

文章编号:1001-9081(2005)07-1526-03

安全关联分析相关技术的研究

高 雷^{1,2}, 肖 政^{1,2}, 韦 卫^{1,3}, 孙育宁^{1,3}

(1. 中国科学院 计算技术研究所, 北京 100080; 2. 中国科学院 研究生院, 北京 100039;

3. 联想研究院, 北京 100085)

(gaoleia@lenovo.com)

摘 要:着重研究网络安全集中管理系统中的关联分析技术,对其通用体系结构及其关键分析技术(产生式关联、即时关联等)、研究趋势(模式抽取、部署架构等)进行了探讨,并提出了基于层级式规则的关联分析解决方案。

关键词:安全关联分析架构;产生式关联;即时关联;引擎部署;模式抽取;层级式规则

中图分类号: TP393.07 **文献标识码:** A

Research on the techniques of security events correlation

GAO Lei^{1,2}, XIAO Zheng^{1,2}, WEI Wei^{1,3}, SUN Yun-ning^{1,3}

(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China;

2. Graduate School, Chinese Academy of Sciences, Beijing 100039, China;

3. Lenovo Corporation of Research, Beijing 100085, China)

Abstract: The events correlation techniques in security integration management systems were introduced. A normal architecture of the correlation engine was introduced, and some discussions on the critical technologies and the main achievements in the field were put forward. The directions of the technology development were analyzed and evaluated, such as pattern obtainment, engine distribution and performance promotion. At last, a solution based on hierarchical rules to correlate events was presented.

Key words: architecture of the security events correlation; causal correlation; temporal correlation; engine distribution; pattern abstraction; hierarchical rules

0 引言

常用的 IDS, FireWall, 各种异常检测器等安全工具各从特定的角度满足了一部分的安全需求,但远远未达到网络管理员对网络整体安全现状进行简单、直观、全面掌握的要求。一方面由于单个工具的局限,常因为不能识别正常行为而引发误报;另一方面,单个攻击行为引发多个重复告警时,给管理员做出正确判断带来困难;更重要的是,网络攻击行为日趋复杂化、分布化,一个攻击过程由多个攻击步骤构成,多个步骤又完全可能在不同的地方实施,依靠单个的事件日志,太过琐碎、无法反映整个攻击行为的全貌,因而也就无法捕捉到那些有计划、有步骤的复杂攻击行为。

最理想的方案是将 IDS, FireWall, 异常检测工具等全部集成于同一系统中,并能对不同类型的事件日志进行全局分析,从而达到对网络安全状况(包括当前遭受的攻击、网络流量、访问路径甚至会话特征^[1]等等)的全面监控。但这样内存、磁盘容量需求都将十分庞大,另外对处理性能也是一个挑战,如果达不到实时或近于实时要求,那么安全分析毫无意义。

一个变通的方案是保持 IDS, FireWall, 异常检测工具在网络节点中的部署不变,而采用一个中心节点集中接收这些节点的安全事件,并对这些信息作关联分析处理,以减少误

报、避免重复报警、增加攻击检测率。这种解决方案目前已被一些研究组织和商业公司所采用。

1 关联分析技术研究发展现状

1.1 关联分析的概念

本文所指的关联分析即对分布式节点(IDS, FireWall 和 Anomaly Detectors)提供的安全事件进行综合分析,以方便管理人员全面监控网络安全状况的技术。

这一技术通过关联来自于不同地点、不同层次、不同类型的安全事件,从而发现真正的安全风险,达到对当前安全态势的准确、实时评估,并根据预先制定策略作出快速的响应。

这里,关联分析要解决的问题有:

- 1) 联系可能的安全场景分析单个告警事件,以避免虚警;
- 2) 对相同、相近的告警事件作处理,以避免重复告警;
- 3) 提高分析的实时性,以利于及时响应;
- 4) 挖掘深层次、复杂的攻击行为,达到识别有计划的攻击,从而增加攻击检测率。

1.2 常见的关联分析架构

关联分析的组成部分包括:预处理、关联分析知识库、关联分析引擎、关联分析结果显示以及响应处理。

如图 1, IDS, VPN, FireWall 和 Router 是网络中呈分布式

收稿日期:2005-01-02; 修订日期:2005-03-06

基金项目:国家 863 计划项目(2002AA142030); 国家 863 计划项目(2003AA148020)

作者简介:高雷(1980-),男,湖北武汉人,硕士研究生,主要研究方向:网络安全; 肖政(1976-),男,安徽巢湖人,博士研究生,主要研究方向:网络安全管理、计算机安全体系结构; 韦卫(1964-),男,河南郑州人,研究员,博士,主要研究方向:网络安全、密码学、密码协议安全体系结构; 孙育宁(1965-),男,江苏南京人,研究员,博士,主要研究方向:计算机应用、网络应用、系统测试、性能优化。

部署的节点,这些节点也可以是用于测量网络其他有用信息(比如流量、访问路径等)的检测器,它们把检测结果以告警事件(Events)的形式发往集中分析引擎。当各类检测器生成的告警事件格式不一致时,在发往分析引擎之前,需要将事件作归一化处理,即最终引擎收到的将是描述格式统一的事件。

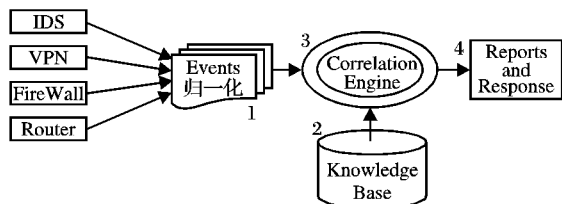


图1 关联分析架构图

构架中核心的部件是2和3,部件2一般是对事件关联关系的定义、事件之间推理规则的描述等,它是部件3工作的基础。部件3依照部件2提供的知识体系对接收事件进行关联分析处理,最后将分析结果提交给管理员,并尽可能对其中一部分作响应处理。

关联分析总体上可以看作是对下级事件产生器提交的结果汇报的综合分析,从某种意义上讲,关联分析引擎本身也变成了一个监测器。它跟基于规则的IDS的工作原理是比较一致的,不过它的“规则”更加复杂。

1.3 关联分析中的关键技术

关联分析包括三个阶段:预处理、分析、结果提交并作响应。

1.3.1 预处理技术

1) 合并与分级

合并技术用于处理重复事件,一般放在检测端完成,便于减轻引擎处理性能上的压力。分级用于划分事件级别,便于提高告警分析的针对性。

2) 归一化

当事件来自异构的安全设备时,格式往往不相同,在提交给引擎作集中分析前,需要进行格式化。可供参考的描述标准有 IODEF, IDMEF 等。其中 IDMEF 格式采用 XML 格式表达,能够在不同产品之间传递和共享,是一种较通用的标准。归一化也一般放在检测器端完成。

1.3.2 关联分析技术

在实际网络环境下,攻击者在实施攻击的过程中,他的扫描行为、口令试探行为、访问文件行为、会话、流量往往会在不同的安全工具上留下相应的特征。关联分析正是要依靠下辖的IDS节点、VPN网关、防火墙、路由器等安全设备提供的这些安全特征信息来对网络安全全局状况作出判断。

1) 关联关系

事件之间的关联关系可分为产生式关联与即时式关联两种。

所谓产生式关联,即根据事件之间的直接因果关系,推断出众多事件的根本原因。其基本原理为,假设: $A \rightarrow B$, 且 $B \rightarrow C$, 则 $A \rightarrow C$ 。(“ \rightarrow ”代表产生关系)。

对于产生式关联有几种不同的处理方法:

基于编码^[2] 根据现象与原因之间的产生关系,生成编码表,然后依照编码表对既存现象进行匹配,从而找到根原因。

基于谓词逻辑^[3~5] 基于现象间的逻辑产生关系,由叶子现象推导根现象。

基于对象^[6] 力图通过专用对象描述语言来描述网络环境中抽象对象(包括物理设备、抽象链接、应用程序等)之间的关系,并以此为依托建立起类似 SNMP 协议 MIB 库的事件库,分析中则通过症状发生节点寻找到公共根节点。

产生式关联是针对那些彼此之间有直接因果关系的事件,

但事实上这种简单的关联关系并不能描述全部的事件关系。

例如某 A,B 事件间的关系表述如下:“在事件 A 发生后,两分钟内发生了事件 B”。在这种情况下 A,B 事件之间不存在直接因果关系,因而无法用产生式关系来描述,但这种事件的发生模式又的确是安全分析人员所关注的。我们把事件之间类似这样的关系称为即时(Temporal)关系。

即时关联通常以专用语言作形式化描述,并针对这样的描述语言定制专用编译器^[5](Compiler)。参照图1来讲,就是其中的 Knowledge Base 是由形式化语言来描述的,引擎的分析过程事实上是一个编译处理的过程,网络安全事件成为这个过程的输入参数。

2) 信息间的关联交互技术

为了增强关联分析的针对性,可以增加与被管系统的交互环节,以判断告警事件是否是真正潜在的危险。

这包括事件与系统弱点信息之间的关联印证^[8]、事件与系统日志之间的关联印证。前者以目标系统的漏洞状况来判断告警的必要性。后者则以目标系统日志来印证告警危险是否属实。

3) 聚类分析技术

基于关联关系的分析技术可以看作是从细节上着手分析复杂现象的本质,而聚类分析^[9]技术则是从全局出发获取网络状态的全貌。

比如对于IDS告警事件,可以使用告警类型、源地址、目标地址和严重程度作为指标,设定不同的条件进行统计分析,对于源地址和目标地址还可以在的层次上给出聚类分析结果,比如子网 10.100.1.* 上的聚类分析结果与子网 10.100.*.* 上的聚类分析结果往往是大不相同的。

1.3.3 关联后处理技术

1) 关联结果呈现

完成关联分析后,首要的事情就是把结果呈现给管理员,常用的手段有统计图、报表、网络攻击拓扑等。

其中网络攻击拓扑是一种较为复杂的技术,即在发现攻击时,实时展现网络拓扑结构,直观地得到攻击者在网络中大致活动位置等信息。

2) 响应处理

系统不单是要让管理员看到实时分析结果,还应有应急响应手段,比如针对某攻击关闭相应端口、提醒用户打补丁、邮件通知等。

1.4 关联分析的研究发展趋势

1.4.1 获取特征模式

目前对于获取模式,较先进的解决方案是基于机器学习^[10~12]。其技术思路是由机器来获得获取待验证的特征模式(在数据挖掘领域中,这种模式就是挖掘算法规则),然后通过智能裁决或人工参与来确认真正有价值的工业实用安全规则。

1.4.2 合理的引擎部署架构

集中式的分析引擎在管理的节点数量比较庞大时,性能问题会变得比较突出。针对这个问题,目前出现了几种分布式引擎部署架构。

1) 级联

在级联架构中,各子节点分别管理一个子域,然后向父节点进行汇报(比如 IBM Tivoli 的管理框架),这样就减轻了子节点的信息负载,不过对父节点的处理能力要求则较高,而且子节点之间信息无法有效共享。

2) 协同

在协同架构^[13]中,各节点在网络中处于平等地位,通过有效的调度算法来管理节点间的交换行为,既要让全域内各节点都能充分通信,又要节制在同一时刻节点之间信息交换的规模。

1.4.3 高性能的分析技术

安全关联分析的内在要求是要尽量达到实时性,即能尽快将出现的异常网络状况反馈给管理员。

目前已出现了一种基于状态机的安全分析技术。它试图以不同的安全状态作为状态机节点,以不同的安全事件的发生或安全性状的出现作为状态跃迁的条件,构建网络安全状态的整体视图,通过已发生的安全事件来判断当前网络安全状况。

2 基于层级式规则的关联分析

2.1 层级式规则的基本思想

基于产生关系的关联分析方法其编码方式、逻辑推理方式虽然简单,但无法描述所有的安全事件关系;而基于即时关系的关联分析方法中,若采用形式化描述语言,则需要针对语言编写专用编译器,又使得工业实施上技术过于复杂。

层级式规则正是作者在所研究的安全集中管理项目中为了弥补前两种分析方法的不足而设计的。一方面,利用层级式规则可以表述较复杂的事件关系;另一方面在分析技术上仍然基于传统的匹配技术来实施关联,达到简洁高效。

2.2 层级式规则的表述形式

事件是规则表述的基础,首先根据实际情况定制事件归一化后的数据属性字段,如:事件类型、事件严重级别、发生时间、目标设备类型、协议类型、源地址、目的地址、源端口号、目的端口号等,这些是安全分析所关注的事件属性。

为了表述复杂的事件关系,将安全事件的发生序列定义为模式。单个事件构成的模式称为原子模式,原子模式由事件的数据字段应满足的条件构成。多个事件序列构成的模式称为复合模式,复合模式可由原子模式通过组合关系(顺序关系、与关系、或关系)构成,并由适当的限定条件定义。模式与模式之间通过关联条件构成规则。

为了达到规则定义的通用性和可扩展性,使用 XML 格式定义规则。

2.3 层级式规则的表述样例

下面给出一个黑客入侵目标系统并通过局域网共享资源散布病毒的规则描述样例。原子模式有密码猜测、共享搜索,复合模式有密码攻击、资源扫描。当两个复合模式均发生而且遭受密码攻击的目标是共享资源扫描的发起者时,认为目标系统已被入侵,并试图通过内部共享资源传播病毒。

```
<atomic_list>
  <atomic_pattern>
    <id>001</id>
    <name> guess_passwd </name>
    <definition> eventtype = "passwd_fail" </definition>
  </atomic_pattern>
  <atomic_pattern>
    <id>002</id>
    <name> share_search </name>
    <definition> protocol = "NetBIOS" port_to = "139"
    </definition>
  </atomic_pattern>
</atomic_list>
<compound_list>
  <compound_pattern>
```

```
<id>1001</id>
<name> passwd_attack </name>
<definition> seq(10, guess_passwd) </definition>
<keep_same> addr_to = "001: addr_to" </keep_same>
<time_out> >120 </time_out>
</compound_pattern>
<compound_pattern>
  <id>1002</id>
  <name> source_scan </name>
  <definition> seq(10, share_search) </definition>
  <keep_same> addr_from = "002: addr_from"
  </keep_same>
  <time_out> >120 </time_out>
</compound_pattern>
</compound_list>
<rule_list>
  <rule>
    <definition> and(1001, 1002) </definition>
    <correlation> cond = "1001: addr_to = 1002: addr_from"
    </correlation>
    <time_out> >600 </time_out>
    <response>
      <mail> smtp = "10.100.4.80" from = "mike@big.com"
        to = "admin@small.com" subject = "virus alert"
        content = "numda virus occurred"
      </mail>
    </response>
  </rule>
</rule_list>
```

2.4 层级式规则的关联匹配算法

基于上述形式的规则描述,给出如下关联分析匹配算法:

- 1) 全局维护一个原子模式队列(AQ),每一个原子模式对应于一个或几个复合模式,而每一个复合模式对应于一个或几个规则,该原子模式队列、原子模式对应的复合模式、复合模式对应的规则,都根据初始规则描述文件生成。
- 2) 全局维护一个复合模式匹配队列(CMQ)和一个规则匹配队列(RMQ),初始化这两个队列为空。
- 3) 接收新事件,根据原子模式队列依次进行原子模式匹配。若不成功,跳到7)。
- 4) 原子模式匹配成功。若复合模式匹配队列中不存在与该原子模式对应的复合模式记录,则生成相应复合模式匹配记录到复合匹配队列;若已存在,进行复合模式匹配;若复合模式匹配不成功,跳到7)。
- 5) 复合模式匹配成功。从复合模式匹配队列中删除该匹配记录。若规则匹配队列中不存在与该复合模式对应的规则记录,则生成相应规则匹配记录到规则匹配队列。若已存在,进行规则匹配;若规则匹配不成功,跳到7)。
- 6) 规则匹配成功。基于规则定义,作出响应处理。从规则匹配队列中删除该记录。
- 7) 检查复合匹配队列以及规则匹配队列,若存在超时,将记录从匹配队列中删除;返回3)重复执行。

匹配算法的基本思想:根据规则描述文件得到所有的原子模式、与原子模式对应的复合模式、与复合模式对应的规则;根据事件对原子模式进行匹配;若匹配成功,转而进行复合匹配;若复合匹配成功,转而进行规则匹配;匹配成功则生成告警并根据规则定义进行响应处理。

2.5 层级式规则的优点

- 1) 规则描述上具有良好的扩展性,能充分描述工业环境中的事件关联关系。

4 结语

基于系统状态集合的攻击模型以攻击对目标系统的影响为出发点,对攻击过程进行描述。首先,根据系统本身提供的服务和攻击行为的后果可以归纳出多种安全相关的系统状态。由于系统可能会同时具有这些状态中的一种或多种,所以使用状态的集合来表示系统。攻击行为对系统造成的影响实质上是状态集合的变化,因此,将攻击行为定义为前提条件、初始状态、完成状态的三元组。最后,以攻击行为的有序排列表示攻击的过程。同其他攻击模型相比较,基于系统状态集合的模型是针对攻击过程检测和安全预警设计的,它具有如下优点:

1) 模型不仅仅刻画组成攻击过程的各个攻击行为之间的关联,而且反映了攻击过程对系统造成的危害。所以容易应用该模型进行安全预警。

2) 模型只考虑导致系统状态集合增加新的、更加危险的状态的攻击行为,并且对系统状态集合影响相同的攻击行为会被定义为同一种攻击行为。因而,模型能够对大量攻击行为进行有效的归纳,减少安全预警不需要的冗余信息。

3) 模型以整个被保护系统为描述对象,而不仅仅是针对单一主机。因此,模型能够更好地适应实际需求。同时,模型的形式化程度高,易于程序实现。

4) 模型对攻击行为的定义使得它可以充分利用目前已有的攻击行为研究成果。实际上,可以比较容易地将目前存在的各种漏洞描述转换成模型中对攻击行为的描述。

本文还提出一种使用基于系统状态集合的攻击模型进行攻击过程检测和安全预警的方法,该方法将实际的攻击过程同已知的攻击过程进行不完全匹配,根据匹配的程度及系统的安全要求预测系统将会具有的危险等级。

(上接第1528页)

例如:对事件关系进行描述时,使用 timeout 字段可以定义事件之间在时间域上的相互关系,而它在传统的产生关系中是无法表述的。

另外,可以使用语句将一个模式中的字段信息与其他模式总的字段信息进行关联,前面表述样例中就使用 <keep_same> addr_to = "001; addr_to" </keep_same> 这样的语句表达“本模式关注所有那些目标地址跟 001 模式目标地址相同的事件”。通过这样的描述方法,十分简单有效地表达了事件之间的关联关系。

2) 与专业化形式描述语言相比较而言,层级式规则降低了技术实施的难度。

首先,使用 XML 表达层级式规则的系统免去了编写专用语言编译器的环节,另外,目前 Java 语言以及 Linux 下的 C 语言 Lib 库都对 XML 语言提供了强大的支持,不需要系统开发额外的规则文法检查及规则文法解析工具。

3) 规则描述中包含了响应描述,有利于分析及响应的一体化。

参考文献:

- [1] HAINES J, RYDER DK. Validation of sensor alert correlators[J]. IEEE Security & Privacy, 2003, 1(1): 46-56.
- [2] KLIGER S, YEMINI S. A coding approach to event correlation[A]. Proceedings of 4th International Symposium on Integrated Network Management (IFIP/IEEE) [C]. Santa Barbara, CA, 1995.
- [3] GRUSCHKE B. Integrated event management: event correlation using dependency graphs[A]. DSOM'98[C], 1998.

参考文献:

- [1] VAN DOORN L. Computer Break-ins: A Case Study [A]. NLUUG proceedings [C], 1992.
- [2] BOULANGER A. Catapults and grappling hooks: The tools and techniques of information warfare [J/OL]. IBM Systems Journal, 1998, 37(1).
- [3] Common Vulnerabilities and Exposures [EB/OL]. <http://cve.mitre.org>, 2003-08.
- [4] 王晓程, 刘恩德, 谢小权. 攻击分类研究与分布式网络入侵检测系统[J]. 计算机研究与发展, 2001, 38(6): 727-734.
- [5] SCHNEIER B. Attack Trees: Modeling Security Threats [J]. Dr. Dobbs's Journal, 1999, 12(24): 21-29.
- [6] MOORE AP, ELLISON RJ, LINGER RC. Attack modeling for information security and survivability [R]. CMU/SEI-2001-TN-001, 2001.
- [7] TIDWELL T, LARSON R, FITCH K, et al. Modeling Internet Attacks [A]. Proceedings of the 2001 IEEE Workshop on Information Assurance and Security [C], 2001. 54-59.
- [8] HUANG M-Y, WICKS TM. A Large-scale Distributed Intrusion Detection Framework Based on Attack Strategy Analysis [A]. Proceedings of the First International Workshop on Recent Advances in Intrusion Detection [C], 1998.
- [9] MCDERMOTT J. Attack Net: Penetration Testing [A]. The 2000 New Security Paradigms Workshop [C], 2000. 5-22.
- [10] STEFAN J, SCHUMACHER M. Collaborative attack modeling [A]. Proceedings of SAC [C], 2002.
- [11] ILGUN K, KEMMERER A, PORRAS P. State Transition Analysis: A Rule-Based Intrusion Detection Approach [J]. IEEE Transactions on Software Engineering, 1995, 21(3): 181-199.
- [12] BACE R, MELL P. Intrusion Detection Systems [A]. NIST Special Publication on Intrusion Detection Systems [C], 2001.

- [4] HASAN M, SUGLA B, VISWANATHAN R. A conceptual framework for network management event correlation and filtering systems [A]. Proceedings of the Sixth IFIP/IEEE International Symposium on Integrated Management [C], 1999.
- [5] LIU G, MOK AK, YANG EJ. Composite event for network event correlation [A]. Proceedings of IM'99 [C], 1999.
- [6] OHSIE D, MAYER A, KLIGER S. Event modeling with the MODEL language: A tutorial introduction [EB/OL]. http://www.smarts.com/resources/code_tpapers_model.pdf, 2004-12.
- [7] CUPPENS F, MIEGE A. Alert correlation in a cooperative intrusion detection framework [A]. Proceedings of the 2002 IEEE Symposium on Security and Privacy [C], 2002.
- [8] GULA R. Correlating IDS alerts with vulnerability information [EB/OL]. http://www.tenablesecurity.com/white_papers/va-ids.pdf, 2004-12.
- [9] DEBAR H, WESPI A. Aggregation and Correlation of Intrusion-Detection Alerts [A]. RAID 2001, LNCS 2212 [C], 2001. 85-103.
- [10] LEE W, STOLFO SJ. A framework for constructing features and models for intrusion detection systems [J]. ACM Transactions on information and system security, 2000, 3(4): 227-261.
- [11] STOLFO SJ, LEE W. Data mining-based intrusion detectors: An overview of the Columbia IDS project [J]. SIGMOD Record, 2001, 30(4): 5-14.
- [12] LEE W, STOLFO SJ. Real time data mining-based intrusion detection [A]. Proceedings of DISCEX III [C], 2001.
- [13] LOCASIO ME, PAREKH JJ, STOLFO S. CUCS-012-04, Collaborative distributed intrusion detection [R], 2004.