网络安全事件关联分析技术与工具研究

琚安康 郭渊博 朱泰铭 王 通
(信息工程大学 郑州 450001) (数学工程与先进计算国家重点实验室 郑州 450001)

摘 要 当前,以APT为代表的新型网络安全攻击事件频发并造成了巨大危害,其定制性、隐蔽性、持续性等特点使得传统攻击检测方法难以奏效。然而,随着大数据技术的日益发展,对各类安全相关事件及系统运行环境信息进行了有效关联,使得有效识别这类攻击和威胁成为可能,安全事件关联分析技术也随之应运而生。首先阐述了安全事件关联分析技术的重要性及其目标意义;然后对现有的安全事件关联分析技术进行了综述,从基于属性特征的关联分析、基于逻辑推理的关联分析、基于概率统计的关联分析、基于机器学习的关联分析等方面,分析描述了现有各种安全事件关联分析技术的机理及其优缺点;最后对现有的开源安全事件关联分析软件进行了综述,从应用场景、编程语言、用户接口以及关联方法等角度进行了综合比较。

关键词 关联分析,特征属性,逻辑推理,概率统计,机器学习
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.02.004

Survey on Network Security Event Correlation Analysis Methods and Tools

JU An-kang GUO Yuan-bo ZHU Tai-ming WANG Tong
(Information Engineering University, Zhengzhou 450001, China)
(State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China)

Abstract At present, the frequency of the new network security attacks events represented by APT is increasing, and it is more harmful to the enterprise information infrastructure. The new types of attack have the characteristics of customization, concealment and continuity, and these make it more difficult for traditional detection methods to detect or predict these deep-hidden attacks in time. However, with the development of big data technology, people can correlate the information about security events and system running environment effectively, and this makes it possible to detect new types of attack and threat. In this paper, we expounded the importance of security event correlation analytics, and then discussed the existing correlation analysis techniques from the aspect of event attributes, logical reasoning, statistics and machine learning. Finally we introduced several commonly used open-source correlation analysis software, and synthetically compared them in application scenarios, programming language, user interface, and the correlation method used.

Keywords Correlation analysis, Feature attributes, Logical reasoning, Statistics, Machine learning

1 引言

随着网络技术的发展,网络安全威胁的方式层出不穷,病毒、蠕虫、后门、木马等网络攻击方式越来越多,逐渐受到人们的广泛关注。为了保证网络系统的安全运行,在网络中广泛使用了防火墙、入侵检测系统、漏洞扫描系统、安全审计系统等安全设备。但是这些传统方法在应用中也存在一定问题。由于这些设备存在功能单一、各自为政、不能协同工作的特点,因此产生的这些安全事件信息中含有大量的重复报警和误报警,而且各类安全事件之间分散独立,缺乏联系,相互之间缺乏联系的安全事件无法给安全管理员提供在攻击时序上和地域上真正有意义的指导。同时,APT(Advanced Persistent Threat)攻击的出现给企业和相关机构带来了巨大威胁,从典型案例来看,APT攻击具有极强的隐蔽能力和针对性,传统的检测方法无法有效检测出APT攻击,而且海量网络安全数据的出现也带来了新的挑战,不仅增加了检测的时间成本,也对检测模式提出了新的要求。

在实际中,大部分的安全事件并不是孤立产生的,它们之间存在着一定的时序或因果联系。安全事件关联分析是指结合安全事件的运行环境,对原来相对孤立的低层的网络安全事件数据集进行关联整合,并通过过滤、聚合等手段去伪存真,发掘隐藏在这些数据之后的事件之间的真实联系,辅助识别网络威胁和复杂的攻击样式,生成高级层面的安全场景,为网络安全态势感知及攻击发现提供可靠的知识支撑。

近年来,网络安全事件关联分析技术得到了领域专家的高度关注,国内外研究人员提出了多种有意义的网络安全事

到稿日期:2016-01-27 返修日期:2016-05-25 本文受国家自然科学基金(61501515)资助。
琚安康(1995—),男,硕士生,主要研究方向为信息安全,E-mail:jusissp@yeah.net;郭渊博(1975—),男,博士,教授,博士生导师,主要研究方向为信息安全;朱泰铭(1991—),男,硕士生,主要研究方向为信息安全;王 通(1988—),男,硕士生,主要研究方向为信息安全。

件关联分析方法，也有一些文献对此进行了综述。例如，文献[1]从基于模型的角度分析比较了现有安全关联分析技术，并将其分为基于相似度、基于序列和基于实例的方法等。其中基于相似度的关联技术又包括基于属性[12]和基于时间关系[15]两种方法；基于序列的关联技术包括基于因果条件[6]、图算法[30]、代码书[40]、马尔可夫模型[34]、贝叶斯网络[31]、神经网络[35]等方法；而基于实例的方法[13-14]往往需要基于先验知识。文献[2]根据对知识库的依赖程度，从基于统计方法[44]、基于知识[11]以及基于相似度[10]3 种关联方法的角度对警报关联技术进行了分析。其中基于统计方法的关联技术不需要关于攻击场景的先验知识，可以用于检测新出现的攻击方式；基于知识的关联技术主要分为基于攻击场景的知识、基于因果关系的知识和混合知识等；基于相似度的关联技术采用的方法主要是将安全报警与现有的数据进行对比，结合过去的方法对当前遇到的问题进行处理。文献[3]则从关联分析过程的视角，对规一化处理、数据聚集、关联分析、误差分析、攻击策略分析、优先级确定等 6 个具体实现步骤的实现方法进行了分析总结，并对各自的优缺点进行了比较。

总体上讲，一方面，网络安全事件关联分析已经取得了很大进展；但另一方面，其相关研究还不成熟，相关技术的侧重点和应用场景都有着很大差别，技术点相对还比较零散，技术体系还不完整，理论与应用方面也存在着一定脱节。本文试图对此领域的研究现状进行梳理，从技术与实现两个角度对相关研究成果进行比较全面的分析总结，目的是为此领域的研究人员提供一定的参考和借鉴。

本文首先从基于属性特征、基于逻辑推理、基于概率统计和基于机器学习等方面对安全事件关联分析的不同技术方法进行阐述，然后对现有的开源安全事件关联分析工具进行分析比较。

2　基于属性特征的安全事件关联分析技术

基于属性特征的关联分析技术是指从事件自身的角度出发，分析属性特征之间的关联特性，并以此为基础配置关联策略，根据事件属性对其进行匹配检测。由于各种类型的安全事件本身具有特有的特征，因此此类方法要求人们对事件有较为深刻的理解，较多地依赖于专家知识，无法对未知问题进行有效关联。

2.1　有限状态机

有限状态机(Finite-State Machine，FSM)又称有限状态自动机，简称状态机，是表示有限个状态以及在这些状态之间的转移和动作等行为的数学模型。通常 FSM 包含以下几个要素：状态管理、状态监控、状态触发、状态触发后引发的动作等。通常，一个有限状态机可定义如下：

- 输入事件集合 I(输入字母表)
- 输出事件集合 O(输出字母表)
- 状态集合 S
- 初始状态 $S_0 \in S$
- 状态转换函数 $f: I \times S \rightarrow S \times (O \cup \varepsilon)$

有限状态机是输出取决于过去输入和当前输入的时序逻辑，根据状态转换函数定义下一状态，并生成可能的输出事件。一般来说，除了输入部分和输出部分外，有限状态机还含有一组存储有限状态机的内部状态的寄存器，常被称为状态寄存器。在有限状态机中，状态寄存器的下一个状态不仅与输入信号有关，而且还与该寄存器的当前状态有关，因此有限状态机又可以认为是组合逻辑和寄存器逻辑的一种组合。其中，寄存器逻辑的功能是存储有限状态机的内部状态；而组合逻辑又可以分为次态逻辑和输出逻辑两部分，次态逻辑的功能是确定有限状态机的下一个状态，输出逻辑的功能是控制有限状态机的输出。

在实际应用中，根据有限状态机是否使用输入信号，设计人员经常将其分为 Moore 型有限状态机和 Mealy 型有限状态机两种类型[29]。Moore 型有限状态机的输出信号仅与当前状态有关，即可以把 Moore 型有限状态的输出看成是当前状态的函数；Mealy 型有限状态机的输出信号不仅与当前状态有关，而且还与所有的输入信号有关，即可以把 Mealy 型有限状态机的输出看成是当前状态和输入信号的函数。对异常行为的识别通常包括两方面的内容：异常发现和异常定位。基于有限状态机的关联分析引擎通常应用于第一阶段，即异常发现阶段。

将有限状态机技术应用到网络安全事件关联分析中的本质在于某个安全事件会带来多个可能的事件序列。基于状态机的关联分析模型[17-18]的优点在于确定了系统状态及转换函数后，描述系统各行为动作时是清晰明确的，具有极强的逻辑约束性，适合于逻辑性较强的系统应用场景。在网络入侵检测方面，可用于业务逻辑清楚且安全要求较高的应用场景，但是也因为其强逻辑性，难以区分攻击事件和人员误操作，所以会带来通用性不好的缺点；而且在确定了状态转换逻辑之后，不支持场景的动态变化，针对不同的应用场景必须重新建立状态模型，且当环境发生变化时需要进行适当调整。

2.2　基于规则

基于规则的关联方法[19-21]是最易实现且也是效率最高的方法，但其缺点也很明显：使用不灵活且配置困难。在实际应用中，通常根据指定的条件动作关系设计和制定关联规则，即为每个规则指定一个条件(如“在不到五分钟内事件 A 至少发生十次，且在不超过一分钟后 B 事件发生”)和相应的操作(如“向管理员发送报警消息”)。对一个规则的评估由相应的输入事件触发，这些规则通常被称为事件条件操作(ECA)规则。

基于规则的关联方法通过提供更细粒度的关联分析减轻场景定义不明确的问题，通过匹配系统必备组件(前提条件)和后果(后置条件)的攻击步骤来关联告警信息。从这个意义上讲，基于规则的关联可被视为特殊的基于场景的关联分析技术。

基于规则的关联分析方法在具体实现时常利用一些改进的 RETE 算法[8-9]。1982 年 Forgy 提出 RETE 算法[7]之后，RETE 算法一直都是目前基于正向链推理最有效的算法，RETE 算法采用控制逻辑评估关联规则，控制逻辑的作用是持续监测所有规则，直到找到第一个匹配的规则。

基于规则的关联分析方法的好处在于使用的简单的 if-then 风格语言与自然语言很相似，便于理解和制定，且可以

准确侦测到底层安全事件之间的关联关系;但是缺点也很明显:规则库依赖于有特定领域经验的专家知识,无法关联出未定义的告警事件。

2.3 基于 codebook

基于 codebook 的方法[40]采用编码技术对安全事件进行关联分析,需要有足够大的编码本来确定某个安全事件的关联类型。如果编码本太大会带来不必要的冗余,若过小又不足以提供足够的信息来区分不同的安全事件。基于 codebook 的关联模型的优点在于匹配过程中不需要专家知识将问题和事件联系起来。事实上,codebook 关联模型更像是一种算法机制,它将检测和标识系统中的异常事件看作编码问题,将由问题关联出的完整事件集视作某个问题的"代码"。因而事件关联的进程仅仅是对一系列监测症状进行"解码"的过程,即判定哪个问题的"代码"最大限度地匹配观测症状。如图 1 所示,事件 W,X 和 Z 作为观测事件,分别针对问题 A,B 和 C 作向量编码,当确定问题类型时,若监测到事件 X 和 Z 发生,从编码本中可以找到匹配问题类型为 A;如果监测值没有对应问题类型,则选取 Hamming 距离最近的问题向量作为匹配向量。

	A	B	C
W	0	1	0
X	1	0	0
Z	1	0	1

图 1 Codebook 关联矩阵示例

基于 codebook 的关联分析分为两个阶段。

第一阶段是编码阶段。在这一阶段,首先将每一个征兆或告警与问题联系起来,然后选择监测事件集,生成有效的"代码"以进行问题标识,这个过程的结果被称作 codebook,其实质是生成一个"问题-征兆"矩阵。然后再将 codebook 进行优化以用于监测,使得其能够区分不同的事件,且能同时保证达到所期望的噪声容错度。由于在网络系统中,事件可能被丢失和延迟,也可能产生错误的告警,或可能存在未检测到的安全事件,因此必须使关联算法在有噪声的情况下准确地标识问题。

第二阶段是解码阶段。在这一阶段,关联分析引擎对告警事件流进行解码,将事件流和 codebook 进行比较和分析,找出最接近观测症状的"代码",从而最终确定发生的问题类型。借助于事件知识模型进行预处理以降低实时事件关联分析的复杂性,并且其固有的纠错能力为事件关联提供了一定程度的容错性,因此基于 codebook 技术的关联分析引擎在速度和性能上具有很大的优势,另外,还可以引入概率模型以适应不确定性环境。基于 codebook 的安全事件关联分析技术还有一个重要优势,即它具有关联跨越不同管理对象事件的能力。但其不足在于需要事先构建 codebook,且编码结果直接影响关联性能,codebook 构建方法的约束性限制了此方法的推广应用。

3 基于逻辑推理的安全事件关联分析技术

基于逻辑推理的关联分析技术是指从事件之间的关联关系出发,合理选择和有效运用相关知识,利用专家知识进行推理,最后完成问题求解,常用的技术方法是实例推断和模型推断。基于逻辑推理关联技术的核心是推理控制策略的设计,即如何合理地选择所需知识。采用此类方法的优势在于可有效解决复杂问题,但是消耗计算资源较多且实时处理效率不高。

3.1 实例推断

基于实例推断的思想源于现实生活中的应用场景,在现实生活中,一些类似情形总是重复发生,处理某一特定情形的方法在其他情形中也能适用,而这些类似的情形并非要与该特定情形完全一致。因此,当试图解决一个问题时,都是从曾经经历过的类似案例出发,基于实例推断的关联技术主要依据这一思想,利用类比推理的方法得到新问题的近似解答,再加以适当修正,使之完全适合新问题。

基于实例推断的关联技术[23-24]是指:将解决过的成功案例作为知识存储起来,遇到新问题时,在实例库中查找相似的案例,对其修正后作为新问题的解决方案。一般地,基于实例推断的关联分析体系的结构由 5 个部分组成,其中包括 1 个实例库和 4 个功能模块,4 个功能模块分别是输入模块、检索模块、修改模块和处理模块。首先,输入模块接收用户提供的问题描述;接着由检索模块在实例库中寻找与之匹配的事例,如果能找到完全匹配的事例,那么就应用该事例的解,问题就迎刃而解。如果找不到完全匹配的事例,检索模块就在实例库中找一个最近似的事例,然后由修改模块对该事例的解作适当的修正即可满足当前问题的要求,其结果是得到一个新问题的解;一旦问题被解决,则处理模块根据解决方案对问题进行处理。

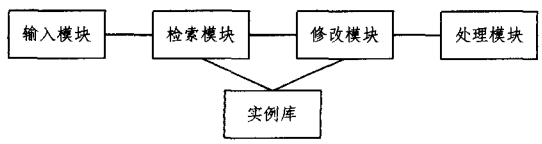

图 2 基于实例推断的关联分析体系结构

采用上述方法对安全事件进行关联分析的优势在于具有学习能力,可以解决一定的未知问题,但是时间效率低,不适用于实时事件关联。

3.2 模型推断

基于模型推断的关联技术[26]将系统中的每一个部件都表示为一个模型,需要对模型进行结构描述和行为描述,以及违背上述两种描述的异常定义,这种方法的缺点在于解决问题的复杂性较高。在基于模型的关联分析系统中,每个管理对象都有一个模型作为其副本与之相联系,一个模型实际上就是一个软件模块,事件之间的关联通过各个模型相互协作实现。

虽然基于模型推断的关联技术在实现上会使用基于规则的方法,但是与基于规则的方法针对于特定行为的事件模式不同,基于模型推断的关联技术更关注一系列事件序列或模型状态,在这种意义上更接近于基于 FSM 的方法。

在应用中,基于模型推断的关联技术较多应用于逻辑回路的错误诊断。例如,在实际网络中一般都会周期性地对某一路由器发出命令以检测其是否正常工作,在基于模型的推理系统中,这一过程实际上是通过该路由器的模型周期性地对路由器发出命令来实现。

4　基于概率统计的安全事件关联分析技术

基于概率统计的安全事件关联分析技术是指从事件发生概率和统计数据角度出发，对报警信息间的关系以概率形式进行刻画，揭示网络安全事件的时序和因果关系。这种方法的优势在于现有统计分析方法已经较为成熟，但缺点也很明显，即需要借助专家知识进行验证和性能调优，因此针对未知攻击方式和存在大量冗余报警的事件集合的关联效果较差。

4.1　投票机制

通过投票机制可以定位网络中的错误或异常[28]。通常，选票（由不同的节点事件表示）不能提供故障位置的确切信息，但其可以指示一定的方向和范围来辅助决策。正如文献[4]指出，在故障定位过程中，基于投票机制的关联分析引擎事先获知网络的拓扑结构，并且可以计算出每个网络组成元素的投票数量，从而得到故障发生的可能位置。

文献[5]描述了一个使用投票机制的示例场景，在该场景中，从一个可能的消息类型中确定一个故障信息（触发了具有脆弱性设备中的一个程序漏洞），采用神经网络的方法检查每个子网，每一个神经网络确定哪一个消息类型最有可能是该消息的错误消息并进行投票，最后根据所有消息类型的选票确定最有可能的错误消息类型。

4.2　依赖图

基于依赖图（Dependency Graphs）的关联分析技术[30]通过将收集到的安全警报映射到基于时序信息的图形中，将警报之间的关系表示为一个有向图，其中节点集表示警报，这些节点相连的边表示连接警报（节点）的时空关系，也称为报警关联图。

报警关联图具有下面几个优点：1）从管理角度看，图是比较简单的模型，即图的形式可以很好地反映节点之间的关联关系；2）在关联图上的操作实现具有鲁棒性，即添加或删除对象和依赖关系都是简单的原子任务；3）图表是一种自然的分布式结构，易于管理，可以由不同的管理员独立添加或删除对象和从属关系。

在使用依赖图的同时往往也会结合决策树模型。决策树是一种树状结构，用于揭示数据中的结构化信息，利用该结构可以将大型记录集分割为相互连接的小记录集，通过每一次连续分割，结果集中的成员彼此之间变得越来越相似。使用决策树算法便于将数据规则可视化，构造决策树的过程所需的时间也比较短，且输出的结果容易理解。决策树具有分类精度高、操作简单以及对噪声数据具有较好健壮性的优点。C4.5 和 ID3 是常见的决策树算法方法，但是 C4.5 决策树算法在生成树的过程中，需要对样本集进行多次扫描和排序，导致算法的效率比较低；另外，C4.5 算法只适用于可以驻留于内存的数据集，当训练集超过内存的容纳能力时，程序就无法运行，这使得该算法对硬件的要求比较高。

4.3　贝叶斯网络

贝叶斯网络模型，也被称为信度网络（或简称为贝叶斯网），是最强大的概率图形（GMs）之一，用于表示关于不确定性的域间知识。贝叶斯网络主要由两个部分组成。

1）一个图形组件组成的向无环图（DAG），图的顶点表示事件，边表示事件和事件之间的关系。

2）由数值组件组成的概率依赖关系，DAG 的每个节点表示在其父母的上下文中条件概率分布的不同环节之间的关系，图中的每个节点代表一个随机变量，而节点之间的边代表相应的随机变量的概率依赖关系。

贝叶斯网络由几个参数定义，即是由先验概率的父节点状态和一组与子节点相关联的条件概率表（CPT）组成的，条件概率表反映子节点和其父节点之间的先验知识。

贝叶斯网络用于解决告警关联问题时是有明显的优势。首先，关联分析的处理速度快；其次，可以通过填充条件概率表，合并先验知识和专家知识；再次，便于引入新数据以发现未观测到的变量概率；另外，可以通过网络传播更新适应新的数据和知识。但是其也存在一定的缺点，即这种方法需要大量的训练活动以取得先验概率，且需要依赖于专家知识。此外，基于贝叶斯网络的概率推理是 NP 难问题，针对大型网络实际上很难实施有效的解决方案。

4.4　马尔科夫模型

马尔科夫模型（Markov Models）[34]是由离散状态和状态转移概率矩阵组成的随机模型，在此模型中的事件被假定遵循马尔科夫特性，模型的下一个状态只取决于当前状态，而不依赖于之前事件的顺序。在马尔科夫模型的定义中，需事先设定好状态间的跃迁概率和初始状态概率，这些参数可以静态定义，也可以通过对数据集进行训练得到。马尔科夫模型经过训练得到定义相关的概率，即通过对一连串的事件进行评估从而获得概率值，将概率与门限值进行对比，从而确定事件之间是否存在关联性。隐马尔可夫模型（Hidden Markov Models，HMMs）中的状态序列不可见，由于攻击过程的行为序列有较强的先后顺序，一个步骤必须在另一步骤之后才能达到攻击的目的，因此隐马尔科夫模型更适合于结合攻击的先决条件来解决定位多步攻击问题。

一般来说，基于马尔科夫模型的关联分析技术更适用于解决连续性质的问题，马尔科夫模型的主要缺点是需要经过适当的训练，且性能依赖于参数调优。

5　基于机器学习的安全事件关联分析技术

基于机器学习的安全事件关联分析技术应用数据挖掘和机器学习的方法训练数据集，生成事件关联规则，通过关联分析得到新型攻击事件模式，是一种可实时运行的事件关联方法。这种方法的优点在于可以自动地为安全事件建立关联模型，为分析管理大量报警信息节省了时间，其结果提供的信息便于分析人员阅读。缺点在于需要对数据进行训练，可能造成结果线程过于庞大，而不存在于线程中的数据则无法进行关联，影响到最后的分析结果的准确性。

5.1　神经网络

人工神经网络（ANN）通过大量相互关联的处理单元即神经元（Neurons）共同作用来解决具体问题，该模型主要受人脑中的神经系统模型的启发。各神经元之间相互关联，每个神经元可被看作是一个简单的自动处理单元，本地提供内存和单向通道，使其与其它神经元进行通信。神经网络通常用于复杂关系的建模或数据输入与输出之间存在非线性依赖关系的场景。

在神经网络中最重要的问题是学习阶段，通过不断调整神经元间的连接强度(权重)，直到整体网络获得所需训练集的观测结果。神经网络的训练方法主要有两种：监督训练和无监督训练。无监督训练靠系统自身找到最优性能点，系统自身不受外部的影响，监督训练则需要网络提供样本输入和输出模式，基于训练样本集进行迭代学习，直到达到符合要求的最佳工作点或达到预先设置的阈值。

基于人工神经的网络安全事件关联技术的优势在于：1)人工神经网络具有抗噪声输入的特点；2)这种方法更便于推广结果的性能，这意味着一个经过良好训练的网络既能学习相同类型的数据，也可以对从未见过的数据进行有效分类；3)人工神经网络事先整合了专家知识，可以直接处理获取的数据，无需建立域规则和实例；4)一旦训练完成，可以构造快速、准确且有高精度的近实时应用。然而，基于人工神经网络的关联方法自身也存在一些缺陷，比如在训练过程中可能采取长会话的方式调整权重；此外，也没有特定的规则来确定网络的层数和每层神经元的数目。因此，在训练阶段，需要经历多次试错的过程，直到网络最终稳定。

由于人工神经网络具有非线性逼近性、自适应性、容错性、较好的并行特性和实时特性等诸多优点，将人工神经网络技术应用于入侵关联分析，可以较好地解决传统方法难以解决的问题，提高系统检测率并减少误报率。

5.2 支持向量机

神经网络与支持向量机(Support Vector Machine，SVM)都源自于感知器(Perceptron)。感知器是由 Rosenblatt 于1958 年发明的线性分类模型，可有效进行线性分类，但现实中的分类问题往往是非线性的，神经网络与支持向量机在感知器的基础上进一步发展，可解决非线性分类问题。经典感知器模型如图 3 所示。

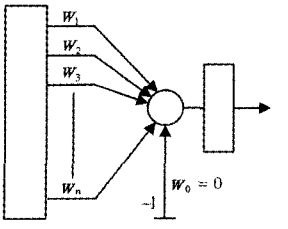

图 3 经典感知器模型

1986 年，Rummelhart 与 McClelland 发明了神经网络的学习算法 BackPropagation。后来，Vapnik 等人于 1992 年提出了支持向量机算法[16]。神经网络是多层(通常是三层)的非线性模型，支持向量机则利用核技巧把非线性问题转换成了线性问题。

支持向量机算法建立在统计学习理论的基础上，能从大量训练数据中选出很少的一部分用于构建模型，通常对维数不敏感，SVM 更适合于小样本数据的机器学习，更符合实际高速网络中的分类场景。另一方面，SVM 能在很大程度上克服传统机器学习(神经网络、决策树等)的维数灾难和局部极小等问题，具有比较好的泛化能力，具有良好泛化性能的入侵检测系统对其实际应用有着重要的意义，因此 SVM 也是研究构建入侵检测系统的重要算法。

表 1 网络安全事件关联分析技术

关联类别	技术方法	主要特点	应用场景	文献
属性特征	有限状态机	行为动作明确清晰且具有极强的逻辑约束性；不够灵活，不支持场景的动态变化	逻辑性较强的系统应用场景	[17-18]
	基于规则	最易实现且效率最高；使用不灵活，配置困难，规则库依赖于专家知识	通用场景	[19-21]
	codebook	关联匹配速度快，性能较好，关联过程不需要专家知识；编码生成过程复杂	异常事件匹配识别	[40]
逻辑推理	实例推断	具有学习能力；时间效率低，不适用于实时事件关联	非实时处理环境	[23-25]
	模型推断	模块化协作；结构描述、行为描述和异常定义	逻辑回路错误诊断	[26]
概率统计	投票机制	指示范围辅助决策；无法给出确切信息	分布式网络下的网络错误或异常定位	[28]
	依赖图	直观反映关联关系，易于管理；生成算法效率较低，依赖资源	反映事件因果的偏序关系，生成操作信息流图	[30,41-42]
	贝叶斯网络	综合了先验知识和专家知识，且可随环境变化更新；需要大量训练，依赖专家知识，针对大型网络效果较差	小型网络环境	[31-33,44-46]
	马尔科夫模型	可检测未知入侵，适用于连续事件流和实时检测；训练代价高，性能依赖于参数调优	多步攻击定位问题	[34,47-49]
机器学习	神经网络	非线性逼近性、鲁棒性以及容错性，自适应性，抗噪声输入，便于推广；调整权重和生成稳定网络的过程需要大量试错	快速、准确且有精度要求的近实时应用	[35-37,43]
	支持向量机	克服了维数灾难和局部极小等问题，具有较好的泛化能力	小样本、高速网络应用	[27,37-39]

6 网络安全事件关联分析工具

开源软件如今已经成为软件产业中不容忽视的开发模式，开源运动的影响力在不断扩大，获得了蓬勃发展；开源软件因其性价比高、可以根据需求进行改进的特点，受到开发者的广泛青睐。上文对网络安全事件关联分析方法进行了综合比较，列举出了各种方法的特点和适用环境，在具体应用中可结合现有的开源关联分析软件开展具体研究。下面对几款常用的关联分析软件进行介绍，并对其功能进行分析对比。

6.1 Swatch

Swatch[50]是一个由 Perl 编写的通用公共许可证(GPL)下的开源日志监控工具，它可以配置简单的规则进行日志处理，每一个规则包含一个正则表达式，采用模式匹配的方法对日志进行处理，对匹配的日志消息采取指定的动作，如在屏幕上输出、发送邮件或执行外部程序等。虽然开发者没有将 Swatch 描述为关联分析软件，但其自身功能支持简单的事件的关联操作，例如可规定阈值的范围，或制定基于时间窗的规则等。Swatch 也可用于在线应用环境中分析日志信息，即通

过实时的方式直接从一个程序的输出或 syslog 消息中读取日志消息来进行关联分析。

6.2 SEC

SEC(Simple Event Correlator)是由爱沙尼亚的 RistoVaarandi 开发的一个功能强大的事件关联引擎[51]，它完全由 Perl 编写，能够处理各种不同的事件关联任务。SEC 事实上是用 Perl 偏写的一款脚本程序，它从一个文件读取输入流，同时应用规则模式匹配相应操作，查找与规则相适应的输入流，SEC 的特性使得它非常适用于大量不同事件的关联任务，其设计初衷就是给开发者提供一个免费的、具有平台独立性的、轻量级的事件关联工具。SEC 关联处理过程示例如图 4 所示。

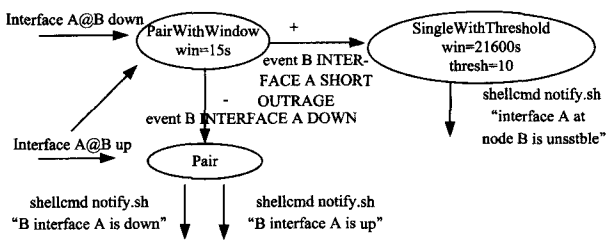

图 4　SEC 关联处理过程示例

6.3 OSSEC

OSSEC 是一款开源的基于主机的入侵检测系统(HIDS)，其由一个核心应用(一个代理程序)和一个基于 Web 的用户界面组成。根据 OSSEC 网站的说明[52]，它具备日志分析、文件完整性检查、策略监控、rootkit 检测、实时报警以及联动响应等功能。同时，OSSEC 支持多数操作系统，如 Linux，Windows，MacOS，Solaris，HP-UX，AIX 等；OSSEC 也可以从不同的设备和应用程序分析日志，如 Cisco 路由器，微软 Exchange 服务器，Nmap，OpenSSH 等。OSSEC 可能的输出包括记录为 syslog 的事件、数据库存储、发送电子邮件、生成报告等，用户也可通过 Web 界面访问查询。

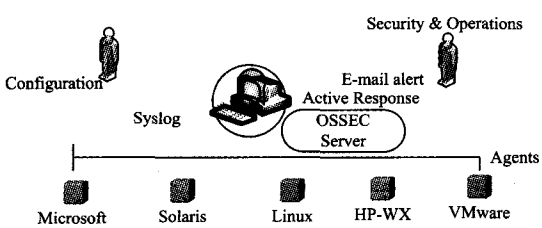

图 5　OSSEC 应用框架

6.4 OSSIM

OSSIM (Open Source Security Information Management)，顾名思义，是一款开源的安全信息管理软件[53]，是针对各种开源组件的集成(如 Nmap，Nessus，Snort，Nagios，OSSEC 等)为一体的综合性安全软件套件，其目的是为一个组织或单位提供集中式的、有组织的、更优良的安全事件监控的侦测和信息显示，进而提高系统的整体安全检测性能。OSSIM 开发者在网站上指出，开发人员依靠现有的开源产品，并添加了一些额外的工具，最重要的特性是增加了对通用关联引擎逻辑指令的支持。OSSIM 通过将开源产品进行集成，提供一种能够实现安全监控功能的基础平台，目的是提供一种集中式的、有组织的、能够更好地进行监测和显示的框架式系统。

OSSIM 的功能共划分为 9 个层次，各个层次之间无缝连接，底层的数据为上层的处理提供信息来源。图 6 是 OSSIM 所提供的功能层次结构图。

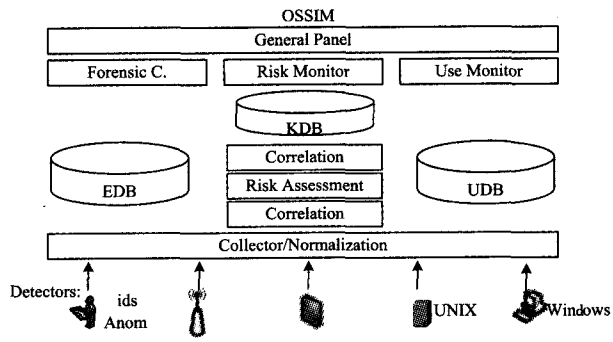

图 6　OSSIM 层次结构图

6.5 Drools

Drools 是一个用 Java 编写的开源规则引擎管理系统[54]，Drools 网站对此项目描述如下：Drools 是一个业务规则管理系统(BRMS)和增强的规则引擎，主要采用一种基于 JVM 实现的 ReteOO 算法。重要的是 Drools 提供声明式编程且具有足够的灵活性，是一款基于 Web 开发的工具，可提高开发人员的生产效率。

Drools 可以将复杂多变的规则从硬编码中解放出来，以规则脚本的形式存放在文件中，使得规则的变更不需要修正代码重启机器就可以立即在线上环境中生效。

6.6 Esper

Esper 是通过 Java 语言编写建立的一个开放源码的 CEP 和 ESP 工具包[55]。它的主要功能是对网络事件进行关联分析，其支持的 CEP 和 ESP 处理方式包括逻辑和时态操作、事件过滤、聚合、速率限制、阈值检测、排序或合并等。

Esper 是一种集中式的流式数据处理方式。它采用内存数据库，与传统的关系数据库相比，有更好的查询性能，更适合 CEP 应用场景。它提供两种机制来处理事件：1)通过状态机来实现基于表达式的事件模式匹配，这种事件处理的方法是匹配期望存在的事件、不存在的事件或事件的组合；2)通过 EPL 语句来实现事件流查询，这种事件处理的方法提供了过滤、滑动窗口、聚合、连接和分析等函数，EPL 采用视图将构造的数据放入到一个事件流中并驱动数据的流动，在数据流动的过程中对数据进行处理，最后得到需要的结果。

Esper 体系结构如图 7 所示。

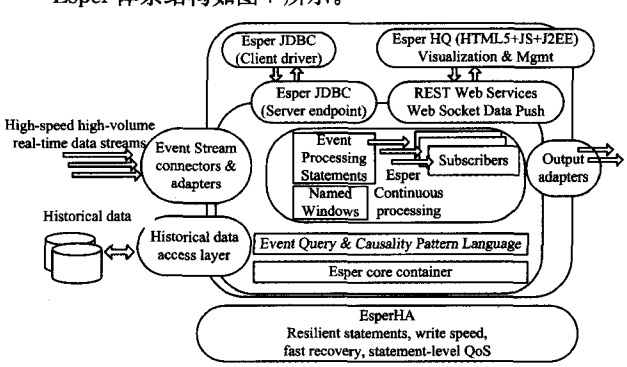

图 7　Esper 体系结构

表 2 对上述几款开源关联分析软件的应用场景、编程语言、用户接口以及所用关联方法进行了对比。

表 2 开源关联分析软件对比

开源软件	应用场景	编程语言	用户接口	关联方法
Swatch	日志监控	Perl	命令行	文本规则
SEC	通用场景	Perl	命令行	文本规则
OSSEC	基于日志的 IDS	C,PHP	Web 接口	Xml 规则
OSSIM	安全管理	C,PHP Python	Web 接口	资产和脆弱性分析
Drools	规则引擎	Java	API	DRL,DSL
Esper	CEP,ESP	Java	API	类 SQL 规则

结束语 网络安全事件关联分析是检测 APT 等新兴攻击手段的重要方法，也是系统整体安全态势预测的关键技术。本文对现有研究成果进行了分析总结，从事件属性特征、逻辑推理、概率统计和机器学习 4 个方面分别介绍了其技术特点，并对几款常用的开源关联分析软件进行了分析比较。

现有网络安全事件关联分析技术的局限性包括：关联方法的适应性和自动化程度不够，而且检测的准确度有待进一步提高；安全事件关联分析的事件类型仍不够全面，而且在大数据计算环境下数据处理能力大大提高，提出一种更有效的关联分析方法来处理多源网络安全事件是下一步研究的关键。

参 考 文 献

[1] SALAH S. A model-based survey of alert correlation techniques [J]. Computer Networks, 2013, 57(5): 1289-1317.

[2] AL-MAMORY S O, ZHANG H L. A Survey on IDS Alerts Processing Techniques[C]// 6th WESEAS International Conference on Information Security and Privacy. Tenerfe, Spain, 2017.

[3] SADODDIN R, GHORBANI A. Alert correlation survey: framework and techniques[C]// Conference on Privacy, Security and Trust. 2006.

[4] GUPTA R K, CHO S Y. A Correlation-Based Approach for Real-Time Stereo Matching[M]// Advances in Visual Computing. Springer Berlin Heidelberg, 2010: 129-138.

[5] ANTONELLO M, PRETTO A, MENEGATTI E. Fast Incremental Objects Identification and Localization using Cross-correlation on a 6 DoF Voting Scheme[C]// Special Session on Active Robot Vision. 2014: 499-504.

[6] XIAO S, ZHANG Y, LIU X, et al. Alert Fusion Based on Cluster and Correlation Analysis[C]// Proceedings of the International Conference on Convergence and Hybrid Information Technology, Daejeon, South Korea. 2008: 163-168.

[7] FORGY C L. Rete: A fast algorithm for the many pattern/many object pattern match problem[J]. Artificial Intelligence, 1982, 19 (82): 17-37.

[8] GU X D, GAO Y, HUANG J. Rete Algorithm: Current Issues and Future Challenge[J]. Computer Science, 2012, 39(11): 8-12. (in Chinese)
顾小东，高阳，黄峻. Rete 算法：研究现状与挑战[J]. 计算机科学，2012，39(11)：8-12.

[9] WEN J R, WANG Y L, LIU W. Improved algorithm for RETE supporting multiple types of imperfect metric[J]. Computer Engineering and Applications, 2015, 51(15): 48-55. (in Chinese)
文举荣，王永利，刘伟. 支持多类型瑕疵度量的 RETE 改进算法[J]. 计算机工程与应用，2015，51(15)：48-55.

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

[11] 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.

[12] ZHUANG X, XIAO D, LIU X, et al. Applying Data Fusion in Collaborative Alerts Correlation[C]// International Symposium on Computer Science and Computational Technology, 2008 (ISCSCT'08). IEEE, 2008: 124-127.

[13] YAN R Y. DDoS Attacks Detection Method Based on Traffic Matrix and KalmanFilter[J]. Computer Science, 2014, 41(3): 176-180. (in Chinese)
颜若愚. 基于流量矩阵和 Kalman 滤波的 DDoS 攻击检测方法[J]. 计算机科学，2014，41(3)：176-180.

[14] VALDES A D J, SKINNER K. Probabilistic alert correlation: Springer Berlin Heidelberg, US 7917393 B2[P]. 2011.

[15] AHMADINEJAD S H, JALILI S. Alert Correlation Using Correlation Probability Estimation and Time Windows[C]// International Conference on Computer Technology and Development. IEEE, 2009: 170-175.

[16] VAPNIK V. SVM method of estimating density, conditional probability, and conditional density[C]// The 2000 IEEE International Symposium on Circuits and Systems, 2000. IEEE, 2000: 749-752.

[17] PARSI S K. Implementing network intrusion detection on a multi-threading FSM[D]. Dissertations & Theses-Gradworks, 2007.

[18] MASTANI S A. Reduced Merge_FSM Pattern Matching Algorithm for Network Intrusion Detection[J]. International Journal on Recent Trends in Engineering & Technolo, 2014, 10(2): 117-122.

[19] ILGUN K, KEMMERER R A, PORAS P A. State transition analysis: a rule-based intrusion detection approach[J]. IEEE Transactions on Software Engineering, 1995, 21(3): 181-199.

[20] YANG Y, MCLAUGHLIN K, LITTLER T, et al. Rule-based intrusion detection system for SCADA networks[C]// Renewable Power Generation Conference (RPG 2013), 2nd IET. IET, 2013: 1-4.

[21] PERERA G. Rules Based *Monitoring and ntrusion Detection* System: US20150326604[P]. 2015.

[22] EILAND E E, EVANS S C, MARKHAM T S, et al. Intrusion detection using MDL compression: US, US8375446B2[P]. 2013.

[23] ESMAILI M, BALACHANDRAN B, SAFAVI-NAINI R, et al. Case-Based Reasoning for Intrusion Detection[C]// Proceedings of the 12th Annual Computer Security Applications Conference. IEEE Computer Society, 1996: 214-223.

[24] LONG J, SCHWARTZ D, STOECKLIN S. Application of Case-

Based Reasoning to MultiSensor Network Intrusion Detection [C]// Proceedings of the 4th WSEAS international conference on Computational intelligence, man-machine systems and cybernetics. World Scientific and Engineering Academy and Society (WSEAS). 2005.

[25] ZENG R G, GUAN X H, ZAN X, et al. Case-Based Reasoning for Intrusion Detection Correlation Analysis[J]. Computer Engineering & Applications, 2006, 42(4): 138-141.

[26] CHEN B, LING Y U, XIAO J M. An Application of Simulated Annealing Algorithm in Model-Based Reasoning Intrusion Detection[J]. Journal of University of Electronic Science & Technology of China, 2005, 34(1): 36-39.

[27] CHEN R C, CHEN S P. An intrusion detection based on support vector machines with a voting weight schema[C]// International Conference on Industrial. Springer-Verlag, 2007: 1148-1157.

[28] TRAN T P, TSAI P, JAN T, et al. Network Intrusion Detection using Machine Learning and Voting techniques[M]. Machine Learning, 2010: 267-290.

[29] BOROWIK B, KARPINSKYY M, LAHNO V, et al. Machines Moore and Mealy[M]// Theory of Digital Automata. Springer Netherlands, 2013: 143-171.

[30] RUBIN D E, MITAL V, BECKMAN B C, et al. Dependency graph in data-driven model: US, US8352397[P]. 2013.

[31] GUMUS F, SAKAR C O, EREDM Z, et al. Online Naive Bayes classification for network intrusion detection[C]// 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2014: 670-674.

[32] VARUNA S, NATESAN P. An integration of k-means clustering and naïve bayes classifier for Intrusion Detection[C]// 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN). IEEE, 2015: 1-5.

[33] FARID D M, HARBI N, RAHMAN M Z. Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection[J]. International Journal of Network Security & Its Applications, 2010, 2(2): 52-58.

[34] JIA I A, YANG C F, et al. An Intrusion Detection Method Based on Hierarchical Hidden Markov Models[J]. Wuhan University Journal of Natural Sciences, 2007, 12(1): 135-138.

[35] VOLLMER T, MANIC M. Title: Computationally Efficient Neural Network Intrusion Security Awareness[C]// 2nd International Symposium on Resilient Control Systems, 2009(ISRCS'09). IEEE, 2009: 25-30.

[36] MACKENZIE M R, TIEU A K. Hermite neural network correlation and application[J]. IEEE Transactions on Signal Processing, 2004, 51(12): 3210-3219.

[37] GILMORE M R, JONES S E, FOSTER J C, et al. Sung Intrusion Detection: Support Vector Machine and Neural Networks [C]// ASME 2002 Pressure Vessels and Piping Conference. American Society of Mechanical Engineers, 2002: 277-281.

[38] RAO X, DONG C X, YANG S Q. An Intrusion Detection System Based on Support Vector Machine[J]. Journal of Software, 2003, 14(4): 798-803. (in Chinese)

饶鲜，董春曦，杨绍全. 基于支持向量机的入侵检测系统[J]. 软件学报，2003，14(4)：798-803.

[39] YANG K H, SHAN G L, ZHAO L L. Correlation Coefficient Method for Support Vector Machine Input Samples[C]// 2006 International Conference on Machine Learning and Cybernetics. IEEE, 2006: 2857-2861.

[40] KLIGER S, YEMINI S, YEMINI Y, et al. A coding approach to event correlation[C]// Proceedings of the Fourth International Symposium on Integrated Network Management IV. Chapman & Hall, Ltd., 1995: 266-277.

[41] GRUSCHKE B. Integrated Event Management: Event Correlation Using Dependency Graphs[C]// Distributed Systems, Operations and Management. 1998.

[42] ROSCHKE S, CHENG F, MEINEL C. A New Alert Correlation Algorithm Based on Attack Graph[M]// Computational Intelligence in Security for Information Systems. Springer Berlin Heidelberg, 2011: 58-67.

[43] ZHU B, GHORBANI A A. Alert Correlation for Extracting Attack Strategies[J]. International Journal of Network Security, 2006, 3(3): 244-258.

[44] STEINDER, MAŁGORZATA, SETHI, et al. Probabilistic Fault Localization in Communication Systems Using Belief Networks [C]// IEEE/ACM Transactions on Networking. 2004: 809-822.

[45] MARCHETTI M, COLAJANNI M, MANGANIELLO F. Identification of correlated network intrusion alerts [M] // 2011 Third International Workshop on Cyberspace Safety and Security (CSS). IEEE, 2011: 15-20.

[46] HARAHAP E, SAKAMOTO W, NISHI H. Failure prediction method for Network Management System by using Bayesian network and shared database[C]// 2010 8th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT). IEEE, 2010: 1-6.

[47] SHI Z, XIA Y. A Novel Hidden Markov Model for Detecting Complicate Network Attacks [C] // 2010 IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS). IEEE, 2010: 312-315.

[48] KELLOGG J, MCNEELY A, RUFFO B, et al. Alert Correlation and Prediction Using Data Mining and HMM[J]. Isecure, 2011, 3: 77-102.

[49] ZAN X, GAO F, HAN J, et al. A Hidden Markov Model Based Framework for Tracking and Predicting of Attack Intention[C]// International Conference on Multimedia Information Networking and Security. IEEE, 2009: 498-501.

[50] Swatchwebsite[OL]. http://sourceforge.net/projects/swa-tch.

[51] SEC-simple event correlator[OL]. http://kodu.neti.ee/~risto/sec.

[52] OSSEC community. Ossec website[OL]. http://ossec.net.

[53] OSSIM community. Ossim website [OL]. http://www.ossim.org.

[54] Drools community. Drools website [OL]. http://www.jboss.org/drools.

[55] EsperTech. Esper website[OL]. http://www.espertech.com.