基于依赖搜索树的即时通信网络安全漏洞识别方法

李 立[1]
LI Li

摘 要 常规的即时通信网络安全漏洞识别方法，以漏洞分类与检测为主，忽略了字符串参数被污染导致的搜索过程中表现的依赖性，影响最终的识别结果。为此，设计了基于依赖搜索树的即时通信网络安全漏洞识别方法。从漏洞报告中提取通信网络安全漏洞特征，并将特征转换为向量形式，确保特征提取的完整性。基于依赖搜索树构建网络安全漏洞识别函数，根据网络安全漏洞特征划分漏洞类别，通过依赖搜索树的大小与复杂度，表达漏洞之间的依赖关系，从而避免漏洞识别错误识别。识别即时通信网络格式化字符串漏洞，对即时通信网络进行 hook 操作，并将格式化字符串函数中参数与控制符数量不同的漏洞识别出来，避免字符串参数被污染的问题。实验结果表明，宏查准率、宏查全率以及宏调和平均数等指标均超过了 0.97，保持了较高的识别水准，甚至在 FXJ 识别中体现了“1”的识别水准、网络安全漏洞识别效果更佳，验证了所提出方法的网络安全漏洞识别性能。

关键词 依赖搜索树；即时通信网络；安全漏洞；识别方法

doi：10.3969/j.issn.1672-9528.2024.02.033

0 引言

即时通信网络是一种实时通信系统，允许两人或多人使用网络实时传递文字消息、文件、语音与视频交流。伴随着即时通信的广泛应用，网络安全问题也日益凸显。恶意攻击者可能利用通信协议中的漏洞，窃取用户的敏感信息，破坏通信系统的正常运行。针对通信网络安全问题，研究人员设计了多种漏洞识别方法。其中，基于数据挖掘的即时通信网络安全漏洞识别方法，与基于改进残差网络的即时通信网络安全漏洞识别方法的应用较为广泛。

吕国曙等人[1]提出基于数据挖掘的电力系统网络安全漏洞识别方法，通过机器学习构建特征构建关联规则挖掘模型，根据统计学特征对电力系统网络状态进行特征抽取，但由于特征抽取过程易受到外界干扰，字符串参数易被污染。任美丽等人[2]提出基于改进残差网络的光通信网络漏洞识别研究，引入注意力机制捕捉网络漏洞特征，但是网络漏洞特征捕捉量大，未构建分类器导致分类性能不佳，特征提取完整性不高。田雪骄[3]基于改进遗传算法的通信网络漏洞数据识别方法，根据遗传算法对通信网络漏洞进行特征求解，引入变异算子提高算法的搜索效率和全局优化能力，但受蚁群算法全局检索影响，实验迭代次数高，识别效率下降。陆璐[4]基于胶囊网络和注意力机制的智能合约漏洞识别方法，考虑引入胶囊网络的胶囊层结构，以捕捉智能合约中的重要特征表示。在胶囊网络中引入注意力机制，以增强模型对重要代码片段和关键特征的关注度，但由于增强模型使得碎片化数据增多，与字符串发生交叠，出现字符串参数误识别问题。

依赖搜索树是一种高效的处理模型，它可以系统地表达和解决复杂的问题。针对漏洞识别字符串参数易受污染、识别精准度差等问题。本文引入依赖搜索树对即时通信网络安全漏洞进行识别，有效提升漏洞识别效果。

1 即时通信网络安全漏洞的依赖搜索树识别方法设计

1.1 提取通信网络安全漏洞特征

本文将通信网络的历史漏洞报告作为基础数据，从中提取通信网络安全漏洞特征，并将特征转换为向量形式，确保特征提取的完整性。将网络安全漏洞特征分为三类，元特征、文本特征、代码特征[5]。提取的代码特征为漏洞文件的代码属性，将危险函数或外部输入函数、漏洞添加或删除的代码行数、漏洞涉及的隐私信息数量、漏洞的复杂度等特征提取出来，确保漏洞特征提取的全面性[6]。漏洞的复杂度特征表示为：

$$H_n(P)=-\sum_{k=1}^{n}(p_k\times\log_n p_k) \tag{1}$$

式中：P 为通信网络安全漏洞；n 为漏洞修改的所有文件数；p_k 为添加或删除的代码行数；k 为代码；$H_n(P)$ 为安全漏洞 P 在文件 n 中的复杂度。提取的元特征为通信网络中预定义的

1. 中国联合网络通信有限公司鄂州市分公司 湖北鄂州 436000

非文本字段，例如通信时间、优先级、上次活跃时间等。提取的文本特征为通信数据的文本内容，将文本中 5 ～ 10 个字词提取出来，识别漏洞潜在的语义 [7]。将漏洞的各类特征提取出来之后，作为识别模型的输入数据，确保漏洞识别的全面性。

1.2 网络安全漏洞识别

本文根据网络安全漏洞特征划分漏洞类别，并通过依赖搜索树的大小与复杂度，表达漏洞之间的依赖关系，从而避免漏洞识别失误的问题。给定 $M=\{A,B\}$，表示即时通信子网络，节点 m 为依赖搜索树的根节点，子网中其他节点集合为 K[8]。构造以 m 为根节点且包含所有节点的最大依赖值生成依赖树 T。从漏洞当前特征搜索到源节点 m 的最大依赖值表示为：

$$M[u]=\max\{M[m]\big|m\in K\} \tag{2}$$

式中：$M[u]$ 为漏洞当前特征 u 搜索到源节点 m 的最大依赖值；$M[m]$ 为节点 m 在子网络的位置。搜索路径表示为：

$$P[u]=M[u]\cdot C(u,m) \tag{3}$$

式中：$P[u]$ 为漏洞当前特征 u 搜索到源节点 m 的搜索路径；$C(u,m)$ 为漏洞当前特征 u 搜索到源节点 m 的路径长度。重复多次 u ～ m 的搜索，找出最短路径 [9]。并构造出路径中漏洞识别的损失函数，公式如下：

$$L=-\sum_{k=1}^{K}C(u,m)\cdot P[u] \tag{4}$$

式中：L 为依赖搜索树搜索路径中漏洞识别的损失函数。将漏洞识别的指数衰减率设定为0.9，构建出安全漏洞识别函数，表达式如下：

$$\delta_m=L\left(p_k,C(u,m),M[u],W\right) \tag{5}$$

式中：δ_m 为安全漏洞识别函数表达式；W 为更新后的 BiLSTM 输出。此时的输出结果受到点对点通信方式影响 [10]，格式化字符串参数容易被污染，为此，本文对即时通信网络进行 hook 操作，并将格式化字符串函数中参数与控制符数量不同的漏洞识别出来，避免字符串参数被污染的问题 [11]。格式化字符串函数在执行覆盖操作时，覆盖返回地址攻击情况如图 1 所示。

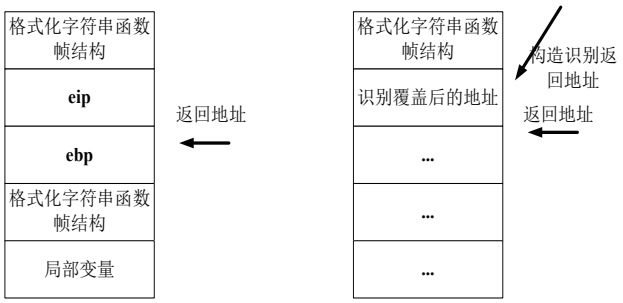

图 1 覆盖返回地址攻击示意图

如图 1 所示，左侧的框图为正常识别程序，右侧框图为存在字符串漏洞的识别程序。在字符串漏洞的程序中，符号变量有关的参数结合漏洞点上下文信息，识别异常字符串 [12]。将正常识别程序上的字符串消除，进一步识别漏洞程序，当字符串与覆盖变量有关的参数为正常值时，则漏洞不存在，有效地避免了识别失误的情况 [13]。就此实现了即时通信网络安全漏洞识别，有效提升漏洞识别精准度。实现伪代码如下。

```
plaintext 复制代码
定义节点集合 M = {A, B} 表示即时通信子网络
定义节点 m 为依赖搜索树的根节点
定义子网中其他节点集合为 K[7]

构造依赖搜索树 T，以 m 为根节点且包含所有节点
初始化漏洞当前特征为 current_feature
初始化最短路径为 infinite
while (true) {
      // 搜索漏洞依赖路径
    search_path = 搜索路径 (current_feature, m)

    // 计算路径长度
    path_length = 计算路径长度 (search_path)

   // 更新最短路径
   if (path_length < 最短路径 ) {
         最短路径 = path_length
   }

   // 计算漏洞识别损失函数
   loss = 计算损失函数 (search_path)

   // 更新漏洞识别函数
   vulnerability_function = 更新漏洞识别函数 (loss, 0.9)

   // 进行 hook 操作，避免字符串参数污染
   perform_hook_operation()

    // 识别异常字符串，消除正常字符串，进一步识别漏洞程序
   identify_abnormal_strings()
   eliminate_normal_strings()
   further_identify_vulnerability()

   // 判断漏洞是否存在
   if ( 字符串与覆盖变量有关的参数为正常值 ) {
         漏洞不存在
         break
   }
   }
  输出：最短路径，漏洞识别结果
```

上述伪代码结合了给定的段落信息，并使用了相应的描述来进行代码的编写 [14]。需要注意的是，由于具体实现细节和上下文不完全清楚，该伪代码可能需要进一步的细化和调

整。此外，还需根据实际应用场景进行具体实现和测试[15]。

2 实验

为了验证本文设计的方法是否满足即时通信网络安全漏洞识别需求，本文对上述方法进行了实验分析。最终的实验结果则以文献 [1] 基于数据挖掘的即时通信网络安全漏洞识别方法、文献 [2] 基于改进残差网络的即时通信网络安全漏洞识别方法，以及本文设计的基于依赖搜索树的即时通信网络安全漏洞识别方法进行对比的形式呈现。具体的实验准备过程以及最终的实验结果如下所示。

2.1 实验过程

本次实验选择的硬件环境为处理器 Intel(R) Core(TM) i9-10900K，主频 3.70 GHz，内存 32.0 GB。软件环境为：操作系统 Windows 10，编程语言 Python 3.7，集成开发环境 Spyder，可以满足本次实验需求。选择 CIRA-CIC-DoHBrw-2020 作为即时通信网络漏洞数据集，实验流程如图 2 所示。

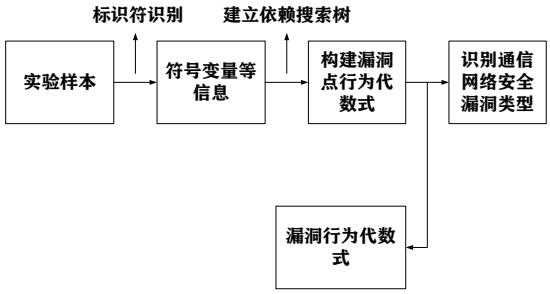

图 2 通信网络安全漏洞识别流程图

如图 2 所示，将数据集输入到实验样本中，缺失数据存在两个特征，分别为“ResponseTimeTimeMedian”和“ResponseTimeTimeSkewFromMedian”，两个特征均取 -1 时，将特征“ResponseTimeTimeMedian”缺失值填充为 1，“ResponseTimeTimeSkewFromMedian”缺失值填充为 0。从数据集中选取 30 个变量信息，包含 29 个漏洞特征属性与 1 个类别标签，并添加一列全零向量，使本次实验数据更加真实。

2.2 性能指标

为了验证本文方法的安全漏洞识别效果，通过宏查准率、宏查全率以及宏调和平均数进行实验验证。

（1）宏查准率

宏查准率具体计算公式为：

$$D=\frac{\sum_{i=1}^{N}P_i}{\delta_m N} \tag{6}$$

式中：D 为宏查准率，能够平衡网络安全漏洞类别，避免识别失误的问题；P_i 为第 i 类安全漏洞的查准率；N 为安全漏洞类别。

（2）宏查全率

宏查全率计算公式为：

$$R_l=\frac{\sum_{i=1}^{N}R_i}{\delta_m N} \tag{7}$$

式中：R_l 为宏查全率，能够确保漏洞识别的全面性；R_i 为第 i 类安全漏洞的查全率。

（3）宏调和平均数

宏调和平均数表示为：

$$F_n=\frac{D+R_l}{2} \tag{8}$$

式中：F_n 为宏调和平均数，能够均衡 D 与 R_l，F_n 越高，即时通信网络安全漏洞识别效果越佳。

根据以上指标进行通信网络安全漏洞识别效果验证。

2.3 实验结果

在上述实验条件下，本文随机选取出 Socket、RTC、SSL 等通信协议、加密协议漏洞，FXJ、Chro 等管理漏洞。并以 P_i、D、R_l、F_n 等指标，作为识别性能的最终评价指标。在其他条件均已知的情况下，将文献 [1] 基于数据挖掘的即时通信网络安全漏洞识别方法的性能指标、文献 [2] 基于改进残差网络的即时通信网络安全漏洞识别方法的性能指标以及本文设计的基于依赖搜索树的即时通信网络安全漏洞识别方法的性能指标进行对比。

如图 3 所示，Socket 为即时通信网络安全的 WebSocket 协议漏洞；RTC 为即时通信网络安全的 WebRTC 协议漏洞；SSL 为即时通信网络安全的 SSL/TLS 加密协议漏洞；FXJ 为即时通信管理漏洞；Chro 为即时通信客户端漏洞。

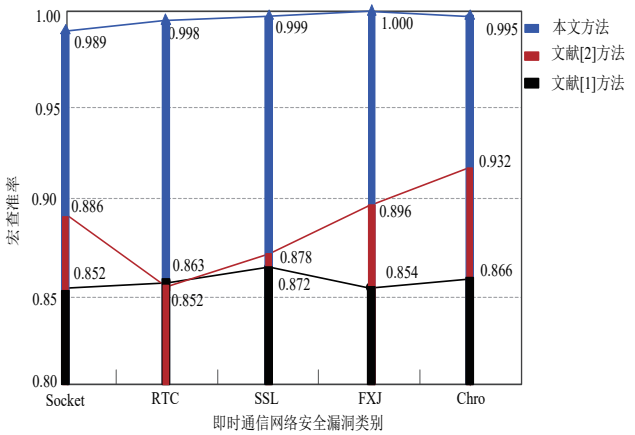

图 3 宏查准率结果

如表 1 所示，在其他条件均一致的情况下，使用文献 [1] 基于数据挖掘的即时通信网络安全漏洞识别方法之后，D 均超过了 0.85，R_l 在 0.75 ～ 0.88 的范围内变化，F_n 在 0.79 ～ 0.85 的范围内变化。由此可见，使用该方法之后，

网络安全漏洞识别存在不准确、不全面的问题，影响通信网络的安全使用。

表 1 实验结果

安全漏洞识别指标	即时通信网络安全漏洞类别				
	Socket	RTC	SSL	FXJ	Chro
文献 [1] 基于数据挖掘的即时通信网络安全漏洞识别方法					
D	0.852	0.863	0.872	0.854	0.866
F_n	0.803	0.843	0.844	0.794	0.819
R_l	0.762	0.862	0.873	0.851	0.823
文献 [2] 基于改进残差网络的即时通信网络安全漏洞识别方法					
D	0.886	0.854	0.878	0.896	0.932
F_n	0.899	0.920	0.916	0.900	0.929
R_l	0.902	0.934	0.946	0.877	0.936
本文设计的基于依赖搜索树的即时通信网络安全漏洞识别方法					
D	0.989	0.998	0.999	1.000	0.995
F_n	0.979	0.990	0.999	1.000	0.999
R_l	0.982	0.991	0.999	1.000	0.999

使用文献 [2] 基于改进残差网络的即时通信网络安全漏洞识别方法之后，D 在 0.85 ～ 0.95 的范围内变化，R_l 在 0.87 ～ 0.95 的范围内变化，F_n 在 0.89 ～ 0.93 的范围内变化。R_l 始终保持着一种平衡关系，安全漏洞识别性能较之文献 [1] 存在较大的改善。

使用本文设计的基于依赖搜索树的即时通信网络安全漏洞识别方法之后，D、R_l、F_n 等指标均超过了 0.97，保持了较高的识别水准，甚至在 FXJ 识别中体现了“1”的识别水准，网络安全漏洞识别效果更佳，符合本文研究目的。

3 结语

近些年来，互联网技术的快速发展，即时通信已成为人们日常生活中的重要部分。即时通信网络支持在线状态显示、消息提醒、文件传输等功能是一种快速、高效、便捷的通信方式。在黑客技术日益壮大的条件下，网络中标识符成为网络安全漏洞的主要攻击目标。因此，本文利用依赖搜索树，设计了即时通信网络安全漏洞识别方法。从漏洞特征、识别模型、字符串漏洞识别等方面，找出漏洞的依赖关系，按照依赖关系的大小排序逐个识别，真正意义上提高了即时网络的安全性。

参考文献：

[1] 吕国曙 , 鞠磊 . 基于数据挖掘的电力系统网络安全漏洞识别方法 [J]. 电工技术 , 2023(2):49-51.

[2] 任美丽 , 孟亮 , 李婷 . 基于改进残差网络的光通信网络漏洞自动辨识研究 [J]. 激光杂志 , 2022,43(12):133-138.

[3] 田雪骄 , 刘芸江 , 樊璞 . 基于改进遗传算法的通信网络漏洞测试数据生成研究 [J]. 中国电子科学研究院学报 , 2023, 18(6): 525-530.

[4] 陆璐 , 赖锦雄 . 基于胶囊网络和注意力机制的智能合约漏洞检测方法 [J]. 华南理工大学学报 (自然科学版), 2023, 51(5): 36-44.

[5] 朱丽娜 , 马铭芮 , 朱东昭 . 基于图神经网络和通用漏洞分析框架的 C 类语言漏洞检测方法 [J]. 信息网络安全 , 2022, 22(10): 59-68.

[6] 杨诗雨 , 桂畅旎 . 美国网络安全和基础设施安全局（CISA）网络安全漏洞治理政策分析 [J]. 中国信息安全 , 2022(6): 34-39.

[7] 赵波 , 上官晨晗 , 彭小燕 , 等 . 基于语义感知图神经网络的智能合约字节码漏洞检测方法 [J]. 工程科学与技术 , 2022, 54(2): 49-55.

[8] 胡建伟 , 赵伟 , 崔艳鹏 , 等 . 一种改进 ASTNN 网络的 PHP 代码漏洞挖掘方法 [J]. 西安电子科技大学学报 , 2020, 47(6): 164-173.

[9] 张和伟 , 王奉章 . 基于被动分簇算法的即时通信网络安全漏洞检测方法 [J]. 智能计算机与应用 , 2023, 13(7): 119-122.

[10] 田恬恬 , 朱倩倩 . 基于改进支持向量机的无线通信网络安全漏洞智能预警方法 [J]. 长江信息通信 , 2023, 36(6): 76-78.

[11] 易宏银 . LTE 无线通信网络安全漏洞及防御措施的研究 [J]. 移动信息 , 2023,45(4):131-133.

[12] 王晓燕 , 刘荷花 , 徐国华 . 基于深度信念网络的光通信网络数据异常识别研究 [J]. 激光杂志 , 2023, 44(2): 149-153.

[13] 李俊州 , 高春艳 . 基于 ESN 神经网络的光通信网络安全态势辨识研究 [J]. 激光杂志 , 2023, 44(5): 91-95.

[14] 张光华 , 刘永升 , 王鹤 , 等 . 基于 BiLSTM 和注意力机制的智能合约漏洞检测方案 [J]. 信息网络安全 , 2022(9): 46-54.

[15] 李鑫 , 杜景林 , 陈子文 , 等 . 基于双通道的智能合约漏洞检测方法 [J]. 科学技术与工程 , 2023, 23(34): 14651-14659.

【作者简介】

李立（1980—），男，湖北恩施人，本科，工程师，研究方向：承载网络的规划、维护和优化、网络安全等。

（收稿日期：2023-12-01）