Digital Investigation 11 (2014) S114–S123

Contents lists available at ScienceDirect

Digital Investigation

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

Multidimensional investigation of source port 0 probing

Elias Bou-Harb*, Nour-Eddine Lakhdari, Hamad Binsalleeh, Mourad Debbabi

The National Cyber-Forensics and Training Alliance (NCFTA) Canada, Concordia University, Concordia Institute for Information Systems Engineering, Montreal, Quebec, Canada H3G1M8

Keywords:
Port 0 probing
Darknet
Malware
Passive DNS
Data correlation

A B S T R A C T

During November 2013, the operational cyber/network security community reported an unprecedented increase of traffic originating from source port 0. This event was deemed as malicious although its core aim and mechanism were obscured. This paper investigates that event using a multifaceted approach that leverages three real network security feeds that we receive on a daily basis, namely, darknet, passive DNS and malware data. The goal is to analyze such event from the perspectives of those feeds in order to generate significant insights and inferences that could contribute to disclosing the inner details of that incident. The approach extracts and subsequently fingerprints such malicious traffic from the received darknet data. By executing unsupervised machine learning techniques on the extracted traffic, we disclose clusters of activities that share similar machinery. Further, by employing a set of statistical-based behavioral analytics, we capture the mechanisms of those clusters, including their strategies, techniques and nature. We consequently correlate the sources with passive DNS in order to investigate their maliciousness. Moreover, to examine if the sources are malware contaminated, we execute a correlation mechanism between the darknet data and the malware feeds. The outcome reveals that such traffic indeed is reconnaissance/probing activities originating from three different horizontal scans utilizing packets with a TCP header length of 0 or packets with odd flag combinations. The results as well demonstrate that 28% of the scanning sources host malicious/blacklisted domains as they are often used for spamming, phishing and other fraud activities. Additionally, the outcome portrays that the bot probing sources are infected by 'Virus.Win32.Sality'. By correlating various evidence, we confirm that such malware specimen is in fact responsible for part of the source port 0 probing event. We concur that this work is a first attempt ever to comprehend the machinery of such unique event and we hope that the community could consider it as a building block for auxiliary analysis and investigation.

Introduction

Probing is often defined as the task of scanning enterprise networks or Internet wide services in an attempt to search for vulnerabilities or ways to infiltrate IT assets. It is typically considered a significant cyber security concern due to the fact that it is commonly the primary stage of an intrusion attempt that enables an attacker to remotely locate, target, and subsequently exploit vulnerable systems. For instance, hackers have employed probing techniques to identify numerous misconfigured proxy servers at the New York Times to access a sensitive database that disclosed more than 3000 social security numbers (New York Times internal network hacked). Further, the United States Computer Emergency Readiness Team (US-CERT) revealed that attackers had performed coordinated probing activities to

* Corresponding author. Tel.: +1 5146495049.
E-mail address: e_bouh@encs.concordia.ca (E. Bou-Harb).

http://dx.doi.org/10.1016/j.diin.2014.05.012
1742-2876/

fingerprint WordPress sites and consequently launched their targeted attacks (WordPress sites targeted). Moreover, it was disclosed that hackers had leveraged sophisticated scanning events to orchestrate multiple breaches of Sony's PlayStation Network taking it offline for 24 days and costing the company an estimated $171 million (PlayStation network outage caused by 'external intrusion'). More alarming, a recent incident reported that attackers had escalated a series of "surveillance missions" against cyber-physical infrastructure operating various US energy firms that permitted the hackers to infiltrate the control-system software and subsequently manipulate oil and gas pipelines (Iran hacks energy firms). Although, on one hand, Panjwani et al. (2005) concluded that a momentous 50% of attacks against cyber systems are preceded by some form of probing activity, however, on the other hand, such observed activities might simply reflect the "background radiation" of various Internet-scale random scanning activities, remnants of past worm/virus outbreaks, or other malware activities on the global Internet at large (Moore et al., 2004). Indeed, it is known in the cyber security and digital forensics communities that it is a daunting task to infer and uncover the intention, mechanism and the nature of the perceived probing activities (Jin et al., 2007).

Background

On November 2nd, 2013, security researchers at Cisco Systems reported that their worldwide deployed sensors have detected a massive increase in TCP source port 0 traffic.[1] They further elaborated that the magnitude observed by the sensors was elevated by 20 times than typical traffic originating from the same port and transport protocol on other days. According to the researchers, such event renders the largest spike in network traffic originating from TCP source port 0 in the last decade. In a follow-up discussion,[2] the researchers noted that such port, according to its RFC, is engineered to be reserved, and that such traffic could be used to fingerprint various operating systems. Additionally, the security researchers speculated about the aim, mechanism and source of that traffic by stating that such rare event could be some sort of a research project, a malware infected probing botnet, a targeted reconnaissance event aiming to launch an immediate or a prolonged malicious task, or even a broken embedded device or a new piece of malware with a bug in its scanning code.

The event was interestingly also observed by DShield/Internet Storm Center (ISC).[3] ISC data comprises of millions of intrusion detection log entries gathered daily from sensors covering more than 500 thousand Internet Protocol (IP) addresses in over 50 countries. As shown in Fig. 1, the event was apparent on four days, namely, November 2nd,[4] November 21st, November 24th and November 25th, 2013. The latter fact is particularly demonstrated by the peaks of the TCP ratio of port 0 on those specific days.

[1] http://tinyurl.com/pds443n.
[2] http://tinyurl.com/n8j58hs.
[3] http://www.dshield.org/port.html.
[4] Coinciding with Cisco reports although not quite as significant.

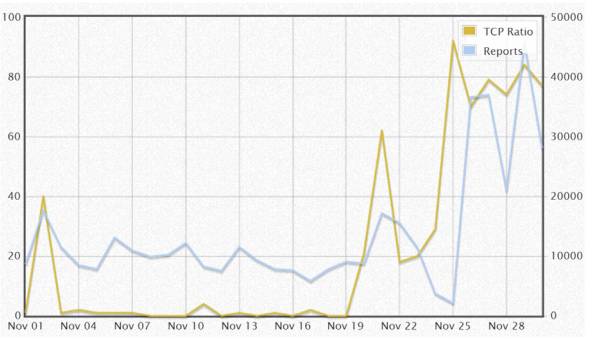

Fig. 1. The source port 0 event as observed by DShield/Internet Storm Center.

Contributions

Motivated by the requirement to shed the light on that incident in order to generate inferences and insights that could contribute in disclosing the inner details of such an unprecedented event, we frame the paper's contributions as follows:

- Proposing a multifaceted approach that leverages three real network security feeds. The approach exploits darknet data (i.e., Internet traffic destined to half a million routable yet unallocated IP addresses) to extract, analyze and uncover the machinery of such traffic. Further, the approach employs correlation between the latter and passive DNS (i.e., Internet-wide authoritative DNS responses) to study the maliciousness of the such traffic. Moreover, the proposed approach correlates darknet-extracted traffic with malware feeds to answer questions related to contamination and attribution. To the best of our knowledge, 1) the proposed approach that correlates those three feeds in an effort to understand a cyber event has never been attempted before and 2) the yielded outcome from adopting such an approach related to this specific event is unique in the literature.
- Employing 1) machine learning data clustering techniques to partition the port 0 traffic according to similar machinery and 2) a set of novel behavioral analytics that scrutinize such traffic to capture the behavior of the sources.
- Evaluating the proposed approach using 30 GB of real darknet traffic, 1.4 billion DNS records and 30 million malware analysis reports.

Organization

The remaining of this paper is organized as follows. In the following section, we present the proposed approach. Specifically, we elaborate on how source port 0 traffic is extracted and fingerprinted from darknet traffic in addition to presenting the analytics that are used to disclose the machinery of such traffic. Further, we discuss the goals and