The Journal of Systems and Software 106 (2015) 28–41

Contents lists available at ScienceDirect

The Journal of Systems and Software

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

Information infrastructure risk prediction through platform vulnerability analysis

Aristeidis Chatzipoulidis*, Dimitrios Michalopoulos, Ioannis Mavridis

Department of Applied Informatics, University of Macedonia, 156 Egnatia Street, 54636 Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 6 December 2013
Revised 31 March 2015
Accepted 12 April 2015
Available online 18 April 2015

Keywords:
Zero-day risk
Kolmogorov-Smirnov test
Bayesian belief network (BBN)

ABSTRACT

The protection of information infrastructures is important for the function of other infrastructure sectors. As vital parts for the information infrastructure operation, software-based platforms, face a series of vulnerabilities and threats. This paper aims to provide a complementary approach to existing vulnerability prediction solutions and launch the measurement of zero-day risk by introducing a risk prediction methodology for an information infrastructure. The proposed methodology consists of four steps and utilizes the outcomes of a proper analysis of security measurements provided by specifications from the Security Content Automation Protocol. First, we identify software platform assets that support an information infrastructure and second we measure the historical rate of vulnerability occurrences. Third, we use a distribution fitting procedure to estimate the statistical correlation between empirical and reference probability distributions and verify the statistical significance of the distribution fitting results with the Kolmogorov—Smirnov test. Fourth, we develop conditional probability tables that constitute a Bayesian Belief Network topology as means to enable risk prediction and estimation on security properties. The practicality of the risk prediction methodology is demonstrated with an implementation example from the electronic banking sector. The contribution of the proposed methodology is to provide auditors with a proactive approach about zero-day risks.

1. Introduction

In our modern society, sectors like banking, finance, government services, and communication technologies, rely on information infrastructures to perform operational activities. Each information infrastructure relies on software and hardware platforms that constitute assets vital for the maintenance of business goals and information assurance. However, information infrastructures are constantly under threat due to a number of challenges, such as the accelerating change of technology, open networks, third party dependencies, outsourcing risk, stakeholder involvement and government requirements for stricter regulation through compliance and policies (Koons and Minoli, 2010; Masera and Fovino, 2007; Theoharidou et al., 2010; Veríssimo et al., 2006). In this regard, the concept of information infrastructure protection (IIP) is of critical importance to ensure that all sectors continue to function and interoperate in an optimum way (Rinaldi et al., 2001).

The growing concern toward zero-day vulnerabilities and threats that derive from software-based platforms and challenge the sustainability of an information infrastructure is the source of the main issue this paper tries to address. By zero-day vulnerabilities and threats we mean: a) vulnerabilities that have zero-day awareness and are unknown and b) threats that exploit zero-day vulnerabilities (Bilge and Dumitras, 2012; Zhang et al., 2011). Another matter of concern is the risk from exploiting zero-day vulnerabilities. Due to lack of a formal definition on zero-day risk in the literature, we define zero-day risk as the uncertain loss caused by a damage event and the corresponding impact for each security property, namely confidentiality, integrity and availability. Challenged by the fact that one can predict only what can be measured, we approach the above issues from a security metrics' perspective, as means to measure and prioritize vulnerability severity. Despite marked progress in security metrics, in terms of vulnerability scoring systems, the scoring of vulnerabilities is still a hot topic in research (Liu et al., 2012). This fact, combined with threats increasing in significance (Aburrous et al., 2010), creates the need to develop prediction techniques as defenses to zero-day risk.

In this paper, inspired by research on vulnerability prediction, we attempt to predict zero-day risk. We believe that early recognition of zero-day risk is in favor of the IIP, because precaution is better than cure, allowing for resource optimization and sustainable allocation of security controls. To the best of our knowledge, there is a lack in literature about measuring zero-day risk since most of the

* Corresponding author. Tel.: +302310891841; fax: +302310891812.
E-mail addresses: chataris@uom.gr (A. Chatzipoulidis), dimich@uom.gr (D. Michalopoulos), mavridis@uom.gr (I. Mavridis).

http://dx.doi.org/10.1016/j.jss.2015.04.062
0164-1212/© 2015 Elsevier Inc. All rights reserved.

published research focus on prediction of zero-day vulnerabilities alone. In this respect, the contribution of this paper is a novel risk prediction methodology as a proactive approach to IIP. The research goal is to aid the auditor during decision making in response to potential risks on a real-time basis and in advance of public disclosure, applying appropriate protective actions. The proposed methodology aim at offering promising results based on stochastic approaches and standardized metrics. To this extent, we use specifications from the Security Content Automation Protocol (SCAP) as means to enable automated software platform and vulnerability identification, quantitative measurement and comparison of results (Waltermire et al., 2011).

The paper is organized as follows: Section 2 is a background on the progress of vulnerability scoring methods. A step-by-step analysis of the proposed methodology is presented in Section 3. Section 4 demonstrates the practicality of the methodology with an implementation example from the e-banking sector. In Section 5 we present related work in terms of vulnerability prediction and in section 6 we develop and demonstrate a two-phase evaluation method for vulnerability prediction. In Section 7 we discuss the proposed methodology and outline limitations. In Section 8 we conclude and suggest future research initiatives.

2. Background

To achieve reliable risk prediction, we are looking for security metrics that a) allow for isomorphic data analysis, b) enable vulnerability, threat and damage measurement and c) use of historic data for prediction purposes. For this reason, we focus on estimating future values of vulnerability occurrences as means to discover the source of zero-day risk (Cavano, 1984; Martinez et al., 2009; Matsuo, 2003). Particularly, we are looking for security metrics that adhere to certain specifications such as a) being standardized, b) enable comparison of results and c) make their formulas available to the public in order to provide evidence of assurance and transparency. Moreover, security metrics should follow certain policies and procedures, provide quantifiable performance measures and emphasize on consistent periodic analysis of the measures data (NIST, 2008). In this respect, published research on the field of vulnerability scoring includes qualitative, quantitative and hybrid methods.

Qualitative methods evaluate the vulnerability severity using a Likert-scale metric type such as low-medium-high-critical. Representative examples are: Symantec (2000), Microsoft (2012), X-Force (1999), Qualys (1999), Secunia (2002), Redhat (2005), Mozilla (2005), Google (2007) and Vupen (2005). The subjectivity of rating values and the unavailability of vulnerability severity formulas to the public implies that the auditor has neither the proper documentation to support the results nor the opportunity to mix and compare them with other methods due to the approximation of ratings and scores. Moreover, these kind of methods are based mainly on subjective judgement and experience rather than an actual estimation of vulnerability characteristics and provide an easy to understand result.

Quantitative methods are more precise in terms of rating vulnerability characteristics and scoring severity which allows them to be more appropriate during comparison of results. Representative examples are the CVSS version 1 (Schiffman and CIAG, 2005) and 2 (Mell et al., 2007), WIVSS (Spanos et al., 2013) and Potential Loss Value (PLV) metric (Wang and Yang, 2012). These kind of methods are based on numerical and statistical techniques to calculate vulnerability severity which implies that they appear more appropriate in terms of analysing vulnerability characteristics compared to qualitative methods, however, staff expertise, increased time and data collection in the desired format are also required.

Nowadays, hybrid methods are considered as one of the best solutions for improving vulnerability scoring. Particularly, the Vulnerability Rating and Scoring System (VRSS) for qualitative rating and quantitative scoring of vulnerabilities has been proposed by Liu and Zhang (2011). In an effort to increase the score diversity even higher and provide a more distinct vulnerability scoring, VRSS improvement enables vulnerability type, in terms of Common Weakness Enumeration (CWE), to prioritize vulnerabilities through an analytic hierarchy process (Liu et al., 2012). In addition, current research proposed a novel approach to software vulnerability prioritization through a combination of fuzzy logic processes. The study highlights the fuzziness of stakeholder involvement and how different weights of evaluation criteria, which are based on CVSS v2 metrics, affect vulnerability prioritization (Huang et al., 2013). These kind of methods use a combination of qualitative and quantitative techniques and appears to offer a more complete and systematic approach to vulnerability severity calculation.

In summary, the method employed to calculate vulnerability severity should provide an objective measurement of vulnerability characteristics and enable vulnerability prioritization. It follows that it is often not feasible to analyze all vulnerabilities that affect the target of analysis at once hence, we require an automated specification which will enable interoperability with other specifications and provide at the same time updated information about vulnerability characteristics. This implies that predicting risk is only as good as the vulnerability data it is built upon.

3. Proposed methodology

In this section, we describe step-by-step the proposed risk prediction methodology through platform vulnerability analysis based on SCAP specifications (Waltermire et al., 2011). The first three steps include the platform vulnerability analysis and the fourth step is the information infrastructure risk prediction process. The steps are the following:

Step 1: Platform identification
Step 2: Vulnerability history
Step 3: Vulnerability prediction
Step 4: Risk prediction

3.1. Platform identification

The objective of the first step is to identify and classify information infrastructure supporting platforms. To achieve this, we use the Common Platform Enumeration (CPE). CPE is a structured naming scheme for Information Technology (IT) systems, platforms and packages. The most common is version 2.2 (Buttner, 2009), a specification which is included in SCAP 1.0 and 1.1 versions. The CPE format describes platforms into specific fields, such as part, vendor, product, version, update, edition and language. CPE evolved into the 2.3 version, part of SCAP 1.2 version, with main differentiation the deployment of four additional, edition-related, fields (Cheikes et al., 2011). A CPE product dictionary represents a list of official CPE names and is provided to the public and supported by the NVD (National Vulnerability Database, 2013), a U.S. government vulnerability data resource.

3.2. Vulnerability history

The objective of the second step is to measure the historical rate of vulnerability occurrences, i.e. the rate by which the vulnerabilities of identified platforms occurred for a defined time period in the past, in our case that of a semester. To achieve this, we retrieve historical data from the NVD, in terms of the Common Vulnerabilities and Exposures (CVE) (Mell and Grance, 2002). In this case, we measure the entry type vulnerabilities. When a new vulnerability is publicly announced, a new CVE identifier is created to represent the vulnerability and CVSS base attributes are computed and added in the NVD. The CVE specification identifies two types of vulnerabilities: entries and candidates.