

# 基于异质图网络的横向移动攻击检测方法

王 天<sup>1,2</sup>, 董 聪<sup>1,2</sup>, 刘 松<sup>1,2</sup>, 田 甜<sup>1,2</sup>, 卢志刚<sup>1,2</sup>, 姜 波<sup>1,2</sup>

<sup>1</sup>中国科学院信息工程研究所, 北京 中国 100093

<sup>2</sup>中国科学院大学网络空间安全学院, 北京 中国 100049

**摘要** 近年来, 随着互联网的高速发展, 高级持续性威胁日益频繁。而横向移动作为其攻击流程的重要一环, 是攻击者进入内网后实施攻击的主要过程, 通常伴随着内部网络的破坏以及机密数据的失窃, 对企业危害巨大。由于其高度的不可预测性和深度的隐蔽性, 传统的入侵检测技术难以应对此类攻击。因此, 本文提出一种基于异质图网络的两阶段横向移动攻击检测方法 HGLM, 通过日志图结构化的方法将横向移动攻击检测转换为一个图上的异常检测任务。首先基于内网的认证日志, 将用户与主机的登录行为图结构化, 构建用户登录图和源主机路径图, 然后在图上进行两阶段异常检测。第一阶段基于用户登录图, 使用以最大化互信息为目标的图模型进行无监督训练, 得到用户在主机间的认证行为特征表示, 再通过局部异常因子算法计算得到部分异常样本; 第二阶段基于源主机路径图和第一阶段得到的少量异常样本, 使用异质图注意力网络算法进行半监督训练, 检测横向移动攻击行为。进一步地, 本文在真实数据集 CMCS Events 上对提出的方法进行了评估和验证。实验结果表明, 本文提出的方法可以在没有样本标签的情况下有效检测横向移动行为, 在数据集上的 AUC 值达到 95.53%, 相比较于传统的 SVM 和 GBDT 模型, HGLM 不需要有标签样本, 且模型的 TPR 有超过 10%以上的大幅提升, 具有高召回率和低误报率。

**关键词** 入侵检测; 横向移动; 图神经网络; 异常检测; 恶意登录

中图法分类号 TP393.08 DOI号 10.19363/J.cnki.cn10-1380/tn.2023.08.12

## Lateral Movement Detection Using Heterogeneous Graph Network

WANG Tian<sup>1,2</sup>, DONG Cong<sup>1,2</sup>, LIU Song<sup>1,2</sup>, TIAN Tian<sup>1,2</sup>, LU Zhigang<sup>1,2</sup>, JIANG Bo<sup>1,2</sup>

<sup>1</sup> Institute of Information Engineering, Chinese Academy of Sciences, Beijing China, 100093

<sup>2</sup> School of Cyber Security, University of Chinese Academy of Sciences, Beijing China, 100049

**Abstract** With the rapid development of the Internet, advanced persistent threats have become more frequent. While, the lateral movement as an important part of its attack cycle, is the main process by which attackers conduct an attack behind the internal network and usually co-occurs with the destruction of internal networks and the theft of confidential data, causing great harm to enterprises. Due to its high unpredictability and depth of concealment, traditional intrusion detection technology is difficult to deal with such attacks. Therefore, we propose a two-phase lateral movement attack detection method HGLM based on heterogeneous graph networks, which converts lateral movement attack detection into an anomaly detection task on the graph by means of log graph structuring. First, based on the authentication log of the internal network, we construct the User Authentication Graph and Host Path Graph to represent the login behavior between users and hosts, and then perform the two-stage anomaly detection on the graphs. In the first stage, we use a graph model with the goal of maximizing mutual information for unsupervised training to learn a characteristic representation of the user's authentication behavior among hosts based on the User Authentication Graph, and then detect some abnormal samples through the Local Outlier Factor algorithm. In the second stage, we use the Heterogeneous Graph Attention Network algorithm to train a semi-supervised model which is used to detect lateral movement attacks based on the Host Path Graph and a small number of abnormal samples obtained in the first stage. Furthermore, our approach is evaluated and verified on the dataset CMCS Events. The experimental results show our approach can effectively detect lateral movement behavior without sample labels, with an AUC value of 95.53% on the dataset. Compared with traditional SVM and GBDT models, HGLM does not need labeled samples, and the TPR of the model has a substantial improvement of more than 10%, with a high recall and low false alarm rate.

**Key words** intrusion detection; lateral movement; graph neural network; anomaly detection; malicious login

通讯作者: 姜波, 博士, 副研究员, Email: jiangbo@iie.ac.cn.

本论文得到国家重点研发计划(No. 2019QY1300, No. 2018YFB0803602), 中国科学院青年创新促进会(No. 2021156), 中国科学院战略性先导 C 类(No. XDC02040100), 国家自然科学基金(No. 61802404)的资助。这项工作也得到了中国科学院网络评估技术重点实验室和北京市网络安全与保护技术重点实验室的部分支持。

收稿日期: 2021-02-01; 修改日期: 2021-03-10; 定稿日期: 2023-08-10

## 1 引言

近年来,随着互联网的高速发展,网络环境变得日益复杂,网络攻击愈发呈现出一种高发频发的态势。其中,高级持续性威胁<sup>[1]</sup>(Advanced Persistent Threat, APT)受益于攻击手法的进步以及攻击组织性的提高,攻击日益频繁。相比于其他攻击,APT 攻击具有更长的潜伏周期以及更大的破坏力,如窃取机密信息<sup>[2]</sup>、破坏电网<sup>[3]</sup>等。其攻击手法也更加全面且能够通过对目标的长期观察开发定制化攻击工具,威胁巨大。因此,对 APT 攻击的检测和防护已成为当前网络安全中亟待解决的问题。

横向移动作为 APT 攻击极为重要的一环,是攻击者进入内网后实施攻击的主要过程。根据 ATT&CK 框架<sup>[4]</sup>,横向移动由攻击者用来进入和控制网络上的远程系统的技术组成。当攻击者成功入侵到网络并建立落脚点后,为了下一步的攻击和收集目标网络的信息,通常都会在网络中进行横向移动,最终获得整个网络的控制权,达成破坏目标网络或基础设施、窃取机密数据或核心知识产权等目的,危害巨大。根据移动媒介的不同可将其分为基于恶意文件的横向移动和基于用户凭证的横向移动。基于恶意文件的方式主要是通过执行和传播恶意文件的方法进行移动,包括利用远程服务漏洞、内部钓鱼等战术,如 APT28 利用 Windows SMB 远程执行代码漏洞进行横向移动<sup>[5]</sup>。基于用户凭证的方式主要是通过窃取凭证伪装成正常用户执行操作进行移动,包括远程服务会话劫持、远程服务连接等战术,如 APT39 通过 RDP 协议在内部网络执行操作<sup>[6]</sup>。后者由于伪装成正常用户进行操作的原因具有高度的隐蔽性,检测难度更大。

为了解决上述问题,本文提出一种基于异质图网络的两阶段横向移动攻击检测方法 HGLM。基于内网的认证日志,将用户与主机的登录行为图结构化,构建用户登录图和源主机路径图,然后在图上进行两阶段异常检测。第一阶段基于用户登录图,使用 Deep Graph InfoMax(DGI)<sup>[7]</sup>进行图上的无监督学习,通过 Local Outlier Factor(LOF)<sup>[8]</sup>算法计算得到部分异常样本;第二阶段基于源主机路径图和第一阶段得到的少量异常样本,使用 Heterogeneous Graph Attention Network(HAN)<sup>[9]</sup>进行图上的半监督学习,得到更多的异常样本,这些检测到的异常样本即判定为横向移动攻击行为。该方法可以在没有样本标签的情况下有效检测横向移动行为,在相关数据集上的 AUC 超过 95%,具有高精确率和低误报率。

本文的结构组织如下。第二章对相关工作进行了阐述。第三章详细说明了本文提出的方法。第四章分析数据集和实验结果。第五章对本文的工作进行总结。

## 2 相关工作

这一部分将简要阐述横向移动攻击检测和基于图的异常检测的研究现状。

### 2.1 横向移动攻击检测

目前,横向移动攻击检测仍处于较为初步的阶段,对横向移动的研究主要是将其转换为内网中异常用户或主机的检测,通过对用户或主机的行为进行建模,检测超出阈值的异常表现。根据检测目标的不同,可以将其分为移动目标型和移动路径型。移动目标型方法主要检测横向移动攻击中攻击者攻陷的用户或主机。Bohara 等人<sup>[10]</sup>提出了一种无监督学习的方法检测横向移动攻击的主机,通过构建主机通信图表示内网中主机之间的通信行为,基于图上提取的特征进行两阶段检测。第一阶段使用主成分分析(Principal Component Analysis, PCA)<sup>[11]</sup>和 k-means<sup>[12]</sup>算法来检测 C&C 受陷主机,第二阶段对 PCA 转换后的数据进行极值分析,以识别受恶意横向移动攻陷的主机,在相关数据集上该方法能够以 90%的召回率检测出异常的横向移动,但同时误报率高达 14%。Powell 等人<sup>[13]</sup>提出了一种基于行为图异常的方法检测受陷主机,基于用户的历史登录活动构建每日登录图,使用非负矩阵分解的方法将图中每个主机分配一个角色,利用主机与角色之间的重构误差检测异常的横向移动行为,通过相关实验验证,该方法的召回率高达 90%,同时误报率低于 5%,但对于登录行为稀疏的用户的检测效果不佳。移动路径型方法则以横向移动攻击中发生的移动路径为检测目标。Siadati 等人<sup>[14]</sup>提出了一种登录结构异常检测方法,通过收集用户正常的登录模式,使用 Market-Basket 变种算法从主机登录结构中提出登录模式,然后使用异常检测的方法检测横向移动,当发生与登录模式不一致的登录行为时,即判定为攻击者通过窃取登录凭证进行的横向移动行为。该方法检测效果良好,但对其他形式的横向移动攻击检测帮助有限。Chen 等人<sup>[15]</sup>提出了一种基于图嵌入的半监督学习方法检测横向移动攻击路径,首先基于原始日志构建带有特征的主机间通信图,其次对图中的主机节点通过预定义的聚合函数聚合其邻接特征,最后使用降噪自编码器<sup>[16]</sup>检测恶意的登录路径,该方法的精确率

可以达到 90% 以上, 但仅考虑了主机行为特征, 而忽略了用户行为模式。综上, 现有的横向移动攻击检测尚处于起步阶段, 在检测方法及性能上仍有较大的提升空间。而内部网络本质上是一张由用户和主机组成的关联异质图, 在异质图上进行横向移动攻击检测还有待研究。

## 2.2 基于图的异常检测

近年来, 基于图的异常检测受到了越来越多的关注, 广泛应用于入侵检测、欺诈识别、安全风控等领域。异常检测是要发现与大部分样本不同的目标样本<sup>[17]</sup>, 以往的研究多聚焦于规则的欧式空间数据, 而现实世界往往是结构化的非欧式空间数据, 常见的如社交网络、交通网络等。因此许多研究工作逐渐集中于基于图的异常检测。根据检测目标图的不同可将其分为静态图异常检测和动态图异常检测。静态图是一种无变化的图, 或者是变化网络图某时刻的一个快照, 基于静态图的异常检测就是根据图的结构和信息, 查找异常的节点或边。如 Akoglu 等人<sup>[18]</sup>提出了基于 *egonet* 检测异常节点的方法, 通过对每个节点提取 *egonet* 并计算相应异常结构的偏离得分, 发现异常节点, 该方法适用于加权图的异常检测, 但没有明确度量规则; Hou 等人<sup>[19]</sup>提出了一种基于异质信息网络识别恶意安卓软件的方法, 通过定义 APP 和 API 的实体及关系, 构建异质信息网络, 并根据多条元路径抽取特征, 使用多核 SVM<sup>[20]</sup>进行模型训练, 检测恶意应用。静态图没有时间属性, 但实际中随着时间的变化, 网络的结构或者属性往往会发生改变。因此动态图的异常检测主要研究随着时间的推移图网络偏离正常演变行为的情况。Ji 等人<sup>[21]</sup>提出了一种针对动态加权图的局部异常检测方法, 首先对样本构建不同时刻的邻域子图, 再通过时间维度上的分析评估, 发现异常样本。该方法可以有效检测动态图中的边异常, 但无法实时检测图中的异常。Zheng 等人<sup>[22]</sup>提出了一种动态预测欺诈用户的方法。该方法首先通过 LSTM-Autoencoder<sup>[23]</sup>将正常用户的行为图映射到低维嵌入空间, 然后采用 GAN<sup>[24]</sup>进行模型训练, 其中生成器用于生成恶意用户样本, 判别器用于区分正常用户和恶意用户, 最后得到一个用于欺诈检测的判别模型。综上, 由于图结构更加贴合现实世界数据, 因此基于图的异常检测的相关工作已经有了一定的积累, 且广泛应用在多个领域。考虑到横向移动攻击作为一个异常检测任务, 涉及内网中用户与主机之间的关联图, 本文从基于图的异常检测的角度出发检测横向移动攻击。

## 3 基于异质图网络的横向移动攻击检测方法

在这一部分, 本文将首先描述发现的横向移动攻击模式, 然后对所提出的基于异质图网络的两阶段横向移动攻击检测方法的结构和执行步骤进行详细阐述。

### 3.1 数据分析

通过对相关横向移动行为分析<sup>[25-27]</sup>, 我们发现横向移动攻击往往表现出行为异常性和主机聚集性。

**行为异常性。**攻击者的横向移动行为往往异于用户历史状态的登录模式。当攻击者通过窃取凭证伪装成正常用户进行操作时, 为了入侵和控制更核心的资产系统, 通常会移动到其历史行为中没有访问或很少访问的主机, 获得更多的控制权, 呈现出不同于用户历史行为模式的异常性。

**主机聚集性。**攻击者的横向移动行为往往集中聚集在少数主机上, 即一台主机上发生多个横向移动攻击。当攻击者成功攻陷一台主机后, 为了资源利用最大化, 他们通常会利用该主机做尽可能多的横向移动攻击行为, 呈现出多个横向移动攻击集中在一个源主机的聚集性。

通过对 CMCS Events 数据集<sup>[28]</sup>进行分析, 我们在图 1 和图 2 分别说明了这两种特性。图 1 展示了用户在一段时间内登录目标主机的行为模式, 横坐标为时间, 纵坐标是目标主机。图中蓝色的点表示正常的登录行为, 橙色的点表示异常的横向移动行为。可以看出, 相较于稀疏的横向移动行为, 用户正常的登录模式均匀地散布在图上, 表明了横向移动攻击的行为异常性。图 2 展示了某一用户在主机之间的移动行为图, 节点表示主机, 边表示用户从源主机登录到目标主机的行为。图中蓝色的点表示正常

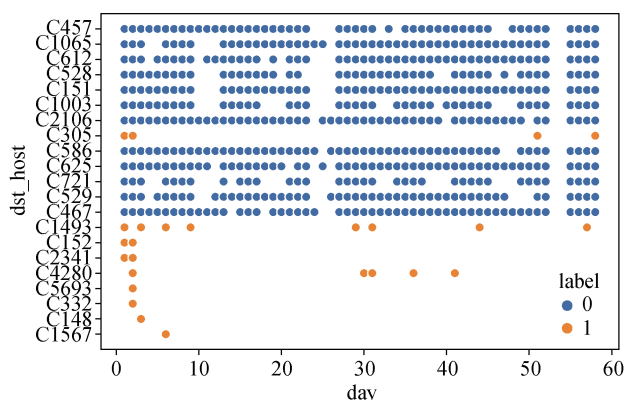


图 1 某一用户在不同主机上的登录行为

Figure 1 The login behavior of a user on different hosts

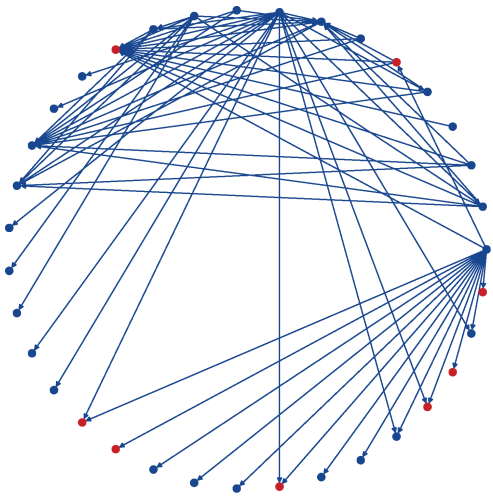


图 2 某一用户在主机间的移动行为

Figure 2 Movement behavior of a user between hosts

登录的目标主机, 红色的点表示横向移动的目标主机。从图中可以发现, 横向移动攻击的源主机往往聚集在相同的主机上, 表明了横向移动的主机聚集性。

基于上述发现, 本文提出了一种基于异质图网络的两阶段横向移动攻击检测方法 HGLM。首先, 基于内网的认证日志, 将用户与主机的登录行为图结构化, 构建用户登录图和源主机路径图, 分别映射横向移动的行为异常性和主机聚集性。其次, 在构建的图上进行两阶段异常检测。第一阶段, 在用户登录图上使用无监督学习方法学习用户在不同主机上的行为特征, 通过异常检测方法得到部分行为异常的主机样本; 第二阶段基于源主机路径图和第一阶段得到的少量异常样本, 使用图上的半监督学习方法, 检测更多的横向移动攻击行为。本方法的架构流程图如图 3 所示。

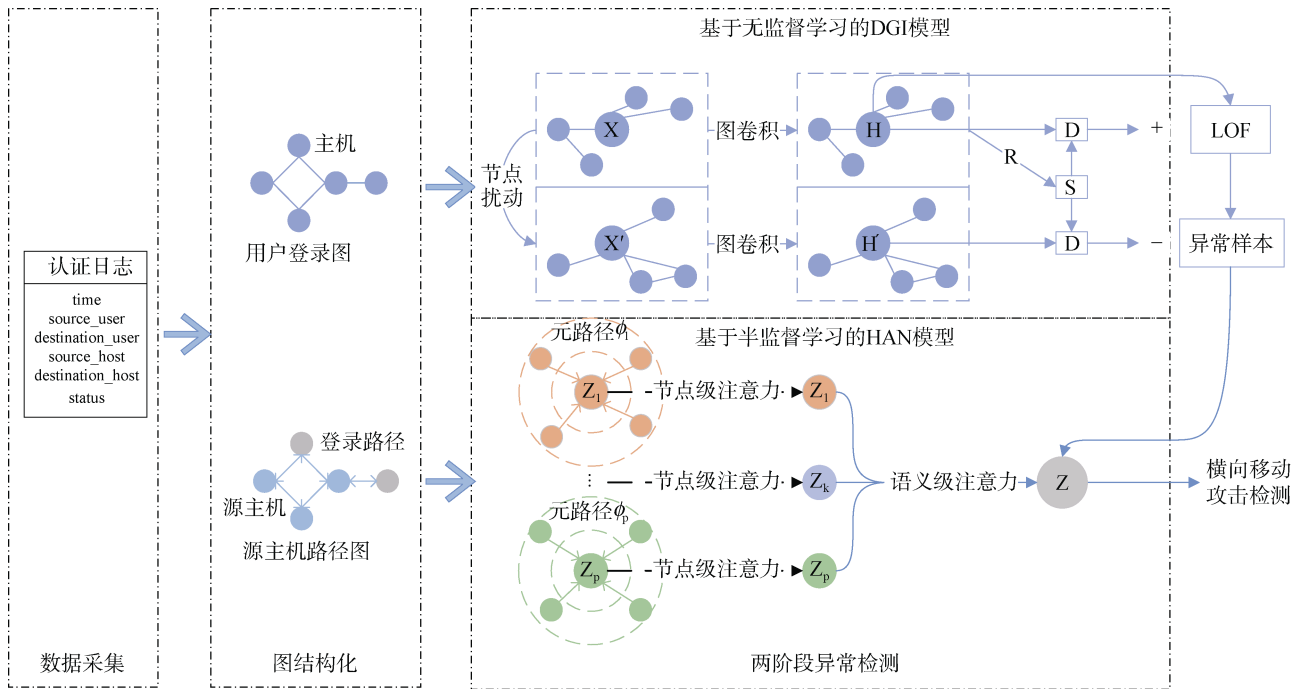


图 3 HGLM 的架构流程图

Figure 3 An overview of the HGLM architecture

### 3.2 日志图结构化

这部分将详细阐述基于内网认证日志构建用户登录图和源主机路径图的全部过程。

#### 3.2.1 数据预处理

日志图结构化的第一步是对内网的认证日志进行预处理。认证日志通常包含认证时间、源用户、目标用户、源主机、目标主机和认证状态等属性, 如表 1 所示。原始日志信息冗余驳杂, 因此需要将其处理成符合横向移动攻击场景的格式。首先, 由于攻击者通常利用受陷用户从一台主机横向移动到另一台主机, 因此我们只需关注源用户和目标用户相同的

认证事件。其次, 横向移动攻击至少涉及两台主机, 因此我们需要对源主机和目标主机相同的认证事件

表 1 认证日志属性

Table 1 Authentication log attributes

属性	说明
time	认证事件发生的时间
source_user	源用户
destination_user	目标用户
source_host	源主机
destination_host	目标主机
status	认证成功/失败

进行过滤。预处理流程如算法 1 所示。给定认证日志数据集  $D$ , 遍历其中的每一条认证事件, 将源用户与目标用户相同且源主机与目标主机不同的事件筛选出来, 得到处理后的数据集  $D_I$ 。

---

**算法 1.** 认证日志预处理.

---

```

输入: 数据集  $D=\{r_1, \dots, r_n\}$ 
输出: 处理后的数据集  $D_I$ 
SET  $D_I=\{\}$ 
FOR EACH  $r_i \in D$  DO
    IF  $r_i.source\_user = r_i.destination\_user$  AND
 $r_i.source\_host \neq r_i.destination\_host$  THEN
         $D_I = D_I \cup r_i$ 
    END IF
END FOR

```

---

### 3.2.2 用户登录图

用户登录图(user authentication graph, UAG)是一张无向同质图, 表示用户一定时间内在主机间的登录行为模式。定义图  $G=(V,E,F)$ , 图中节点  $V$  表示主机, 边  $E$  表示用户在主机间的登录连接。通过将滑动窗口下用户在主机上的登录次数作为特征  $F$  赋予到节点, 连接边上无特征, 得到一个带有特征的用户登录图网络。具体地, 给定数据集  $D_I$ 、用户  $u$  和滑动窗口长度  $L$ , 首先在  $D_I$  中筛选出属于用户  $u$  的认证事件, 得到数据集  $D_u$ 。其次, 根据滑动窗口长度  $L$  将数据分为多个时间窗口, 用于计算不同窗口下用户在主机上的登录次数特征  $F$ 。最后, 遍历  $D_u$  中的每一条认证事件, 将源主机和目标主机添加到图中的节点  $V$ , 并添加一条源主机与目标主机的连接到图中的边  $E$ (节点和边若添加重复则忽略), 同时将  $F$  中源主机与目标主机在对应窗口下的登录次数加一, 遍历结束即得到用户  $u$  的带有特征的用户登录图  $G_u=(V,E,F)$ 。用户登录图的构建流程如算法 2 所示。

---

**算法 2.** 用户登录图的构建.

---

```

输入: 数据集  $D_I$ , 用户  $u$ , 滑动窗口长度  $L$ 
输出: 用户登录图  $G_u=(V,E,F)$ 
SET  $D_u=\{\}, V=E=F=\phi$ 
FOR EACH  $r_i \in D_I$  DO
    IF  $r_i.source\_user = u$  THEN
         $D_u = D_u \cup r_i$ 
    END IF
END FOR
FOR EACH  $r_i \in D_u$  DO

```

---

```

         $V.add(r_i.source\_host)$ 
         $V.add(r_i.destination\_host)$ 
         $E.add(<r_i.source\_host, r_i.destination\_host>)$ 
         $F[r_i.source\_host, r_i.time/L] += 1$ 
         $F[r_i.destination\_host, r_i.time/L] += 1$ 
    END FOR
     $F = normalization(F)$ 
    RETURN  $G_u=(V,E,F)$ 

```

---

### 3.2.3 源主机路径图

源主机路径图(host path graph, HPG)是一个有向异质图, 表示用户到目标主机的登录路径与源主机之间的发生关系。定义图  $G=(V,E,F)$ , 图中定义有两种类型的节点, 一类表示源主机  $V_{src}$ , 一类表示用户到目标主机的登录路径  $V_{path}$ , 边也存在两种类型, 一类为发送边  $E_{send}$ , 从源主机节点指向用户到目标主机的登录路径节点, 表示用户从源主机登录到目的主机; 另一类为依托边  $E_{on}$ , 从用户到目标主机的登录路径节点指向源主机节点, 表示用户到目标主机的登录路径发生在源主机上, 这两种类型的边是对称的。通过将滑动窗口下登录路径在源主机上的发生次数和统计特征  $F_{statistic}^{[29]}$  赋予到节点上, 边仅表示连接关系, 得到一个源主机路径图网络。具体地, 给定数据集  $D_I$ 、滑动窗口长度  $L$  和统计特征  $F_{statistic}$ , 遍历  $D_I$  中每一条事件, 将源主机添加  $V_{src}$ , 将用户和目标主机拼接成登录路径  $path$  作为节点添加到  $V_{path}$ , 并将由源主机指向登录路径的连接边添加到  $E_{send}$ , 对称地将由登录路径指向源主机的连接边添加到  $E_{on}$ , 对滑动窗口登录次数特征的计算同用户登录图。最后, 对图中登录路径类型的节点  $V_{path}$  进行遍历, 将统计特征  $F_{statistic}$  追加到节点上, 同时对源主机节点  $V_{src}$  赋予独热编码特征, 得到带有特征的源主机路径图  $G_p=(V,E,F)$ 。本文使用的统计特征如表 2 所示。源主机路径图的构建流程如算法 3 所示。

表 2 统计特征

Table 2 Statistical features

特征	说明
success_count	用户到该目标主机认证的成功次数
fail_count	用户到该目标主机认证的失败次数
path_rate	用户到该目标主机的认证次数占用户总认证次数的比率
time_interval_median	用户到该目标主机认证事件发生的时间间隔平均值
time_interval_max	用户到该目标主机认证事件发生的时间间隔最大值
time_interval_min	用户到该目标主机认证事件发生的时间间隔最小值

---



**算法 3. 源主机路径图的构建.**


---

输入: 数据集  $D_I$ , 滑动窗口长度  $L$ , 统计特征  $F_{statistic}$

输出: 源主机路径图  $G_p=(V,E,F)$

SET  $V=\{V_{src}, V_{path}\}, E=\{E_{send}, E_{on}\}$

FOR EACH  $r_i \in D_I$  DO

$V_{src}.add(r_i.source\_host)$

$path=(r_i.source\_user, r_i.destination\_host)$

$V_{path}.add(path)$

$E_{send}.add(<r_i.source\_host, path>)$

$E_{on}.add(<path, r_i.source\_host>)$

$F[path, r_i.time/L] += 1$

END FOR

FOR EACH  $v \in V_{path}$  DO

$F[v].append(F_{statistic}[v])$

END FOR

$F[V_{src}] = \text{onehotencoder}(V_{src})$

$F = \text{normalization}(F)$

RETURN  $G_p=(V,E,F)$

---

**3.3 基于图的两阶段异常检测**

这部分将详细阐述利用基于图的两阶段异常检测方法检测横向移动攻击的整体过程。第一阶段, 基于UAG利用DGI学习用户在不同主机上登录模式的隐层表示, 通过异常检测得到部分异常样本; 第二阶段, 基于HPG和第一阶段的少量样本利用HAN学习在源主机上不同登录路径的行为模式, 检测更多的横向移动攻击行为。

**3.3.1 基于无监督学习的异常登录行为检测**

用户登录图本质上是一个无向同质图网络, 节点之间通过主机间的登录行为连接。基于横向移动攻击的行为异常性发现, 异常样本往往不同于用户的历史行为模式, 因此第一阶段使用DGI算法进行图上的无监督学习, 通过学习用户在不同主机上的登录模式, 得到UAG节点的隐层表示。

DGI将Deep InfoMax(DIM)<sup>[30]</sup>引入图网络领域, 是一种图上的无监督学习方法。DIM认为传统的基于最小化重构误差的方式学得样本特征并不是最好的, 为了学习样本最独特的特征, 它使用互信息<sup>[31]</sup>作为损失函数:

$$I(X;Y) = \int_Y \int_X p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx dy \quad (1)$$

其中,  $X$  为输入,  $Y$  为输出, 通过最大化局部特征和全局特征的互信息训练得到样本的隐层特征表示。具体到图网络领域, 对于一个图网络, 每一个节点的特征向量就是该节点的局部特征, 通过图卷积核编

码器来学习节点的隐层向量, 而全局特征通过 *readout* 函数获得。通过对节点加以扰动得到负样本, 使用一个判别器对“样本对”进行打分, 最后得到节点的隐层表示。

在本文中, 对于UAG, 我们使用Graph Convolutional Network(GCN)<sup>[32]</sup>作为图卷积核编码器, 对于每个节点的特征向量  $\vec{h}_i$ , 通过GCN将邻接节点的信息整合起来:

$$H^{(t+1)} = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(t)} \theta) \quad (2)$$

其中,  $H^{(t)}$  表示进行  $t$  层卷积后节点的特征,  $\hat{A}$  是带有自循环的邻接矩阵,  $\hat{D}$  是  $\hat{A}$  的对角度矩阵,  $\theta$  是参数,  $\sigma$  是激活函数。而全局特征  $\vec{s}$  通过使用平均 *readout* 函数  $R$  得到:

$$R(H) = \frac{1}{N} \sum_{i=1}^N \vec{h}_i \quad (3)$$

其中,  $N$  表示图中节点数量。令  $(\vec{h}_i, \vec{s})$  为正样本对, 负样本  $(\vec{h}_i, \vec{s}')$  则通过保持图的邻接矩阵不变而对特征矩阵进行节点扰动(随机乱序)得到, 最后通过一个线性二元分类器对“样本对”打分, 训练得到节点的隐层特征表示。基于DGI学得样本特征, 我们使用LOF算法检测得到部分异常点, 用于第二阶段的半监督学习。

**3.3.2 基于半监督学习的横向移动攻击检测**

源主机路径图本质上是一个有向异质图网络, 图中有两种类型的节点: 源主机节点  $V_{src}$  和路径节点  $V_{path}$ , 以及两种类型的边: 发送关系  $E_{send}$  和依托关系  $E_{on}$ 。基于横向移动攻击的主机聚集性发现, 发生横向移动攻击的路径往往聚集在相同的源主机上。因此第二阶段使用HAN进行图上的半监督学习, 基于第一阶段的少量样本, 通过学习路径节点之间的关联, 检测更多的横向移动路径。

HAN将注意力机制引入异质图, 是一种图上的半监督学习方法。在同质图中, 图注意力网络(Graph Attention Network, GAT)<sup>[33]</sup>提出了用注意力机制聚合邻接节点特征:

$$e_{ij} = a([W\vec{h}_i \parallel W\vec{h}_j]), j \in N_i \quad (4)$$

$$\alpha_{ij} = \frac{\exp(\text{Leaky ReLU}(e_{ij}))}{\sum_{k \in N_i} \exp(\text{Leaky ReLU}(e_{ik}))} \quad (5)$$

$$\vec{h}_i' = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W \vec{h}_j\right) \quad (6)$$

其中,  $e_{ij}$  表示节点  $i$  和  $j$  的相似系数,  $[\parallel]$  表示向量

拼接操作,  $\alpha_{ij}$  表示 *softmax* 操作,  $\bar{h}_i$  表示节点  $i$  通过注意力机制聚合后的特征向量,  $\sigma$  表示激活函数。节点在特征学习中通过计算对其邻接节点的注意力系数聚合邻接特征。而 HAN 将注意力机制引入到了异质图中, 包括节点级注意力和语义级注意力。通过定义图上的元路径(Meta-path), 节点级注意力主要学习其元路径上邻接节点的权重, 而语义级注意力学习基于不同元路径的权重。最后, 通过相应的聚合操作得到最终的节点表示。

在本文中, 对于 HPG, 我们定义两条元路径: 从路径节点到源主机节点的元路径 ( $v_{path}, e_{on}, v_{src}$ ) 和从路径节点到源主机节点再到路径节点的元路径 ( $v_{path}, e_{on}, v_{src}, e_{send}, v_{path}$ )。基于两条元路径使用 GAT 计算节点级注意力:

$$e_{ij}^{\Phi} = att_{node}(h_i, h_j; \Phi) \quad (7)$$

$$\alpha_{ij}^{\Phi} = soft \max(e_{ij}^{\Phi}) \quad (8)$$

$$z_i^{\Phi} = \sigma(\sum_{j \in N_i^{\Phi}} \alpha_{ij}^{\Phi} h_j) \quad (9)$$

其中,  $\Phi$  表示元路径,  $e_{ij}^{\Phi}$  表示相似系数,  $\alpha_{ij}^{\Phi}$  表示注意力系数,  $z_i^{\Phi}$  表示节点  $i$  的聚合特征。基于来自两条元路径的聚合特征  $Z_{\Phi_1}$  和  $Z_{\Phi_2}$  计算语义级注意力:

$$\omega_{\Phi_i} = \frac{1}{|M|} \sum_{i \in M} q^T \cdot \tanh(W \cdot z_i^{\Phi} + b) \quad (10)$$

$$\beta_{\Phi_i} = \frac{\exp(\omega_{\Phi_i})}{\sum_{i=1}^P \exp(\omega_{\Phi_i})} \quad (11)$$

$$Z = \sum_{i=1}^P \beta_{\Phi_i} \cdot Z_{\Phi_i} \quad (12)$$

其中,  $\omega_{\Phi_i}$  表示元路径  $\Phi_i$  的重要度系数,  $M$  表示元路径集合,  $q$  为注意力向量,  $\beta_{\Phi_i}$  表示元路径  $\Phi_i$  的权重,  $Z$  表示聚合后的节点特征。基于第一阶段的少量样本, 对图中对应的路径节点进行标注, 使用交叉熵损失函数进行半监督训练, 检测异常的路径节点, 发现横向移动攻击行为。

## 4 实验和结果

### 4.1 数据集

由于对横向移动攻击检测的研究尚处于初级阶段, 相关数据集极为匮乏, 许多研究工作使用的数据集由于信息敏感而不对外公开。因此, 为了有效验证 HGLM 的性能, 在数据集方面本文选择了大部分研究工作所使用的公开数据集, 即洛斯阿拉莫斯国

家实验室(Los Alamos National Laboratory, LANL)于 2015 年公开的多源复杂网络安全事件数据集 (Comprehensive, Multi-Source Cyber-Security Events, CMCS Events)<sup>[34]</sup>。数据集包含了洛斯阿拉莫斯实验室内部网络中连续 58 天的多源日志事件数据。我们仅使用其中的认证日志和红队攻击数据。认证数据是收集自 Windows 主机、服务器以及 Active Directory 域服务器的身份验证事件, 每条数据表示一个在给定时间下的认证事件, 其包含的属性如表 3 所示。而红队攻击数据来自于经过授权的专业人员进行的恶意认证活动, 其中的每条数据表示从认证日志中获取的红队攻击认证事件。为了有效验证本文提出的方法, 我们选择了前 9 天的数据作为实验数据集, 涵盖实验室内网中 12425 个用户, 14067 个主机以及 152468274 条认证事件, 其中包含有红队攻击的 323 条恶意事件。

表 3 认证事件包含的属性

Table 3 Authentication event attributes

属性	说明
time	时间
source user@domain	指定域的源用户
destination user@domain	指定域的目标用户
source computer	源主机
destination computer	目标主机
authentication type	认证类型
logon type	登录类型
authentication orientation	认证方向
success/failure	验证成功或失败

### 4.2 数据集预处理

由于原始认证事件的冗余驳杂, 我们首先对数据集进行预处理以符合横向移动攻击场景的格式。首先通过第 3 章的算法 1 将源用户与目标用户相同且源主机与目标主机不同的事件筛选出来; 其次删除不需要的认证类型、登录类型和认证方向属性; 最后将与红队攻击入侵的用户不相符的认证事件过滤掉, 得到预处理后的认证数据。该数据集包含 51 个内网用户, 1835 个主机以及 344557 条认证事件, 其中, 横向移动攻击事件占比仅 0.09%, 同时包含跨越时间段的横向移动攻击行为, 符合横向移动攻击高隐蔽性和低频率性的场景。通过对预处理后的数据集进行图结构化, 基于构建的用户登录图和源主机路径图进行图上的两阶段异常检测。

### 4.3 评估指标

由于横向移动攻击检测是一个二分类的异常检

测问题, 因此我们选择召回率  $TPR$  与误报率  $FPR$  作为模型检测性能的评估指标。其计算公式如下:

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

$$FPR = \frac{FP}{FP + TN} \tag{14}$$

其中,  $TP$ 、 $FP$ 、 $TN$ 、 $FN$  代表了样本真实值与模型预测值之间的关系, 具体如表 4 所示。此外, 为了消除正负样本不平衡的影响, 我们使用  $AUC$  评估模型整体的表现效果, 使用精确率  $precision$  与准确率  $accuracy$  评估模型在不同条件下的表现, 其计算如下:

$$precision = \frac{TP}{TP + FP} \tag{15}$$

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{16}$$

表 4 评估指标的含义

Table 4 The meaning of evaluation indicators			
样本真实值	模型预测值		
		正	负
	正	TP	FN
	负	FP	TN

实验中, 我们选择 Bohara 等人<sup>[10]</sup>提出的无监督检测模型作为对照的基准模型。该模型通过构建主机间通信图抽取特征, 利用 PCA 等无监督算法检测横向移动攻击, 最终效果  $TPR$  达到 92.39%,  $FPR$  为 14.06%。同时, 我们也对比了在相同数据特征上不同算法的表现性能。

4.4 实验结果

为了验证本文所提方法的有效性, 我们在数据集上进行了大量的实验。具体实验如下: 首先, 基于预处理后的数据集构建 UAG 和 HPG。通过将认证事件按照用户进行分组, 我们对每一个用户构建 UAG, 之后基于所有认证事件构建 HPG。最后得到 51 个 UAG 和 1 个 HPG, 其中部分 UAG 和 HPG 分别如表 5 和表 6 所示。其次, 构建并训练 DGI 和 HAN 模型, 进行两阶段异常检测。第一阶段, 构建 DGI 图网络模型进行无监督训练, 通过 LOF 对学得的节点的隐层特征进行异常检测得到部分异常主机样本, 将用户与主机拼接起来作为第二阶段的有标签恶意样本。其中, DGI 使用的超参数范围如下: 隐层特征维度  $n\_hidden=[4,8,16,32]$ , 图卷积层数  $n\_layers=[2,3,4]$ , 学习率  $learning\_rate=[1e-3,1e-4]$  以及正则化  $dropout=[1e-1,2e-1]$ 。第二阶段, 构建 HAN 图网络模型进行半监督训练, 对路径节点进行二分类检测横

向移动攻击样本。其中, HAN 使用的超参数范围如下: 隐层特征维度  $n\_hidden=[4,8,16,32]$ , 学习率  $learning\_rate=[1e-3,5e-3]$  以及正则化  $dropout=[1e-1,2e-1]$ 。通过参数调优, 模型的  $AUC$  值可以达到 95.53%, 且召回率  $TPR$  为 96.88%, 误报率  $FPR$  为 5.81%。与对照的基准模型对比, HGLM 在  $TPR$  和  $FPR$  上都有显著提升, 尤其是误报率的大幅降低。而与已有方法对比, 我们的方法不需要样本标签, 并可以超过大部分有监督的检测方法, 整体效果对比如表 8 所示, 其中我们的模型使用的最优超参数如表 7 所示。此外, 我们也对比了模型在不同用户上的表现, 如图 4 所示, 可以发现对于大部分用户, 模型的召回率可以超过 95%, 误报率低于 5%。

表 5 部分用户登录图对比

Table 5 Comparison of some user Authentication graphs

用户	节点数	边数
U1653	1,004	1029
U1723	182	405
U66	167	285
U748	91	346
U737	83	209
U293	74	351

表 6 源主机路径图

Table 6 The host path graph

源主机节点数	路径节点数	发送边数	依托边数
1310	1651	5537	5537

表 7 模型最优超参数

Table 7 Model optimal hyperparameters

超参数	DGI	HAN
$n\_hidden$	16	32
$n\_layers$	2	×
$learning\_rate$	$1e-3$	$5e-3$
$dropout$	$2e-1$	$2e-1$

表 8 横向移动攻击检测模型的性能比较

Table 8 Comparison of performance of lateral movement attack detection models

模型	是否需要样本标签	TPR	FPR	AUC
SVM	supervised	71.43%	3.32%	84.07%
GBDT	supervised	85.71%	4.01%	90.85%
GCN	semi-supervised	82.14%	4.12%	89.01%
DGI	unsupervised	75%	3.53%	85.73%
HGLM	unsupervised	96.88%	5.81%	95.53%



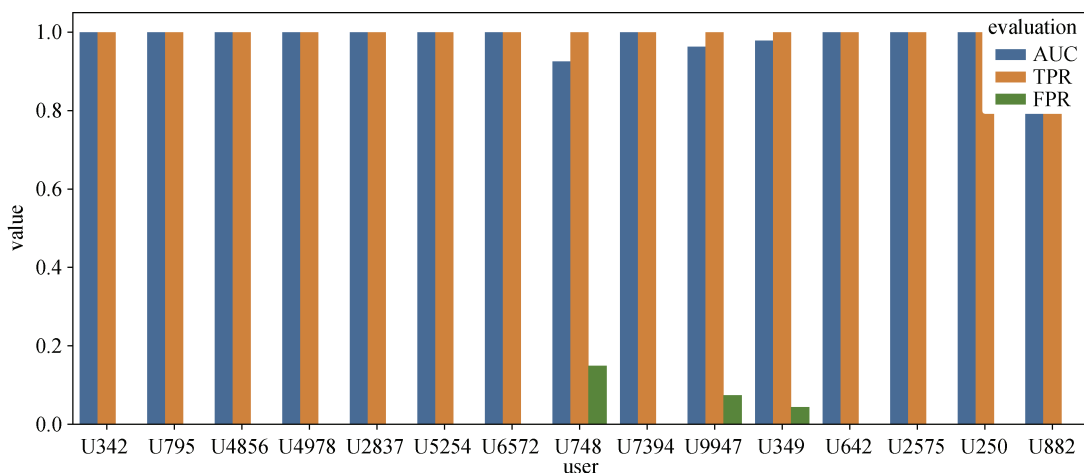


图4 HGLM模型在不同用户上的结果对比

Figure 4 Comparison of HGLM model results on different users

为了评估所提方法在不同条件下的检测性能, 我们通过调节用于第一阶段检测的用户所占整体用户的比例和用于第二阶段检测的有标签样本占整体样本的比例, 评估这两个超参数对于检测性能的影响。

**用户比率。**用户比率控制着第一阶段用于构建图并作训练的用户占整体用户的比例。保持其他设置不变, 通过改变用户比率, 我们评估模型在仅有部分数据的条件下的检测性能。如图5所示, 可以看出, 当用户比率达到30%以上, 模型的精确率就可以超过90%, 而随着用户比率的增大, 数据增加, 模型的检测性能越来越好。

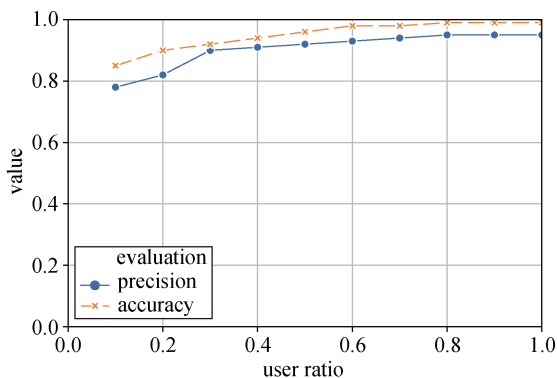


图5 用户比率对HGLM模型效果的影响

Figure 5 The effect of user ratio on HGLM model performance

**标签比率。**标签比率控制着用于第二阶段检测的有标签样本占整体样本的比例。保持其他设置不变, 通过改变用于训练的有标签正常样本的数量, 结合第一阶段的有标签异常样本, 我们评估模型在仅有部分标签的条件下的检测性能。如图6所示, 可以发现, 随着有标签样本的比率增大, 模型的精确

率逐步增加, 当比率为0.3的时候, 模型效果达到最佳, 而当标签比率继续增大, 有标签正常样本越来越多, 增大了误报的可能性, 模型精确率有所下降。

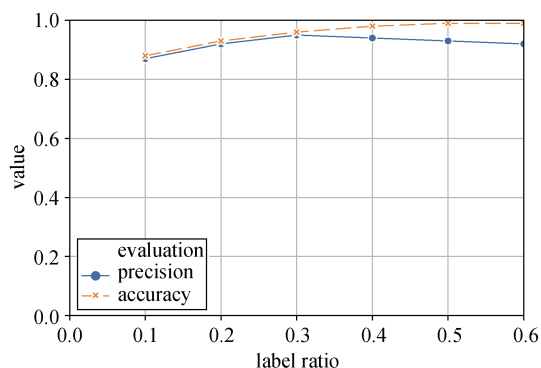


图6 标签比率对HGLM模型效果的影响

Figure 6 The effect of label ratio on HGLM model performance

## 5 结论

为了有效地检测横向移动攻击, 本文提出一种基于异质图网络的两阶段横向移动攻击检测方法HGLM。通过对横向移动行为进行分析, 我们发现了其具有行为异常性和主机聚集性特征, 因此我们使用了图神经网络模型学习横向移动的行为模式, 在图上进行异常检测, 发现横向移动攻击。具体地, 我们使用内网的认证日志, 将用户与主机的认证事件图结构化, 构建用户登录图和源主机路径图, 然后在图上进行两阶段异常检测。第一阶段基于用户登录图, 使用DGI模型进行无监督训练, 学习主机的行为特征表示, 通过异常检测方法得到部分异常样本; 第二阶段基于源主机路径图和第一阶段得到的少量异常样本, 使用HAN模型进行图上的半监督学

习,发现横向移动攻击。本文提出的方法可以在没有样本标签的情况下有效检测横向移动行为,在CMCS Events数据集上的AUC值超过95%,部分用户的TPR达到100%,FPR为0,效果超过了大部分有监督学习的方法。

本方法在抽取特征的时候仅考虑了认证事件,维度相对比较简单,后续将考虑流量、进程等多个维度,进一步提高横向移动的检测率。同时,为了更全面地验证本方法的效果,后续也将考虑在多个数据集上进行实验以查看效果。

**致 谢** 感谢中国科学院网络测评技术重点实验室的各位老师和同学提出的有益建议。感谢审稿专家和编辑部老师对本文提出的有益建议及指导。

## 参考文献

- [1] Binde B, McRee R, O'Connor T J. Assessing outbound traffic to uncover advanced persistent threat[J]. *SANS Institute. Whitepaper*, 2011, 16.
- [2] Li Z, Wang Y, Wen S, et al. Evil Chaincode: APT Attacks Based on Smart Contract[C]. *International Conference on Frontiers in Cyber Security*, 2020: 178-196.
- [3] Case D U. Analysis of the cyber attack on the Ukrainian power grid[J]. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 2016, 388.
- [4] Oosthoek K, Doerr C. SoK: ATT&CK Techniques and Trends in Windows Malware[C]. *International Conference on Security and Privacy in Communication Systems*. Springer, Cham, 2019: 406-425.
- [5] Shuya M. Russian Cyber Aggression and the New Cold War[J]. *Journal of Strategic Security*, 2018, 11(1): 1-18.
- [6] Singh S, Sharma P K, Moon S Y, et al. A Comprehensive Study on APT Attacks and Countermeasures for Future Networks and Communications: Challenges and Solutions[J]. *The Journal of Supercomputing*, 2019, 75(8): 4543-4574.
- [7] Veličković P, Fedus W, Hamilton W L, et al. Deep Graph Infomax[EB/OL]. 2018: arXiv: 1809.10341. <https://arxiv.org/abs/1809.10341>
- [8] Breunig M M, Kriegel H P, Ng R T, et al. LOF: Identifying Density-Based Local Outliers[C]. *The 2000 ACM SIGMOD international conference on Management of data*, 2000: 93-104.
- [9] Wang X, Ji H Y, Shi C, et al. Heterogeneous Graph Attention Network[C]. *WWW'19: The World Wide Web Conference*, 2019: 2022-2032.
- [10] Bohara A, Noureddine M A, Fawaz A, et al. An Unsupervised Multi-Detector Approach for Identifying Malicious Lateral Movement[C]. *2017 IEEE 36th Symposium on Reliable Distributed Systems*, 2017: 224-233.
- [11] Wold S, Esbensen K, Geladi P. Principal Component Analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1/2/3): 37-52.
- [12] Pham D T, Dimov S S, Nguyen C D. Selection of  $K$  in  $K$ -Means Clustering[J]. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2005, 219(1): 103-119.
- [13] Powell B A. Detecting Malicious Logins as Graph Anomalies[J]. *Journal of Information Security and Applications*, 2020, 54: 102557.
- [14] Siadati H, Memon N. Detecting Structurally Anomalous Logins within Enterprise Networks[C]. *The 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017: 1273-1284.
- [15] Chen M Y, Yao Y P, Liu J R, et al. A Novel Approach for Identifying Lateral Movement Attacks Based on Network Embedding[C]. *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications*, 2019: 708-715.
- [16] Vincent P, Larochelle H, Lajoie I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. *Journal of Machine Learning Research*, 2010, 11: 3371-3408.
- [17] Chandola V, Banerjee A, Kumar V. Anomaly Detection[J]. *ACM Computing Surveys*, 2009, 41(3): 1-58.
- [18] Akoglu L, McGlohon M, Faloutsos C. Oddball: Spotting anomalies in weighted graphs[C]. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, Berlin, Heidelberg, 2010: 410-421.
- [19] Hou S F, Ye Y F, Song Y Q, et al. HinDroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network[C]. *The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017: 1507-1515.
- [20] Alam S, Kang M, Pyun J Y, et al. Performance of Classification Based on PCA, Linear SVM, and Multi-Kernel SVM[C]. *2016 Eighth International Conference on Ubiquitous and Future Networks*, 2016: 987-989.
- [21] Ji T, Yang D, Gao J. Incremental local evolutionary outlier detection for dynamic social networks[C]. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Berlin, Heidelberg, 2013: 1-15.
- [22] Zheng P P, Yuan S H, Wu X T, et al. One-Class Adversarial Nets for Fraud Detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 1286-1293.
- [23] Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised Learning of Video Representations Using LSTMS[C]. *The 32nd International Conference on International Conference on Machine Learning - Volume 37*, 2015: 843-852.
- [24] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [25] Amrouche F, Lagraa S, Kaiafas G, et al. Graph-Based Malicious Login Events Investigation[C]. *2019 IFIP/IEEE Symposium on Integrated Network and Service Management*, 2019: 63-66.
- [26] Hagberg A, Lemons N, Kent A, et al. Connected Components and

- Credential Hopping in Authentication Graphs[C]. *2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems*, 2015: 416-423.
- [27] Fawaz A, Bohara A, Cheh C, et al. Lateral Movement Detection Using Distributed Data Fusion[C]. *2016 IEEE 35th Symposium on Reliable Distributed Systems*, 2016: 21-30.
- [28] Kent A D. Cyber security data sources for dynamic network research[M]. *Dynamic Networks and Cyber-Security*. 2016: 37-65.
- [29] Kaiafas G, Varisteas G, Lagraa S, et al. Detecting Malicious Authentication Events Trustfully[C]. *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018: 1-6.
- [30] Hjelm R D, Fedorov A, Lavoie-Marchildon S, et al. Learning Deep Representations by Mutual Information Estimation and Maximization[EB/OL]. 2018: arXiv: 1808.06670. <https://arxiv.org/abs/1808.06670>
- [31] Viola P, Alignment by Maximization of Mutual Information[J]. *International Journal of Computer Vision*, 1997, 24(2): 137-154.
- [32] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[EB/OL]. 2016: arXiv: 1609.02907. <https://arxiv.org/abs/1609.02907>
- [33] Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[EB/OL]. 2017: arXiv: 1710.10903. <https://arxiv.org/abs/1710.10903>
- [34] Kent A D. Comprehensive, multi-source cyber-security events data set[R]. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2015.



**王天** 于 2018 年在吉林大学软件工程专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读硕士学位。研究领域为网络安全态势感知、网络攻击检测等。Email: wangtian@iie.ac.cn



**董聪** 于 2017 年在天津大学信息管理与信息系统(保密方向)专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读博士学位。研究领域为网络安全态势感知、网络攻击检测等。Email: dongcong@iie.ac.cn



**刘松** 于 2018 年在中国科学院大学计算机技术专业获得硕士学位。现任中国科学院信息工程研究所工程师。研究领域为网络安全态势感知、数据存储等。Email: liusong1106@iie.ac.cn



**田甜** 于 2017 年在中国科学院大学计算机技术专业获得硕士学位。现任中国科学院信息工程研究所工程师。研究领域为网络安全态势感知、数据可视化与可视分析等。Email: tiantian@iie.ac.cn



**卢志刚** 于 2010 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所正高级工程师, 中国科学院网络空间安全学院副教授。研究领域为网络安全态势感知、网络攻击检测、移动终端安全等。Email: luzhigang@iie.ac.cn



**姜波** 于 2016 年在中国科学院大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为网络安全态势感知、知识图谱、数据挖掘等。Email: jiangbo@iie.ac.cn