CAMBRIDGE

Cambridge University Press
978-1-107-18577-7 — Darkweb Cyber Threat Intelligence Mining
John Robertson , Ahmad Diab , Ericsson Marin , Eric Nunes , Vivin Paliath , Jana Shakarian , Paulo Shakarian
Excerpt
More Information

1

Introduction

Recently, the online market for exploit kits, malware, botnet rentals, tutorials, and other hacking products has continued to evolve, and what was once a rather hard-to-penetrate and exclusive market—whose buyers were primarily western governments [95]—has now become more accessible to a much wider population. Specifically, the darknet—portions of the Internet accessible through anonymization protocols such as Tor and i2p—has become populated with a variety of markets specializing in such products [94, 2]. In particular, 2015 saw the introduction of darknet markets specializing in zero-day exploit kits, designed to leverage previously undiscovered vulnerabilities. These exploit kits are difficult and time consuming to develop—and often are sold at premium prices.

The explosive increase in popularity of exploit markets and hacker forums presents a valuable opportunity to cyber defenders. These online communities provide a new source of information about potential adversaries, consequently forming the nascent cyber threat intelligence industry. Pre-reconnaissance cyber threat intelligence refers to information gathered prior to a malicious actor interacting with a defended computer system. To provide a concrete example demonstrating the importance of pre-reconnaissance cyber threat intelligence, consider the case study shown in Table 1.1. A Microsoft Windows vulnerability was identified in February 2015. Microsoft's public press release regarding this vulnerability was essentially their way of warning customers of a security flaw. At the time of its release, there was no publicly known method to leverage this flaw in a cyber-attack (i.e., an available exploit). However, about a month later, an exploit was found to be on sale in a darknet exploit marketplace. It was not until July when FireEye, a major cybersecurity firm, identified that the Dyre Banking Trojan, designed to steal credit card information, exploited this particular vulnerability. This vignette illustrates how threat warnings gathered from the darknet can provide valuable information for security

Table 1.1. *Exploit example*

Timeline	Event
February 2015	Microsoft identifies Windows vulnerability MS15-010/CVE 2015-0057 for remote code execution. There was no publicly known exploit at the time the vulnerability was released.
April 2015	An exploit for MS15-010/CVE 2015-0057 was found on a darknet market on sale for 48 BTC (around $10,000–15,000 at the time).
July 2015	FireEye identified that the Dyre Banking Trojan, designed to steal credit card number, actually exploited this vulnerability.[1]

professionals in the form of early-warning threat indicators. Between Dyre and the similar Dridex banking trojan, nearly 6 out of every 10 global organizations were affected, a shocking statistic.[2]

In another instance, 17-year-old hacker Sergey Taraspov from St. Petersburg, Russia, along with a small team of hackers, allegedly wrote a piece of malware that targeted point-of-sale (POS) software and sold it for $2,000 on a Russian forum-cum-marketplace. This malware was, in turn, used by around forty individuals to steal over 110 million American credit card numbers in the "Target" data breach of 2013.[3]

It is now possible, and quite common, to leverage data-mining and machine-learning techniques to make sense out of large quantities of data. After further motivating the importance of cyber threat intelligence and discussing online hacker communities in detail, we will discuss specifically how data-mining and machine-learning techniques can be applied to the cyber threat intelligence domain. Using these techniques, we will be able to gain additional insight into the structure of online hacker communities as well as the behavior of individuals within them. We will also draw from the artificial intelligence literature to build threat models, informed from the data mined from hacker communities, to provide system-specific cyber intelligence.

This book is intended to give an overarching view into the burgeoning field of cyber threat intelligence. The remainder of the book is structured as follows: Chapter 2 will further motivate the use of cyber threat intelligence by organizations, discussing and addressing some of the difficulties in realizing wide-scale cyber threat intelligence adoption. Chapter 3, will discuss, in detail, the online

[1] https://www.fireeye.com/blog/threat-research/2015/07/dyre_banking_trojan.html
[2] https://www.fireeye.com/blog/threat-research/2015/06/evolution_of_dridex.html
[3] http://www.nbcnews.com/news/world/skilled-cheap-russian-hackers-power-american-cybercrime-n22371

hacker communities from which a lot of cyber threat intelligence is derived. Chapter 4 will introduce techniques to build a large-scale scraping and parsing infrastructure to gather data from darknet communities, discussing some of the associated challenges as well as the performance of various data-mining and machine-learning techniques in the context of gathering cyber threat intelligence. Chapter 5 presents a number of case studies that illustrate how the collected data can be translated to actionable, real-world cyber threat intelligence and uses unsupervised learning techniques to cluster products from darknet markets into specific categories.

The next two chapters (Chapter 6 and 7) introduce more sophisticated models that use the aggregated data from the darknet in interesting ways to provide rich threat intelligence. Chapter 6 frames the host defense scenario as a security game, presenting a game theoretic framework that informs the attacker model with real-world darknet exploit data and is capable of making system-specific policy recommendations. The model presented in Chapter 7 also leverages exploit information, but in the context of defending industrial control systems (ICS): IT infrastructure that controls physical systems (electricity, water, industrial machinery, etc.).

Chapter 8 wraps up the book, discussing ongoing work as well as the unique challenges associated with sociocultural modeling of cyber threat actors and why they necessitate further advances in artificial intelligence—particularly with regard to interdisciplinary efforts with the social sciences.