入侵告警信息聚合与关联技术综述

李祉岐[1] 黄金垒[2] 王义功[2] 胡 浩[2*] 刘玉岭[3,4]

[1](北京国网思极网安科技有限公司 北京 100071)

[2](信息工程大学 河南 郑州 450001)

[3](中国科学院软件研究所可信计算与信息保障实验室 北京 100190)

[4](中国科学院大学网络空间安全学院 北京 100190)

摘 要 告警聚合与关联是入侵检测研究的一个关键问题,可以有效解决 IDS 在实际应用中存在大量重复告警和高误报率的不足。介绍告警聚合和关联的重要性,对现有告警聚合和关联技术进行深入分析比较;总结归纳现有告警聚合与关联的体系结构与应用准则;对当前研究面临的重要技术难题与发展趋势进行展望。

关键词 网络安全 入侵检测 告警聚合 告警关联

中图分类号 TP393.8 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.04.046

INTRUSION ALERT INFORMATION AGGREGATION AND CORRELATION TECHNOLOGY: A SURVEY

Li Zhiqi[1] Huang Jinlei[2] Wang Yigong[2] Hu Hao[2*] Liu Yuling[3,4]

[1](*Beijing State Grid Siji Network Security Technology Limited Company, Beijing 100071, China*)

[2](*Information Engineering University, Zhengzhou 450001, Henan, China*)

[3](*Trusted Computing and Information Assurance Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*)

[4](*School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100190, China*)

Abstract Alert aggregation and correlation is a key issue in intrusion detection research, which can effectively solve the shortcomings of IDS in practical applications, such as a large number of repeated alerts and high false alert rate. We introduced the importance of alert aggregation and correlation, and analyzed and compared the existing alert aggregation and correlation techniques. Then we summarized the existing architecture and application criteria of alert aggregation and correlation, and looked forward to the important technical problems and development trend of current research.

Keywords Network security Intrusion detection Alert aggregation Alert correlation

0 引 言

入侵检测作为网络安全的重要支撑技术已广泛部署应用,但仍存在众多不足[1]:(1) 入侵检测系统 IDS 产生大量重复的告警(又称报警),分析和处理成本较高;(2) 漏告警和误告警严重;(3) 告警间相互孤立且层次较低,反映了单步攻击动作,无法展现攻击渗透过程,难以刻画攻击的真实意图。告警数据处理的质量严重影响 IDS 的实用性,给管理员分析和管理这些告警信息带来了很大的困难,并直接影响 IDS 在实际应用中的有效性。

告警信息处理可以分为:告警聚合(alert aggregation)和告警关联(alert correlation)[2],利用以上两个步

收稿日期:2018-09-26。国家自然科学基金项目(61471344);国家高技术研究发展计划项目(2015AA016006);国家重点研发计划课题(2016YFF0204002,2016YFF0204003);郑州市科技领军人才项目(131PLJRC644);"十三五"装备预研领域基金项目(6140002020115);CCF-启明星辰"鸿雁"科研计划项目(2017003)。**李祉岐**,硕士,主研领域:网络安全,项目管理。**黄金垒**,硕士。**王义功**,硕士。**胡浩**,博士。**刘玉岭**,副研究员。

骤可以实现下列目标[3]:(1) 减少或消除冗余的告警,利用告警聚合降低冗余告警数量;(2) 减小漏报率,通过联合分析不同 IDS 产生的告警,可以降低漏报率;(3) 降低误报率,关联属于同一攻击场景的告警事件,降低随机或独立事件产生的误告警;(4) 重建攻击场景,将属于同一攻击场景的告警事件关联分析,识别入侵意图和手段,把握入侵过程的全貌,增强对攻击行为的理解,为风险评估[4]、入侵响应[5]、防御决策[6]等奠定基础;(5) 增加 IDS 检测范围,在异构复杂网络环境中,单个 IDS 的检测范围和能力受限,通过部署多个 IDS,汇总和关联不同 IDS 生成的告警信息,扩大安全防御和应急响应的范围和效率。

在上述分析的基础上,本文重点关注告警聚合和关联技术新进展,总结归纳其体系架构与应用准则,分析当前面临的技术难题,并指明未来研究的发展方向。

1　告警聚合与关联概述

数据融合[7]最初应用于军事领域(声纳解释系统),对从单个和多个数据源获取的数据在一定准则(如时序等)下进行自动分析、综合等操作,以完成所需的评估和决策任务。下面介绍告警聚合与关联的相关定义及实施过程。

1.1　相关定义

结合国内外研究成果,告警聚合和关联的一般定义如下:

定义 1　原始告警:由 IDS 等安全设备直接检测到,且未经深入分析和处理的告警。

定义 2　超告警:由多个原始告警合并生成的告警。

定义 3　告警聚合:由同一安全事件引起的特征相同或相似的告警合并生成的一个告警。

定义 4　告警关联:将同一攻击过程的多个单步攻击动作所诱发的告警联系在一起,重建攻击线程。

定义 5　攻击线程:多个单步攻击动作按一定次序构成整个攻击过程。

定义 6　攻击场景:由一个或多个攻击线程组成,刻画系统面临的安全威胁和攻击者意图。

1.2　聚合与关联层次

告警信息处理是一个复杂问题,文献[8]将告警信息的聚合和关联过程分为以下四个层次,以降低告警分析处理的复杂度。随着告警信息处理水平的提高,告警质量不断提高,如图 1 所示。

(1) 预处理　首先将异构 IDS 产生的告警信息格式进行标准化处理,并完成攻击类型、时间戳等参数的统一映射,以便深入分析和处理告警信息。

(2) 单个 IDS 告警聚合　将不同 IDS 产生的具有特征相似性的多个告警在时间维度上进行纵向融合操作,通常在独立的 IDS 系统中进行,用于将具有由单个 IDS 发出的相似或相同特征的多个告警组合成一个元告警,主要用于时间轴上的纵向融合操作。

(3) 多个 IDS 告警聚合　将多个同构或异构 IDS 检测到的告警合并,对不同源告警在横向空间维度上相互补充、印证,以发现理解攻击行为的本质。

(4) 告警关联　在告警聚合完成之后,通过关联告警数据,将由相同攻击线程生成的逻辑相关告警链接,以重建入侵场景,更深入剖析攻击者的真实意图与入侵路径,为风险评价和入侵响应提供数据参考。

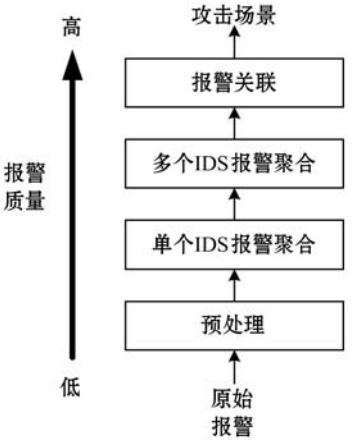

图 1　聚合与关联过程

上述四个层次之间无明显界限,基本目标是提升告警信息的质量,增强 IDS 的准确性和可用性。

2　聚合算法分析

2.1　预处理

告警预处理主要对告警语义及语法进行标准化处理,需要对告警信息的格式进行统一。IDMEF[9]是由入侵检测工作组(IDWG)制定的一种标准草案,采用面向对象的方法描述安全设备产生的告警信息格式。该草案以 XML 语言来描述(RFC4765[10]),刻画了其数学模型,同时阐述了该模型的基本原理及应用方法。其目的是定义安全设备(安全检测或响应系统)的数据交换格式和协议。

依据 IDMEF 标准化后的告警信息属性包括告警类型、时间戳、检测器 ID、端口号、IP、优先级等。由于告警信息处理系统中,传感器的时钟存在差异,通常借鉴网络时间协议(Network Time Protocol)[11]实现传感器时钟的同步。

告警预处理所涉及的算法较少,本文重点对告警聚合与关联算法进行分析。

2.2 告警聚合

告警聚合主要用于对原始告警信息进行处理,通过降低漏报和误报,实现告警精简,并为告警关联提供告警数据支撑。

2.2.1 基于属性相似度的方法

通常由相同攻击类型引起的告警信息特征(或属性)具有相似性,因此可以通过比较告警属性的相似性来聚合告警信息。

Valdes 等[12]采用概率统计的方法,定义属性相似度函数,设计属性特征多元匹配算法,计算告警的整体相似性,以将告警信息进行分组和归并,从而达到减少重复告警的目的。Debar 等[13]提出一种隐式聚合组件,建立了 TACC 模型,摘取新告警的部分属性,并与历史事件进行匹配,以聚合告警。该方法将告警信息投射到三条轴,即源地址、目的地址和类型,然后根据不同的场景类型,如果包含相同属性的告警信息数目达到阈值,则将其合并成一条元告警。

龚俭等[14]提出了一种基于攻击类型、空间与时间特征的冗余告警消减算法,可以消除大量重复告警,但存在信息损失且规则制定较为困难,不利于告警信息的后续处理(攻击线程发现等)。陈志文等[15]设计了一种基于动态聚合时间窗口和最大聚合数量窗口相结合的告警聚合算法,但同样存在告警信息损失的不足。

基于属性相似度的告警聚合方法通过知识定义相似度函数、权重等,能够更好地对已知攻击类型告警进行聚合,且算法计算效率高,具有较好的实时性,但告警属性相似度度量和权重分配很大程度依赖于专家知识,是一个由专家经验维护的系统。

2.2.2 基于专家经验的方法

相比于基于统计方法计算告警的相似性,Cuppens 等[16-17]提出聚类稳定性的概念,基于专家系统定义告警相似性,定义谓词逻辑,通过专家规则定义告警字段的相似性,该方法更加依赖于专家知识。Zhang 等[18]设计决策支持模型,定义告警聚合规则,通过人机交互为告警聚合提供依据,其本质仍是基于专家知识的方法。

2.2.3 基于数据挖掘的方法

由于基于相似度和专家经验的告警聚合方法都过分依赖于专家知识,许多学者对此进行了改进。数据挖掘[19](又称知识发现)指从海量不完整,嘈杂的随机数据中自动搜索隐藏特殊关系信息的过程,便于数据拥有者理解和使用。聚类是数据挖掘的方法之一,以无监督的方法将样本化分成群或类。聚类技术应用于告警聚合领域,主要方法有:

(1) 基于层次的告警聚合算法。Julish[20-21]认为现有的聚合技术并不能很好地解决误报率高的问题,作者将属性泛化 AOI(attribute-oriented induction)应用于聚类技术,并提出一种新的采用概念聚类的启发式告警聚合方法,以一般到特殊的方式分解告警属性,对告警信息进行分层,比较两个告警的不同点,并将具有相同产生原因 RC(root cause)的告警聚合成一个元告警,然后根据 RC 过滤误告警。该方法在聚合告警的同时,过滤了误告警,但该方法并不区分告警不同的属性,而且假设所有的类具有固定的大小。Mamory 等[22]采用半监督的方法,总结了 AOI 相关技术,并对 Julish 方法的过泛化问题进行研究,引入最近的共同祖先 NCA(nearest common ancestor)的概念,提出一种概率聚类算法聚合告警信息。

基于层次的告警聚合算法适用于任意形状的聚类,且能控制不同层次的聚类粒度,但计算复杂度较高,且存在无法回溯的不足,适用于离线场合(对日志文件的分析等)。

(2) 基于人工神经网络的聚合算法。人工神经网络 ANN(artificial neural network)[23-24]模拟人脑的工作过程,对大规模复杂的非线性系统具有良好的模拟能力。广泛应用于告警聚合的神经网络算法有自组织神经网络 SOM(self-organizing map)、自联想神经网络 AA(auto-associative)与学习向量机 LVQ(learning vector quantization)。

SOM 是 Kohonen[25]提出的一种无人监督有竞争的人工神经网络算法,通常应用于样本分类,排序和检测等领域。Kumar 等[26]将 SOM 应用于告警聚类,将告警的属性特征作为模型输入。在训练阶段,将每一个神经元的权向量映射到二维空间。该方法使用以下公式来计算告警之间的距离(相异性),其中,x 为特征集$(\alpha_1,\alpha_2,\cdots,\alpha_n)$,$w_k$为权向量,$w_g$为最佳匹配单元。

$$\|x-w_g\|=\|x-w_k\|$$

LVQ[27]是一种有监督的竞争式的神经网络模型,为提高聚类的准确性,使用有标签的数据训练样本。Wang 等[28]将 LVQ 应用于告警聚合流程,将告警集分成五类,即 normal、probing、DOS、U2R、R2L。作为有监督的竞争式的神经网络模型,LVQ 使用有标签的数据训练样本,其聚合准确率高于 SOM,且学习速度高于 BP 模型(学习速度慢且难以收敛),可以很好地处理大规模告警集。

AA[29]是一种无监督的前馈神经网络模型,具有

对称拓扑结构,相比于 SOM 算法,其优点是不需要先验知识,而是在训练数据时通过自组织的方式获得,但聚类精度相比不高。Smith 等[30]结合了 AA 算法与 SOM、EM 算法,设计实现了告警聚合系统,借鉴误差反向传播算法训练数据,系统结构分为三层,包含 n 个输入层单元,n 个输出层单元,以及 $j(j<n)$ 个隐含层单元。通过下列公式得到每一层第 i 个单元的输出,其中,y_i表示神经单元 i 在接收到上一层神经单元的信号 k_i后的输出。

$$Q_i = f\left(\sum_{k=1}^{k_g} w_{ik} y_k\right)$$

采用神经网络的告警聚合方法,通常基于并行结构,处理速度快,在噪声的环境下鲁棒性和容错能力强,但诸多训练参数需要设定,如拓扑结构、权值和阈值的初始化等,由于不能观察学习过程,输出结果难以解释,进而可能影响结果的可信性和可用性。

2.2.4　基于模式搜索的方法

Siraj 等[31]提出一种基于 IUR(improved unit range)的混合聚类模型,使用主成分分析法 PCA(principal component analysis)和期望最大化 EM(expectation maximization)算法合并相似告警,以减少告警的数量。比较了 EM 算法和其他无监督机器学习算法(SOM,K-mean,FCM)的性能,并通过实验证明 EM 算法性能最优(90.33% IPCA)[32]。不同的是,Siraj 使用了主成分分析法,提升了告警聚合的精度和速度,而 Hofmann 的方法建立数据流模型,告警聚合的速度快,可以用于实时告警聚合。

基于模式搜索的方法具有高聚合精度并充分考虑噪声数据,但该方法假设告警属性概率分布彼此独立。而且,概率分布的更新和存储开销很大,并且不适用于大规模数据告警集。

2.2.5　基于均方误差的方法

该方法在聚类技术中是最直观和经常使用的,对密集分布、类之间距离较远时效果较好。典型的算法有 K-means 算法和 ISODATA 算法[33-37]。K-means 算法使用欧几里德距离来测量相似度,并使用平方误差 dist 和作为聚类评估指标,可用下式计算:

$$dist = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Tjhai 等[34]提出一个告警聚合架构,包括聚合和分类阶段,分别使用 SOM 和 K-means 算法。SOM 算法具有自动聚类、容错性、特征映射和可视化等优点,但 SOM 算法本身不足以区分不同的类,为了改善这个不足,采用 K-means 算法,以确定类之间的边界,同时将预定义告警簇的数目作为 SOM 算法的输入向量数。然而,Fatma 等[35]认为将 K-means 算法与 SOM 算法相结合仍有不足,即需要人为定义告警频率和时间间隔以确定正确的分类,其提出的方法很好地解决了该问题,第一步采用基于 K-means 的 SOM 算法或基于 FCM 的神经核气算法将一定时间段内的告警聚合成元告警,且实验表明基于 FCM 的神经核气算法聚合效果优于基于 K-means 的 SOM 算法。第二步为过滤误告警,将元告警二分成正确告警和误告警,并提供了三种解决思路,即基于 K-means 的 SOM 算法、SVM、DT,且实验表明 SVM 性能优于另两种方法(由于其较高的检测率)。

ISODATA 算法[36]基于 K-means 算法对聚类结果进行组合和分割,并使用误差平方作为聚类准则,通过迭代改进聚类效果。在算法运行期间,聚类中心数量动态变化,通过重复校正以获得更合理的聚类数量 K,显著降低了人为误差。Man 等[37]采用 ISODATA 算法以解决 IDS 的告警洪流和大量重复告警的问题。在随机选定初始聚类中心的基础上,将全部样本按相似度准则进行聚类,对聚类后的结果再次求解均值作为下一轮聚类的中心,通过多次的反复迭代完成最终的告警聚类。

基于均方误差的告警聚合方法具有简单高效的特点,然而,预先给出的簇的数量对聚合结果具有很大影响并且对噪声敏感。

2.2.6　基于进化算法的方法

进化算法可以有效解决许多传统算法无法解决的复杂问题,例如,随意性、复杂非线性力学、多峰函数等。典型的用于告警聚合的进化算法有遗传算法 GA 和人工免疫系统 AIS[38-39]。

(1) *遗传算法*　遗传算法 GA[38]借鉴生物进化理论中的自然选择和遗传机制来搜索最优解,其优点是不依赖样本分布的先验知识,并且不受初始解选择的影响。Wang 等[39]将其应用于 IDS 告警聚合领域,并比较了 GA 算法以及改进的 GA 算法(IGA)。Bahrbegi 等[40]采用七种不同和遗传算法用于告警聚合,并对比分析了实验结果。

(2) *人式免疫系统*　人体免疫系统[41]是一种自适应系统,可自动识别、自我组织,具有学习、记忆和适应动态环境变化和自组织特征,主要功能是识别自体和非自体细胞。文献[42-43,45]将人工免疫系统应用于告警聚合,可以在告警生成阶段有效过滤误告警,减少告警数量。基于进化算法的告警聚合方法较好地解决了均方误差方法难以获得合理聚类中心的不足,抗噪能力强,且基于人工免疫的聚合算法可以有效过

滤误告警,但该方法需要合适的训练集。

2.2.7 基于哈希函数的方法

根据 Mohamed 在文献[46]所述,基于数据挖掘的方法在告警聚合时存在两点不足:一是在属性归纳(泛化)过程中,假设告警可以聚合是因为它们拥有相同的特征或来自于同一个源;二是距离度量增加了错误的可能,且计算开销大(因为要计算每个告警以及类之间的距离)。Mohamed 采用一种基于散列函数的聚合方法,从告警中选择三个属性(目的 IP、告警 ID、时间戳),并通过 MD5 算法进行运算,得到哈希值。将新告警的哈希值与已有的告警(或类)进行匹配,若匹配成功,则将告警聚合;否则,创建一个新类。基于哈希函数,构建哈希表,使得告警特征比较的复杂度大大降低,但采用哈希值难以区分部分告警信息,聚合精度不高。

3 关联算法分析

告警关联通过对告警聚合结果进一步分析处理,目的是挖掘攻击意图,重建攻击场景。告警关联一般需要先验知识的支持,依据对先验知识的依赖强弱,可以将告警关联技术分为强依赖和弱依赖两类。

3.1 强依赖算法

根据知识表示的重点不同,强依赖算法可分为下列四种。

3.1.1 基于模型语言的关联方法

该方法特点主要是在知识库中明确攻击事件关联关系。早期采用攻击序列模板的方式表示关联关系,文献[47]提出 LAMBDA 语言,将告警之间的关系表示为 $R = r_1 \circ r_2 \circ \cdots \circ r_n$,其中 $\circ$ 表示运算符,在整个序列 E 中,定义了告警间的串行、并行等五种关系,利用运算符描述攻击序列模板,并采用字符匹配算法完成告警关联。攻击序列模板表示的关联关系相对简单,且不易修改扩充。文献[48]和文献[49]分别采用 CAML 语言和显示关联方法,采用模块化类过程语言表示告警间的时间关系、属性关系以及前后继承关系等,表达能力增强。

此类方法的一个重要应用是事件关系模板的获取,常用的方法有:基于专家知识构建关系模板,采用机器学习建模的方法,以及数据挖掘算法挖掘事件的频繁模式等,此处不作详述。

基于模型语言的关联方法需要人工描述攻击场景,具有简单高效的特点,但只能识别已知的攻击场景,且该方法对于漏告警比较敏感,存在漏报时无法将告警信息有效关联。

3.1.2 基于事件的前因/后果的关联方法

该方法依据告警事件产生的前因后果确定其关联关系,文献[50]利用这一思路来确定告警之间的关联关系,并在过滤误报、发现攻击意图以及挖掘新的攻击模式方面实际效果良好。Hu 等[51-52]基于多步攻击各步骤间的因果关联关系,在时空两个维度对告警信息进行融合,提出了基于攻击预测的安全态势量化方法,辅助安全管理员从整体上把握网络安全状态的变化趋势。

基于事件前因/后果的关联分析方法可以全面挖掘事件对关联关系的影响,但并非所有的关系都能够准确反映攻击意图。此外,这种方法构建知识库的成本很高,对漏报的容忍度很低。

3.1.3 基于 Petri 建模的关联方法

文献[53-54]提出了一种基于攻击行为的 Petri 建模关联方法,以区分告警事件和入侵动作。虽然模型中也包含前因和结果的概念,但核心是针对影响危害,将动作的前因和结果均量化为与影响危害的关系。告警关联结果是获取受危害资源和相应的置信水平,能够有效减少告警数目,而且该方法对告警信息抽象的层次较高,可以用于安全态势评估等领域。但整个模型参数较多,构建代价较大;而且该方法无法给出告警事件的关联序列,影响了对告警的深度分析和响应。

3.1.4 基于隐马尔可夫模型的关联方法

文献[55]将隐马尔夫模型 HMM(hidden Markov model)引入告警关联,实验结果表明,HMM 适用于攻击步骤的时间跨度变化大、有多种相似类型、难以准确提供复杂攻击训练数据的应用场景。HMM 模型适用于分析随机观测序列和隐藏的状态转移序列间的关系,观察序列对应攻击诱发的告警信息,而状态序列对应攻击步骤,状态转移概率描述了攻击步骤间的关联性。文献[56]中,作者利用 HMM 对告警进行关联并分类,实验结果优于决策树和神经网络。胡浩等[57]利用吸收马尔可夫链 AMC(absorbing Markov chain)描述多步攻击的状态转移过程,实现了入侵意图和路径的预测,但该方法侧重静态分析,如何结合实时告警信息,提升应急响应的时效还有待进一步研究。

3.2 弱依赖算法

弱依赖算法主要是从历史样本数据中提取攻击或告警关联关系,典型的方法有以下两种。

3.2.1 基于相似度的关联方法

基于相似度的关联方法把告警信息属性泛化为特征向量,通过计算属性相似度来分析告警间的整体相

似度,将当前告警与历史模板中告警的相似度大于阈值且相似度最高的告警进行关联;否则,构建新的关联队列并将当前告警加入队列中的首位置。

文献[58]利用模糊综合评判法设计告警关联算法,基于有监督的确信度学习过程实现误告警过滤,其定义告警关联度计算公式为:

$$SIM(a_{new},a_{old}) = \sum_{i=1}^{n} k_i \cdot r_{ij}$$

式中:$SIM(a_{new},a_{old})$表示告警之间的整体相似性,k_i为属性的权值,r_{ij}为模糊隶属度权值。

通过数据挖掘可以获得相似度函数,文献[59]将告警分为发现、扫描、提升特权、拒绝服务和隐蔽攻击五种类型。对于不同的告警类型 i 和 j,定义类型间的转换关系为 R_{ij},采用 Sigmod 函数 $\sigma_{ij}(\Delta t) = 1/(1 + e^{\alpha_{ij}+\beta_{ij}\Delta t})$分析告警的时间关系;将源 IP 地址间的关系分解为五种转换概率,即 $P_{ij}(r)$,$r = 0,8,16,24,32$。参数值可由挖掘告警数据集获取,告警的相似度函数表示为 $R_{ij} \cdot \sigma_{ij}(\Delta t) \cdot P_{ij}(r)$。

基于相似度函数的方法,具有不依赖专家先验知识的优势,但需要在大量样本数据训练的基础上确定各项参数值。此外,算法获得的逻辑关联告警仅具有相似的空间和时间特征,但难以刻画告警间的本质关系,不利于分析攻击者意图。

3.2.2　基于时间序列的关联方法

Qin 等[60]认为攻击的因果关系反映在告警信息上是时间序列的联系,采用自回归模型 AM(autoregressive model)设计了时间序列分析方法来发现新的攻击关系,同时利用了贝叶斯信念网络模型降低计算开销。类似地,殷其雷等[61]利用 Apriori 算法分析告警数据的关联规则,挖掘告警间的时序关系。

该方法有利于发现新的告警关系,且无需大量样本数据,但实际应用中,该方法对噪声告警非常敏感,由于不能简单依据随机分布的特征排除所有噪声,研究设计鲁棒性更强的时间序列关联方法具有重要意义。

3.3　对比分析

从现实应用角度出发,利用以下四个指标对上述告警关联方法进行综合比较,如表 1 所示。

(1) 在专家知识需求方面,是否依赖专家知识是评估关联算法的重要指标,尽管强依赖算法的应用更好。然而,需要大量的先验知识,难以维护,并且弱依赖性算法相对灵活。

(2) 在发现新攻击序列方面,除了基于模型语言和隐马尔可夫模型的告警关联方法外,其他关联方法还具有发现新攻击序列的能力。强依赖算法发现未知攻击的能力受限于领域知识,弱依赖算法基于报警统计特征,因此发现未知攻击序列的能力较弱。

(3) 在算法健壮性方面,健壮性的强弱主要指算法对误报和漏报的适应能力。基于事件的因果关系和 Petri 建模方法可以在一定程度上消除误报并推测漏告警。

(4) 在协同检测能力方面,复杂网络环境中,对某个目标的攻击可能由多个黑客协同完成。分析算法的基本原理,修改基于模型语言方法中源地址关系、修改基于 Petri 建模方法中角色的概念、以及在基于 HMM 模型的方法中按目的地址组织告警序列,以上三种方法具备一定的检测协同能力。

表 1　关联算法对比分析

关联算法	专家知识	新攻击序列	健壮性	协同攻击检测
模型语言	√	×	×	√
前因和后果	√	√	√	×
Petri 建模	√	√	√	√
隐马尔可夫模型	√	×	×	√
相似度	×	√	×	×
时间序列	×	√	×	×

4　应用分析

告警信息处理具有一定的复杂性,采用单一策略无法取得理想的结果,应灵活选择告警聚合与关联体系结构和算法。

4.1　体系结构

目前,入侵告警信息聚合与关联系统的体系结构包括集中式、层次式、分布式三种类型,如图 2 所示,集中式结构适用于较小的网络环境,而层次式和分布式结构适用于中大规模的网络。

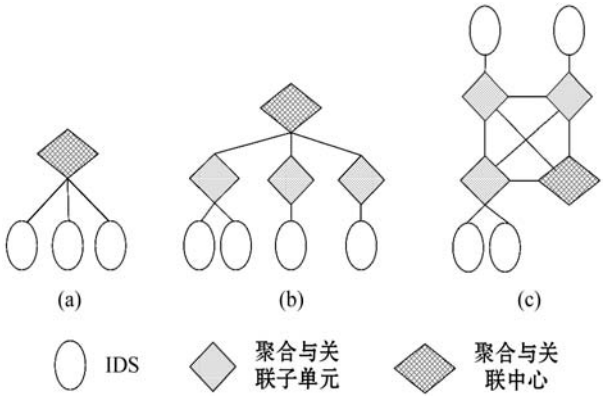

图 2　聚合与关联系统结构

(1) 集中式体系结构如图 2(a)所示。集中式告警聚合与关联系统中,中心处理单元将多源 IDS 告警数据进行聚合和关联,能够有效缩减告警处理时间,然而随着告警数量的增多,系统传输开销急剧增加,且中心处理单元易被来不及处理的告警信息"淹没"。

(2) 层次式体系结构如图 2(b)所示。鉴于集中式结构的缺陷,研究人员提出了一种分层告警聚合和关联结构,以定义几个不同等级的关联区域。每个聚合和关联子单元仅处理特定区域中的告警信息,然后将结果发送给上一级聚合和关联单元。该结构能够有效降低中心处理单元的工作量和网络通信负载,同时在一定程度上增强容错性。然而,由于中央单元的存在,整个聚合和关联系统在可扩展性方面没有得到实质性改进。

(3) 分布式体系结构如图 2(c)所示。在分布式结构中,中央处理单元用于维护整个系统,而其他处理单元不分级别,各自处理一定范围内的告警信息,并结合聚合和关联结果依据实际情况发送给其他单元。由于告警和处理单元间的相关性是已知的,因此能够显著降低网络负荷,同时整体关联效果良好。虽然没有了中心处理单元,系统可扩展性和安全性等得到了明显提升,但该结构实施较为复杂,维护成本高。

4.2 应用准则

告警聚合和关联算法各有利弊,分别适用于不同的环境或场合。在实际应用中,应遵循以下准则:

(1) 告警聚合和关联是告警信息处理的中间层次,对下是目标,对上是手段,其结果应利于高层安全策略,即有利于告警信息的进一步分析,如风险评估、态势感知、入侵响应等。

(2) 依据应用场景使用不同的聚合和关联算法。不同算法适用于不同场合,各有优势,算法间的互补性很强,系统可以根据安全要求综合使用不同的算法,以取得更好的聚合与关联效果。

5 结 语

近年来,入侵告警聚合与关联技术取得了长足的发展,很大程度上提高了 IDS 的性能和可用性。但仍面临诸多挑战,包括:(1) 告警信息数量与质量的矛盾,如何有效减少误报和漏报,并提高系统的检测率是一个难点。(2) 异构安全设备的告警聚合与关联,由于不同安全设备检测技术和攻击分类方法不同,在告警聚合和关联时,不仅要统一告警信息格式,还要实现整个系统内的攻击统一分类和映射。(3) 告警的深入处理,在告警聚合和关联基础上,如何联合不同的处理系统识别攻击意图,并与风险评估和入侵响应的结合也是应用难点。(4) 告警聚合与关联的实时性、统一规范的描述语言等也是未来研究的难点。

伴随网络技术的迅速发展,目前网络攻击呈现出多样化和复杂化的特点,对于告警 = 聚合和关联技术提出了更高的要求,呈现出以下发展趋势:(1) 智能处理方法,灵活应对复杂网络环境和攻击方法。(2) 关联规则挖掘,关联规则如何产生并应用于告警关联、分析网络态势等。(3) 可视化,协助安全管理员分析隐藏在大量告警消息背后的攻击模式和意图。(4) 态势评估,如何对告警信息进行量化评价,从而分析网络态势,得到系统更准确可靠的安全状态。(5) 入侵响应,判断真实的入侵和攻击行为,减少响应次数和经济成本。

简而言之,告警聚合与关联技术要求准确、实时、适应性强和扩展性好等,更有效地提高 IDS 和其他安全设备的适用性,以增强网络整体的安全保护和应急响应能力。

参 考 文 献

[1] 李文彬. 计算机网络安全与防护技术探究[J]. 软件导刊, 2015(5):152 - 153.

[2] Sadighian A, Zargar S T, Fernandez J M, et al. Semantic-based context-aware alert fusion for distributed intrusion detection systems[C]//Risks and Security of Internet and Systems (CRiSIS), 2013 International Conference on. IEEE, 2013:1 - 6.

[3] Liu T, Zhao Y C, Liu Y, et al. An alert fusion-based smart grid attack detection method[J]. Journal of Shandong University, 2014, 49(9): 35 - 40.

[4] 胡浩, 叶润国, 张红旗等, 面向漏洞生命周期的安全风险度量方法[J]. 软件学报, 2018, 29(5): 1213 - 1229.

[5] Hu H, Liu Y, Zhang H, et al. Security metric methods for network multistep attacks using AMC and big data correlation analysis[J]. Security and Communication Networks, 2018, Article ID 5787012: 1 - 14.

[6] Hu H, Liu Y L, Zhang H Q, et al. Optimal network defense strategy selection based on incomplete information evolutionary game[J]. IEEE Access, 2018, 6: 29806 - 29821.

[7] 刘同明. 数据融合技术及其应用[M]. 北京:国防工业出版社, 1998.

[8] 穆成坡, 黄厚宽, 田盛丰,等. 基于模糊综合评判的入侵检测报警信息处理[J]. 计算机研究与发展, 2005, 42(10):1679 - 1685.

[9] Curry D, Debar H. Intrusion Detection Message Exchange Format Data Model and Extensible Markup Language(XML)

Document Type Definition[EB/OL]. 2004. http://www.ietf.org/internet-drafts/draft-ietf-idwg-idmef-xml-11.txt.

[10] The intrusion detection message exchange format [EB/OL]. http:www.ietf.org/rfc/ rfc4765.txt.

[11] Mills D L. Network time protocol(version 3) specification, implementation and analysis [EB/OL]. http: www.faqs.org/rfcs/rfc1305.html.

[12] Valdes A, Skinner K. Probabilistic alert correlation[M]. Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, 2001:54-68.

[13] Debar H, Wespi A. Aggregation and correlation of intrusion-detection alerts[C]//Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection. Springer-Verlag, 2001:85-103.

[14] 龚俭, 梅海彬, 丁勇,等. 多特征关联的入侵事件冗余消除[J]. 东南大学学报:自然科学版, 2005, 35(3):366-371.

[15] 陈志文, 王开云, 姜建国. 网络入侵检测系统的告警合成算法设计[J]. 信息与电子工程, 2005(3):182-185.

[16] Cuppens F. Managing alerts in a multi-intrusion detection environment[C]//Computer Security Applications Conference,2001. ACSAC 2001. Proceedings 17th Annual. IEEE, 2001: 22-31.

[17] Cuppens F, Mige A. Alert correlation in a cooperative intrusion detection framework[C]//IEEE Symposium on Security & Privacy IEEE Computer Society. IEEE, 2002:202-215.

[18] Zhang Y, Huang S, Wang Y. IDS alert classification model construction using decision support techniques[C]//Computer Science and Electronics Engineering, International Conference on. IEEE, 2012:301-305.

[19] 王梦雪. 数据挖掘综述[J]. 软件导刊, 2013, 12(10): 135-137.

[20] Julisch K. Mining alarm clusters to improve alarm handling efficiency[C]//Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual. IEEE, 2001: 12-21.

[21] Julisch K. Clustering intrusion detection alarms to support root cause analysis[J]. Proceedings of ACM Transactions on Information & System Security Ny Usa, 2003, 6(4):443-471.

[22] Al-Mamory S O, Zhang H. Intrusion detection alarms reduction using root cause analysis and clustering[J]. Computer Communications, 2009, 32(2): 419-430.

[23] 丁硕, 常晓恒, 巫庆辉. 基于自组织特征映射神经网络的聚类分析[J]. 信息技术, 2014(6): 18-21.

[24] Macedo M N Q, Galo J J M, Almeida L A L D, et al. Demand side management using artificial neural networks in a smart grid environment[J]. Renewable & Sustainable Energy Reviews, 2015, 41(41): 128-133.

[25] Kohonen T. The self-organizing map[J]. Neurocomputing, 1998, 78(9): 1-6.

[26] Kumar M, Siddique S, Noor H. Feature-based alert correlation in security systems using self-organizing maps[C]//Proceedings of SPIE—The International Society for Optical Engineering, 2009.

[27] Demuth, Howard B. Neural network design[M]. China Machine Press, 2002.

[28] Wang J X, Wang Z Y, Dai K. A PCA-LVQ Model for intrusion alert analysis[M]//Intelligence and Security Informatics. Springer Berlin Heidelberg, 2006:715-716.

[29] Ravi V, Yadav A. Privacy preserving data mining using general regression auto-associative neural network: application to regression problems [M]//Swarm, Evolutionary, and Memetic Computing. Springer International Publishing,2014: 618-624.

[30] Smith R, Japkowicz N, Dondo M, et al. Using unsupervised learning for network alert correlation[C]//Conference on Canadian Society for Computational Studies of Intelligence. Springer-Verlag, 2008:308-319.

[31] Siraj M M, Maarof M A, Hashim S Z M. Intelligent alert clustering model for network intrusion analysis[J]. International Journal of Advances in Soft Computing & Its Applications, 2009, 1(1):33-48.

[32] Hofmann A, Sick B. Online Intrusion Alert Aggregation with Generative Data Stream Modeling[J]. Dependable & Secure Computing IEEE Transactions on, 2011, 8(2):282-294.

[33] 王茜, 刘胜会. 改进 K-means 算法在入侵检测中的应用研究[J]. 计算机工程与应用, 2015(17):124-127.

[34] Tjhai G C, Furnell S M, Papadaki M, et al. A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm[J]. Computers & Security, 2010, 29(6):712-723.

[35] Fatma H, Mohamed L. A two-stage technique to improve intrusion detection systems based on data mining algorithms [C]//Modeling, Simulation and Applied Optimization(ICMSAO),2013 5th International Conference on. IEEE, 2013: 1-6.

[36] 陈平生. K-means 和 ISODATA 聚类算法的比较研究[J]. 江西理工大学学报, 2012, 33(1):78-82.

[37] Man D, Yang W, Wang W, et al. An alert aggregation algorithm based on iterative self-organization[J]. Procedia Engineering, 2012, 29:3033-3038.

[38] 叶安新, 邓大勇. 基于改进量子遗传算法的聚类算法[J]. 计算机仿真, 2013, 30(4):275-278.

[39] Wang J, Cui B. Clustering IDS alarms with an IGA-based approach[C]//Communications, Circuits and Systems,2009. ICCCAS 2009. International Conference on. IEEE, 2009:586-590.

[40] Bahrbegi H, Ahrabi A, Mirnia M, et al. A new system to evaluate GA-based clustering algorithms in intrusion detection alert management system[C]//Nature and Biologically Inspired Computing(NaBIC), 2010 Second World Congress on, 2010, pp. 115-120.
[41] 莫宏伟. 人工免疫系统[M]. 科学出版社, 2009.
[42] 王慧. 基于危险理论的网络入侵检测系统研究[J]. 计算机仿真, 2010, 27(6):159-162.
[43] 白鹏翔, 张清华, 段富,等. 基于模糊规则的免疫算法在网络入侵中的应用[J]. 计算机工程与设计, 2015(12):3246-3249.
[44] Mahboubian M, Udzir N I, Subramaniam S, et al. An alert fusion model inspired by artificial immune system[C]//Cyber Security, Cyber Warfare and Digital Forensic(CyberSec), 2012 International Conference on. IEEE, 2012:317-322.
[45] 彭凌西, 谢冬青, 付颖芳,等. 基于危险理论的自动入侵响应系统模型[J]. 通信学报, 2012, 33(1):136-144.
[46] Mohamed A B, Idris N B, Shanmugum B. Alert correlation using a novel clustering approach[C]//International Conference on Communication Systems & Network Technologies. IEEE Computer Society, 2012:720-725.
[47] Cuppens F, Mige A. Alert correlation in a cooperative intrusion detection framework[C]//IEEE Symposium on Security & Privacy IEEE Computer Society. IEEE, 2002:202-215.
[48] Cheung S, Lindqvist U, Fong M W. Modeling multistep cyberattacks for scenario recognition[C]//DARPA Information Survivability Conference and Exposition, 2003. Proceedings. IEEE, 2003:284-292.
[49] Bouzar-Benlabiod L, Benferhat S, Boubana-Tebibel T. Integrating security operator knowledge and preferences to the alert correlation process[C]//Machine and Web Intelligence(ICMWI), 2010 International Conference on. IEEE, 2010:416-420.
[50] Ning P, Cui Y, Reeves D S, et al. Techniques and tools for analyzing intrusion alerts[J]. Acm Transactions on Information & System Security, 2004, 7(2):274-318.
[51] Hu H, Liu Y, Yang Y, et al. New insights into approaches to evaluating intention and path for network multistep attacks[J]. Mathematical Problems in Engineering, 2018, Article ID 4278632: 1-13.
[52] Hu H, Zhang H, Liu Y, et al. Quantitative method for network security situation based on attack prediction[J]. Security and Communication Networks, 2017, Article ID 3407642: 1-19.
[53] Yu D, Frincke D. A novel framework for alert correlation and understanding[M]//Applied Cryptography and Network Security. Springer Berlin Heidelberg, 2004:452-466.
[54] Du J, Wu Z, Chen M. Attack modeling using colored petri net and alerts correlation algorithms design[J]. Journal of Chongqing University, 2011, 34(4):118-124.
[55] Ourston D, Matzner S, Stump W, et al. Applications of hidden Markov models to detecting multi-stage network attacks[C]//Hawaii International Conference on System Sciences. 2003:73-76.
[56] 唐梦楠. 网络入侵事件检测及攻击行为预测的方法研究[D]. 西安:西安电子科技大学, 2014.
[57] 胡浩, 刘玉岭, 张红旗,等, 基于吸收马尔可夫链的网络入侵路径预测方法[J]. 计算机研究与发展,2018,55(4):831-845.
[58] 穆成坡, 黄厚宽, 田盛丰,等. 基于模糊综合评判的入侵检测报警信息处理[J]. 计算机研究与发展,2005,42(10):1679-1685.
[59] Dain O, Cunningham R K. Fusing a heterogeneous alert stream into scenarios[M]//Applications of Data Mining in Computer Security. Springer US, 2002:103-122.
[60] Qin X, Lee W. Discovering novel attack strategies from INFOSEC alerts[J]. Advances in Information Security, 2004, 31:439-456.
[61] 殷其雷, 吴平平. 基于 Apriori 算法的攻击行为时序关联规则检测方法[J]. 计算机安全, 2014(9):2-7.

(上接第 240 页)

[17] Sinha R,Swearingen K. Comparing recommendation made by online systems and friend[C]//Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries,2001:10-16.
[18] Li Y M, Wu C T, Lai C Y. A social recommender mechanism for ecommerce: combining similarity, trust, and relationship[J]. Decision Support Systems,2013,55(3):740-752.
[19] 罗辛,欧阳元新,熊璋,等.通过相似度支持度优化基于 K 近邻的协同过滤算法[J].计算机学报,2010,33(8):1438-1445.
[20] 邓爱林,朱扬勇,施伯乐.基于项目评分预测的协同过滤推荐算法[J].软件学报,2003,14(9):1621-1628.

(上接第 285 页)

[9] 牛培峰, 吴志良, 马云鹏,等. 基于鲸鱼优化算法的汽轮机热耗率模型预测[J]. 化工学报, 2017, 68(3):1049-1057.
[10] Awouda A E A , Mamat R B . Refine PID tuning rule using ITAE criteria[C]//2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE). IEEE, 2010:171-176.
[11] Heck D, Saccon A, Wouw N V D, et al. Guaranteeing stable tracking of hybrid position-force trajectories for a robot manipulator interacting with a stiff environment[J]. Automatica, 2016,63:235-247.