事件关联分析算法及其实现技术

李思广 赵振然

（周口职业技术学院 周口 河南 466000）

摘 要: 本文分析了事件关联的方法，探讨了几种事件关联算法，提出了事件关联实现技术。

关键词: 入侵事件; 事件关联; 算法; 实现技术

0 引言

入侵检测的根本任务就是对提取到的庞大数据进行分析，以期从中找到入侵的痕迹。通过对大量黑客的一系列入侵行为进行分析，我们知道，黑客入侵行为并不是独立的，而都是有所关联的。从攻击的角度出发，事件之间的关联是指它们是否是同一个攻击行为所产生的，这种攻击行为包括单个简单攻击行为或由一系列攻击步骤组成的复杂攻击行为。事件关联技术是将各种数据来源进行综合分析，把分散的异常事件痕迹关联起来，最终给出完整的事件描述。

1 入侵事件描述

从安全防护观点来看，入侵事件是指能够引起入侵行为的发生，或具有潜在入侵行为的一切活动、变化和事情。多个入侵事件即组成入侵过程。网络安全事件通常有两种类型：基本事件和组合事件。其中，基本事件被网管系统事先定义，其检测机制通常被嵌入到系统实现中。组合事件由基本事件或其它组合事件构成，每一个构成组合事件的事件叫做元素事件。一个基本事件可以包含多个属性，因此，每一个基本事件可用一个包含这些属性值的多元组表示。

基本事件：可用一个三元组表示为 $E(name,msg,time)$ 其中，name 表示事件名，msg是事件的基本内容，time指事件产生的时间。

组合事件：由在给定的时隙内发生的一组满足一定的时序或逻辑关系的事件组成，可表示为 $E = f(E_1,E_2,E_3,\cdots)$ 其中， E_i 是事件E的一个元素事件，是基本事件， f 代表它们之间的逻辑关系。

2 事件关联的方法

2.1 冗余关系关联法[1]

对于冗余关系所采用的事件关联分析方法主要是依据事件攻击特征(Attack_Specif)中相关属性之间的相似度。这里，相似度采用概率统计的方法来计算。选取Attack_Specif中的4个有意义的属性加上Attack_Id字段属性构成所谓相似属性集:

S=(S d_id，S s_ip，S d_ip，S d_port，S a_type)

其中S i依次分别对应:数据采集器标识符detect_id、源IP地址source_ip、目的IP地址dest_ip、目的端口号dest_port、攻击类型Attack_Id。假设X和Y代表两个事件，它们的相似属性集分别为(X d_id，X s_ip，X d_ip，X d_port，X a_type)，(Y d_id，Y s_ip，Y d_ip，Y d_port，Y a_type)，则X和Y的相似度

$$SIM(X,Y)=\frac{\sum_{i=1}^{S}W_i SIM(X_i,Y_i)}{\sum_{i=1}^{S}W_i}$$

式中，SIM(X_i，Y_i)是对应属性间的相似度，W_i是对应的期望权值，即各个具体属性相似度在整体相似度中的所占比重值。通常情况下需要设定一个阈值β，如果SIM(X,Y)> β，就表示事件X和Y是冗余关系。对单个属性间的相似度计算，分别采用相应的规则。

2.2 因果关系关联法[1]

判断事件间因果关系所采用的事件关联分析方法主要基于攻击事件模型E的三个字段:Attack_Precond、Attack_Postcond、Attack_Specif。基本思想是:寻找一个攻击事件的前因(Attack_Precond)和另一个攻击事件的后果(Atack_Postcond)之间是否存在逻辑联系，如果存在联系，就表明这两个攻击事件是关联的。

一般来说，先进行冗余关系的分析，将重复的多个事件聚合为一个事件，再进行因果关系的分析，这样做可以减少不必要的重复计算。

3 事件关联算法

事件关联算法根据其对先验知识的依赖程度可以划分为两类:有指导关联算法和无指导关联算法。

3.1 有指导关联算法

有指导的关联算法[2-3]指的是关联算法在先验知识的帮助下，完成事件关联的工作。先验知识的不同表示方法导致了不同的关联算法。

（1）基于攻击序列模板的事件关联方法

基于攻击序列模板的事件关联方法[2]是最早用于报警事件的关联行为研究的一种方法，关联过程中所需要的先验知识表现为攻击序列模板，其形式一般为：$E = e_1 \text{ op } e_2 \text{ op} \cdots e_i \text{ op} \cdots e_n \text{ op}$。其中$e_i$为报警事件，op代表为了说明$e_i$和$e_j$之间关系的不同的运算符，在整个事件序列E中，隐含着不同报警事件发生的一个时序关系，即e_{i+1}是e_i的后继。不同的文献中，为了有效表示攻击序列模板，采用了不同的形式，攻击序列模板形式的不同也导致了不同的关联算法。

（2）基于单个攻击发生的前提和后果的关联方法

基于单个攻击发生的前提条件和后果的事件关联方法[3-4]强于基于攻击序列模板的事件关联方法，其安全知识一般表示为三元组:(Attack, Prerequisites, Consequences)，其中，Attack代表攻击动作名，Prerequisites代表攻击发生的前提条件而Consequences代表攻击发生后给整个系统所造成的影响。其关联算法的总体思想就是用攻击att_i发生后的后续结果和攻击att_j发生的前提条件去进行匹配，如果能够全部或部分匹配，则表明攻击att_i和att_j是具有因果联系的，从而可完成两者之间的关联工作。基于单个攻击发生的前提条件和后果的事件关联方法是现在研究最多的一种关联方法。

3.2 无指导关联算法

无指导的关联算法[5-6]在整个事件关联的过程中，不需要任何先验知识的帮助。文献[5-6]中所发表的成果总体上可以划分为两类:基于近似度函数的关联算法和基于统计分析的关联算法。

（1）基于近似度函数的事件关联方法

基于近似度函数的关联算法[5-6]主要通过把报警事件定义为一个实体，形成描述单个报警事件内容的向量，然后通过定义的计算事件e_1和与关联队列中的e_2之间相似性函数来计算他们之间的近似度，如果当前发生的报警事件其与已发生的报警事件之间的近似度大于预定义的阈值，则其与近似度比较大的事件实体完

成关联，否则，则创建新的关联队列来容纳事件e_1，并将其做为该队列的首事件。

文献[5]中作者提出的计算报警事件之间的关联度的计算公式，如下所示:

$$SIM(X,Y)=\frac{\sum_j S_j SIM(X_j,Y_j)}{\sum_j S_j}$$

其中$SIM(X_j,Y_j)$用来计算属性X_j 、Y_j之间的相似性。由于报警事件的属性的类型的多样性，难于找到一个统一的方法来计算报警事件E_a的属性X_j和报警事件E_b的X_j之间的关联度，为此，作者根据不同的属性类型，定义了不同的计算方法，包括定义相似性矩阵，距离函数等，通过计算SIM(X,Y)来完成报警事件X、Y之间的关联工作。

（2）基于统计分析的关联算法

基于统计分析的关联算法主要通过已经收到的历史报警信息来建立关于报警事件的预测模型，然后，通过训练出的预测模型去计算其与哪个正处于关联过程中的攻击序列最接近，从而完成整个事件关联的工作。

文献[6]中提出一种基于时间序列分析的关联方法，该方法引入了时间序列分析的预测方法。整个算法过程如下:

首先定义时间间隔T，然后把该时间间隔划分成N份，并利用聚类方法把发生在时间段i内的报警事件聚合成事件A_i，从而产生报警事件集合$\{A_1,\cdots,A_i,\cdots,A_n\}$，文中，把事件$A_i(1\le i\le n)$定义为时间序列分析中的时间序列变量，然后引入AR模型:

$$y(k)=\sum_{i=1}^{p}\theta_i y(k-i)+e_0(k)$$

和 AR 模型:

$$y(k)=\sum_{i=1}^{p}\theta_i y(k-i)+\sum_{i=1}^{p}\beta_i x(k-i)+e_0(k)$$

并利用公式 $g=\frac{(R_0-R_1)/p}{R_1/(T-2p-1)}-F(p,T-2p-1)$（其中，$R_0=\sum_{k=1}^{r}e_0^2(k)$，$R_1=\sum_{k=1}^{r}e_1^2(k)$）计算新发生的报警事件所对应事件序列变量y(k)和最近发生的报

警事件所对象的时间序列变量x(k)之间的g值，如果y(k)与x(k)之间的g值最大，则y(k)所对应的报警事件与x(k)所对应的报警事件具有最大的关联可能性，从而完成事件关联的整个工作。

4 事件关联实现技术

4.1 基于规则的推理

基于规则的推理(RBR, Rule-Based Reasoning)又称为基于规则的专家系统、产生式系统和黑板系统等，它是最早出现的一种事件关联技术。

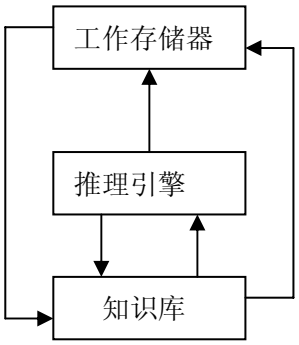

图4-1 RBR系统工作原理

如图4-1所示，基于规则的推理系统一般由一个工作存储器(Working Memory)、一个推理引擎(Inference Engine)和一个知识库(Knowledge Base)组成。3个组成部分所处的3个层次分别代表了数据层、控制层和知识层。工作存储器通过具体的网络管理协议如SNMP、CMIP，收集被监视网络的各种信息，其中包括网络拓扑信息和被监视网元的状态信息。当网络中发生故障时，工作存储器通过分析这些信息识别出网络进入错误状态。知识库中包含从人类领域专家那里得到的专家知识。知识库有两个功能：（１）尽可能地确定网络中到底发生了什么问题；（２）当某一特定问题发生时，指出系统所要执行的动作。知识库中的专家知识是基于规则的，即所有知识都采用"if-then"或者“condition-action"规则集的形式。与知识库合作的推理引擎将日前网络的状态与知识库中规则的条件部分进行比较，以决定该规则是否被采用。当条件满足时，输出规则的后项。在最简单的情况下，一条规则就可以判断出网络故障的根源。实际上这种情况十分少见，多数时候推理引擎需要将当前得到的结论作为条件在知识库中进行多次推理，最终得出故障结论。

4.2 基于事例的推理

基于事例推理(CBR, Case-Based Reasoning)的思想源于现实生活。在现实生活中，一些情形总是重复发生，处理某一特定情形的方法在类似的情形中也能适用，而这些类似的情形并非要与该特定情形完全一致。因此，我们当试图解决一个问题时，都是从曾经经历过的类似事例出发的。

CRB技术就是根据这一思想将过去成功的事例存入事例库，遇到新问题时，在事例库中寻找类似的过去事例，利用类比推理方法得到新问题的近似解答；再加以适当修改，使之完全适合新问题。

CRB体系结构如图4-2所示[7]。它由5个部分组成，其中包括一个事例库和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对问题的描述。接着由检索模块到事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，检索模块就在事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就可满足当前问题的要求，其结果是得到一个完整的解。一旦问题被解决，则处理模块将新的事例加入到事例库中，以备以后的应用。建造和扩充事例库需定义词汇，从领域专家那里收集事例并将其装入事例库。

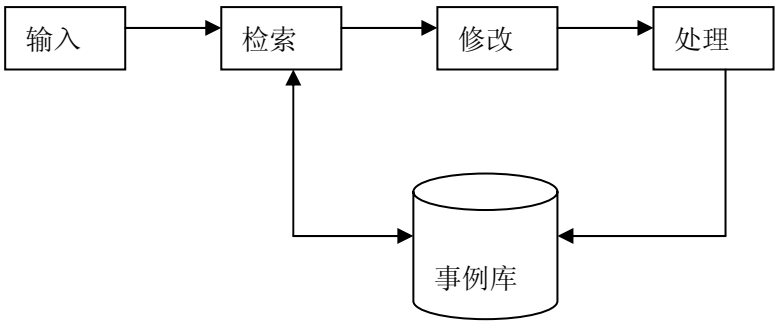

图4-2 CRB体系结构图

值得注意的是在系统刚开始运行的时候，一般很少能找到完全匹配的事例，但随着事例库的增长，这种巧合会越来越多，因此系统的效率也会逐渐提高。另外，由于新问题的解可以自动加入到事例库中去，以后如果再遇到同样的问题，系统就不会重复上述步骤，而是直接得到一个完全匹配的解，因此基于事例推理的系统具有自学习的能力。

4.3 基于模型的推理

在基于模型的推理(Model-Based Reasoning)系统中，每个被管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处于网络管理系统(NMS)中的事件相关器建立在面向对象的模型之上。作为它所模仿的网元的代表，面向对象的模型有自身属性、该模型与其他模型之间的关系以及该模型的行为等要素。

在基于模型的推理系统中，事件关联是模型之间协作的结果。网络管理系统和被管网元之间的通信是通过NMS中的事件关联器和每个被管网元的模型之间的通信实现的，被管网元和被管网元之间的通信是通过被管网元的模型之间的通信实现的。这样，模型之间的关系反映出它们所代表的被管网元之间的关系。

每个模型通过与自身所表示的被管网元以及与其它模型之间进行通信，分析白身所表示的网元是否发生故障。因此，网元的故障首先由模拟该网元的模型识别出，然后报告给网络管理系统。例如，在实际网络中NMS一般都会周期性的对某一路由器发出Ping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的对路由器发出Ping命令来实现的。

5 结束语

上面提出的基于规则的推理虽然优点明显，但也存在以下缺陷:一是规则必须完全匹配，但是预先为一个大型网络确定所有的故障规则是非常困难甚至是不可能的；二是规则不易维护，由于网络变更频繁，需要不断添加、修改和删除规则；三是事件发生时，每次都必须执行有关的规则，所以基于规则的系统比较慢，性能比较低。而基于事例的推理则克服了RBR的许多不足:在RBR中，知识的单元是规则，而在CBR中知识的单元是事例；在RBR中，检索是基于对规则的完全匹配，而在CBR中，检索是基于对事例的部分匹配； RBR适用干网络规模不大时微小事件的重复出现，而CBR适用于大规模网络复杂问题的整体解决。在基于模型的推理系统中，事件关联是模型之间协作的结果，其效率是非常高的。

参考文献

[1]陈晓苏 尹宏斌等.入侵检测中的事件关联分析[J].华中科技大学学报，2003，31（4）：31-32.

[2]Valdes A，Skinner K.Probabilistic alert correlation[C] //Fourth

International Workshop on the Recent Advances in Intrusion Detection (RAID’2001), Davis, USA, October2001.

[3]Cuppens and Miege 2002 CUPPENS, F. AND MIEGE, A. 2002. Alert correlation in a cooperative intrusion detection framework[A]. In:Proceedings of the 2002 IEEE Symposium on Security and Privacy[C],2002.

[4]KRUGEL C, TOTH T, KERER C. Decentralized Event Correlation for Intrusion Detection[Z].4th International Conference on Information Security and Cryptology(ICISC), 2001.

[5]NING P, CUI Y, REEVES DS. Analyzing intensive intrusion alerts via correlation[A].In: Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection[C].Zurich, Switzerland, 2002.

[6]YU D, FRINCKE D. A Novel Framework for Alert Correlation and Understanding(Springer-Verlag)[A].International Conference on Applied Cryptography and Network Security(ACNS)[C], 2004.452-466.

[7]彭熙，李艳等.事件关联策略的实现及其应用研究[J].计算机工程与设计, 2003,24(10):17-18.