# 1
# Introduction

With the wide spread of cyber-attack incidents, such as those recently experienced by Facebook, Dow Jones, AMC Networks, T-Mobile, Disney+, and US Customs and Border Protection [82], cyber-security has become a serious concern for organizations. A security bulletin published by the Kaspersky Lab [86] reported that 2,672,579 cyber-attacks were repelled daily by the company in 2019 (30 per second on average), which reflects the average activity of criminals involved in the creation and distribution of cyber-threats. Worldwide, cyber-attacks cost organizations an estimated US600 billion in 2017 [68] (0.8% of global income), and US5.2 trillion in additional costs and lost revenue are expected until 2024 [2].

A credible explanation for this threatening scenario, which is corroborated by many cyber-security researchers [145], is that malicious hackers are increasingly using the World Wide Web (the Web) to share knowledge and achieve their goals. Many works detail how threat actors rely on online hacker communities to (1) identify software vulnerabilities, (2) create or purchase exploits, (3) choose a target and recruit collaborators, (4) obtain access to the infrastructure needed, and (5) plan and execute the attack [152], making what was once a hard-to-penetrate market accessible to a much wider population [126, 145]. Although this hacker behavior helps to produce a huge amount of malware, it also provides intelligence for defenders, as the information shared by hackers on online communities can be leveraged as precursors to various types of cyber-attacks [10, 11, 51, 83, 148]. Thus, a deep understanding of the adversaries present in those environments can help organizations deal with the risk of attacks, moving their perspective toward a more proactive, intelligence-driven security.

The state of the art for cyber-security primarily provides information on what is already deployed by cyber-attackers, setting a dominant viewpoint of cyber-defense that solely focuses on the defender's environment [144, 145].



https://doi.org/10.1017/9781108869003.003 Published online by Cambridge University Press

For example, consider the current technologies used for quantifying, monitoring, and managing cyber-risk. On one hand, most of those tools only consider technical characteristics of the defender's networked assets and their vulnerabilities, such as the number of public-facing hosts, the number of vulnerable software products they run, and the ease of exploiting these software vulnerabilities. These tools use a wide range of data sources, such as security advisory databases, software vendor websites, penetration testing tools, and vulnerability databases. On the other hand, most of those security tools focus on detecting abnormalities already present in the defender's systems and networks, such as traditional firewalls, intrusion detection systems, spam filtering, and technologies that leverage anomaly detection or threat signatures using data shared among different organizations. This security perspective has a main shortcoming: the lack of a holistic consideration of the attacker's activity. Fortunately, in the recent few years, significant interest has grown towards tools, systems, and analyses for understanding and monitoring the hacker activity using cyber-threat intelligence gathered from online hacker communities. This trend arose as a response to the growing broad realization of the importance of the attacker's role, allowing organizations to effectively deploy security measures and allocate resources by exploring the ability to anticipate future cyber-attacks. The speed of those contemporary attacks, along with the high costs of remediation for recovering from the reputation, revenue, or data losses produced, overall incentives avoidance and impacts limitation over response. For instance, according to the National Cybersecurity Institute at Excelsior College (NCI), about 60% of small and mid-sized businesses that are seriously breached go out of business within six months [89], strengthening the security specialists' claim toward prevention.

In order to understand adversarial settings and achieve proactive cyber-threat intelligence, important information should be continuously mined from the ever-evolving malicious hacking communities, especially from the main platforms leveraged by malicious hackers to share knowledge: forums and marketplaces. By analyzing data about market dynamics within those communities, the rise and fall of particular personalities and venues, the nature of the conversations that take place in the forums, and the overall evolution of these communities, organizations can design more accurate attack prediction systems. Those attack predictions, in turn, can lead to a variety of strategic decisions to avoid infections, including prioritizing certain patches, discontinuing the use of a piece of software, purchasing or developing software, and segregating certain computers from the rest of the network. Thus, the explosive increase in popularity of exploit markets and hacker forums existing in and across all layers of the Web (the surface-web, the deepweb, and

https://doi.org/10.1017/9781108869003.003 Published online by Cambridge University Press

the so-called darkweb) also has a bright side for cyber-security. The cyber-defenders can now leverage the hackers' digital traces existing in those environments to get valuable insights into evolving cyber-threats and into a pending cyber-offensive well before malicious activity is detected on a target system [101, 102, 127, 155].

For instance, the WannaCry ransomware attack directed against hospitals in the United Kingdom and numerous other worldwide targets was discussed several weeks prior on a darkweb forum [176]. Hackers likely involved in this attack discussed the number of unpatched machines, the exploit to be used, the industry verticals, and the method of attack (ransomware) – which encrypts the hard drive data and demands a ransom payment to have it recovered (see the WannaCry's home screen in Figure 1.1). The discussions were made in several languages, and medical institutions were chosen as prime targets based on the history of paid ransom from similar institutions. WannaCry's example demonstrates how malicious hacker platforms provide valuable information regarding the capabilities and intent of threat actors. The attack did not come out of nowhere: it exploited a known flaw in Microsoft Windows that hackers knew was left unpatched on many machines. They just needed to share the

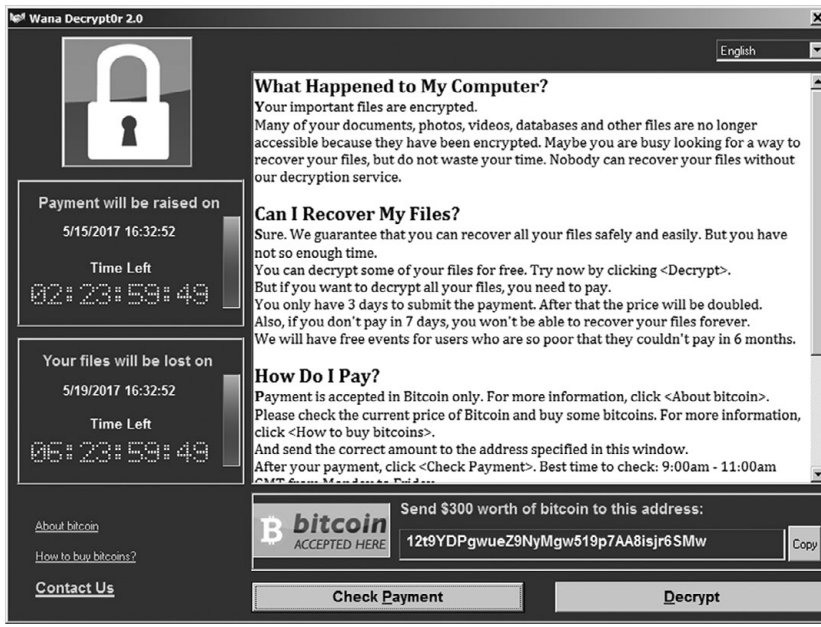


Figure 1.1 WannaCry's home screen with instructions for victims of the cyber-attack.

https://doi.org/10.1017/9781108869003.003 Published online by Cambridge University Press

information and get online help to troubleshoot obstacles. Thus, by looking in the right place at the right time, defenders can mine the interest of malicious hackers, adopting a proactive behavior that can protect their assets before the attackers operate.

This book is intended to give an overarching view into how to explore malicious hacker communities to achieve proactive cyber-threat intelligence. After introducing those communities and giving general information on the cyber-security domain, we conduct a series of studies that demonstrate how artificial intelligence, machine learning, and social network analysis techniques can be used to make sense out of large quantities of hacker community data for security purposes. We divide those studies into two parts. In the first part, formed by Chapters 3, 4, and 5, we use those techniques to gain additional insight into the structure of online hacker communities as well as the behavior of their members. We focus there on scrutinizing the threat actors creating and distributing malicious code online and getting knowledge about dynamic reputation systems, user engagement, and highly specialized groups of hackers that can aid in the identification of credible threats. In the second part, formed by Chapters 6, 7, 8, and 9, we leverage those techniques to effectively predict future cyber-threats, either by identifying exploits-in-the-wild, predicting enterprise-targeted external cyber-attacks, or finding at-risk systems. We focus there on analyzing online hacker communication to find confident patterns of attack behavior, leveraging those patterns to anticipate future cyber-incidents. Table 1.1 summarizes the two main parts of this book.

Throughout the book, we use data collected from a commercial version of the system that we proposed in [126, 145]. This system, currently maintained and provided by Cyber Reconnaissance, Inc. (CYR3CON) [42] through API, is responsible for gathering cyber-threat intelligence from various social platforms of the Web. We query data originally hosted across multiple network

Table 1.1 *The two main parts of the book*

| | | |
|---|---|---|
| Understanding the Behavior of Malicious Hackers | Chapter 3 | Mining Key-hackers |
| | Chapter 4 | Reasoning about Hacker Engagement |
| | Chapter 5 | Uncovering Communities of Malware and Exploit Vendors |
| Predicting Imminent Cyber-threats | Chapter 6 | Identifying Exploits in the Wild Proactively |
| | Chapter 7 | Predicting Enterprise-Targeted External Cyber-attacks |
| | Chapter 8 | Bringing Social Network Analysis to Aid in Cyber-attack Prediction |
| | Chapter 9 | Finding At-Risk Systems without Software Vulnerability Identifiers (CVEs) |

https://doi.org/10.1017/9781108869003.003 Published online by Cambridge University Press

protocols (the surface-web, the deepweb, and the "darkweb"), informing in each chapter what type of data is being used for the corresponding study. In addition to data obtained directly from the hacker discussions on forums and marketplaces, data from other resources, including social media platforms, Chan sites, paste sites,[1] exploit archives, vulnerability databases, and bug bounty programs,[2] are obtained through this system.

The remainder of the book is structured as follows. Chapter 3 leverages content, social network, and seniority analysis to mine key-hackers on hacking forums, identifying reputable individuals who are likely to succeed in their cyber-criminal goals. Next, as hackers often use Web platforms to advertise and recruit collaborators, Chapter 4 analyzes forum engagement by predicting where and when hackers will post a message in the near future given their recurrent interactions with other hackers. After that, Chapter 5 demonstrates how vendors of malware and malicious exploits organically form hidden organizations on online markets, analyzing the similarity of product offerings in different networks.

The next four chapters (Chapters 6, 7, 8, and 9) directly measure the risk of cyber-attacks, either by predicting exploits-in-the-wild, anticipating cyber-incidents at particular enterprises, or conducting assessment of threats to particular systems. Particularly, Chapter 6 predicts if exploits are going to be used in the wild by analyzing multiple sources of threat intelligence generated after vulnerability disclosures. Differently, Chapter 7 describes a temporal logical framework to learn rules that correlate malicious hacking activity with real-world cyber-incidents, leveraging these rules for predicting enterprise-targeted external cyber-attacks. With the same prediction goal, Chapter 8 measures network features and user/thread posting statistics to hypothesize that the interaction dynamics focused on a set of specialized users and the attention broadcast by them to others can be relevant to generating cyber-attack warnings. Finally, Chapter 9 looks at online hacker discussions to identify platforms, vendors, and products likely to be at risk, gathering indicators regarding the hacker capability of targeting systems to conduct the corresponding threat assessment.

Chapter 10 wraps up the book, presenting the overall contributions of all studies conducted, the challenges overcome, and future research directions to further empower cyber-security. The latter is done by addressing current limitations of the book as well as by discussing new models and methods that will contribute to enhancing proactive cyber-threat intelligence.

[1] Type of online content hosting service where users can store plain text, e.g., source code snippets.
[2] Deal offered by sites and organizations by which individuals can receive recognition and compensation for reporting bugs.

https://doi.org/10.1017/9781108869003.003 Published online by Cambridge University Press