利用 NVD 漏洞数据库挖掘网络攻击效果

胡 影 郑康锋 杨义先

(北京邮电大学网络与交换技术国家重点实验室信息安全中心 北京 100876)

摘 要 网络攻击效果评估技术,主要研究在目标网络环境下,如何对网络攻击的效果和能力进行评测。攻击效果评估模型的研究,需要对大量典型的攻击手段进行分析,提取出基本的网络攻击效果。本文采用 NVD(National Vulnerability Database)漏洞数据库挖掘网络攻击效果。首先对 NVD 数据库的数据进行预处理,去掉不相关的字段和不完整的数据,分解 cvss_vector 字段,提取攻击效果,转换成 NAED(Network Attack Effects Database)数据库;然后在 NAED 数据库的基础上,进行攻击效果频度分析和关联分析,提取出具有典型性、发展性、明确性和独立性的攻击效果。

关键词 美国国家漏洞数据库,网络攻击效果,机密性,完整性,可用性

Network Attack Effects Mining Based on National Vulnerability Database

HU Ying ZHENG Kang-Feng YANG Yi-Xian

(Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract Network attack effect evaluation, pays attention to how to evaluate effect and abilityof network attacks in the target environment. It is important to pick up essential attack effects based on analysis of most typical computer attacks for network attack effect evaluation model. The paper makes data mining for network attack effects based on National Vulnerability Database (NVD). Firstly, the NVD data are pretreated and converted to NAED(Network Attack Effect Database), including deleting irrelated vulnerability attributes and imperfect data, decomposing 'cvss_vector', and building attack effects. Then, the paper analyzes frequency and correlation of these attack effects, and gains representative, developing, definite and independent network attack effects.

Keywords National Vulnerability Database(NVD), Network attack effects, Confidentiality, Integrity, Availability

1 前言

随着信息网络的迅速发展,网络攻击技术也层出不穷,出现了诸如拒绝服务、扫描探测、恶意软件、伪装欺骗等攻击手段。网络攻击效果评估技术,主要研究在目标网络环境下,如何对网络攻击的效果和能力进行评测。该技术的研究,有利于提高网络设备的防御能力,检验信息系统安全。

要建立网络攻击效果评估模型,首先需要对大量典型的攻击手段进行分析,提取出基本的网络攻击效果。考虑到网络攻击多是利用信息系统的脆弱性进行攻击,所以我们可以结合目前主流的漏洞数据库(Vulnerability Database)分析攻击效果。

目前主要的漏洞数据库和漏洞列表[1],包括 CVE[2]、Bugtraq[3]、NVD[4]、CERT/CC[5]、X-Force[6]、eEye[7]、SANS[8]等。CVE(Common Vulnerability and Exposures),是由 MITRE 公司建立的一个标准化漏洞命名列表,但 CVE 不是一个漏洞数据库,只是一个标准漏洞字典。Bugtraq,是由 SecurityFocus 组织收集和发布的漏洞信息库,漏洞属性描述较完备,信息更新及时,且提供了较详细的攻击方法。但是 Bugtraq 漏洞库的属性字段之间没有很好的关联。NVD(National Vulnerability Database)是美国国家漏洞数据库,它综合了美国政府可用的各种漏洞资源,参考了安全行业的多种漏洞描述,漏洞信息描述全面。CERT/CC 库,是由计算机应急响应中心 CERT/CC 收集和发布的安全漏洞库,该库提供了一个表征漏洞严重程度的属性,描述的漏洞都经过严格验证,但漏洞个数比较少,且更新较慢。X-Force 库是 ISS 公司发布的漏洞库,该库在漏洞属性描述方面没有特殊之处,但它更新得较为及时。此外,还有诸如 eEye、SANS、Cisco、Microsoft 等各个厂商和研究机构发布的漏洞列表,主要对一些较为严重的漏洞或某类软件的漏洞进行了描述。

用于挖掘网络攻击效果的漏洞数据源,应该具有较全面的漏洞属性描述,收录了大量典型的漏洞记录,漏洞内容更新及时,并且数据库字段和关系定义清晰。考虑到以上要求,本文选择 NVD 漏洞数据库作为挖掘网络攻击效果的数据源。NVD 漏洞数据库[4],兼容 CVE 漏洞命名标准,对漏洞提供 CVSS(Common Vulnerability Scoring System)打分,并可综合 OVAL(Open Vulnerability Assessment Language)查询。

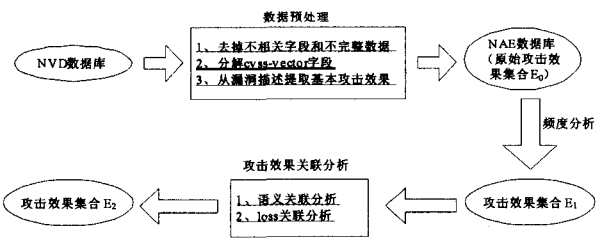

图 1 NVD 数据库挖掘网络攻击效果流程

胡 影 博士,主要研究方向:安全评估,网络攻防技术。

本文首先对NVD数据库的数据进行预处理，转换成用Access存储的NAED(Network Attack Effect Database)数据库；然后利用SPSS软件在NAED数据库的基础上，对原始网络攻击效果集合E_0进行效果频度分析，提取出典型且具有发展性的攻击效果，构成网络攻击效果集合E_1；最后在E_1的基础上，进行效果关联分析，提取出意义明确且独立的网络攻击效果集合E_2。整个挖掘过程如图1所示。

2 数据预处理

2.1 NVD漏洞数据库字段过滤

NVD数据库的主要字段，包括漏洞名称、类型、描述、发现日期、公布日期、修改日期、严重程度、CVSS分数、CVSS向量、解决方法、效果、类型、参考、软件等。

我们收集的原始NVD漏洞数据库，以XML格式记录了2004到2007四年的漏洞记录。考虑到挖掘网络攻击效果的需要，我们只保留一些与效果比较相关的字段，如表1所示。

表1 保留属性列表

数据库字段	字段含义
name	CVE-ID名称
seq	CVE序列号码
severity	漏洞严重程度
cvss_score	CVSS分数
cvss_vector	CVSS向量
desc	漏洞描述
loss_types	漏洞攻击效果类型
vuln_types	漏洞类型
range	漏洞利用特征

2.2 CVSS_Vector字段分解

NVD数据库的cvss_vector字段，描述了CVSS基本评价中各个因素的取值。CVSS的基本评价因素，可参见文[9]。其中Impact Bias，用来衡量被攻击目标对机密性、完整性、可用性哪方面更敏感。在NVD数据库的CVSS打分中，不考虑目标的偏向性，将B都取值为Normal。

考虑到CVSS_Vector的评价因素，与攻击效果的挖掘有关，所以将各个评价因素从Vector中分解出来，而由于B的取值不具有代表性，因而不考虑在内。

2.3 攻击效果提取

NVD数据库的desc字段，含有比较详细的漏洞描述。我们通过分析，发现描述的格式通常是："…allows remote/local attackers/users to … via/by…"，"to"之后通常表示该漏洞被利用的攻击效果，"via/by"之后通常表示所用的攻击方法。因而，我们可通过编写Perl程序对desc字段里"allow … to"和"via/by"之间的内容进行提取，得到每个漏洞对应的基本攻击效果描述。

例如CVE-2005-0012漏洞，漏洞描述为"Format string vulnerability in the a_Interface_msg function in Dillo before 0.8.3-r4 allows remote attackers to execute arbitrary code via format string specifiers in a web page."，通过过滤"allows remote attackers…via"可提取出该漏洞的效果为"execute arbitrary code"。

2.4 NAED数据库

对NVD的XML漏洞数据文件，首先进行数据预处理，去掉不相关的字段和不完整的数据，分解cvss_vector字段，提取出基本的攻击效果。然后采用Altova公司的XMLSpy工具将NVD漏洞文件转换成Access数据库NAED(Network Attack Effect Database)，其数据表之间的关系如图2所示。

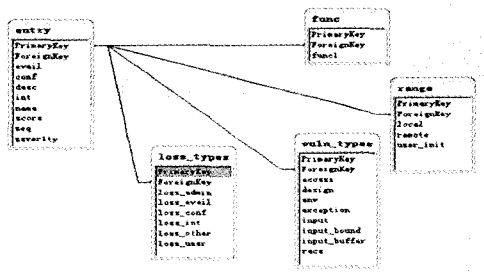

图2 NAED漏洞数据库关系图

2.5 攻击效果挖掘准则

挖掘的网络攻击效果，要求具有以下几个特点：

(1)典型性：代表典型的网络攻击效果；

(2)发展性：反映网络攻击的发展趋势；

(3)明确性：攻击效果内涵明确，不模糊；

(4)独立性：攻击效果含义相互独立。

NVD数据库进行预处理转换成NAED数据库后得到的网络攻击效果集合，是攻击效果的原始粗糙集合E_0。要得到最终的攻击效果，我们还需要在E_0的基础上，进行频度分析和关联分析。

3 数据分析

3.1 攻击效果频度分析

通过对E_0进行初步分析，发现原始效果集合E_0，根据效果所用的动词，可以分为：拒绝服务类、Execute类、Bypass类、Spoof类、Obtain/Read/Access类、Inject类、Gain类、Conduct类、Modify/Write类、Create类、Delete类、Trick类等。

对各类的效果分别进行统计频度分析，我们可以得到2004年至2007年每年效果的频度大小和趋势变化。从中选择平均频度较大，且有新的增长趋势的效果。

以Execute类攻击为例，Execute类主要包括9个攻击效果，它们从2004年到2007年的频度比例(%)如图3所示。此处"频度比例"，表示该效果出现的次数与该类效果出现的次数的比值。

对于Execute类9种攻击效果，从频度比例选择出平均值较高，且在2007年仍有一定数量出现的效果，得到Execute类效果={execute arbitrary code, execute arbitrary SQL, execute arbitrary PHP}，而且，由图3可以看出，"execute arbitrary code"四年内的平均频率和2007年的最新频率都是最高的；"execute arbitrary SQL"的平均频率居于第二，但是最新频率却低于"execute arbitrary PHP"，这说明从总体来看，SQL注入的漏洞在减少，而针对PHP的漏洞攻击在增多。

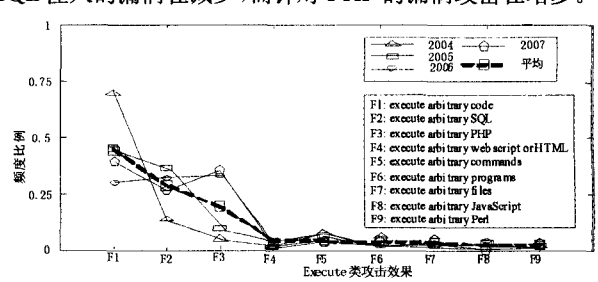

图3 Execute类攻击效果的频度分析

与Execute类效果频度分析相同，我们分别对各类效果

进行频度统计分析，提取出平均频度比例和最新频度比例(2007年)都比较高的攻击效果，得到具有典型性和发展性的效果集合 E_1。

3.2 攻击效果关联分析

在得到的效果集合 E_1 的基础上，对攻击效果进行关联分析，去掉含义不明确的效果，将互相关联的效果合并或分解，使效果保证明确性和独立性。

攻击效果之间的关联分析，包括语义关联分析和 loss 关联分析两种方法。语义关联分析，通过分析哪些效果经常同时出现，来建立语义关联。Loss 关联分析，对于逻辑意义相似的效果，通过漏洞数据库的 loss_type 和 cvss_vector，来进一步判断这些效果是合并还是分开。

(1)语义关联分析

通过分析，我们得到一些语义关联规则，例如：

• Bypass authentication，bypass access restrictions，之后多伴随着获取了某种权限或可执行其他攻击；

• Access，表示对实体的访问，经常伴随着获得对实体的某种权限。

• 拒绝服务经常和执行任意代码同时出现。

相应地，对于 Bypass 类的效果，可以只关注对防御工具的规避，而 Access 类攻击的效果可以体现在获取权限上。

(2)loss 关联分析

采用贝叶斯统计分析的方法，分析效果字段和 loss_type 字段的关联可能性，或者效果字段和 cvss_vector 中的因素字段的关联可能性。以 Bypass 类攻击效果和 loss_type 中的 avail、conf、int 的关联分析为例，如图 4 所示。

由此可见，Bypass 类的攻击效果与 Availability、Confidentiality 和 Integrity 都有关，而且与 Bypass 类攻击效果关联最强的是 Integrity，其次是 Confidentiality，最后是 Availability。

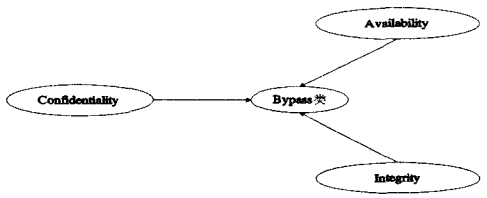

图 4 Bypass 类攻击效果和 loss_type 类 CIA 关联图

通过分析得到一些 loss 关联分析的规则，例如：

• Bypass authentication 和 Bypass access authentication 的效果取值类似；

• Read、Include、List 和 Download 类多针对机密性进行破坏；

• modify，write，overwrite，upload 针对完整性；

• delete、create、execute 针对 CIA；

• Spoof 类攻击效果，主要破坏完整性，也破坏可用性和机密性；

• 获得权限类的攻击效果，对 CIA 都有影响。

3.3 提取的攻击效果

根据攻击效果关联分析，在攻击效果集合 E_1 的基础上得到最终过滤的效果集合 E_2。另一方面，攻击效果关联分析，也有助于我们将攻击效果集合进行分类。最终的效果如表 2 所示。

表 2 最终提取的攻击效果集合 E_2

类别	主要效果
探测信息类	探测存活主机，探测端口和服务，探测操作系统指纹等
拒绝服务类	消耗系统计算资源，消耗系统存储资源，消耗网络带宽资源等
数据破坏类	篡改文件，篡改内存，篡改注册表，篡改帐户等
入侵控制类	非法执行程序，权限提升，开启后门等
信息欺骗类	电子邮件欺骗，DNS 欺骗，证书或密钥欺骗等

结束语 本文以 NVD 漏洞数据库作为数据源，挖掘网络攻击效果。通过频度分析和关联分析，挖掘网络攻击效果，如表 3 所示。

采用 NVD 漏洞数据库挖掘网络攻击效果的过程中，由于原始的攻击效果是从漏洞的描述字段进行提取，因此存在一些形式不一样、但是含义相同的效果。所以，在攻击效果挖掘中，需要进行大量的人工分析过滤工作。进一步的研究中，可以将提取的攻击效果规范入库，在此基础上还可进行聚类分析，或者针对 CIA 的取值进行更详细的关联分析，得到攻击效果与 CIA 的关联系数。

此外，从 NVD 漏洞数据库去挖掘网络攻击效果，也存在两个缺点。一是漏洞利用攻击的效果，可以从漏洞的描述得到，即只要成功利用就有此效果，比较固定，但还有很多网络攻击手段却有更多的随机性；另一方面，NVD 数据库中 CVSS 对 CIA 的评价中，只定义了三个等级(无、中等和全面控制)，量化得不够细，且评价时也可能包括了漏洞利用的潜在后果。

因而，进一步的网络攻击效果评估研究，可以一方面在 NVD 漏洞数据库的基础上进一步深入挖掘，也可以在结合其他攻击库的基础上进行挖掘。

参 考 文 献

1 张永铮，方滨兴，迟悦. 计算机弱点数据库综述与评价. 计算机科学，2006，33(8)：19～21，49

2 MITRE. Common Vulnerabilities and Exposures. http://www.cve.mitre.org

3 SecurityFocus Bugtraq Vulnerability Database. http://xforce.com/bid

4 National Vulnerability Database. http://nvd.nist.gov

5 CERT/CC. CERT/CC Vulnerability Notes Database. http://www.kb.cert.org/vuls

6 Internet Security Systems. X-Force Vulnerability Database. http://xforce.iss.net

7 eEye. Vulnerability Archive. http://www.eeye.com

8 SANS. System Administration，Networking and Security. http://www.sans.org

9 Chambers J T，Thompson J W. Common Vulnerability Scoring System Final Report and Recommendations. 2004

10 Krsul I V. Software Vulnerability Analysis：[Ph. D. dissertation]. Purdue University，1998