广西科学院学报
Journal of Guangxi Academy of Sciences

2009,25(4):278~281
Vol. 25,No. 4 November 2009

基于聚类的触发式网络监控模型研究*
Research on the Clustering-based Triggering Network Monitoring Model

李陶深,陆宇昊,葛志辉
LI Tao-shen,LU Yu-min,GE Zhi-hui

(广西大学计算机与电子信息学院,广西南宁 530004)
(School of Computer, Electronics and Information, Guangxi University, Nanning, Guangxi, 530004,China)

摘要:根据网络传输数据量大,实时性高的特点,提出一种基于聚类的触发式网络监控模型。该模型将系统划分为网络用户层、实时监控层、数据分析层、触发控制层、前端应用管理层等五个层次,将数据处理分为信息采集、数据初步分析、数据聚类、结论确定和响应阶段,依靠各模块实现各阶段的任务和功能。模型各部件分布式部署,可各自承担数据处理和传输存储任务,提高了整体运行效率。

关键词:网络监控 结构模型 触发式 聚类
中图法分类号:TP393.08 **文献标识码**:A **文章编号**:1002-7378(2009)04-0278-04

Abstract: A clustering-based triggering network monitoring model is proposed in this paper. This model divides the system according to different levels. It uses the proxy interactive mechanism to realize the integration of the network data acquisition, storage, analysis, maintenance and network control and management. The data processing is divided into the information acquisition, data preliminary analysis, data clustering, conclusion and response. Every functional module is distributedly deployed, which can respectively process the data, transmission and storage. It raises the effectiveness of the whole system.

Key words: network monitoring, structural model, triggering, clustering

随着网络技术的发展和互联网普及率的提高,网络规模迅速激增,信息化水平越来越高。有效管理日益复杂的网络及其使用,掌握网络运行状况和用户使用网络的状态,对网络行为实施实时监管,都是网络研究必须面对的问题。

网络上许多安全事件的发生和安全隐患的存在都与管理不善有关[1]。因此,只有不断完善和加强各种安全管理手段,才能有效保证网络系统的安全。当前,连接到互联网的主机数目猛增,使得网络规模不断扩大,网络应用又已经深入各个关键领域,大量网络攻击充斥于这个庞大的网络系统,内部网的安全受到越来越多的挑战。因此,网络行为实时监控系统的研究和应用是十分必要的。利用实时监控系统对各种网络行为实时追踪,有效防范违规网络行为,是网络安全研究的一个方向。

聚类是数据挖掘领域中最成功的一种利用无监督学习过程获取知识的方法[2],已经初步应用于入侵检测系统[3]、网络安全监控的用户行为分析[4]、网络安全事件计算[5]、网络安全系统的警报聚类[6]等方面。

本文鉴于网络传输数据具有传输量大、实时性高等特点,提出一种新的基于聚类的触发式网络监控模型 CTNMM (Clustering-based Triggering Network Monitoring Model)。

1 CTNMM 模型的设计目标

CTNMM 模型的设计是从网络信息实时捕获和数据挖掘中的聚类技术两方面入手,通过充分利用一切网络数据分析用户状态,对网络实施有效监

收稿日期:2009-10-10
作者简介:李陶深(1957-),男,教授,博士,主要从事网络计算与信息安全、分布式数据库、CAD理论与应用研究。
*广西科技创新能力与条件建设项目(桂科能 07109008-006-Z),广西自然科学基金项目(桂科自 0832056)资助。

控。作为置于内部网各个关键节点的触发式网络监控系统,应该是合理分工、协调工作,能高效、准确地掌握用户网络访问状态,只有这样才能为触发式监控提供条件,为智能化决策提供依据。而且,触发式监控和智能化决策依赖于对已有信息的准确挖掘和科学的系统运行机制。因此,CTNMM 模型的设计目标是实时性、准确性和触发式监控。

1.1 实时性

实时性是指随时能捕获网络中传输的原始数据,减小丢包率,连续、完整的原始数据能够真实记录和反映用户一贯和实时的网络访问情况。为此,在模型的设计中,应使之能快速从实时网络信息中提取行为的有效特征,加快触发响应时间。同时,应保证各功能单元间的数据顺畅交互,尽可能避免某一处理环节的瓶颈,提高整体效率和性能。

1.2 准确性

准确的监控分析包括对原始数据从网络底层协议到高层协议的准确分析、数据包内容的精确查找、历史数据挖掘的准确性、用户网络访问的精准判断等。分析用户行为规律和访问趋势,对用户单个或多个连续行为做进一步判断分析,也是准确性要求的具体表现。

1.3 触发式监控

CTNMM 模型的触发式监控是由 1～2 个典型网络特征粗粒度地触发一级响应,从而启动细粒度的监控分析,比对聚类分析结果,得出较精确的判断并做出响应。这样的监控可以尽可能缩小最终防范响应的时间与用户行为的时间差,将危害程度降到最小,使整个网络的运行状况和用户的每一步行为都在掌控之中,提高了系统的控制能力。

2 CTNMM 模型总体架构

2.1 CTNMM 模型的结构

改进和扩展文献[7,8]的监控结构模型,我们设计了 CTNMM 模型的总体架构(如图 1 所示)。

CTNMM 为五层结构模型,包含有网络用户层、实时监控层、数据分析层、触发控制层和前端应用管理层。

网络用户层为数据源层次,由所有的用户终端组成,即为 CTNMM 的监控对象。

实时监控层的主要任务是监听任何与网络用户层的用户终端有关的网络实时数据,根据一定的规则存储。该层次首先在网络数据汇集的关键节点实时监听,对数据进行解析,筛选有效数据缓存到数据缓冲区,数据缓冲区是数据处理和中转的临时缓存空间。从数据缓冲区中调取的数据须经过数据的协议分析,将原始数据解封装并进行统一格式的处理,再由数据转存模块发送至数据分析层存储。

数据分析层主要由数据接收、数据分析和数据库组所组成。该层根据系统的不同需求,完成各种类型的数据分析,与触发控制层的交互较多。

触发控制层根据触发条件触发警报和响应。

前端应用管理层实现前端可视化的配置、管理和系统响应显示。

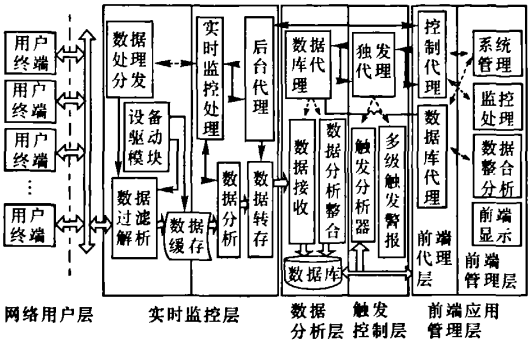

图 1 CTNMM 模型结构

◇:数据流;←┈→:控制指令;↔,→:内部通信。

2.2 CTNMM 模型的工作流程

(1)从网卡中监听数据,而且监听完整帧数据,对碎片数据不予保存。

(2)解析监听到的数据,将解析后的数据分 2 个位置保存,分别是数据库中和缓存中。缓存中的数据为进一步提取特征值提供方便。

(3)提取数据基本特征值。根据定义的相关特征组成,针对若干条目同时提取,并通过触发器随时检测提取特征值后的数据集。模型可以对触发器各个特征值设置阈值,根据用户行为的典型特征来粗略判定用户是否违规。若超过阈值,则将此用户定义为疑似违规用户,并启动响应进行进一步判定。

(4)对数据实施聚类分析和比对分析。通过对数据解析存储的原始数据库中的数据进行聚类分析,得出包含所有数据信息的知识库,并将该用户行为与其他所有数据进行比对。若当前行为尚未出现过,则保存在未知数据库中,其与原始数据库统一格式,并可供聚类分析在未来调用从而将该行为保存。若判定行为违规,即在精确分析中确认违规用户,可以采取响应措施对该用户进行限制或进一步监控。

3 CTNMM 模型的数据分析与处理

CTNMM 模型的网络监控过程主要分为 4 个

阶段：网络实时数据信息采集阶段、数据初步分析整理阶段、数据聚类综合分析阶段、结论确定和响应阶段。根据不同阶段，数据的分析流程如图2所示。

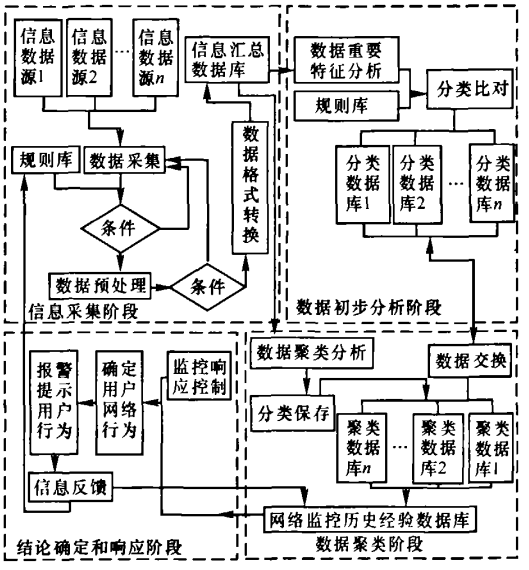

图2 数据分析流程

3.1 信息采集阶段

从分布于网络各关键节点的多个信息数据采集源获取原始的实时网络数据，制定相应规则和处理条件对数据进行预处理，并将各种数据转换成格式统一的规范数据，存储到信息汇总数据库。

3.2 数据初步分析阶段

分析多种网络访问行为表现的典型特征，重点分析在网络管理中违规和禁止的网络行为，制定典型特征的规则库。分析实时网络数据时，将从实时原始数据分析得到的实时特征与规则库分类进行比对，初步判断终端用户的网络访问行为。

3.3 数据聚类综合分析阶段

对信息采集阶段得到的原始数据汇总作为数据源，分析网路行为的各种特征，从而分析各种网络行为类别的精细结果，聚类结果作为实时数据的分析标准。此阶段的聚类分析与初步分析阶段的数据进行交互，初步分析阶段分析典型特征，而聚类阶段分析的是全面的特征，因此，聚类阶段的数据比对更为全面、复杂。

3.4 结论确定和响应阶段

经过数据初步分析和聚类综合分析阶段，监控系统发出监控响应控制的指令，由系统综合分析确定用户行为，发出警报提示。监控系统的判断决策正确与否需要证实，信息反馈到各有关规则库和数据库记录或适当修正存储结果，便于监控系统提高准确率。

CTNMM模型中的任何数据分析结果都是为了对网络行为做出正确判断和决策，触发具有警示作用的模块，从而最终采取适当的响应控制措施，有效管理网络。因此，该模型体现的是历史与实时，及分析、决策与报警的关系。该模型的分析、决策和警报等功能单元及处理数据的流程如图3所示。分析器由协议分析器、历史数据分析器和实时数据分析器组成，负责数据协议分析开始到终端用户网络访问行为分析整个过程的数据分析。历史数据分析器将历史数据进行特征提取分析和聚类分析，这些非实时操作在系统后台完成。实时数据分析器提取实时网络数据的特征，比对历史数据聚类结果，从而分析出终端用户最有可能的网络访问行为。控制模块发出指令控制历史数据分析器和实时数据分析器的启动和运行，分析结果通过数据存储模块与数据库组交互。决策器由决策判断模块、决策处理模块和显示模块组成，在分析器与警报响应之间起承上启下的核心决策作用。分析器与决策器的分析结果和决策有交互过程。决策判断模块收到分析器传送的分析结果后，判断分析层次，通过调出有关数据包内容、关键字比对或人工控制等方法，得到决策结果。决策处理模块根据决策结果，触发警报或网络设备模块。显示模块负责与决策处理同步的屏幕显示。警报与响应采用两级警报体系，粗粒度分析与决策启动一级警报，细粒度分析与决策启动二级警报。

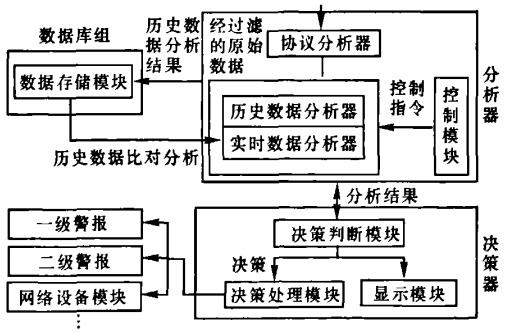

图3 CTNMM中分析、决策与报警的关系

分析器有初步分析和精细分析的功能，初步分析后，决策器判断触发一级警报，决策器反馈信息给分析器进行精细分析，分析结果再次传送至决策器，由决策器决定是否触发二级警报，给出网络控制管理的适当策略或通过网络设备接口操作网络设备。

4 模型测试与性能分析

下面以触发响应模块的测试为例，说明

CTNMM 模型的性能。触发响应模块的测试可以看作是网络监控系统的整体性能测试,因为监控系统所有的数据分析和处理都是围绕着系统能否正确判断用户的网络行为并做出响应,因此触发响应的正确率、误报率和漏报率等指标就是监控系统整体性能的标准。

触发响应模块测试分析是在进行聚类分析得到结果之后,对系统的连续测试。我们选取 3 种情形进行测试:(1)仅监控一个网络终端用户,而且该用户的网络访问行为在一个时间段内是单一的,如先浏览网页,然后使用 P2P 下载工具,下载完成后再使用即时聊天工具聊天;(2)多 IP 监控,每个用户的网络访问行为在一个时间段内是单一的;(3)多 IP 监控,每个用户的网络访问是单行为或多行为交叉的。每种情形的测试时间均为 1 小时,监控范围是 3 个计算机实验室,从监控响应正确率、误报率、漏报率等指标对触发响应模块进行评价。

当测试完成时,我们逐一核对响应警报,对所有数据包都进行了扫描,统计正确率、误报率和漏报率。3 次测试的总体情况如表 1 所示。

表 1 触发响应模块测试结果

情形	监控子网段	同时监控 IP 数量	平均每秒数据包总数	正确率(%)	误报率(%)	漏报率(%)
1	1	1	58.6	88.37	11.63	25.97
2	3	16	1089.2	87.16	12.84	32.58
3	3	16	1206.5	73.59	26.41	39.81

从表 1 所示的实验测试结果来看,触发响应的正确率是较高的。在情况 1 和情况 2 中,监控系统一旦触发了一级警报锁定用户行为,经过进一步的分析之后,基本上能识别用户网络行为属于哪类违规操作。对于情况较复杂的情况 3,其正确率相对于情况 1 和情况 2 略低一些。这说明触发响应模块的特点有二:一是可利用聚类分析结果分析实时数据,触发系统的警报和响应,该模块是建立在数据分析基础上的可视化模块;二是该模块的二级响应体系能够根据网络访问行为的少量典型特征触发一级警报,再启动精细的数据分析,较精确判断用户行为。该模块具有的技术特点,使系统能够有针对性地缩小范围,迅速锁定可疑的网络行为,重点分析可疑网络行为,判断是否为违规的网络访问。

实时网络数据的特征与聚类中心的距离比对值表明实时网络行为的特征与聚类类别的相似程度,当两者的相似程度在一定范围内,监控系统的触发响应模块做出响应。从实验测试的 3 种情形(表 1)可以看出,系统的漏报率均在 40%以下,触发响应模块能够识别大多数网络行为是否为违规行为以及属于哪一类行为。因此,触发式网络监控系统对于网络行为的监控是比较有效的。

在测试过程中,系统的触发响应模块对用户行为响应后,判断用户进行了哪种类型操作的准确率是较高的,误报率较低。然而,用户的网络行为往往不是单一的,多种网络访问同时进行,在这种情况下,较难获取用户某一行为单一的网络特征,使系统判断偶尔出现偏差,因此,在测试中系统的漏报率还较高。

5 结束语

本文在比较当前研究现状的基础上,就网络监控中的监控模型、触发响应机制等问题进行了研究,设计实现了一种基于聚类的触发式网络监控的结构模型。该模型将系统划分为网络用户层、实时监控层、数据分析层、触发响应层、前端应用管理层,共 5 个层次,可以各自承担数据处理和传输存储任务,较好地适应了网络监控的环境和特点,提高了整体运行效率。

参考文献:

[1] Zhou Wanlei, Yang Xiang. Network and system security [J]. Journal of Network and Computer Applications, 2009, 32(2): 345-346.

[2] 肖敏,韩继军,肖德宝,等. 基于聚类的入侵检测研究综述[J]. 计算机应用,2008,21(s1):34-38,42.

[3] Safaa O Al-Mamory, Zhang Hongli. Intrusion detection alarms reduction using root cause analysis and clustering [J]. Computer Communications, 2009, 32(2): 419-430.

[4] 王勇,王洁,王明华,等. 基于序列聚类的事件流数据特征分析[J]. 计算机工程,2008,34(12):34-36.

[5] 孙美凤,彭艳兵,龚俭,等. 自然着色聚类过程中的网络安全事件计算[J]. 计算机学报,2007,30(10):1787-1797.

[6] 韩宗芬,杨志玲,储杰,等. 一种用于网络安全系统的报警聚类与关联模型[J]. 计算机工程与科学,2005,27(10):8-9.

[7] 王振宇,邓锦福. 基于 Netfilter 的 DNS 实时监控统计系统的设计[J]. 计算机应用研究,2009,26(4):1487-1490.

[8] 包加桐,唐鸿儒. 基于智能体架构的网络行为监控集成系统[J]. 计算机应用与软件,2008,25(11):274-277.

(责任编辑:邓大玉)