1. Categorizing and Predicting Invalid Vulnerabilities on Common Vulnerabilities and Exposures

Qiuyuan Chen (1); Lingfeng Bao (2); Li Li (3); Xin Xia (3); Liang Cai (1)

Source: *2018 25th Asia-Pacific Software Engineering Conference (APSEC)*, p 345-54, 2018; **ISBN-13:** 978-1-7281-1970-0; **DOI:** 10.1109/APSEC.2018.00049; **Conference:** 2018 25th Asia-Pacific Software Engineering Conference (APSEC), 4-7 Dec. 2018, Nara, Japan; **Publisher:** IEEE Computer Society, Los Alamitos, CA, USA

Author affiliation: (1) Zhejiang University, College of Computer Science and Technology, China (2) Zhejiang University City College, School of Computer and Computing Science, China (3) Monash University, Faculty of Information Technology, Melbourne, VIC, Australia

Abstract: To share vulnerability information across separate databases, tools, and services, newly identified vulnerabilities are recurrently reported to Common Vulnerabilities and Exposures (CVE) database.Unfortunately, not all vulnerability reports will be accepted. Some of them might get rejected or be accepted with disputations.In this work, we refer to those rejected or disputed CVEs as invalid vulnerability reports. Invalid vulnerability reports not only cause unnecessary efforts to confirm the vulnerability but also impact the reputation of the software vendors. In this paper, we aim to understand the root causes of invalid vulnerability reports and build a prediction model to automatically identify them.To this end, we first leverage card sorting to categorize invalid vulnerability reports, from which six main reasons are observed for rejected and disputed CVEs, respectively.Then, we propose a text mining approach to predict the invalid vulnerability reports. Our experiments reveal that the proposed text mining approach can achieve an AUC score of 0.87 for predicting invalid vulnerabilities. We also discuss the implications of our study: our categorization can be used to guide new committer to avoid these traps; some root causes of invalid CVEs can be avoided by using automatic techniques or optimizing reviewing mechanism; invalid vulnerability reports data should not be neglected. (0 refs)

Inspec controlled terms: data mining - database management systems - security of data - text analysis
Uncontrolled terms: vulnerability information - newly identified vulnerabilities - invalid vulnerabilities - common vulnerabilities and exposures database - CVE - AUC score
Classification Code: C6130S Data security - C6170K Knowledge engineering techniques - C6160 Database management systems (DBMS) - C6130D Document processing techniques
IPC Code: G06F15/18 - G06F17/21 - G06F17/30 - G06F21/00 - G06N5/04
Treatment: Practical (PRA)
Database: Inspec
Data Provider: Engineering Village