DEFENDING AGAINST MALICIOUS WEBSITES: THEMED THREATS, DETECTION, AND LAW-ENFORCEMENT

by

MIR MEHEDI AHSAN PRITOM, M.S.

DOCTORAL DISSERTATION
Presented to the Graduate Faculty of
The University of Texas at San Antonio
In Partial Fulfillment
Of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

COMMITTEE MEMBERS:
Ravi Sandhu, Ph.D., Co-Chair
Shouhuai Xu, Ph.D., Co-Chair
Xiaoyin Wang, Ph.D.
Greg B. White, Ph.D.
Mimi Xie, Ph.D.

THE UNIVERSITY OF TEXAS AT SAN ANTONIO
College of Sciences
Department of Computer Science
August 2022

DEDICATION

First, thanks to the Almighty Allah (SWT) for giving me this opportunity to go through this challenging yet beautiful journey and experience. I want to dedicate this work to my family, who brought me up and helped me to become the person I am today. Especially, I would like to thank my dearest mother, who taught me how to be resilient and fight back in life even when I have limited resources. Next, I would like to thank my beloved wife, who always kept pushing and encouraging me throughout this journey. Additionally, I would like to dedicate this work to my "to be" firstborn, who has brought blessings into our lives even before his landing on planet earth. Lastly, I know my father would be proud to see this from a better place.

ACKNOWLEDGEMENTS

I would like to thank my co-advisor Dr. Shouhuai Xu for not only being my supervisor but also a mentor to lead me in this journey. Dr. Xu has guided me to develop myself as an independent researcher through his constant thoughtful feedback. I would like to thank my co-advisor Dr. Ravi Sandhu without whose guidance and support it would have been a tough call to stay at UT San Antonio and finish my Doctoral degree on time. Next, I would like to earnestly thank all of my committee members: Dr. Gregory White, Dr. Xiaoyin Wang, and Dr. Mimi Xie for their immense support and counsel. Moreover, I would like to thank Dr. Raymond M. Bateman from US ARL-South Cyber for his constant direction, guidance, and encouragements. Additionally, I am deeply grateful to the Computer Science Department for supporting me financially and giving me the lecturing opportunity to develop my teaching as well as soft skills. Thus, after this wonderful 3 and a half years at UTSA, I can say I am a proud Roadrunner.

Additionally, I would like to thank Dr. Abdur Razzaque, my undergraduate thesis supervisor and first mentor, who always believed in me and planted the seeds for pursuing this degree. Besides, I would like to thank Dr. Moinul Hossain (Sohan), an inspiring person, a staunch friend, for his efforts to motivate me to thrive in my professional research career and providing timely advice. Furthermore, I would like to thank my high school teachers, college lecturers, private tutors, colleagues, co-authors, and seniors as it would not have been possible to come this far without their guidance and assistance. Moreover, during this doctoral expedition, I have made some great friends, whom I can not thank enough for always being so supportive, which eventually helped me stay fit mentally and physically. I want to remember my dear friend, Imtiaz Ikram Alis, whom I always admired for his positive energy, but lost him in a tragic road accident in 2017. To sum up, I realize this journey made me the better person that I am today, for which I am forever indebted to everyone involved.

This research was supported in part by NSF Grant #2122631 (#1814825) and ARO Grant #W911NF-17-1-0566.

"This Doctoral Dissertation was produced in accordance with guidelines which permit the inclusion as part of the Doctoral Dissertation the text of original papers, submitted for publication. The Doctoral Dissertation must still conform to all other requirements explained in the "Guide for the Preparation of a Doctoral Dissertation at The University of Texas at San Antonio." It must include a comprehensive abstract, a full introduction and literature review, and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported. It is acceptable for this Doctoral Dissertation to include as chapters authentic copies of papers already published, provided these meet type size, margin, and legibility requirements. In such cases, connecting texts, which provide logical bridges between different manuscripts, are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student's contribution to the work and acknowledging the contribution of the other author(s). The approvals of the Supervising Committee which precede all other material in the Doctoral Dissertation attest to the accuracy of this statement."

August 2022

DEFENDING AGAINST MALICIOUS WEBSITES: THEMED THREATS, DETECTION, AND LAW-ENFORCEMENT

Mir Mehedi Ahsan Pritom, M.S.
The University of Texas at San Antonio, 2022

Supervising Professors: Ravi Sandhu, Ph.D. and Shouhuai Xu, Ph.D.

Malicious websites have become a main cyber threat. Despite the substantial effort made by researchers and practitioners, some fundamental problems regarding effective defenses against these attacks remain open, such as: What are the emerging trends of malicious websites? How should we cope with the new trends? What do we need to do to help law-enforcement deal with them? This dissertation addresses these problems by making three contributions. The first contribution is to characterize the emerging themed threats including themed malicious websites, which represent one trend as evidenced by the many malicious websites exploiting the COVID incident. The characterization offers a deep understanding of the attacks, which leads to the second contribution, namely the investigation of how to detect the emerging themed malicious websites exploiting the COVID-19 incident. While the preceding two contributions are from a purely technological point of view, the third contribution investigates the gap between technology and law-enforcement with respect to malicious websites. Understanding and addressing the gap is essential because we anticipate that the law-enforcement eventually needs to be involved in dealing with malicious websites, if not already. For this purpose, we focus on investigating how to support the law-enforcement dealing with blacklisted websites while highlighting two important factors: one is the trustworthiness of Machine Learning methods in predicting malicious websites, which is important because blacklists are not perfect; and the other is how to interpret or explain the individual predictions (possibly in the court), which existing black-box ML models can not provide. The resulting methodology could be leveraged to cope with malicious websites towards ultimately eliminating, or at least ad-

equately mitigating, them. Finally, we present case studies with real-world datasets to show the usability and efficacy of the proposed methods.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Dissertation Motivation

Cyberattacks exploiting websites as a medium have become a persistent problem, triggering the need to cope with malicious websites. There are many proposed systems and methods to detect and tackle malicious websites. However, there is a never ending arms race between the attackers and the defenders, namely that attackers attempt to find new evasive techniques and tactics to avoid detection and takedown by defenders, whereas defenders aim to detect and cope with increasingly more sophisticated malicious websites. As a consequence, it has become increasingly challenging to design effective defenses against the increasingly sophisticated attacks that attempt to exploit their weaknesses [209].

In the past decade, Machine Learning (ML) and Deep Learning (DL) based models became very popular in many contexts of cybersecurity as computers can process significantly faster than what was offered 20 years ago [27]. Moreover, data-driven approaches, also known as *data analytics*, has been very popular as big data handling and analysis became a reality [27]. Even from the point of view of malicious website detection and classification, ML and DL based models are often proposed to cope with the problem of malicious website detection (see, for example, [9, 10, 32, 40, 44, 57, 76, 133, 165, 185, 196, 197, 220–222, 244]); more discussion on related prior studies will be presented in the respective technical chapters.

Despite the many studies, we encounter the issue that the existing methods have not adequately addressed the characterization and detection of emerging threats of themed malicious websites, which have seemingly become a trend and can possibly impose a bigger threat [5]. This motivates us to conduct a systematic study on this emerging perspective of malicious websites. Specifically, we propose characterizing the landscape of themed malicious websites and develop methodologies to detect them with supervised machine learning model [6].

Moreover, we observe that traditional ML based approaches are usually black-box models and hard to explain why they make certain classification [142, 186]. This means that these models are

not adequate to support decision-making in the real world because they force decision-makers to trust their classifications. Indeed, the use of ML based black-box models have become questionable [103] because of the trust and transparency issues have significantly reduced the adoption of ML and DL based methods in practice. The state-of-the-art is that there is very little knowledge on how well these models would be when applied to cope with websites in the wild; for example, adversarial examples of malicious websites [220] can easily evade advanced models that aim to detect malicious websites. This issue motivates us to adapt ML interpretability (also known as explinability) and probabilistic prediction to the context of detecting malicious websites, as a first step towards quantifying the trust one can put on the classifications made by ML models.

This dissertation further introduces a new dimension of the problem of defense against malicious websites, namely the law-enforcement perspective. It aims to help the law-enforcement in coping with malicious blacklisted websites, which may be given by third-parties or by ML based models. However, these given malicious websites may not be truly malicious because the third-parties and ML models may not be trustworthy. This perspective is an important matter because the law-enforcement needs to make reliable decisions in coping with these websites; for example, a false-positive may do more damage than good (e.g., the law-enforcement may get sued by wrongly take down a legitimate website or disrupt regular business of legitimate websites). To deal with the problem, we propose using probabilistic classification and interpretability to enhance the trustworthiness of the classifications made by ML models [143].

1.2 Dissertation Aims and Objectives

The aims of the dissertation are in three-fold:

- Conducting research to protect web users from emerging themed cyberattacks and malicious websites.
- Investigating effective and efficient defenses against themed malicious websites by actively detecting them.

- Studying technical means to support cyber defenders (e.g., blacklist authorities, law-enforcement, domain hosting companies) with effective defense decision-making tools incorporating trust and transparency for handling malicious websites.

The main objectives of the dissertation are to devise defense tools against malicious websites, including:

- Characterizing the landscape of themed attacks and malicious websites exploiting COVID-19 as themes.
- Detecting themed malicious websites exploiting COVID-19 related themes those are registered in large volumes to amplify attack effects.
- Proposing a framework to support law-enforcement decision-making to take effective actions against truly malicious attacker-owned websites given by some blacklists.

1.3 Dissertation Contributions

The main contributions of this dissertation are summarized as follows:

- The dissertation provides a characterization of the landscape of cyberattacks exploiting themed malicious websites and their level of sophistication in terms of **t**echniques, **t**actics, and **p**rocedures (TTP) by conducting a retrospective analysis. This research has been published as [5].
- The dissertation proposes a novel methodology for detecting emerging malicious websites which are themed with COVID-19 pandemic incident. It also provides a real-world case study on the characterization and data-driven detection of COVID-19-themed malicious websites. However, the proposed methodology can also be adapted for other X-themed malicious websites. This research has been published as [6].
- The dissertation proposes a novel framework for supporting law-enforcement decision-making while dealing with blacklisted websites. It demonstrates how to distinguish the truly mali-

cious (e.g., attacker-owned or attacker-operated) websites, which should be taken down, from the websites which are owned or operated by legitimate users but have been compromised and then abused to wage attacks, meaning that these websites' owner or operators should be notified for cleaning up rather than taking them down. The framework leverages interpretable ML models to provide probabilistic predictive classification and local interpretation for individual prediction outcome (e.g., why a particular website is classified as truly malicious rather than compromised or benign). This enhances the overall trust and transparency for effective decision-making by the law-enforcement. This research has been submitted for review as [179].

1.4 Dissertation Organization

The rest of this dissertation is organized as follows: Chapter 2 presents the characterization of the landscape of cyberattacks exploiting themed malicious websites and other themed threats and a brief introduction on the available defense strategies.

Chapter 3 investigates ML based detection system for characterizing and detecting newly emerged COVID-19 themed malicious websites in the wild. During the evolution of COVID-19 pandemic, the attackers have exploited the opportunity because of the elevated uncertainty and confusion on web users and the increase in remote working tendency. For example, there are a large volume of new websites that were registered with a theme aligning with some form of services or products related to COVID-19.

Chapter 4 presents a framework to support the law-enforcement in decision-making when dealing with blacklisted websites. This framework provides answer to the question "what to do?" when the law-enforcement is provided with blacklisted URLs as input. The final outcome of this framework incorporates probabilistic class prediction and explanation (i.e., interpretability) on why a ML model predicts a website as truly malicious vs. compromised but legitimate.

Chapter 5 concludes the dissertation with a concise discussion on the limitations of the presented ideas and their results. This chapter also highlights the many areas that need to be further

investigated from a broader context of cybersecurity. The importance of the problem justifies that these directions should be further and systematically investigated. Thus, the hope is that the dissertations will inspire many more studies in the future.

CHAPTER 2: CHARACTERIZING THE LANDSCAPE OF EMERGING THEMED THREATS

Themed cyberattacks are an emerging new trend of attacks that use a specific event to target a group of victim users. Especially, during the COVID-19 pandemic we observe a rise in this type of themed cyberattacks mostly leveraging various social engineering tactics to exploit user trust with various COVID-19 related themes [11, 205]. In particular, work-from-home has become a new norm for employees during the pandemic, which enhanced the attack surface significantly and created a plenty of opportunities for the attackers. Despite the fact that COVID-19 pandemic can equally impact innocent people and cyber criminals, it is ironic to see surges in cyberattacks leveraging COVID-19 as a theme, dubbed *COVID-19 themed cyberattacks* or *COVID-19 attacks* for short, which represent a new emerging threat and has got the attention after the beginning of the pandemic [105]. In this chapter, we make one of the earliest step towards fully characterizing the landscape of these attacks along with themed malicious websites [6,217] and the sophistication level of these attacks via the Cyber Kill Chain model. We also briefly explore the solution space of defenses against these attacks.

2.1 Chapter Introduction

The COVID-19 (Coronavirus) pandemic has had a huge impact on the global society and economy. It attacks everyone, including both the innocent people and the cyber criminals. Ironically, we have witnessed surges in cyberattacks leveraging COVID-19 as a theme, dubbed *COVID-19 themed cyberattacks* or *COVID-19 attacks* for short. We observe that global events such as the COVID-19 pandemic has created scopes for attackers to victimize users capitalize on the situation with these emerging themed threats when tensions, fear, and confusions are all around society [23,136]. Previously we have seen a rise in attacks like spear-phishing, malware down-loaders, scamming threats, impersonating official representatives, credit card scamming, dating scams, telephony scams, and frauds during holiday seasons or tax seasons [56, 109, 166, 199]. But, due to the unprecedented

volume of work-from-home (WFH) policies, COVID-19 has extend the attack surface and higher ROI (return on investments) opportunities for cyberattackers. For example, there is a 32 times increase in the malware and phishing websites from February 25, 2020 to March 25, 2020 [190]; Google has been reported to be blocking around 240 million COVID-19 related spam emails and 18 million phishing and malware emails daily [88] during that time; there is a 148% increase in ransomware attacks in March 2020 over February 2020 [168]. Moreover, different APT groups are eyeing to spread data exfiltration malware through emails, SMS, and social media links. Most attackers mainly aim to gain financial benefits and gather credentials or sensitive personal information from the victims, which they can later use for get-in enterprise networks or sell on the *DarkWeb* [85]. Additionally, we see a surge in Zoom-bombing attacks during COVID-19 which leverage vulnerabilities in remote meeting software used for remote work at a unprecedented volume [24]. The situation is further exacerbated since home computers or devices are often less protected than their enterprise counterparts. Indeed, a CheckPoint survey [175] has reports that 55% of security professionals are concerned with remote access and 47% are concerned with their employees using shadow IT systems from their home. In the beginning, COVID-19 attackers have mainly targeted the finance, healthcare, government, media streaming, retail business, and COVID-19 research sectors. In response, experts have recommended using multi-factor authentication for critical transactions, virtual private networks for remote access, and regularly patching and updating software as immediate solutions [141]. In brief, we find COVID-19 attacks are an emerging threat and a new phenomenon that is here to stay longer and possibly be imitated in future events. Hence, it is important to understand them thoroughly to pave a way for effective defense.

Chapter Contributions. In this chapter, we make a first step towards understanding COVID-19 themed cyberattacks. Specifically, we explore five classes of these attacks, namely themed malicious websites, themed malicious emails, themed malicious mobile apps, themed malicious messaging, and themed misinformation. In order to characterize these attacks, we map them to the Cyber Kill Chain model [90]. We show that they can use multiple attack techniques to achieve multiple attack goals. We find that COVID-19 attackers have been professional rather than oppor-

tunistic and have been heavily employing various social-engineering attack techniques. We further explore the solution space of defenses against COVID-19 attacks. Since COVID-19 attacks do have their counterparts that are not specific to the COVID-19 incidents, our focus is on exploring the COVID-19 specific aspects. To the best of our knowledge, this is the first systematic characterization of COVID-19 attacks and defenses, which can be adapted to cope with any "X-themed cyberattacks" that may emerge in the future, where X can be any kind of social incidents (e.g., election, natural or man-made disaster, war).

Chapter Outline. Section 2.2 characterizes COVID-19 attacks along with complexity of attacks using themed malicious websites. Section 2.3 explores the defense solution space. Section 2.4 discusses the closely related works. Finally, section 2.5 concludes the chapter.

2.2 Characterizing COVID-19 Attacks

We characterize 5 classes of COVID-19 attacks: malicious websites, malicious emails, malicious mobile apps, malicious messaging, and misinformation. For this purpose, we collect existing news reports and blogs on relevant cyberattacks, manually verify them, and propose mapping them to the Lockheed Martin's Cyber Kill Chain [90], which is a model consisting of the following 7 stages as shown in figure 2.1. (i) *Reconnaissance* (Recon), which corresponds to pre-attack plannings, finding vulnerabilities, collecting possible victims, and setting attack goals. (ii) *Weaponization*, which corresponds to setting up attack propagation mediums, injecting malicious contents into the mediums, and setting up traps to fool the identified victims. (iii) *Delivery*, which corresponds to the attacker's penetration into a victim's system through some entry point. (iv) *Exploitation*, which corresponds to the wage of actual attacks against a victim's system. (v) *Installation*, which corresponds to the installation of malicious payloads on a victim's system. (vi) *Command-and-Control* (C2), which corresponds to the attacker's use of remote access to the victims' systems. (vii) *Objectives*, which corresponds to the accomplishment of the attacker's pre-determined goal.

COVID-19 themed attacks have mainly targeted the *finance*, *healthcare*, *government*, *media*

Figure 2.1: Stages of the Cyber Kill Chain model [127]

streaming, *retail business*, and *COVID-19 research* sectors. Often times these sectors become a target partly because the employees switch to the practice of work from home (WFH). This not only makes the end points (e.g., computers at home) an ideal "stepping stone" for the attackers because they are typically less secure than the enterprise computers that are protected by IT professionals, but also enables the attackers to leverage man-in-the-middle and social engineering attacks because there are no strong end-to-end (i.e., home-to-enterprise) authentications [55, 202]. Moreover, we see a surge in Zoom-bombing attacks during COVID-19 which leverage vulnerabilities in remote meeting software used for remote work at a unprecedented volume [24]. The health care sector, including hospitals, remains to be an important target of COVID-19 themed cyberattacks as they are overwhelmed with COVID-19 patients [20]. The government, including city counselor and governor's offices, is also targeted by COVID-themed scams and social engineering attacks, perhaps because they are dealing with many urgent purchases of medical items [100]. The media streaming sector is targeted for phishing, scams, and social engineering attacks as they are getting more user attentions for alternative recreation during stay-at-home orders [157]. The medical research centers on COVID-19 are also targeted by COVID-19 themed attacks by state backed

cyber criminals [215].

In summary, attackers appear to have been quickly adapted to target services that remain virtually operating often by victims working from vulnerable home networks during the COVID-19 pandemic. We discuss the 5 classes of COVID-19 themed attacks in the following subsections.

2.2.1 COVID-19 Themed Malicious Websites

Attackers have abused websites to wage COVID-19 attacks to steal login credentials, sell fake medications related to COVID-19, and inject malicious payloads into these themed websites to distribute malware [25, 71, 98]. We map these attacks to the Cyber Kill Chain model as follows.

(i) *Reconnaissance:* An attacker selects target audience, chooses a COVID-19 related target theme, searches for cheap and unregulated domain registration and web hosting services, and sets attack goals. (ii) *Weaponization:* An attacker registers new websites with COVID-19 related names. For example, an attacker may register websites with typo-squatting names to mimic legitimate websites related to COVID-19 (e.g., CDC, WHO, FDA) [181]; an attacker may register websites to imitate legit Virtual Private Network (VPN) software or remote communication software; an attacker may register domains to offer fake legal services related to COVID-19; an attacker may change an existing phishing website to accommodate COVID-19 themes; an attacker may register fake media streaming domains; and an attacker may register fake donation websites. (iii) *Delivery:* An attacker hosts COVID-19 themed malicious websites mentioned above. (iv) *Exploitation:* Victims visit malicious websites, and then trust the fake forms or download malicious payloads to their devices. (v) *Installation:* A victim may provide sensitive information to a malicious website or intentionally/unintentionally install malware. (vi) *C2:* An attacker remotely controls victims' infected computers, for example instructing its agents (e.g., malicious websites, downloaded malware) to send the stolen data/credentials to the attacker. (vii) *Objectives:* An attacker gets sensitive credentials, encrypts a victim's computer, or gets ransom payment.

2.2.2 COVID-19 Themed Malicious Emails

Attackers have abused emails to wage COVID-19 attacks to send phishing, spamming, scamming, malicious attachments, and malicious websites [182]. We map these attacks to the Cyber Kill Chain model as follows.

(i) *Reconnaissance:* An attacker selects target audience, generates and profiles email lists, selects a target topic for COVID-19 themed lures, and sets an attack goal. (ii) *Weaponization:* An attacker creates fake typo-squatting email addresses imitating legitimate entities (e.g., CEO, Netflix support team, medical doctors), writes malicious emails with legitimate logo (e.g., WHO, hospital logo) and authority names, writes emails with COVID-19 related information and offers, writes emails with malicious attachments [73, 194], writes fake COVID-19 donation scam emails [4], writes emails with fake financial relief payments [25], writes emails with blackmailing schemes (e.g., threatening languages) [195], writes emails to lure victims to provide personal information or pay fees for false unemployment training and certification [188]. (iii) *Delivery:* An attacker sends the aforementioned emails to the target audience. (iv) *Exploitation:* A victim trusts an email received from an attacker, clicks its malicious links, opens its attachments, or downloads its malicious contents. (v) *Installation:* A victim replies to the attacker with sensitive personal information or installs malicious content on its computer either intentionally or unintentionally. (vi) *C2:* An attacker establishes connections with victim's devices through C2 channels, for example, to instruct the compromised computers to send back sensitive data. (vii) *Objectives:* An attacker encrypts a victim's computer, receives ransom payment, or receives sensitive information.

2.2.3 COVID-19 Themed Malicious Mobile Apps

Attackers have abused mobile apps to wage COVID-19 attacks to distribute malware and steal information from the victims [80]. Google and Apple have taken steps during this pandemic to reject publishing of COVID-19 related mobile apps from unauthorized entities [191]. Despite these efforts to secure reputed app stores, malicious apps could still get published and remain undetected as many third party app stores do not have proper reviewing and regulation for publishing apps.

Reports showing third-party app stores are eight times more likely to contain malicious apps than than Google Play store [107]. We map the attack of COVID-19 themed malicious mobile apps to the Cyber Kill Chain model as follows.

(i) *Reconnaissance:* An attacker selects target audience (e.g., based on geographical region), selects a COVID-19 themed topic/service (e.g., tracing, tracking, maps, VPN, remote meeting, COVID-19 guidelines, COVID-19 test information), finds and selects profitable unregulated app stores, and sets attack goals. (ii) *Weaponization:* An attacker creates fake mobile apps with typo-squatted app names and legitimate logos to imitate authentic apps, repackages existing COVID-19 themed legitimate apps with malware or ransomware (e.g., banking Trojan, spyware) to trick users [80]. (iii) *Delivery:* An attacker uploads malicious apps into the unregulated app stores or code repositories, and advertises these mobile apps through websites pop-ups. (iv) *Exploitation:* A victim trusts an malicious app and downloads the app. (v) *Installation:* A victim installs the downloaded malicious app on an mobile device. (vi) *C2:* An attacker remotely controls victims' compromised mobile devices to send sensitive user data to the C2 server. (vii) *Objectives:* An attacker encrypts a victim's mobile device, gets a ransom payment, or steals a victim's private information (e.g., login credentials, crypto wallet passwords), breaches user privacy (e.g., location).

2.2.4 COVID-19 Themed Malicious Messaging

Attackers have abused messaging services to wage COVID-19 attacks (e.g., phishing, malware, spamming, and scamming) [47]. COVID-19 has increased the usage of mobile devices which create more incentives for attackers. These attacks are similar to malicious email attacks, but are unique in that messaging can offer more emotional and persuasive live chats. We map them to the Cyber Kill Chain model as follows.

(i) *Reconnaissance:* An attacker selects target audience (e.g., based on demography, geography, severity of COVID-19 infections), collects phone and social media contacts, selects target platform (e.g., Facebook, WhatsApp, Twitter), chooses a COVID-19 themed topic (e.g., fake cures, products, services), and sets attack goals. (ii) *Weaponization:* An attacker writes persuasive and

emotional messages (e.g., asking for COVID-19 donations) to trick victims, creates fake social media profiles, and creates social media groups to lure target audience. (iii) *Delivery:* An attacker sends malicious messages, website links, and attachments through messaging to targeted victims, sends scams mentioning fines for leaving home during stay-at-home orders [47], sends fraud messages with free subscription lures for media streaming services [98], sends messages to sell low-quality supplies (e.g., masks, gloves, fake cures, and illegal chemical materials) [83], sends COVID-19 related lucrative offers (e.g., giveaways, loans, lawyer help, food stamps, stimulus check updates, news guidelines), and sends crafted misinformation messages with fake claims and made up evidence. (iv) *Exploitation:* A victim trusts a received message and falls victim to it by clicking its malicious links, downloading its malicious contents, and forwarding it to other users. (v) *Installation:* A victim intentionally or unintentionally installs the malicious payload on an messaging device (e.g. Android mobile phone). (vi) *C2:* An attacker establishes channels (e.g., reply messages, servers connected to a phishing webpage) to remotely control the compromised messaging devices, for example, to receive victims' sensitive information. (vii) *Objectives:* An attacker gets victims' sensitive information or makes lateral movements in victims' networks.

2.2.5 COVID-19 Themed Misinformation

Attackers have waged COVID-19 attacks to spread misinformation, which includes false or inaccurate information (e.g., hoaxes, rumors, or propaganda [92]). Examples include: "COVID-19 is invented in a Chinese lab [18]"; "5G is spreading COVID-19 [207]", "Black are immune to COVID-19 [184]", "X can cure COVID-19" where X can be a drug or food items (i.e., Ginger) [212], or "Wearing a mask causes you to inhale too much carbon dioxide, which can make you sick" or "Wearing a mask can result in getting pneumonia" [**?**]. Social media and messaging platforms further increase the impact of such misinformation. The term *Infodemic* has even been coined because of this [45]. We map the COVID-19 themed misinformation attack to the Cyber Kill Chain as follows.

(i) *Reconnaissance:* An attacker analyzes the characteristics of targeted audience (e.g., ethnic-

ity, demography or nationality), identifies vulnerable divisions in society, selects themed topics, and sets attack goals. (ii) *Weaponization:* An attacker writes fake COVID-19 themed statements and mix them with false evidence and out-of-context truths, creates fake groups in social networking platforms, creates themed memes, creates bots in social media (e.g., Twitter) to propagate misinformation, and infiltrates into social media groups containing targeted ethnic audience. (iii) *Delivery:* An attacker posts and shares COVID-19 related misinformation (e.g., narratives, memes, images, and hashtags through social media groups and messaging apps) and publishes fake news on paid online news/tabloids, and/or keeps posting to a larger audience with bots to amplify the impact. (iv) *Exploitation:* A victims (e.g., social media user) reads and forwards misinformation messages. (v) *Installation:* A victims gets to believe the misinformation which goes viral. (vi) *C2:* An attacker may generate fake real-life incidents/experience posts on social media related to COVID-19. (vii) *Objectives:* An attacker succeeds when bringing more division, mistrust, health crisis, and chaos in society, and possibly earns money from the crisis.

2.2.6 Systematizing COVID-19 Themed Cyberattacks

We systematize COVID-19 attacks by mapping them to their attack techniques and attack goals, and by contrasting their Cyber Kill Chain models.

Mapping Attacks, Techniques and Goals

Figure 2.2 depicts the mapping between the COVID-19 attacks, the attack techniques they use, and their attack goals. We observe that one attack may use multiple attack techniques. For example, a COVID-19 themed malicious website attack may use a range of attack techniques, including phishing, malware, ransomware, vaccine scams, donation scams, masks scams, testing scams, and VPN scams. Moreover, a COVID-19 themed malicious website attack may have multiple goals. On the other hand, one goal can be achieved by using various kinds of attack techniques, which may be waged through multiple classes of attacks. This means that when an attacker attempt to achieve an attack goal, the attacker can choose attacks and attack techniques in a cost-effective, if

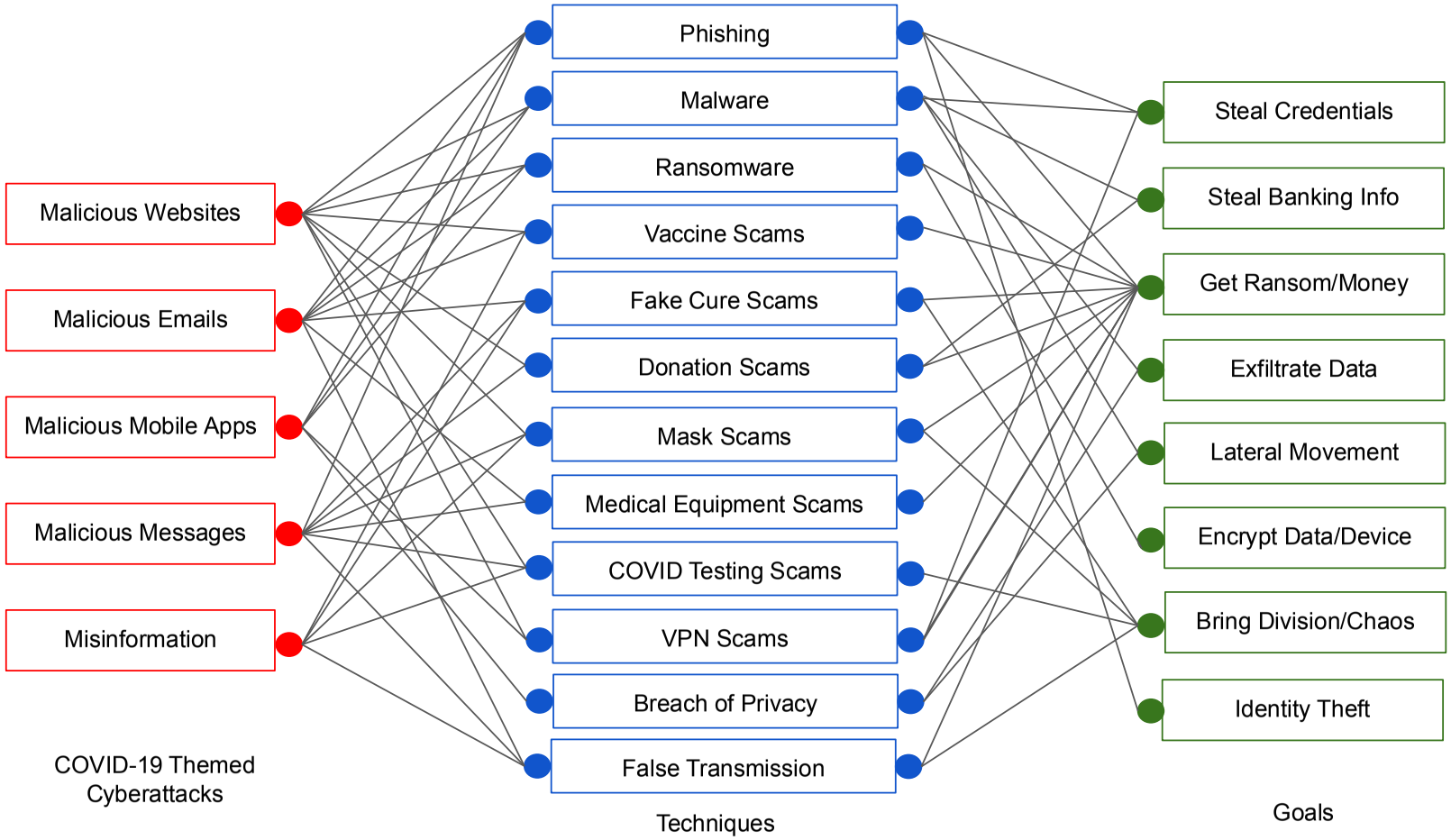

Figure 2.2: Systematizing COVID-19 attacks (red), attack techniques (blue), and attack goals (green)

not optimal, fashion. For example, each attack may incur some cost or risk (e.g., the cost for using phishing via COVID-19 themed malicious websites and COVID-19 themed malicious emails may be different), and may have different success probabilities (e.g., phishing via COVID-19 themed malicious websites may be more or less successful than phishing via COVID-19 themed malicious emails). This would allow an intelligent attacker to wage the cost-effective or event optimal attack. A systematic framework for achieving this type of attacker decision-making is beyond the scope of the present paper.

Insight 1. *A COVID-19 attack may use multiple attack techniques to achieve multiple attack goals, and an attack goal may be achieved by using multiple attack techniques that can correspond to multiple attacks. This flexibility allows the attacker to choose cost-effective, if not optimal, attacks in order to achieve a certain attack goal.*

Systematizing Attacks via Their Cyber Kill Chains

Figure 2.3 depicts the Cyber Kill Chain mappings of the aforementioned 5 classes of COVID-19 attacks, which are represented by different colors. We observe that in each stage of the Cyber Kill Chain, there can be multiple *tactics* (e.g., "select target audience" and "choose COVID-19 theme topic" at the reconnaissance stage). We observe that the 5 classes of COVID-19 attacks would use some common tactics at some stages as well as their distinct tactics at other stages. For example, "select target audience" at the reconnaissance stage is a tactic that can be used by the 5 classes of attacks, but "find unregulated app stores" is a tactic that would be unique to the COVID-19 themed malicious apps attack. We also observe that the *exploitation* stage almost always leverages victims' mistrust in social engineering, which highlights that human factor remains to be a critical vulnerability in COVID-19 attacks, which reinforces the importance of seeking effective defenses against such attacks [149].

Insight 2. *COVID-19 attacks can be very sophisticated, rather than only opportunistic, which means that effective defense must be designed on a deeper understanding about the attack tactics that can be used in each stage of the attack (i.e., knowing the attacker better).*

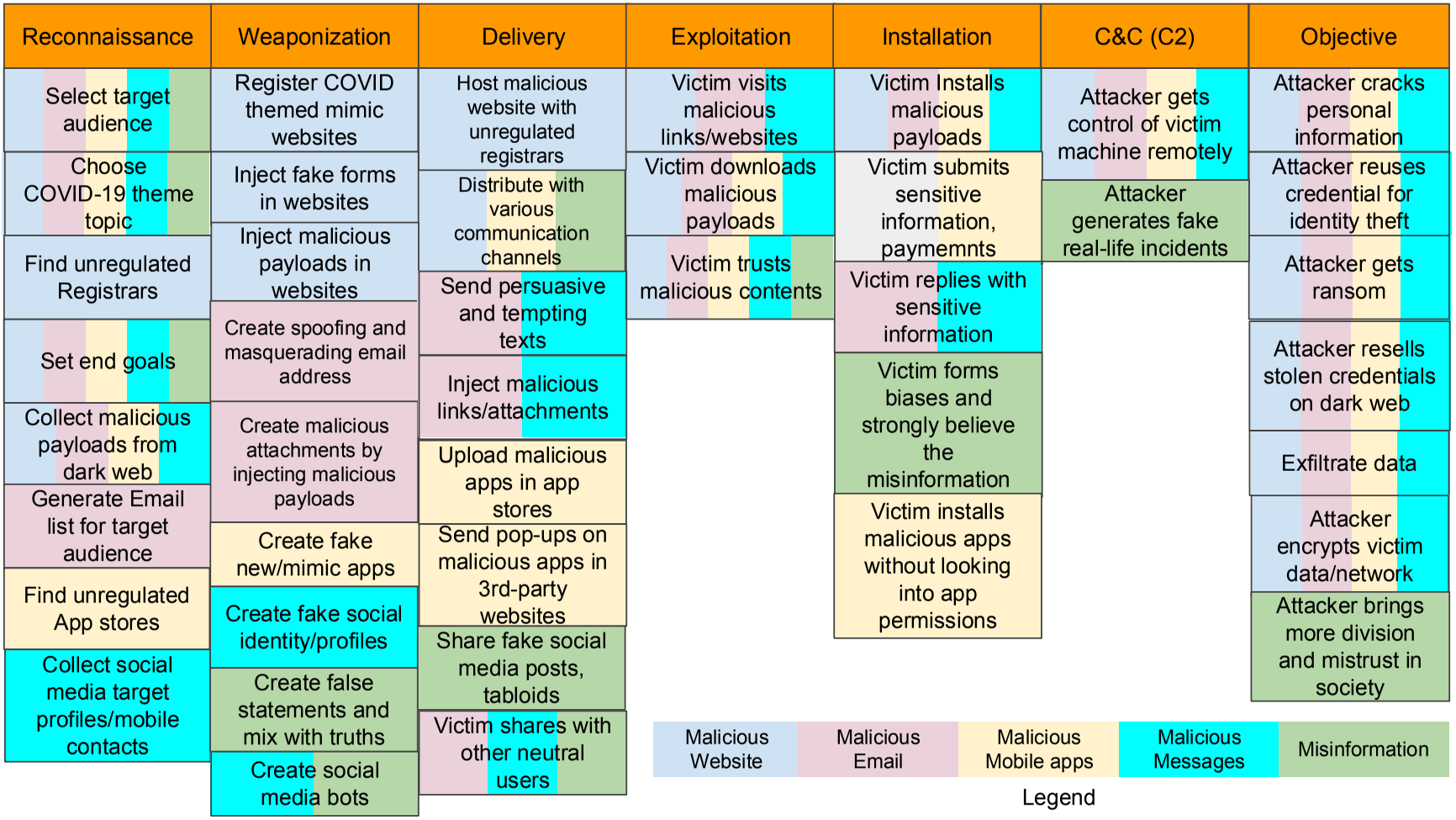

Figure 2.3: Systematizing the Cyber Kill Chains for COVID-19 themed cyberattacks, which are coded in colors (see Legend).

2.3 Exploring the Defense Space

The preceding characterization of COVID-19 attacks guides us to explore defense strategies against them, with an emphasis on *what-to-leverage* when designing defense systems. The investigation of these proposed approaches is beyond the scope of the present paper. This is because each approach needs to be investigated separately, with corresponding experiments.

2.3.1 COVID-19 Malicious Websites Defense

We propose four approaches to defending against COVID-19 themed malicious websites. The first approach is to leverage various website contents pertinent to COVID-19. What is unique to content-based detection of COVID-19 themed malicious websites is the COVID-19 related features, such as the presence or absence of keywords in website names (e.g., *coronavirus*, *COVID-19*, *masks*, *n95*, and *test*). The second approach is to leverage website environment, including URLs' information. For example, typo-squatting URLs or mimicking fake websites can be detected by analyzing URLs information and website screenshots. The third approach is to leverage websites' age information. Since COVID-19 themed malicious websites would be created after the outbreak of the COVID-19 pandemic, hinting that the lifetime of many such websites would be short. The fourth approach is to leverage effective training to make users more skeptical about website contents.

2.3.2 COVID-19 Malicious Emails Defense

We propose three approaches to defending against COVID-19 themed malicious emails. The first approach is to filter emails by searching COVID-19 themed keywords in their subject lines and contents. Examples of such keywords include: *COVID-19 cures*, *COVID-19 guidelines*, and *COVID-19 offers*. The second approach is to verify the sender email address to detect email masquerading [16]. The third approach is to leverage email content, for example by analyzing their attachments, links and texts.

2.3.3 COVID-19 Malicious Mobile Apps Defense

We propose four approaches to defending against COVID-19 themed malicious apps. The first approach is to leverage computer vision to proactively examine newly published app's logos, especially when they are similar to, if not exactly the same as, the logos of some popular legitimate apps. The second approach is to analyze the content of apps to detect the malicious ones (e.g., repackaged apps). For this purpose, static analysis, dynamic analysis, and their combinations may be utilized. The third approach is to examine the string edit distance of app names with respect to some popular ones. The fourth approach is to train users to improve their awareness of malicious apps according to some best practices in using mobile apps securely [137].

2.3.4 COVID-19 Malicious Messaging Defense

We propose three approaches to defending against COVID-19 themed malicious messaging. The first approach is to leverage message content to check if a message contains suspicious content (e.g., the presence of URLs, emoticons, special characters, and COVID-19 themed keywords). The second approach is to detect persuasive messages waging social engineering cyberattacks. This may be achieved by analyzing texts and leveraging human factors and psychological means [**?**]. The third approach is to train users to improve their awareness of COVID-19 themed malicious messages.

2.3.5 COVID-19 Misinformation Defense

We propose four approaches to defending against COVID-19 themed misinformation attacks. The first approach is to use fact-checking to detect fake news (or social media posts), perhaps by aggregating similar news reports from credible sources and AI or machine learning techniques. For example, create supervised ML models for detection based on labeled social media posts leveraging NLP techniques for any specific themed misinformation. The second approach is to use central repositories to host COVID-19 related information and resources (e.g., Facebook's COVID-19 Information Center). The third approach is to train and educate users to improve their skills and

capabilities in recognizing fake misinformation. The fourth approach is to leverage crowdsourcing, namely encouraging or incentivizing users to report COVID-19 suspicious misinformation posts and links.

2.4 Related Work

The problem of COVID-19 attacks has started to receive attention from the research community right after the COVID-19 outbreak. There are studies on the types of cyberattacks and their overall trends amid the COVID-19 pandemic [47, 71, 98, 160]. Moreover, there are studies on specific cyberattacks and cybercrimes themed with various COVID-19 related topics such as mobile malware ecosystem and their detection [80, 105], themed domain campaigns and their detection [217], themed phishing attacks and their detection [3, 7, 8], themed malware [158], fake social media posts themed with health information and their detection [74, 178], and themed cryptocurrency scams [218]. When compared with these studies, we aim at systematically characterizing the landscape of the COVID-19 attacks, the usage of themed malicious websites, including their sophistication through the the Kill Chain [90] and exploring the space of defenses against these attacks, which is not systematically understood by looking at the individual studies separately. It is essential to understand the landscape of these themed cyberattacks before compiling any effective defense action plans and designing guidelines for safeguarding users from these threats.

2.5 Chapter Summary

We have explored the landscape of COVID-19 themed cyberattacks and defenses. We discussed 5 classes of attacks and mapped them to the Cyber Kill Chain model. We explored defense strategies against these attacks. Although the study is geared towards COVID-19 themed cyberattacks, the exploration and landscape can be adapted to future X-themed cyberattacks exploiting future events (e.g., election, natural or man-made disasters). It is also interesting to rigorously model these attack-defense interactions in the Cybersecurity Dynamics framework [37, 145, 227, 235].

CHAPTER 3: DATA-DRIVEN DETECTION OF THEMED MALICIOUS WEBSITES

Themed malicious website is a emerging threat that is leveraged by attackers to craft their website based attacks with a popular theme. The attackers carry out these attacks often using some variation of social engineering techniques to penetrate user trust and attention. During COVID-19, it has hit hard on the global community, and organizations are working diligently to cope with the new norm of "work from home". This high volume of remote work is unprecedented and creates opportunities for cyber attackers to penetrate home computers. Attackers have been leveraging websites with COVID-19 related names, dubbed *COVID-19 themed malicious websites*. We have observed a large volume of COVID-19 themed domain names registered where many are later found to be involved in shady activities. This sudden rise in the volume of COVID-19 themed websites bring the attention of the community to focus on underlying infrastructures and ways to deal with them. These themed malicious websites mostly contain false information, fake forms, fraudulent payments, scams, or malicious payloads to steal sensitive information or infect victims' computers. In this chapter, we present a data-driven study to characterize the underlying hosting infrastructures and detect these COVID-19 themed malicious websites. Our characterization study shows that attackers are agile and are deceptively crafty in designing geolocation targeted websites, often leveraging popular domain registrars and top-level domains (TLDs). Our detection study shows that the Random Forest classifier can detect COVID-19 themed malicious websites based on the defined domain lexical and WHOIS features, achieving a 98% accuracy with a reasonable 2.7% false-positive rate.

3.1 Chapter Introduction

The COVID-19 pandemic has incurred many new cyber attack vectors. Many of these cyber attacks incorporate COVID-19 themed factors into phishing, malware, and scamming schemes for various malicious goals (e.g., monetary benefits, stealing credentials, stealing credit card numbers,

or identity theft). For example, there is reportedly a 148% increase in ransomware attacks in March 2020 compared with February 2020 [168], where many attacks are initiated by themed malicious websites abusing victims' trust.

This chapter focuses on one emerging attack vector, namely malicious websites leveraging a theme or *themed malicious websites*. We choose COVID-19 themed malicious websites as a representative for the case study on the detection of themed malicious websites [248]. As organizations incorporate the "work from home" policy, the consequences of COVID-19 themed malicious websites can be significantly amplified because home computers are often more vulnerable to attack than work computers. During the COVID-19 pandemic, many people lost their jobs and are affected by mental health issues, which causes excessive pressures. These pressures may make average users even more vulnerable to social engineering attacks waged via COVID-19 themed malicious websites. This increases the motivation of the importance of understanding and defending against COVID-19 themed malicious websites, which is a new problem that has not been studied before in a systematic way.

Chapter contributions. In this chapter, we make the following contributions. First, we propose a methodology for characterizing and detecting COVID-19 themed malicious websites through a data-driven approach. To the best of our knowledge, this is the first study on *data-driven* characterization and detection of COVID-19 themed malicious websites. Second, we apply the methodology to specific datasets to draw the following insights: (i) some attackers may be incentivized to use cheaper registrars for registering COVID-19 themed malicious websites; (ii) attackers often abuse popular top-level domains for their COVID-19 themed malicious websites; (iii) attackers are agile in waging the COVID-19 themed malicious website attack; (iv) attackers are crafty in using COVID-19 themed keywords, and geographical information in creating COVID-19 themed malicious website domain names; (v) the small degree of data imbalance does not have any significant impact in the effectiveness of detecting COVID-19 themed malicious websites; and (vi) COVID-19 themed malicious website detectors must consider WHOIS features and Random Forest performs better than K-nearest neighbor, decision tree, logistic regression, and support vector machine.

Chapter Outline. The rest of the chapter is organized as follows. Section 3.2 explores the research questions and the methodology, which guide us to characterize and detect COVID-19 themed malicious websites. Section 3.3 reports the experiments and results. Section 3.4 explores the related work. Section 3.5 concludes the Chapter.

3.2 Methodology

Next, our methodology for *data-driven* characterization and detection of COVID-19 themed malicious websites is centered at answering a range of research questions.

3.2.1 Characterization Methodology

In order to characterize COVID-19 themed malicious websites, we address 4 Research Questions (RQs):

- RQ1: Which WHOIS registrars are most abused to launch COVID-19 themed malicious websites?
- RQ2: Which Top Level Domains (TLDs) are most abused by COVID-19 themed malicious websites?
- RQ3: What trends are exhibited by COVID-19 themed malicious websites?
- RQ4: Which theme keywords are mostly abused by attackers, and how?

We consider WHOIS information because it has shown to be useful in the era prior to the COVID-19 pandemic [222, 223]. Answering the preceding questions will deepen our understanding of COVID-19 themed malicious website attacks.

3.2.2 Detection Methodology

We propose leveraging machine learning to detect COVID-19 themed malicious websites and answer:

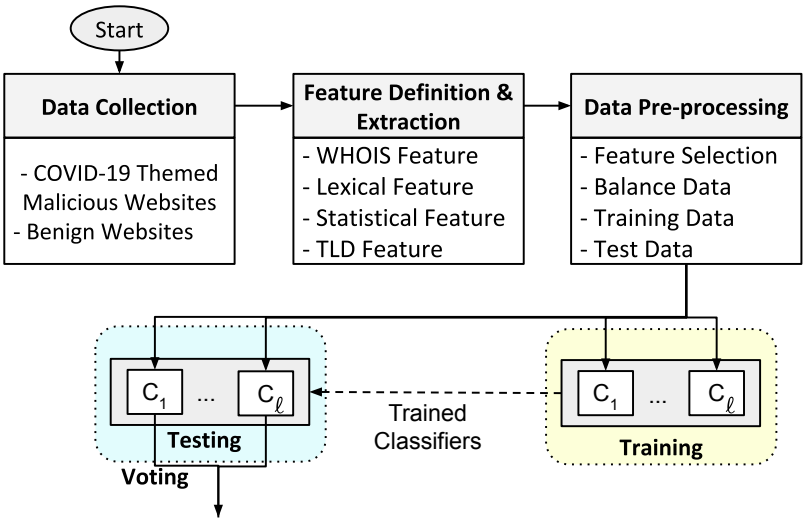

Figure 3.1: Methodology for detecting COVID-19 themed malicious websites

- RQ5: Which classifier is competent in detecting COVID-19 themed malicious websites?
- RQ6: What is the impact of WHOIS features on the classifier's effectiveness?

In order to answer these questions, we need to train detectors. Figure 3.1 highlights the methodology for detecting COVID-19 themed malicious websites. The methodology can be decomposed into the following modules: data collection, feature definition and extraction, data pre-processing, classifier training, and classifier test.

Data about websites need to be collected from reliable sources. The collected data may need enrichment to provide more information, as what will be illustrated in our case study. Then, features may be defined to describe these websites. In the case of using deep learning (which requires much larger datasets), features may be automatically learned. One may consider a range of classifiers, which are generically called C_i's in Figure 3.1. As shown in Figure 3.1, one can use classifiers individually or an ensemble of them (e.g., via a desired voting scheme, such as weighted vs. unweighted majority voting). In the simple form of unweighted majority voting, a website is classified as malicious if majority of the classifiers predict it as malicious; otherwise, it is classified as benign.

In order to evaluate the effectiveness of the trained classifiers, we propose adopting the standard metrics, including: accuracy (ACC), false-positive rates (FPR), false-negative rates (FNR), and

$F1$-score. Specifically, let TP be the number of true positives, TN be the number of true negatives, FP be the number of false positives, and FN be the number of false negatives. Then, we have ACC $= \frac{TP+TN}{TP+TN+FP+FN}$, FPR $= \frac{FP}{FP+TN}$, FNR $= \frac{FN}{FN+TP}$, and $F1$-score $= \frac{2TP}{2TP+FP+FN}$.

3.3 Case Study

Our case study applies the methodology to specific datasets.

3.3.1 Data Collection

Our dataset of COVID-19 malicious website examples are obtained from what was published between 2/1/2020 and 5/15/2020 by two sources: (i) CheckPhish [33], which contains 131,761 malicious websites waging scamming attacks related to COVID-19; and (ii) DomainTools [60], which contains 157,579 malicious websites waging malware, phishing, and spamming attacks related to COVID-19. The union of these two sets leads to a total of 221,921 malicious websites, denoted by $D_{malicious}$, owing to the fact that 67,419 websites belong to both sets. For obtaining benign websites, we use the top 250,000 websites from Cisco's Umbrella 1 million websites dataset [87] on 05/16/2020, denoted by D_{benign}, which is a source of reputable websites. We compile a merged dataset denoted by $D_{initial} = D_{malicious} \cup D_{benign}$.

In order to collect WHOIS information of a website, we use the python library `whois 0.9.7` to query the WHOIS database on 8/7/2020. We observe that 42,540 (or 19.17%) out of the 221,921 malicious websites have no WHOIS information available, and 93,082 (or 37.2%) out of the 250,000 benign websites have no WHOIS information available. This means that the presence/absence of WHOIS information does not indicate that a website is malicious or not.

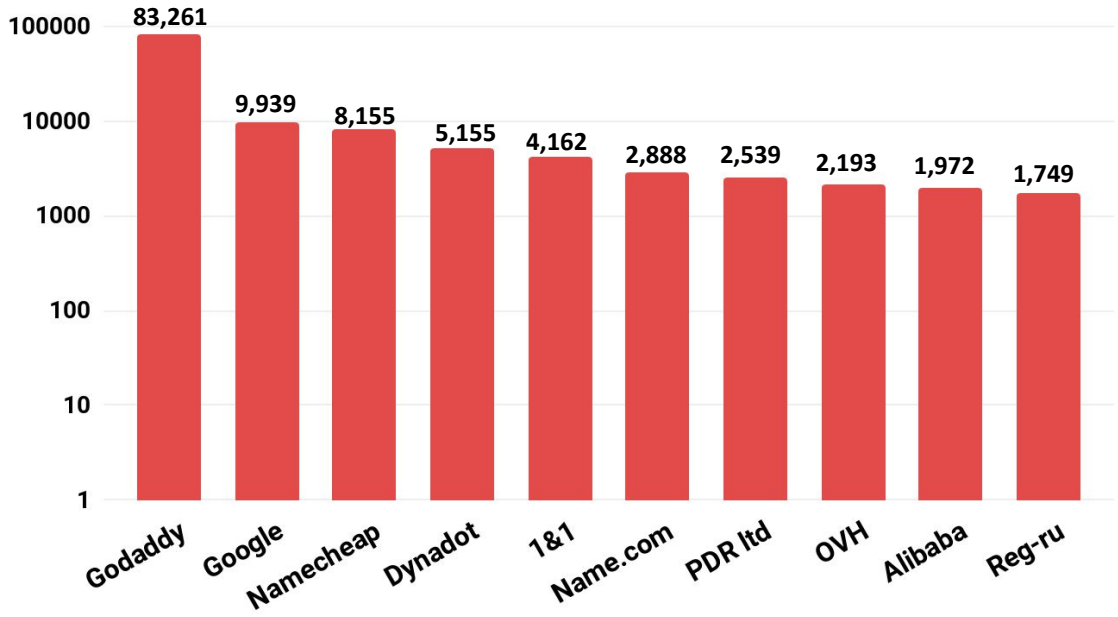

Figure 3.2: Top 10 abused WHOIS registrars of COVID-19 themed malicious websites (the y-axis is in the log-scale).

3.3.2 Characterization Case Study

Answering RQ1: Identifying the WHOIS registrars that are most abused to launch COVID-19 themed malicious websites

For this purpose, we use a subset of $D_{malicious}$ set, denoted by $D'_{malicious}$, which contains 171,901 malicious websites with WHOIS *registrar_name* information available.

Figure 3.2 depicts the top 10 abused registrars, which are ranked according to the absolute number of COVID-19 themed websites in $D'_{malicious}$ that are respectively registered by them. We observe that `Godaddy` is the most frequently abused registrar, followed by `Google` and `Namecheap`. This finding inspires us to analyze if there is any financial incentive behind the use of a specific registrar. The cost registering a `.com` domain in the first year, is: `Godaddy` for $11.99, `Google` for $9, `Namecheap` for $8.88, `Dynadot` for $8.99, `1&1` for $1, `name.com` for $8.99, `PDR Ltd` for $35, `OVH` for $8.28, `Alibaba` for $7.99, `Reg-ru` for $28. This suggests that some attackers might have considered registrar `1&1` because it is the cheapest, while some attackers use reputed registrars.

Insight 3. *Some attackers may be incentivized to use cheaper registrars but some of the other don't.*

Answering RQ2: Which Top Level Domains (TLDs) are most abused by COVID-19 themed malicious websites?

In order to answer this question, we use the original dataset $D_{malicious}$, which contains 221,921 COVID-19 themed malicious websites with corresponding TLD information.

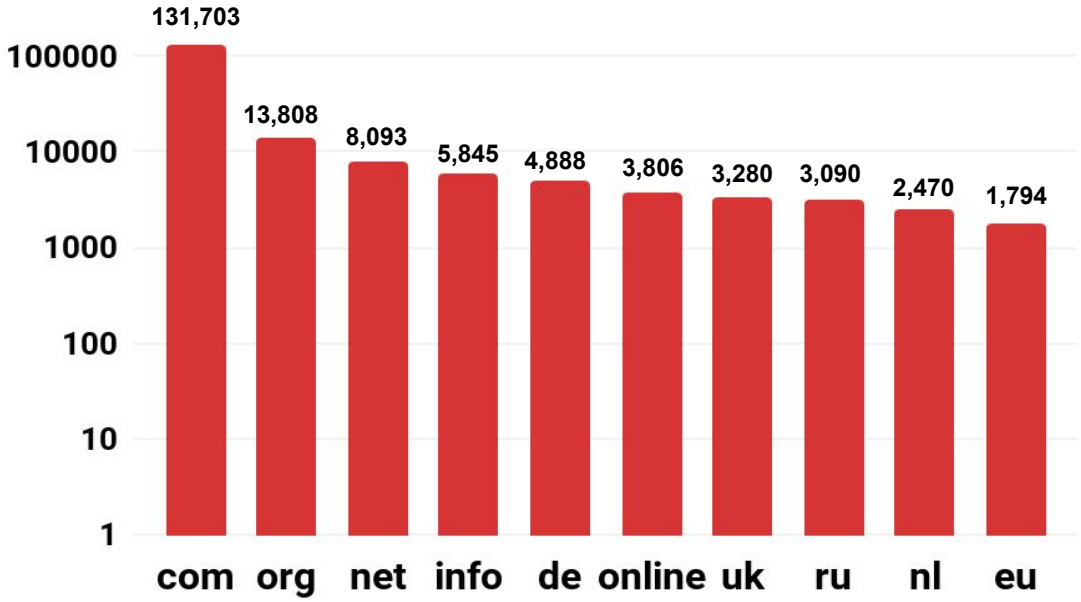

Figure 3.3: Top 10 abused TLDs of COVID-19 themed malicious websites (the y-axis is in the log-scale).

Figure 3.3 depicts the top 10 abused TLDs, which are ranked according to the absolute number TLDs for COVID-19 themed malicious websites. We make the following observations. First, `.com` hosts the highest number of malicious websites, followed by `.org` and `.net`. Second, 5 of the top 10 abused TLDs correspond to country-level `ccTLDs`, including `.de`, `.uk`, `.ru`, `.nl` and `.eu`.

Insight 4. *Attackers often abuse popular TLDs.*

Answering RQ3: What trends are exhibited by COVID-19 themed malicious websites?

In order to answer this question, we use the dataset $D_{malicious}$ mentioned above. Figure 3.4 depicts the trend of malicious websites, leading to two observations. First, there is a discrepancy between the daily numbers of websites that are reported by the two sources. According to CheckPhish, the number of COVID-19 themed malicious websites reaches the peak on 03/25/2020, with 18,495 malicious websites; according to DomainTools, the number of COVID-19 themed malicious websites reaches a peak on 03/20/2020, with 3,981 malicious websites. This data indicates that there are reporting inconsistencies among sources and many COVID-19 themed malicious websites are created at the early stage of the pandemic when *uncertainties* are maximum. Second, the number of COVID-19 themed malicious websites, by and large, has been decreasing since the last week of March 2020 (i.e., two weeks after the pandemic declaration), leading to about 1,000 websites per day during the first week of May 2020 (i.e., about two months after pandemic declaration). However, there is still oscillation. One possible cause is that the attackers have been waiting to create new COVID-19 themed malicious websites based on the pandemic's new developments (e.g., vaccine).

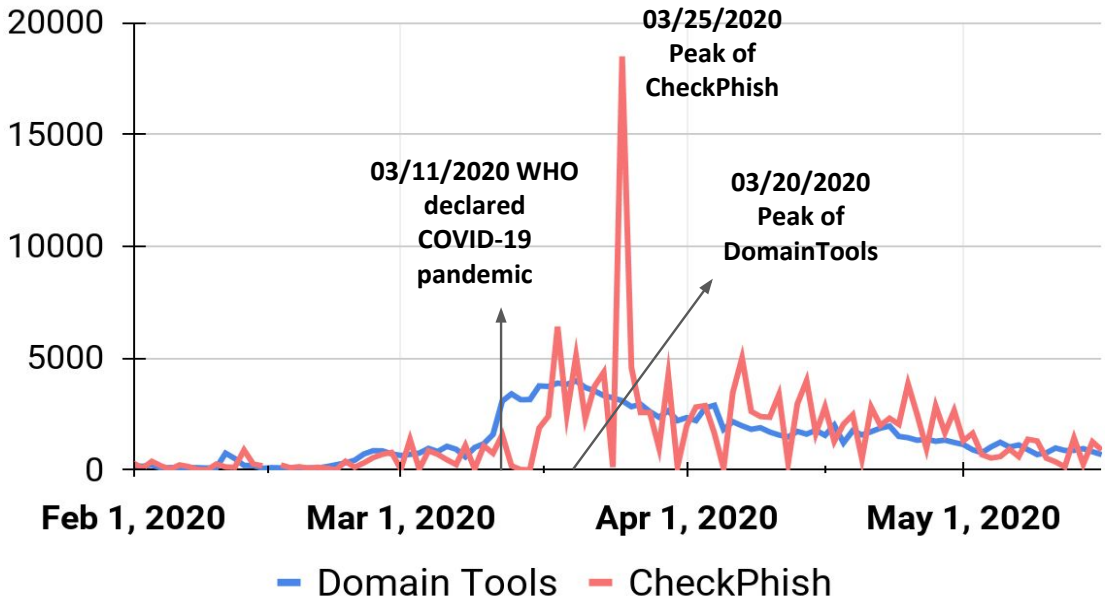

Figure 3.4: Trends of COVID-19 themed malicious website.

Insight 5. *Inconsistencies in reporting mechanisms, attackers are agile in creating COVID-19 themed malicious websites.*

Answering RQ4: Which theme keywords are mostly abused by attackers, and how?

In order to answer this question, we analyze the dataset $D_{malicious}$ mentioned above. We use the python library `wordninja` with English Wikipedia language model [12] to split domain name strings and extract COVID-19 themed keywords. We observe that 4 keywords (i.e., *covid*, *corona*, *covid19*, and *coronavirus*) are most widely used as expected; they are followed by *mask*, *quarantine*, *virus*, *test*, *facemask*, *pandemic*, and *vaccine*. We extract more than 19,000 keywords. A further analysis of the domain names reveals that attackers create COVID-19 themed malicious websites with names containing geographical attributes. For example, `coronaviruspreventionsanantonio.com`, `coronavirusprecentionhouston.com`, and `coronaviruspreventiondallas.com` use a combination of city name and a COVID-19 themed keyword. Moreover, we observe the existence of COVID-19 themed "parking" websites, which have no content at the present time but might be used for upcoming COVID-19 themes.

Insight 6. *Attackers are crafty in using COVID-19 themed keywords and geographical information in creating COVID-19 themed malicious website domain names.*

3.3.3 Detection Case Study

Given $D_{initial}$, the detection case study proceeds as follows.

Feature Definition and Extraction

We define features according to the following aspects of websites: WHOIS (F1-F4), domain name lexical information (F5-F9), statistical information (F10), and Top-Level Domain or TLD (F11).

- Current WHOIS registration lifetime (F1): This is the number of days that has passed since a website's registration, with respect to the date when this feature's value is extracted (e.g., 08/07/2020 in our case).

- Remaining WHOIS expiration lifetime (F2): This is the number of remaining days before a website's WHOIS registration expires, with respect to the date when this feature's value is extracted (e.g., 08/07/2020 in our case).

- Number of days since last WHOIS update (F3): This is the number of days elapsed since a website's last update with respect to the date when this feature's value is extracted (e.g., 08/07/2020 in our case).

- WHOIS registrar reputation (F4): We propose measuring a WHOIS registrar's reputation as $\frac{n}{|D_{benign}|}$, where n is the number of benign websites in D_{benign} that are registered by this particular registrar and $|D_{benign}|$ is the size of set D_{benign}.

- Number of dots in domain name (F5): This is the number of dots (character '.') in the domain name. For example, domain `any.com` has 1 dot.

- Domain hyphen count (F6): This is the number of hyphens ('-') in a domain name.

- Domain vowel count (F7): This is the number of vowels (i.e., *a*, *e*, *i*, *o*, *u*) in a domain name.

- Domain digits percentage (F8): This is the ratio of the number of digits (0-9) in a domain name to the number of characters including digits.

- Domain unique alphabetic-numeric characters count (F9): This is the total number of unique alphabetic and numeric characters (i.e., a-z, A-Z, 0-9) in a domain name.

- Domain entropy (F10): This is the Shannon entropy [214] of the domain name (i.e., a kind of statistical information), which is computed based on the frequency of characters in the domain name.

- TLD Reputation (F11): We propose measuring a TLD's reputation as $\frac{m}{|D_{benign}|}$, where m is the number of websites in D_{benign} that contain this particular TLD.

Data Pre-Processing

Given that some websites may not have information for the features, it is important to consider different scenarios. In our example, we propose considering two datasets that can be derived from $D_{initial}$ because some websites do not have information for the WHOIS features.

- Dataset $D_1 \subset D_{initial}$ consists of websites for which WHOIS information is available (i.e., features F1-F4 are available). D_1 contains 21,749 websites in total, including 16,411 COVID-19 themed malicious websites and 5,338 benign websites.

- Dataset $D_2 \subset D_{initial}$, where $D_1 \cap D_2 = \emptyset$, consists of websites for which WHOIS information is absent (i.e., features F1-F4 are entirely missing). D_2 contains 135,621 websites, including 42,540 malicious websites and 93,081 benign websites. For each website belonging to D_2, only values of the 7 features (i.e., F5-F11) are available.

Table 3.1: Relative importance of features in D_1 with respect to the random forest method.

Feature	Importance	Feature	Importance
F1	0.429	F7	0.080
F2	0.094	F8	0.009
F3	0.131	F9	0.028
F4	0.065	F10	0.029
F5	0.065	F11	0.068
F6	0.003		

Since only D_1 contains all WHOIS information, We use it for feature selection study. For this purpose, we use the *random forest classification feature importance* method [101] (with the 80-20 splitting of training-test data) to find the important features. Table 3.1 depicts the relative importance of the features in D_1. We observe that F6 and F8 have a very small relative importance (i.e., < 0.01) when compared to the others, suggesting that hyphens and digits are equally used in malicious or benign domain names. Hence, we will eliminate F6 and F8 in the rest study of D_1.

In order to see whether or not the feature selection result is impacted by the data imbalance of D_1 (with the malicious:benign ratio being 3.1:1), we explore two widely-used methods: (i)

oversampling the minority class to replicate some random examples; and (ii) *undersampling* the majority class to remove some random examples. At first, we do the 80-20 splitting of training-test data, and then change the malicious:benign ratio in the training set, while keeping the test set intact. We wish to identify the ratio that achieves the highest $F1$-score. In what follows we only report the results of Random Forest because it outperforms the other classifiers for the original dataset D_1.

Table 3.2: Impact of the malicious:benign ratio on the effectiveness of the Random Forest classifier with *Oversampling* and *Undersampling*, where D_1 with ratio 3.1:1 is the original D_1.

Dataset	Method	*Ratio*	ACC	FPR	FNR	$F1$-score
D_1	(none)	3.1:1	0.980	0.030	0.017	0.987
D_1	Oversample	2:1	0.980	0.030	0.018	0.986
D_1	Oversample	1.67:1	0.980	0.027	0.017	0.988
D_1	Oversample	1.43:1	0.979	0.028	0.019	0.986
D_1	Oversample	1.25:1	0.979	0.028	0.018	0.986
D_1	Oversample	1.11:1	0.979	0.027	0.019	0.986
D_1	Oversample	1:1	0.979	0.026	0.019	0.986
D_1	Undersample	2:1	0.977	0.023	0.022	0.985
D_1	Undersample	1.67:1	0.976	0.023	0.025	0.984
D_1	Undersample	1.43:1	0.975	0.023	0.025	0.984
D_1	Undersample	1.25:1	0.972	0.020	0.031	0.981

Table 3.2 shows the impacts of the malicious:benign ratio in the training set. We observe that the oversampling-incurred ratio 1.67:1 leads to the highest $F1$-score (and the second best FPR and lowest FNR), while under-sampling never performs better than the original data ratio in terms of accuracy and $F1$-score. This can be explained by the fact that the latter eliminates useful information. This prompts us to use oversampling to achieve the 1.67:1 ratio when training classifiers, which turns D_1 into D_1' (i.e., the training set is augmented).

Figure 3.5 further highlights the *confusion matrix* of the experiment one the same test set but corresponding to D_1 and D_1', which shows a slight improvement in detection when augmenting the training set with oversampling.

Insight 7. *The data imbalance issue does not affect the model performance significantly in this*

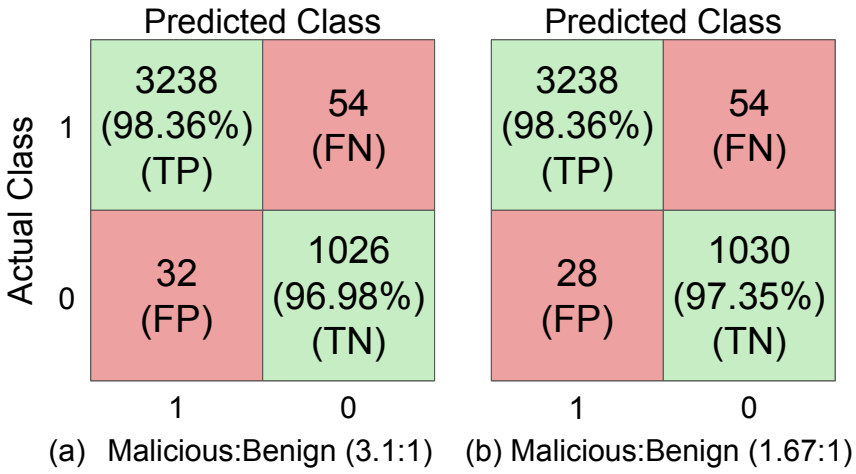

Figure 3.5: Confusion matrix for (a) D_1 with 3.1:1 malicious:benign ratio in the training data and (b) D_1' with 1.67:1 ratio in the training data.

case, perhaps because the degree of imbalance is not severe enough.

Training and Test

Having addressed the issue of feature selection and data imbalance, we consider the following classifiers: Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). Specifically, we use the python `sklearn` module to import the following classifier algorithms: (i) Random Forest or RF with parameters `n_estimator`=100 (i.e., 100 trees in a forest) and `criterion`=*'entropy'* (i.e., entropy is used to measure information gain); (ii) K-Nearest Neighbor or KNN, with parameters `n_neighbors`=8 (i.e., 8 of neighbors are considered), `metric`=*'minkowski'* with $p = 2$ (i.e., the Minkowski metric with $p = 2$ measures the distance between two feature vectors), and the rest parameters are the default values; (iii) Decision Tree or DT with default parameters; (iv) Logistic Regression or LR with default parameters; (v) Support Vector Machine or SVM with `linear` kernel and other default parameters. For voting the outputs of the five classifiers mentioned above, we use the `VotingClassifier()` function and set `voting`=*'hard'* (i.e., majority voting). We always considering the 80-20 splitting of the scaled training-test data.

Answering RQ5 and RQ6

In order to answer RQ5 and RQ6, we conduct the following experiments, where we use the 80-20 train-test splitting of D_1 and then augmenting the training set as mentioned above. Our experiments are conducted on a virtual machine on `https://www.chameleoncloud.org/`, running CentOS 7 on a machine of an x86_64 processor with 48 cores and CPU frequency 3.1 GHz.

- Experiment (Exp.) 1: Use the lexical, statistical, and TLD features (i.e., F5, F7, F9-F11) only, while ignoring the WHOIS features. (This experiment is equally applicable to D_2, which is not reported owing to space limitation.)
- Experiment (Exp.) 2: Use the WHOIS features (i.e., F1-F4), while ignoring all other features.
- Experiment (Exp.) 3: Use both lexical and WHOIS features (i.e., F1-F5, F7, F9-F11).

Table 3.3 summarizes the experimental results with a range of classifiers and the actual time spent on training a model and classifying the entire test set. We make several observations. First, for a specific classifier, using WHOIS features alone (Exp. 2) almost always leads to significantly higher effectiveness than using lexical features alone (Exp. 1), except for Logistic Regression. Second, for a fixed classifier, using both lexical and WHOIS features together (i.e., Exp. 3) always performs better than using lexical or WHOIS features alone. Third, among the classifiers considered, Random Forest performs the best in every metric in each experiment. In particular, Random Forest (i.e., non-linear classifier) achieves a better performance than the Ensemble method because there are classifiers (e.g., Logistic Regression and SVM) that are substantially less accurate than the other classifiers and therefore "hurt" the voting results. Fourth, Decision Tree has the fastest execution time, followed by KNN and Random Forest, while Logistic Regression is the slowest and causes a delay for the voting ensemble. To understand the generalizability, when conducting Exp. 1 on the augmented D_2' with the benign:malicious ratio at 1.25:1, we observe that Random

Table 3.3: Experimental results on dataset D_1' with a range of classifiers (with oversampling), their total CPU times for training and test: Exp. 1 uses lexical features only; Exp.2 uses WHOIS features only; Exp. 3 uses both lexical and WHOIS features.

Exp.	Classifier	ACC	FPR	FNR	$F1$-score	Execution Time(s)
1	RF	0.924	0.150	0.052	0.950	0.48
2	RF	0.977	0.025	0.023	0.985	0.59
3	RF	0.980	0.027	0.017	0.988	0.64
1	KNN	0.887	0.199	0.086	0.925	0.40
2	KNN	0.949	0.034	0.056	0.966	0.25
3	KNN	0.947	0.031	0.060	0.964	0.30
1	DT	0.917	0.151	0.061	0.945	0.07
2	DT	0.973	0.045	0.022	0.982	0.08
3	DT	0.974	0.051	0.019	0.983	0.14
1	LR	0.885	0.216	0.082	0.924	20.30
2	LR	0.883	0.362	0.038	0.926	23.03
3	LR	0.918	0.178	0.051	0.946	44.40
1	SVM	0.888	0.220	0.078	0.925	1.69
2	SVM	0.881	0.373	0.038	0.924	1.68
3	SVM	0.920	0.164	0.054	0.946	2.38
1	Ensemble	0.916	0.171	0.056	0.945	21.40
2	Ensemble	0.962	0. 031	0.041	0.974	24.75
3	Ensemble	0.970	0.035	0.028	0.980	45.70

Forest outperforms other models by achieving a 0.947 accuracy, a 0.066 FPR, a 0.041 FNR, and a 0.947 F1-score.

Insight 8. *COVID-19 themed malicious website detectors must consider WHOIS features; and Random Forest performs the best among the classifiers that are considered.*

3.4 Related Work

Although the problem of COVID-19 themed malicious websites has not been investigated until now, the problem of malicious websites has been studied in the literature prior to the COVID-19 pandemic. The problem of detecting malicious URLs generated by domain generating algorithms has been investigated in [123]. The problem of detecting phishing websites has been addressed via various approaches, including: the descriptive features-based model [46], the lexical and HTML features-based model [31], the HTML and URL features-based model [115], and the natural language processing and word vector features-based model [187]. The problem of detecting malicious websites has been addressed via the following approaches: leveraging application and network layers information [222], leveraging image recognition [126], leveraging generic URL features [93, 135], leveraging character-level embedding or keyword-based recurrent neural networks [2, 206, 238], the notion of adversarial malicious website detection [223]. However, these studies do not consider features pertinent to the COVID-19 pandemic, which are we leverage. Nevertheless, the present study fall under the umbrella of cybersecurity data analytics [37, 145, 172, 227, 241, 242], which in turn belong to the Cybersecurity Dynamics framework [34, 35, 169, 230, 247].

3.5 Chapter Summary

We have presented the first systematic study on *data-driven* characterization, and detection of COVID-19 themed malicious websites. We presented a methodology and applied it to a specific dataset. Our experiments led to several insights, highlighting that attackers are *agile*, *crafty*, *economically incentivized* in waging COVID-19 themed malicious websites attacks. Our experiments

show that Random Forest can serve as an effective detector against these attacks, especially when WHOIS information about websites in question is available. This highlights the importance of domain registrars to collect more information when registering domains in future.

CHAPTER 4: SUPPORTING LAW-ENFORCEMENT IN COPING WITH BLACKLISTED WEBSITES

Cyber attackers have long abused web domains and URLs to carry out various attacks such as Phishing, web scamming, and malware attacks. In order to defend against these attacks, URL blacklisting has been widely used. However, this approach has significant weaknesses, especially from a law-enforcement policing point of view. In particular, the law-enforcement does not know what to do with a blacklist because it is unclear what needs to be done (e.g., shutting down a host or domain) due to the subtleties associated with the problem. Predictive policing on the blacklist is one way to deal with this problem and minimize the cyberattacks proactively [89, 156, 167]. In order to help the law-enforcement in dealing with blacklisted websites, we propose a novel framework based on Machine Learning (ML) while providing the law-enforcement with probabilistic classification and interpretability of the predictions made by the interpretable model. Our probabilistic classification and interpretability measures provide a basis for law-enforcement trustworthy decision-making against the attacker-owned malicious websites with proper justification and remove the black-box nature of traditional ML-based approaches. Experimental results show that the framework is practical and has further potential to tackle website maliciousness.

4.1 Chapter Introduction

Websites have been widely abused as a medium for propagating cyberattacks [41, 135, 222]. One simple defense against these threats is to use URL and domain blacklists, which are client-side interventions and often provided by third-party vendors (e.g., Phishtank, Google Safe Browsing, URLhaus). However, this does not completely eliminate the threat because some users may not use such services and the malicious or compromised domains or hosts are still on the loose. Moreover, these blacklists are far from perfect [102] because they are neither *complete*, meaning that they do not contain all of the malicious websites [17,99,200], nor *accurate*, meaning that they contain many false-positive websites (including the compromise ones that were malicious in the past but have

already been cleaned up) [222]. Additionally, there are website domains those are often reused in blacklists within multiple URLs in between days, which indicates domain-level intervention is necessary to mitigate those attacks. Another defense is to use Machine Learning (ML) models to proactively detect malicious websites (see, e.g., [79, 135, 222]). Moreover, there are also some third-party vendors (e.g., Netcraft [1]) providing takedown services on user requests for protecting against cybercrimes (e.g., cybersquatting). of abusing domains that are imitating a user's brand to provide user protection against cybercrimes.

However, there is one important perspective that has not been investigated in the literature, namely *law-enforcement*, as evidenced by FBI shutting down botnets [97]. We envision that law-enforcement will be, if not already, authorized to take actions against malicious websites. This introduces a new dimension of the problem because the law-enforcement must treat detected malicious websites carefully. For example, the law-enforcement can be authorized to shut down a malicious website owned or operated by a malicious party, but may only be authorized to notify the owner or the operator of a website which itself is compromised and then abused by an attacker to wage further attacks. Moreover, oftentimes we observe that attackers reuse the same domains and hosts for new URL based attacks causing the same domain or hostname to appear in a URL blacklist [150]. This phenomena further encourages us to consider the law-enforcement perspective, as more higher level such as domain or host level intervention is more effective than client-side interventions. This call for studies on helping the law enforcement in distinguishing between malicious (i.e., attacker-owned) and compromised (i.e., legitimate party-owned) websites to take actions.

Chapter Contributions. In this paper we make three contributions. First, we initiate the study of the law-enforcement perspective when coping with malicious websites. This turns out to be a challenge because of the dynamic nature of web domains and complexity of web hosting infrastructures. This prompts us to introduce a novel framework to help the law-enforcement to cope with malicious websites. The framework highlights the importance of using *interpretable* (i.e., explainable) ML, while considering the probabilistic *uncertainty* associated with the prediction

outcomes of ML models. The framework integrates a ML intepretability system, such as InterpretML [164], to provide explanations and probabilistic predictions to the law-enforcement (e.g., why is a website predicted as malicious, and what is the likelihood it is indeed malicious?).

Second, we investigate how to choose the entity for action: domain vs. hostname. To our knowledge, this is the first time to propose a principle method for making such decisions.

Third, we conduct a case study on evaluating an instance of the framework with a real-world URL blacklist. Experimental results show that we achieve a 86% accuracy with a 0.92 F-1 score, while providing local explainability (i.e., interpretation) for the individual prediction outcomes for each input blacklisted website.

Chapter Outline. Section 4.2 presents the problem statement and background. Section 4.3 presents our framework and novel techniques. Section 4.4 describes our case study and results. Section 4.5 discusses related prior studies. Section 4.6 summarizes the chapter.

4.2 Problem Statement

Suppose the law-enforcement is authorized to take actions against malicious websites. The problem is: *Given a URL that is blacklisted (i.e., deemed malicious), what should the law-enforcement do?* While intuitive, there are technical subtleties.

4.2.1 Technical Subtleties Encountered by Law-Enforcement

Subtlety 1: URL structure is complicated. Figure 4.1 illustrates the URL structure, including: a *protocol* name (`https` in this example), a *hostname* (`mail.example.com`), and a *URL path* (`mail/u/0`) possibly along with some queries within the URL path. A hostname consists of a *domain name* (`example.com`, also refereed to as a 2LD) and possibly a *subdomain name* (`mail`, referred to as a 3LD). A hostname is sometimes referred to as a Fully Qualified Domain Name (FQDN) [146] and mapped to one or multiple IP addresses by the DNS server, while noting that one IP address may be mapped to multiple hostnames (i.e., shared hosting [161]). A domain name must contain a *Top-Level Domain* (TLD) (e.g., `.com`) within it. A higher level subdomain,

say Fourth-Level-Domain (4LD) (e.g., `z.x.example.com`), could be resolved to the same IP address as 3LD `x.example.com` or 2LD `example.com`. In this paper, we refer to the 3LD or higher level (if present) of a URL as *hostname* and the 2LD as *domain name*, while noting that these two become the same in the absence of a subdomain name within the website URL in question.

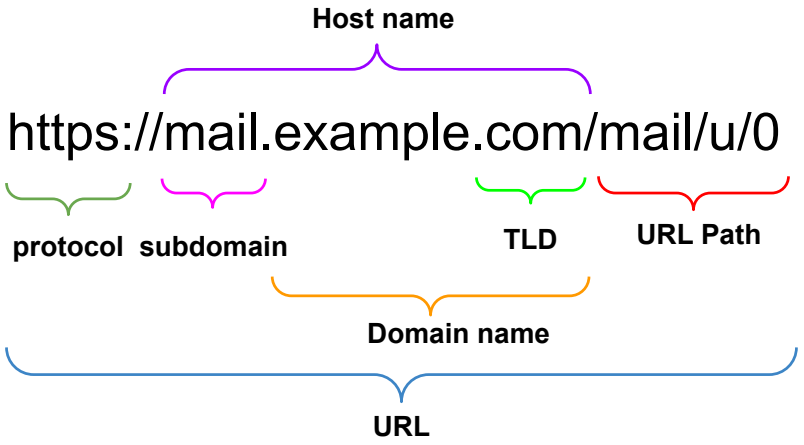

Figure 4.1: Illustration of the structure of URLs

The complexity of URL structure makes it unclear what the law-enforcement should do to a malicious website or URL. To see this, consider a URL with a path name. In this case,

- shutting down the specific URL (including the path name), for example by filtering web traffic corresponding to the URL, may not be effective because the attacker can easily create other URL paths with the same hostname;
- shutting down the corresponding port is no good idea because the host (i.e., web server) corresponding to the URL may be malicious (i.e., the host can continue to wage attacks via other ports);
- shutting down the entire domain (e.g., `example.com` in this case) or the entire hostname (e.g., *mail.example.com*) corresponding to the URL without considering its ownership is no good idea because there might be many benign subdomains and URLs blueassociated with the domain or the hostname, which will be affected and deemed as false positives.

To further illustrate the problem, let us look at the structure of a website using web-hosting vs.

domain-hosting as shown in Figure 4.2. In this example, the host `sites.example.com` is not malicious and should not be shutdown even though some URL(s) with the hostname are malicious. In the case of domain-hosting shown in Figure 4.2(B), multiple hostnames are created under the benign domain `example.com`. If one hostname, say `site1.example.com`, is created by an attacker to publish malicious contents by abusing the hosting service, then `site1.example.com` should be shut down but the other subdomains. However, if domain `example.com` is not associated with any hosting service, then we can safely assume that `example.com` is owned and/or operated by an attacker and therefore should be shut down.

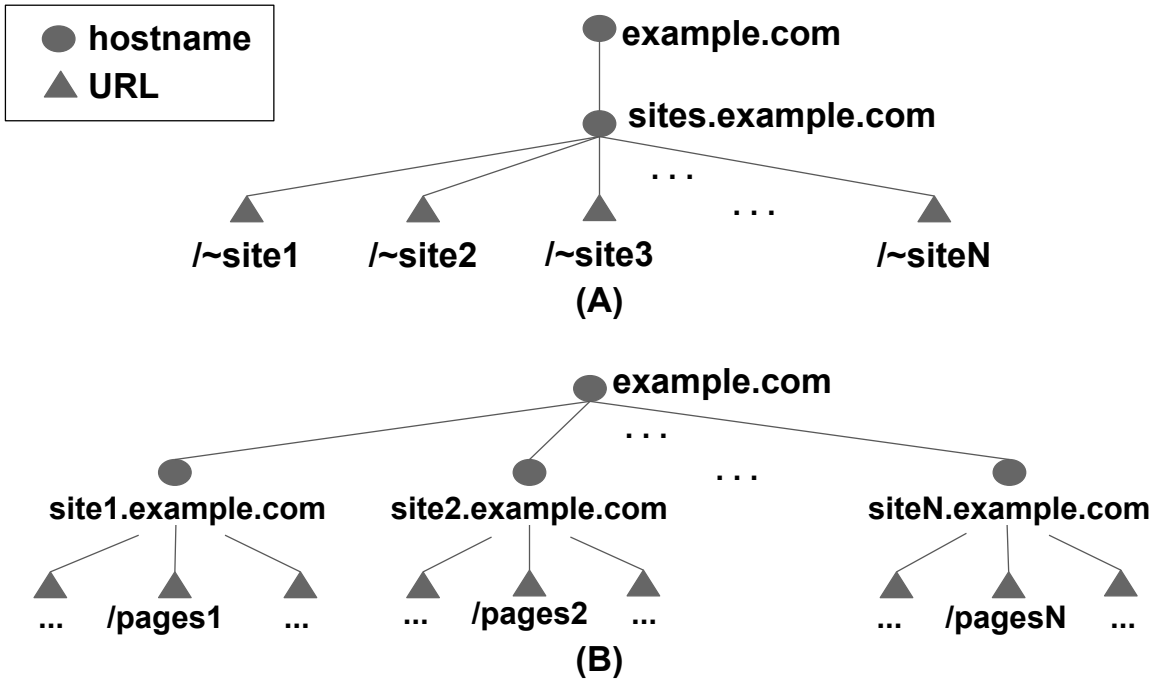

Figure 4.2: The structural difference between (A) web-hosting and (B) domain-hosting within an example domain named `example.com` (adapted from [39])

Subtlety 2: Trustworthiness of given malicious websites. Blacklists may contain false-positives [222], meaning that the law-enforcement cannot blindly trust or shut down blacklisted websites. Instead, the law-enforcement must leverage other means to examine whether a blacklisted entity (i.e., host or domain) is indeed malicious before taking any actions.

Subtlety 3: IP Address-based blacklisting is no good idea. Nowadays sharing IP addresses and hosting is very popular. With shared IP addresses, many domains may resolve to the same IP

address. If one of the hostnames is malicious, it does not mean the other hostnames that resolve to the same IP address are malicious. This makes it challenging to use IP address-based blocking [155, 161, 213]. In order to further highlight the ambiguity, we randomly pick a hostname `mail[.]weddingstaffcompanies[.]com`[1] from the publicly available PhishTank blacklist; the hostname is labeled as *suspicious* by McAffe PC security when accessed from Google Chrome. The hostname is resolved to IP address is 207.38.88.153. Then, we do reverse DNS lookup to get a FQDN `usloft5543[.]serverprofi24[.]com`, which is different from the input hostname. In order to see what other domains or hostnames are mapped to IP address 207.38.88.153, we query Robtex.com and find that at least 44 domains or hostnames are associated with it. However, we do not find any other domains or hostnames associated with this IP address in the blacklist. In this case, if the law-enforcement blocks IP address 207.38.88.153, then the other 43 hostnames will be affected, which is not justified. Therefore, IP address-based blocking is no feasible solution.

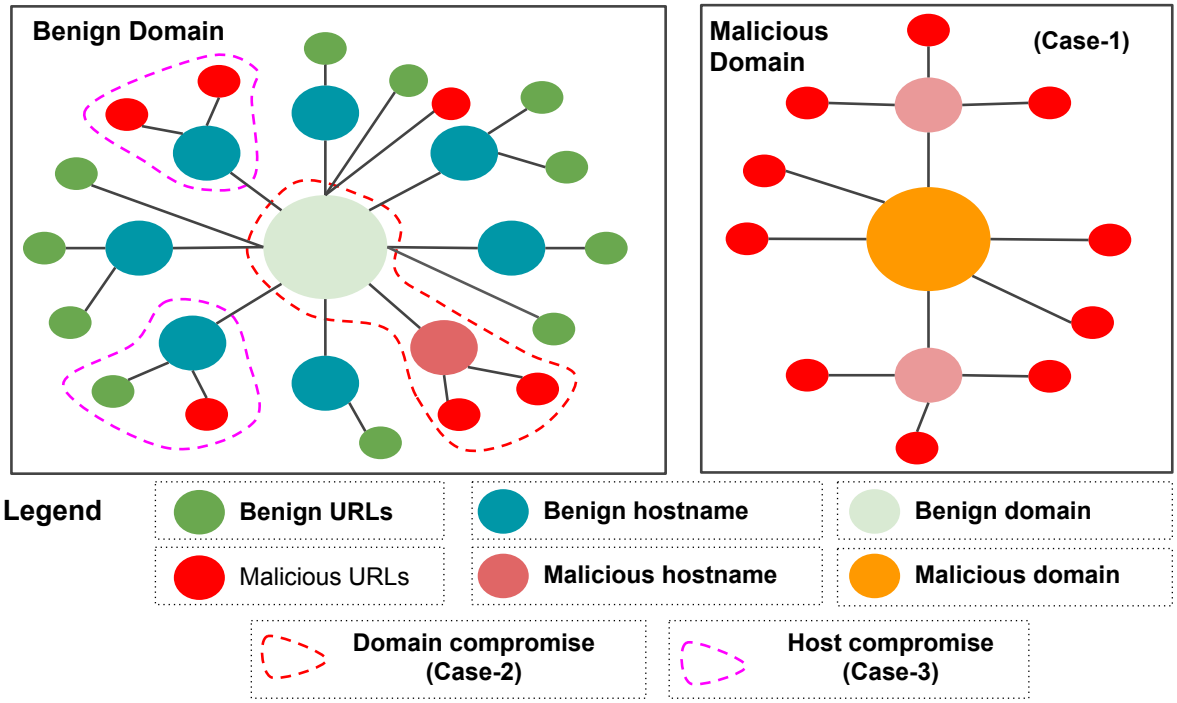

Figure 4.3: Structure of websites including domain, hostname, and URL(s)

Subtlety 4: Complications encountered when taking actions against malicious domains or

[1][.] is used to safeguard reader from clicking possibly malicious website

hostnames. Figure 4.3 highlights the mapping of hostnames and URLs in a domain name. It shows multiple cases: (case-1) The law-enforcement encounters a malicious domain and associated malicious hostnames and URL paths. The law-enforcement can justifiably shut down the domain. (case-2) The law-enforcement encounters a benign domain that has been compromised to create malicious hostname. The law-enforcement needs shut-down the malicious hostname and notify the legitimate domain owners to let them clean up the compromises. (case-3) The law-enforcement encounters a benign hostname associated with a benign domain which is compromised and abused to create malicious URLs. The law-enforcement should notify the hostname owner to clean up and put the URL in blocklist without shutting down the domain or hostname. In summary, it is essential to support the law-enforcement with various kinds of details.

4.2.2 Research Questions (RQs)

The preceding subtleties prompts to revise the research problem as as follows: This leads to the following research questions (RQs) regarding a blacklisted URL:

- RQ1 (*URL characterization*): At what entity level(s), such as domain and/or host, should the law-enforcement take actions?
- RQ2 (*quantitative classification*): What is the likelihood that the entity (e.g., domain or hostname) corresponding to a blacklisted URL is malicious or victim (i.e., compromised and abused to wage attacks)?
- RQ3 (*prediction interpretability*): Why an entity associated with a blacklisted URL is predicted as malicious or compromised?
- RQ4 (*law-enforcement actions*): What should the law-enforcement do when the answers to RQ2 and RQ3 are not satisfactory or convincing?

4.3 Framework

To address the RQs mentioned above, we propose a framework, which is highlighted in Figure 4.4. The framework has the following modules: *blacklist collection*, *characterizing*, *labeling*, *feature analysis & extraction*, *training interpretable ML models*, *probabilistic classification*, and *decision-making*. At a high level, these modules work together to address the aforementioned RQ1-RQ4 as follows. To address RQ1, we propose extracting the hostname and domain namefrom a given blacklisted URL, while finding out if the domain or hostname is associated with any known hosting service. In practice, one of the two following scenarios happen often: (i) the law-enforcement is often given blacklisted URLs without being told why and how the URLs are blacklisted; (ii) the law-enforcement is told how the URLs are blacklisted but without being given any explanation on why they are deemed malicious and/or the probability that the associated websites are indeed malicious. This prompts us to propose that the law-enforcement should build their own systems to analyze blacklisted URLs to address RQ2-RQ4. In particular, addressing RQ2 requires to quantifying the probabilistic uncertainty (e.g., probabilistic class prediction) associated with ML models and addressing RQ3 requires to using interpretable ML models for explaining individual prediction outcome. Lastly, addressing RQ4 helps the law-enforcement in dealing with truly malicious or attacker-owned entities.

Notations. A URL blacklist $\mathcal{L}$ contains n URLs, denoted by $\mathcal{L} = \{u_1, \ldots, u_n\}$. For URL $u_i \in \mathcal{L}$, we denote the associated hostname by h_i and domain name by d_i, a corresponding entity for i-th URL is denoted as en_i. Table 4.1 summarizes the main notations used in the paper.

4.3.1 Blacklist Collection Module

This module collects the selected URL blacklist $\mathcal{L}$, which does not contain any labels (e.g., malicious, compromise) for the URLs. We observe the trends and distributions of the URLs within the blacklist to better understand the characteristics and size of the blacklist.

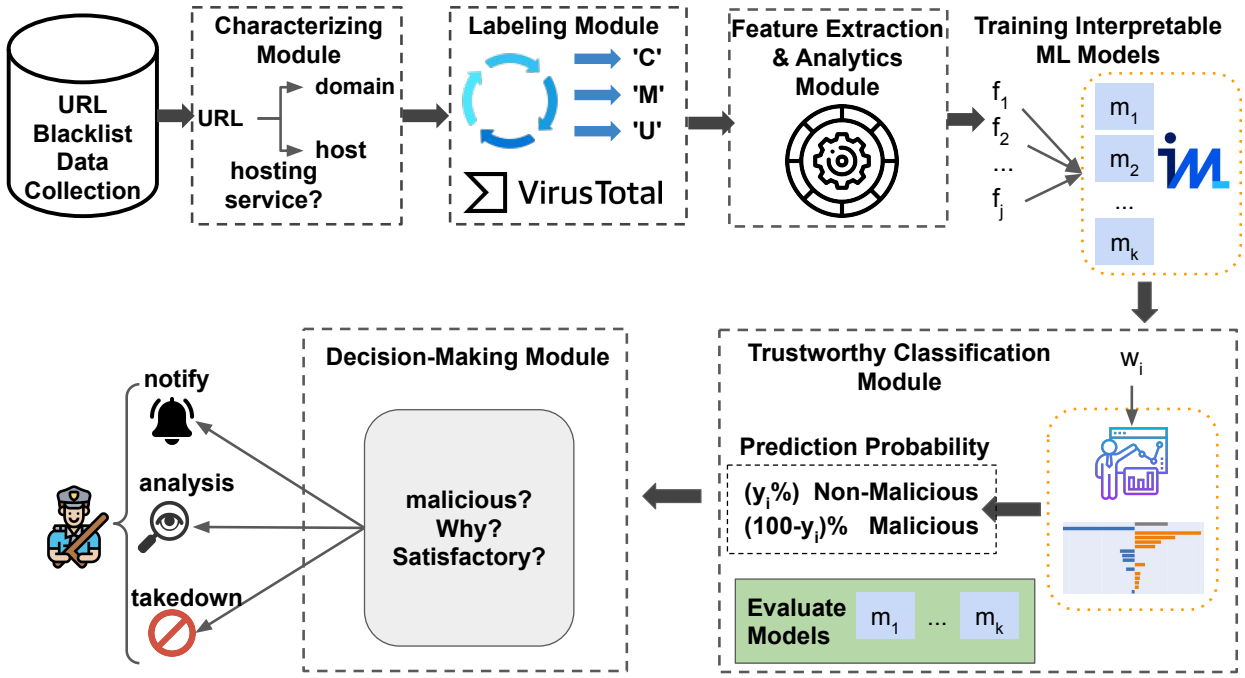

Figure 4.4: The Framework with six modules.

Table 4.1: Summary of notations used in the paper

Notation	Meaning
$u_i, \mathcal{L}$	i-th URL on blacklist $\mathcal{L}$
d_i, h_i, en_i	Domain, host, and entity for the URL u_i, respectively
$\mathrm{CL} = \{\mathrm{C}, \mathrm{M}\}$	Class label compromised (C) vs. malicious (M)
y_i	Predicted probability for maliciousness of entity en_i
$\mathcal{M}$	Any supervised machine learning model
F_i	Feature vector for i-th entity
E_i	Explanation set for i-th entity
ϕ_j^i	Impact of the j-th feature on classification of i-th entity
$\phi_{j,k}^i$	Impact of the j-th and k-th features together on classification of i-th entity
D_{entity}	Total unique entities (websites) without association to any public/private hosting services
$D_{labeled}$	Labeled ground-truth entities

4.3.2 Characterizing Module

This module characterizes the URLs on a blacklist to select the appropriate entity en_i based on their domain structure. It should first extract the hostname h_i and the domain name d_i from the input blacklisted URL u_i. Then, we check if the hostname h_i (3LD or higher level subdomain) is associated with any known hosting services, if yes then we notify the hostname for cleaning up and no quantification is necessary; otherwise, we check if the domain name d_i is associated with any known hosting services, if not then the selected entity $en_i = d_i$; otherwise then we further check if $d_i = h_i$ (meaning hostname same as domain or no subdomain in URL), if yes, then we notify the domain and no quantification is required (because domain is legitimate, URL is created with malicious path); otherwise, when $h_i \neq d_i$, we select entity $en_i = h_i$ for quantification. The flow chart is presented in Figure 4.5. The URLs that are characterized as to quantify hostname or quantify domain names are the ones compiled as the $D_{unlabeled}$, and going as an input to the next module.

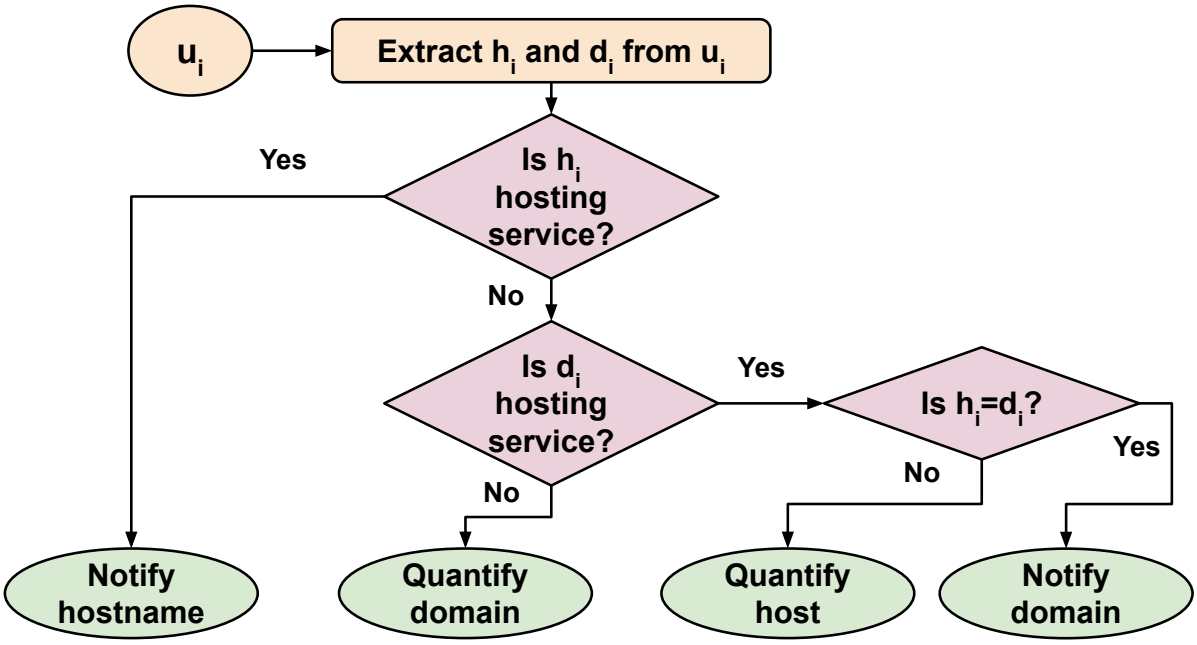

Figure 4.5: Flowchart for Characterizing Module.

4.3.3 Labeling Module

This module produces a labeled dataset for training ML models, ideally automated. This module takes a hostname or domain name as input, depending on the entity chosen to quantify by the characterizing module. It labels the entity via a third-party threat engine such as the VirusTotal [208]. There may be four possible labels: malicious (M); compromised (C) meaning that the entity is itself a victim, namely compromised and then abused to wage attacks; unknown (U); and not available (N). It may be necessary to preprocess the input from such third-party services, and the preprocessing method may be specific to the third-party services. In any case, we propose only considering the malicious (M) and the compromised (C) labels for analyzing the law-enforcement perspective.

4.3.4 Feature Extraction & Analytics Module

Feature Extraction

This module extracts feature representations of URLs, including but are not limited to what have been extracted by the *characterizing* module. Denote the resulting feature representation of URL u_i by $F_i = \{f_1^i \dots f_k^i\}$, where k is the total number of features. We propose using the following features:

1. Brand-name in hostname and domain name ($f_1 \in \{0, 1\}$): This feature indicates whether a hostname and domain name string contains popular brand names according to some list(s) of reputable domains, such as Alexa [91] or Tranco [106]. Any domain or hostname involved in the a blacklist should have $f_1 = 0$.

2. Twisted brand-name in hostname ($f_2 \in \{0, 1\}$): This feature indicates if the domain name or hostname contains any twisted string of the top 5k brand names, such as typo-squatting, combo squatting, or homographing. For example, `amaz0n-pay.example.com` contains a twisted version of a top-brand name *amazon* and therefore has been assigned value 1, otherwise 0.

3. Number of dots in a hostname (f_3): This is the number of dot characters (i.e., '.') in a hostname. For example, hostname `ab.c-d.df.com` has 3 dots.

4. Number of hyphens in hostname (f_4): This is the number of hyphen characters (i.e., '-') in the hostname. For example, hostname `ab.c-d.df.com` has 1 hyphen.

5. Digit ratio in hostname (f_5): This is the ratio of the number of digits in a hostname to the length of a hostname. For example, hostname `a12.c34-d1.df.com` has 5 digits, meaning a digit ratio of $5/17$.

6. Number of unique alphabetic-numeric characters in hostname (f_6): This is the number of unique alphabetic characters or digits in a hostname except the TLD part, which is excluded because it often consists of letters. For example, hostname `ab12.c34-d1.df.com` has 9 unique alphabetic-numeric characters (i.e., `a, b, c, d, f, 1, 2, 3, 4`) other than the TLD (`.com`).

7. Hostname length (f_7): This is the length of a hostname.

8. Number of tokens in hostname (f_8): This is the number of tokens in a hostname after tokenizing with hyphen '-' and dot '.', except the TLD part which is excluded because it is typically one token. For example, hostname `ab12.c34-d1.df.com` has 4 different tokens (i.e., `ab12`, `c34`, `d1`, and `df`). This feature highlight the hostnames that contain many '-' and/or '.' characters.

9. Length of the longest token (f_9): This is the length of the longest token in a hostname. For example, the longest token `ab12.c34-d1.df.com` is 4 (i.e., `ab12`).

10. Number of redirects (f_{10}): Attackers often use redirects to deceive victims. This feature reports the number of redirects associated with a hostname.

11. Number of passive DNS queries (f_{11}-f_{12}): These features describe the number of records found for individual DNS record types, such as DNS 'A' and 'NS' records in the global

passive DNS database from CIR.CL [48], respectively. For a passive DNS query, an 'A' record indicates an IP addresses and a 'NS' record indicates the corresponding name server. These features report the record counts for each record type in the passive DNS database. A high number in these records indicate the website has more historical presence in the Internet, thus deemed more reputable.

12. Presence of self-resolving name servers ($f_{13} \in \{0, 1\}$): It indicates if the associated domain has any self-resolving name server or not. For example, if domain `'example.com'` has name server `ns1.example.com`, the domain is self-resolving and this feature value is 1; otherwise, its value is 0.

13. Domain ranking (f_{14}):This numeric feature measures the reputability of a domain name with respect to some list of reputable websites (e.g., Tranco [106]). In our case study we will present an example.

14. Hostname ranking (f_{15}): This numeric feature measures the reputability of a hostname with respect to some list of reputable hosts (e.g., Tranco [106]). It can be assigned in the same fashion as f_{14}. In our case study we will present an example.

15. Number of subdomains (f_{16}): This is the number of subdomains associated with a domain name corresponding to a URL. For example, given domain name $d =$"`wixsite.com`", one can query all active subdomains of the form of sub`.wixsite.com` and then count the number of such subdomains.

After extracting these features, it is worth mentioning that a preprocess may be needed to eliminate the correlated features (if applicable) before training ML models, because it is well-known that highly correlated features affect model predictions. Removing unnecessary features may also improve ML models' performance. Moreover, other blacklist dataset specific features can be added to complement the existing generic features for extending the framework.

4.3.5 Training Interpretable ML Model Module

This module trains for a ML model $_x$ to predict any given URL for law-enforcement purposes. Here, x denotes the corresponding interpretable ML model. It takes feature vectors as input to train a ML classifier with uncertainty quantification through probability, while providing interpretability of predictions. We reiterate that one should use the ML methods that (i) are interpretable, so that the law-enforcement can understand why a URL is deemed malicious, and (ii) can quantify the confidence or uncertainty associated with a prediction. Both are important for justifying law-enforcement actions against a blakclisted website.

4.3.6 Probabilistic Classification with Interpretation

A trained ML model makes probabilistic predictions on URLs, while interpreting its predictions. The probabilistic classification can be evaluated using the standard metrics, such as accuracy, AUC score, precision, recall, and F-1 scores [169]. For a given feature vector representing an entity, the law-enforcement uses the classifier to predict the probability that the corresponding entity (i.e., host or domain) as malicious (M) or compromised (C). Moreover, there will be a explanation set $E_i = \{\phi^i_j\}_j \cup \{\phi^i_{j,k}\}_{(j,k)}$ corresponding to the i-th entity. Both the probabilistic predictions and interpretations will be leveraged by the *Decision-Making* module.

4.3.7 Decision-Making Module

This module leverages the output of the probabilistic classification and interpretation module to help the law-enforcement make decisions. At a high level, if an entity associated with a URL, is predicted as malicious (M) with a high probability and a satisfactory interpretation, the law-enforcement should shut down the entity. If it is predicted as compromise (C) with high probability and a satisfactory interpretation, then the law-enforcement should notify the corresponding host or domain owners to clean up. Otherwise, if the probability is not high (e.g., $< 70\%$) or the interpretation is not satisfactory enough then the law enforcement should send the URL to human analysts for further analysis.

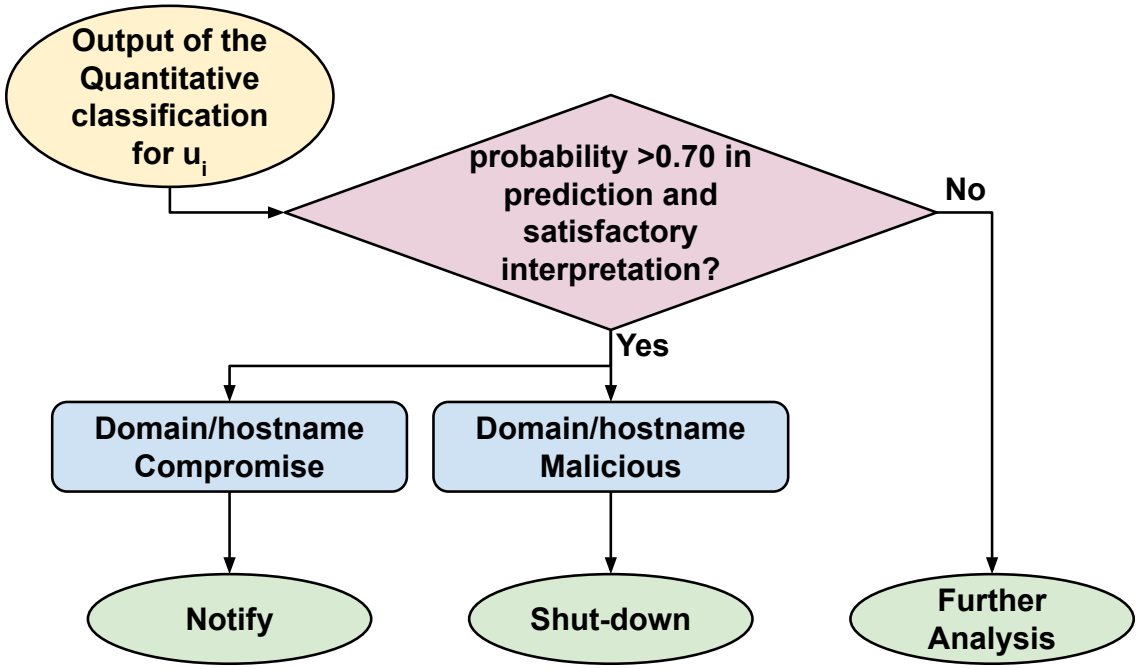

Figure 4.6: Flowchart of the decision-making Module.

Figure 4.6 highlights the flowchart, which can be understood via the following example. If the length of an entity's name string is the main contributor to the maliciousness prediction and the length is less than 10 characters, then it is not conclusive to shut down the entity. If the number of unique alphanumeric characters contained in the entity's name string is an indicator and this number is large, while there are other indicators (e.g., a large digit ratio, a large number of tokens in the entity's name, a large number of redirects, or a large number of passive DNS queries), then it becomes satisfactory that the website is malicious and should be shutdown.

4.4 Case Study

Now we present a case study on applying the framework to help the law-enforcement to cope with blacklisted URLs. For this purpose, we use a blacklist, *PhishTank* because it is open-sourced and easy to reproduce research results and widely-used in literature for similar research purposes. Moreover, PhishTank does not provide any categorization for either malicious or compromised websites, thus making it suitable for the proposed *law-enforcement* perspective.

4.4.1 Blacklist Collection

We collect the blacklisted URLs from PhishTank during May 4 - 10, 2021. These 7 days blacklist URL distributions are presented in figure 4.7, where we have collected 12,843 unique unlabeled URLs referred as D_{all} within these seven days combined, which is used as an input for the labeling process. The figure 4.7 also depicts that there are on average around 9,700 unique URLs each day that remains constant (e.g., duplicate URLs) within the blacklist where the number of newly added (e.g., new URLs) or deleted (e.g., Removed URLs) URLs on each day is ranged between 400 to 800 unique URLs. The cumulative is calculated by adding the 'new URLs' on each day but without removing any of the 'removed URLs' from the blacklist.

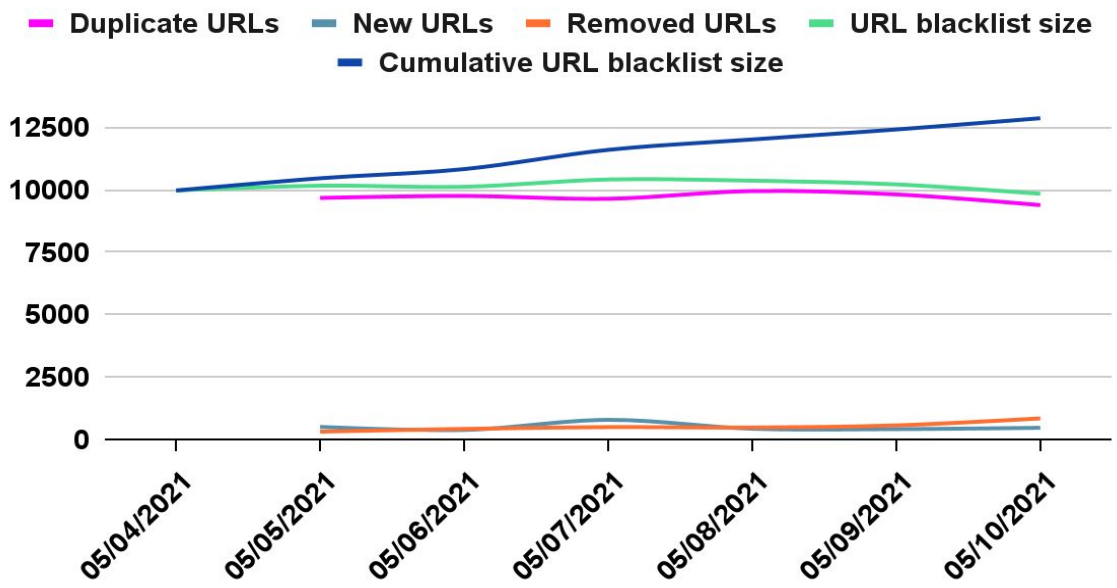

Figure 4.7: The distribution of the URLs in PhishTank blacklist during 7 days (May 4 - 10, 2021).

4.4.2 Characterizing Module

This module answers RQ1 by characterizing the URLs to determine the appropriate entities on which the law-enforcement should act upon based on the associated hosting services. In this study, we curate a list of 61 publicly known reputable hosting services. We extract hostnames and domain

names from the 12,863 URLs, leading to a total of 7,482 unique hostnames. This means that many URLs have the same hostname but different paths. Among the 7,482 hostnames, there are 8 hostnames that are directly related to some known hosting services and they are notified of the URLs if encountered. For the rest 7,474 (= 7,482-8) hostnames, 2,707 are only domain name (i.e., meaning $h_i = d_i$); among the rest 4,767 hostnames (where $h_i \neq d_i$), 1,590 have their domains associated with hosting services, meaning that the law-enforcement should select entity $en_i = h_i$ to take actions; for the rest 3,177 (=4,767-1,590) hostnames, their domains are not related to any hosting services, meaning that the law-enforcement should select entity $en_i = d_i$ to take actions. In total, we get 6,195 unique entities, where 4,605 are unique domains (derived from the 5,884 = 3,177+2,707 hostnames) and 1,590 are unique hostnames to quantify through the framework.

4.4.3 Labeling Module

In our case study we leverage VirusTotal to infer the labels of the selected entities corresponding to the URL in PhishTank blacklist because not all entities are bad. We use the following heuristics to approximate the ground-truth dataset of the URLs on $\mathcal{L}$. For a given URL, we use its entity (i.e., hostname or domain name), which is the output of the *characterizing* module, to query VirusTotal. An entity is deemed malicious (M) if 3 or more VirusTotal detectors say it is malicious; an entity is deemed compromised (C) if no VirusTotal detectors deem it as malicious; otherwise, an entity is disregarded because we consider binary classification. In our experiment, the initial blacklist size is $|\mathcal{L}| = 12{,}863$, from which we obtain a total of 6,195 unique entities, which (equivalently, their corresponding URLs in $\mathcal{L}$) are denoted by D_{entity}. Among the 6,195 entities, 968 are labeled as compromised (C) denoted by D_{com} and 4,017 are labeled as malicious (M) denoted by D_{mal}, leading to a total of 4,985 entities, denoted by $D_{labeled} = D_{com} \cup D_{mal}$; while noting that the other 1,210 entities are disregarded.

4.4.4 Feature Extraction & Analytics Module

Extracting Feature Values. For the feature of brand name in hostname (f_1), we curate a list of top 5,000 domains from Tranco [106] on the day of blacklisting and extracted the domain part (e.g., "`example`" from `example.com`) as the brand name. We only consider the brand names with string length greater or equal to 4 because shorter ones can include a lot of noises (e.g., 'fb' is a brandname of Facebook but other benign legitimate domains / hostnames such as 'fbox' could also include 'fb' as a sub-string, which causes ambiguity). If any brand name is present in the hostname or (sub)domain name, we set feature $f_1 = 1$, and $f_1 = 0$ otherwise. For feature f_2, we use dnsTwist [59] to generate typo-squatted domain names based on the top 5,000 brand names mentioned above. This leads to 24,213,971 twisted domains. Among these twisted names, we keep the ones with length greater than or equal to 4 for the same reason as mentioned above and assign $f_2 = 1$ if twisted brand name is present, and $f_2 = 0$ otherwise.

For deriving the values of features f_3 to f_9, we use the hostname and domain name contained in $D_{labeled}$. For feature f_{10}, we use the python `requests` module with an entity en_i as input. If entity is unreachable, then we set $f_{10} = 0$. For features f_{11} and f_{12}, we query the CIR.CL passive DNS database [48] for the 'A' record and 'NS' record corresponding to the entity in $D_{labeled}$, which. gives us a hint on the corresponding host's or domain's past activities. We sum up all the 'A' and 'NS' record counts for f_{11} and f_{12}, respectively. For feature f_{13}, we use the python module `dns.resolver` to resolve the name servers corresponding to the domain name and cross-check if the self-resolving name server is present ($f_{13} = 1$) or not ($f_{13} = 0$).

For feature f_{14}, we extract the Tranco rank list corresponding to the dataset time-frame. We set f_{14} to be the rank of the domain name if it is on the list of Tranco; otherwise, we set is to be the lowest rank or 6,000,000 because it is the size of the Tranco list. For feature f_{15}, we also use the Tranco list and the hostname for determining the value of f_{15}.

For feature f_{16}, we rely on a third-party open-source penetration testing tool *sublist3r*, which uses OSINT [140] to query from search engines (e.g., Yahoo, Bing, Baidu, and Ask) and other threat intelligence feeds (e.g., Netcraft, ThreatCrowd, DNSDumpster, and ReverseDNS).

Feature Values Analysis. We first analyze the correlations between features. By analyzing the Pearson correlation [19] among the numeric features, we find no significant correlations between the numeric features since their correlation coefficient is less than 0.5 between all pair of features.

4.4.5 Training Interpretible ML Model Module

We use the labelled data, $D_{labeled}$, for training and testing interpretible ML classification models. We randomly select 80% of data, $D_{train} \in D_{labeled}$, for training and the remaining 20% of data, $D_{test} = D_{labeled} - D_{train}$, for testing. The training set ($|D_{train}|$=3988) contains 3,219 malicious entities and 769 compromised entities, while the test set ($|D_{test}|$=997) contains 798 malicious entities and 199 compromised entities.

In our experiments, we consider the following ML models: Explainable Boosting Machine (EBM) [96, 129], Decision Tree (DT) and Random Forest (RF). Since the last two models are well-known, we only briefly review EBM. EBM a tree-based cyclic gradient boosting model [129], which improves the generative additive mode (GAM) [239]. In a GAM, the model outcome is $f(\mathcal{E}[y_i]) = \beta_0 + \sum \phi_j^i$, meaning that summation of individual feature's contribution is added to the model along with a co-efficient β_0; in EBM, the model outcome is $f(\mathcal{E}[y_i]) = \beta_0 + \sum \phi_j^i + \sum \phi_{j,k}^i$, which contains another summation of the pair-wise interactive feature contributions $\phi_{j,k}$ where $j \neq k$. These feature contributions can be deemed as the explanation set, denoted as $E_i = \{\phi_1, \ldots, \phi_{16}, \phi_{1,2}, \ldots, \phi_{15,16}\}$ where 16 is the total number of actual features for this case study, which quantifies the contributions of individual features as well as pairs of features for the predicted label of entity en_i. Here, EBM is chosen because EBM provides the probabilistic quantitative classification along with the interpretability at the individual prediction outcome both of which are required in the proposed framework. Moreover, explainability provides transparency and trust in classification of highly imbalanced datasets.

4.4.6 Probabilistic Classification With Interpretation Module

This module answers both RQ2 and RQ3 through the use of probabilistic label prediction for a given entity, and the corresponding visualization of an explanation set, respectively. In our experiments, we use the InterpretML platform's visualization to particularly answer the RQ3. Table 4.2 presents the EBM model and the other ML models' (i.e., accuracy, AUC score, weighted precision, weighted recall, and weighted F-1 score). We observe that EBM performs better than the two other models. Moreover, EBM offers interpretations for individual predictions, which is important to the law-enforcement for effective decision-making on certain entities. Therefore, we will focus on EBM in the rest of the paper. The following Decision-Making module provide examples for use cases of the framework.

Table 4.2: Performance Metrics of the ML Models on D_{test}

Models	Acc	AUC	Precision	Recall	F-1 Score
EBM	$85 \pm 1\%$	0.87	0.86	0.98	0.92
RF	$84 \pm 1\%$	0.83	0.88	0.93	0.86
DT	82 ± 1	0.72	0.88	0.88	0.88

4.4.7 Decision-Making Module

This module answers the RQ4 by aiding the law-enforcement with the appropriate decision support based on the satisfaction with the quantitative predictive classification and the visualized explanation set (i.e., interpretation). For blacklisted entity en_i, in this module the law-enforcement receives a prediction probability y_i for the entity to be malicious or compromised, together with an explanation set E_i as visualized through InterpretML platform shows top contributing features for that prediction. In what follows, we use examples to show how the system can indeed support the law-enforcement decision-making process.

Example 1 (Malicious $\rightarrow$ Takedown). There are maliciously registered websites abusing hosting services such as the `inmotionhosting.com`. Figure 4.9 shows one example. In this example, the EBM classifier predicts the hostname in question as malicious (M), with the explanation that several indicators—such as the number of name server records (22272), digits ratio (0.10), entity

length (29.00), unique alphanumeric characters (15.00) —make significant contribution to the malicious prediction; whereas, a few other features—such as the number of A records, number of dots, domain rank, number of tokens, self-resolving NS, max token length, and number of hyphen—contribute to indicate that the hostname is compromised (C). However, we observe that the prediction by the interpretable EBM classifier is made as malicious (label value 1) with a high probability 0.868. As a result of both the high prediction probability and the strong explanations, the law-enforcement should *takedown* this hostname `secure285[.]inmotionhosting.com` and possibly notify the domain owner `inmotionhosting.com` for clean up because it is a hosting service provider.

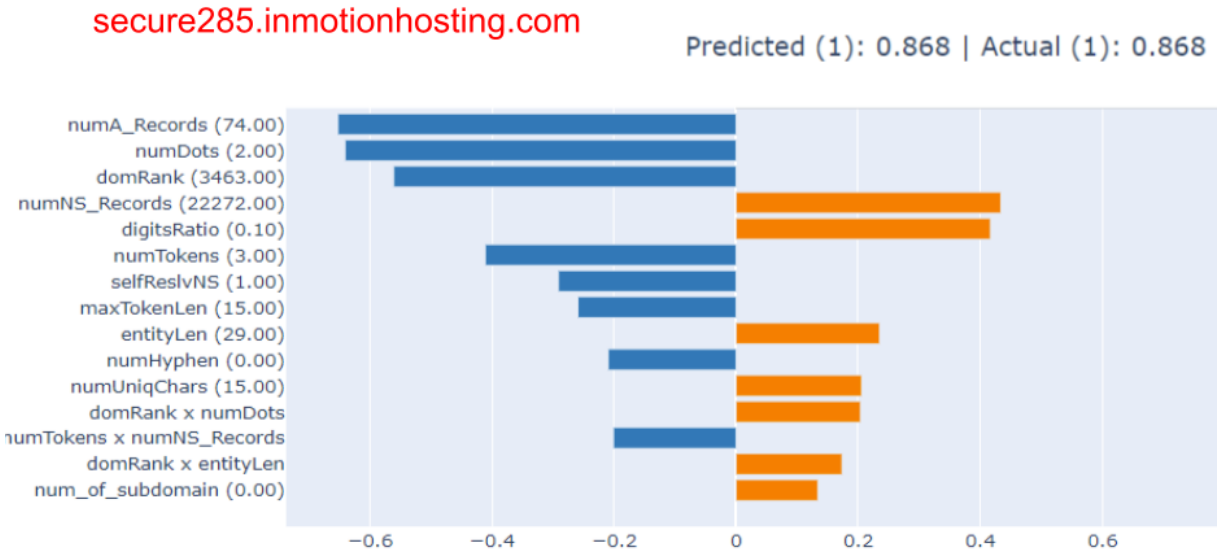

Figure 4.8: Example 1: malicous → takedown.

Example 2 (Compromise → Notify). Oftentimes there are many legitimate business website that gets compromised by attackers and abused to wage cyberattacks, which place them into blacklists. But it is very important that law enforcement identify this legitimate entities and try to notify them asap for cleaning up. Figure 4.9 shows one example of a compromise case where the framework is predicting the input entity `123formbuilder.com`. In this example, the EBM classifier predicts domain name in question as compromised (C) with explanations that indicators—such as hostname ranking (7070), number of A records (96), number of tokens (2.00), number of dots (1.00) —make significant contribution towards the compromise prediction. There are few indicators—such as

digit ratio and (domain rank × entity length) —make some contributions towards the malicious (M) prediction. In this particular case, the prediction probability for class C is 0.916 which is very high and trustworthy along with the strong host rank features. Hence, it is quite trustworthy that the domain name is compromised in this case and thus be notified by law-enforcement.

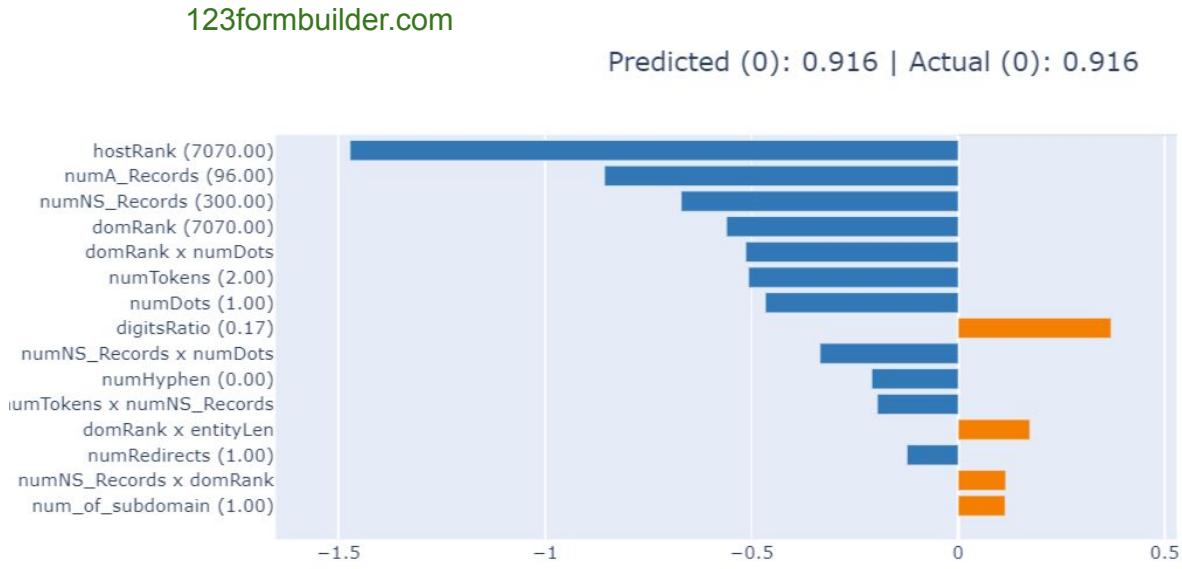

Figure 4.9: Example 2: Compromised → Notify.

Example 3 (Compromise → Further Analysis). In this example, the domain name

`whyymedia[.]com.au`corresponds to available domain under the hosting domain `h2osupportservice`
It is definitely not a malicious website for now, but we do not know for sure it it is a safe site in the future. It is blacklisted by PhishTank. As highlighted in Figure 4.10, the EBM classifier predicts it as compromised (C) rather than malicious (M), while only 3 features contribute to supporting the prediction that the domain is M rather than C. However, the probability of class C prediction is 0.611, which infers there is uncertainty associated with this prediction. Thus based on the decision-making module recommendations as shown in the flowchart described in Figure 4.6, the law-enforcement should not take actions (e.g., notify or takedown) without conducting a further analysis. Our manual verification on a later date, which is more than 10 days after the URL is blacklisted by PhishTank, confirms that the associated domain name is indeed legitimate but may

be taken by attackers in future. Hence, further analysis will justify the corresponding action.

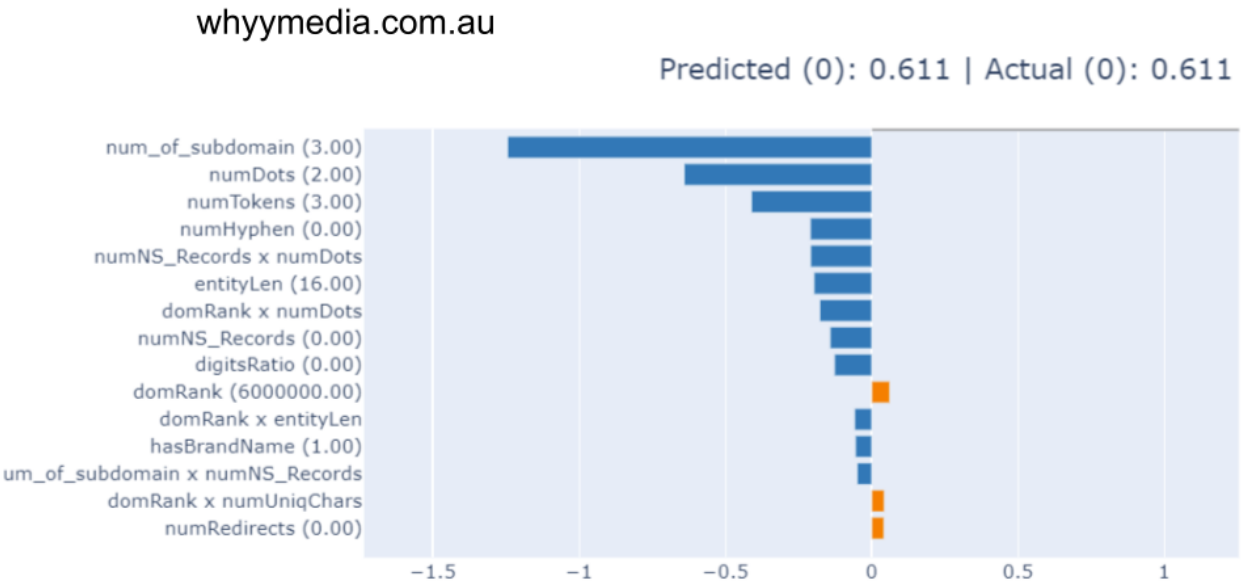

Figure 4.10: Example 3: Compromised → Further Analysis.

Example 4 (Malicious → Further Analysis). Figure 4.11 shows an example, where `sctrlgin[.]com` is blacklisted by PhishTank. The EBM classifier predicts it as malicious (M) with a probability of 0.675. Moreover, the classifier does not give a convincing interpretation on the prediction. Specifically, only few features—such as the domain rank (6000000) and number of redirects (0.00)—contribute to predicting it as M, which is not convincing enough. This would give the law-enforcement a low confidence in the prediction, suggesting that the law-enforcement should conduct a further analysis on the domain name in question.

4.5 Related Works

Although the *law-enforcement* perspective is a comparatively newer dimension in dealing with blacklisted websites, the problem of malicious websites has been extensively investigated (see, e.g., [134, 219]). While the literature studies are loosely related to ours, it's worth discussion by divided them into the following categories.

First, few recent studies mention the notion of *compromised websites* in the same sense as ours (i.e., they are owned by legitimate users) [57, 138, 165]. However, these studies do not consider the

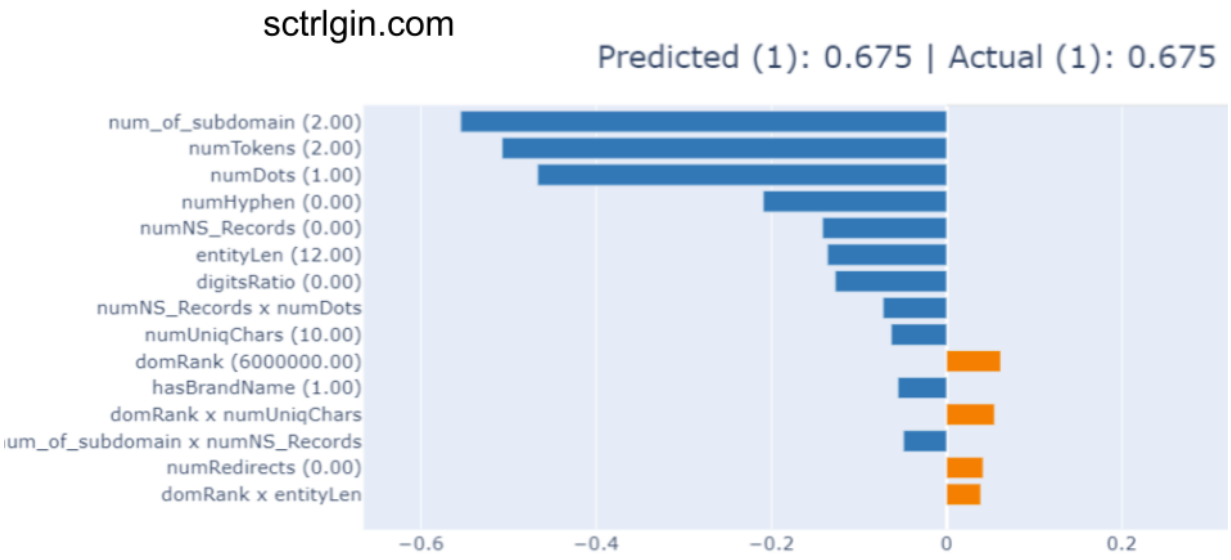

Figure 4.11: Example 4: malicious → further analysis.

notion of *compromised hostname* and *compromised domains* separately.

Second, from an interpretability point of view, most studies deal with the detection of malicious website or phishing URLs via black-box ML models, which often provide highly accurate models but lack interpretability and cannot be used for law-enforcement (e.g., takedown, notify) purposes. Recently, Silva et al. [57] use the LIME explanation method on the random forest model to provide global feature explanations but not individual predictions. In our case study, we use the EBM model for individual prediction interpretations, which the EBM model is also used to detect phishing URLs [84].

Third, there are studies focusing on the detection of phishing webspages or URLs [49, 116].

Fourth, there are studies on leveraging new kinds of information to detect malicious domains. For example, Bilge et al. present EXPOSURE [22], which leverages passive DNS analysis to detect malicious domains; their study inspires us to use the CIR.CL passive DNS dataset and incorporate 'A' records and 'NS' records as features in this study. Another proactive defense tool, PREDATOR [79], aims to detect domain abuses at the time of registration, which is effective against bulk registration events.

Fifth, with the increase in adoption of data-driven methods few recent studies [75, 198] have proposed using web and data analytics to detect malicious fast-flux and cloaking-based web do-

main and campaigns.

Sixth, there are studies on sophisticated attacks and defenses. For example, attackers may re-register expired benign domains to exploit their residual trust and evade reputation based detection (i.e., domain drop-catch).

Seventh, the notion of *adversarial malicious website detection* has been studied in [224]. Moreover, Studies [17, 99, 200] investigate how impersonation and combo-squatting can evade detection by blacklists for a long time, further justifying the incompleteness of blacklists.

Eighth, study [26] presents the importance of deigning cybersecurity for protecting people from malicious actors.

Ninth, study [62] highlights a systematic survey on the usage of data-mining and analytics techniques to aid law-enforcement policing in different aspects.

4.6 Chapter Summary

We presented a framework to help the law-enforcement deal with blacklisted websites, namely taking the appropriate data-driven decisions in terms of (i) whether a domain or host should be treated as the entity for action and (ii) whether the entity should be shut down or notified to the stakeholder. The framework leverages interpretable ML techniques and quantitative classification, which is essential in website maliciousness where datasets are highly imbalanced. Our case study shows that the framework is useful, the results are accurate, but there are rooms for improvement. In the future, the framework can be enhanced to incorporate ways to analyze quality of the interpretations of ML models and to understand why misclassification happens for a complete law-enforcement decision-support system framework to cope with blacklisted websites.

CHAPTER 5: DISCUSSION AND CONCLUSION

In this Chapter we discuss the limitations of the dissertation study, future research directions, and dissertation conclusion.

5.1 Limitations of the Dissertation Study

The dissertation study has a number of limitations, in terms of the characterization of themed threats, the detection of themed malicious websites, and the support to law enforcement in coping with malicious websites. These limitations represent some research directions.

5.1.1 Limitations on Characterizing the Landscape of Themed Threats

The study on characterizing the landscape of themed threats, which was presented in Chapter 2, has the following limitations. First, it only uses the Cyber Kill Chain to map the attack steps for themed attacks. Nevertheless, there are other security frameworks, such as the MITRE ATT&CK. It is interesting to map these frameworks to themed threats and compare which framework may provider deeper insights.

Second, the characterization study focused on the COVID-19 incident as a case study showing how attackers have exploited emerging incidents as themes. However, there are other kinds of incidents that can be exploited to wage themed attacks. These incidents may be different from the COVID-19 one in some sense, and these differences are not systematically studied in the dissertation (e.g., how the differences may lead to different characteristics). Two candidate themes are election and wars, such as the ongoing Ukraine War. While a full-fledged investigation is beyond the scope of the present dissertation, in what follows we sketch what should be considered when investigating these issues.

- In the case of election-themed attacks, the attack vector would be very similar to, if not exactly the same as, the COVID-19 themed attacks, including malicious websites (e.g., themed phishing, malicious payloads in themed websites), malicious emails (e.g., malicious attach-

ments), malicious messages (e.g., persuasive texts), malicious mobile apps (e.g., mimic apps), and misinformation/disinformation through social and online media campaigns. However, there would be differences in terms of the attacker's objectives and the scale of the attack propagation. For example, election-themed attacks typically target one specific country or region. Moreover, these attacks could be waged by nation-state attackers (e.g., Russian interference in the 2016 election through fake news and troll behavior in social media [14, 131]). These attacks often do not just attempt to steal credentials, get ransom, or conduct lateral movements; rather they also attempt to achieve more strategic goals, such as manipulating election results [21], bringing chaos and divisions to societies [13], breaching sensitive data, suppressing the voters' turn-out and making the legitimacy of an election or even the entire democracy system questionable [54].

- In the case of war-themed attacks, the targeted victims also belongs to a specific region (or country). Moreover, websites, emails, and/or messages are themed with attractive war lures (e.g., Russo-Ukraine war updates). These attack vectors are often linked to phishing websites or malicious payloads (e.g., malware, ransomware, adware). The objectives of the attackers include both financial gains and state-sponsored strategic interest [108]. In addition to credential stealing/harvesting (e.g., passwords, username) and economic gains (e.g., fake fundraising scams, fake RedCross donation scams, ransomware attacks), war-themed attacks may focus on cyber espionage (e.g., hacking / attacking against critical infrastructure), Internet disruption (e.g., government websites defacement), and war-related disinformation websites [104]. Moreover, the targeted victims may be primarily geared toward top government officials or high-risk individuals because they might hold sensitive or critical information that may be exploited to change the trajectory of wars. In terms of techniques, these attacks may leverage APTs and malware through themed file attachments linked in emails or messages [108]. To characterize war-themed attacks in the case of the ongoing Russo-Ukraine war, one may group them into two categories: *pro-Russia* and *pro-Ukraine*; this categorization is not relevant to, for example, the COVID-19 themed attacks.

The preceding discussion highlights the sort of adaptations that would need to be made when adapting the characterization study of the COVID-19 themed threats to other kinds of themed threats. Going beyond the two example incidents or themes mentioned above, we observe that attacks leveraging more hyped incidents would reach out to a larger population of victims. Again, this provides a natural prioritization for defenders when defense resources are limited.

Third, the defense and solution spaces are discussed briefly as most of the attack types (e.g., Phishing, Frauds) are well known and some variation of social engineering techniques to victimize user trust by creation a false sense of trust and cash-in the popular interest at the time of the event. It would be exciting research direction to systematically and quantitatively investigate the role of social engineering attacks in these contexts; for this purpose, good first steps have been presented in [128, 147, 148, 183].

5.1.2 Limitations on the Detection of Themed Malicious Websites

The proposed detection methodology of themed malicious websites presented in Chapter 3 has several limitations, ranging from its way of inferring ground-truth labels, potential data imbalance issue, incompleteness of data, ML model interpretability, potential inadequacy in accommodating language differences, and potential incapability in dealing with other kinds of themed threats.

First, it uses a heuristic method to determine the ground truth. This heuristic method can only approximate the ground truth because the data sources (i.e., CheckPhish and DomainTools feeds in this case) may contain some errors, which we could not verify when conducting the study. One research direction to alleviate the problem is to leverage the more advanced methods that have been proposed for a similar purpose [28–30, 61].

Second, it could not avoid the data imbalance problem, meaning that the resulting detectors or classifiers may be slightly biased towards the majority class even after the oversampling of the minority class. However, this resembles the real-world scenarios where malicious websites and benign websites are almost never equal.

Third, it only considers the host-based WHOIS and URL lexical features, but not the website

contents or the network layer features because of the retrospective nature of the study.

Fourth, interpretable ML is not used in the detection case study, which can be leveraged as shown in the corresponding law-enforcement framework (Chapter 4) to explain individual detection and actions against the themed malicious websites.

Fifth, it shows that the python library `wordninja` [12] can make bad splits at times (e.g., when a domain name is seemingly in English characters but actually in another language). This can impact the retrieval of keywords in domain names written in other languages. This means that future research needs to extend the study to incorporate other languages to split domain names into meaningful words in other languages.

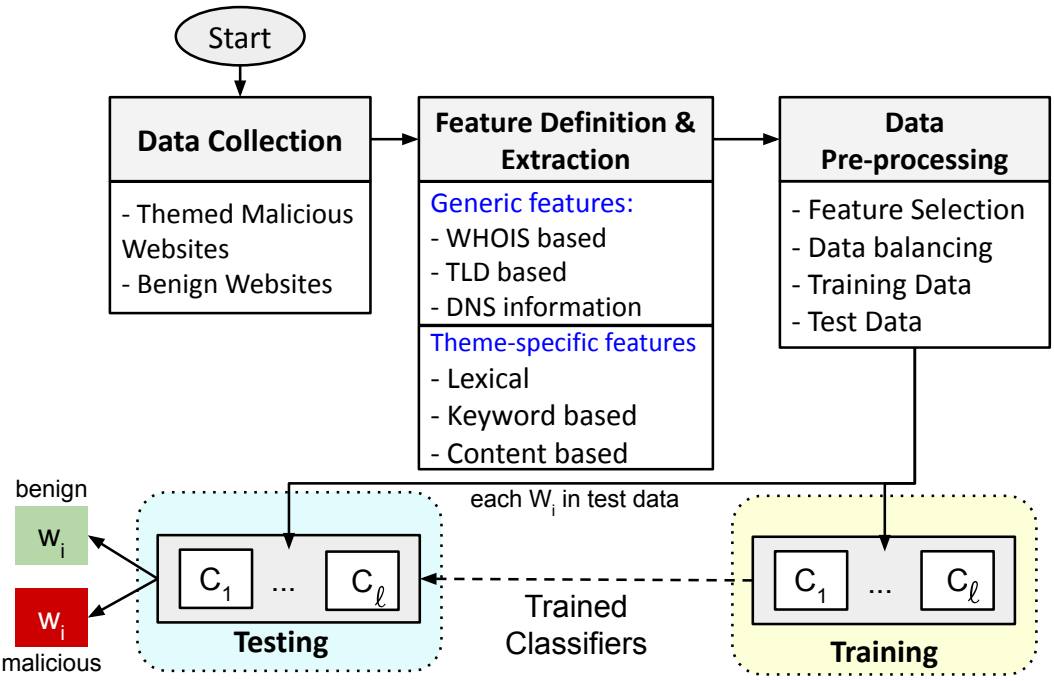

Figure 5.1: Adapted the methodology for detecting COVID-19 themed malicious websites (Figure 3.1) to detect other kinds of themed malicious websites

Sixth, we observe that there is a similar trend of themed domains and websites associated with the ongoing Russia-Ukraine war [51, 174]. Although the dissertation does not investigate specifically how to detect other themed attacks than the COVID-19 one, our detection methodology can be adapted to detect other kinds of themed malicious websites. Figure 5.1 highlights how the methodology (as shown in Figure 3.1) we used to detect COVID-19 themed malicious websites to

detect the other types of themed malicious websites. The key adaptation is in the *Feature Definition & Extraction* component, by distinguishing *generic features* (e.g. WHOIS, TLD, and DNS information) from *theme-specific features* (e.g., domain name lexical information, keyword, and web page content information). For example, theme-specific features would be adapted to reflect the theme, such as pandemic vs. war vs. election. When further adapting the framework to detect themed attacks that exploit emails rather than websites, the generic features would need to be adapted to reflect the email infrastructure rather than the web infrastructure.

5.1.3 Limitations on the Study on Supporting Law-enforcement

The law-enforcement framework presented in Chapter 4 has some limitations, ranging from the ML model interpretability, the potential inadequacy of the method in determining ground-truth labels, the incomplete information that is used to describe websites. Nevertheless, the present study presents, to the best of our knowledge, the first step in *automating* the (hypothetical) process that would be used by the law-enforcement in coping with malicious websites. We are not aware of any documents on how the law-enforcement has been taking further actions against themed malicious websites after identifying some of them. This is may be caused by the lack of laws which authorize the law-enforcement to take further actions. It is worth mentioning that while there have been authorized law-enforcement actions on tearing down botnets [52, 58, 159], it is not clear whether such actions can be automatically extended to accommodating themed malicious websites.

First, the case study is outlined with only the EBM model [164] as a representative of interpretable ML models because we want to provide explanation for individual prediction outcomes. Even though the literature states that EBM is often more intelligible and higher performing than the Random Forest model, which is often successfully used in malicious website detection [125], we have not compared the EBM model results with the other explainable ML methods.

Second, in the current framework we rely on VirusTotal (VT) for labeling websites, but VT is not perfect [204], meaning that the resulting labels may have some ambiguity. We observed

that in some cases where the website is already removed from the blacklists and the corresponding domain is deleted, in such cases the VT considers the website as benign, but in reality their domain name lexical features should indicate malicious behavior. This may be addressed by adapting the methods presented in [28–30, 61].

Third, we do not use any WHOIS features because of potential concerns in relation to the GDPR [203], which hinders our model performance. However, this also gives us the opportunity to observe model performances for classifying truly malicious websites from the compromised ones without using sensitive WHOIS information (i.e., private information).

Fourth, we do not analyze or leverage website contents for this study. It would be interesting to investigate whether accommodating website contents would substantially increase the detection capability.

Fifth, the quality of explanations and further evaluation need to be quantified in future studies. This represents a big challenge which does not appear to have been even recognized by the research community.

5.2 Future Research Directions

In addition to addressing the preceding limitations of the present dissertation study, there are exciting directions for future research.

First, data-driven defenses against malicious websites falls under the umbrella of *cybersecurity data analytics*, or more specifically machine learning based data analytics, which is on par with other kinds of cybersecurity data analytics, such as: algorithmic data analytics [67, 68, 70, 72, 81, 82], statistical data analytics [64,65,69,170,171,173,201,226,227,240,243], or information theory based data analytics [37]. Machine learning, especially deep learning, based cybersecurity data analytics have many applications, such as malware / attack detection (including adversarial malware detection) [66, 94, 95, 110–113, 151–154], software vulnerability detection (including adversarial software vulnerability detection) [118–122, 249, 250], software attribution (including adversarial software attribution) [117], cyber threat hunting [132, 139, 144, 192], anomaly detection from host

/ network logs [15, 176, 177, 193, 216], and time series prediction or forecasting (especially multivariate time series forecasting) [63]. Given all these kinds of cybersecurity data analytics, it would be interesting to unify them into a single theory of cybersecurity data analytics.

Second, cybersecurity data analytics are often geared towards certain cybersecurity metrics, such as attack rates [65, 69, 170, 171, 173, 201, 226, 240, 243] or breach rates [64, 227]. Although the problem of cybersecurity metrics (and quantification) is known to be notoriously difficult, or hard problems [50, 162, 189], significant progresses have been made recently [34–36, 38, 42, 43, 86, 145, 163, 169, 180, 210, 211, 236, 245]. This prompts an exciting future research direction: What kinds of cybersecurity data analytics are suitable for investigating, such as forecasting website maliciousness, and what kinds of cybersecurity metrics? What kind of metrics are suitable to assess the quality of the individual interpretations (i.e., explanations)? Is the explanation stable enough with data perturbations? Can user study of the law-enforcement model reveal the quality of the explanation? Can we provide confidence measures for individual probabilistic prediction outcomes, meaning if a model predict any website as malicious with probability p, can we determine what is the confidence or reliability of such prediction with a score $c \in [0, 1]$?

Third, it is known that cybersecurity data analytics is one pillar, which is on par with the other two pillars in cybersecurity metrics and cybersecurity first-principle modeling, under the umbrella of the broader Cybersecurity Dynamics framework [229–231, 235]. There are several families of first-principle cybersecurity dynamics models, such as reactive and preventive vs. adaptive vs. proactive vs. active, with a rich body of results, especially rigorously characterizing the evolution of the global security state of a network under respective kinds attack-defense-use interactions [36, 53, 77, 78, 114, 124, 130, 225, 228, 232–234, 237, 246, 247]. However, these theoretical studies often make some assumptions. Cybersecurity data analytics or data-driven studies can be applied to validate or invalidate these models and/or the assumptions that are made by them. Therefore, it is an interesting research direction to answer the following question: What kinds of cybersecurity data analytics are suitable for validating or invalidating what kinds of first-principle cybersecurity dynamics models? How should we build first-principle cybersecurity dynamics models to

characterize the evolution of (themed) malicious websites? How should we build first-principle cybersecurity dynamics models to characterize the arms races in the context of (themed) malicious websites? How can we build a more complete and automated trustworthy Decision-Support System (DSS) for cyber defenders to act reliably against (themed) malicious websites?

5.3 Dissertation Conclusion

In summary, the dissertation achieves the goal of characterizing themed malicious websites their attack sophistication, providing methods on how to detect these themed malicious websites, and presenting a framework to support the law-enforcement in dealing with blacklisted websites. It shows how the Cyber Kill Chain model can be leveraged to characterize the threats of themed malicious websites as well as their attack tactics. It introduces a methodology to detect themed malicious websites and domains by using supervised machine learning models (e.g., Random Forest). It investigates how to help the law-enforcement in coping with blacklisted websites, which may contain false-positives in the sense that a blacklisted may be actually benign, or actually owned or operated by a legitimate user but has been compromised and then abused to wage attacks. It shows how interpretable machine learning (e.g., Explainable Boosting Machine) can be leveraged to provide a more trustworthy and transparent law-enforcement decision-making mechanism via quantitative probabilistic class prediction as well as individual prediction outcome explanations (e.g., local explanations based on features).

From a conceptual point of view, the dissertation highlights that trustworthy decision-making by both the law-enforcement and the defenders who use machine learning models to detect or classify malicious websites should have causality and interpretability grounds to make trustworthy decisions.

We hope the present dissertation, especially the future research directions outlined above, will inspire many more studies gearing towards a more robust and trustworthy Decision-Support System (DSS) to take effective measures against malicious websites.

BIBLIOGRAPHY

[1] Netcraft site take down service. https://www.netcraft.com/. 2022. Accessed May 1, 2022.

[2] Farhan Douksieh Abdi and Lian Wenjuan. MALICIOUS URL DETECTION USING CONVOLUTIONAL NEURAL NETWORK, December 2017.

[3] Hossein Abroshan, Jan Devos, Geert Poels, and Eric Laermans. Covid-19 and phishing: Effects of human emotions, behavior, and demographics on the success of phishing attempts during the pandemic. *IEEE Access*, 9:121916–121929, 2021.

[4] ED MURPHY AC, JARED MYERS. Technical analysis: Hackers leveraging covid-19 pandemic to launch phishing attacks, fake apps/maps, trojans, backdoors, cryptominers, botnets ransomware. https://www.carbonblack.com/2020/03/19/technical-analysis-hackers-leveraging-covid-19-pandemic-to-launch-phishing-attacks-trojans-backdoors-cryptominers-botnets-ransomware/, 2020. accessed on 5 June, 2020.

[5] Mir Mehedi Ahsan Pritom, Kristin M. Schweitzer, Raymond M. Bateman, Min Xu, and Shouhuai Xu. Characterizing the landscape of covid-19 themed cyberattacks and defenses. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 1–6, 2020.

[6] Mir Mehedi Ahsan Pritom, Kristin M. Schweitzer, Raymond M. Bateman, Min Xu, and Shouhuai Xu. Data-driven characterization and detection of covid-19 themed malicious websites. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 1–6, 2020.

[7] Ali F Al-Qahtani and Stefano Cresci. The covid-19 scamdemic: A survey of phishing attacks and their countermeasures during covid-19. *IET Information Security*, 2022.

[8] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, and Imtiaz Khan. Phishing attacks: A recent comprehensive study and a new anatomy. *Frontiers in Computer Science*, 3, 2021.

[9] Eihal Alowaisheq, Peng Wang, Sumayah A. Alrwais, Xiaojing Liao, Xiaofeng Wang, Tasneem Alowaisheq, Xianghang Mi, Siyuan Tang, and Baojun Liu. Cracking the wall of confinement: Understanding and analyzing malicious domain take-downs. *Proceedings 2019 Network and Distributed System Security Symposium*, 2019.

[10] Alhanoof Alwaghid. *A Study of Malware Behaviour of Webpages*. PhD thesis, Auckland University of Technology, 2019.

[11] Ahmed Alzahrani. Coronavirus social engineering attacks: Issues and recommendations. *International Journal of Advanced Computer Science and Applications*, 11(5), 2020.

[12] Derek Anderson. wordninja 2.0.0. https://pypi.org/project/wordninja/, 2019. accessed on 12 June, 2020.

[13] Adam Badawy, Aseel Addawood, Kristina Lerman, and Emilio Ferrara. Characterizing the 2016 russian ira influence campaign. *Social Network Analysis and Mining*, 9(1):1–11, 2019.

[14] Adam Badawy, Emilio Ferrara, and Kristina Lerman. Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pages 258–265. IEEE, 2018.

[15] Tim Bai, Haibo Bian, Abbas Abou Daya, Mohammad A Salahuddin, Noura Limam, and Raouf Boutaba. A machine learning approach for rdp-based lateral movement detection. In *2019 IEEE 44th Conference on Local Computer Networks (LCN)*, pages 242–245. IEEE, 2019.

[16] Shahryar Baki, Rakesh Verma, Arjun Mukherjee, and Omprakash Gnawali. Scaling and effectiveness of email masquerade attacks: Exploiting natural language generation. In *Proc. ACM AsiaCCS*, page 469–482, 2017.

[17] Anirban Banerjee, Md Sazzadur Rahman, and Michalis Faloutsos. Sut: Quantifying and mitigating url typosquatting. *Computer Networks*, 55(13):3001–3014, 2011.

[18] Eliza Barclay. Why these scientists still doubt the coronavirus leaked from a chinese lab. https://www.vox.com/2020/4/23/21226484/wuhan-lab-coronavirus-china, 2020. accessed on 5 June, 2020.

[19] Jacob Benesty, Jingdong Chen, and Yiteng Huang. On the importance of the pearson correlation coefficient in noise reduction. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(4):757–765, 2008.

[20] Jenni Bergal. Hospital hackers seize upon coronavirus pandemic. https://www.nextgov.com/cybersecurity/2020/04/hospital-hackers-seize-upon-coronavirus-pandemic/164605/, 2020. accessed on 5 June, 2020.

[21] Hal Berghel. Oh, what a tangled web: Russian hacking, fake news, and the 2016 us presidential election. *Computer*, 50(9):87–91, 2017.

[22] Leyla Bilge, Sevil Sen, Davide Balzarotti, Engin Kirda, and Christopher Kruegel. Exposure: A passive dns analysis service to detect and report malicious domains. *ACM Trans. Inf. Syst. Secur.*, 16(4):14:1–14:28, April 2014.

[23] Marzieh Bitaab, Haehyun Cho, Adam Oest, Penghui Zhang, Zhibo Sun, Rana Pourmohamad, Doowon Kim, Tiffany Bao, Ruoyu Wang, Yan Shoshitaishvili, et al. Scam pandemic: How attackers exploit public fear through phishing. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–10. IEEE, 2020.

[24] FBI Boston. Fbi warns of teleconferencing and online classroom hijacking during covid-19 pandemic. https://www.fbi.gov/contact-us/field-offices/boston/news/press-releases/fbi-warns-of-teleconferencing-and-online-classroom-hijacking-during-covid-19-pandemic, 2020. accessed on 11 June, 2020.

[25] Thomas Brewster. Coronavirus scam alert: Watch out for these risky covid-19 websites and emails. https://www.forbes.com/sites/thomasbrewster/2020/03/12/coronavirus-scam-alert-watch-out-for-these-risky-covid-19-websites-and-emails/, 2020. accessed on 11 June, 2020.

[26] Alex Cadzow. Are we designing cybersecurity to protect people from malicious actors? In Tareq Ahram, Waldemar Karwowski, and Redha Taiar, editors, *Human Systems Engineering and Design*, pages 1038–1043, Cham, 2019. Springer International Publishing.

[27] Alvaro A Cardenas, Pratyusa K Manadhata, and Sreeranga P Rajan. Big data analytics for security. *IEEE Security & Privacy*, 11(6):74–76, 2013.

[28] J. Charlton. *Inferring Malware Detector Metrics in the Absence of Ground-Truth*. PhD thesis, Department of Computer Science, University of Texas at San Antonio, 2021.

[29] J. Charlton, P. Du, J. Cho, and S. Xu. Measuring relative accuracy of malware detectors in the absence of ground truth. In *Proc. IEEE MILCOM*, pages 450–455, 2018.

[30] John Charlton, Pang Du, and Shouhuai Xu. A new method for inferring ground-truth labels and malware detector effectiveness metrics. In Wenlian Lu, Kun Sun, Moti Yung, and Feng Liu, editors, *Science of Cyber Security - Third International Conference, SciSec 2021, Virtual Event, August 13-15, 2021, Revised Selected Papers*, volume 13005 of *Lecture Notes in Computer Science*, pages 77–92. Springer, 2021.

[31] M. Chatterjee and A. Namin. Detecting phishing websites through deep reinforcement learning. In *Proc. IEEE COMPSAC*, pages 227–232, 2019.

[32] Moitrayee Chatterjee and Akbar Siami Namin. Deep reinforcement learning for detecting malicious websites. *arXiv preprint arXiv:1905.09207*, 2019.

[33] CheckPhish. Covid-19 (coronavirus) phishing & scam tracker. https://checkphish.ai/coronavirus-scams-tracker, 2020. accessed on 15 May, 2020.

[34] H. Chen, J. Cho, and S. Xu. Quantifying the security effectiveness of firewalls and dmzs. In *Proc. HoTSoS'2018*, pages 9:1–9:11, 2018.

[35] H. Chen, J. Cho, and S. Xu. Quantifying the security effectiveness of network diversity. In *Proc. HoTSoS'2018*, page 24:1, 2018.

[36] Huashan Chen, Hasan Cam, and Shouhuai Xu. Quantifying cybersecurity effectiveness of dynamic network diversity. *IEEE Transactions on Dependable and Secure Computing*, 2021.

[37] Y. Chen, Z. Huang, S. Xu, and Y. Lai. Spatiotemporal patterns and predictability of cyber-attacks. *PLoS One*, 10(5):e0124472, 05 2015.

[38] Y. Cheng, J. Deng, J. Li, S. DeLoach, A. Singhal, and X. Ou. Metrics of security. In *Cyber Defense and Situational Awareness*, pages 263–295. 2014.

[39] Daiki Chiba, Mitsuaki Akiyama, Takeshi Yagi, Kunio Hato, Tatsuya Mori, and Shigeki Goto. Domainchroma: Building actionable threat intelligence from malicious domain names. *Computers & Security*, 77:138–161, 2018.

[40] Daiki Chiba, Ayako Akiyama Hasegawa, Takashi Koide, Yuta Sawabe, Shigeki Goto, and Mitsuaki Akiyama. Domainscouter: Analyzing the risks of deceptive internationalized domain names. *IEICE TRANSACTIONS on Information and Systems*, 103(7):1493–1511, 2020.

[41] Daiki Chiba, Takeshi Yagi, Mitsuaki Akiyama, Toshiki Shibahara, Tatsuya Mori, and Shigeki Goto. Domainprofiler: toward accurate and early discovery of domain names abused in future. *International Journal of Information Security*, 17(6):661–680, 2018.

[42] J. Cho, P. Hurley, and S. Xu. Metrics and measurement of trustworthy systems. In *Proc. IEEE MILCOM*, 2016.

[43] J. Cho, S. Xu, P. Hurley, M. Mackay, T. Benjamin, and M. Beaumont. Stram: Measuring the trustworthiness of computer-based systems. *ACM Comput. Surv.*, 51(6):128:1–128:47, 2019.

[44] Hyunsang Choi, Bin B Zhu, and Heejo Lee. Detecting malicious web links and identifying their attack types. *WebApps*, 11(11):218, 2011.

[45] Salvador Rodriguez Christina Farr. Facebook, amazon, google and more met with who to figure out how to stop coronavirus misinformation. https://www.cnbc.com/2020/02/14/facebook-google-amazon-met-with-who-to-talk-coronavirus-misinformation.html, 2020. accessed on 4 June, 2020.

[46] Orestis Christou, Nikolaos Pitropakis, Pavlos Papadopoulos, Sean McKeown, and William J. Buchanan. Phishing url detection through top-level domain analysis: A descriptive approach. In *ICISSP*, 2020.

[47] Ben Collier, Shane Horgan, Richard Jones, and Lynsay A Shepherd. The implications of the covid-19 pandemic for cybercrime policing in scotland: a rapid review of the evidence and future considerations. 2020.

[48] Luxembourg Computer Incident Response Center. Passive dns. https://www.circl.lu/services/passive-dns/, 2020. Accessed on 1 August, 2021.

[49] Igino Corona, Battista Biggio, Matteo Contini, Luca Piras, Roberto Corda, Mauro Mereu, Guido Mureddu, Davide Ariu, and Fabio Roli. Deltaphish: Detecting phishing webpages in compromised websites. In *European Symposium on Research in Computer Security*, pages 370–388. Springer, 2017.

[50] INFOSEC Research Council. Hard problem list. `http://www.infosec-research.org/docs\_public/20051130-IRC-HPL-FINAL.pdf`, 2007.

[51] CSC. How to manage the online effects of the ukraine war. https://www.cscdbs.com/blog/how-to-manage-the-online-effects-of-the-ukraine-war/, 2022. accessed on 2 July, 2022.

[52] Christian Czosseck, Gabriel Klein, and Felix Leder. On the arms race around botnets-setting up and taking down botnets. In *2011 3rd International Conference on Cyber Conflict*, pages 1–14. IEEE, 2011.

[53] G. Da, M. Xu, and S. Xu. A new approach to modeling and analyzing security of networked systems. In *Proc. HotSoS'14*, pages 6:1–6:12, 2014.

[54] DANON and THE PROOFPOINT THREAT INSIGHT TEAM. 2020 election threats: An overview of our research. https://www.proofpoint.com/us/blog/threat-insight/2020-election-threats-overview-our-research, 2020. accessed on 1 July, 2022.

[55] Jessica Davis. Hackers, apts exploiting covid-19 with phishing attacks, fraud schemes. https://healthitsecurity.com/news/hackers-apts-exploiting-covid-19-with-phishing-attacks-fraud-schemes, 2020. accessed on 11 June, 2020.

[56] Daniel De Roux, Boris Perez, Andrés Moreno, Maria del Pilar Villamil, and César Figueroa. Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 215–222, 2018.

[57] Ravindu De Silva, Mohamed Nabeel, Charith Elvitigala, Issa Khalil, Ting Yu, and Chamath Keppitiyagama. Compromised or attacker-owned: A large scale classification and study of hosting domains of malicious urls. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*, 2021.

[58] Joseph Demarest. Taking down botnets. https://www.fbi.gov/news/testimony/taking-down-botnets, 2014. accessed on 3 July, 2022.

[59] DNSTwist.it. Dns twist. https://github.com/elceef/dnstwist. Last accessed on 5th October, 2021.

[60] DomainTools. Free covid-19 threat list - domain risk assessments for coronavirus threats. https://www.domaintools.com/resources/blog/free-covid-19-threat-list-domain-risk-assessments-for-coronavirus-threats, 2020. accessed on 14 May, 2020.

[61] P. Du, Z. Sun, H. Chen, J. H. Cho, and S. Xu. Statistical estimation of malware detection metrics in the absence of ground truth. *IEEE T-IFS*, 13(12):2965–2980, 2018.

[62] Matthew Edwards, Awais Rashid, and Paul Rayson. A systematic survey of online data mining technology intended for law enforcement. *ACM Comput. Surv.*, 48(1), sep 2015.

[63] X. Fang, M. Xu, S. Xu, and P. Zhao. A deep learning framework for predicting cyber attacks rates. *EURASIP J. Information Security*, 2019:5, 2019.

[64] Z. Fang, M. Xu, S. Xu, and T. Hu. A framework for predicting data breach risk: Leveraging dependence to cope with sparsity. *IEEE T-IFS*, 16:2186–2201, 2021.

[65] Zijian Fang, Peng Zhao, Maochao Xu, Shouhuai Xu, Taizhong Hu, and Xing Fang. Statistical modeling of computer malware propagation dynamics in cyberspace. *Journal of Applied Statistics*, pages 1–26, 2020.

[66] Gabriel C. Fernandez and Shouhuai Xu. A case study on using deep learning for network intrusion detection. In *2019 IEEE Military Communications Conference (MILCOM'2019)*, pages 1–6, 2018.

[67] E. Ficke, K. Schweitzer, R. Bateman, and S. Xu. Analyzing root causes of intrusion detection false-negatives: Methodology and case study. In *MILCOM*, 2019.

[68] Eric Ficke. Reconstructing alert trees for cyber triage, 2022.

[69] Eric Ficke, Kristin M. Schweitzer, Raymond M. Bateman, and Shouhuai Xu. Characterizing the effectiveness of network-based intrusion detection systems. In *2018 IEEE Military Communications Conference, MILCOM 2018, Los Angeles, CA, USA, October 29-31, 2018*, pages 76–81, 2018.

[70] Eric Ficke and Shouhuai Xu. Apin: Automatic attack path identification in computer networks. In *IEEE ISI'2020*, 2020.

[71] Marites V. Fontanilla. Cybercrime pandemic. *Eubios Journal of Asian and International Bioethics*, 30(4):161–165, 2020.

[72] Richard Garcia-Lebron, Kristin Schweitzer, Raymond Bateman, and Shouhuai Xu. A framework for characterizing the evolution of cyber attacker-victim relation graphs. In *IEEE Milcom'2018*. 2018.

[73] Sergiu Gatlan. Azorult malware infects victims via fake protonvpn installer. https://www.bleepingcomputer.com/news/security/azorult-malware-infects-victims-via-fake-protonvpn-installer/, 2020. accessed on 5 June, 2020.

[74] K. GRADOŃ. Crime in the time of the plague: Fake news pandemic and the challenges to law-enforcement and intelligence community. *Society Register*, 4(2):133–148, 2020.

[75] Ziji Guo and Yong Guan. Active probing-based schemes and data analytics for investigating malicious fast-flux web-cloaking based domains. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9, 2018.

[76] Mee Lan Han, Byung Il Kwak, and Huy Kang Kim. Cbr-based decision support methodology for cybercrime investigation: Focused on the data-driven website defacement analysis. *Security and Communication Networks*, 2019, 2019.

[77] Y. Han, W. Lu, and S. Xu. Characterizing the power of moving target defense via cyber epidemic dynamics. In *HotSoS*, pages 1–12, 2014.

[78] Y. Han, W. Lu, and S. Xu. Preventive and reactive cyber defense dynamics with ergodic time-dependent parameters is globally attractive. *IEEE TNSE*, 8(3):2517–2532, 2021.

[79] Shuang Hao, Alex Kantchelian, Brad Miller, Vern Paxson, and Nick Feamster. Predator: proactive recognition and elimination of domain abuse at time-of-registration. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1568–1579, 2016.

[80] Ren He, Haoyu Wang, Pengcheng Xia, Ling-Ling Wang, Yuanchun Li, Lei Wu, Yajin Zhou, Xiapu Luo, Yao Guo, and Guoai Xu. Beyond the virus: A first look at coronavirus-themed mobile malware. *ArXiv*, abs/2005.14619, 2020.

[81] Songlin He, Eric Ficke, Mir Mehedi Ahsan Pritom, Huashan Chen, Qiang Tang, Qian Chen, Marcus Pendleton, Laurent Njilla, and Shouhuai Xu. Blockchain-based automated and robust cyber security management. *J. Parallel Distributed Comput.*, 163:62–82, 2022.

[82] Songlin He, Eric Ficke, Mir Mehedi Ahsan Pritom, Huashan Chen, Qiang Tang, Qian Chen, Marcus Pendleton, Laurent Njilla, and Shouhuai Xu. Method and system for blockchain-based cyber security management. Patent Application, 2022.

[83] Rebecca Heilweil. Coronavirus scammers are flooding social media with fake cures and tests. https://www.vox.com/recode/2020/4/17/21221692/digital-black-market-covid-19-coronavirus-instagram-twitter-ebay, 2020. accessed on 11 June, 2020.

[84] Paulo R Galego Hernandes, Camila P Floret, Katia F Cardozo De Almeida, Vinícius Camargo Da Silva, João Paulo Papa, and Kelton A Pontara Da Costa. Phishing detection using url-based xai techniques. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 01–06. IEEE, 2021.

[85] Micah Hoffman. Dark web. https://www.sans.org/security-awareness-training/resources/dark-web, 2019. accessed on 10 June, 2020.

[86] J. Homer, S. Zhang, X. Ou, D. Schmidt, Y. Du, S. Rajagopalan, and A. Singhal. Aggregating vulnerability metrics in enterprise networks using attack graphs. *J. Comput. Secur.*, 21(4):561–597, 2013.

[87] Dan Hubbard. Cisco umbrella 1 million. https://umbrella.cisco.com/blog/cisco-umbrella-1-million, 2016. accessed on 13 June, 2020.

[88] Shane Huntley. Findings on covid-19 and online security threats. https://www.blog.google/threat-analysis-group/findings-covid-19-and-online-security-threats/, 2020. accessed on 5 June, 2020.

[89] Alice Hutchings, Richard Clayton, and Ross Anderson. Taking down websites to prevent crime. In *2016 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–10, 2016.

[90] Eric M Hutchins, Michael J Cloppert, and Rohan M Amin. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, 1:80, 2011.

[91] Alexa Internet Inc. Global top sites. https://www.alexa.com/topsites, 2008, Dec.

[92] Storyful Intelligence. Misinformation and disinformation. https://storyful.com/thought-leadership/misinformation-and-disinformation/, 2018. accessed on 5 June, 2020.

[93] H. M. Junaid Khan, Q. Niyaz, V. K. Devabhaktuni, S. Guo, and U. Shaikh. Identifying generic features for malicious url detection system. In *Proc. IEEE UEMCON*, pages 0347–0352, 2019.

[94] E. Kartaltepe, J. Morales, S. Xu, and R. Sandhu. Social network-based botnet command-and-control: Emerging threats and countermeasures, 2010 (manuscript in submission).

[95] Erhan J. Kartaltepe, Jose Andre Morales, Shouhuai Xu, and Ravi S. Sandhu. Social network-based botnet command-and-control: Emerging threats and countermeasures. In *ACNS*, pages 511–528, 2010.

[96] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[97] Shubhdeep Kaur and Sukhchandan Randhawa. Dark web: A web of crimes. *Wireless Personal Communications*, 112(4):2131–2158, 2020.

[98] Navid Ali Khan, Sarfraz Nawaz Brohi, and Noor Zaman. Ten deadly cyber security threats amid covid-19 pandemic, May 2020.

[99] Panagiotis Kintis, Najmeh Miramirkhani, Charles Lever, Yizheng Chen, Rosa Romero-Gómez, Nikolaos Pitropakis, Nick Nikiforakis, and Manos Antonakakis. Hiding in plain sight: A longitudinal study of combosquatting abuse. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 569–586, 2017.

[100] KRISTIN M. KRAEMER. Social engineering scam hits washington county government. https://www.govtech.com/security/Social-Engineering-Scam-Hits-Washington-County-Government.html, 2020. accessed on 5 June, 2020.

[101] Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*, volume 26. Springer, 2013.

[102] Marc Kührer, Christian Rossow, and Thorsten Holz. Paint it black: Evaluating the effectiveness of malware blacklists. In *International Workshop on Recent Advances in Intrusion Detection*, pages 1–21. Springer, 2014.

[103] Himabindu Lakkaraju and Osbert Bastani. " how do i fool you?" manipulating user trust via misleading black box explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 79–85, 2020.

[104] Ravie Lakshmanan. Ukraine war themed files become the lure of choice for a wide range of hackers. https://thehackernews.com/2022/05/ukraine-war-themed-files-become-lure-of.html, 2022. accessed on 1 July, 2022.

[105] Harjinder Singh Lallie, Lynsay A. Shepherd, Jason R.C. Nurse, Arnau Erola, Gregory Epiphaniou, Carsten Maple, and Xavier Bellekens. Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers Security*, 105:102248, 2021.

[106] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, NDSS 2019, February 2019.

[107] Robert Lemons. Lessons from the war on malicious mobile apps. https://www.darkreading.com/mobile/lessons-from-the-war-on-malicious-mobile-apps/d/d-id/1333946, 2019. accessed on 3 June, 2020.

[108] Billy Leonard. Update on cyber activity in eastern europe. https://blog.google/threat-analysis-group/update-on-cyber-activity-in-eastern-europe/, 2022. accessed on 1 July, 2022.

[109] Michael Levi. Trends and costs of fraud. In *Fraud*, pages 37–48. Routledge, 2016.

[110] D. Li, Q. Li, Y. Ye, and S. Xu. Enhancing robustness of deep neural networks against adversarial malware samples: Principles, framework, and aics'2019 challenge. In *AAAI-2019 Workshop on Artificial Intelligence for Cyber Security (AICS'2019)*.

[111] D. Li, T. Qiu, S. Chen, Q. Li, and S. Xu. Can we leverage predictive uncertainty to detect dataset shift and adversarial examples in android malware detection? In *The 2021 Annual Computer Security Application Conference (ACSAC)*, 2021.

[112] Deqiang Li, Qianmu Li, Yanfang Ye, and Shouhuai Xu. A framework for enhancing deep neural networks against adversarial malware. *IEEE Trans. Netw. Sci. Eng.*, 8(1):736–750, 2021.

[113] Deqiang Li, Qianmu Li, Yanfang (Fanny) Ye, and Shouhuai Xu. Arms race in adversarial malware detection: A survey. *ACM Comput. Surv.*, 55(1), nov 2021.

[114] X. Li, P. Parker, and S. Xu. A stochastic model for quantitative security analyses of networked systems. *IEEE TDSC*, 8(1):28–43, 2011.

[115] Yukun Li, Zhenguo Yang, Xu Chen, Huaping Yuan, and Wenyin Liu. A stacking model using url and html features for phishing webpage detection. *Future Gener. Comput. Syst.*, 94:27–39, 2019.

[116] Yukun Li, Zhenguo Yang, Xu Chen, Huaping Yuan, and Wenyin Liu. A stacking model using url and html features for phishing webpage detection. *Future Generation Computer Systems*, 94:27–39, 2019.

[117] Z. Li, Q. Chen, C. Chen, Y. Zou, and S. Xu. Ropgen: Towards robust code authorship attribution via automatic coding style transformation. In *Accepted to International Conference on Software Engineering (ICSE'2022)*.

[118] Z. Li, D. Zou, S. Xu, Z. Chen, Y. Zhu, and H. Jin. Vuldeelocator: A deep learning-based fine-grained vulnerability detector. *IEEE TDSC*, 2021.

[119] Z. Li, D. Zou, S. Xu, H. Jin, H. Qi, and J. Hu. Vulpecker: an automated vulnerability detection system based on code similarity analysis. In *Pro. ACSAC'16*, pages 201–213, 2016.

[120] Z. Li, D. Zou, S. Xu, H. Jin, Y. Zhu, and Z. Chen. Sysevr: A framework for using deep learning to detect software vulnerabilities. *IEEE Transactions on Dependable and Secure Computing*, 2021.

[121] Z. Li, D. Zou, S. Xu, X. Ou, H. Jin, S. Wang, Z. Deng, and Y. Zhong. Vuldeepecker: A deep learning-based system for vulnerability detection. In *Proc. NDSS'18*, 2018.

[122] Zhen Li, Jing Tang, Deqing Zou, Qian Chen, Shouhuai Xu, Chao Zhang, Yichen Li, and Hai Jin. Robustness of deep learning-based vulnerability detectors: Attack anddefense. under review, 2021.

[123] Y. Liang and X. Yan. Using deep learning to detect malicious urls. In *Proc. IEEE ICEI*, pages 487–492, 2019.

[124] Z. Lin, W. Lu, and S. Xu. Unified preventive and reactive cyber defense dynamics is still globally convergent. *IEEE/ACM ToN*, 27(3):1098–1111, 2019.

[125] Chunlin Liu, Lidong Wang, Bo Lang, and Yuan Zhou. Finding effective classifier for malicious url detection. In *Proceedings of the 2018 2nd International Conference on Management Engineering, Software Engineering and Service Sciences*, pages 240–244, 2018.

[126] D. Liu, J. Lee, W. Wang, and Y. Wang. Malicious websites detection via cnn based screenshot recognition*. In *International Conf. on Intelligent Computing and its Emerging Applications*, pages 115–119, 2019.

[127] Lockheed Martin. Cyber kill chain. `https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html`, (Accessed June 25, 2022).

[128] Theodore Longtchi, Rosana Montañez Rodriguez, Laith Al-Shawaf, Adham Atyabi, and Shouhuai Xu. Sok: Why have defenses against social engineering attacks achieved limited success? *CoRR*, abs/2203.08302, 2022.

[129] Yin Lou, Rich Caruana, and Johannes Gehrke. Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–158. ACM, 2012.

[130] W. Lu, S. Xu, and X. Yi. Optimizing active cyber defense dynamics. In *Proc. GameSec'13*, pages 206–225, 2013.

[131] Luca Luceri, Silvia Giordano, and Emilio Ferrara. Detecting troll behavior via inverse reinforcement learning: A case study of russian trolls in the 2016 us election. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 417–427, 2020.

[132] Nataliia Lukova-Chuiko, Andriy Fesenko, Hanna Papirna, and Sergiy Gnatyuk. Threat hunting as a method of protection against cyber threats. In *IT&I*, pages 103–113, 2020.

[133] Justin Ma, Lawrence K Saul, Stefan Savage, and Geoffrey M Voelker. Beyond blacklists: learning to detect malicious web sites from suspicious urls. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1245–1254, 2009.

[134] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Beyond blacklists: Learning to detect malicious web sites from suspicious urls. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 1245–1254, New York, NY, USA, 2009. Association for Computing Machinery.

[135] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Learning to detect malicious urls. *ACM TIST*, 2(3):30:1–30:24, 2011.

[136] Katelyn Wan Fei Ma and Tammy McKinnon. Covid-19 and cyber fraud: Emerging threats during the pandemic. *Journal of Financial Crime*, 2021.

[137] Yury Magali. 10 security best practices for mobile device owners. https://www.cdillc.com/10-security-best-practices-mobile-device-owners/. accessed on 8 August, 2020.

[138] S. Maroofi, M. Korczyński, C. Hesselman, B. Ampeau, and A. Duda. Comar: Classification of compromised versus maliciously registered domains. In *2020 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 607–623, 2020.

[139] Kassidy Marsh and Samira Eisaloo Gharghasheh. Fuzzy bayesian learning for cyber threat hunting in industrial control systems. In *Handbook of Big Data Analytics and Forensics*, pages 117–130. Springer, 2022.

[140] Muhammad Ahmed Masud. An open source intelligence (osint) framework for online investigations, June 2019.

[141] Rick McElroy. What is the cyber security equivalent of washing your hands for 20 seconds? https://www.enterprisetimes.co.uk/2020/04/15/what-is-the-cyber-security-equivalent-of-washing-your-hands-for-20-seconds/, 2020. accessed on 31 May, 2020.

[142] Amy McGovern, Ryan Lagerquist, David John Gagne, G Eli Jergensen, Kimberly L Elmore, Cameron R Homeyer, and Travis Smith. Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, 100(11):2175–2199, 2019.

[143] Albert Meijer and Martijn Wessels. Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, 42(12):1031–1039, 2019.

[144] Md Nazmus Sakib Miazi, Mir Mehedi A. Pritom, Mohamed Shehab, Bill Chu, and Jinpeng Wei. The design of cyber threat hunting games: A case study. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6, 2017.

[145] J. Mireles, E. Ficke, J. Cho, P. Hurley, and S. Xu. Metrics towards measuring cyber agility. *IEEE T-IFS*, 14(12):3217–3232, 2019.

[146] P. V. Mockapetris. Rfc1034: Domain names - concepts and facilities, 1987.

[147] Rosana Montanez, Edward Golob, and Shouhuai Xu. Human cognition through the lens of social engineering cyberattacks. *Frontiers in Psychology*, 11:1755, 2020.

[148] Rosana Montañez, Adham Atyabi, and Shouhuai Xu. *Cybersecurity and Cognitive Science*, chapter Social Engineering Attacks and Defenses in the Physical World vs. Cyberspace: A Contrast Study. Elsevier, 2022.

[149] Rosana Montañez, Edward Golob, and Shouhuai Xu. Human cognition through the lens of social engineering cyberattacks. *Frontiers in Psychology*, 11:1755, 2020.

[150] Tyler Moore and Richard Clayton. Evil searching: Compromise and recompromise of internet hosts for phishing. In *International Conference on Financial Cryptography and Data Security*, pages 256–272. Springer, 2009.

[151] J. Morales, M. Main, W. Luo, S. Xu, and R. Sandhu. Building malware infection trees. In *Proceedings of the 2011 6th International Conference on Malicious and Unwanted Software (MALWARE'11)*, pages 50–57, 2011.

[152] Jose Andre Morales, Areej Al-Bataineh, Shouhuai Xu, and Ravi S. Sandhu. Analyzing DNS activities of bot processes. In *4th International Conference on Malicious and Unwanted Software, MALWARE 2009, Montréal, Quebec, Canada, October 13-14, 2009*, pages 98–103, 2009.

[153] Jose Andre Morales, Areej Al-Bataineh, Shouhuai Xu, and Ravi S. Sandhu. Analyzing and exploiting network behaviors of malware. In *SecureComm*, pages 20–34, 2010.

[154] Jose Andre Morales, Erhan J. Kartaltepe, Shouhuai Xu, and Ravi S. Sandhu. Symptoms-based detection of bot processes. In *MMM-ACNS*, pages 229–241, 2010.

[155] Giovane CM Moura, Ramin Sadre, and Aiko Pras. Bad neighborhoods on the internet. *IEEE communications magazine*, 52(7):132–139, 2014.

[156] Ishmael Mugari and Emeka E. Obioha. Predictive policing and crime control in the united states of america and europe: Trends in a decade of research and the future of predictive policing. *Social Sciences*, 10(6), 2021.

[157] Phil Muncaster. Hackers target netflix and disney+ with covid19 phishing. https://www.infosecurity-magazine.com/news/hackers-target-netflix-disney/, 2020. accessed on 25 June, 2020.

[158] Sunil Kumar Muttoo and Shikha Badhani. An analysis of malware detection and control through covid-19 pandemic. In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 637–641, 2021.

[159] Yacin Nadji, Manos Antonakakis, Roberto Perdisci, David Dagon, and Wenke Lee. Beheading hydras: performing effective botnet takedowns. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 121–132, 2013.

[160] Rennie Naidoo. A multi-level influence model of covid-19 themed cybercrime. *European Journal of Information Systems*, 0(0):1–16, 2020.

[161] Amirreza Niakanlahiji, Mir Mehedi Pritom, Bei-Tseng Chu, and Ehab Al-Shaer. Predicting zero-day malicious ip addresses. In *Proceedings of the 2017 Workshop on Automated Decision Making for Active Cyber Defense*, SafeConfig '17, pages 1–6, New York, NY, USA, 2017. ACM.

[162] David Nicol, Bill Sanders, Jonathan Katz, Bill Scherlis, Tudor Dumitra, Laurie Williams, and Munindar P. Singh. The science of security 5 hard problems (august 2015). `http://cps-vo.org/node/21590`.

[163] S. Noel, , and S. Jajodia. *A Suite of Metrics for Network Attack Graph Analytics*, pages 141–176. Springer International Publishing, 2017.

[164] Harsha Nori, Samuel Jenkins, Paul Koch, and Rich Caruana. Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223*, 2019.

[165] Sophie Le Page, Guy-Vincent Jourdan, Gregor von Bochmann, Iosif-Viorel Onut, and Jason Flood. Domain classifier: Compromised machines versus malicious registrations. In *ICWE*, 2019.

[166] Sharbani Pandit, Roberto Perdisci, Mustaque Ahamad, and Payas Gupta. Towards measuring the effectiveness of telephony blacklists. In *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*. The Internet Society, 2018.

[167] Mandeep Pannu, Iain Kay, and Daniel Harris. Using dark web crawler to uncover suspicious and malicious websites. In Tareq Z. Ahram and Denise Nicholson, editors, *Advances*

in Human Factors in Cybersecurity, pages 108–115, Cham, 2019. Springer International Publishing.

[168] JIM TREINEN PATRICK UPATHAM. Amid covid-19, global orgs see a 148% spike in ransomware attacks; finance industry heavily targeted. https://www.carbonblack.com/2020/04/15/amid-covid-19-global-orgs-see-a-148-spike-in-ransomware-attacks-finance-industry-heavily-targeted/, 2020. accessed on 10 June, 2020.

[169] M. Pendleton, R. Garcia-Lebron, J. Cho, and S. Xu. A survey on systems security metrics. *ACM Comput. Surv.*, 49(4):62:1–62:35, 2016.

[170] Chen Peng, Maochao Xu, Shouhuai Xu, and Taizhong Hu. Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics*, 44(14):2534–2563, 2017.

[171] Chen Peng, Maochao Xu, Shouhuai Xu, and Taizhong Hu. Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics*, 44(14):2534–2563, 2017.

[172] Chen Peng, Maochao Xu, Shouhuai Xu, and Taizhong Hu. Modeling multivariate cybersecurity risks. *Journal of Applied Statistics*, 45(15):2718–2740, 2018.

[173] Chen Peng, Maochao Xu, Shouhuai Xu, and Taizhong Hu. Modeling multivariate cybersecurity risks. *Journal of Applied Statistics*, 0(0):1–23, 2018.

[174] Kathleen Persighetti, Laura Teixeira, da Rocha, and Avinash Shende. Newly observed domains and the ukraine war. https://blogs.infoblox.com/cyber-threat-intelligence/cyber-threat-advisory/newly-observed-domains-and-the-ukraine-war/, 2022. accessed on 2 July, 2022.

[175] Check Point. A perfect storm: the security challenges of coronavirus threats and mass remote working. https://www.blog.google/threat-analysis-group/findings-covid-19-and-online-security-threats/, 2020. accessed on 5 June, 2020.

[176] Brian A Powell. Detecting malicious logins as graph anomalies. *Journal of Information Security and Applications*, 54:102557, 2020.

[177] Mir Mehedi A. Pritom, Chuqin Li, Bill Chu, and Xi Niu. A study on log analysis approaches using sandia dataset. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6, 2017.

[178] Mir Mehedi A. Pritom, Rosana Montanez Rodriguez, Asad Ali Khan, Sebastian A. Nugroho, Esra'a Alrashydah, Beatrice N. Ruiz, and Anthony Rios. Case study on detecting COVID-19 health-related misinformation in social media. *CoRR*, abs/2106.06811, 2021.

[179] Mir Mehedi Ahsan Pritom and Shouhuai Xu. Supporting law-enforcement to cope with blacklisted websites: Framework and case study. In *2022 IEEE Conference on Communications and Network Security (CNS)*, pages 1–10, 2022.

[180] A. Ramos, M. Lazar, R. H. Filho, and J. J. P. C. Rodrigues. Model-based quantitative network security metrics: A survey. *IEEE Communications Surveys Tutorials*, 19(4):2704–2734, 2017.

[181] PrivSec Report. Typosquatting & duplication of pharmaceutical domain – possibly used for phishing activity. https://gdpr.report/news/2020/05/06/typosquatting-duplication-of-pharmaceutical-domain-possibly-used-for-phishing-activity/, 2020. accessed on 9 June, 2020.

[182] RiskIQ. Covid-19 cybercrime update. https://www.riskiq.com/blog/analyst/covid19-cybercrime-update/, 2020. accessed on 3 June, 2020.

[183] Rosana Montanez Rodriguez and Shouhuai Xu. Cyber social engineering kill chain. In *manuscript under review*, 2022.

[184] Janell Ross. Coronavirus outbreak revives dangerous race myths and pseudoscience. https://www.nbcnews.com/news/nbcblk/coronavirus-outbreak-revives-dangerous-race-myths-pseudoscience-n1162326, 2020. accessed on 12 June, 2020.

[185] Ethan M. Rudd and Ahmed Abdallah. Training transformers for information security tasks: A case study on malicious url prediction, 2020.

[186] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[187] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, and Banu Diri. Machine learning based phishing detection from urls. *Expert Systems with Applications*, 117:345 – 357, 2019.

[188] Lisa Weintraub Schifferle. Looking for work after coronavirus layoffs? https://www.consumer.ftc.gov/blog/2020/04/looking-work-after-coronavirus-layoffs, 2020. accessed on 11 June, 2020.

[189] National Science and Technology Council. Trustworthy cyberspace: Strategic plan for the federal cybersecurity research and development program. `https://www.nitrd.gov/SUBCOMMITTEE/csia/Fed\_Cybersecurity\_RD\_Strategic\_Plan\_2011.pdf`, 2011.

[190] Menlo Security. Sophisticated covid-19–based phishing attacks leverage pdf attachments and saas to bypass defenses. https://www.menlosecurity.com/blog/sophisticated-covid-19-based-phishing-attacks-leverage-pdf-attachments-and-saas-to-bypass-defenses, 2020. accessed on 5 June, 2020.

[191] Ian Sherr. Apple, google, amazon block nonofficial coronavirus apps from app stores. https://www.cnet.com/news/apple-google-amazon-block-nonofficial-coronavirus-apps-from-app-stores/, 2020. Accessed on 2 June, 2020.

[192] Yash Shukla. *Threat Hunting Using a Machine Learning Approach*. PhD thesis, Dublin, National College of Ireland, 2020.

[193] Hossein Siadati and Nasir Memon. Detecting structurally anomalous logins within enterprise networks. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1273–1284, 2017.

[194] Sudeep Singh. Coronavirus-themed document targets brazilian users. https://www.zscaler.com/blogs/research/coronavirus-themed-document-targets-brazilian-users, 2020. Accessed on 10 June, 2020.

[195] Bridget Small. Scam emails demand bitcoin, threaten blackmail. https://www.consumer.ftc.gov/blog/2020/04/scam-emails-demand-bitcoin-threaten-blackmail, 2020. accessed on 11 June, 2020.

[196] Kyle Soska and Nicolas Christin. Automatically detecting vulnerable websites before they turn malicious. In *USENIX Security Symposium*, 2014.

[197] Jan Spooren, Davy Preuveneers, Lieven Desmet, Peter Janssen, and Wouter Joosen. On the use of dgas in malware: An everlasting competition of detection and evasion. *SIGAPP Appl. Comput. Rev.*, 19(2):31–43, aug 2019.

[198] Oleksii Starov, Yuchen Zhou, Xiao Zhang, Najmeh Miramirkhani, and Nick Nikiforakis. Betrayed by your dashboard: Discovering malicious campaigns via web analytics. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 227–236, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[199] Guillermo Suarez-Tangil, Matthew Edwards, Claudia Peersman, Gianluca Stringhini, Awais Rashid, and Monica Whitty. Automatically dismantling online dating fraud. *IEEE Transactions on Information Forensics and Security*, 15:1128–1137, 2019.

[200] Ke Tian, Steve T. K. Jan, Hang Hu, Danfeng Yao, and Gang Wang. Needle in a haystack: Tracking down elite phishing domains in the wild. In *Proceedings of the Internet Measurement Conference 2018*, IMC '18, page 429–442, New York, NY, USA, 2018. Association for Computing Machinery.

[201] Van Trieu-Do, Richard Garcia-Lebron, Maochao Xu, Shouhuai Xu, and Yusheng Feng. Characterizing and leveraging granger causality in cybersecurity: Framework and case study. *EAI Endorsed Trans. Security Safety*, 7(25):e4, 2020.

[202] Filip Truta. Hackers actively exploiting enterprise vpn bugs amid covid-19 telework trend, says dhs. https://securityboulevard.com/2020/03/hackers-actively-exploiting-enterprise-vpn-bugs-amid-covid-19-telework-trend-says-dhs/, 2020. accessed on 11 June, 2020.

[203] European Union. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj, 2016, April.

[204] Pelayo Vallina, Victor Le Pochat, Álvaro Feal, Marius Paraschiv, Julien Gamba, Tim Burke, Oliver Hohlfeld, Juan Tapiador, and Narseo Vallina-Rodriguez. Mis-shapes, mistakes, misfits: An analysis of domain classification services. In *Proceedings of the ACM Internet Measurement Conference*, pages 598–618, 2020.

[205] B. R. Chandavarkar Venkatesha Sushruth, K. Rahul Reddy. Social engineering attacks during the covid-19. *SN Computer Science*, 2(78), 2021.

[206] Rakesh Verma and Avisha Das. What's in a url: Fast feature extraction and malicious url detection. In *Proc. ACM IWSPA'17*, page 55–63, 2017.

[207] James Vincent. Conspiracy theorists say 5g causes novel coronavirus, so now they're harassing and attacking uk telecoms engineers. https://www.theverge.com/2020/6/3/21276912/5g-

conspiracy-theories-coronavirus-uk-telecoms-engineers-attacks-abuse, 2020. accessed on 5 June, 2020.

[208] VirusTotal. Virustotal online service. https://www.virustotal.com, 2020.

[209] David Wall. *Cybercrime: The transformation of crime in the information age*, volume 4. Polity, 2007.

[210] L. Wang, S. Jajodia, and A. Singhal. *Network Security Metrics*. Springer, 2017.

[211] L. Wang, S. Jajodia, A. Singhal, P. Cheng, and S. Noel. k-zero day safety: A network security metric for measuring the risk of unknown vulnerabilities. *IEEE TDSC*, 11(1):30–44, 2014.

[212] World Health Organization (WHO). Coronavirus disease (covid-19) advice for the public: Myth busters. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters, 2020. accessed on 12 June, 2020.

[213] Nimesha Wickramasinghe, Mohamed Nabeel, Kenneth Thilakaratne, Chamath Keppitiyagama, and Kasun De Zoysa. Uncovering IP address hosting types behind malicious websites. *CoRR*, abs/2111.00142, 2021.

[214] Wikipedia. Shannon entropy. https://en.wiktionary.org/wiki/Shannon-entropy, 2020. accessed on 3 June, 2020.

[215] Davey Winder. Fbi says foreign states hacked into u.s. covid-19 research centers: Report. https://www.forbes.com/sites/daveywinder/2020/04/17/fbi-says-foreign-states-hacked-into-us-covid-19-research-centers-report/147cd0e73c29, 2020. accessed on 5 June, 2020.

[216] Yanna Wu, Fucheng Liu, and Yu Wen. Malicious login detection using long short-term memory with an attention mechanism. In *IFIP International Conference on Digital Forensics*, pages 157–173. Springer, 2021.

[217] Pengcheng Xia, Mohamed Nabeel, Issa Khalil, Haoyu Wang, and Ting Yu. Identifying and characterizing covid-19 themed malicious domain campaigns. In *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, CODASPY '21, page 209–220, New York, NY, USA, 2021. Association for Computing Machinery.

[218] Pengcheng Xia, Haoyu Wang, Xiapu Luo, Lei Wu, Yajin Zhou, Guangdong Bai, Guoai Xu, Gang Huang, and Xuanzhe Liu. Don't fish in troubled waters! characterizing coronavirus-themed cryptocurrency scams, 2020.

[219] L. Xu, Z. Zhan, S. Xu, and K. Ye. Cross-layer detection of malicious websites. In *ACM CODASPY*, pages 141–152, 2013.

[220] L. Xu, Z. Zhan, S. Xu, and K. Ye. An evasion and counter-evasion study in malicious websites detection. In *IEEE CNS*, pages 265–273, 2014.

[221] Li Xu. Detecting and characterizing malicious websites, 2014.

[222] Li Xu, Zhenxin Zhan, Shouhuai Xu, and Keying Ye. Cross-layer detection of malicious websites. In *Third ACM Conference on Data and Application Security and Privacy (CODASPY'13)*, pages 141–152, 2013.

[223] Li Xu, Zhenxin Zhan, Shouhuai Xu, and Keying Ye. An evasion and counter-evasion study in malicious websites detection. In *Proc. IEEE CNS*, pages 265–273, 2014.

[224] Li Xu, Zhenxin Zhan, Shouhuai Xu, and Keying Ye. An evasion and counter-evasion study in malicious websites detection. In *IEEE Conference on Communications and Network Security (CNS'14)*, pages 265–273, 2014.

[225] M. Xu, G. Da, and S. Xu. Cyber epidemic models with dependences. *Internet Mathematics*, 11(1):62–92, 2015.

[226] M. Xu, L. Hua, and S. Xu. A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics*, 59(4):508–520, 2017.

[227] M. Xu, K. M. Schweitzer, R. M. Bateman, and S. Xu. Modeling and predicting cyber hacking breaches. *IEEE T-IFS*, 13(11):2856–2871, 2018.

[228] M. Xu and S. Xu. An extended stochastic model for quantitative security analysis of networked systems. *Internet Mathematics*, 8(3):288–320, 2012.

[229] S. Xu. Emergent behavior in cybersecurity. In *Proc. HotSoS*, pages 13:1–13:2, 2014.

[230] S. Xu. Cybersecurity dynamics: A foundation for the science of cybersecurity. In *Proactive and Dynamic Network Defense*, pages 1–31. 2019.

[231] S. Xu. The cybersecurity dynamics way of thinking and landscape (invited paper). In *ACM Workshop on Moving Target Defense*, 2020.

[232] S. Xu, W. Lu, and L. Xu. Push- and pull-based epidemic spreading in networks: Thresholds and deeper insights. *ACM TAAS*, 7(3), 2012.

[233] S. Xu, W. Lu, L. Xu, and Z. Zhan. Adaptive epidemic dynamics in networks: Thresholds and control. *ACM TAAS*, 8(4), 2014.

[234] S. Xu, W. Lu, and Z. Zhan. A stochastic model of multivirus dynamics. *IEEE Transactions on Dependable and Secure Computing*, 9(1):30–45, 2012.

[235] Shouhuai Xu. Cybersecurity dynamics. In *Proc. HotSoS'14*, pages 14:1–14:2, 2014.

[236] Shouhuai Xu. Sarr: A cybersecurity metrics and quantification framework (keynote). In *Science of Cyber Security - Third International Conference (SciSec'2021)*, volume 13005 of *Lecture Notes in Computer Science*, pages 3–17. Springer, 2021.

[237] Shouhuai Xu, Wenlian Lu, and Hualun Li. A stochastic model of active cyber defense dynamics. *Internet Mathematics*, 11(1):23–61, 2015.

[238] W. Yang, W. Zuo, and B. Cui. Detecting malicious urls via a keyword-based convolutional gated-recurrent-unit neural network. *IEEE Access*, 7:29891–29900, 2019.

[239] Mingsheng Ying. Additive models of probabilistic processes. *Theoretical Computer Science*, 275(1-2):481–519, 2002.

[240] Z. Zhan, M. Xu, and S. Xu. Characterizing honeypot-captured cyber attacks: Statistical framework and case study. *IEEE T-IFS*, 8(11), 2013.

[241] Zhenxin Zhan, Maochao Xu, and Shouhuai Xu. Characterizing honeypot-captured cyber attacks: Statistical framework and case study. *IEEE Transactions on Information Forensics and Security*, 8(11):1775–1789, 2013.

[242] Zhenxin Zhan, Maochao Xu, and Shouhuai Xu. Predicting cyber attack rates with extreme values. *IEEE Transactions on Information Forensics and Security*, 10(8):1666–1677, 2015.

[243] Zhenxin Zhan, Maochao Xu, and Shouhuai Xu. Predicting cyber attack rates with extreme values. *IEEE T-IFS*, 10(8):1666–1677, 2015.

[244] Kunsan Zhang, Wen Ji, Nan Li, Yiting Wang, and Shengyang Liao. Detection of malicious domain name based on dns data analysis. In *Journal of Physics: Conference Series*, volume 1544, page 012169. IOP Publishing, 2020.

[245] M. Zhang, L. Wang, S. Jajodia, A. Singhal, and M. Albanese. Network diversity: A security metric for evaluating the resilience of networks against zero-day attacks. *IEEE Trans. Inf. Forensics Secur.*, 11(5):1071–1086, 2016.

[246] R. Zheng, W. Lu, and S. Xu. Active cyber defense dynamics exhibiting rich phenomena. In *Proc. HotSoS*, 2015.

[247] R. Zheng, W. Lu, and S. Xu. Preventive and reactive cyber defense dynamics is globally stable. *IEEE TNSE*, 5(2):156–170, 2018.

[248] Zeljka Zorz. Spotting and blacklisting malicious covid-19-themed sites. https://www.helpnetsecurity.com/2020/04/07/covid-19-malicious-sites/, 2020. accessed on 12 August, 2020.

[249] D. Zou, S. Wang, S. Xu, Z. Li, and H. Jin. μvuldeepecker: A deep learning-based system for multiclass vulnerability detection. *IEEE TDSC*, 2020.

[250] Deqing Zou, Yawei Zhu, Shouhuai Xu, Zhen Li, Hai Jin, and Hengkai Ye. Interpreting deep learning-based vulnerability detector predictions based on heuristic searching. *ACM Trans. Softw. Eng. Methodol.*, 30(2), March 2021.

VITA

Mir Mehedi Ahsan Pritom is a Ph.D. candidate in the Computer Science Doctoral Program at UT San Antonio (UTSA). He joined the program in January 2019 (Spring term) as an international student from Bangladesh after completing his Master of Science degree in Information Technology with security and privacy concentration from the University of North Carolina at Charlotte (UNCC). Prior to that, Mir completed his Bachelor of Science in Computer Science and Engineering from the University of Dhaka, Bangladesh back in 2014. During his Ph.D. journey, Mir has published multiple peer-reviewed articles in top journals and conferences, which are related to this present dissertation. His latest conference paper, titled "Supporting Law-enforcement in Coping with Blacklisted Websites: Framework and Case Study," is submitted to the IEEE conference on Communication and Network Security (CNS) 2022 (under review). Mir also co-authored a journal paper titled "Blockchain-Based Automated and Robust Cyber Security Management" in the Journal of Parallel and Distributed Computing (JPDC), 2022. Moreover, his papers "Characterizing the Landscape of COVID-19 themed Cyber Attacks and Defenses" and "Data-Driven Characterization and Detection of COVID-19 Themed Malicious Websites" used in this dissertation, are both published in the 2020 IEEE International Conference on Intelligence and Security Informatics (ISI 2020). His previous research has been published in peer-reviewed ACM SafeConfig 2017 (ACM CCS workshop), ICCCN 2017 conference, and International Journal of Distributed Sensor Networks (IJDSN). At UTSA, Mir has worked as both Research Assistant and Teaching Assistant to excel in his future career in academia. Before joining UTSA, he worked as a teaching assistant at UNCC for several CS and IT Security courses. Prior to that, Mir also worked as a Software Engineer at the Samsung Research and Development Institute in Bangladesh. After this Doctoral degree, Mir will be joining as a tenure-track Assistant Professor of Computer Science at Appalachian State University (ASU) to start his career in academic research and teaching in the field of Computing and Cybersecurity.