Computers & Security 153 (2025) 104366

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Full length article

DomainDynamics: Advancing lifecycle-based risk assessment of domain names

Daiki Chiba [*], Hiroki Nakano, Takashi Koide

NTT Security Holdings Corporation & NTT Corporation, Tokyo, Japan

ARTICLE INFO

Keywords:
Malicious domain name detection
domain name lifecycle
temporal analysis
risk prediction
Cybersecurity
Machine learning

ABSTRACT

The persistent threat of malicious domains in cybersecurity necessitates robust detection systems. Traditional machine learning approaches often struggle to accurately assess domain name risks due to their static analysis methods and lack of consideration for temporal changes in domain attributes. To address these limitations, we developed DomainDynamics, a novel system that evaluates domain name risks by analyzing their lifecycle phases. This study provides a comprehensive evaluation and refinement of the DomainDynamics framework. The system creates temporal profiles for domains and assesses their attributes at various stages, enabling informed, time-sensitive risk assessments. Our initial evaluation, involving over 85,000 malicious domains, achieved an 82.58% detection rate with a low 0.41% false positive rate. We expanded our research to include benchmarking against commercial services, feature significance analysis using interpretable AI techniques, and detailed case studies. This investigation not only validates the effectiveness of DomainDynamics but also reveals temporal indicators of malicious intent. Our findings demonstrate the advantages of lifecycle-based analysis over static methodologies, providing valuable insights for practical cybersecurity applications.

1. Introduction

The persistent threat of cyber-attacks using malicious domain names requires robust detection mechanisms. Identifying inherently malicious domains is crucial for effective cyber defense strategies. Numerous studies have explored the application of machine learning (ML) to detect such domains (see Section 2.1), often by analyzing extensive datasets to discern patterns indicative of malicious intent in new domain names.

While these ML-based approaches show promise, they have limitations. Their static analysis methods often fail to capture the dynamic changes in domain attributes over time, leading to false positives when previously malicious domains become benign or false negatives when legitimate domains are compromised. These challenges stem from the inability of traditional assessment methods to incorporate the temporal context and lifecycle changes of domain behavior.

To address these limitations, we introduced DomainDynamics in our previous work (Chiba et al., 2025), an advanced system capable of assessing domain name risks at any point in their lifecycle. DomainDynamics constructs a timeline for each domain name, extracts characteristics at various points in time, and employs ML to determine risk based on these temporal features. This approach significantly enhances the precision and timeliness of malicious domain detection.

To clarify, the primary goal of DomainDynamics is to assist security administrators in proactively identifying domain names that are likely to be used for malicious activities in the near future. By analyzing the lifecycle changes of domain names, our system provides time-sensitive risk assessments that enable administrators to update security policies, such as blocklists, before an attack occurs. This proactive approach enhances the effectiveness of cyber defense strategies by allowing for timely interventions.

Our initial research on DomainDynamics showed promising results, accurately predicting domain names that would be used in attacks within 7 days with a detection rate of 82.58% and a low false positive rate of 0.41%. This represented a significant improvement over previous studies. Building on these encouraging findings, we have conducted extensive additional research and experimentation to further validate and enhance the system's capabilities.

In this expanded study, we have broadened our analysis of related work, examining 51 key studies in the field of malicious domain detection, up from the original 12 studies. This comprehensive review has allowed us to position DomainDynamics more clearly within the existing research and highlight its unique contributions.

We have also conducted more rigorous evaluations of DomainDynamics' performance across various scenarios and datasets. Our expanded experiments include testing with different training datasets and

[*] Corresponding author.
E-mail address: daiki.chiba@ieee.org (D. Chiba).

https://doi.org/10.1016/j.cose.2025.104366
Received 20 November 2024; Received in revised form 26 December 2024; Accepted 6 February 2025
Available online 20 February 2025

periods to optimize system performance. Additionally, we have compared DomainDynamics with VirusTotal, a leading commercial security service, to benchmark its capabilities against industry standards.

To provide deeper insights into the system's decision-making process, we have implemented advanced explainable AI techniques, specifically SHAP (SHapley Additive exPlanations). This has allowed us to analyze the relative importance of different features in our model, revealing the critical role of lifecycle-based features in risk assessment.

Furthermore, we have presented comprehensive case studies that demonstrate DomainDynamics' effectiveness in detecting true positives and true negatives, as well as analyzing false positives. These real-world examples provide valuable insights into the system's practical applications and limitations.

The structure of this paper is as follows: Section 2 provides a comprehensive review of prior research on malicious domain name detection, highlighting the challenges this research aims to address with motivating examples. Section 3 details the design and implementation of DomainDynamics. Section 4 describes our expanded evaluation experiments, including performance comparisons, feature analysis, and case studies. Section 5 discusses the ethical considerations, limitations, and future directions of this research. Finally, Section 6 summarizes our findings and their implications for cybersecurity practices.

2. Background and related work

This section reviews past research on malicious domain names and explains the challenges this study aims to address, using illustrative examples.

2.1. Detecting malicious domains

Table 1 summarizes 51 key studies on malicious domain detection (2010—2023), detailing venue, year, targets, inputs, ML features, and outputs. Based on the table, we will describe each study in order of their target, specifically what they aim to detect within malicious domain names.

Malware Domains. Malware domains, crucial for command-and-control (C2) operations and malware dissemination, represent the field's earliest research focus. Early studies leveraged ISP-level DNS data to evaluate them, including dynamic DNS reputation systems like Notos (Antonakakis et al., 2010). Kopis (Antonakakis et al., 2011) and FluxBuster (Perdisci et al., 2012) monitor DNS traffic to identify malware domains through DNS query patterns and network-wide DNS traffic, respectively. Segugio (Rahbarinia et al., 2015) uses these data to construct bipartite graphs for identifying unknown malware domains. Some approaches also utilize corporate network logs for detection. Manadhata et al. proposed a system to detect malware domains using graph inference by constructing a host-domain graph from a company's proxy logs (Manadhata et al., 2014). Oprea et al. developed a framework based on belief propagation inspired by graph theory, focusing on early detection of malware domains at the enterprise level (Oprea et al., 2015). MADE, a security analysis system using ML, addresses the problem of prioritizing malware domains in enterprise networks (Oprea et al., 2018). HinDom detects malware domains based on a heterogeneous information network and inference classification methods (Sun et al., 2019).

Algorithmically Generated Domains. Algorithmically generated domains (AGDs) are domain names created by domain generation algorithms (DGA) within malware domains (Almashhadani et al., 2020; Liang et al., 2022). Yadav et al. proposed a method to detect AGDs by analyzing the distribution of characters and bigrams in domain names (Yadav et al., 2010). Pleiades detects the emergence of DGA-based malware by collecting non-existent domain (NXDomain) responses from DNS traffic and using a combination of clustering and classification (Antonakakis et al., 2012). Krishnan et al. developed a method to identify randomly generated domains from NXDomain responses using statistical methods and detect clients infected with DGA-based malware (Krishnan et al., 2013). Phoenix identifies AGDs using features based on characters and IP addresses (Schiavoni et al., 2014).

Phishing Domains. Phishing domain studies, leveraging Certificate Transparency (CT) logs, target sites in preparation (Drichel et al., 2021), phishing kit characteristics (Bijmans et al., 2021), and content-independent identification (Sabah et al., 2022; Bozkir et al., 2023). Research using visual characteristics of phishing sites includes VisualPhishNet (Abdelnabi et al., 2020), which uses triplet convolutional neural networks; Phishpedia (Lin et al., 2021), which accurately recognizes logos on screenshots; and PhishIntention (Liu et al., 2022), which visually extracts the intent to capture credentials on phishing sites. Other methods include monitoring domain names containing phishing-related strings (Marchal et al., 2012), identifying phishing sites from squatting domain candidates (Tian et al., 2018), and automated collection of social engineering attacks including phishing (Koide et al., 2020). Kim et al. proposed a network-based inference method to detect disguised phishing URLs (Kim et al., 2022). SpamHunter (Tang et al., 2022) and CrowdCanary (Nakano et al., 2023) extract new phishing sites using security-related posts on social media. Bitaab et al. reported on the construction of a dataset and classifier for large-scale detection of fraudulent e-commerce sites (Bitaab et al., 2023).

Squatting Domains. Squatting domains mimic legitimate ones, with studies focusing on detection and characteristics (Koide et al., 2023). Research on typosquatting (e.g., `eaxmple[.]test`) addresses domains created by typographical errors (Szurdi et al., 2014; Agten et al., 2015). Studies on combosquatting (e.g., `example-login[.]test`) examine deception based on visual appearance by adding keywords to legitimate brand names (Kintis et al., 2017; Roberts et al., 2019). Recent research has focused on detection methods and survey results for homograph IDN (e.g., `exаmple[.]com`) using internationalized domain names (IDN) (Liu et al., 2018; Suzuki et al., 2019; Chiba et al., 2019; Quinkert et al., 2019).

Compromised Domains. Compromised domains, initially legitimate, are later misused by attackers. DomainChroma first highlighted the need to distinguish whether a domain name is compromised as a separate axis from being malicious, producing practical blocklists from malicious domains (Chiba et al., 2018). This concept has been further developed in COMAR (Maroofi et al., 2020) and as a separate ML system (Silva et al., 2021). Research also addresses shadowed domains, which are specifically created subdomains of compromised domains (Liu et al., 2017).

Malicious Domains. Malicious domain research broadly targets various domains, not limited to specific categories (Zhao et al., 2019). EXPOSURE (Bilge et al., 2011) and DomainProfiler (Chiba et al., 2016) are representative studies that perform domain reputation analysis by detecting general malicious domains using DNS data and ML. WarningBird (Lee and Kim, 2012) detects various malicious domains/URLs appearing in social media posts. PREDATOR (Hao et al., 2016) and Premadoma (Spooren et al., 2019) focus on measures against malicious domains at domain registrars, aiming to detect them at the time of registration based on available features. Several approaches passively monitor DNS traffic and use graphs to detect malicious domains (Khalil et al., 2016; Nabeel et al., 2020). MalRank detects malicious domains as a graph inference problem based on logs obtained from security information and event management (SIEM) (Najafi et al., 2019).

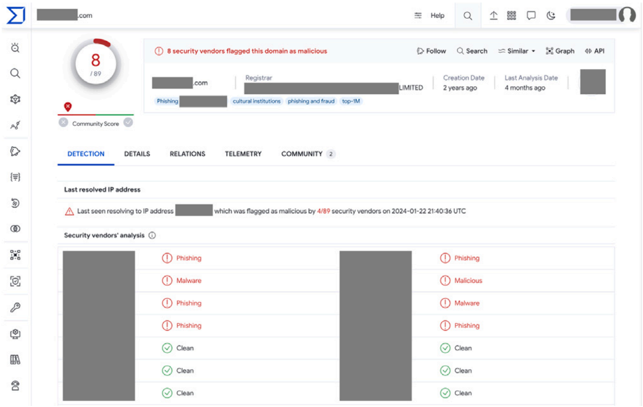
2.2. Comparison with related work

DomainDynamics offers a novel approach to malicious domain detection by focusing on the lifecycle changes of domain names. Unlike traditional methods that rely on static analysis of domain features at a single point in time, our system constructs a temporal profile for each domain, enabling a dynamic risk assessment that adapts to changes over time. This lifecycle-based approach addresses a critical gap in

Table 1

Comparison of detection methods in major studies on malicious domain name detection from 2010 to 2023, highlighting the unique capability of DomainDynamics to indicate temporal risk changes.

				Input			Features					Output	
Literature	Venue	Year	Target	Domain	Path	Content	Lexical	Context	Resource	User	Change	Point	Change
Notos Antonakakis et al. (2010)	Security	2010	Malware Domains	✓			✓		✓			✓	
Yadav et al. (2010)	IMC	2010	Algorithmically Generated Domains	✓			✓		✓			✓	
EXPOSURE Bilge et al. (2011)	NDSS	2011	Malicious Domains	✓			✓		✓	✓		✓	
Kopis Antonakakis et al. (2011)	Security	2011	Malware Domains	✓					✓	✓		✓	
FluxBuster Perdisci et al. (2012)	TDSC	2012	Malware Domains	✓					✓	✓		✓	
Pleiades Antonakakis et al. (2012)	Security	2012	Algorithmically Generated Domains	✓			✓		✓	✓		✓	
Marchal et al. (2012)	RAID	2012	Phishing Domains	✓			✓					✓	
WarningBird Lee and Kim (2012)	NDSS	2012	Malicious Domains	✓	✓			✓	✓			✓	
Krishnan et al. (2013)	DSN	2013	Algorithmically Generated Domains	✓						✓		✓	
Phoenix Schiavoni et al. (2014)	DIMVA	2014	Algorithmically Generated Domains	✓			✓		✓			✓	
Manadhata et al. (2014)	ESORICS	2014	Malware Domains	✓						✓		✓	
Szurdi et al. (2014)	Security	2014	Typosquatting Domains	✓		✓	✓	✓	✓			✓	
Oprea et al. (2015)	DSN	2015	Malware Domains	✓					✓	✓		✓	
Segugio Rahbarinia et al. (2015)	DSN	2015	Malware Domains	✓					✓	✓		✓	
Agten et al. (2015)	NDSS	2015	Typosquatting Domains	✓		✓	✓	✓	✓			✓	
DomainProfiler Chiba et al. (2016)	DSN	2016	Malicious Domains	✓			✓		✓		✓	✓	
Khalil et al. (2016)	ASIACCS	2016	Malicious Domains	✓					✓			✓	
PREDATOR Hao et al. (2016)	CCS	2016	Malicious Domains	✓			✓		✓		✓	✓	
Woodpecker Liu et al. (2017)	CCS	2017	Shadowed Domains	✓			✓		✓			✓	
Kintis et al. (2017)	CCS	2017	Combosquatting Domains	✓			✓					✓	
DomainChroma Chiba et al. (2018)	COSE	2018	Compromised Domains	✓			✓		✓			✓	
FANCI Schüppen et al. (2018)	Security	2018	Algorithmically Generated Domains	✓			✓					✓	
Liu et al. (2018)	DSN	2018	Homograph IDNs	✓			✓					✓	
Tian et al. (2018)	IMC	2018	Phishing Domains	✓		✓	✓	✓				✓	
MADE Oprea et al. (2018)	ACSAC	2018	Malware Domains	✓	✓		✓		✓	✓		✓	
HinDom Sun et al. (2019)	RAID	2019	Malware Domains	✓			✓		✓	✓		✓	
ShamFinder Suzuki et al. (2019)	IMC	2019	Homograph IDNs	✓			✓					✓	
DomainScouter Chiba et al. (2019)	RAID	2019	Homograph & Combosquatting IDNs	✓					✓			✓	
Premadoma Spooren et al. (2019)	ACSAC	2019	Malicious Domains	✓			✓		✓			✓	
MalRank Najafi et al. (2019)	ACSAC	2019	Malicious Domains	✓					✓	✓		✓	
Roberts et al. (2019)	CCS	2019	Target Embedding Domains	✓			✓		✓			✓	
Quinkert et al. (2019)	CNS	2019	Homograph IDNs	✓		✓	✓	✓	✓			✓	
Zhao et al. (2019)	Access	2019	Malicious Domains	✓			✓					✓	
COMAR Maroofi et al. (2020)	EuroS&P	2020	Compromised Domains	✓		✓	✓	✓	✓	✓		✓	
StraySheep Koide et al. (2020)	ASIACCS	2020	Social Engineering URLs	✓	✓	✓	✓	✓				✓	
VisualPhishNet Abdelnabi et al. (2020)	CCS	2020	Phishing Websites			✓		✓				✓	
Nabeel et al. (2020)	TOPS	2020	Malicious Domains	✓			✓		✓	✓		✓	
MaldomDetector Almashhadani et al. (2020)	COSE	2020	Algorithmically Generated Domains	✓			✓					✓	
Drichel et al. (2021)	ARES	2021	Phishing Domains	✓			✓		✓			✓	
Bijmans et al. (2021)	Security	2021	Phishing Domains	✓			✓	✓	✓			✓	
Phishpedia Lin et al. (2021)	Security	2021	Phishing Websites	✓	✓	✓		✓				✓	
Silva et al. (2021)	Security	2021	Compromised Domains	✓	✓		✓		✓			✓	
Sabah et al. (2022)	RAID	2022	Phishing Domains	✓			✓		✓		✓	✓	
PhishIntention Liu et al. (2022)	Security	2022	Phishing Websites	✓	✓	✓		✓				✓	
Kim et al. (2022)	CCS	2022	Phishing URLs	✓	✓		✓		✓			✓	
SpamHunter Tang et al. (2022)	CCS	2022	Phishing URLs			✓		✓		✓		✓	
HAGDetector Liang et al. (2022)	COSE	2022	Algorithmically Generated Domains	✓			✓					✓	
CrowdCanary Nakano et al. (2023)	ARES	2023	Phishing URLs	✓	✓	✓	✓	✓		✓		✓	
GramBeddings Bozkir et al. (2023)	COSE	2023	Phishing URLs	✓	✓		✓					✓	
PhishReplicant Koide et al. (2023)	ACSAC	2023	Generated Squatting Domains	✓			✓					✓	
Beyond Phish Bitaab et al. (2023)	S&P	2023	Fraudulent e-Commerce Websites	✓	✓	✓	✓	✓	✓	✓		✓	
DomainDynamics	-	-	**Malware & Phishing Domains**	✓					✓		✓	✓	✓



[redacted].com [redacted]

8 / 89

8 security vendors flagged this domain as malicious

[redacted].com

Registrar [redacted]LIMITED

Creation Date 2 years ago

Last Analysis Date 4 months ago

Phishing [redacted] cultural institutions phishing and fraud top-1M

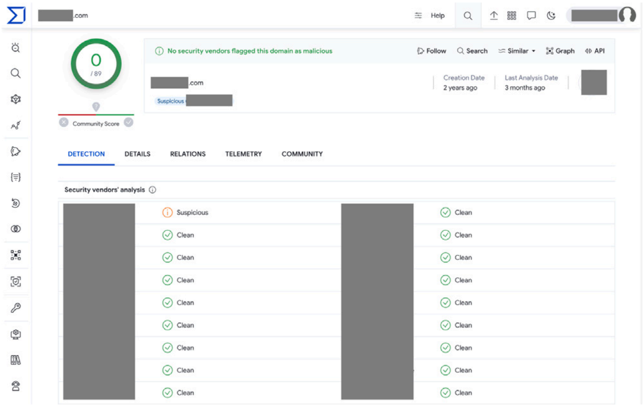
Last resolved IP address

Last seen resolving to IP address [redacted] which was flagged as malicious by 4/89 security vendors on 2024-01-22 21:40:36 UTC

Security vendors' analysis

[redacted]	Phishing	[redacted]	Phishing
	Malware		Malicious
	Phishing		Malware
	Phishing		Phishing
	Clean		Clean
	Clean		Clean
	Clean		Clean

Fig. 1. VirusTotal assessment results for a domain name previously involved in malicious activities, now inactive. Despite its current inactivity, the figure shows the domain still being labeled as Malicious by several AI-powered engines, highlighting the persisting influence of past evaluations and the challenge of accurately reassessing the risk of domain names over time.



[redacted].com [redacted]

0 / 89

No security vendors flagged this domain as malicious

[redacted].com

Creation Date 2 years ago

Last Analysis Date 3 months ago

Suspicious [redacted]

Security vendors' analysis

[redacted]	Suspicious	[redacted]	Clean
	Clean		Clean
	Clean		Clean
	Clean		Clean
	Clean		Clean
	Clean		Clean
	Clean		Clean
	Clean		Clean
	Clean		Clean

Fig. 2. VirusTotal assessment results for a domain name that transitioned from legitimate use to malicious activity after a change in ownership. The figure illustrates that most security engines assess the domain as Clean, despite its current use for malicious purposes. This discrepancy underpins the need for improved methods to detect the dynamic risk profile of domain names as their usage evolves.

existing methods, which often fail to capture the evolving nature of malicious domains.

Table 1 provides a comparative overview of DomainDynamics against 51 key studies in the field. While many approaches utilize historical data, DomainDynamics is unique in its ability to model temporal risk variations for individual domain names. For instance, studies like EXPOSURE (Bilge et al., 2011) and DomainProfiler (Chiba et al., 2016) analyze domain features but do not provide a continuous risk timeline. Our approach allows for the identification of when a domain transitions from benign to malicious, offering a proactive rather than reactive security measure.

Furthermore, DomainDynamics leverages a combination of WHOIS, SOA, and TLS records, providing a comprehensive view of a domain's lifecycle. This contrasts with methods that focus solely on lexical features (e.g., FANCI Schüppen et al., 2018) or content-based analysis (e.g., VisualPhishNet Abdelnabi et al., 2020), which may be susceptible to evasion techniques or limited in their applicability to domains without accessible web content. By incorporating lifecycle data, DomainDynamics offers a more robust and scalable solution for detecting malicious domains.

Direct numerical comparison with related work is challenging due to the lack of publicly available source code and datasets. Many studies in the field do not provide open access to their implementations or the data used for evaluation, making it difficult for third parties to reproduce or directly compare results. In this research, later in Section 4.6, we have re-implemented DomainProfiler (Chiba et al., 2016) based on the description in the original paper and utilized VirusTotal (VirusTotal, 2024a) as baselines for comparison, as they represent the most accessible and representative approaches for evaluating domain name risk.

2.3. Motivating examples

We present two domain names as motivating examples to illustrate the challenges this study seeks to address. The first is a domain name previously involved in malicious activities but has since changed ownership, ceased to be involved in attacks, and had its DNS records deleted, rendering it currently inactive. The second is a domain name that once offered legitimate services but was later acquired through drop catching, changed ownership, and is now being used for malicious purposes. The studies listed in Table 1 lack open access for real-time risk assessment of current domain names. Therefore, we use VirusTotal (VirusTotal, 2024a) to obtain risk assessments for these examples. VirusTotal provides the latest risk assessments for domain names from 89 different security engines and is widely referenced in both industry and academia. Fig. 1 shows the output of VirusTotal for the first domain name, which was previously malicious and then non-malicious. Out of 89 engines, 8 security engines detected the domain name as malicious. Fig. 2 shows the output of VirusTotal for the second domain name, which was previously non-malicious and then malicious. Out of 89 engines, no security engines detected the domain name as malicious. The first domain name, now inaccessible worldwide, should pose no risk, yet VirusTotal results, particularly from AI-powered engines, labeled it as Malicious. This suggests that past evaluations continue to exert a strong influence. The second domain name, actively used for malicious purposes, should be considered risky, but VirusTotal's engines mostly assessed it as Clean. These examples lead us to the following research question:

RQ: How can considering the lifecycle changes of domain names enhance the accuracy of their risk assessments?

As demonstrated by these examples, domain names can vary significantly over time in terms of registrants, DNS settings, and web content. Consequently, a single risk assessment at one point—especially the most recent—may not suffice for an accurate evaluation. Unlike image classification tasks, where a dog remains classified as a dog, or binary malware classification, where malware retains its label, domain names require risk assessments that adapt over time. To draw an analogy, it would be akin to evaluating a completely different restaurant with a new owner based on reviews of a restaurant that has closed. The core issue is that judgments are made without considering the timing of the results—whether they pertain to the era of the previous establishment or that of the new one.

3. Proposed system

To address the challenges identified in the previous section, we propose DomainDynamics, a novel system that uniquely models the lifecycle phases of domain names to reflect risk variations over time. Unlike traditional approaches that may use historical data without accounting for its temporal context, DomainDynamics leverages a lifecycle-based analysis to consider how domain attributes evolve. This allows for more accurate and timely risk assessments by capturing the dynamic nature of domain names.

3.1. Goal and scope

This research primarily aims to develop a system capable of assessing the risk of malicious domain names by considering their temporal context, thus enabling risk determinations at any given moment, from the past to the future. As outlined in Section 2.3, this system is designed to reduce incidents by preventing continuous false positive detections for domain names that were once malicious but are no longer active. In contrast, it can swiftly identify a domain name that has transitioned from being used for legitimate services to being exploited for malicious purposes, thereby preventing false negatives.

For an overview of the scope of this research and its relationship with previous studies, refer to Table 1.

Target. Previous research has generally focused on malicious domains, with some studies specifically targeting malware domains, algorithmically generated domains, phishing domains, typosquatting domains, combosquatting domains, and homograph IDNs, among others. In contrast, this research specifically focuses on malware and phishing domains.

Input. Regarding the input to the system, most past studies have concentrated on domain names, though some have also utilized the full path of URLs or content such as text or screenshots from websites. In this research, the input is limited solely to domain names.

Features. A key distinction of this research from previous work is the combination of features used in ML. Table 1 categorizes features into five groups: Lexical, Context, Resource, User, and Change. Lexical features originate from the character strings of domain names themselves, including n-grams, domain name length, and the entropy of the domain name string. Context features derive from content accessed via the web, such as website text, screenshots, or logo images. Resource features pertain to information related to domain names, such as WHOIS records, IP addresses, DNS records, TLS certificates, and web server data. User features are derived from user interactions and communications, such as the domain names accessed through an organization's proxy server or data from social media posts. Change features relate to long-term changes in domain-related information, such as past registration history (PREDATOR Hao et al., 2016) or the transition of inclusion in Top Lists (DomainProfiler Chiba et al., 2016). This research employs features from both Resource and Change categories to enable risk assessment throughout the entire lifecycle of a domain name. This approach is chosen because reliance on content, such as the web-accessible content of domain names, can hinder risk prediction if the content is cloaked or if it is not feasible to monitor and follow all content changes for each domain name in a scalable manner.

Output. The output of most prior research is limited to a binary classification indicating whether a domain name is malicious at a specific point in time. In contrast, our research presents a significant advancement: the generation of a continuous risk timeline. This timeline not only provides the risk assessment at the evaluation moment but also predicts the risk evolution from the past into the upcoming future for each domain name. Unlike the static point-in-time analysis commonly found in existing literature, our system dynamically models the risk associated with domain names, reflecting changes in their lifecycle. This continuous risk timeline offers a more nuanced understanding of domain risk, accommodating the fluid nature of domain usage and threat levels over time.

Threat Model and Practical Applications. DomainDynamics operates under the threat model where attackers often repurpose domain names over time, making static risk assessments inadequate. Our system is designed for use by security teams to monitor domain names continuously and adjust security measures accordingly. Rather than informing end-users directly about potential future risks, DomainDynamics aids security administrators in making informed decisions about which domains to monitor more closely or preemptively block, thereby enhancing organizational security without disrupting user experience unnecessarily.

Summary. Refer to Table 1 for a comparison of DomainDynamics with prior research. While previous studies have incorporated temporal changes as features (refer to the Change column in Table 1), none have demonstrated the temporal risk variations for individual domain names as achieved by DomainDynamics (refer to the Output column in Table 1). Additionally, DomainDynamics is unique in its focus on malware and phishing domains (see the Target column in Table 1) and is distinguished by its use of a combination of resource and change features, independent of content (see the Features column in Table 1). *In summary, DomainDynamics stands apart from conventional research in terms of both the features it uses and the outputs it generates.*

Listing 1: WHOIS Record Example

```
Domain Name: example.test
Registry Domain ID: 123456789_DOMAIN
Registrar WHOIS Server: whois.example.test
Registrar URL: http://www.example.test
Updated Date: 2024-01-01T00:00:00Z
Creation Date: 2000-01-01T00:00:00Z
Registrar Registration Expiration Date: 2025-01-01T00:
    00:00Z
Registrar: Example Registrar, Inc.
Registrar IANA ID: 55555
Registrar Abuse Contact Email: abuse@example.test
Registrar Abuse Contact Phone: +1.5555555555
Domain Status: clientTransferProhibited
Domain Status: serverTransferProhibited
Registry Registrant ID: REDACTED FOR PRIVACY
Registrant Name: REDACTED FOR PRIVACY
Registrant Street: REDACTED FOR PRIVACY
...
```

3.2. System overview

DomainDynamics, the system proposed in this paper, is designed to achieve the goals and scope outlined above. Fig. 3 presents a diagrammatic overview of the system. DomainDynamics operates in two distinct phases: Training and Predicting.

During the Training phase, domain names are input. The process begins with the construction of a Domain Timeline (❶), followed by the extraction of Features for each point on the Domain Timeline (❷). Using ground truth data, the points in time when the domain was involved in attacks are determined, and Labels are assigned accordingly (❸). Subsequently, an ML Model is trained (❹).

In the Predicting phase, a Target Domain Name is input, and the Domain Timeline and Features are processed similarly to the Training phase (❶, ❷). The ML Model trained earlier is then used to predict the points at which the Target Domain Name might be exploited for attacks (❺).

The subsequent sections detail each step of the process.

3.3. ❶ Building domain timeline

This step is fundamental to both the Training and Predicting phases. A domain name is input, and a historical database pertinent to the domain name is referenced to construct a Domain Timeline. The Domain Timeline represents the lifecycle changes of a domain name, capturing its evolution over time.

To track these lifecycle changes, the study utilizes historical databases of WHOIS records, SOA records, and TLS certificates. For WHOIS records, changes in Registrar, Creation Date, Expiry Date, and Domain Status are monitored. In SOA records, alterations in MNAME, RNAME, and SERIAL are tracked. For TLS certificates, especially pertinent for Web/HTTPS use, changes in Issuer C, Issuer CN, Issuer O, Validity Not Before, Validity Not After, and Subject CN are followed. The following provides detailed definitions and examples for each of these components.

WHOIS Records. WHOIS records contain crucial information about domain registration. As shown in Listing 1, the key elements we focus on are:

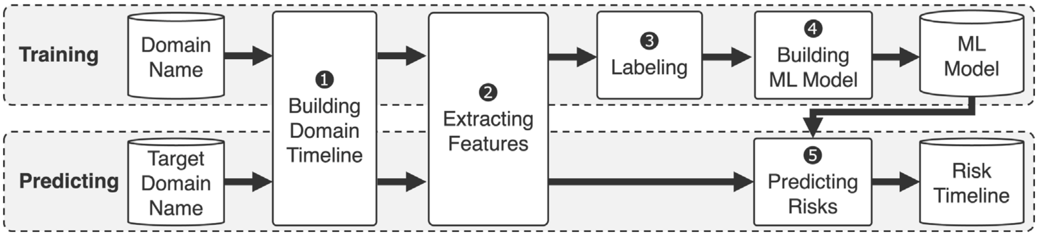


Fig. 3. DomainDynamics system overview, detailing Training phase for model development and Predicting phase for risk assessment of target domain names.

Listing 2: SOA Record Example

```
example.test.  IN  SOA  ns1.example.test.  admin.
     example.test. (
                    2003080800 ; SERIAL
                    28800      ; REFRESH
                    7200       ; RETRY
                    604800     ; EXPIRE
                    86400      ; MINIMUM
                    )
```

Listing 3: TLS Certificate Example

```
Subject: CN = www.example.test, O = Example Inc., L =
     Los Angeles, ST = California, C = US
Issuer: CN = ExampleCert SHA2 Secure Server CA, O =
     ExampleCert Inc, C = US
Validity:
    Not Before: Jan  1 00:00:00 2024 GMT
    Not After : Jan 31 23:59:59 2025 GMT
Public Key:
    Algorithm: RSA
    Modulus: 00:9c:5c:...
    Exponent: 65537 (0x10001)
Signature Algorithm: SHA256-RSA
```

- Registrar: The organization or company that manages the registration of domain names (e.g., Example Registrar, Inc.). It is important to note that the Registrar is distinct from the Registrant, who is the person or organization that actually registers the domain name.
- Creation Date: The date when the domain name was registered (e.g., 2000-01-01T00:00:00Z).
- Expiry Date (or Registrar Registration Expiration Date): The date when the domain name registration expires (e.g., 2025-01-01T00:00:00Z).
- Domain Status: The current status of the domain name (e.g., clientTransferProhibited).

SOA Records. SOA (Start of Authority) records contain essential information about a domain. As illustrated in Listing 2, the key elements we track are:

- MNAME: The name of the primary name server for the domain (e.g., ns1.example.test).
- RNAME: The email address of the administrator responsible for the domain (e.g., admin.example.test).
- SERIAL: The serial number of the SOA record (e.g., 2003080800).

TLS Certificates. TLS certificates are crucial for secure web communications. As shown in Listing 3, the key elements we monitor are:

- Issuer C: The country of the certificate issuer (e.g., US).
- Issuer CN: The common name of the issuer (e.g., ExampleCert SHA2 Secure Server CA).
- Issuer O: The organization of the issuer (e.g., ExampleCert Inc).
- Validity Not Before: The date when the certificate becomes valid (e.g., Jan 1 00:00:00 2024 GMT).
- Validity Not After: The date when the certificate expires (e.g., Jan 31 23:59:59 2025 GMT).
- Subject CN: The common name of the subject (e.g., www.example.test).

By tracking changes in these elements over time, we can construct a comprehensive Domain Timeline that reflects the lifecycle and evolution of a domain name.

To construct the Domain Timeline, we utilized publicly available and commercial services that provide historical data for WHOIS records, SOA records, and TLS certificates. Services such as DomainTools (DomainTools, 2024) and SecurityTrails (SecurityTrails, 2024) offer comprehensive historical WHOIS and DNS data, including SOA records, accessible via their APIs or data export features. For TLS certificates, platforms like crt.sh (Anon, 2024) aggregate Certificate Transparency logs, providing public access to historical TLS certificate data for domain names. These resources are readily available and regularly updated, making it feasible to construct accurate Domain Timelines without the need to build a database from scratch. By leveraging these existing services, our system can efficiently obtain the necessary historical data for a large number of domain names.

A schematic representation of the Domain Timeline for a domain name is depicted in the upper part of Fig. 4. Due to space constraints, not all items are shown in the figure, but examples of actual outputs from DomainDynamics are provided in Fig. A.6 in Appendix.

The domain name featured in Fig. 4 had been used for legitimate purposes for over 10 years but was captured by an attacker, left idle for about a year to avoid easy detection, and then abruptly put to malicious use, based on an actual case. In the WHOIS Domain Timeline in Fig. 4, we can see that the Registrar changed from `Foo Inc` to `Bar Ltd` at a certain point, the Creation Date, which had been maintained at `2009-10-10`, expired and was re-registered on `2021-12-28`, and the Expiry Date, which had been extended annually from `2019-10-10` to `2021-10-10`, then expired. Additionally, the SOA Domain Timeline in Fig. 4 shows that MNAME changed from `ns.foo.test` to `ns.suspended.test`, and later to `ns.bar.test`. Furthermore, the TLS Domain Timeline in Fig. 4 indicates transitions in Issuer O and Not After, revealing that during the period when `Foo Inc` was the Registrar, the domain started using certificates from `BazCert Corp`, but there were periods of certificate expiration leading to gaps, and the certificates expired before the domain itself did. After the Registrar changed to `Bar Ltd`, it took some time before certificates from `QuxCert Inc` began to be activated.

3.4. ❷ Extracting Features

This step is integral to both the Training and Predicting phases, where features of domain names are extracted at each point in time by referring to the Domain Timeline. In essence, features for a single domain name are extracted at multiple temporal points. These features will later serve as attributes to predict the future risk associated with a domain name at a specific point in time. Table 2 enumerates the features extracted from the Domain Timeline. The intuition behind

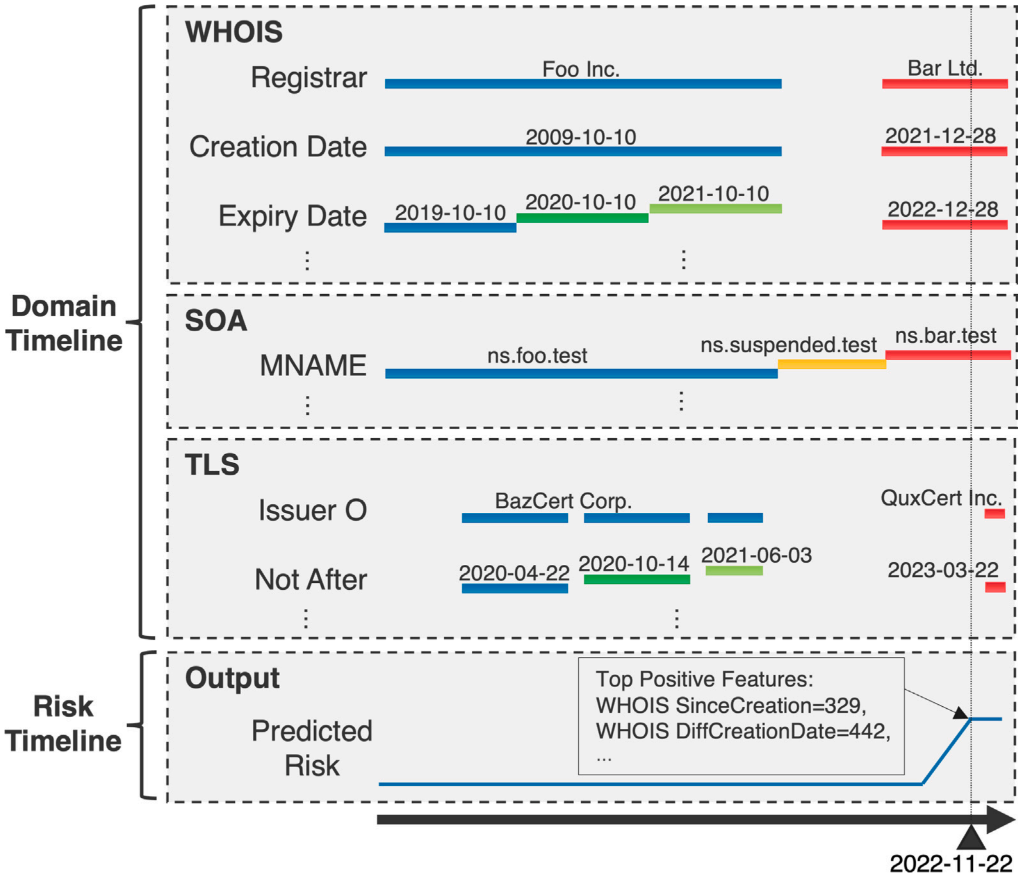


Fig. 4. A depiction of a domain's lifecycle through a Domain Timeline (top) and the corresponding Risk Timeline with predicted probabilities of maliciousness (bottom).

Table 2
Comprehensive list of features extracted from the Domain Timeline with examples of values for a given date.

Category	No.	Feature	Dimension	Example Vectors at 2022-11-22 in Fig. 4
①WHOIS Records	1	Unique WHOIS records count	4	Registrars = 2, CreationDates = 2, ExpiryDates = 4, ...
	2	Days between WHOIS dates	3	SinceCreation = 329, UntilExpiry = 36, ...
	3	Domain status	14	clientTransferProhibited
	4	Days between WHOIS updates	4	Min = 351, Mean = 393, Median = 367, Max = 461
	5	Days difference between WHOIS records	3	DiffCreationDate = 4462, DiffExpiryDate = 444, ...
	6	WHOIS records change	2	DiffRegistrar = 1, DiffStatus = 1
②SOA Records	7	Unique SOA records count	3	MNAME = 3, ...
	8	Days between SOA dates	1	SinceLastUpdate = 265
	9	Days between SOA updates	4	Min = 141, Mean = 472, Median = 472, Max = 803
	10	Days difference between SOA records	1	DiffDate = 141
	11	SOA records change	2	DiffMNAME = 1, ...
③TLS Certificates	12	Unique TLS certificates count	6	issuerO = 1, NotAfter = 3, ...
	13	Days between TLS dates	2	UntilNotAfter = −421, ...
	14	Days between TLS updates	4	Min = 31, Mean = 74, Median = 59, Max = 117
	15	Days difference between TLS certificates	2	DiffNotAfter = 116, ...
	16	TLS certificates change	4	issuerO = 0, ...

using WHOIS, SOA, and TLS records as feature sources lies in their ability to encapsulate different aspects of a domain's lifecycle. WHOIS Records provide a foundational glimpse into the domain's registration and administrative lineage, crucial for identifying ownership changes or inactivity that may signal misuse. SOA Records reflect operational readiness within the DNS hierarchy, marking key lifecycle events that could indicate a domain's transition to active use or susceptibility to malicious exploitation. TLS Certificates signal a domain's readiness for secure online operations, with changes in certificates often preceding the active deployment of web services, including those for nefarious purposes. At a designated date in the Domain Timeline, features are generated solely from data available prior to that point in time, from the perspective of WHOIS records, SOA records, and TLS certificates.

3.4.1. ① *WHOIS Records*

No. 1: Unique WHOIS records count. Variations in the registrar or domain name status can indicate changes in the domain's lifecycle. Features are thus designed to effectively capture such histories. Specifically, the count of unique values historically registered for items such as Registrar, Creation Date, Expiry Date, and Domain Status in the WHOIS records is employed as a feature. For example, at the point in time of 2022-11-22 in Fig. 4, there are two unique values registered historically for Registrar, Foo Inc and Bar Ltd, resulting in a Unique WHOIS records count of 2. Actual examples of feature vectors are provided in Table 2.

No. 2: Days between WHOIS dates. These features are designed to quantify the time elapsed since lifecycle changes at each feature extraction point. Specifically, the number of days from dates such as the

Creation Date, Expiry Date, and Updated Date in the WHOIS records to the point in time is used as a feature.

No. 3: Domain status. At each feature extraction point, the most recent domain status (e.g., clientTransferProhibited) from the WHOIS record is used as a feature to determine whether the domain name is in a state that is normatively usable from a lifecycle perspective.

No. 4: Days between WHOIS updates. These features are designed to leverage lifecycle changes by assessing the frequency of WHOIS record updates up to each feature extraction point. Specifically, the minimum, mean, median, and maximum number of days between updates across multiple historical WHOIS records are used as features.

No. 5: Days difference between WHOIS records. To discern whether WHOIS changes are routine updates for legitimate purposes, the difference in days between the dates (Updated Date, Creation Date, Expiry Date) of the most recent WHOIS record and the preceding one is used as a feature at each feature extraction point.

No. 6: WHOIS records change. To ascertain whether the most recent WHOIS change was for legitimate reasons, a boolean feature is employed at each feature extraction point to determine whether the Registrar and Domain Status of the most recent WHOIS record differ from the previous one.

3.4.2. ② *SOA Records*

No. 7: Unique SOA records count. Alterations in the DNS SOA records can signify lifecycle shifts for a domain name. Features are designed to effectively capture such histories. Specifically, the number of distinct values historically registered for items such as MNAME, RNAME, and SERIAL in the SOA records is utilized as a feature.

No. 8: Days between SOA dates. This feature helps understand the time elapsed since lifecycle changes at each feature extraction point. Specifically, the number of days from the last updated date of SOA record to the point in time is used as a feature.

No. 9: Days between SOA updates. These features leverage lifecycle changes by determining the frequency of SOA record updates up to each feature extraction point. Specifically, the minimum, mean, median, and maximum number of days between updates across multiple historical SOA records are used as features.

No. 10: Days difference between SOA records. To distinguish whether SOA changes are regular updates for legitimate purposes, the difference in days between the dates of the most recent SOA record and the previous one is used as a feature at each feature extraction point.

No. 11: SOA records change. To identify whether the most recent SOA change was for legitimate reasons, a boolean feature is used at each feature extraction point to determine whether the MNAME and RNAME of the most recent SOA record differ from the preceding one.

3.4.3. ③ *TLS Certificates*

No. 12: Unique TLS certificates count. Shifts indicated by changes in TLS certificates are critical to understanding the domain's lifecycle. Features are therefore designed to effectively capture these histories. Specifically, the count of unique values historically registered for items such as Issuer C, Issuer CN, Issuer O, Validity Not Before, Validity Not After, and Subject CN in the TLS certificates is used as a feature.

No. 13: Days between TLS dates. These features gauge the time elapsed since lifecycle changes at each feature extraction point. Specifically, the number of days from each date such as Validity Not Before and Validity Not After in the TLS certificates to the point in time is used as a feature.

No. 14: Days between TLS updates. These features utilize lifecycle changes by assessing the frequency of TLS record updates up to each feature extraction point. Specifically, the minimum, mean, median, and maximum number of days between updates across multiple historical TLS certificates are used as features.

No. 15: Days difference between TLS certificates. To determine whether TLS changes are routine updates for legitimate purposes, the difference in days between the dates (Validity Not Before, Validity Not After) of the most recent TLS record and the one prior is used as a feature at each feature extraction point.

No. 16: TLS certificates change. To discern whether the most recent TLS change was for legitimate reasons, a boolean feature is employed at each feature extraction point to determine whether the Issuer C, Issuer CN, Issuer O, and Subject CN of the most recent TLS record differ from the previous one.

3.5. ❸ *Labeling*

Labeling is a step exclusive to the training phase, wherein ground truth labels are assigned to indicate the risk associated with domain names at each point in time. The original purpose of DomainDynamics was to predict the likelihood of a domain name being used for an attack in advance.

In this study, ground truth labels denote whether a domain name will be used for an attack within N days following a specific point in time. For each temporal point extracted in the Feature Extraction step, if there is verified evidence that the domain name was used for an attack within the subsequent N days, a Malicious/Positive label is assigned; otherwise, a Non-malicious/Negative label is applied. The parameter N is adjustable and will be evaluated in detail in Section 4.3.

For instance, if N is set to 30 and it is known that a domain name hosted a phishing site on January 30, 2024, then features of that domain name prior to January 1, 2024, would receive a Non-malicious/Negative label, as there was no attack within the following 30 days. Conversely, features after January 1, 2024, would be labeled Malicious/Positive because the domain would be used for an attack within the next 30 days.

3.6. ❹ *Training ML Model*

This step, which is part of the training process, involves creating a supervised ML model using the features and labels of domain names associated with each time point. In supervised ML, a model is initially trained with known data (in this case, the features and labels at each time point), and then this model is used to predict the risk for new, unseen data (in this case, the features of the Target Domain Name).

Several algorithms are suitable for supervised ML; given that the problem is defined with temporal features for domain names and binary labels (Positive/Negative), any binary classification algorithm is applicable. The performance of various algorithms will be assessed in Section 4.3.

3.7. ❺ *Predicting Risks*

During the prediction phase, the ML Model trained earlier is used to estimate the risk of the Target Domain Name being exploited in an attack within the next N days. This estimation is based on the features at each time point for the Target Domain Name. The risk at each point is represented by a prediction probability generated by the ML Model.

The prediction probability is a continuous value ranging from 0 to 1, with values closer to 0 suggesting Non-malicious/Negative and those closer to 1 indicating Malicious/Positive. The risk for each time point of the Target Domain Name is presented as a Risk Timeline, as depicted at the bottom of Fig. 4.

In addition to predicting the risk probability, this step incorporates techniques from XAI (eXplainable Artificial Intelligence) to elucidate the features influencing each probability prediction. In this research, SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) is utilized as the XAI method. SHAP quantifies the contribution of each input feature to the prediction probability provided by the ML Model, assigning SHAP Values that reflect the impact of each feature.

Using SHAP Values, the top contributing features, both positive and negative, are identified for each risk prediction at every time point. For example, as demonstrated in the Risk Timeline of Fig. 4, hovering over the prediction probability at a specific time point reveals the top contributing features that influenced the probability at that moment.

4. Evaluation

In this section, we conduct a performance evaluation to answer the following three questions, assessing the effectiveness of DomainDynamics:

Q1: Can DomainDynamics predict the risk of a domain name being used in an attack before it occurs?

Q2: What are the intrinsic characteristics of DomainDynamics?

Q3: How does DomainDynamics compare to baseline systems in terms of detection capability?

To address Q1, we compiled a dataset of domain names that were actually used in attacks and evaluated our system's predictive performance. For Q2, we explore the impact of system-required parameters and the influence of features on predictions, characterizing DomainDynamics. In response to Q3, we introduce new evaluation metrics that account for the temporal change in the risk of domain names and present a comparative performance analysis of DomainDynamics against two baseline systems.

4.1. Metrics

We introduce new evaluation metrics that consider the temporal changes in domain name risk, enabling a more nuanced understanding of detection capabilities. We redefine and utilize the following four metrics:

True Positive (TP): A case is considered a True Positive if a malicious domain name, actually used in an attack (e.g., malware, phishing), is detected before the attack takes place.

False Negative (FN): A case is considered a False Negative if a malicious domain name, actually used in an attack, is not detected before the attack takes place.

False Positive (FP): A case is considered a False Positive if a benign domain name, not used in an attack, is incorrectly identified as malicious at least once during the observation period.

True Negative (TN): A case is considered a True Negative if a benign domain name, not used in an attack, is never falsely identified as malicious during the observation period.

These metrics can be expressed mathematically as follows:

$TP : \forall t \in T_{attack}, y_t = 1, \hat{y}_t = 1$

$FN : \forall t \in T_{attack}, y_t = 1, \hat{y}_t = 0$

$FP : \forall t \in T_{all}, y_t = 0, \hat{y}_t = 1$

$TN : \forall t \in T_{all}, y_t = 0, \hat{y}_t = 0$

Here, T_{attack} denotes any point before an attack occurs, T_{all} represents any point during the entire observation period, y_t is the actual label at time t, and $\hat{y}_t$ is the predicted label at time t.

Additionally, we define the following four ratio-based metrics:

False Positive Rate (FPR): FPR is the proportion of benign domain names, not used in an attack, that are falsely judged to be malicious at least once during the period, defined as $FPR = \frac{FP}{FP+TN}$.

Precision: Precision is the proportion of domain names identified as malicious that were actually used in an attack, defined as $Precision = \frac{TP}{TP+FP}$.

Recall or True Positive Rate (TPR): Recall/TPR is the proportion of domain names actually used in an attack that were correctly identified as malicious, defined as $Recall/TPR = \frac{TP}{TP+FN}$.

F1 Score: The F1 Score is the harmonic mean of Precision and Recall, defined as $F1 = \frac{2\times Precision\times Recall}{Precision+Recall}$.

Accuracy: Accuracy is a measure of overall correctness of the model's predictions across all classes. It is defined as the ratio of correctly predicted instances (both true positives and true negatives) to the total number of instances. The formula for Accuracy is given by $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$.

The rationale for establishing such metrics is to accurately assess detection performance considering the temporal dynamics of domain name risk. In traditional studies, as documented in Table 1, domain names collected over a certain period using blocklists and similar methods are labeled "Malicious", even if they are no longer active in attacks thereafter. Counting predictions of these domain names as TPs can lead to an inflated detection rate. Furthermore, our new metrics are particularly stringent for DomainDynamics, as any single error made during the observation period results in an FP or FN.

Table 3
Summary of ground truth datasets used for training and predicting domain names, covering an 18-month period from April 1, 2022, to September 30, 2023.

Dataset	Ground truth	For training	For predicting	# FQDNs
Malware	Malicious	✓	✓	3410
Phishing	Malicious	✓	✓	30,711
CrowdCanary	Malicious	✓	✓	51,502
Tranco Top10k	Non-malicious		✓	4358

Table 4
Breakdown of Malware Families in the Malware Dataset.

Rank	Malware family	# FQDNs
1	Necurs	587
2	Locky	583
3	SharkBot	247
4	Monero	243
5	MyDoom	233
6	Ramnit	200
7	QakBot	153
8	Bedep	143
9	Flubot	129
10	DirCrypt	119
11	Zloader	76
12	Orchard	75
13	Pitou	64
14	Dromedan	47
15	Sphinx	39
16	MetaStealer	38
17	GOZ	37
18	PadCrypt	35
19	Tinba	35
20	Proslikefan	33

4.2. Dataset

To accurately evaluate the system using the new metrics described previously, a ground truth dataset is essential. Such a dataset must accurately reflect periods when a given domain name was not used for attacks and when it began to be used for attacks. Therefore, in this study, we carefully prepared three types of malicious ground truth datasets and one type of benign ground truth dataset. The details of each dataset are summarized in Table 3.

Malware. The malware dataset comprises domain names that were confirmed to function as command and control (C2) for malware and botnet operations at the time of observation, indicating when a C2 domain name was active. This dataset was obtained through a partnership with a managed security service provider and includes unique C2 fully qualified domain names (FQDNs) of 47 malware families, collected over 18 months from April 2022 to September 2023, totaling 3410 entries. Table 4 provides a detailed breakdown of the malware families included in the Malware Dataset. The dataset comprises a diverse range of malware, including Necurs, Locky, SharkBot, Monero, and MyDoom, which are among the most prevalent families observed.

Phishing. The phishing dataset contains domain names that were active as phishing sites at the time of observation. Specifically, we utilized commercial OpenPhish (OpenPhish, 2024) data and targeted those with evidence of operation, such as screenshots taken at specific timestamps. This dataset includes 30,711 unique FQDNs over the same 18-month period.

CrowdCanary. CrowdCanary is a dataset comprising domain names that users have definitively reported as being used in phishing emails

Table 5
Performance of DomainDynamics with different parameters, highlighting selected parameters.

	Parameter	FPR (↓better)	Precision (↑better)	Recall (↑better)	F1 (↑better)	Accuracy (↑better)
Feature Set	WHOIS	0.96%	99.29%	35.52%	52.32%	48.82%
	SOA	0.60%	99.66%	46.57%	63.48%	57.64%
	TLS	1.88%	99.33%	73.85%	84.71%	78.93%
	WHOIS+SOA	**0.32%**	99.85%	58.35%	73.66%	67.01%
	WHOIS+TLS	0.69%	99.78%	81.80%	89.90%	85.47%
	SOA+TLS	0.78%	99.73%	77.25%	87.06%	81.85%
	WHOIS+SOA+TLS	0.41%	**99.87%**	**82.58%**	**90.40%**	**86.14%**
Labeling Period	**7**	**0.41%**	**99.87%**	82.58%	90.40%	86.14%
	14	0.73%	99.77%	82.84%	90.52%	86.28%
	30	1.10%	99.66%	85.82%	92.23%	88.56%
	60	3.49%	99.07%	97.92%	98.49%	97.63%
	90	3.53%	99.06%	**98.24%**	**98.64%**	**97.87%**
ML Algorithm	Decision Tree	9.59%	96.93%	80.35%	87.86%	82.45%
	Random Forest	1.93%	99.38%	81.89%	89.79%	85.28%
	LightGBM	1.24%	99.60%	82.21%	90.08%	85.68%
	XGBoost	**0.41%**	**99.87%**	**82.58%**	**90.40%**	**86.14%**
Training Dataset	Malware	**0.05%**	**99.98%**	51.77%	68.21%	61.86%
	Phishing	0.46%	99.85%	81.29%	89.62%	85.12%
	CrowdCanary	0.09%	99.97%	78.48%	87.93%	82.97%
	Malware+Phishing	0.46%	99.85%	82.13%	90.13%	85.77%
	Malware+CrowdCanary	0.18%	99.94%	78.90%	88.18%	83.28%
	Phishing+CrowdCanary	0.57%	99.82%	82.55%	90.36%	86.08%
	Malware+Phishing+CrowdCanary	0.41%	99.87%	**82.58%**	**90.40%**	**86.14%**
Training Period	**3 months**	**0.41%**	**99.87%**	82.58%	90.40%	86.14%
	6 months	0.55%	99.83%	84.11%	91.30%	87.32%
	9 months	0.60%	99.81%	**84.18%**	**91.33%**	**87.36%**
	12 months	0.69%	99.78%	82.85%	90.53%	86.30%

or phishing websites at the time of observation. CrowdCanary, a system recently proposed in literature (Nakano et al., 2023), monitors text and images posted on platforms like X/Twitter to extract phishing reports, and we received a research dataset from the authors. We used 42,756 posts collected over 18 months, with the criteria that the tweets were from security experts, included images (indicating that the sites were accessible at the time), and contained URL or domain name information in the tweet text (ensuring high reliability of the information). This dataset includes domains identified in spam campaigns, thus broadening the scope of our analysis beyond targeted phishing attacks. Some posts contain multiple domain names, resulting in a total of 51,502 unique FQDNs.

These three malicious datasets will be used for training and evaluating the ML model's predictions. To prevent temporal leakage—where future data (the ground truth) would be used to predict past data—in the evaluation, we will use data up to March 2023 for training and data from April 2023 to September 2023 for prediction. It is important to note that there are no overlapping FQDNs between these three datasets, and naturally, no FQDNs overlap between the training and prediction sets.

Tranco Top10k (Non-malicious). The ground truth for benign FQDNs consists of domain names that were consistently ranked in the top 10,000 of the Tranco list (Pochat et al., 2019) during the 18 months from April 2022 to September 2023. To ensure the absence of malicious domains, each domain's website content was manually checked. These were used only as a reference for evaluation during prediction as FQDNs that were never used for attacks during the period and were not used for training. As outlined in Section 3.5, DomainDynamics considers the lifecycle of a domain name, labeling it as Non-malicious before it is used for an attack and as Malicious when it is used for an attack. Therefore, it should be emphasized that FQDNs that have always been benign are not necessarily required for training the ML model, and in this case, they are not used. The Tranco Top10k is prepared solely to evaluate whether DomainDynamics is causing obvious false positives.

Data Availability. We have made a portion of the evaluation datasets available for academic research purposes. Due to contractual obligations with our data providers, we cannot release the entire dataset publicly. However, interested researchers can request access to a subset of the data for validation and further study. The process for obtaining access, along with the terms and conditions, is detailed at https://domaindynamics.github.io/.

4.3. Performance with variable parameters

We evaluated the performance of DomainDynamics by varying parameters adjustable within the system, utilizing the metrics mentioned in Section 4.1. The parameters in question include the Feature Set, Labeling Period, ML Algorithm, Training Dataset, and Training Period, totaling five. Due to space constraints, we report only the results of evaluations with other parameters held constant, while varying the target parameter.

Feature Set. As explained in Section 3.4, DomainDynamics prepares three types of feature sets: WHOIS records, SOA records, and TLS certificates. We evaluated the difference in performance when using combinations of these feature sets.

Table 5 summarizes the detection performance when changing the feature set to WHOIS, SOA, TLS, WHOIS+SOA, WHOIS+TLS, SOA+TLS, and WHOIS+SOA+TLS (ALL). For evaluation metrics, a lower FPR is preferable, while higher values in Precision, Recall/TPR, F1, and Accuracy are desirable. It should be noted that Precision and Recall/TPR are generally in a trade-off relationship, so it may not be possible to optimize both simultaneously. From the perspective of FPR, the combination of WHOIS+SOA shows the best results with 0.32. However, when using WHOIS+SOA, the Recall/TPR is relatively low at 58.35%, resulting in an F1 of only 73.66%, indicating a modest proportion of malicious domain names detected before an attack begins. When using WHOIS+SOA+TLS (ALL), the Recall/TPR is the highest at 82.58%, resulting in the highest F1 of 90.40% and Accuracy of 86.14%. Although the FPR is higher compared to WHOIS+SOA, it remains a low value of 0.41%, leading us to conclude that WHOIS+SOA+TLS (ALL) is the most effective parameter set.

Labeling Period. We evaluated the performance of DomainDynamics by varying the labeling period N as per the procedure in Section 3.5. This serves to investigate how well the system performs in predicting future risks. Table 5 summarizes the detection performance when changing N to 7, 14, 30, 60, and 90 days. The results show that at $N = 7$, the FPR is the lowest at 0.41% and the Precision is the highest at 99.87%. Increasing N results in a higher FPR, which is expected as a longer forecasting period increases the likelihood of making incorrect predictions. Conversely, for Recall/TPR and F1, except for the case of $N = 14$, the values increase as N increases because a longer forecasting period provides more opportunities to detect malicious domain names before attacks begin. This result indicates a trade-off, and the choice of parameter will be determined by the acceptable FPR. However, considering the volume of data handled in this study (tens of thousands of cases or more), an FPR of 1% or more would result in an excessive number of false positives. Therefore, we select $N = 7$ with an FPR of 0.41% as the optimal parameter.

ML Algorithm. We evaluated the performance of DomainDynamics by varying the ML algorithms used for binary classification. For this study, we selected and compared four types of ML algorithms: Decision Tree (Quinlan, 1986), Random Forest (Breiman, 2001), LightGBM (Ke et al., 2017), and XGBoost (Chen and Guestrin, 2016). We chose these algorithms because they are all tree-based, which means they do not require feature scaling, and we expect them to be robust to differences in feature sets and variations in training data type and period. Additionally, by selecting algorithms with significantly different characteristics, we can compare their respective features. Specifically, Decision Tree is a basic tree model, Random Forest is an ensemble learning algorithm combining many decision trees, and LightGBM and XGBoost are sequential tree-building algorithms known as gradient boosting trees. For adjustable parameters of each selected algorithm, we used Optuna (Akiba et al., 2019) to search for parameters that minimize FPR and selected the best ones. Table 5 summarizes the performance for each ML algorithm. The results are clear: XGBoost performed best across all metrics, with an FPR of 0.41%, Precision of 99.87%, Recall/TPR of 82.58%, F1 of 90.40%, and Accuracy of 86.14% leading us to conclude that XGBoost is the most suitable algorithm for our experimental setup.

Training Dataset. We evaluated the performance of DomainDynamics by varying the combinations of malicious ground truth included in the Training Dataset. We assessed different combinations of the three datasets outlined in Table 3: Malware, Phishing, and CrowdCanary, including individual use and combinations such as Malware+Phishing, Malware+CrowdCanary, Phishing+CrowdCanary, and Malware+Phishing+CrowdCanary (ALL). Note that the Predicting Dataset in all cases include all three datasets: Malware, Phishing, and CrowdCanary, and we are evaluating how the combination of Training Dataset affects detection performance. Table 5 summarizes the performance for each Training Dataset. The results show that when Malware is used alone as the Training Dataset, the FPR is extremely low at 0.05%, and the Precision is very high at 99.98%, but the Recall/TPR is 51.77%, and the F1 is only 68.21%, as expected. This is because when only Malware is used as the Training Dataset, the model learns only the time-series changes used by Malware domains and cannot adapt to a wide range of patterns used by domains for other attacks. For other patterns, there is no significant difference, but from the perspective of Recall/TPR, F1, and Accuracy, the ALL (Malware+Phishing+CrowdCanary) pattern shows the highest values. This is expected because using all three datasets allows learning the time-series changes of domain names used in each type of attack. We select this pattern as the most effective parameter for our study. While we have evaluated the performance changes when varying the combination of Training Dataset here, we will evaluate the detection breakdown of Malware, Phishing, and CrowdCanary included in the Predicting Dataset (e.g., which dataset's domain names are detected well?) in comparison with the baselines in Section 4.6.

Training Period. As shown in Table 3, the three datasets, Malware, Phishing, and CrowdCanary, were collected over a period of 18 months from April 2022 to September 2023. We evaluated whether performance improves by using longer-term training data. Specifically, as shown in Table 6, we assessed performance by changing the Training Period to 3 months (from January to March 2023), 6 months (from October 2022 to March 2023), 9 months (from July 2022 to March 2023), and 12 months (from April 2022 to March 2023). Table 6 also shows the breakdown of Training and Predicting for each pattern. Table 5 summarizes the differences in performance when creating ML models with different Training Period patterns. The results indicate no significant difference when changing the Training Period, but from the perspective of FPR and Precision, 3 months shows the best results, while from the perspective of Recall/TPR, F1, and Accuracy, 9 months shows the best results. This suggests that using more recent data allows for learning the latest trends to better control FPR. Although all Training Periods are within a selectable range of detection performance, we select 3 months as the optimal parameter because it achieves over 90% in F1 while keeping FPR low.

Table 6
Breakdown of training and predicting datasets.

Usage	Dataset	Period	# FQDNs
Training	Malware	2023-01-01 – 2023-03-31 (3 months)	361
Training	Phishing	2023-01-01 – 2023-03-31 (3 months)	3935
Training	CrowdCanary	2023-01-01 – 2023-03-31 (3 months)	3763
Training	Total	2023-01-01 – 2023-03-31 (3 months)	8059
Training	Malware	2022-10-01 – 2023-03-31 (6 months)	738
Training	Phishing	2022-10-01 – 2023-03-31 (6 months)	9997
Training	CrowdCanary	2022-10-01 – 2023-03-31 (6 months)	9881
Training	Total	2022-10-01 – 2023-03-31 (6 months)	20,616
Training	Malware	2022-07-01 – 2023-03-31 (9 months)	1046
Training	Phishing	2022-07-01 – 2023-03-31 (9 months)	17,463
Training	CrowdCanary	2022-07-01 – 2023-03-31 (9 months)	23,525
Training	Total	2022-07-01 – 2023-03-31 (9 months)	42,034
Training	Malware	2022-04-01 – 2023-03-31 (12 months)	1414
Training	Phishing	2022-04-01 – 2023-03-31 (12 months)	22,793
Training	CrowdCanary	2022-04-01 – 2023-03-31 (12 months)	44,967
Training	Total	2022-04-01 – 2023-03-31 (12 months)	69,174
Predicting	Malware	2023-04-01 – 2023-09-30 (6 months)	1996
Predicting	Phishing	2023-04-01 – 2023-09-30 (6 months)	7918
Predicting	CrowdCanary	2023-04-01 – 2023-09-30 (6 months)	6535
Predicting	Total	2023-04-01 – 2023-09-30 (6 months)	16,449

Summary. In summary, DomainDynamics demonstrates its best performance with the Feature Set WHOIS+SOA+TLS (ALL), a Labeling Period of 7 days, the XGBoost ML Algorithm, the Malware+Phishing+CrowdCanary (ALL) Training Dataset, and a Training Period of 3 months. In subsequent evaluations, we will employ DomainDynamics trained with these parameters. It should be noted that we operated DomainDynamics on a standard Linux machine (4-core CPU, 16 GB memory) and confirmed that it functions sufficiently fast without the need for a GPU, as it utilizes conventional binary classification ML. It took on average only 0.73 s to input a domain name and output the risk prediction result with the best performance parameter combination and even when running in single-threaded mode without parallelization, proving that it is practically efficient. This processing time encompasses the entire workflow, including the construction of the domain timeline from historical data. Furthermore, the storage required for processing over 85,000 domains was less than 2 GB, which includes debug logs and raw data. This demonstrates the efficiency of the system, both in terms of processing speed and storage utilization, highlighting the effectiveness of our proposed features derived from raw WHOIS, SOA, and TLS logs.

4.4. Analysis of features

To better understand the behavior of DomainDynamics, we investigated the features that contributed to the model's output. As indicated by the results in Section 4.3, DomainDynamics performs optimally when utilizing XGBoost, which employs gradient boosting trees with multiple decision trees, rather than a single, simple decision tree. However, the ensemble nature of the model obscures the rationale behind its final output. To elucidate the features that contributed to the risk assessment by DomainDynamics, we utilized the SHAP values output, as discussed in Section 3.7.

SHAP (SHapley Additive exPlanations) is a method that explains the contribution of each feature to a specific prediction result, producing SHAP values that quantify this contribution. Positive SHAP values

Table 7
Comprehensive ranking and scores of features influencing DomainDynamics predictions, derived from SHAP value analysis, highlighting the most significant contributors to both positive/malicious and negative/non-malicious predictions.

Rank	Feature contributed to positive/malicious predictions	Score	Rank	Feature contributed to negative/non-malicious predictions	Score
1	③TLS, No. 13, Days between TLS dates	85,363	1	③TLS, No. 13, Days between TLS dates	–83,161
2	②SOA, No. 8, Days between SOA dates	30,689	2	②SOA, No. 8, Days between SOA dates	–36,663
3	①WHOIS, No. 2, Days between WHOIS dates	24,023	3	①WHOIS, No. 2, Days between WHOIS dates	–25,834
4	②SOA, No. 7, Unique SOA records count	9972	4	②SOA, No. 7, Unique SOA records count	–12,173
5	③TLS, No. 12, Unique TLS certificates count	7650	5	③TLS, No. 15, Days difference between TLS certificates	–4895
6	②SOA, No. 9, Days between SOA updates	5296	6	③TLS, No. 12, Unique TLS certificates count	–4757
7	③TLS, No. 14, Days between TLS updates	4657	7	③TLS, No. 14, Days between TLS updates	–4362
8	③TLS, No. 15, Days difference between TLS certificates	4230	8	②SOA, No. 9, Days between SOA updates	–3788
9	①WHOIS, No. 3, Domain status	3583	9	①WHOIS, No. 3, Domain status	–3193
10	①WHOIS, No. 4, Days between WHOIS updates	1918	10	①WHOIS, No. 4, Days between WHOIS updates	–3033
11	②SOA, No. 10, Days difference between SOA records	900	11	①WHOIS, No. 1, Unique WHOIS records count	–1177
12	③TLS, No. 16, TLS certificates change	574	12	②SOA, No. 10, Days difference between SOA records	–633
13	①WHOIS, No. 5, Days difference between WHOIS records	567	13	①WHOIS, No. 5, Days difference between WHOIS records	–525
14	①WHOIS, No. 1, Unique WHOIS records count	479	14	①WHOIS, No. 6, WHOIS records change	–475
15	①WHOIS, No. 6, WHOIS records change	425	15	③TLS, No. 16, TLS certificates change	–384
16	②SOA, No. 11, SOA records change	283	16	②SOA, No. 11, SOA records change	–250

suggest that the feature has pushed the model's prediction outcome in a positive direction, while negative values indicate a negative contribution. We employed TreeSHAP (Lundberg et al., 2020) to examine the features that contributed positively and negatively to the output of DomainDynamics. Specifically, for each domain name in the Predicting Dataset, we calculated the contribution of each feature to the decision made by XGBoost, expressed as SHAP values, and aggregated these values to determine the overall impact of each feature on the decisions for all domain names in the Predicting Dataset.

Table 7 shows that the features contributing to the output of DomainDynamics are not biased towards any specific set of the three feature sets: WHOIS, SOA, and TLS; individual features from each of these sets appear at the top of the list. This aligns with the findings from Section 4.3, which indicate that accuracy improves when ALL (WHOIS+SOA+TLS) Feature Set is used. Notably, features such as No. 13 (Days between TLS dates), No. 8 (Days between SOA dates), and No. 2 (Days between WHOIS dates) are prominent, appearing in the top three for both positive and negative contributions. These features, specifically designed to capture the importance of being aware of lifecycle changes—a problem first acknowledged in this study—are playing a significant role in determining the risk of domain names, as anticipated. Moreover, these features are novel, as they have not been used in prior research, as shown in Table 1.

On the other hand, it is evident that some individual features contribute less significantly (e.g., features ranked below 11). While it might be possible to eliminate such low-contributing features, doing so would not notably reduce computational load or execution time, given that data from WHOIS, SOA, and TLS is already being processed. Furthermore, since these features have a non-zero score—meaning they do contribute to the risk assessment of certain domain names—we did not consider such reductions in this study.

4.5. Performance on specific domain sets: Less-common TLDs and IDNs

To further understand the performance of DomainDynamics on various types of domains, we analyzed its effectiveness on specific domain sets, namely domains with less-common Top-Level Domains (TLDs) and Internationalized Domain Names (IDNs). First, we provide the distribution of TLDs in the Predicting Dataset. Table 8 shows the top 20 TLDs and the number of domains for each dataset (Malware, Phishing, and CrowdCanary). As shown in the table, the .com TLD is the most prevalent across all datasets.

Table 8
TLD Distribution in Malicious Predicting Dataset.

Rank	TLD	Malware	Phishing	CrowdCanary	Total
1	.com	575	3329	1716	5620
2	.cfd	0	70	1645	1715
3	.cn	0	77	993	1070
4	.top	17	301	513	831
5	.org	584	169	30	783
6	.xyz	110	255	396	761
7	.id	0	373	0	373
8	.icu	0	60	289	349
9	.net	30	269	20	319
10	.shop	0	84	219	303
11	.info	125	138	10	273
12	.ru	89	140	1	230
13	.pw	222	5	1	228
14	.cc	11	71	109	191
15	.co	0	151	15	166
16	.br	0	156	1	157
17	.pl	0	152	0	152
18	.click	0	139	1	140
19	.online	8	114	1	123
20	.cyou	0	24	98	122

Next, we evaluated the performance of DomainDynamics on domains with common TLDs (specifically .com) versus those with less-common TLDs (all TLDs other than .com). Table 9 presents the results, showing the FPR, Precision, Recall, F1-score, and Accuracy for each set. The results indicate that the performance of DomainDynamics is comparable between the two sets, although there is some variation in specific metrics. For example, the Recall for .com domains is slightly higher than for other TLDs across all datasets in some cases. However, the F1-score, which balances Precision and Recall, remains high for both sets.

Finally, we analyzed the presence and performance of DomainDynamics on IDNs within the Predicting Dataset. The Phishing Dataset contained 56 IDNs out of 7918 domains. Among these, 48 were correctly identified as True Positives, while 8 were False Negatives, resulting in a Recall of 85.71%. This demonstrates that DomainDynamics is capable of detecting malicious IDNs with high accuracy, although the relatively small number of IDNs in the dataset suggests that further evaluation with a larger IDN dataset would be beneficial.

4.6. Comparison with baseline systems

In comparison to DomainDynamics, this study evaluates the performance against previous research and existing commercial security services using the same dataset and metrics.

Table 9
Performance of DomainDynamics on Common (.com) and Less-Common TLDs.

Predicting dataset	TLD	FPR (↓better)	Precision (↑better)	Recall (↑better)	F1 (↑better)	Accuracy (↑better)
Malware	.com	0.47%	97.26%	**74.09%**	**84.11%**	**94.82%**
Malware	Others	**0.33%**	**99.41%**	71.57%	83.22%	87.37%
Phishing	.com	0.47%	99.50%	72.45%	83.85%	84.15%
Phishing	Others	**0.33%**	**99.83%**	**78.10%**	**87.64%**	**84.24%**
CrowdCanary	.com	0.47%	99.27%	**95.69%**	**97.45%**	**97.98%**
CrowdCanary	Others	**0.33%**	**99.87%**	93.42%	96.54%	95.14%
Malware+Phishing+CrowdCanary	.com	0.47%	99.73%	79.72%	88.61%	85.87%
Malware+Phishing+CrowdCanary	Others	**0.33%**	**99.93%**	**84.06%**	**91.31%**	**86.31%**

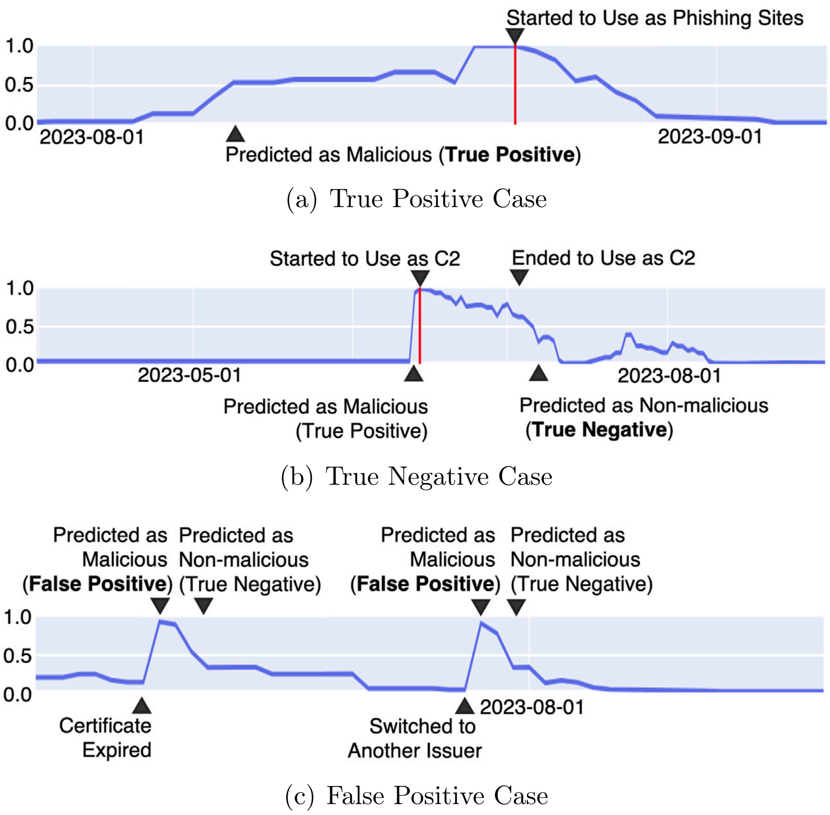
Table 10
Comparative performance metrics of DomainDynamics against baseline systems (DomainProfiler and VirusTotal). Delay indicates the average time difference (in days) between the prediction of a malicious domain and its actual use in an attack.

Predicting dataset	System	FPR (↓better)	Precision (↑better)	Recall (↑better)	F1 (↑better)	Accuracy (↑better)	Delay (↓better)
Malware	**DomainDynamics**	0.41%	98.77%	**72.29%**	**83.48%**	**91.01%**	**−14.05**
	DomainProfiler	**0.00%**	**100.00%**	4.71%	9.00%	70.07%	−12.95
	VirusTotal	0.05%	96.55%	2.81%	5.45%	69.44%	−7.95
Phishing	**DomainDynamics**	0.41%	99.70%	**75.73%**	**86.08%**	**84.20%**	**−13.47**
	DomainProfiler	**0.00%**	**100.00%**	10.81%	19.51%	42.47%	−3.20
	VirusTotal	0.05%	99.38%	4.05%	7.79%	38.10%	−0.39
CrowdCanary	**DomainDynamics**	0.41%	99.71%	**94.02%**	**96.78%**	**96.25%**	**−3.11**
	DomainProfiler	**0.00%**	**100.00%**	5.95%	11.24%	43.58%	−0.70
	VirusTotal	0.05%	99.08%	3.29%	6.37%	41.96%	−0.32
ALL	**DomainDynamics**	0.41%	99.87%	**82.58%**	**90.40%**	**86.14%**	**−9.13**
	DomainProfiler	**0.00%**	**100.00%**	8.14%	15.06%	27.38%	−3.16
	VirusTotal	0.05%	99.66%	3.60%	6.95%	23.78%	−1.08

DomainDynamics. DomainDynamics employs the optimal parameters identified in Section 4.3 and reports the performance when evaluating Malware, Phishing, and CrowdCanary in the Predicting Dataset individually as well as their combined performance in Table 10. The results for DomainDynamics in Table 10 for ALL (Malware+Phishing+CrowdCanary) correspond to those when the Training Period in Table 5 is set to three months. When comparing individual results for Malware, Phishing, and CrowdCanary in Table 10, it is evident that the detection performance (F1) is superior in the order of Malware, Phishing, and CrowdCanary. This indicates that the difficulty of assessing the risk of domain names varies depending on the type of attack. In the context of this study, Delay in Table 10 represents the average time difference (in days) between the date when a system identified a domain as malicious and the date when the domain was actually used in an attack. Delay is calculated only for True Positives, i.e., instances where the system correctly identified a domain as malicious before it was used in an attack. A negative Delay value indicates that the system successfully detected the malicious domain *before* the attack occurred, demonstrating the system's proactive detection capability. A Delay of 0 means the detection occurred on the same day as the attack. The average Delay for DomainDynamics is −14.05 days for Malware, −13.47 days for Phishing, −3.11 days for CrowdCanary, and −9.13 days for ALL datasets. These results demonstrate that DomainDynamics can proactively detect malicious domains, on average, 9.13 days before they are used in attacks.

Baseline (DomainProfiler). First, we consider the comparison with previous research. It is important to note, as described in Section 2, that DomainDynamics is the first to provide outputs that consider changes in risk, which was not possible in conventional research, hence a direct comparison with prior work is inherently limited. Specifically, conventional studies have reported accuracy evaluation results by labeling domain names collected over a certain period as "Malicious" using blocklists, etc., even if they were no longer used in attacks after a certain point. If predicted as malicious, they are considered true positives. However, this study does not use such metrics for evaluation; instead, we assess and compare with our metrics (Section 4.1) centered on whether detection occurs prior to an attack. Moreover, the previous studies mentioned in Table 1 are neither open source nor provide publicly available labeled data for constructing ML models, making them effectively inaccessible to third parties. Therefore, we re-implemented DomainProfiler based on the description in the original paper (Chiba et al., 2016), which is the most representative of the previous research with such conventional labeling, and applied it to our dataset to evaluate the performance of DomainProfiler. We obtained the historical output of DomainProfiler for the corresponding domain names. We then calculate the metrics based on these results. The corresponding results can be found in the row for Baseline (DomainProfiler) in Table 10. As a result, DomainProfiler achieved an FPR of 0% and a precision of 100% on our dataset. However, for ALL (Malware+Phishing+CrowdCanary), the Recall/TPR was only 8.14%, the F1 score was limited to 15.06%, and the Accuracy was only 27.38%. The average Delay for DomainProfiler is −12.95 days for Malware, −3.20 days for Phishing, −0.70 days for CrowdCanary, and −3.16 days for ALL datasets. Compared to DomainDynamics, DomainProfiler is less effective in proactive detection especially for Phishing and CrowdCanary.

Baseline (VirusTotal). In evaluating the performance of commercial security services, we consider VirusTotal, a platform that, as of January 2024, can collate detection results from up to 89 different security engines for any given domain name. Our comparison with DomainDynamics is based on the aggregate performance of these engines as reported by VirusTotal. It is important to recognize that the engines included in VirusTotal's analysis may not be identical to their respective commercial versions, which might influence their detection capabilities (VirusTotal, 2024b). For this study, each domain name's detection status was assessed through the VirusTotal API, classifying a domain as malicious if detected by any of the engines, and non-malicious otherwise. This approach reflects our focus on the collective capability to identify threats rather than the performance of individual engines. Notably, one engine, OpenPhish, directly correlates with the phishing dataset utilized in our research. To eliminate potential bias, we excluded OpenPhish from our analysis, relying on the remaining 88 engines for a comprehensive evaluation. The results, as detailed in the row for Baseline (VirusTotal) in Table 10, reveal that VirusTotal



(a) True Positive Case

(b) True Negative Case

(c) False Positive Case

Fig. 5. Illustrative Risk Timelines for a True Positive, True Negative, and False Positive case as determined by DomainDynamics. Each subfigure provides insight into the temporal risk assessment capabilities of DomainDynamics.

achieved a False Positive Rate (FPR) of 0.05% on our dataset. Despite utilizing 88 engines, its overall precision for detecting Malware, Phishing, and CrowdCanary incidents was 99.66%, with a Recall/True Positive Rate (TPR) of only 3.60%, resulting in an F1 score of 6.95% and Accuracy of 23.78%. The average Delay for VirusTotal is −7.95 days for Malware, −0.39 days for Phishing, −0.32 days for CrowdCanary, and −1.08 days for ALL datasets. VirusTotal is less effective than DomainDynamics and DomainProfiler in proactive detection. Even with the collective insights of 88 engines, VirusTotal detected a significantly lower proportion of attacks before their commencement compared to DomainDynamics, highlighting the added value of our lifecycle-aware approach to risk prediction for domain names.

4.7. Real-world deployment

We implemented DomainDynamics in the operational environment of a real-world security service provider, applying it to assess the risk of a vast number of domain names observed in a large-scale network and those newly registered in domain registries.

Data Sources We acquired domain names from two types of data sources for risk assessment using DomainDynamics. The first source was domain names observed through passive DNS traffic, collected from 66 DNS servers installed across 18 countries on global Tier 1 networks. We obtained 55.57 million domain names that were newly resolved during the 28 days from November 27, 2022, to December 24, 2022, and assessed their risk in real-time with DomainDynamics. The second source was newly registered domain names observed in domain registries, collected using Zonefiles (Zonefiles, 2024). Over the same 28-day period, we daily acquired newly registered domain names, totaling 3.86 million, and assessed their risk with DomainDynamics.

Verification Method We verified domain names identified as risky by DomainDynamics using three methods. First, we used the VirusTotal API to check daily for up to one month after detection by DomainDynamics whether those domain names were identified by any of the 88 detection engines. Second, we verified whether the domain names were C2 domains generated on the day of detection by known DGAs using a commercial Threat Intelligence service utilized by the security service provider, checking daily for up to one month after detection. Third, we conducted daily web crawls for up to one month after detection by DomainDynamics. Whenever web access was achieved, we took screenshots to manually check whether the sites were phishing sites. However, there are constraints: the first and second methods are challenging to conduct daily on all items due to the high licensing costs of the APIs used, and the third method is difficult to implement on a large scale and daily in an ethically considerate and scalable way. Therefore, we randomly sampled 100,000 domain names detected by DomainDynamics and verified them daily for up to one month following detection.

Verification Results First, as a result of verification using the VirusTotal API, out of the 100,000 domain names assessed for risk by DomainDynamics, 6937 were detected by at least one of VirusTotal's engines after detection. Although the number is not large due to the generally low detection rate of VirusTotal, as shown in Section 4.6, it was confirmed that there indeed are domain names among those detected by DomainDynamics that are used in malicious activities and judged malicious. Next, as a result of verification using Threat Intelligence, out of the 100,000 domain names assessed for risk by DomainDynamics, 16,674 were confirmed to be C2 domain names generated by DGAs and used after detection. This result indicates that DomainDynamics can preemptively detect C2 domain names generated by DGAs. Lastly, as a result of verification through web crawling, out of the 100,000 domain names assessed for risk by DomainDynamics, 1552 were confirmed to be operating as phishing sites through screenshots taken after detection. Despite the low frequency of web crawls and only

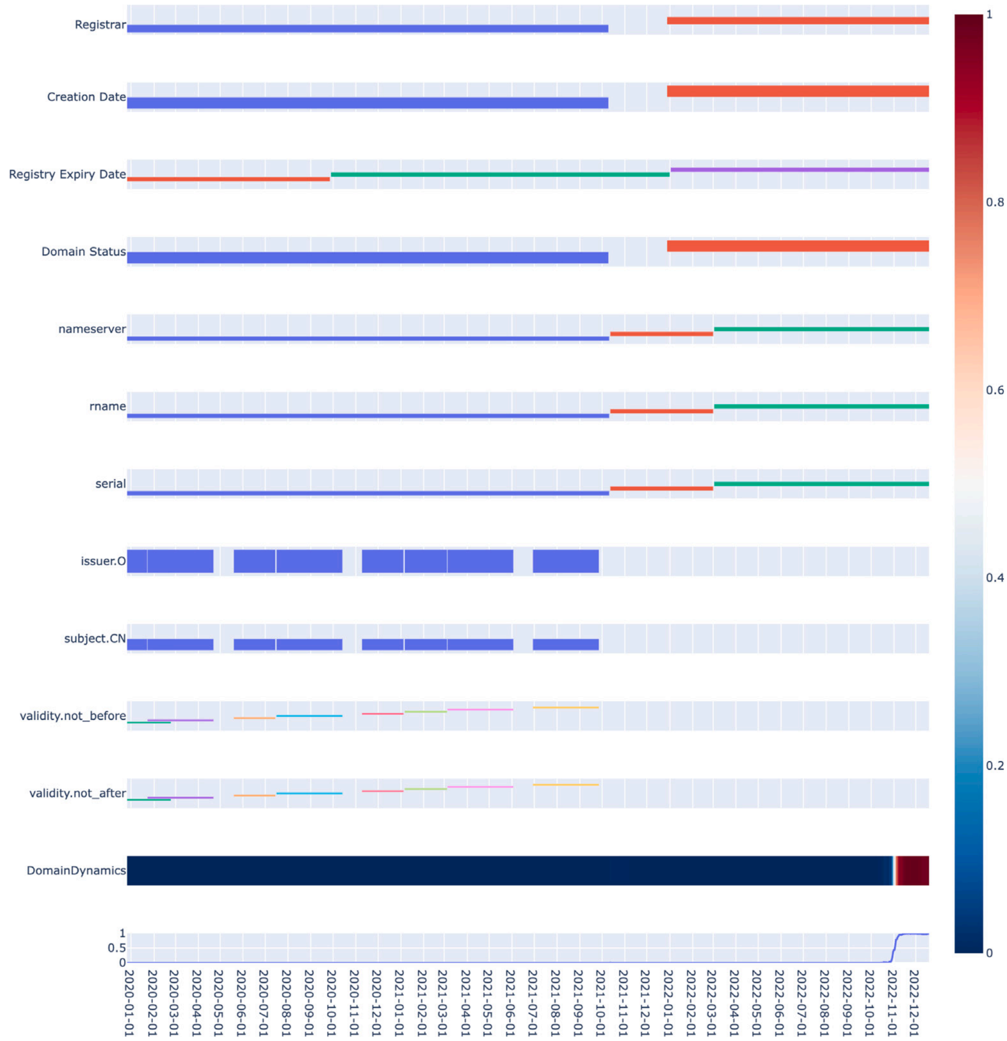


Fig. A.6. Example output from DomainDynamics illustrating the Domain Timeline and Risk Timeline of a domain name that transitioned from legitimate use to malicious activity. The timeline captures the evolution of the domain through changes in WHOIS records, SOA records, and TLS certificates over a span of two years.

accessing domains directly without specifying URL paths, DomainDynamics was shown to preemptively detect a certain number of phishing sites.

4.8. Case study

We discuss some case studies that illustrate actual risk assessments obtained from DomainDynamics.

True Positive Case. A domain name that was previously legitimate but then re-registered (drop-catch) after expiration and subsequently used as a phishing site was correctly identified. Fig. 5(a) shows the Risk Timeline generated by DomainDynamics for the relevant domain name. The vertical red line indicates the first date the domain was used in an attack, and it is clear that a high risk was predicted beforehand. Due to space constraints, full Domain Timeline is presented in Fig. A.7 in Appendix. In fact, this domain name remained dormant for nearly a year after re-registration before being utilized as a phishing site just prior to the expiration of the re-registered domain name. This suggests that the attacker deliberately avoided using the domain to evade detection mechanisms that target recently registered domain names. Upon examining the SHAP output from DomainDynamics, it was evident that the feature No. 2 (Days between WHOIS updates) was a top-ranking feature, accurately capturing the attacker's strategy of targeting the period just before expiration, thus successfully predicting the risk in advance.

True Negative Case. We examine a case involving a C2 domain that served as a connection point for malware. This C2 domain, generated by a DGA, was only active as a C2 for a specific period. Fig. 5(b) displays the Risk Timeline from DomainDynamics for the domain in question. Due to space constraints, full Domain Timeline is shown in Fig. A.8 in Appendix. DomainDynamics' risk prediction for this domain correctly identified it as high risk one day before its use for malicious activity, resulting in a True Positive. When the DGA was no longer active, the risk was assessed as low, resulting in a True Negative. DomainDynamics was capable of predicting risk changes for each requested date without specific knowledge of individual DGAs, thus accurately determining the risk of domain names generated by such algorithms.

False Positive Case. We present a typical False Positive case. A domain name used for targeted advertising was erroneously assessed as high risk on two dates, resulting in False Positives. Fig. 5(c) illustrates the Risk Timeline from DomainDynamics for the domain in question. Due to space constraints, full Domain Timeline is presented in Fig. A.9 in Appendix. The recurrent cause for the incorrect high-risk assessments was the expiration of the domain's TLS certificate without renewal within the validity period, followed by an update the next day, triggering a Positive response due to feature No. 13 concerning TLS characteristics. Specifically, the second occurrence was due to a delay when the issuer O transitioned from Cloudflare to Google. Most

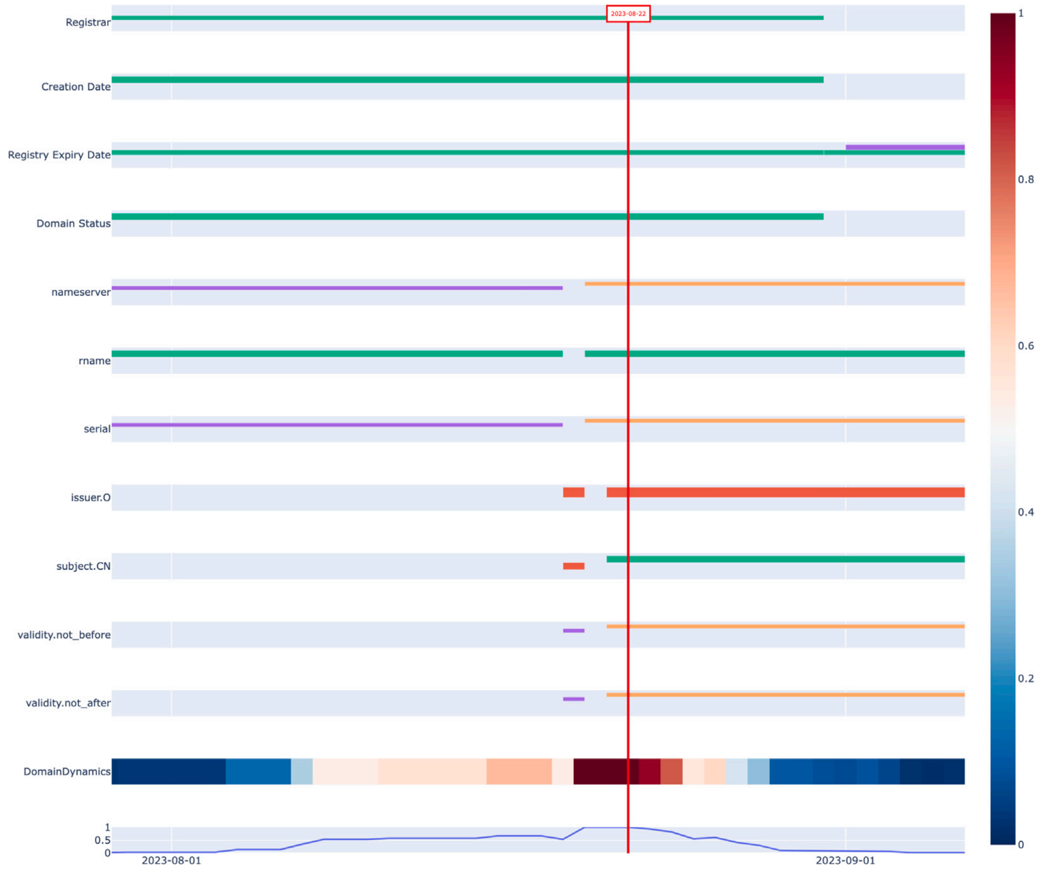


Fig. A.7. Extended Risk Timeline for a True Positive Case: This figure presents a detailed timeline showcasing the risk assessment of a domain name that was initially legitimate, later re-registered, and subsequently used for phishing. The Risk Timeline highlights the period of dormancy after re-registration and demonstrates DomainDynamics' accurate prediction of high risk before the domain was used in an attack.

misclassifications observed in our evaluation (0.69% overall) were attributable to such certificate changes, which skewed DomainDynamics' scores toward the positive. Changes in certificates, such as expiration or issuer transitions, are risk indicators and are incorporated into the ML model based on patterns of malicious domain names, which explains these misclassifications. To reduce these false positives, one could consider excluding TLS certificate changes as a feature, but this would compromise the detection of true positives. We chose not to pursue this approach, as we achieved an acceptably low false positive rate.

5. Discussion

This section discusses the ethical considerations and limitations of this study.

5.1. Ethical consideration

Our study raises no ethical concerns. We did not use any user-related data. In the deployment described in Section 4.7, we utilized Passive DNS data, which resides between cache and authority, containing information solely on the resolved domain names without revealing any data about the users who initiated the name resolutions. Furthermore, our approach to web crawling in Section 4.7 was carefully managed to avoid ethical concerns, including implementing a delay of at least three seconds between accesses to the same server. Consequently, our organization did not view our research as involving human subjects, thus we did not seek IRB approval, in compliance with the ethical considerations outlined in our study.

5.2. Limitation

We acknowledge several limitations of this research, which are outlined below.

Necessity of Historical Data. DomainDynamics relies on a historical database of WHOIS records, SOA records, and TLS certificates to construct a Domain Timeline, as it predicts not just a single point of risk but changes in risk over multiple points in time. Although using commercial security services simplifies the process of accessing large-scale historical databases, for reproducibility purposes, WHOIS records, SOA records, and TLS certificates can be independently verified by anyone, thus reducing concerns regarding the accessibility of data.

Accuracy of WHOIS Records. The implementation of GDPR has resulted in restrictions on the availability of WHOIS records (Lu et al., 2021). However, the information restricted by GDPR pertains to the domain registrant, not the name of the Registrar itself or details such as Creation Date, Expiry Date, and Domain Status, which were utilized in DomainDynamics in the evaluation presented in this paper. Therefore, the GDPR restrictions did not impact the availability of the information necessary for our study.

Lack of Web Content Analysis. While web content analysis is a valuable technique for detecting malicious domains, DomainDynamics takes a different approach by focusing on lifecycle data. We believe that relying on web content has inherent limitations. First, it is reactive rather than proactive, as analyzing content comes at the end of a domain's lifecycle. Second, it makes the detection accuracy dependent on the timing of observation and the distribution of accessible data. By not relying on web content, DomainDynamics avoids these limitations and enables earlier detection, potentially before a domain is actively used for malicious purposes. This approach also enhances scalability

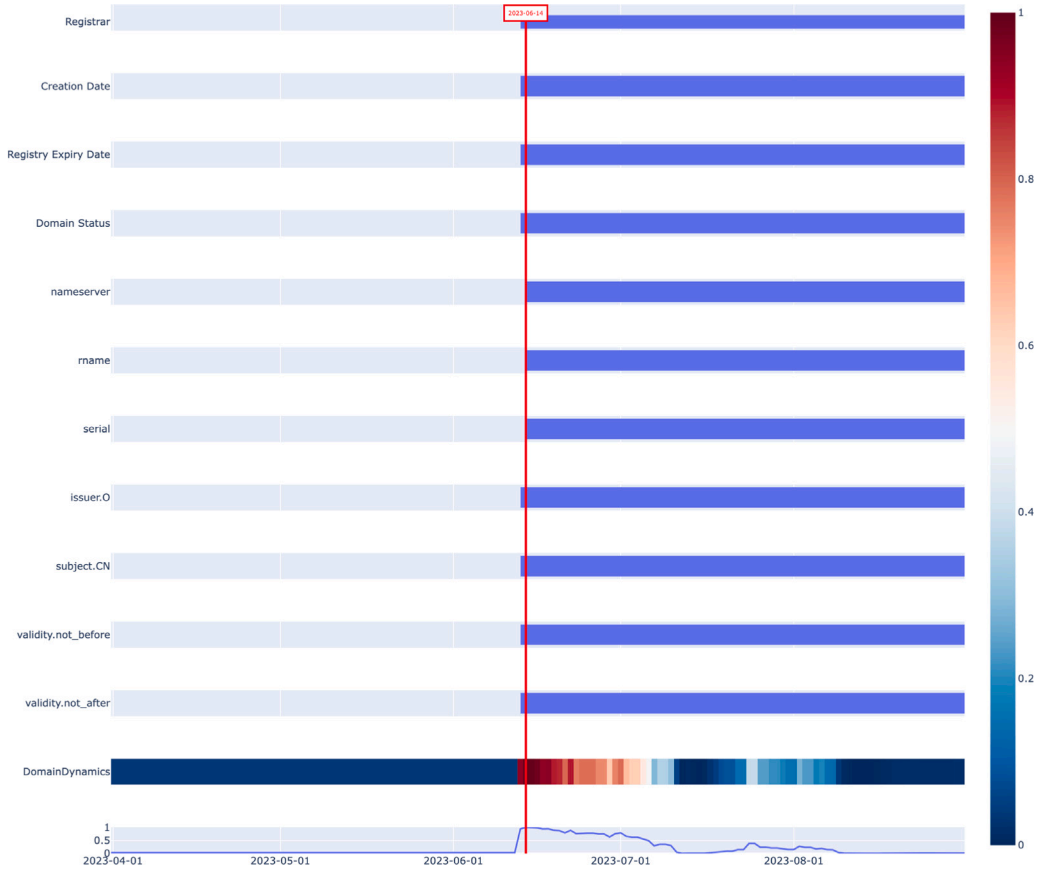


Fig. A.8. Extended Risk Timeline for a True Negative Case: This figure provides an in-depth view of the risk assessment for a Command and Control (C2) domain generated by a Domain Generation Algorithm (DGA). The timeline illustrates DomainDynamics' effective risk prediction, flagging the domain as high risk one day before malicious use and accurately identifying it as low risk when the DGA was inactive, resulting in a true negative.

for large-scale monitoring, as it does not require the resource-intensive process of crawling and analyzing web content for every domain.

Lack of Benign Training Data. Our approach differs fundamentally from studies that aim to perform binary classification between malicious and non-malicious/benign domains. DomainDynamics focuses on identifying when a malicious domain transitions to being actively used for malicious activities. Including benign domains in the training phase could lead to overfitting, where the model learns to identify top-ranked or popular domains rather than those with malicious intent. The low False Positive Rate (FPR) achieved on our test set, even with domains exhibiting unusual but legitimate lifecycle changes (such as changes in certificate providers), demonstrates the practicality and effectiveness of our approach. It shows that we can accurately identify non-malicious domains as non-malicious throughout their lifecycle, even without explicitly training on benign data.

Limitations of Lifecycle-Based Approach. We acknowledge that DomainDynamics may not be able to detect compromised legitimate domains where only the content is modified, without any changes to WHOIS, SOA, or TLS records. For example, if a legitimate domain's Content Management System (CMS) is exploited to alter the website content for malicious purposes, but the domain's registration and DNS records remain unchanged, DomainDynamics might not flag it as high-risk. This limitation highlights that DomainDynamics is designed to complement, not replace, other detection methods that focus on content analysis or behavioral anomalies.

Potential for Evasion Detection. By disclosing the details of DomainDynamics, there is a risk that attackers might devise strategies to avoid leaving traces in historical data, thereby evading detection, or to alter the features that contribute to risk prediction. Nevertheless, in the current Internet environment, it is practically impossible to prepare a domain name for user access without leaving any trace in the historical data of WHOIS records, SOA records, and TLS certificates used by DomainDynamics, without being utilized as features by our system.

Difficulty in Inferring Attack Objectives. While DomainDynamics excels at assessing the risk level of a domain name at specific points in its lifecycle, it does not categorize the nature of the threat (e.g., malware, phishing). This limitation stems from its methodological approach, which focuses on temporal risk assessment rather than detailed threat classification. In practical terms, irrespective of the threat being malware or phishing, the critical factor is the ability to preemptively identify and neutralize harm by either blocking access to high-risk domain names or preemptively taking action against them. Thus, the granular identification of attack types, while informative, is secondary to the primary goal of mitigating risk through lifecycle-based domain analysis.

6. Conclusion

This research presents an expanded study of DomainDynamics, a novel system designed to improve the detection of malicious domain names through a lifecycle-aware approach. Building upon our initial work (Chiba et al., 2025), we have conducted a comprehensive evaluation and enhancement of the system. DomainDynamics constructs timelines for domain names and applies machine learning to analyze characteristics at different lifecycle stages, effectively addressing the common issues of false positives and false negatives in traditional detection methods.

Our expanded evaluation, which included a substantial dataset of over 85,000 known malicious domain names, rigorous comparisons with commercial services, and detailed feature importance analysis, has

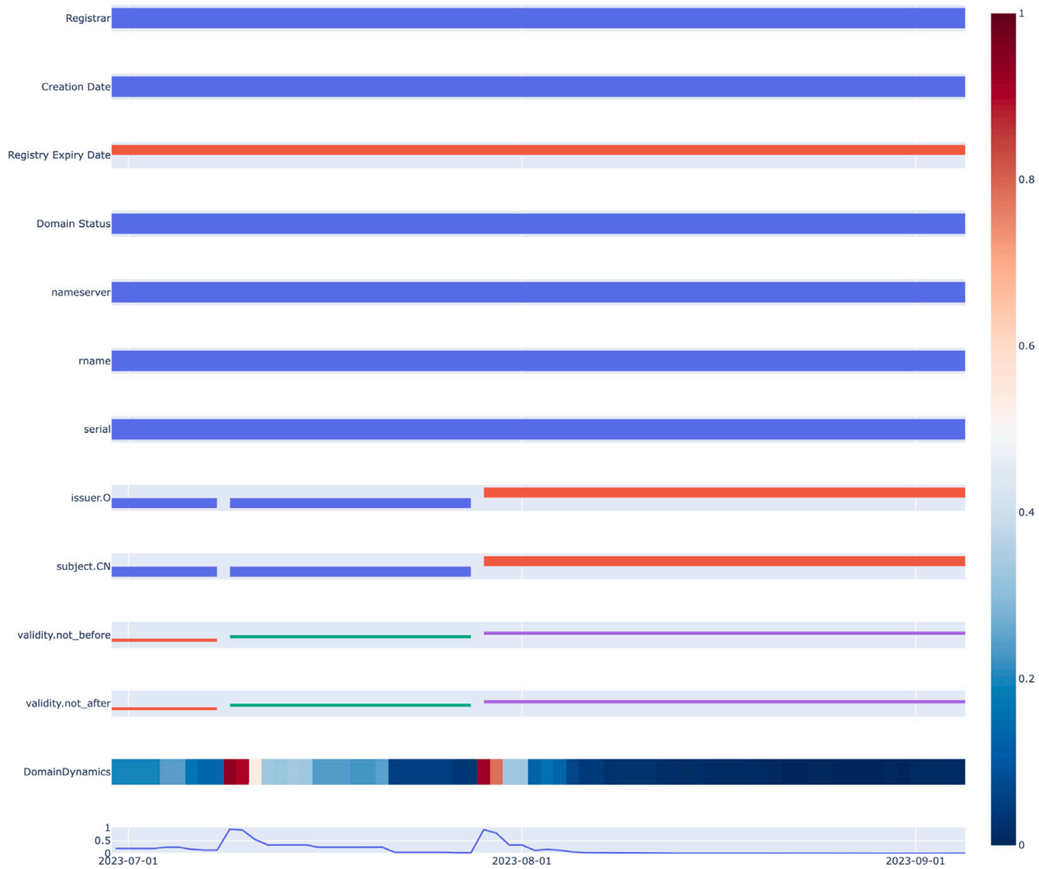


Fig. A.9. Extended Risk Timeline for a False Positive Case: This figure depicts the risk assessment of a domain used for targeted advertising that was mistakenly classified as high risk due to TLS certificate expiration and subsequent renewal. The timeline highlights two instances of incorrect high-risk assessments tied to TLS certificate changes, including a transition between issuers. The figure underscores the challenges in balancing the sensitivity of the ML model to detect true positives while minimizing false positives due to benign certificate changes.

further validated and refined the system's effectiveness. DomainDynamics achieved a detection rate of 82.58% within a seven-day forecast period, while maintaining a low false positive rate of 0.41%. This performance significantly surpasses that of prior studies and commercial services, marking a notable advancement in cybersecurity.

In conclusion, DomainDynamics represents a significant advancement in proactive cybersecurity, offering a robust tool to predict and mitigate cyber threats more effectively. The insights gained from this expanded study have practical implications for improving the operational efficiency of cybersecurity teams and enhancing the overall resilience of digital infrastructures. As cyber threats continue to evolve, the lifecycle-aware approach demonstrated by DomainDynamics provides a promising direction for future research and development in malicious domain detection. Future work will focus on incorporating new data sources, such as advanced DNS features and other threat intelligence feeds, to further enhance the system's accuracy and broaden its applicability. We also plan to explore techniques for addressing emerging threats, including those that exploit new domain registration patterns and those that operate with minimal lifecycle changes. Additionally, we will investigate methods for improving the interpretability of the model's predictions, providing more actionable insights for security analysts.

CRediT authorship contribution statement

Daiki Chiba: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Hiroki Nakano:** Writing – review & editing, Data curation. **Takashi Koide:** Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Supplementary material

Fig. A.6 shows the full output of DomainDynamics which includes Domain Timeline and Risk Timeline. Fig. A.7 shows the full output of DomainDynamics for the True Positive case study. Fig. A.8 shows the full output of DomainDynamics for the True Negative case study. Fig. A.9 shows the full output of DomainDynamics for the False Positive case study.

Data availability

Data will be made available on request.

References

Abdelnabi, S., Krombholz, K., Fritz, M., 2020. Visualphishnet: Zero-day phishing website detection by visual similarity. In: Ligatti, J., Ou, X., Katz, J., Vigna, G. (Eds.), CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13 2020. ACM, pp. 1681–1698. http://dx.doi.org/10.1145/3372297.3417233.

Agten, P., Joosen, W., Piessens, F., Nikiforakis, N., 2015. Seven months' worth of mistakes: A longitudinal study of typosquatting abuse. In: 22nd Annual Network and Distributed System Security Symposium. NDSS 2015, San Diego, California, USA, February 8-11 2015, The Internet Society, URL https://www.ndss-symposium.org/ndss2015/seven-months-worth-mistakes-longitudinal-study-typosquatting-abuse.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Teredesai, A., Kumar, V., Li, Y., Rosales, R., Terzi, E., Karypis, G. (Eds.), Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, KDD 2019, Anchorage, AK, USA, August 4-8 2019, pp. 2623–2631. http://dx.doi.org/10.1145/3292500.3330701.

Almashhadani, A.O., Kaiiali, M., Carlin, D., Sezer, S., 2020. Maldomdetector: A system for detecting algorithmically generated domain names with machine learning. Comput. Secur. 93, 101787. http://dx.doi.org/10.1016/J.COSE.2020.101787.

Anon, 2024. Sectigo, crt.sh - certificate search. URL https://crt.sh.

Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., Feamster, N., 2010. Building a dynamic reputation system for DNS. In: 19th USENIX Security Symposium, Washington, DC, USA, August (2010) 11-13, Proceedings. USENIX Association, pp. 273–290, URL http://www.usenix.org/events/sec10/tech/full_papers/Antonakakis.pdf.

Antonakakis, M., Perdisci, R., Lee, W., Vasiloglou, N., Dagon, D., 2011. Detecting malware domains at the upper DNS hierarchy. In: 20th USENIX Security Symposium, San Francisco, CA, USA, August (2011) 8-12, Proceedings. USENIX Association, URL http://static.usenix.org/events/sec11/tech/full_papers/Antonakakis.pdf.

Antonakakis, M., Perdisci, R., Nadji, Y., Vasiloglou, N., Abu-Nimeh, S., Lee, W., Dagon, D., 2012. From throw-away traffic to bots: Detecting the rise of dga-based malware. In: Kohno, T. (Ed.), Proceedings of the 21th USENIX Security Symposium, Bellevue, WA, USA, August (2012) 8-10. USENIX Association, pp. 491–506, URL https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/antonakakis.

Bijmans, H.L.J., Booij, T.M., Schwedersky, A., Nedgabat, A., van Wegberg, R., 2021. Catching phishers by their bait: Investigating the dutch phishing landscape through phishing kit detection. In: Bailey, M.D., Greenstadt, R. (Eds.), 30th USENIX Security Symposium, USENIX Security 2021, August 11-13 2021. USENIX Association, pp. 3757–3774, URL https://www.usenix.org/conference/usenixsecurity21/presentation/bijmans.

Bilge, L., Kirda, E., Kruegel, C., Balduzzi, M., 2011. EXPOSURE: finding malicious domains using passive DNS analysis. In: Proceedings of the Network and Distributed System Security Symposium, NDSS 2011, San Diego, California, USA, 6th February - 9th 2011. The Internet Society, URL https://www.ndss-symposium.org/ndss2011/exposure-finding-malicious-domains-using-passive-dns-analysis.

Bitaab, M., Cho, H., Oest, A., Lyu, Z., Wang, W., Abraham, J., Wang, R., Bao, T., Shoshitaishvili, Y., Doupé, A., 2023. Beyond phish: Toward detecting fraudulent e-commerce websites at scale. In: 44th IEEE Symposium on Security and Privacy. SP 2023, San Francisco, CA, USA, May 21-25 2023, IEEE, pp. 2566–2583. http://dx.doi.org/10.1109/SP46215.2023.10179461.

Bozkir, A.S., Dalgic, F.C., Aydos, M., 2023. Grambeddings: A new neural network for URL based identification of phishing web pages through n-gram embeddings. Comput. Secur. 124, 102964. http://dx.doi.org/10.1016/J.COSE.2022.102964.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Krishnapuram, B., Shah, M., Smola, A.J., Aggarwal, C.C., Shen, D., Rastogi, R. (Eds.), Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17 2016. ACM, pp. 785–794. http://dx.doi.org/10.1145/2939672.2939785.

Chiba, D., Akiyama, M., Yagi, T., Hato, K., Mori, T., Goto, S., 2018. DomainChroma: Building actionable threat intelligence from malicious domain names. Comput. Secur. 77, 138–161. http://dx.doi.org/10.1016/j.cose.2018.03.013.

Chiba, D., Hasegawa, A.A., Koide, T., Sawabe, Y., Goto, S., Akiyama, M., 2019. DomainScouter: Understanding the risks of deceptive IDNs. In: 22nd International Symposium on Research in Attacks, Intrusions and Defenses. RAID 2019, Chaoyang District, Beijing, China, September 23-25 2019, USENIX Association, pp. 413–426, https://www.usenix.org/conference/raid2019/presentation/chiba.

Chiba, D., Nakano, H., Koide, T., 2025. DomainDynamics: Lifecycle-aware risk timeline construction for domain names. In: 22nd IEEE Consumer Communications & Networking Conference. CCNC 2025, Las Vegas, NV, USA, January 10-13 2025, IEEE.

Chiba, D., Yagi, T., Akiyama, M., Shibahara, T., Yada, T., Mori, T., Goto, S., 2016. DomainProfiler: Discovering domain names abused in future. In: 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. DSN 2016, Toulouse, France, June 28 - July 1 2016, IEEE Computer Society, pp. 491–502. http://dx.doi.org/10.1109/DSN.2016.51.

DomainTools, 2024. Domaintools. URL https://www.domaintools.com.

Drichel, A., Drury, V., von Brandt, J., Meyer, U., 2021. Finding phish in a haystack: A pipeline for phishing classification on certificate transparency logs. In: Reinhardt, D., Müller, T. (Eds.), ARES 2021: The 16th International Conference on Availability, Reliability and Security, Vienna, Austria, August 17-20 2021. ACM, pp. 59:1–59:12. http://dx.doi.org/10.1145/3465481.3470111.

Hao, S., Kantchelian, A., Miller, B., Paxson, V., Feamster, N., 2016. PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration. In: Weippl, E.R., Katzenbeisser, S., Kruegel, C., Myers, A.C., Halevi, S. (Eds.), Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28 2016. ACM, pp. 1568–1579. http://dx.doi.org/10.1145/2976749.2978317.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T., 2017. Lightgbm: A highly efficient gradient boosting decision tree. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9 2017, Long Beach, CA, USA. pp. 3146–3154, URL https://Proceedings.Neurips.Cc/Paper/2017/Hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.Html.

Khalil, I., Yu, T., Guan, B., 2016. Discovering malicious domains through passive DNS data graph analysis. In: Chen, X., Wang, X., Huang, X. (Eds.), Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security. AsiaCCS 2016, Xi'an, China, May 30 - June 3 2016, ACM, pp. 663–674. http://dx.doi.org/10.1145/2897845.2897877.

Kim, T., Park, N., Hong, J., Kim, S., 2022. Phishing URL detection: A network-based approach robust to evasion. In: Yin, H., Stavrou, A., Cremers, C., Shi, E. (Eds.), Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. CCS 2022, Los Angeles, CA, USA, November 7-11 2022, ACM, pp. 1769–1782. http://dx.doi.org/10.1145/3548606.3560615.

Kintis, P., Miramirkhani, N., Lever, C., Chen, Y., Gómez, R.R., Pitropakis, N., Nikiforakis, N., Antonakakis, M., 2017. Hiding in plain sight: A longitudinal study of combosquatting abuse. In: Thuraisingham, B., Evans, D., Malkin, T., Xu, D. (Eds.), Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03 2017. ACM, pp. 569–586. http://dx.doi.org/10.1145/3133956.3134002.

Koide, T., Chiba, D., Akiyama, M., 2020. To get lost is to learn the way: Automatically collecting multi-step social engineering attacks on the web. In: Sun, H., Shieh, S., Gu, G., Ateniese, G. (Eds.), ASIA CCS '20: The 15th ACM Asia Conference on Computer and Communications Security, Taipei, Taiwan, October 5-9 2020. ACM, pp. 394–408. http://dx.doi.org/10.1145/3320269.3384714.

Koide, T., Fukushi, N., Nakano, H., Chiba, D., 2023. PhishReplicant: A language model-based approach to detect generated squatting domain names. In: Annual Computer Security Applications Conference. ACSAC 2023, Austin, TX, USA, December 4-8 2023, ACM, pp. 1–13. http://dx.doi.org/10.1145/3627106.3627111.

Krishnan, S., Taylor, T., Monrose, F., McHugh, J., 2013. Crossing the threshold: Detecting network malfeasance via sequential hypothesis testing. In: 2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks DSN, Budapest, Hungary, June 24-27 2013. IEEE Computer Society, pp. 1–12. http://dx.doi.org/10.1109/DSN.2013.6575364.

Lee, S., Kim, J., 2012. Warningbird: detecting suspicious urls in twitter stream. In: 19th Annual Network and Distributed System Security Symposium. NDSS 2012, San Diego, California, USA, February 5-8 2012, The Internet Society, URL https://www.ndss-symposium.org/ndss2012/warningbird-detecting-suspicious-urls-twitter-stream.

Liang, J., Chen, S., Wei, Z., Zhao, S., Zhao, W., 2022. Hagdetector: Heterogeneous DGA domain name detection model. Comput. Secur. 120, 102803. http://dx.doi.org/10.1016/J.COSE.2022.102803.

Lin, Y., Liu, R., Divakaran, D.M., Ng, J.Y., Chan, Q.Z., Lu, Y., Si, Y., Zhang, F., Dong, J.S., 2021. Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages. In: Bailey, M.D., Greenstadt, R. (Eds.), 30th USENIX Security Symposium, USENIX Security 2021, August 11-13 2021. USENIX Association, pp. 3793–3810, URL https://www.usenix.org/conference/usenixsecurity21/presentation/lin.

Liu, D., Li, Z., Du, K., Wang, H., Liu, B., Duan, H., 2017. Don't let one rotten apple spoil the whole barrel: towards automated detection of shadowed domains. In: Thuraisingham, B., Evans, D., Malkin, T., Xu, D. (Eds.), Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. CCS 2017, Dallas, TX, USA, October 30 - November 03 2017, ACM, pp. 537–552. http://dx.doi.org/10.1145/3133956.3134049.

Liu, R., Lin, Y., Yang, X., Ng, S.H., Divakaran, D.M., Dong, J.S., 2022. Inferring phishing intention via webpage appearance and dynamics: A deep vision based approach. In: Butler, K.R.B., Thomas, K. (Eds.), 31st USENIX Security Symposium, USENIX Security 2022, Boston, MA, USA, August 10-12 2022. USENIX Association, pp. 1633–1650, URL https://www.usenix.org/conference/usenixsecurity22/presentation/liu-ruofan.

Liu, B., Lu, C., Li, Z., Liu, Y., Duan, H., Hao, S., Zhang, Z., 2018. A reexamination of internationalized domain names: The good, the bad and the ugly. In: 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. DSN 2018, Luxembourg City, Luxembourg, June 25-28 2018, IEEE Computer Society, pp. 654–665. http://dx.doi.org/10.1109/DSN.2018.00072.

Lu, C., Liu, B., Zhang, Y., Li, Z., Zhang, F., Duan, H., Liu, Y., Chen, J.Q., Liang, J., Zhang, Z., Hao, S., Yang, M., 2021. From WHOIS to WHOWAS: A large-scale measurement study of domain registration privacy under the GDPR. In: 28th Annual Network and Distributed System Security Symposium. NDSS 2021, Virtually, February 21–25 2021, The Internet Society, URL https://www.ndss-symposium.org/ndss-paper/from-whois-to-whowas-a-large-scale-measurement-study-of-domain-registration-privacy-under-the-gdpr/.

Lundberg, S.M., Erion, G.G., Chen, H., DeGrave, A.J., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67. http://dx.doi.org/10.1038/S42256-019-0138-9.

Lundberg, S.M., Lee, S., 2017. A unified approach to interpreting model predictions. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9 2017, Long Beach, CA, USA. pp. 4765–4774, URL https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Manadhata, P.K., Yadav, S., Rao, P., Horne, W.G., 2014. Detecting malicious domains via graph inference. In: Kutylowski, M., Vaidya, J. (Eds.), Computer Security - ESORICS 2014-19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11 2014. Proceedings, Part I. In: Lecture Notes in Computer Science, vol. 8712, Springer, pp. 1–18. http://dx.doi.org/10.1007/978-3-319-11203-9_1.

Marchal, S., François, J., State, R., Engel, T., 2012. Proactive discovery of phishing related domain names. In: Balzarotti, D., Stolfo, S.J., Cova, M. (Eds.), Research in Attacks, Intrusions, and Defenses - 15th International Symposium, RAID 2012, Amsterdam, the Netherlands, September (2012) 12-14. Proceedings. In: Lecture Notes in Computer Science, vol. 7462, Springer, pp. 190–209. http://dx.doi.org/10.1007/978-3-642-33338-5_10.

Maroofi, S., Korczynski, M., Hesselman, C., Ampeau, B., Duda, A., 2020. COMAR: classification of compromised versus maliciously registered domains. In: IEEE European Symposium on Security and Privacy. EuroS & P 2020, Genoa, Italy, September 7-11 2020, IEEE, pp. 607–623. http://dx.doi.org/10.1109/EUROSP48549.2020.00045.

Nabeel, M., Khalil, I.M., Guan, B., Yu, T., 2020. Following passive DNS traces to detect stealthy malicious domains via graph inference. ACM Trans. Priv. Secur. 23 (4), 17:1–17:36. http://dx.doi.org/10.1145/3401897.

Najafi, P., Mühle, A., Pünter, W., Cheng, F., Meinel, C., 2019. Malrank: a measure of maliciousness in siem-based knowledge graphs. In: Balenson, D. (Ed.), Proceedings of the 35th Annual Computer Security Applications Conference. ACSAC 2019, San Juan, PR, USA, December 09-13 2019, ACM, pp. 417–429. http://dx.doi.org/10.1145/3359789.3359791.

Nakano, H., Chiba, D., Koide, T., Fukushi, N., Yagi, T., Hariu, T., Yoshioka, K., Matsumoto, T., 2023. Canary in Twitter mine: Collecting phishing reports from experts and non-experts. In: Proceedings of the 18th International Conference on Availability, Reliability and Security. ARES 2023, Benevento, Italy, 29 2023-1 2023, ACM, pp. 6:1–6:12. http://dx.doi.org/10.1145/3600160.3600163.

OpenPhish, 2024. OpenPhish. URL https://openphish.com/.

Oprea, A., Li, Z., Norris, R., Bowers, K.D., 2018. MADE: security analytics for enterprise threat detection. In: Proceedings of the 34th Annual Computer Security Applications Conference. ACSAC 2018, San Juan, PR, USA, December 03-07 2018, ACM, pp. 124–136. http://dx.doi.org/10.1145/3274694.3274710.

Oprea, A., Li, Z., Yen, T., Chin, S.H., Alrwais, S.A., 2015. Detection of early-stage enterprise infection by mining large-scale log data. In: 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. DSN 2015, Rio de Janeiro, Brazil, June 22-25 2015, IEEE Computer Society, pp. 45–56. http://dx.doi.org/10.1109/DSN.2015.14.

Perdisci, R., Corona, I., Giacinto, G., 2012. Early detection of malicious flux networks via large-scale passive DNS traffic analysis. IEEE Trans. Dependable Secur. Comput. 9 (5), 714–726. http://dx.doi.org/10.1109/TDSC.2012.35.

Pochat, V.L., van Goethem, T., Tajalizadehkhoob, S., Korczynski, M., Joosen, W., 2019. Tranco: A research-oriented top sites ranking hardened against manipulation. In: 26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February (2019) 24-27. URL https://www.ndss-symposium.org/ndss-paper/tranco-a-research-oriented-top-sites-ranking-hardened-against-manipulation/.

Quinkert, F., Lauinger, T., Robertson, W.K., Kirda, E., Holz, T., 2019. It's not what it looks like: Measuring attacks and defensive registrations of homograph domains. In: 7th IEEE Conference on Communications and Network Security. CNS 2019, Washington, DC, USA, June (2019) 10-12, IEEE, pp. 259–267. http://dx.doi.org/10.1109/CNS.2019.8802671.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106. http://dx.doi.org/10.1023/A:1022643204877.

Rahbarinia, B., Perdisci, R., Antonakakis, M., 2015. Segugio: efficient behavior-based tracking of malware-control domains in large ISP networks. In: 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. DSN 2015, Rio de Janeiro, Brazil, June 22-25 2015, IEEE Computer Society, pp. 403–414. http://dx.doi.org/10.1109/DSN.2015.35.

Roberts, R., Goldschlag, Y., Walter, R., Chung, T., Mislove, A., Levin, D., 2019. You are who you appear to be: A longitudinal study of domain impersonation in TLS certificates. In: Cavallaro, L., Kinder, J., Wang, X., Katz, J. (Eds.), Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. CCS 2019, London, UK, November 11-15 2019, ACM, pp. 2489–2504. http://dx.doi.org/10.1145/3319535.3363188.

Sabah, M.A., Nabeel, M., Boshmaf, Y., Choo, E., 2022. Content-agnostic detection of phishing domains using certificate transparency and passive DNS. In: 25th International Symposium on Research in Attacks, Intrusions and Defenses. RAID 2022, Limassol, Cyprus, October 26-28 2022, ACM, pp. 446–459. http://dx.doi.org/10.1145/3545948.3545958.

Schiavoni, S., Maggi, F., Cavallaro, L., Zanero, S., 2014. Phoenix: dga-based botnet tracking and intelligence. In: Dietrich, S. (Ed.), Detection of Intrusions and Malware, and Vulnerability Assessment - 11th International Conference, DIMVA 2014, Egham, UK, July 10-11 2014. Proceedings. In: Lecture Notes in Computer Science, vol. 8550, Springer, pp. 192–211. http://dx.doi.org/10.1007/978-3-319-08509-8_11.

Schüppen, S., Teubert, D., Herrmann, P., Meyer, U., 2018. FANCI : Feature-based automated nxdomain classification and intelligence. In: Enck, W., Felt, A.P. (Eds.), 27th USENIX Security Symposium, USENIX Security 2018, Baltimore, MD, USA, August (2018) 15-17. USENIX Association, pp. 1165–1181, URL https://www.usenix.org/conference/usenixsecurity18/presentation/schuppen.

SecurityTrails, 2024. Securitytrails. URL https://securitytrails.com.

Silva, R.D., Nabeel, M., Elvitigala, C., Khalil, I., Yu, T., Keppitiyagama, C., 2021. Compromised or attacker-owned: A large scale classification and study of hosting domains of malicious urls. In: Bailey, M., Greenstadt, R. (Eds.), 30th USENIX Security Symposium, USENIX Security 2021, August 11-13 2021. USENIX Association, pp. 3721–3738, URL https://www.usenix.org/conference/usenixsecurity21/presentation/desilva.

Spooren, J., Vissers, T., Janssen, P., Joosen, W., Desmet, L., 2019. Premadoma: an operational solution for DNS registries to prevent malicious domain registrations. In: Balenson, D. (Ed.), Proceedings of the 35th Annual Computer Security Applications Conference. ACSAC 2019, San Juan, PR, USA, December 09-13 2019, ACM, pp. 557–567. http://dx.doi.org/10.1145/3359789.3359836.

Sun, X., Tong, M., Yang, J., Liu, X., Liu, H., 2019. Hindom: A robust malicious domain detection system based on heterogeneous information network with transductive classification. In: 22nd International Symposium on Research in Attacks, Intrusions and Defenses. RAID 2019, Chaoyang District, Beijing, China, September 23-25 2019, USENIX Association, pp. 399–412, URL https://www.usenix.org/conference/raid2019/presentation/sun.

Suzuki, H., Chiba, D., Yoneya, Y., Mori, T., Goto, S., 2019. ShamFinder: An automated framework for detecting IDN homographs. In: Proceedings of the Internet Measurement Conference. IMC 2019, Amsterdam, the Netherlands, October 21-23 2019, ACM, pp. 449–462. http://dx.doi.org/10.1145/3355369.3355587.

Szurdi, J., Kocso, B., Cseh, G., Spring, J., Félegyházi, M., Kanich, C., 2014. The long "taile" of typosquatting domain names. In: Fu, K., Jung, J. (Eds.), Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22 2014. USENIX Association, pp. 191–206, URL https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/szurdi.

Tang, S., Mi, X., Li, Y., Wang, X., Chen, K., 2022. Clues in tweets: Twitter-guided discovery and analysis of SMS spam. In: Yin, H., Stavrou, A., Cremers, C., Shi, E. (Eds.), Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. CCS 2022, Los Angeles, CA, USA, November 7-11 2022, ACM, pp. 2751–2764. http://dx.doi.org/10.1145/3548606.3559351.

Tian, K., Jan, S.T.K., Hu, H., Yao, D., Wang, G., 2018. Needle in a haystack: Tracking down elite phishing domains in the wild. In: Proceedings of the Internet Measurement Conference 2018. IMC 2018, Boston, MA, USA, October 31 - November 02 2018, ACM, pp. 429–442, URL https://dl.acm.org/citation.cfm?id=3278569.

VirusTotal, 2024a. Virustotal. URL https://www.virustotal.com/.

VirusTotal, 2024b. Why don't you have statistics comparing antivirus performance? URL https://docs.virustotal.com/docs/antivirus-stats.

Yadav, S., Reddy, A.K.K., Reddy, A.L.N., Ranjan, S., 2010. Detecting algorithmically generated malicious domain names. In: Allman, M. (Ed.), Proceedings of the 10th ACM SIGCOMM Internet Measurement Conference. IMC 2010, Melbourne, Australia - November 1-3 2010, ACM, pp. 48–61. http://dx.doi.org/10.1145/1879141.1879148.

Zhao, H., Chang, Z., Wang, W., Zeng, X., 2019. Malicious domain names detection algorithm based on lexical analysis and feature quantification. IEEE Access 7, 128990–128999. http://dx.doi.org/10.1109/ACCESS.2019.2940554.

Zonefiles, 2024. Zonefiles. URL https://zonefiles.io/.