Determination of features of cyber-attack goals based on analysis of data in open security data sources

To cite this article: E V Doynikova et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 734 012160

View the article online for updates and enhancements.
Determination of features of cyber-attack goals based on analysis of data in open security data sources

E V Doynikova¹, A V Fedorchenko¹ and R O Kryukov²

¹St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14 Line, St.Petersburg, 199178, Russia
²A.F. Mozhaysky's Military-Space Academy, 13, Jdanovskaya av., St. Petersburg, 197198, Russia

E-mail: doynikova@comsec.spb.ru

Abstract. The paper analyzes security databases including attack patterns database, vulnerability database and weaknesses database. Special attention is given to the schema that underlies the attack patterns database and specifies its objects and relations between them. This scheme is used for selection of features that characterize different classes of cyber-attack goals. The paper outlines metrics of security related objects, such as attacks, weaknesses and vulnerabilities provided by different schemas, the classes of cyber-attack goals, and analyzes relations between different goals and features of cyber-attacks. The experiments demonstrated dependency between the values of selected features and their applicability for determination of different classes of cyber-attack goals.

1. Introduction

Currently the targeted cyber-attacks provide serious security risks for organizations. Complexity of protection from attacks of this type is related with the fact that they differ by the detailed preliminary analysis of weaknesses of specific organization and can include big number of steps devoted to exploration of its infrastructure, inconspicuous intrusion, attack itself, and identification prevention. At the same time complexity of such attacks (multi steps) can be the basis for their timely detection and prevention of their development in case of existence of security monitoring system within the organization. One of the essential aspects for the timely attack response and prevention of similar attacks in the future is both correct determination of goal of the detected attack and forecasting of potential attack goals within the system under analysis.

This research proposes an approach that incorporates determination of features of different attack types, analyzed system monitoring for detection of these features and forecasting attack goal using machine learning methods. This paper considers the stage of determination of features of different attack goals for the further attack goal forecasting. To this moment research community developed and filled a lot of sources of security related data that contain description and characteristics of various security related objects. In particular, in this research the following objects are analyzed: software and hardware and corresponding database National Vulnerability Database (NVD) [1] where this information is provided in the CPE (Common Platform Enumeration) format; vulnerabilities of the software and hardware and appropriate database NVD where this information is provided in the CVE (Common Vulnerabilities and Exposures) format [2]; weaknesses of information systems and
appropriate database CWE (Common Weakness Enumeration) [3]; and cyber-attacks patterns and appropriate database CAPEC (Common Attack Pattern Enumeration and Classification) [4]. The research object in this research is the set of characteristics provided in the listed databases and relations among them. They are used to select features specific for different attack goals. Earlier in [5] the authors proposed common security metrics ontology for determination of attack goals. This paper is devoted to specification of metrics and features for attack goals classification using open security data on the basis of the previously proposed ontology. Besides, it describes the experiments that analyze the values of different characteristics provided in the outlined databases for discovering the correlations among them and different categories of attacks.

The paper is organized as follows. Section 2 describes the selected security data sources. Section 3 proposes an approach for features selection and classification by the cyber-attack goals. It also provides the results of experiments. The paper ends with conclusion and description of future work.

2. Open databases as sources of security data

We understand a security database as any source of data for security management. Currently rather long list of security data sources can be constructed, including sources of data on cyber-attacks (CAPEC), weaknesses (CWE), software and hardware (NVD), vulnerabilities (CVE, NVD, “Open Source Vulnerabilities Data Base” (OSVDB) [6], “Vulnerability Notes Database” (VND) [7], SecurityFocus project with BugTraq [8], and IMB X-Force [9]), exploits (“Exploit DataBase”, EDB [10], Metasploit [11]), and configurations of the software and hardware (NVD). As it was noticed in the introduction, in this research the following databases were selected for the detailed analysis: NVD, CAPEC and CWE. We selected these databases because of the connections among them and software and hardware [5]. Namely, software and hardware provided in NVD in the CPE format have vulnerabilities represented in the CVE format and references to the weaknesses provided in the CWE database for each vulnerability. Weaknesses provided in the CWE database, in their turn, have references to the attack patterns provided in the CAPEC database. Thus, we assumed that the possible attack goals for the specific system can be determined considering objects and vulnerabilities of this system, namely, using relation between CVE and CAPEC through CWE.

The vulnerability database NVD is the most complete dictionary of software and hardware, their configurations and vulnerabilities. One of its advantages is detailed description of vulnerabilities using the set of characteristics. These characteristics can be divided in three groups: the characteristics that describe vulnerability; metrics; and the characteristics that are references to other data sources. The first group incorporates a vulnerability description and characteristics from the CVE format. The second group incorporates CVSS scores of versions 2.0 and 3.0 for vulnerabilities. These scores can take values from 0 to 10 and consist from the impact subscore and exploitability subscore. The subgroup of CVSS metrics of version 2.0 in the NVD data feeds additionally includes the privileges obtained in case of vulnerability exploitation (all, users, or other) and the metric that represents if interaction with a user is required for the vulnerability exploitation. The third group incorporates the references to the vulnerable products and their versions in the CPE format, the references to the vulnerable configurations in the CCE format, the references to the weakness category in the CWE database, and references to the descriptions in other sources and to the possible solutions.

For this research the most interesting characteristics are CVSS metrics, references to the vulnerable products and configurations, and references to the weaknesses in CWE. The CVSS scores can be used for the attack goal determination as soon as they represent impact and privileges obtained in case of vulnerability exploitation, that, in their turn, are among main characteristics of the attack goal. The references to the vulnerable products and configurations allow connecting vulnerabilities and information system under analysis. The references to the weaknesses allow connecting the analyzed system with its weaknesses, potential attacks and attack goals through its vulnerabilities.

The CWE database is the most complete weaknesses database, i.e. the database of software errors that can lead to weaknesses [3]. The weaknesses are described using the fields specified by the CWE scheme [12]. The most interesting for this research are the following fields: related attack patterns (it
includes references to the attack patterns in other databases) as soon as this field allows connecting CWE entities (weaknesses) with CAPEC entities (attacks); observed example field (includes references to the real life examples of weaknesses exploitation, usually there are vulnerabilities in the CVE format) can be used to connect NVD and CWE databases, but rather small number of real life examples limits using of this field. Other interesting fields can be used to determine different classes of attack goals. These fields are as follows: applicable platforms (it represents languages, operation systems, architectures, paradigms, and technologies where weakness can appear), if operation system is specified then it can be represented using CPE format, but there is a small number of named operation systems represented using CPE format in the CWE database, thus this field can’t be used for direct connection of software and hardware with CWE; common consequences field describes consequences of weakness exploitation including scope (what security property can be violated), impact (technical impact in case of successful vulnerability exploitation), and likelihood (how likely it is that these consequences will be, and not others); functional area specifies functional areas of software where weakness can appear (this field can take following values: authentication, authorization, code libraries, counters, cryptography, error handling, interprocess communication, file processing, logging, memory management, networking, number processing, program invocation, protection mechanism, session management, signals, string processing, or functional area independent); affected resources field is used to specify system resources that can be affected while weakness exploitation. In the CWE database the Category field is used for weaknesses classification. This field can be used for determination of attack goals considering system weaknesses.

The *attack patterns database CAPEC* incorporates methods and ways of exploitation of software or security system flaws used by malefactors for implementation of information security threat [4]. In CAPEC an hierarchical structure of attack patterns representation is used: at the lower level there are attack patterns of detailed abstraction (lower level description of the approach used for attack), they are included in the set of attack patterns of the standard abstraction (detailed description of the approach used for attack, i.e. specific method of attack is described and its implementation), the attack patterns of standard abstraction are incorporated in the attack patterns of meta abstraction (high level description of the approach used for attack, i.e. the method is described but not a specific implementation of the attack), they are aggregated in the attack categories based on some common characteristic. The categories, in their turn, are included in the attack patterns catalogue. The abstraction level is specified by the attack pattern element in scope of the CAPEC scheme [13, 14]. The last version of CAPEC database incorporates 519 attack patterns, including 61 pattern of meta abstraction, 159 attack patterns of standard abstraction and 299 patterns of detailed abstraction.

In the CAPEC database there are two main attack patterns views (they are specified by the CAPEC scheme element) – by attack technique (it incorporates 9 main categories that represent different attack techniques) and by attack scope (it incorporates 6 main categories that represent different attack scopes). The attack patterns catalogue is generated depending on the selected view. The CAPEC scheme specifies attack patterns representation and describes values and meanings of their parameters [13, 14]. The following key elements of scheme can be outlined: entities; elements; reference elements; enumerations; and attributes.

The entities of the CAPEC database are catalog, view, category, attack pattern. They define main content of the database. The catalog of attack patterns is root scheme element. The catalog is used for the numbered storing of attack patterns. It is organized using views and categories. The views represent the perspective from which the attack patterns set is generated. The categories are collections of attack patterns joint by the common characteristic. The catalog, views and categories are specified, in their turn, using the set of elements and attributes. The attack patterns are abstractions that describe how an attack is executed including the difficulties an attacker may face, techniques for their overcoming and recommended methods for attack response. The database structure is constructed using reference elements. They allow constructing hierarchical structure of the database and connecting similar patterns within the database and outside of it. The possible values of elements and attributes are specified by the enumerations.
From our point of view, for attack goal determination the most interesting element is Consequences element of attack patterns. It is specified by the following subelements: scope (violated security property), impact (technical impact in case of the successful attack), likelihood (likelihood that certain consequences will occur), note (contains additional comments), consequence ID. Scope can take the following values: confidentiality, integrity, availability, access control, accountability, authentication, authorization, non-repudiation, other. Technical impact can take the following values: modify data, read data, unreliable execution, resource consumption, execute unauthorized commands, gain privileges, bypass protection mechanism, hide activities, alter execution logic, other.

3. Common approach and experiments
The relationships between the different objects of the subject area constitutes the basis of our approach. Namely, there are the relationships between an attack, an attacker, an attacked object, its vulnerabilities and weaknesses, attack goals, and relationships between the appropriate metrics and classification of their values. To represent these relationships, we developed the security metrics ontology [15]. For this research the ontology fragment that represents the listed objects is essential, especially the fragment related with two metrics that can be used to classify attack goal, namely Scope and Technical Impact. It is provided in figure 1.

![Figure 1. Ontology fragment with attack metrics.](image-url)

To prove that attack goals can be classified using the Scope and Technical Impact metrics we analyzed the CAPEC database. It contains views with following identifiers: 1000, 3000, 2000, 282, 283, 284, 553. Totally they contain 568 attack patterns. For 307 attack patterns the elements Scope and Technical Impact have values. As it was mentioned above, Scope and Technical Impact can take different values. There are 37 different combinations of Scope and Technical Impact values for the attack patterns in the CAPEC database. These combinations can determine different attack goals. Absolute number of such combinations in the CAPEC database, considering repetitions, is 1290.

In figure 2 the statistics of joint using of possible values of Scope and Technical Impact considering the case when within the same pattern several pairs of Scope and Technical Impact values (i.e. different cases of attack consequences can occur) can exist is represented. The sum of the frequencies is 2.271126760563381. In figure 3 the statistics of joint using of possible values of Scope and Technical Impact considering the null pairs (i.e. the values of Scope and Technical Impact are
“None”) is represented. The sum of the frequencies is 4.201954397394137, as soon as all observed pairs of Impact and Scope are considered. In figure 4 the common statistics of joint using of possible values of Scope and Technical Impact is represented. The sum of the frequencies tends to 0.

The Cramér’s V for nominal features (measure of association between two nominal variables) between Scope and Technical Impact excluding patterns with empty values of Scope and Technical Impact is 0.42293458362479625. The Cramér’s V considering patterns with empty values of Scope and Technical Impact is 0.5209203125184796. It allows concluding that though these metrics are not directly dependent they correlate. Conducted experiments allow concluding that the attack goal is defined by the scope and impact as connected metrics.

**Figure 2.** The statistics of joint using of possible values of Scope (vertically) and Technical Impact (horizontally), within the same pattern several pairs of Scope and Technical Impact values can exist.

**Figure 3.** The statistics of joint using of possible values of Scope (vertically) and Technical Impact (horizontally) considering the null pairs.

**Figure 4.** The common statistics of joint using of possible values of Scope (vertically) and Technical Impact (horizontally).
In the future work we plan evolve an analysis using the rest features represented in CAPEC database and the features of weaknesses represented in CWE database. Besides, we plan to use in our research the MITRE Att@ck database that includes entities describing tactics and techniques of attackers.

4. Conclusion
The paper analyzed the attack patterns database CAPEC, vulnerabilities database NVD and weaknesses database CWE. An analysis of the CAPEC scheme allowed outlining the metrics that characterize different classes of attack goals and developing on their basis the ontology of metrics and attack goals. The experiments with CAPEC data shown an existence of the correlation between some metrics and possibility of their application for determination of attack goals classes. In the future work we plan to use conducted analysis of NVD and CWE databases, and their relations with CAPEC database, to select additional features of attack goals classes and to conduct appropriate experiments. Besides, we plan to analyze MITRE Att&ck database to outline additional features, to extend the proposed ontology and to develop the methods for determination of attack goals considering the ongoing values of the selected features on the basis of the proposed ontology.

References
[1] National Vulnerability Database Retrieved from: https://nvd.nist.gov/
[2] Common Vulnerabilities and Exposures Retrieved from: //cve.mitre.org/
[3] Common Weakness Enumeration Retrieved from: https://cwe.mitre.org/
[4] Common Attack Pattern Enumeration and Classification Retrieved from: https://capec.mitre.org/
[5] Doynikova E and Kotenko I 2018 IOP Conference Series: Materials Science and Engineering 450 052006
[6] Open Source Vulnerabilities Data Base Retrieved from: https://blog.osvdb.org
[7] Vulnerability Notes Database Retrieved from: //www.us-cert.gov
[8] Official web site of SecurityFocus company. BugTraq project. Access: http://securityfocus.com
[9] IBM X-Force Exchange. Project of IBM company Retrieved from: https://exchange.xforce.ibmcloud.com/
[10] Exploit DataBase Retrieved from: https://www.offensive-security.com/community-projects/the-exploit-database/
[11] Metasploit Retrieved from: https://www.metasploit.com/
[12] Common Weakness Enumeration schema Retrieved from: https://cwe.mitre.org/documents/schema/
[13] CAPEC Scheme Documentation. Official web-site of MITRE corporation Retrieved from: https://capec.mitre.org/documents/schema/index.html
[14] CAPEC Scheme Retrieved from: https://capec.mitre.org/data/xsd/ap_schema_v3.1.xsd
[15] Doynikova E, Fedorchenko A and Kotenko I 2019 Ontology of metrics for cyber security assessment Proceedings of the 14th International Conference on Availability, Reliability and Security p 8