doi :10.3969/j.issn.1671-1122.2011.09.030

# 一种网络安全事件关联分析的专家系统研究

王雯霞，贾焰，韩伟红，徐镜湖，郑黎明

（ 国防科技大学计算机学院，湖南长沙 410073 ）

**摘 要：**该文针对现有入侵检测系统存在误报率高、漏报率高等问题，提出了一种用于网络安全事件关联分析的专家系统。该方法对共性知识库进行分层立体化建模以提高关联分析性能，提供资产信息和漏洞信息分析模块来提高对重点设备、网络区域、网络安全事件的关注度，并对冗余信息进行剪枝、去重。同时，在专家系统中引入时间流，从而提高系统的实时性。通过真实环境下的实验分析说明该方法能有效提高关联分析性能，具有易添加、易扩充等优势。

**关键词：**网络安全；关联分析；专家系统；时间流

**中图分类号：**TP393.08 **文献标识码：**A **文章编号：**1671-1122（2011）09-0097-04

## An Expert System Based on Network Security Event Correlation

WANG Wen-xia, JIA Yan, HAN Wei-hong, XU Jing-hu, ZHENG Li-ming

( *School of Computer Science ,National University of Defense Technology ,Changsha Hunan* 410073*, China* )

**Abstract:** It puts forward an expert system based on network security event correlation, to solve the problem that there are high false alarm and missed alarm rate existing in IDS. This paper presents abstraction modeling for knowledge base to advance the performance of correlation analysis, and presents assets information and vulnerability information analysis module to increase the attention to important equipment, network area and network security event, also presents pruning to optimize the redundance. On the other hand, it introduces time stream into expert system to improve the real time action. The system has been applied in real condition, and the results of experiments show that the system can effectively advance the performance of correlation analysis, and can easily extend.

**Key words:** networking security; alert correlation; expert system; time stream

## 0 引言

入侵检测技术存在高误报和高漏报问题[1]。如何科学、合理、统一地组织信息，并快速、有针对性地将它们与已有的知识进行匹配，从而高效、准确地从这些信息中提取出我们关注的网络安全事件成为了入侵检测技术发展的关键。

将关联分析[2]与传统专家系统[3]相结合，是入侵检测技术的一个发展方向。Onera Toulouse 使用专家系统评估关联分析中的相似度[4]。Steven Cheung 等提出了一种基于专家系统的攻击场景识别技术[5]，并建立了 CAM 原型。吴志超将专家系统应用到网络故障诊断模块中，对告警信息进行关联分析[6]。

综上所述，国内外研究者提出了多种可行的解决方法，但依然存在诸多方面的不足，例如：面对海量网络信息，现有关联分析技术性能较低，而从算法上改进出现瓶颈；没有内建的、处理有序数据的能力是专家系统的致命伤，但网络安全的应用环境对实时性有很高的要求。

为了解决上述问题，首先将知识库进行分层立体化建模，从而将高频率发生的共性操作集中，一方面精简知识库，另一方面为低频率、大深度操作删减掉大量冗余信息，极大地提高性能；其次，关联分析部分，添加资产配置和漏洞信息分析模块，通过关联重点设备、网络区域数据，去掉不需要关注的信息，来降低开销；再次，对破坏过程行为进行匹配来聚合网络安全事件，用于确认安全事件是否真实发生，判断攻击是否成功，以消除误报警，计算告警可信度（风险值）来识别误报警，有效识别高风

**收稿时间：**2011-07-15

**基金项目：**国家高技术研究发展计划 (863 计划）（2011AA010702）、江苏省基金（BK2010131）

**作者简介：**王雯霞（1987-），女，四川绵竹人，硕士研究生，主要研究方向：信息网络安全；贾焰（1961-），女，教授，博士生导师，主要研究方向：网络信息安全、数据挖掘、社会网络；韩伟红（1973-），女，副教授，主要研究方向：网络信息安全、数据库与数据挖掘；徐镜湖（1985-），博士研究生，主要研究方向：网络信息安全、数据库与数据挖掘；郑黎明（1983-），博士研究生，主要研究方向：网络信息安全、数据库与数据挖掘。

险告警；然后，采用剪枝操作来降低冗余开销；最后，添加时间流分析机制，满足实时要求。

## 1 系统结构

网络安全事件关联分析的专家系统是国家高技术研究发展计划（863 计划）网络安全态势分析与预测系统的核心部分，对分布式入侵检测系统、专家系统、实时处理、知识表示、模式匹配、关联分析等技术进行深入的研究，集成各项先进技术优势，生成正确的报警与警告，降低误报率和漏报率，提高网络安全告警的性能。

网络安全事件关联分析专家系统基于产生式系统实现，由知识库、规则库、推理机、工作存储器、数据库（包括用户数据库、知识数据库、事件数据库等）、知识获取机制、用户解释界面等部分组成，其中，基于关联分析的知识库和规则库是本系统的核心。

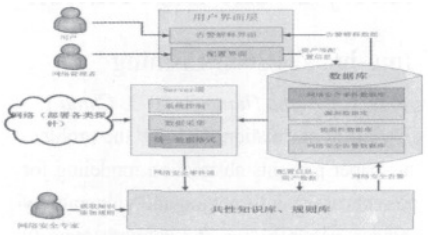

图1 系统结构图

本系统整体流程为图 1 中，Server 控制端将各探针采集到的数据进行整理、统一数据格式，形成网络安全事件流，提交给推理机，插入到工作存储器中。推理机的调度器跟据工作存储器中的事件、当前状态及有关信息，在规则库中进行模式匹配，并通过冲突解决策略管理可能冲突的规则的执行顺序。规则的执行会修改工作存储器的状态，推理机再根据这种新的状态驱动新的查找，如此反复，直到没有事件可以匹配规则，推理结束。在推理过程中，如果有事件流与告警规则匹配，则生成网络安全告警，并写入数据库中。用户界面通过调用数据库，可以生成告警解释说明报告。

为了提高报警性能，在知识库和规则库中，对已有告警规则进行分析，抽象出它们的共性，将这些共性分层、分类，搭建金字塔型知识库，实现关联分析过程从扁平到立体化，从而尽早将大量冗余信息过滤掉。

## 2 知识表示

基于关联分析的专家系统知识库有以下定义。

定义 1：网络安全事件，是在网络环境中，各探针检测到的数据提交到服务端，进行集成、整理、规范化后形成的信息。

每个网络安全事件的属性为 15 元组，格式如下，<class_id,calss_sid,category_id,category_sid,src_ip,src_port,dst_ip,dst_port,priority,reliability,startTime,is_collate_attack,is_collate_cluster,is_collate_fine_grained_cluster,id>。其中，class_id 为网络安全事件类型标识，包括木马、蠕虫、扫描、暴力破解、拒绝服务等类型，calss_sid 为其细分类；category_id 为网络安全事件破坏过程行为类型，包括特征、获取权限、信息泄露等类型 ,category_sid 为其细分类；src_ip、src_port 为发起攻击主机的 IP 地址及端口；dst_ip、dst_port 为攻击目标的 IP 地址及端口；priority（事件优先级 [0–5]）如果一个攻击实时发生产生的危害；reliability（事件可信度 [0–10]）攻击实际发生的可能性；startTime 为安全事件发生时间；is_collate_attack 标记该网络安全事件是否激活一个新的 Attack；is_collate_cluster 标记该网络完全事件是否进行了某类统计；is_collate_fine_grained_cluster 标记该网络完全事件是否进行了某类细粒度统计；id 为网络安全事件在系统中的编号。

定义 2：Attack（Alarm）网络安全攻击（网络安全告警），Attack 是由网络安全事件激活的某一类网络安全攻击，经过关联分析，如果与预定破坏过程匹配，那么它就是网络安全告警 (Alarm)。

每个网络安全攻击的属性为 16 元组，格式如下，<class_id,calss_sid,category_id,category_sid,src_ip,src_port,dst_ip,dst_port,priority,reliability,risk,startTime,is_alarm,step,srcDstList,id>。其中，属性 class_id,calss_sid,category_id,category_sid,src_ip,src_port,dst_ip,dst_port,priority,reliability, startTime, id 与网络安全事件的属性的意义基本一致，这些属性的值都是由激活本 Attack 的网络安全事件赋予的；Risk（攻击风险 [0–10]）：网络安全攻击发生的风险评估，由相关网络安全事件的 Priority 、Reliability，以及攻击主机的 asset（资产重要度 [0–5]）计算得到；is_alarm 为标示该 Attack 是否已经确认为网络安全告警（Alarm）；step 标示破坏过程具体匹配到哪一步行为；srcDstList 记录破坏工程中每一步相关的攻击主机和被攻击主机的信息。

定义 3：网络安全事件统计（Cluster），跟据破坏过程需要，进行对网络安全事件的统计。

每个网络安全事件统计的属性为 11 元组，格式如下，<class_id, category_id, src_ip,src_port,dst_ip,dst_port,startTime,count,id,is_special,is_formal>。其中，src_ip,src_port,dst_ip,dst_port 跟据统计目标可选，因此分类包括源 IP、目的 IP、源目的 IP、目的端口等；count 为计数值；is_special 标示特殊统计情况；is_formal 标示对应攻击是否生成告警。

定义 4：网络安全事件细粒度统计（fine_grained_cluster），格式为 <class_id, category_id,category_sid,src_ip,src_port,dst_ip,dst_port,startTime,count,id,is_special,is_formal>, 相比网络安全事件统计（Cluster），加强对破坏过程行为的统计，即 category_sid。

## 3 基于关联分析的推理过程

关联分析过程是通过破坏过程行为匹配来聚合网络安全

事件,计算告警可信度(风险值)来识别误报警,融合在知识库、规则库中。分为三层，从顶层到底层，每一层的抽象度依次递减，共性也依次递减，直到底层详细规则。每一层、每一类根据需要建立对应的统计项，并根据统计值及已有知识判断是否进入下一层。层数越底层，统计项越详细，这样在上层时可以不用过多考虑不确定的因素，保证上层的效率。另一方面，由于上层过滤掉了多余的信息，这样使得大量的底层详细规则不用面对过多的信息，提高底层的效率。

### 3.1 推理流程及算法

统计项与网络安全事件之间、统计项与网络安全攻击之间、细粒度统计项与网络安全事件之间、细粒度统计项与网络安全攻击之间都进行时间聚类，提高系统时效性。

具体实现上，网络安全事件流进入推理机进行关联分析(图 2)，包括网络安全事件脆弱性关联、网络安全事件与资产关联、网络安全事件与事件关联。其中，脆弱性关联与资产关联在系统中为第一层，它们可将用户特别关心以及特别不关心的信息筛选出来，为事件与事件关联排除掉大量信息。为了进一步提高系统性能，将事件与事件关联分为粗、细粒度两层，分别为系统的第二、三层，第二层是对攻击破坏过程行为进行整理后，总结出行为的共性进行关联，第三层则是对攻击的破坏过程具体行为关联。推理过程主要采用正向推理的控制策略进行基于时间序列规则的因果关联。

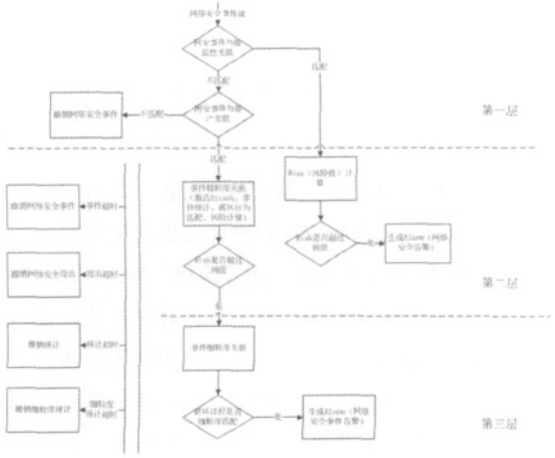


图2 算法流程图

在第二层中，网络安全事件要进行的行为包括与统计计数规则匹配、与激活新的网络安全攻击规则匹配，如果匹配成功，并且新的统计项与网络安全攻击的破坏行为规则匹配，那么会生成新的统计项，新统计项的统计结果会与网络安全攻击的破坏过程下一行为规则匹配，直到还有网络安全事件攻击破坏行为过程匹配的规则。具体算法见表 1。

第二层结束时，已可以判断是否有网络安全告警生成，保证告警时效性，进入第三层，则是对告警进行进一步加强确定，保证告警准确性，降低误报率。

表1 第二、第三层算法

| 第二层 | 第三层 |
|---|---|
| While（流中存在 event 或者 cluster 或者 Attack）<br>{<br>If（event 符合某 cluster 统计条件）<br>Then 该 cluster 计数加一<br>If（event 符合某 Attack 生成条件）<br>Then 生成新的 Attack，生成该攻击破坏过程第一步的行为<br>If（某 cluster 符合某 Attack 破坏过程行为）<br>Then 该 Attack 破坏行为进入下一步，生成新的 cluster<br>If（某 cluster 符合某 Attack 破坏过程行为最后一步）<br>Then 计算 risk 值<br>If（某 attack 的 risk 值大于阈值）<br>Then 推理进入第三层<br>} | While(流中存在 event 或者 Fine-grained-cluster 或者 Attack)<br>{<br>If（某 attack 的 risk 值大于阈值并进入第三层）<br>Then 跟据其破坏过程详细行为生成各 Fine-grained-cluster<br>If（event 符合某 Fine-grained-cluster 统计条件）<br>Then 该 Fine-grained-cluster 计数加一<br>If（若干 Fine-grained-cluster 符合某 Attack 破坏过程行为）<br>Then 该 Attack 成为告警<br>} |

### 3.2 去冗与剪枝

某些规则产生的结果可能冗余，比如重复统计项，风险值过低的 Attack，各类超时对象。对于这些需要剪枝的对象，也在推理过程中进行适当处理。

一方面，系统规则支持统计项去重判断、上层进入下层判断，进行剪枝、去冗操作。

另一方面，对各类对象分别进行分析，跟据其估计最长有效生命周期，设置相应的时间阀值，调度程序自动判断网络安全事件超时、网络安全攻击超时、统计超时、细粒度统计超时等，并执行撤销任务。

规则实现时语法参考 MYCIN[7] 系统规则表示。

## 4 实验结果

在实验中，如图 3 有台式机一台作为演示机，运行 Win7 操作系统，安装 Firefox 浏览器；笔记本一台作为攻击机，运行 Win7 操作系统，安装攻击程序 AttackReplay；服务器两台，安装 Server 端以及 Web 程序；服务器五台，安装传感器 Snort、Ntop、Nessus、Nagios 等。

为了进行对比，下载开源安全信息管理系统 OSSIM2.0，它是目前一个非常流行和完整的开源安全架构体系。OSSIM 通过将开源产品进行集成，从而提供一种能够实现安全监控功能的基础平台。 它的目的是提供一种集中、有组织的，能够更好地进行监测和显示的框架式系统。

为了实验的公平、合理，两个系统的规则库来源一致。

指定攻击的频率、规范设备关注等，根据这些生成的脚本由攻击机进行攻击。

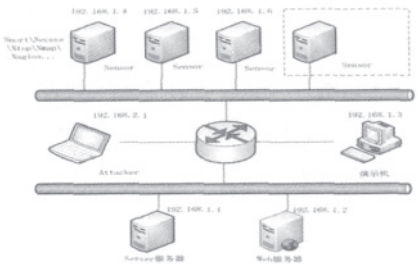

图3 系统部署图

图 4、图 5 中，系统处理网络安全事件性能确实实质上得到很大的提高，分析数据，性能提高最大有 8 倍左右。导致性能有如此大提高的原因，主要有以下三点：

1）相比 OSSIM 系统每次场景匹配时都要进入规则库中读取关联规则，本系统采用专家系统的工作存储器，用于保存运行过程中所需信息（如初始状态、知识和中间状态），不需要多次读取和记录内存，从而大大节省了时间。

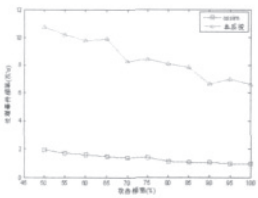

图4 Trojan攻击测试结果对比

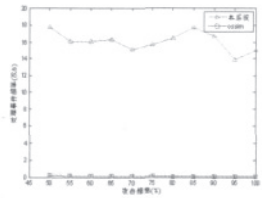

图5 Worms攻击测试结果对比

2）专家系统知识库立体化建模及去冗余模块发挥了作用。分析图 4，随着攻击频率（即生成网络安全告警的网络安全事件样本在所有样本中的比例）的提高，两个系统单位时间处理网络安全事件数目都在下降，但是本系统下降幅度更大，说明本系统的多层次立体结构去除了大量不相关网络安全事件信息，在面对海量攻击时，对提高性能发挥重要作用。

3）时间流技术的引入，弥补了传统专家系统的缺陷，使其具有大量实时数据处理能力。

## 5 结束语

本文提出一种用于网络安全事件关联分析的专家系统。该方法对共性知识库进行分层立体化建模以提高关联分析性能，添加资产信息和漏洞信息分析模块来提高对重点设备、网络区域、网络安全事件的关注度，并对冗余信息进行剪枝、去重。同时，在专家系统中引入时间流，提高系统的实时性。实验表明，基于关联分析的专家系统能够准确、高效地分析隐藏在原始网络安全事件背后的逻辑联系和攻击意图，挖掘其内在联系，除去冗余，剔除误报，减少漏报。（责编 程斌）

**参考文献：**

[1] 罗宁，喻莉．入侵检测技术研究发展 [J]. 计算机与数字工程，2005，33（06）：6．

[2] P.A.Porras，P.G.Neumann. EMERALD:Event Monitoring Enabling Responses to Anomalous Live Disturbances[C].Proc.20th NIST—NCSC National Information Systems Security Conference,1997：353—365.

[3] Michad M.Sebring, Eric Shellhouse,Mary E.Hanna, R Alan Whitehurst. Expert systems in intrusion detection:a case study[C]. Proceedings of the 11th National Computer Security Conference,Baltimore,Maryland,1988：74-81.

[4]F. Cuppens. Managing Alerts in a Multi-Intrusion Detection Environment[C].Prooeedings 17th Computer Security Applications Conference,New Orleans,LA,December 2001.

[5]Steven C, Ulf L, Martin F. Modeling Multistep Cyber Attacks for Scenario Recognition[C]. Proc. of Third DARPA InformationSurvivability Conference and Exposition, Washington, 2003.

[6] 吴志超．使用专家系统进行网络告警相关性分析 [A]. 中国通信学会无线及移动通信委员会、IP 应用与增强电信技术委员会 2007 年度联合学术年会论文集 [C]. 2007．

[7] B G Buchanan, E Shortliffe .Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project[C].1984.

上接第 96 页

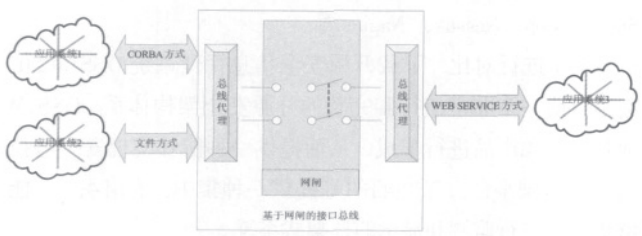


图3 基于网闸的接口总线式电子政务安全模式解决方案

实现系统间安全。

## 4 结束语

电子政务系统的目标是建立开放的、基于标准的统一网络平台。如何在不影响开放性的前提下提升网络信息安全，推动统一标准的网络信息平台建设，正是本文探索的目标。我们基于网闸的安全模式提出了异构数据下的中间件解决模型，并建议制定一体化的、统一标准的接口总线模式解决方案，对于推进电子政务的健康有序发展，具有积极的意义。

需要指出的是，电子政务规划是一个长期的过程，更是一个在建设和应用推进下不断完善的过程。各级政府信息化人员应当担负起对规范化数据和整体发展框架进行研究、定义的重任，利用深入了解业务应用的优势，为各级政府制定出长远的 IT 应用和 IT 安全滚动规划。而且更重要的是，滚动规划必须随着信息化应用水平的不断提升，实践探索中的不断的发现问题，而不断更新，唯有如此，才能真正做到网络信息与网络安全的真正有序发展。（责编 程斌）

**参考文献：**

[1] 李建设，卢辉斌，陈淑清，徐天斌．电子政务系统安全的框架性解决方案 [J]. 计算机工程与设计，2007，7（14）：3486-3488.

[2] 黑广彬．网闸在检察院信息化系统建设中的应用探讨 [J]. 情报探索，2007，8（08）：68-72.

[3] 许云明，李春生．物理隔离网闸原理及应用 [J]. 计算机安全，2005，（12）：26-29.

[4] 童恒庆，聂会琴，李锡韶 .CORBA / COM / EJB 三种组件模型的分析与比较 [J]. 计算机应用研究，2004，21（04）：66-68.

[5] 肖飞．主流中间件技术比较及其在电信行业的应用初探 [J]. 技术与维护，2007，（02）：23-25.