nation-state actors. Le Blond et al. [44] characterized targeted APT attacks against NGO members finding the actors to utilize recently disclosed vulnerabilities in their malware. Urban et. al [80] analyzed 93 APT reports and found that 80% of the APT actors start their attacks by sending phishing emails. Saha et al. conducted a user study utilizing 15 APT expert practitioners and identified that current tools and practices in APT analysis feature significant challenges for threat hunting and attribution [70]. In our study, we reinforce the findings of these prior works regarding the difficulty of the analysis and the sophistication of APT threats.

VIII. Conclusion

In this work, we analyzed the network infrastructure of 405 APT actors spanning over a decade. Using our novel measurement methodology, we were able to identify 3.06 times more historically utilized IP infrastructure compared to that published in threat reports. Our lifecycle analysis suggests that organizations will need to retain network logs for at least 25 months in order to maintain comprehensive historical visibility for forensic purposes in the case of an APT attack. Furthermore, we observed that while APT actors utilize a plethora of different hosting providers, they only re-use a small portion of them, and that over the years, the use of cloud-fronting has increased significantly, making network forensics and attribution harder. Our findings verify prior insights from experts, and we hope to be the basis for increased attention from the community.

IX. Acknowledgements

We would like to thank the anonymous reviewers for their insights and suggestions in improving this work. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001123C0035. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Defense (DoD).

References

[1] Antonia Affinito, Raffaele Sommese, Gautam Akiwate, Stefan Savage, Geoffrey Voelker, Alessio Botta, Mattijs Jonker, et al. Domain name lifetimes: baseline and threats. Network Traffic Measurement and Analysis Conference (TMA), 2022.

[2] Atif Ahmad, Jeb Webb, Kevin C Desouza, and James Boorman. Strategically-motivated advanced persistent threat: Definition, process, tactics and a disinformation model of counterattack. *Computers & Security*, 86:402–418, 2019.

[3] Almuthanna Alageel and Sergio Maffeis. Hawk-eye: holistic detection of apt command and control domains. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, pages 1664–1673, 2021.

[4] OTX AlienVault. The World's First Truly Open Threat Intelligence Community, 2024.

[5] Eihal Alowaisheq. Cracking wall of confinement: Understanding and analyzing malicious domain takedowns. In *The Network and Distributed System Security Symposium (NDSS)*, 2019.

[6] Abdulellah Alsaheel, Yuhong Nan, Shiqing Ma, Le Yu, Gregory Walkup, Z Berkay Celik, Xiangyu Zhang, and Dongyan Xu. {ATLAS}: A sequence-based learning approach for attack investigation. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*, 2021.

[7] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Luca Invernizzi, Michalis Kallitsis, et al. Understanding the mirai botnet. In *26th USENIX security symposium (USENIX Security 17)*, pages 1093–1110, 2017.

[8] Manos Antonakakis, Roberto Perdisci, David Dagon, Wenke Lee, and Nick Feamster. Building a dynamic reputation system for {DNS}. In *19th USENIX Security Symposium (USENIX Security 10)*, 2010.

[9] Afsah Anwar, Yi Hui Chen, Roy Hodgman, Tom Sellers, Engin Kirda, and Alina Oprea. A recent year on the internet: Measuring and understanding the threats to everyday internet devices. In *Proceedings of the 38th Annual Computer Security Applications Conference*, pages 251–266, 2022.

[10] Ignacio Arnaldo, Alfredo Cuesta-Infante, Ankit Arun, Mei Lam, Costas Bassias, and Kalyan Veeramachaneni. Learning representations for log data in cybersecurity. In *Cyber Security Cryptography and Machine Learning: First International Conference, CSCML 2017, Beer-Sheva, Israel, June 29-30, 2017, Proceedings 1*, pages 250–268. Springer, 2017.

[11] Daniel Arp, Erwin Quiring, Feargus Pendlebury, Alexander Warnecke, Fabio Pierazzi, Christian Wressnegger, Lorenzo Cavallaro, and Konrad Rieck. Dos and don'ts of machine learning in computer security. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 3971–3988, 2022.

[12] Athanasios Avgetidis, Omar Alrawi, Kevin Valakuzhy, Charles Lever, Paul Burbage, Angelos D Keromytis, Fabian Monrose, and Manos Antonakakis. Beyond the gates: An empirical analysis of {HTTP-Managed} password stealers and operators. In *32nd USENIX security symposium (USENIX Security 23)*, pages 5307–5324, 2023.

[13] Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. Exposure: Finding malicious domains using passive dns analysis. In *Ndss*, pages 1–17, 2011.

[14] Beth Binde, Russ McRee, and Terrence J O'Connor. Assessing outbound traffic to uncover advanced persistent threat. *SANS Institute. Whitepaper*, 16, 2011.

[15] Xander Bouwman, Victor Le Pochat, Pawel Foremski, Tom Van Goethem, Carlos H Gañán, Giovane CM Moura, Samaneh Tajalizadehkhoob, Wouter Joosen, and Michel Van Eeten. Helping hands: Measuring the impact of a large threat intelligence sharing community. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 1149–1165, 2022.

[16] Benjamin Bowman, Craig Laprade, Yuede Ji, and H Howie Huang. Detecting lateral movement in enterprise computer networks with unsupervised graph {AI}. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses ({RAID} 2020)*, pages 257–268, 2020.

[17] Juan Caballero, Gibran Gomez, Srdjan Matic, Gustavo Sánchez, Silvia Sebastián, and Arturo Villacañas. Goodfatr: A platform for automated threat report collection and ioc extraction. *arXiv preprint arXiv:2208.00042*, 2022.

[18] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[19] Daiki Chiba, Hiroki Nakano, and Takashi Koide. Domaindynamics: Advancing lifecycle-based risk assessment of domain names. *Computers & Security*, 153:104366, 2025.

[20] CyberMonitor. Apt and cybercriminal campaign collection. https://github.com/CyberMonitor/APT_CyberCriminal_Campagin_Collections, 2023.

[21] Team Cymru. Unravelling the Mystery of Bogons: A senior stakeholder and IT professional guide. https://www.team-cymru.com/post/unravelling-the-mystery-of-bogons-a-senior-stakeholder-and-it-professional-guide.

[22] Giorgio Di Tizio, Michele Armellini, and Fabio Massacci. Software updates strategies: A quantitative evaluation against advanced persistent threats. *IEEE Transactions on Software Engineering*, 49(3):1359–1373, 2022.

[23] Domaintools. Continuous Eruption: Further Analysis of the SolarWinds Supply Chain Incident. https://www.domaintools.com/resources/blog/continuous-eruption-further-analysis-of-the-solarwinds-supply-incident/, 2020.

[24] Domaintools. Unraveling Network Infrastructure Linked to the SolarWinds Hack. https://www.domaintools.com/resources/blog/unraveling-network-infrastructure-linked-to-the-solarwinds-hack/, 2020.

[25] Domaintools. Domaintools whois history. https://research.domaintools.com/research/whois-history/, 2023.

[26] Aaron Faulkenberry, Athanasios Avgetidis, Zane Ma, Omar Alrawi, Charles Lever, Panagiotis Kintis, Fabian Monrose, Angelos D Keromytis, and Manos Antonakakis. View from above: Exploring the malware ecosystem from the upper dns hierarchy. In *Proceedings of the 38th Annual Computer Security Applications Conference*, pages 240–250, 2022.

[27] Fortinet. Solar winds cyber attack. https://www.fortinet.com/resources/cyberglossary/solarwinds-cyber-attack, 2024.

[28] Tillson Galloway, Kleanthis Karakolios, Zane Ma, Roberto Perdisco, Angelos Keromytis, and Manos Antonakakis. Practical attacks against dns reputation systems. In *2024 IEEE Symposium on Security and Privacy (SP)*, 2024.

[29] Thomas Geras and Thomas Schreck. The" big beast to tackle": Practices in quality assurance for cyber threat intelligence. In *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses*, pages 337–352, 2024.

[30] Joobin Gharibshah, Tai Ching Li, Andre Castro, Konstantinos Pelechrinis, Evangelos E Papalexakis, and Michalis Faloutsos. Mining actionable information from security forums: the case of malicious ip addresses. *From Security to Community Detection in Social Networking Platforms*, pages 193–211, 2019.

[31] Joobin Gharibshah, Tai Ching Li, Maria Solanas Vanrell, Andre Castro, Konstantinos Pelechrinis, Evangelos E Papalexakis, and Michalis Faloutsos. Inferip: Extracting actionable information from security discussion forums. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 301–304, 2017.

[32] Xueyuan Han, Thomas Pasquier, Adam Bates, James Mickens, and Margo Seltzer. Unicorn: Runtime provenance-based detector for advanced persistent threats. *arXiv preprint arXiv:2001.01525*, 2020.

[33] Grant Ho, Mayank Dhiman, Devdatta Akhawe, Vern Paxson, Stefan Savage, Geoffrey M Voelker, and David Wagner. Hopper: Modeling and detecting lateral movement. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 3093–3110, 2021.

[34] Thorsten Holz, Christian Gorecki, Konrad Rieck, and Felix C Freiling. Measuring and detecting fast-flux service networks. In *Ndss*, 2008.

[35] Md Nahid Hossain, Sadegh M Milajerdi, Junao Wang, Birhanu Eshete, Rigel Gjomemo, R Sekar, Scott Stoller, and VN Venkatakrishnan. {SLEUTH}: Real-time attack scenario reconstruction from {COTS} audit data. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*, pages 487–504, 2017.

[36] Infoblox. Dog Hunt: Finding Decoy Dog Toolkit via Anomalous DNS Traffic. https://blogs.infoblox.com/threat-intelligence/cyber-threat-advisory/dog-hunt-finding-decoy-dog-toolkit-via-anomalous-dns-traffic/.

[37] Deep Instinct. MuddyC2Go – Latest C2 Framework Used by Iranian APT MuddyWater Spotted in Israel. https://www.deepinstinct.com/blog/muddyc2go-latest-c2-framework-used-by-iranian-apt-muddywater-spotted-in-israel.

[38] IPinfo. 'IPinfo. https://ipinfo.io/, 2024.

[39] Hassaan Irshad, Gabriela Ciocarlie, Ashish Gehani, Vinod Yegneswaran, Kyu Hyung Lee, Jignesh Patel, Somesh Jha, Yonghwi Kwon, Dongyan Xu, and Xiangyu Zhang. Trace: Enterprise-wide provenance tracking for real-time apt detection. *IEEE Transactions on Information Forensics and Security*, 16:4363–4376, 2021.

[40] Joseph Khoury, Dorde Klisura, Hadi Zanddizari, GDLT Parra, Peyman Najafirad, and Elias Bou-Harb. Jbeil: Temporal graph-based inductive learning to infer lateral movement in evolving enterprise networks. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 9–9. IEEE Computer Society, 2023.

[41] Isaiah J King and H Howie Huang. Euler: Detecting network lateral movement via scalable temporal link prediction. *ACM Transactions on Privacy and Security*, 26(3):1–36, 2023.

[42] Athanasios Kountouras, Panagiotis Kintis, Chaz Lever, Yizheng Chen, Yacin Nadji, David Dagon, Manos Antonakakis, and Rodney Joffe. Enabling network security through active dns datasets. In *International Symposium on Research in Attacks, Intrusions, and Defenses*, pages 188–208. Springer, 2016.

[43] Pavlos Lamprakis, Ruggiero Dargenio, David Gugelmann, Vincent Lenders, Markus Happe, and Laurent Vanbever. Unsupervised detection of apt c&c channels using web request graphs. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 366–387. Springer, 2017.

[44] Stevens Le Blond, Adina Uritesc, Cédric Gilbert, Zheng Leong Chua, Prateek Saxena, and Engin Kirda. A look at targeted attacks through the lense of an {NGO}. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 543–558, 2014.

[45] Chaz Lever, Robert Walls, Yacin Nadji, David Dagon, Patrick McDaniel, and Manos Antonakakis. Domain-z: 28 registrations later measuring the exploitation of residual trust in domains. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 691–706. IEEE, 2016.

[46] Frankie Li, Anthony Lai, and Ddl Ddl. Evidence of advanced persistent threat: A case study of malware for political espionage. In *2011 6th International Conference on Malicious and Unwanted Software*, pages 102–109. IEEE, 2011.

[47] Fucheng Liu, Yu Wen, Dongxue Zhang, Xihe Jiang, Xinyu Xing, and Dan Meng. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 1777–1794, 2019.

[48] Siôn Lloyd, Carlos Hernandez-Gañan, and Samaneh Tajalizadehkhoob. Towards more rigorous domain-based metrics: quantifying the prevalence and implications of "active" domains. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 539–545. IEEE, 2023.

[49] Mandiant. Highly Evasive Attacker Leverages SolarWinds Supply Chain to Compromise Multiple Global Victims With SUNBURST Backdoor. https://www.mandiant.com/resources/blog/evasive-attacker-leverages-solarwinds-supply-chain-compromises-with-sunburst-backdoor/, 2020.

[50] Alessandro Mantovani, Simone Aonzo, Xabier Ugarte-Pedrero, Alessio Merlo, and Davide Balzarotti. Prevalence and impact of low-entropy packing schemes in the malware ecosystem. In *NDSS*, 2020.

[51] William R Marczak, John Scott-Railton, Morgan Marquis-Boire, and Vern Paxson. When governments hack opponents: A look at actors and technology. In *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, pages 511–525, 2014.

[52] Xianghang Mi, Xuan Feng, Xiaojing Liao, Baojun Liu, XiaoFeng Wang, Feng Qian, Zhou Li, Sumayah Alrwais, Limin Sun, and Ying Liu. Resident evil: Understanding residential ip proxy as a dark service. In *2019 IEEE symposium on security and privacy (SP)*, pages 1185–1201. IEEE, 2019.

[53] Sadegh M Milajerdi, Rigel Gjomemo, Birhanu Eshete, Ramachandran Sekar, and VN Venkatakrishnan. Holmes: real-time apt detection through correlation of suspicious information flows. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1137–1152. IEEE, 2019.

[54] MISP. MISP Galaxy Threat Actors. https://raw.githubusercontent.com/MISP/misp-galaxy/main/clusters/threat-actor.json, 2024.

[55] MITRE. MITRE Groups. https://attack.mitre.org/groups/.

[56] Palo Alto Networks. Strategically Aged Domain Detection: Capture APT Attacks With DNS Traffic Trends. https://unit42.paloaltonetworks.com/strategically-aged-domain-detection/.

[57] Palo Alto Networks. Russia's Gamaredon aka Primitive Bear APT Group Actively Targeting Ukraine. https://unit42.paloaltonetworks.com/gamaredon-primitive-bear-ukraine-update-2021/, 2022.

[58] Palo Alto Networks. 'Russia's Trident Ursa (aka Gamaredon APT) Cyber Conflict Operations Unwavering Since Invasion of Ukraine'. https://unit42.paloaltonetworks.com/trident-ursa/, 2022.

[59] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 2020.

[60] Krebs on Security. Stark Industries Solutions: An Iron Hammer in the Cloud. https://krebsonsecurity.com/2024/05/stark-industries-solutions-an-iron-hammer-in-the-cloud/.

[61] Alina Oprea, Zhou Li, Ting-Fang Yen, Sang H Chin, and Sumayah Alrwais. Detection of early-stage enterprise infection by mining large-scale log data. In *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 45–56. IEEE, 2015.

[62] Pingdom. 'The top 100 web hosting countries'. https://www.pingdom.com/blog/web-hosting-countries-2013/.

[63] Daniel Plohmann, Martin Clauss, Steffen Enders, and Elmar Padilla. Malpedia: a collaborative effort to inventorize the malware landscape. *The Journal on Cybercrime and Digital Investigations*, 3(1):1–19, 2017.

[64] Silent Push. FIN7: Silent Push unearths the largest group of FIN7 domains ever discovered. 4000+ IOFA domains and IPs found. Louvre, Meta, and Reuters targeted in massive global phishing and malware campaigns. https://www.silentpush.com/blog/fin7/.

[65] Md Rayhanur Rahman, Setu Kumar Basak, Rezvan Mahdavi Hezaveh, and Laurie Williams. Attackers reveal their arsenal: An investigation of adversarial techniques in cti reports. *arXiv preprint arXiv:2401.01865*, 2024.

[66] Nanda Rani, Bikash Saha, and Sandeep Kumar Shukla. A comprehensive survey of advanced persistent threat attribution: Taxonomy, methods, challenges and open research problems. *arXiv preprint arXiv:2409.11415*, 2024.

[67] Mati Ur Rehman, Hadi Ahmadi, and Wajih Ul Hassan. Flash: A comprehensive approach to intrusion detection via provenance graph representation learning. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 139–139. IEEE Computer Society, 2024.

[68] Reuters. Crypto's biggest hacks and heists after $1.5 billion theft from Bybit. https://www.reuters.com/technology/cybersecurity/fbi-says-north-korea-was-responsible-15-billion-bybit-hack-2025-02-27/.

[69] Candong Rong, Gaopeng Gou, Mingxin Cui, Gang Xiong, Zhen Li, and Li Guo. Malfinder: An ensemble learning-based framework for malicious traffic detection. In *2020 IEEE Symposium on Computers and Communications (ISCC)*, pages 7–7. IEEE, 2020.

[70] Aakanksha Saha, James Mattei, Jorge Blasco, Lorenzo Cavallaro, Daniel Votipka, and Martina Lindorfer. Expert insights into advanced persistent threats: Analysis, attribution, and challenges. In *Proceedings of the 34th USENIX Security Symposium (USENIX Sec)*, 2025.

[71] Silvia Sebastián, Raluca-Georgia Diugan, Juan Caballero, Iskander Sanchez-Rola, and Leyla Bilge. Domain and website attribution beyond whois. In *Proceedings of the 39th Annual Computer Security Applications Conference*, pages 124–137, 2023.

[72] Snapmaker. Snapmaker controller reaching out to Anubis networks sink hole. https://forum.snapmaker.com/t/snapmaker-controller-reaching-out-to-anubis-networks-sink-hole/31250.

[73] Miroslav Stampar and M Kasimov. maltrail: Malicious traffic detection system, 2019.

[74] Blake E Strom, Andy Applebaum, Doug P Miller, Kathryn C Nickels, Adam G Pennington, and Cody B Thomas. Mitre att&ck: Design and philosophy. In *Technical report*. The MITRE Corporation, 2018.

[75] Shuyan Sun. Meta-analysis of cohen's kappa. *Health Services and Outcomes Research Methodology*, 11:145–163, 2011.

[76] Ke Tian, Steve TK Jan, Hang Hu, Danfeng Yao, and Gang Wang. Needle in a haystack: Tracking down elite phishing domains in the wild. In *Proceedings of the Internet Measurement Conference 2018*, pages 429–442, 2018.

[77] Virus Total. VirusTotal Historical Whois API. https://docs.virustotal.com/reference/domain-resolutions.

[78] Virus Total. VirusTotal. https://www.virustotal.com/en, 2024.

[79] Wiem Tounsi and Helmi Rais. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & security*, 72:212–233, 2018.

[80] Tobias Urban, Matteo Große-Kampmann, Dennis Tatang, Thorsten Holz, and Norbert Pohlmann. Plenty of phish in the sea: Analyzing potential pre-attack surfaces. In *Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS*

2020, Guildford, UK, September 14–18, 2020, Proceedings, Part II 25, pages 272–291. Springer, 2020.

[81] Liang Wang, Antonio Nappa, Juan Caballero, Thomas Ristenpart, and Aditya Akella. Whowas: A platform for measuring web deployments on iaas clouds. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 101–114, 2014.

[82] Johannes Zirngibl, Steffen Deusch, Patrick Sattler, Juliane Aulbach, Georg Carle, and Mattijs Jonker. Domain parking: Largely present, rarely considered! In *Proc. Network Traffic Measurement and Analysis Conference (TMA) 2022*, 2022.

[83] zonefiles.io. Compromised domain list. https://zonefiles.io/compromised-domain-list/, 2024.

Appendix

A. Detailed Atropos Features

Temporal Class (3 features):

- (f_1) **Domain Detection and IP First Seen Date Delta:** The time delta (in days) between the first day the domain name was reported in a threat report and the first day that domain first pointed to the IP. This feature aims to identify the IPs close to the detection of the domain that are more likely to be associated with the actor and remove older or newer IPs that are likely associated with previous or future owners of the domain name.
- (f_2) **Domain Detection and IP Last Seen Date Delta:** The time delta (in days) between the first day the domain name was reported in a threat report and the last day the domain first pointed to the IP. Since the disclosure of the APT domains to the public does not always happen right after their detection or sinkholing, this feature is meant to identify sinkhole and parking infrastructure that an APT domain has been pointed to before its detection and persisted months or even years after its public disclosure.
- (f_3) **IP Lifetime:** The number of days that the domain pointed to the IP address. This feature can help differentiate between short-lived placeholder and testing IPs and longer-lived APT-controlled IPs and parking. For example, in Figure 2, the domains pointed to placeholder parking IPs [82] for a median of 41 days compared to 314 and 366 days for the actor-controlled IPs.

Infrastructure Class (4 features):

- (f_4) **Number of Historical Domains Pointed to the IP:** The total number of historical domains ever pointed to the given IP according to the DNS data source. Similar to our example, this feature is meant to find parking and sinkhole IPs.
- (f_5, f_6) **Mean/Median of Concurrent Domains Pointed to the IP:** The mean and median number of other domains pointed to the IP during the period that the given domain is pointed to the IP. Since IP addresses are volatile over time, these features are meant to capture the infrastructure behavior of a given IP only at the time when the domain was pointed to it.
- (f_7) **Number of Historical Files communicating with the IP:** The total number of historical files that have been communicating with the given IP according to VirusTotal. In our example (Section 2), sinkhole IPs have a median number of 83,128 communicating files on VirusTotal compared to a median of zero for the APT-controlled IP and parking infrastructure. This usually happens because malware dynamic execution will occur after a domain has been sinkholed and VirusTotal will only see the sinkhole IP.

OSINT Class (11 features):

- **Parking Features**. (f_8) Known Parking IP: Whether the IP appears on known parking lists. (f_9) Known Parking Nameserver IP Overlap: Whether the domain is served by a known parking nameserver at the same time as the domain points to the IP for at least 70% of the time (a percentage we manually pick after multiple tests).
- **Sinkhole Features**. (f_{10}) Known Sinkhole IP: Whether the IP appears on known sinkhole lists. In our example, this time period is illustrated by the red-colored infrastructure (Figure 2). (f_{11}) Known Sinkhole Nameserver IP Overlap: Whether a known sinkhole nameserver is serving a domain at the same time as the domain points to the IP for at least 70% of the time.
- **IP Reputation:** These features (f_{12}: IP Reputation, f_{13}: IP Votes Malicious and f_{14}: IP Votes Harmless) take into account the publicly known reputation of an IP based on the votes from the VirusTotal community [78]. Despite these scores not being perfect, they do help in some instances to identify benign IPs that malware actors can point their domain names to gain residual trust.
- **IP Analyses:** These features (f_{15}: IP Analyses Malicious, f_{16}: IP Analyses Suspicious, f_{17}: IP Analyses Undetected, and f_{18}: IP Analyses Harmless), compute the number of URL scanners in VirusTotal that have flagged an IP with the given label.

Domain Name Class (4 features).

- (f_{19}) **Number of Communicating Files:** The number of files that VirusTotal has found to have communicated with the domain.
- (f_{20}) **Number of Downloaded Files:** The number of files that were available to be downloaded by the given domain name according to Virus Total.
- (f_{21}) **Number of Subdomains:** The number of subdomains that were seen according to VirusTotal under the given domain name.
- (f_{22}) **Number of Certificates:** The number of SSL certificates that have been associated with the domain name at some point in time according to VirusTotal.

B. Detailed Featured Importance

Aside from its strong performance, we chose to utilize a Random Forests model when testing Atropos for its good interpretability compared to other models. Table VIII ranks features used by calculating the Mean Decrease of Impurity (MDI) score on an 80-20% split utilizing the *PR* dataset and Active DNS data. As we can see, the top five features include the Median Concurrent Domains on IP, the number of Historic Domains on IP, the Mean Concurrent Domains on IP, Detection, the IP first seen Delta, and the Detection IP last seen Delta, thus highlighting that infrastructure and temporal features are significantly more important.

The strong performance of the top three features can be attributed to their capability to identify parking and sinkhole infrastructure. This reflects on the motivating example of

TABLE VIII: Atropos MDI Feature Importance when trained on *PR* dataset and utilizing Active DNS data with an 80-20% split.

#f	Feature	MDI	#f	Feature	MDI
f_1	Detection and IP Fseen Delta	0.177	f_{12}	IP Reputation	0.012
f_2	Detection and IP Lseen Delta	0.050	f_{13}	# of Malicious Votes	0.038
f_3	IP Lifetime	0.007	f_{14}	# of Harmless Votes	0.009
f_4	# of Historic Domains on IP	0.187	f_{15}	# of Malicious Analyses	0.004
f_5	Mean Concurrent Domains on IP	0.125	f_{16}	# of Suspicious Analyses	0.005
f_6	Median Concurrent Domains on IP	0.149	f_{17}	# of Undetected Analyses	0.011
f_7	# of IP Communicating Files	0.158	f_{18}	# of Harmless Analyses	0.013
f_8	IP is Known Parking	0.018	f_{19}	# of Domain Communicating Files	0.001
f_9	Nameserver is Known Parking	0.019	f_{20}	# of Files Downloaded From Domain	0.003
f_{10}	IP is Known Sinkhole	0.009	f_{21}	# of Domain Subdomains	0.003
f_{11}	Nameserver is Known Sinkhole	0.000	f_{22}	# of Domain Certificates	0.002

SolarWinds we showcased in Section II-B where the actor-controlled IPs had only the SolarWinds domain names pointed to them while parking and sinkhole IPs had more than nine million and 600 other domains pointed to them respectively. The other benefit of these features is that Atropos does not only rely on parking and sinkhole IP and DNS name server lists, which are usually static and can take months or even years to be updated.

The second strongest set of features is the temporal features. This is not a surprise, because as we saw in *SolarWinds* the APT-controlled IPs pointed to the domains a few months before the detection and continued to be the primary destination of the domains until very close to their detection. Atropos can pick up on this temporal aspect and penalize IPs of previous owners that were first seen on the domains very early and IPs of sinkholes that were first seen after the domain detection similarly to Fig. 2.

C. *Generalization for Threat Hunting*

To see whether Atropos can generalize and adapt to different analyst requirements, such as threat hunting for IPs that are non-cloud-fronting and virtual hosting – similar to the labeling methodology of **EA** described in Section IV-B1 –, we modify the *PR* dataset by flipping all the labels of IP addresses with more than 200 concurrent domain names pointed to them as non-APT controlled to imitate *EA* labeling process, changing 63 resource records from APT-controlled to non-APT controlled. We name this dataset *PR-NVH*. We train our model again utilizing *PR-NVH* and report our results in Table IX. We observe that the accuracy and precision of the new model improve compared to those presented in Table VII, meaning that Atropos can be trained on datasets with different requirements and provide accurate results for different use cases that are outside of the scope of our study.

TABLE IX: Evaluation of Atropos trained with and altered PR dataset.

DNS Dataset	Test Set	ROC AUC	F1-Macro	Accuracy	Precision	Recall
Active DNS	EA	87.86%	89.03%	93.58%	85.45%	93.58%
Virus Total	EA	87.29%	88.06%	92.39%	83.22%	92.39%

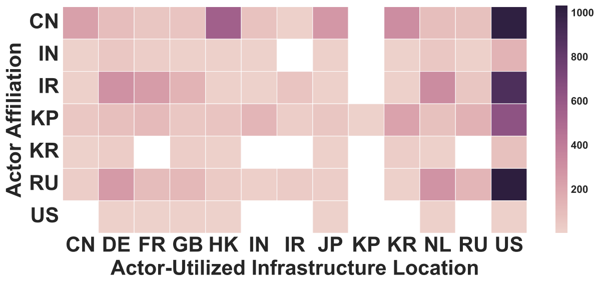

Fig. 8: Number of actor-utilized IPs per country for the top affiliated countries of the APT actors.

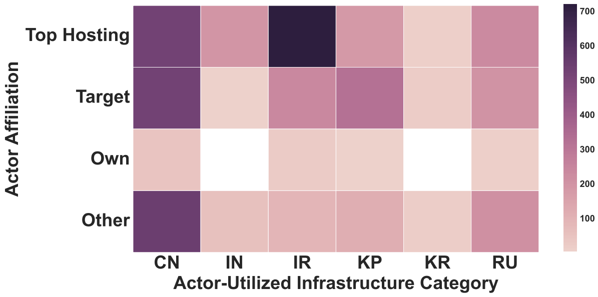

Fig. 9: Number of IPs per category of infrastructure for the top affiliated countries of the APT actors.

D. *Infrastructure Geolocation*

Another important insight that can help us characterize and compare the APT actors is the geolocation of their infrastructure. To that end, we map each actor-utilized IP address to the country where it is most likely located according to IPInfo [38] and then analyze the correlation of the location of the infrastructure with the country affiliation of the actors.

Figure 8, shows a heatmap of the country an actor is affiliated with and the country where the actor-utilized infrastructure is provisioned. In the interest of space, the countries have been limited to the ones with the most publicity and references across our threat reports. We can observe that most of the actor-utilized IPs are provisioned in the USA, with other big hosting provider countries like Germany and the Netherlands to follow. Additionally, we can see that actors from different countries choose to utilize infrastructure with different patterns that in some cases overlap, like the Russian and Iranian APT actors. Their utilization of infrastructure among the US, the Netherlands, Germany, France, and the United Kingdom is more evident and different from that of Chinese actors which, aside from their disproportionate use of US-based infrastructure, also utilize more infrastructure in Hong Kong, Japan, and South Korea.

These findings raise two interesting questions. First, whether the location of the actor-utilized infrastructure correlates with the location of the attack target. Second, whether the location of said infrastructure relates to countries with large hosting providers. To answer these questions, we utilize targeting data from the APT reports and match each domain name and IP with the countries that were identified as targets in the same APT reports. We only use infrastructure for which targeting information is available in this part of our analysis. We also group together countries that are the top 10 largest hosting providers [62] to see if the infrastructure provisioning of the actors is correlated with those aspects.

In Figure 9, we see that APT actors from the top countries mostly provision their infrastructure either in countries that have large hosting providers or in the target countries. Chinese and North Korean actors deploy most of their infrastructure in their target countries while Iranian and Indian groups mostly utilize countries that have large hosting providers. The Chinese actors look to provision infrastructure to countries labeled as ”Other” which is mostly located in Hong Kong and Singapore. Finally, as expected, country-affiliated APT actors rarely provision infrastructure in their own country.