

基于流负载均衡的入侵检测系统

陆磊, 王 锋

(昆明理工大学, 云南 昆明 650051)

摘 要:网络带宽的飞速发展对入侵检测系统的性能提出了更高的要求,单机的网络入侵检测系统性能已跟不上发展的需要。在分析了现有分布式入侵检测系统的基础上,构建了一个高速网络环境下基于负载分发的入侵检测集群系统结构。该集群系统采用对 IP 数据包流识别信息进行 Hash 的方法对负载进行分发,同时引入 Agent 技术中的通信代理实现各检测节点之间的数据共享和交换,解决了检测集群各节点的协作和信息共享。对集群系统进行的测试说明其实现了多机集群的入侵检测系统。

关键词:入侵检测;负载均衡;集群;Hash

中图分类号:TP393.08

文献标识码:A

文章编号:1673-629X(2009)11-0135-04

Intrusion Detection System Based on Flow Load Balance

LU Lei, WANG Feng

(Kunming University of Science and Technology, Kunming 650051, China)

Abstract: With the fast development of network bandwidth, there is a higher requirement for the performance of the intrusion detection system. Single NIDS is facing the problem that the detection of packets processing capacity cannot adaptive to the development of network bandwidth. Based on the study of existing distribute intrusion detection system, present an NIDS cluster architecture as a scalable solution for realizing high-performance. The cluster hashing IP package flow identifier to distribute the load, and uses communication agent from agent technology to share data between detection nodes. Finally the test reveals that the cluster achieved multi-machine intrusion detection system.

Key words: intrusion detection; load balance; cluster; Hash

0 引 言

随着高速互联网络的发展,网络带宽的提高对IDS(入侵检测系统)的性能要求也越来越高。现有的入侵检测系统性能的提高更多的是依赖于硬件性能的提高。应对现今动辄以 Gbps 带宽来衡量的网络数据流量的监测,如果使用单一节点的入侵检测,对数据包的处理能力已不能满足需要,提高 IDS 的性能已迫在眉睫。文中研究通过多机集群来提高 IDS 的性能。

1 相关工作

为了实现高速入侵检测,有的 IDS 关注于系统级别的优化,比如系统的中断联合、IDS 规则库的裁剪等,有的研究^[1,2]关注入侵检测算法的优化和改进。还有的研究^[3]则关注了如何将高带宽的数据流分发到各个入侵检测节点。

现今高速网络环境下的入侵检测集群系统已经很多,它们在设计时考虑的重点各有不同,有些系统基于 Agent 技术来实现分布式的入侵检测,例如文献[4, 5]中提到的系统。这样的系统通过各种类型的 Agent 各自工作与相互间的信息交换使整个集群系统协同工作。而另一种基于 SATA^[6]的系统中,如果其中一个入侵检测节点不能完整地检测到攻击,NetSTAT(SATA 的子部分)将传送部分设定好的包含状态信息的数据到别的检测节点。SATA 工具集是一个基于误用的分布式入侵检测工具集。Prelude^[7]则是一个依赖 IDMEF 标准来交换事件的分布式的 NIDS(网络入侵检测系统),在其系统中,检测节点连接到管理节点上,管理节点负责处理和综合报警信息。管理节点同时可以承担中继服务器的任务,并向一个中央服务器报告信息。

2 集群系统结构设计

在研究相关系统的基础上,设计了入侵检测集群系统要达到的目标:(1)透明:给使用者是在使用单一

网络入侵检测系统的感觉。(2)可扩展性:可以轻松地增加分析节点来适应负载的需要。(3)经济性:要让系统具有很高灵活性并且在经济的硬件平台上实现。(4)易用:操作员可在一台主机上对入侵检测集群进行操作,实现报警日志的综合分析以及调整系统的配置。(5)易于维护:对失效的入侵检测集群的部分,集群系统应该能够不影响其他部分的情况下很快替换掉。

为了达到以上目标,对入侵检测集群系统作了如下系统结构设计(见图 1),包括:前端分发节点,分析节点群,通信代理和一个中央管理节点。

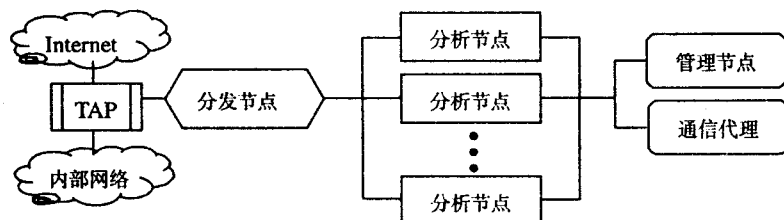


图 1 IDS 集群系统结构

前端分发节点将被监控的网络数据流分发到分析节点群;分析节点群运行着 NIDS,进行流量数据的分析工作,并通过通信代理节点交换信息;管理节点则是整个集群系统的主控制台;分析节点运行于 PC 上,每一个分析节点配有两块网卡,其中的一块负责接收前端分发节点分发的待处理的网络数据流,另一块则用于连接控制节点以及通信代理。每一个分析节点上都装有一个 NIDS,负责分析它们获得的那一部分网络数据流,分析结果是一系列的告警事件,在分析过程中需要全局数据进行分析的时候,它们可以通过通信代理来获取其它 IDS 分析节点上的信息。所有分析节点对数据流的检测都是一致的,它们的检测规则将通过管理节点来分发。

通信代理节点是一个分析节点间交换信息的中转站,它避免了分析节点间直接通信时占用大量网络带宽的现象。

管理节点则是整个入侵检测集群的操作控制终端,它负责综合过滤各分析节点的分析结果,将分析节点的告警事件通过事件解释器产生全局报警日志。同时,它还负责整个集群系统的管理工作,比如新的 IDS 规则的配置和开关某个集群中的节点。

3 负载分发机制

3.1 分发策略

对前端分发节点而言,分发数据流最常见的分发方法^[8]就是基于 IP 数据包的轮循机制,这样每个分析节点分配到同样多的 IP 数据包。但目前多数 NIDS

对流量的检测都是基于一个连接流,而不是基于数据包本身。所以,这样的分发方法并不适合 IDS 集群系统。

基于流的分发策略则是将属于同一个流的 IP 数据包发给相同的分析节点。例如,在基于特征匹配的传统 Snort 入侵检测系统中,它要求被检测的数据必须保证流的完整性,以使其能检测到攻击。并且 IDS 的资源消耗与其所分析的流的数量有关,这样,基于流的分发策略达到了将负载均衡地分发到各 IDS 节点的目的。

为了让分发节点更加简单高效,直接对流识别信息进行 Hash 运算,把 IP 数据包分发到 $\{0, \dots, n-1\}$ 个集合中(n 为分析节点的数量)。不论是 TCP 还是 UDP,流识别信息都是(addr1, port1, addr2, port2)这个 4 元组,利用 Hash 算法对 4 元组的值运算后取模,可以决定 IP 包发往哪个节点。首先,试验了文献[8]中表现较好的 CRC16

算法来进行 Hash 值的运算:

$$h_{CRC}(addr1, port1, addr2, port2) := CRC16 \\ (addr1 + port1 + addr2 + port2) \bmod n$$

利用这个 Hash 算法,可以将一个连接流的所有 IP 数据包交付到同一个分析节点。

设 N_i 为第 i 个时间间隔中数据流的数量。一个理想的分发策略应该将每个节点所分到的流的数量 M_i 满足 $M_i := N_i/n$ 。测试中 $n = 3$ (即 3 个节点),分发节点应用上述 Hash 方法,对预先获取的 6 小时的流量进行测试,取样时间间隔为 5 分钟,图 2 表示的是分析节点间获得流数量的差异。可以看到,以上所述的 h_{CRC} 算法能将流量很好地分配到各节点,与理想情况的标准差 σ 仅为 0.41%。

考虑到对每个包的 4 元组进行 CRC16 计算十分消耗系统资源,测试了另一 Hash 运算:

$$h_{da}(addr1, port1, addr2, port2) := (addr1 + port1 \\ + addr2 + port2) \bmod n$$

这个算法得出的 σ 值为 1.58%,节点间流量差异仍然很小,不会带来明显的负载不均。接着考虑到,对 4 元组进行 Hash 的方法提供了很好的负载分流,但也有些缺点,首先,这样的 Hash 中有端口号,而端口号不一定存在于每个 IP 包中(例如 ICMP 包)。不是每个以太网帧中都有端口信息,所以 4 元组的 Hash 算法不能处理以太网帧。其次,这样的运算需要对帧进行重组,消耗了时间和空间。随后测试了只依赖源地址和目的地址的 Hash 算法,进行测试:

$$h_{CRC_2}(\text{addr1}, \text{addr2}) := CRC16(\text{addr1} + \text{addr2}) \bmod n$$

这样的算法会将两个主机间的所有流量交付到同一 IDS 分析节点来分析,测试结果同样在图 2 中, h_{CRC_2} 的 σ 值为 1.67。

使用 2 元组的 Hash 方式还有一个好处,它减少了检测两个主机通信数据需要的信息交换。例如,端口扫描这类攻击的检测需要的是两个主机间的全部数据包,而只依赖于地址的 Hash 方法可以把这些包分发到同一分析节点上,这样,只需要单一的分析节点就可以完成这一类入侵的检测。在我们的 IDS 集群上,使用了对 2 元组 Hash 的方式。

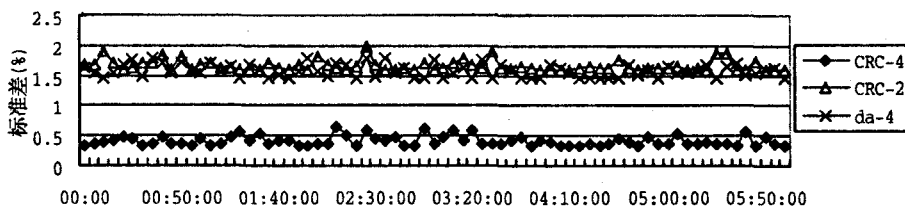


图 2 Hash 分发策略测试

3.2 前端节点

在前端节点的部署中,设计了两种方式可以根据需要来选择:

1) 使用专用硬件。

在分发节点的选择上,可以使用专用的硬件产品来实现,例如利用 FPGA 来实现高带宽网络数据处理的硬件,可以对 FPGA 编程,使其实现我们的 Hash 算法,并实时地对数据包的目的 MAC 地址重写,发送到我们的 IDS 分析节点上。

2) 使用软件的方式。

这种方式所提供的性能要比使用专用硬件低,但软件前端分发节点的性能能满足分发 2Gbps 带宽数据的需要(受限于计算机的硬件性能),选择了 Click^[9]作为分发节点的软件路由,它完全运行于 Linux 内核,并且同样能改写数据包的目的地址。

3.3 失效恢复机制

该入侵检测系统的设计目标之一就是易于维护,所以,在集群系统中,当一个节点失效的时候,对失效节点的恢复工作是简单和高效的。理想情况下,系统应该能自动处理失效节点,不影响其他节点的正常工作。对于前端分发节点,要做到以上的目标必须使用热备份的硬件。对于其他节点的失效处理可以按以下的办法来实现。

分析节点的失效,最直接的影响是其所分析中的那一部分数据将不会被检测。为了解决分析节点的失效问题,在系统配置中,将运行一个附加的热备份节

点。在通常情况下,这个备份节点和其他分析节点的配置是一样的,但备份节点不接受任何数据。当集群系统检测到有分析节点失效时(可以采用心跳检测机制来实现),备份节点将会更改自己的 MAC 地址来取代实效节点,并继续着失效节点的数据分析工作。这个时候,备份节点并没有失效节点的状态信息,需要通过通信代理节点来获得整个集群系统的全局共享信息。

当通信代理节点实效时,与其连接的分析节点在分析数据时将不同其他分析节点交互信息,同样,热备份的通信代理节点将自动地替换实效节点的工作,并连接到所有受影响的分析节点上,与它们同步各分析节点的状态。

当中央管理节点实效时,入侵检测集群会失去其报告和日志记录的功能。但在后台,分析节点和通信代理节点仍将不受影响地继续

运行。尽管在中央管理节点实效阶段入侵日志记录等信息不再汇总到中央管理节点,但各个分析节点仍然可以在本地进行日志记录。同样的,一旦检测到中央管理节点的失效现象,备份的节点将立刻启用,连接分析节点并汇总它们的分析结果。

4 实验与分析

为了比较在不同系统配置环境下的实验结果,实验中使用了一段预先在网络上截取的流量(40GB),所有节点使用的计算机配置为 Intel Pentium 4 2.6GHz, 1 GB RAM。集群系统中,把一个节点配置为中央控制节点,同时也将其配置为通信代理节点,这样,这个节点同时担负管理与内部通信的任务。同时,还配置了 3 个分析节点。

理想情况下,集群系统得出的结果应该和单一节点 IDS 的结果一致。对集群系统和单一节点系统的入侵日志和报警信息进行了比较。得出以下结果:两个系统的报警数量同为 856 个,说明我们的集群系统相较单一系统而言不会产生漏报。

接着,对各节点的 CPU 负载情况来检验前端节点对流量的负载均衡是否有效,并比较了不同类型节点间 CPU 资源的利用情况。如图 3,左图展示的是每个分析节点的 CPU 负载情况,右图展示的是不同类型节点间 CPU 的利用率情况。

实验结果表明我们的入侵检测集群系统能对负载进行有效的分发,并能得出正确的检测结果,满足应用的需求。

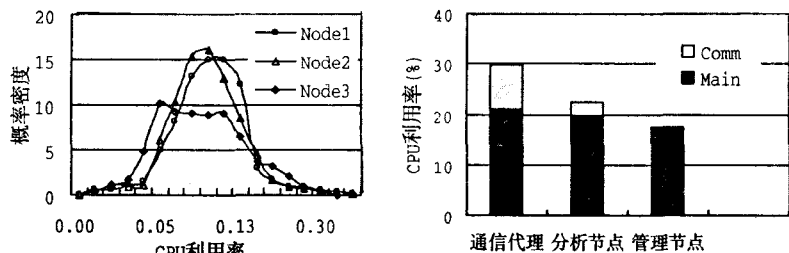


图 3 实验结果

5 结束语

给出了一个在普通硬件条件下建立的可扩展、高性能的网络入侵检测系统,该入侵检测系统由一个分发网络流量的前端分发节点及一个可扩展的分析节点群组成。每一个分析节点对其接收到的数据进行入侵检测,并在需要时和集群系统同步状态和检测信息。该集群系统对于使用者是透明的,感受如同使用单一入侵检测系统。

在集群系统的建立过程中,着重讨论了前端分发节点的多种流量分发策略,最后对该集群系统进行的测试说明其达到了透明、可扩展、易于维护、经济的目标。

对入侵检测系统分析算法的改进使其更加适合分布式的集群系统,减少分析节点间内部数据的交换,有待进一步地研究。

参考文献:

- [1] 甘学士,孙力娟.改进的模式匹配算法及在入侵检测中的应用[J].计算机技术与发展,2006,16(7):150-152.
- [2] 高志森,张铮,李俊.入侵检测中贝叶斯分类器改进的研究[J].计算机技术与发展,2006,16(11):154-155.
- [3] 杨武,方滨兴,云晓春,等.基于骨干网的并行集群入侵检测系统[J].哈尔滨工业大学学报,2004,36(3):273-276.
- [4] 张丹慧,佟振声.基于Agent与数据挖掘的分布式入侵检测系统[J].微机发展(现更名:计算机技术与发展),2004,14(3):125-126.
- [5] 徐长棣,刘方爱.基于P2P和移动代理的入侵检测系统研究[J].计算机技术与发展,2007,17(1):164-166.
- [6] Vigna G, Eckmann S T, Kemmerer R A. The STAT Tool Suite[C]//In:Proc. DARPA Information Survivability Conference and Exposition. [s.l.]:[s.n.],2000.
- [7] Blanc M, Oudot L, Glaume V. Global Intrusion Detection: Prelude Hybrid IDS[R]. [s.l.]:[s.n.],2003.
- [8] Zhiruo C, Zheng W, Zegura E. Performance of Hashing-Based Schemes for Internet Load Balancing[M]. Piscataway: [s.n.],2000:332-341.
- [9] Kohler E, Morris R, Chen B, et al. The Click modular router [J]. ACM Transactions on Computer Systems,2000,18(3):263-297.

(上接第 134 页)

表 3 误报率比较

Classifier	DOS	PRB	U2R	U2L	Average
PCA	0.033	0.028	0.032	0.032	0.031
LDA	0.039	0.038	0.035	0.032	0.036
FUSION	0.017	0.028	0.025	0.018	0.011

4 结束语

通过在决策级融合单分类器的检测结果,提高了检测率,同时降低了误报率。进行特征级融合是今后的研究方向。

参考文献:

- [1] Labid K, Venuri V R. Application of Principal Component Analysis to the Detection and Visualization of Computer Network Attacks[M]//Annals of Telecommunications. France: [s.n.],2005.
- [2] Shyu M L, Chen S C, Sarinapakorn K, et al. A Novel Anomaly Detection Scheme Based on Principal Component Classifier[C]//Proceedings of ICDM Foundation and New Direction of Data Mining workshop. [s.l.]:[s.n.],2003:172-179.

- [3] 王坤,潘继农,张鹏,等.基于主成分分析的异常检测方法研究[J].信息工程大学学报,2004,5(3):56-59.
- [4] 谷雨,邓锦辉,孙剑,等.基于独立成分分析和支持向量机的入侵检测方法[J].西安交通大学学报,2005,39(8):876-879.
- [5] 田捷,杨鑫.生物特征识别技术理论与应用[M].北京:电子工业出版社,2004.
- [6] Kuncheva L. A theoretical study on six classifier fusion strategies[J]. IEEE Trans on PAMI,2002,24(2):281-286.
- [7] Kitter J, Hatef M, Duin R, et al. On combining classifiers[J]. IEEE Trans on PAMI,1998,20(3):226-239.
- [8] David M J. Combining multiple classifiers by averaging or by multiplying[J]. Pattern Recognition,2000,33:1475-1485.
- [9] 王勇,王行愚,张瑞霞.基于D-S证据理论的分布式入侵检测方法研究[J].计算机工程与应用,2004(13):167-169.
- [10] 徐耀红.数据融合理论与应用[M].西安:西安电子科技大学出版社,1998.
- [11] 孙即祥.现代模式识别[M].长沙:国防科技大学出版社,2003.