分布式入侵检测中负载平衡的应用分析与设计

玄加林[1],才书训[2]
(1.东北大学 软件学院,辽宁 沈阳 110004;
2.东北大学 秦皇岛分校,河北 秦皇岛 066004)

摘 要:网络数据传输的随机性导致传统的分布式入侵检测系统各监测点数据处理的不平衡,进而影响系统的整体性能甚至产生检测问题。为解决以上问题,通过对现有的入侵检测系统问题的分析,将负载平衡的思想引入其中并与分布式思想相结合,提出了一种提高入侵检测系统性能的系统结构设计方案。并根据该设计思想实现了一个系统原型。仿真测试表明改进后的设计在丢包率和系统整体性能等方面有较大提高。

关键词:入侵检测系统;系统结构;负载平衡;心跳
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1005-3751(2006)01-0213-04

Application Analysis and Design of Loading Balance in Distributed Intrusion Detection System

XUAN Jia-lin[1], CAI Shu-xun[2]
(1. Software College, Northeastern University, Shenyang 110004, China;
2. Qinhuangdao Branch, Northeastern University, Qinhuangdao 066004, China)

Abstract: Because of the randomness of network data delivery, result in the load is unbalanced in every nodes of the traditional distributed intrusion detection system, then affect the function of the whole system even produce examination problem. Pass to analysis the existing intrusion detection system, import the loading balance and make some improvement on the structure of the system. An archetypal system is made, which improves the rate of throw data package and capability of whole system.

Key words: intrusion detection system ; system structure; loading balance; heartbeat

0 引 言

计算机网络技术的发展与应用使得普通人使用计算机完成各种工作成为可能,同时也使得非法入侵者通过不正当手段获得重要数据的可能性越来越大。随着计算机网络的普及,各种黑客技术层出不穷,攻击事件时有发生,因此如何保证数据的安全性进而保障正常工作的顺利进行成为计算机及网络的发展必须要面对的问题。

1 入侵检测系统的概念及作用

针对不断出现的网络攻击行为,通过多年的研究人们提出了多种不同的网络安全技术,这些技术在现实的应用中起到了重大的作用。传统的网络安全技术主要包括:加密和数字签名机制、身份认证与访问控制机制、认证授权、安全审计、系统脆弱性检测、构筑防火墙系统等等。这些技术都发展得比较成熟,特别是防火墙技术,它能有效地控制内部网络与外部网络之间的访问及数据传送,从而达到保护内部网络的信息不受外部非授权用户的访问和过滤信息的目的,防火墙配置的多样性和防护的有效性使它成为网络安全防线的中流砥柱。然而据统计,全球80%以上的入侵来自于内部[1]。防火墙访问控制可以拒绝未授权用户的访问,却并不能防止已授权访问的用户获取系统中未授权信息;防火墙可以将危险挡在外面,却无法挡住内部的入侵。人们迫切需要一种能解决以上问题,可以主动地去检测、发现和排除安全隐患的网络安全保护措施。因此入侵检测系统作为防火墙后的第二道防线开始崭露头角,成为了安全市场上和研究上新的热点,不仅愈来愈多地受到人们的关注,而且已经开始在各种不同的环境中发挥越来越重要的作用。

入侵检测是对入侵行为的发觉。它通过对计算机网络或计算机系统中的若干关键点收集信息并对其进行分析,从中发现网络或系统中是否有违反安全策略的行为和被攻击的迹象。进行入侵检测的软件与硬件的组合便是入侵检测系统(Intrusion Detection System, IDS)。与其他安全产品不同的是,入侵检测系统需要更多的智能,它必须能将得到的数据进行分析,并得出有用的结果。一个合

收稿日期:2005-04-20
作者简介:玄加林(1981—),男,山东泰安人,硕士研究生,研究方向为计算机安全;才书训,教授,研究方向为计算机安全、数据挖掘、决策支持。

格的入侵检测系统能大大地简化管理员的工作,保证网络的安全运行。它是网络安全技术中不可或缺的一部分,也是对其他安全技术的一个补充。

2 分布式入侵检测系统的一般架构及存在问题

根据入侵检测系统的系统结构框架可以将其分为集中式入侵检测系统和分布式入侵检测系统。

2.1 集中式入侵检测系统

系统通过从主机的网络设备或者缓冲区中提取数据或者通过监听网络数据包获得检测的数据,然后在系统所在的固定的检测点对数据进行处理。在这种体系结构中,系统作为一个单一模块或多个同种结构功能的模块交互运行。

2.2 分布式入侵检测系统

由不同的实体组成,分布在系统的每一个实体都执行自己的任务,各实体之间通过消息或其他机制进行交互。不同实体可能位于不同的物理主机上,也可能在同一主机上。

由于集中式的入侵检测系统的数据分析都集中在一个检测点上进行,因此对检测点的压力较大,当数据量较大时要求检测点要有较快处理速度和较大的存储能力,否则就会由于检测点不能及时处理而产生大量丢包,导致入侵行为不能被有效检测,入侵检测系统也就形同虚设[2]。而分布式入侵检测系统的系统结构恰好能够解决以上的问题,因此成为了目前研究的热点。

分布式入侵检测系统一般采用一个检测中心的模式[3]。即在被检测网络中设置多个网络或主机的检测点,这些检测点固定地对负责的目标网络或主机进行检测,收集目标的数据,并加以初期简单的处理和分析。然后以一种统一的格式发送给检测中心主机,由检测主机对这些数据进行分析,确定采取何种响应措施。这样的系统结构分布性较好,网络整体的检测负担能够有效地分配到不同的检测点上,可以对各检测点进行动态的管理和配置,基于网络的检测点和基于主机的检测点可以互不影响的工作,由于两种检测方式发给检测中心的数据有相同的格式,因此检测中心可以统一进行分析。

但是这样的结构同样也存在一些问题[4]。

首先单点失效的问题特别突出,由于数据最终要发送到检测中心进行处理,因此整个系统的可用性决定于检测中心的可用性,如果检测中心受到安全威胁而失效,则整个检测系统随即瘫痪。

其次,由于检测点所收集到的数据都要发送给检测中心进行处理,因此当检测点较多且数据流量较大时检测数据的传输会增加网络的负担。这直接违背了检测系统的引入不应增加原有系统负担的原则。

再次,现有分布式入侵检测系统的各个检测点的检测目标由管理员固定设置,因此增加了管理员的负担,同时由于网络中不同网段之间的数据流量不同,就存在各检测点之间工作量严重不平衡,部分检测点要应付大量数据甚至不得不丢包,而另一部分检测点处于闲置状态的问题。

3 改进的分布式入侵检测系统结构

为了解决以上分布式入侵检测系统存在的问题,这里提出了一种分布式入侵检测系统框架结构的改进构想。整体结构如图 1 所示。

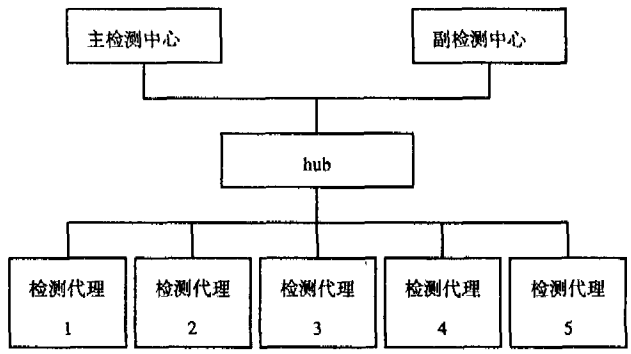

图 1 分布式入侵检测系统框架结构图

首先,系统仍然采用检测中心加检测代理的整体结构,但是检测中心不再进行一般数据的分析,而是成为了整个系统正常运行的协调和控制者。同时提高各检测代理点的级别,使检测代理点不仅具有抓取数据包的功能而且能够对数据包进行检测和响应,这样代理点不必再将收集到的网络数据全部发送到检测中心从而降低了网络数据的传输压力,同时使检测代理点在进入正常工作状态后即使脱离检测中心也能独立运行,使得整个系统不会因为检测中心的故障而瘫痪。

其次,为了防止检测中心的单点失效,采取检测中心冗余设置的方式,在整个系统中设置主检测中心和副检测中心。每个检测中心独立运行,两个检测中心之间定时发送心跳信息,这种信息告诉对方本身处于有效状态,当两个检测中心都正常时,只有主检测中心对系统整体进行控制,而副检测中心仅做与主检测中心同样的数据处理而不真正对系统进行控制。当主检测中心发现副检测中心未能正常工作时向管理员报警,而当副检测中心发现主检测中心未能正常工作时则报警并接替主检测中心的工作,以维持系统的正常运行。

第三,利用控制中心与系统所有的检测代理的交互能力,对整个系统进行系统配置检测库升级等的集中控制,从而保证了系统整体检测能力的一致性,同时也减轻了管理员的工作负担。

第四,采用中心汇总检测的策略,将各代理上发现的有潜在入侵威胁的信息汇总到监控中心进行集中检测。例如,现有的大部分入侵检测系统如果发现同一个外部主机试图连接本地的同一个主机上的多个端口则认为外部主机在对本地主机进行扫描。入侵者通常采用对本地主机上的端口进行小分组,然后通过双循环方式对多台主机的相同的组进行顺序扫描的方式躲避检测系统的检测[5](如图 2 所示)。

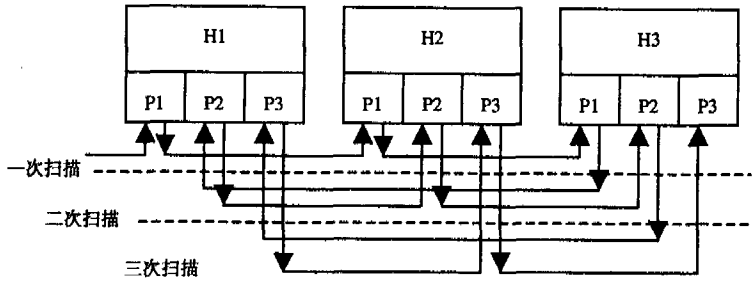

图2　端口扫描示意图

为了对这种扫描方式进行有效的检测,本设计采取如下策略。首先设置一个确认扫描阈值和一个潜在扫描阈值。代理点如果发现远程主机对本地主机的连接大于确认扫描阈值则报警,如果连结数大于潜在扫描阈值而小于扫描阈值则将这部分连接信息发送到检测中心,由检测中心对多个代理点发来的数据进行汇总,如果发现同一外部主机对网内多台主机的部分相同端口进行扫描,则认为入侵行为发生。

最后引入了负载平衡机制,使得在各检测代理上所负责检测的数据不是固定的,而是根据整个网络的状态和不同检测代理的运行状态由控制中心进行动态的分配,使得各代理上的处理负载基本平衡,从而保证了系统整体处于最好的工作状态。同时,负载平衡也附带完成检测代理点故障处置的功能,最大限度地保证了系统的安全性。

4　负载平衡的概念及作用

所谓负载平衡是对集群各节点数据处理的合理分配策略。随着集群技术的应用以及大型网络的发展网络传输速度的不断加快,随之而来的问题就是如何协调一个机群系统中不同的节点之间的负载:当一个新的任务来的时候,如何让负载较轻的节点去处理这个任务;当系统中的一个节点由于某种原因出现崩溃时,如何让别的节点接受它的工作,并保证服务的连续性,不丢失任何数据。由此,负载平衡技术就被提到了一个很重要的地位。

负载平衡技术的主要思想就是如何根据某种算法将网络的业务量平均分配到不同的服务器和网络设备上去,以减轻单台服务器和网络设备的负担,从而提高整个系统的利用率和安全性。

现有的负载平衡究其策略而言可以分为两种[6]:

(1)自主控制策略。

这种策略的主要思想是:由空闲节点逐个向邻接节点请求任务,如果请求到任务,则中止请求,否则就继续询问下一个节点。也有可能所有相邻节点都没有满足请求,请求节点就等待,过一段时间后再向相邻节点发出任务请求。

主要缺点:开始和结束阶段时任务数相对较少,许多任务请求会延迟忙节点的执行。

(2)中央控制策略。

这种策略的主要思想是:节点间的任务调节分配由中心的控制者执行,至于分配给哪个邻接节点则主要取决于节点的负载状态,因此,该策略需要交换处理器的负载信息。

在各种策略中定位每次平衡调度的源节点或目的节点时,可以选择首次适应算法或最佳适应算法。前者只要找到满足平衡调度条件某个节点即可,不需要检查所有邻接节点;而后者每次总检查比较所有邻接节点,挑选一个最佳的节点来完成这次负载调度。

5　入侵检测中负载平衡的实现策略

在入侵检测系统中引入负载平衡策略的目的是:

a.能够使不同的检测代理点的工作量相对平衡,从而使整个系统的性能最优;

b.当有的检测代理出现故障时能够让其他的代理点接收它的工作。

本设计采取中央控制加自主控制的策略,使两种策略相互补充,吸取了各种策略的优点,同时尽量避免不利的影响。由于有中央控制节点负责控制,因此在定位平衡调度的源节点和目标节点方面采用最佳适应算法,从而使系统平衡后性能最优。

具体平衡策略如下:

(1)检测中心维护一个检测代理点和其检测范围(主机或网络)的分配表。当整个检测系统启动时,由于还无法得到整个系统的运行状态,因此检测中心只是简单地将系统所负担的网络以主机数量为指标平均分配到各活动的检测代理点上,并将分配结果记录入分配表。

(2)当有新的检测代理点加入系统时,检测中心运用平衡机制从原有各活动的检测代理点上收集一部分检测网段或主机分配给新代理点,并修改分配表中的数据。

(3)各活动的检测代理点会定时地发送心跳数据[7]给检测中心,心跳数据中包含检测代理点的编号和从上次心跳数据发送到本次发送期间代理点上所处理的数据包总数。检测中心收到各代理的心跳后就得到了两个信息:这个检测代理点处于有效活动状态;这个检测代理点上的数据处理负荷。检测中心可以根据不同检测代理点提供的处理数据包数判断各检测代理点的负载情况,从而为平衡提供数据依据。

考虑到主机数据流量的随机性,监控中心可以保存本次心跳之前的两次心跳发来的处理数据包数量情况,以最近一段时间代理点处理的数据量的加权和表征这个代理点的负载情况,这样就降低了网络流量的不规则突变对统计信息的影响,保证了检测中心对代理点的负载量判断的准确性。

为了防止心跳数据的发送对网络正常数据传输的影响,对心跳数据发送的周期进行合理的设定,降低心跳数据所占带宽,从而使得系统的引入不增加网络的负担,同时统计信息也尽可能地反映了各代理点的负载情况。

(4)设定一个可接受负载变化范围的阈值。检测中心

收到各检测代理的心跳信息后,对比不同检测代理处理数据的总量,如果有个别代理点处理的数据量与全部代理点处理数据量的平均值差距超出了限定的差距阈值,则监控中心会在平均值两侧各找一个差距较大的代理点对其进行平衡(即最佳适应)。

平衡时以高负载点向低负载点让出部分负载的形式实现。平衡的最终目标是两个检测代理点的负载都接近或等于系统检测代理负载平均值。监控中心首先向两个平衡点各发送一个带有标志位的平衡点数据,标志位标示本代理点为高负载代理或低负载代理点,数据中包含了参与平衡的另一代理点的 IP 和整个系统各检测代理的负载平均值(即中央控制策略)。此后监控中心不再参与平衡活动,整个平衡活动以高负载的一方为中心进行(即自主控制策略)。高负载代理点从自身的负载中取出一部分负载,将这些负载的信息和在本检测点上存储的负载的历史统计信息传送给低负载点,由低负载检测代理点接收并开始检测工作。低负载点在启动对这部分的检测后发送完成信号给高负载检测点和检测中心,收到完成信号后,高负载点正式中止对这部分的检测,检测中心则修改分配表中的信息。

在平衡过程中,如果高负载点让出大数据量的被检测负载则需要传输的负载信息一般较大,而如果让出数据量较小的负载则可能会造成两检测代理点的负载点数量相差较大。因此应该从系统负载平均值附近选择适量的负载点。同时为了完成以上的工作各检测代理点应该维护一个记录被检测目标与其数据量的值的对应表(见表 1)。

表 1 被检测点数据量纪录表

被检测目标 1	数据量
被检测目标 2	数据量
被检测目标 3	数据量
……	……

(5)如果在一个心跳周期后检测中心未能收到检测代理点发来的心跳信息,则认为检测代理点未能正常工作,此时检测中心根据分配表中的数据将故障代理的负载分配到其他的有效检测点上去,同时修改分配表并且向管理员报警,从而保证了系统整体的有效性和安全性。

(6)为保证系统内部的数据传输高效性,将系统各部分传输的数据设置为统一格式[8](如图 3 所示),以数据类型标志位来区分不同的数据。

底层数据包首部	数据类型标志位	数据

图 3 系统数据格式

6 总 结

通过对以上策略进行有针对性的实现并在中型网络上进行测试实验,在试验过程中合理选取各种阈值,发现以上的设计对分布式入侵检测系统的整体性能有明显提高,同时也提高了系统整体的安全性。

参考文献:

[1] 张世永.网络信息安全技术[M].北京:科学出版社,2003.

[2] 张铭来,金成飚,赵文耘.网络型入侵检测系统存在的漏洞及其对策研究[J].计算机工程,2002,28(1):172-174.

[3] 史美林,钱 俊,董永乐.入侵检测技术与其发展趋势[R].北京:清华大学计算机系 CSCW 实验室,2002.

[4] 纪祥敏,连一峰,许晓利.入侵检测技术的研究与进展[J].计算机仿真,2004,21(11):129-132.

[5] 王晓程,刘恩德.攻击分类与分布式网络入侵检测系统[J]计算机研究与发展,2001(6):17-20.

[6] 陈华平.分布式动态负载平衡调度的一个通用模型[J]软件学报,1998,9(1):25-29.

[7] 唐正军.网络入侵检测系统的设计与实现[M].北京:电子工业出版社,2002.

[8] Comer D E. Internetworking With TCP/IP[M].北京:电子工业出版社,2003.

(上接第 209 页)

服务,给客户更加个性化的界面,开展有针对性的电子商务,以更好地满足访问者的需求。文中主要介绍了智能搜索引擎系统、网站用户访问分析系统和个性化推荐系统 3 种不同特点的电子商务 Web 挖掘系统。但是,Web 挖掘系统的应用还处于比较初级的阶段,如何让 Web 挖掘系统有更高的智能,还有很多问题需要解决。

参考文献:

[1] 王玉珍 .Web 使用模式挖掘在电子商务中的应用[J].计算机应用研究, 2003,10:155-157.

[2] 李 岩,陈新中,杨炳儒 .基于 Web 挖掘的智能门户搜索引擎的研究[J].计算机工程与应用,2002,38(4):34-36.

[3] 鲍玉斌,王大玲,于 戈 .关联规则和聚类分析在个性化推荐中的应用[J].东北大学学报(自然科学版),2003,24(12):1149-1152.

[4] 曾 春,邢春晓,周立柱 .个性化服务技术综述[J].软件学报,2002,13(10):1952-1961.

[5] Mobsher B. WebPersonalizer: a server side recommender system based on web usage mining[EB/OL]. http://www.cs.depaul.edu/research/technical.asp, 2001.

[6] Asnicar F, Tasso C. ifWeb: a prototype of user model based intelligent agent for documentation filtering and navigation in the World Wide Web[A]. In: Tasso C, Jameson A, Paris C L. Proceedings of the UM 1997 Workshop on Adaptive Systems and User Modeling on the World Wide Web[C]. West Newton, MA: User Modeling Inc, 1997. 3-12.