DOI:10.13245/j.hust.2003.04.011

第31卷 第4期 华 中 科 技 大 学 学 报(自然科学版) Vol.31 No.4
2003年 4月 J. Huazhong Univ. of Sci. & Tech.(Nature Science Edition) Apr. 2003

入侵检测中的事件关联分析

陈晓苏 尹宏斌 肖道举

(华中科技大学计算机科学与技术学院)

摘要:大部分的攻击事件都不是孤立产生的,相互之间存在着某种联系,而这种联系可以抽象为冗余关系和因果关系.当前的大多数入侵检测系统忽略了这种事件之间的关联性,从而暴露出一些问题.针对这些问题,结合这两种事件关系的基本特征,给出了相应的事件关联分析方法,并在此基础上给出了一个事件关联分析器的体系结构设计.

关 键 词:入侵检测;事件关联;原始事件

中图分类号:TP393 文献标识码:A 文章编号:1671-4512(2003)04-0030-04

随着网络结构的日趋复杂,网络规模的日趋庞大,以及入侵手段的日趋多样化,入侵检测系统IDS(Intrusion Detection System)在不断提高检测能力,扩大检测范围的同时,也暴露出了一些问题:**a**. 事件风暴;**b**. 虚警率高;**c**. 上下文关系不明确.

产生上述问题的原因在于当前大部分入侵检测系统仅仅只关注系统所检测到的原始事件,而忽略了隐藏在这些原始事件背后的逻辑联系和攻击意图.如果这种相关的分析工作全部由系统管理员人工来完成,工作量巨大且效率低下,其结果是不仅严重浪费系统管理员的时间和精力,而且无法实现系统的快速判断和响应.显然,在入侵检测系统中引入自动实时的事件关联分析功能具有很强的实际意义.基于这一考虑,本文给出了入侵检测中事件关联分析的方法以及相应事件关联分析器的体系结构.

1 事件关联的相关概念

1.1 引入事件关联的IDS系统组成

在入侵检测系统中引入事件关联分析,实质上是在入侵检测引擎和系统管理界面之间增加一个起到事件再加工作用的事件关联层——事件关联分析器[1],系统组成如图1所示.

事件关联分析器根据给定的关联分析方法对入侵检测引擎检测到的原始事件进行关联分析,并将加工处理后的事件传递给系统管理界面显示出来.采用这种处理模式,一方面可减少系统管理员看到的事件数量,从而降低其工作强度;另一方面,通过提供高度综合、更加全面的事件报告,有利于系统管理员的分析判断.

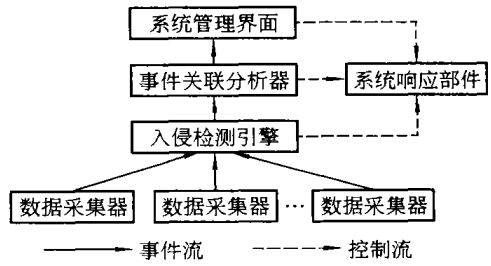

图1 引入事件关联的IDS系统组成

值得强调的是,尽管图1给出的是一个集中式的事件关联结构,但也可以根据具体的环境和要求,采用分布式的体系来部署.

1.2 事件关联性的定义及分类

从攻击的角度出发,事件之间的关联性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和由一系列攻击步骤组成的复杂攻击行为.

根据上述事件关联性的定义,事件之间的关系可以进一步抽象为两种关系:

a. 冗余关系.往往是由一个简单攻击或一个复杂攻击的某个步骤触发多个入侵检测引擎或多次触发一个入侵检测引擎所引起.以最常见的端口扫描行为为例,通常,一次扫描行为会同时针对目标机的多个端口,这样同一个行为可能产生多个警报.同样道理,若一次扫描行为是针对不同子

收稿日期:2002-09-20.

作者简介:陈晓苏(1953-),男,教授,武汉,华中科技大学计算机科学与技术学院(430074).

网的不同主机，则会触发各个子网的 IDS 部件产生警报，等等.将这样一类事件间的联系称之为冗余关系.

b. 因果关系.黑客的攻击行为总是沿一定的步骤和路径来进行的.因此，在一个完整的攻击过程中，由单个攻击步骤触发的警报事件可能存在其前因事件和可能带来的后果事件.例如，攻击者可能会首先进行漏洞扫描，发现有漏洞的主机后，根据获取的漏洞信息，进一步对目标主机进行渗透攻击，在取得主机的控制权后，再去攻击其他主机.概念上，这些事件都属于同一个复杂攻击，它们之间存在某种因果关系.

冗余关系相对来说比较明显，易判断、易处理，可以采用聚合方法将属于冗余关系的事件过滤掉，从而有效减少原始事件的数量；事件之间的因果关系显得更隐蔽，揭示事件之间的因果关系有助于发现贯穿于整个安全系统的攻击模式和入侵趋势，预见下一步将面临的威胁，提前阻止攻击的发生，将系统的安全防御从被动式的“滞后型”转为主动式的“抢先型”.

2　事件描述

2.1　事件建模

要做到有效的事件关联分析，先要对可能存在的攻击事件作清楚、明确的描述，在此基础上，才有可能对各个警报事件进行定位、归类和关联分析.

针对因果关系所具有的隐蔽性特征，从全面揭示事件间复杂的因果关系这一角度出发，本文将攻击事件抽象成一个七元组来描述.

设攻击事件模型用 E 表示，则有

$$E = (\text{Attack\_Id}, \text{Attack\_Name}, \text{Attack\_Precond}, \text{Attack\_Postcond}, \text{Attack\_Specif}, \text{Attack\_Time}, \text{Attack\_Respose}),$$

其中，字段 Attack_Id 为攻击标志，表明攻击类型；Attack_Name 为攻击名称；Attack_Precond 为攻击前提，即攻击实施前所应该满足的条件集合；Attack_Postcond 为攻击后果，即攻击实施后对安全系统造成的所有可能影响的集合；Attack_Specif 为攻击特征，即攻击报文的特征描述；Attack_Time 为检测到攻击所发生时的时间；Attack_Respose 为攻击响应，针对某个攻击，系统所应采取的相应对策.

由于 Attack_Specif 字段的数据来源主要是侦听到的网络攻击报文，因而可以进一步用一个六元组将其特征化.

设一个报文用 P 表示，则有

$$P = (\text{detect\_id}, \text{source\_ip}, \text{dest\_ip}, \text{source\_port}, \text{dest\_port}, \text{payload}),$$

其中，detect_id 表示数据采集器的标识符；source_ip，dest_ip 分别表示源、目的 IP 地址；source_port，dest_port 分别表示源、目的端口号；payload 表示有效载荷.

本文将着重讨论“Attack_Precond”，“Attack_Postcond”和“Attack_Specif”字段，并以此作为事件关联分析的依据.

在实际中，可以考虑用 XML 语言来描述上述这些元组.作为一种表示和交换网络文档及数据的语言，XML 获得了广泛的支持和应用，因此用 XML 语言来描述事件模型可获得较好的灵活性、通用性以及高效率.

2.2　形式化描述

由于 Attack_Precond 和 Attack_Postcond 字段实质上描述了整个安全系统的状态，因而很难用类似于 N 元组这类描述方法来将其特征化.考虑到形式化描述方法具有简单且易于系统具体实现等特征，因而采用某种形式化描述方法来反映系统的安全状态是一条较好的途径.

为便于问题描述，这里设“状态”是用形式化描述语言表达的语义关系的文字说明.同时为了便于计算机处理，以多元逻辑断言(Predicate)作为形式化描述的基础，并相应地进行扩展.例如，二元逻辑断言 IP_Service(ip, service)表示具有某 ip 地址的主机上运行着的 service 服务；IP_OS(ip, os)表示具有某 ip 地址的主机上运行着的 os 操作系统；Attack_DOS(IP)表示具有相应 IP 地址的主机受到了拒绝服务攻击.这些逻辑断言经过适当设计，很容易从字面上理解它们的含义.为了表示更复杂的状态，用逻辑关系符“∧”(与)、“¬”(非)来连接多个逻辑断言.例如，Winnuke 攻击需要两个必要前提：目标主机运行 windows 操作系统，同时开启有 dns 服务.那么用逻辑断言描述这种情况有：

$$\{\text{IP\_OS}(\text{target\_ip}, \text{windows}) \wedge \text{IP\_Service}(\text{target\_ip}, \text{netbios-ssn})\}.$$

考虑到有一些攻击仅仅只是为了获得目标系统的相关信息，比如扫描攻击等.对于这类攻击，采用 Get_Info() 来表示其攻击后果.例如：Get_

Info(IP–Service(ip, service))表明获取了具有某ip地址的主机正运行着service服务的信息.

3 事件关联方法

由于事件之间的关系可以抽象为冗余关系和因果关系两类,因而所采用的事件关联分析方法也就存在差异.

3.1 冗余关系关联分析方法

对于冗余关系所采用的事件关系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[2]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

$$SIM(X, Y)=\sum_{i=1}^{5} W_i SIM(X_i, Y_i)\Big\backslash\sum_{i=1}^{5} W_i,$$

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$SIM(X, Y)\geqslant\beta$,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.例如,对源IP地址的相似度计算,可以设定:

$$SIM(X_{s\_ip}, Y_{s\_ip})=\begin{cases}0 & \text{(完全不相似);}\\ 0.8 & \text{(同一子网);}\\ 1 & \text{(相同IP).}\end{cases}$$

3.2 因果关系关联分析方法

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(Attack–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.这种分析思想的形式化描述为:

设A和B分别代表两个攻击事件,Attack–postcond(A)和Attack–precond(B)分别表示A和B两个攻击事件的攻击后果和攻击前提,predicate表示逻辑断言.

$$\text{Attack–Postcond}(A)=\exp_{A1}\wedge\exp_{A2}\wedge\cdots\wedge\exp_{Am},$$

$$\text{Attack–Precond}(B)=\exp_{B1}\wedge\exp_{B2}\wedge\cdots\wedge\exp_{Bn},$$

其中$\exp_i$具有下列三种形式之一:

$\exp_i=$predicate, $\exp_i=\neg$predicate,

$\exp_i=$Get–info(predicate$|\neg$predicate).

若存在$i\in[1, m]$、$j\in[1, n]$,当满足一定的关联条件$\oint$时有$\exp_i=\exp_j$,或者其中一个以Get–Info( )的形式包含另一个,则一个攻击A与另一个攻击B关联,于是有

$$\text{Attack–Postcond}(A)\wedge\oint\rightarrow\text{Attack–Pre cond}(B).$$

假设事件A为漏洞扫描攻击事件,事件B为Winnuke攻击事件,则有

Attack–Postcond(A)=
Get–Info(IP–OS(ip1, os))∧
Get–Info(IP–Service(ip1, service)),
Attack–Precond(B)=IP–OS(ip2, os)∧
IP–Service(ip2, service).

关联条件$\oint$为:ip1=ip2; os=windows; service=netbios-ssn.

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.

4 事件关联分析器体系结构设计

基于以上所提出的事件描述和关联分析方法,这里就图1中的事件关联分析器给出其详细的体系结构设计,如图2所示.

关联预处理模块:该模块用来完成攻击事件表示格式的标准化处理.它接受各个检测器传来的原始警报事件,然后根据关联特征库产生相应的标准XML攻击描述文件.

XML解析模块:该模块对关联预处理模块产生的XML攻击描述文件进行解析,提取出其中的关键字段信息.

特征提取模块:该模块对提取出的Attack–

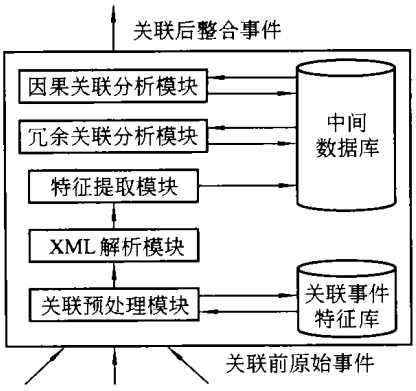

图 2　事件关联分析器的体系结构

Precond、Attack–Postcond 和 Attack–Specif 等关键字段做进一步的特征化处理，主要包括形式化描述的生成和攻击特征属性的提取.

冗余关联分析模块：该模块根据前面几步得到的数据，通过给定的冗余关系关联分析方法，过滤掉属于冗余关系的重复事件，并将关联后的聚合事件提交给上层因果关联分析模块.

因果关联分析模块：该模块根据冗余关联分析模块所提交的事件集，通过给定的因果关系关联分析方法，揭示原始事件之间的因果联系，并将关联后的整合事件提交给系统管理界面.

关联事件特征库：用于存放各种特定事件的前因后果，以支持关联预处理模块构建 XML 攻击事件描述文件.

中间数据库：考虑到 XML 文件的特征提取数据量可能会比较大，将分析结果存在中间数据库中，供关联模块读取，从而起到数据缓冲的作用.

在网络结构的日趋复杂，网络规模的日趋庞大以及入侵技术的日趋成熟的今天，采用事件关联技术，可以有效地减少原始攻击事件处理数量，降低虚警率，突出事件间上下文关系.

参　考　文　献

[1] Poirk Y. Event Correlation. IEEE Potentials, 2001, 20(2): 34~35

[2] Ye Nong, Li Xiaoyang, Chen Qiang, et al. Probabilistic techniques for intrusion detection based on computes audit data. IEEE Transactions on System, Man, and Cybernetics, 2001, 31(4): 266~274

The analysis of event correlation in intrusion detection

Chen Xiaosu　Yin Hongbin　Xiao Daoju

Abstract: The method for the analysis of an event correlation was introduced based on the characteristics of the two kinds of relationships, that is, redundancy relationship and cause and effect relationship. Based on that, the architecture designed for event correlation analysis apparatus was presented. Practice shows that event correlation can decrease number of alert, reduce false alert and discover high-level attack strategies effectively.

Key words: intrusion detection; event correlation; raw event

Chen Xiaosu　Prof.; College of Computer Sci. & Tech., Huazhong Univ. of Sci. & Tech., Wuhan 430074, China.

我校去年获 2 亿元横向科研经费

近年来，我校面向国民经济建设主战场，充分利用自身在科技、教育、卫生领域的优势，为地方经济建设服务，与全国各地方政府或企业广泛开展科技合作. 仅 2002 年，学校新增项目 683 项，其中百万元以上的项目就达 29 项，“赚回”了超过 2 亿元的横向科研经费.