文章编号:1008-0570(2007)05-3-0123-03

Snort报文捕获机制的研究与改进

The research and enhancement of Snort's packet capturing scheme

(鞍山科技大学网络信息中心)黄以卫 战学刚
HUANG YIWEI ZHAN XUEGANG

摘要:在对 Snort 系统的报文捕获机制进行研究的同时,针对 Snort 在高速网络上使用 libpcap 进行报文捕获的缺点提出了一种基于 Sample and hold 采样算法的改进机制,提高了 Snort 系统在高速网络链路上检测异常/入侵活动的效率。
关键字:入侵检测系统; Snort; 报文捕获; libpcap; Sample and hold
中图分类号:TP393.08 文献标识码:A

Abstract:This paper did some research on Snort's packet capturing scheme and proposed a enhanced sample algorithm based on Sample and hold theory due to its disadvantages on high speed network link. This algorithm greatly improved the efficiency of detecting anomalous or intrusive activities.
Key words:IDS, Snort, Packet capture, libpcap, Sample and hold

1 引言

随着互联网网速的飞速发展 Snort 系统中基于 libpcap 库的数据包监听解析方法已经不能满足高效实时检测网络异常/入侵活动的需要。鉴于此,作者对 Snort 的数据抓取及解析机制进行了详细研究并提出了一种更加高效的基于 Sample and hold 的数据包采样捕获机制。

2 Snort 简介

Snort 作为一个基于网络的入侵检测系统,其工作原理是检测网络的原始传输数据,分析捕获的报文,通过匹配入侵行为的特征或者从网络活动的角度检测异常行为。从检测模式而言,Snort 属于是误用检测。该系统对已知攻击的特征模式进行匹配,包括利用工作在网卡混杂模式下的监听功能被动地进行协议分析,以及对一系列报文解释分析特征。从本质上说,Snort 是基于规则检测的入侵检测工具,即针对每一种入侵行为,都提炼出它的特征值并按照规范写成检验规则,从而形成一个规则数据库。然后将捕获的报文按照规则库使用模式匹配算法进行匹配,若匹配成功,则认为该入侵行为成立。在中作者对入侵监测系统中的各种模式匹配的算法进行了详细的研究。

3 Snort 报文捕获的工作模式

Snort 报文捕获功能就是通过其报文捕获和解析子系统实现对网络中报文监听和解析的。包括以下几个主要步骤:(1)需要将网卡设置为混杂模式,混杂模式是指网络上的所有设备都对总线上传送的数据进行侦听,然后将一台计算机的网络连接设置为接收有以太网总线上的数据,实现网络信息包的捕获功能。

(2)调用函数 read()从数据链路层抓获一个包,抓包成功则进行过滤。过滤规则可以由管理员事先设定,将有用数据的包存储在报文缓冲区中,返回数据段的大小。其中,在过滤过程中,将显示被捕获的有用包的相关头信息,如源地址、目的地址、端口值等。过滤后的报文将进行进一步的协议解析、数据分析、登记日志文件等操作,实现管理员预期的各种功能。

4 Snort 报文捕获和报文解析的实现

Snort 中最基本的模块就是报文捕获和解析子系统,该子系统的功能为捕获网络的传输数据并按照 TCP/IP 协议的不同层次将报文进行解析。Snort 利用 libpcap 库函数进行采集数据,该库函数可以为应用程序提供直接从链路层捕获报文的接口函数,并可以设置报文的过滤器来捕获指定的数据。网络数据采集和解析机制是整个 NIDS 实现的基础。其中,最为关键的是要保证高的速度和低的丢包率,这不仅仅取决于软件的效率,还同硬件的处理能力相关。具体来说,报文捕获需要监听网络报文,进行 IP 重组和 TCP/UDP 协议分析,同时也要进行应用层协议数据流分析。一般来说,检测引擎所分析检测的数据都是基于某个网络协议层的数据信息,或是直接采用这些数据的某

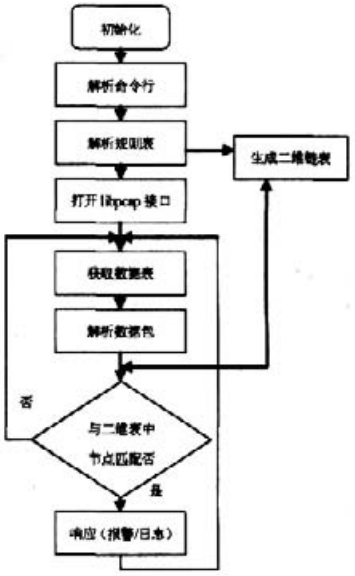

图 1 Snort 报文解析流程图

黄以卫:工程师 硕士研究生

些部分。因此在系统的设计中,除了捕获报文采集数据外,还要对这些信息进行协议分析。协议分析是指将网络上采集到的基于 IP 的数据进行处理,以便得到基于某种协议的数据。libpcap 在网上捕获到的是数据帧,系统还需要对数据帧进行协议分析,协议分析的处理过程为:首先根据预先定义的过滤规则从网络上获取所监听子网上的报文,然后进行 TCP/IP 协议栈由下至上的处理过程,主要是 IP 重组和 TCP/UDP 协议处理,最后进行应用层协议分析。流程图如图 1 所示。

5 Snort 报文捕获机理的优缺点

由于 Snort 采用在基于 libpcap 的网络数据包监听的捕获方法,而由于 libpcap 自身的特点决定了 Snort 只能使轻量级的入侵检测系统。具体表现在以下几个方面。

①在 pcap_loop()的每次循环中,首先通过调用 PacketReceivePacket()函数,从内核缓冲区中把一组数据包读取到用户缓冲区;然后,根据 bpf_hdr 结构提供的该数据包的定位信息,把用户缓冲区的多个数据包逐个提取出来,并依次送入回调函数进行进一步处理。由于用户缓冲区对数据包的处理方式过于复杂,当网络流量比较大时,用户缓冲区处理数据包的速度跟不上内核缓冲区从网卡复制数据的速度,所以新的数据包就会因为内核缓冲区满而被丢弃。

②由于网络带宽的高速发展以及监听系统主机自身的性能也影响到该技术的正常应用。举例来说,网络链路的带宽为 1Gbps 而监听系统的网卡只有 100Mbps,系统抓包的速度跟上数据传输的速度,必然会丢弃很多有用的数据包。这样的系统抓取到的数据肯定会影响到后续的数据解析以及规则匹配过程,从而失去准确性。

5.1 改进 Snort 报文捕获效率的方法

鉴于上述数据包检测机制存在的问题,许多研究成果也提出了一些改进方法。如:使用 NAPI 技术提高数据包捕获的效率、FHW 方法、使用高性能的 Asic、NP 网络处理器以提高报文捕获性能等等。这些方法都在一定的程度上提高了 snort 系统检测异常及网络入侵的效果,但是在应对高速发展的网络带宽方面这些方法还是存在着扩展方面的问题。

6 基于采样的数据包捕获技术的提出

随着网络速度的高速发展,互联网骨干链路的带宽已经达到了 10Gbps,也就是说每一分钟在骨干网络上将会有上百 G 的流量流过,这对基于 libpcap 的逐包捕获机制来说是不现实的。鉴于此,一些专家学者提出了使用采样原理来对网络流量进行捕获,然后对采样结果进行分析来预测网络流量总体行为特征的方法。

6.1 几种常见的采样方法

6.1.1 周期采样(Systematic sampling)

周期采样采用相同的时间间隔对网络数据包进行采集。该方法简单易行但也存在着缺点:如果采集的数据包本身具有周期性的行为,那么采样过程将仅仅得到周期性行为的一部分,这样会使得采样在较大程度上不能真实反映出周期性的采样行为容易产生较大的误差而引起测量失真。

6.1.2 随机附加采样与简单随机采样 (Random additive and simple random sampling)

随机附加采样是由分布函数 G(t) 随机生成采样间隔,采样的好坏取决于分布函数 G(t),这种采集方法具有无偏的特性,但缺点是频域(frequency domain)分析变得复杂;当 G(t) 为指数分布时,采样不可预测。简单随机采样只是简单地在整个网络流量集中采集 n 个数据包来进行分析。

6.1.3 均匀分层随机采样 (Uniform stratified random sampling)

均匀分层随机采样与周期采样方法相似,所不同的是周期采样只采样每个采样周期的第一个数据包,而该采样方法在每一个采样周期内随机采样一个数据包。

6.1.4 泊松采样(Poisson sample)

泊松采样适合任何随机到达过程,它是无偏的、非同步的. RFC2330 中介绍了三种泊松采样方法,方法一:产生 Gi,并等待 Gi 时间;执行一次测量,获取一个采样值. 重复该过程;方法二:产生 Gi,并等待 Gi 时间,执行一次测量,计算出测量所用时间 Mi;产生 Gi + 1,等待 Gi + 1 - Mi 的时间,执行一次测量,计算出测量所用时间 Mi + 1;以此类推;方法三:产生出 G1, G2 …, Gn, 计算出测量时刻序列 T1, T2, …, Tn. 其中, T1 = G1; T2 = G1 +G2; …, Tn = G1 + G2 + …+ Gn, 在每个时刻 Ti 上做一次测量。

几何采样以固定的概率来对数据包进行采集,例如,如果随机数字一律分布在 0 - 1 之间,且比给定概率 P 要小。那么,我们捕获某条链路上的所有包,但只对特定的包记录跟踪文件,几何采样是无偏且不可预测的。

6.1.5 Sample and hold 采样算法

1 算法描述

在文献中该算法的描述如下:建立一个用于统计 flow 的哈希表,哈希表中每一条目都有一个计数器,用于统计符合条件的 flow 的数据包总数。 刚开始的时候哈希表的所有表项及计数器全部置‘0’,以概率 p 对到来的数据包进行采样,如果到来的数据包没有被采样那么什么也没有发生。如果该数据包被采样了,那么就在哈希表中相应的表项记录下该数据包,计数器加‘1’。以后所有符合该 flow 特征的数据包到来时计数器都会进行计数统计操作。具体的操作流程如图 2 所示:

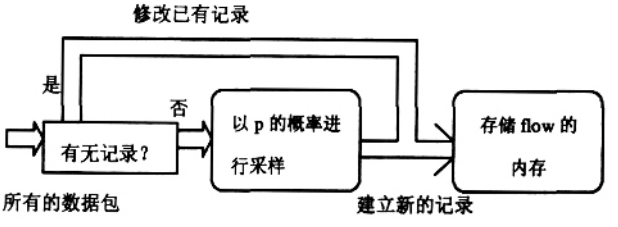

图 2 Sample and hold 采样流程图

2. Sample and hold 算法的效率

Cristian Estan 在中,对 Netflow 基于 Systematic sample 算法与 Sample and hold 采样算法的效率进行了比较,结果如表 1、2 所示。从这两个表中可以明显地看出 Sample and hold 采样算法的优点。

表 1 周期采样准确性分析

Error	1GB	100MB
1%	39.24%	79.03%
3%	1.07%	41.48%

表 2 Sample and hold 采样准确性分析

Error	1GB	100MB
1%	5.6E-32	0.09%
3%	1.8E-94	6.96E-10

6.2 Sample and hold 采样思想在 Snort 报文捕获中的应用

6.2.1 算法描述

在本文中，作者借鉴了 Sample and hold 采样的主要思想并在在做了一定的修改后把它应用到了 Snort 系统报文捕获机制中去。具体的实现如下：以流 flow 为单位建立一个哈希表，然后对于每一个到达的数据包：

①.检查属于这个 flow 的数据包在 hash 表中是否有记录。

②. 如果有，符合该 flow 条件的后续所有的数据包都会被捕获，数据包被捕获后送到 snort 报文解码模块进行报文分析，然后转到步骤 4.

③.如果没有，以概率 p 在 hash 表中建立记录. 到步骤 4.

④处理下一个到达的数据包，采样概率为 p。

6.2.2 选择 sample and hold 算法改进 Snort 报文捕获机制的理由

由于 snort 系统的另外一个重要的功能模块就是协议分析，而协议分析就需要对监测到的数据包进行报文重组，然后再与预设的特征库进行模式匹配以检测是否发生入侵行为，而上文中提到的随机、泊松、几何等采样算法采集到的数据包是随机的，因此在协议分析的时候所得到的信息是不全面的，协议分析也就无从谈起，更谈不上进行入侵活动的检测了。Sample and hold 的思想解决了这个问题，因为其算法的主要思想就是当采样到一个特定的 flow 数据包时其后续的通讯数据将都被采集，这样的话每一次采样都能得到一个完整的网络通讯报文，从而 Snort 系统的协议分析模块就能够对其进行报文重组和分析。

7 结论及改进

通过使用改进后的报文捕获算法，大大地提高了 Snort 系统检测入侵的效率，解决了一些机构部署 Snort 系统在高速网络上的难题。但是，由于采样算法本身精度的影响，该算法只是一个改进的算法。在以后的研究中应该继续对采样的精度上做进一步的研究，以提高 Snort 检测入侵行为与网络异常活动的精度。

创新点：Sample and hold 算法的原理是对采样到的数据及其后续符合同一特征的数据持续进行采集，直到没有符合该特征的数据为止。而 Snort 系统协议分析的特点就是对网络中的每一次通讯过程进行协议重组，这就需要该通讯过程中完整的报文数据，Sample and hold 原理正好与该协议分析的要求相符且 Sample and hold 算法是使用采样的方法来捕捉网络上的通讯数据包，这样就解决了 Snort 逐包采集分析无法适应高速网络监测入侵活动的问题，大大地提高了 Snort 在高速网络上检测入侵/异常活动的效率。

参考文献

[1] 李伟，鲁士文。Snort 数据包捕获性能的分析与改进 2005.22.7

[2]廖俊云,范明钰,王光卫。一种改进的基于 WinPcap 的快速抓包方法

[3]赵念强，鞠时光。入侵检测系统中模式匹配算法的研究[J]微计算机信息，2005.21. 8-3

[4]CHEN Thomas M，HU Lucia. Internet Performance Monitoring [DB/OL]. http://engr.smu.edu/tchen/papers/ProcIEEE-Aug2002.pdf

[5]CLAFFY Kimberly C， POLYSOS Gerorge C. Application of Sampling Methodologies to Network traffic characterization [DB/OL]http://portal.acm.org/affiliated/ft_gateway.cfm?id=166256&type=pdf&coll=ACM&dl=ACM&CFID=15151515&CFTOKEN=6184618

[6]ESTAN Cristian，VARGHESE George. New directions in traffic measurement and accounting[DB/OL]. http://www.acm.org/sigs/sigcomm/sigcomm2002/papers/traffmeas.pdf

[7]PAXSON V，ALMES G，MAHDAVI J. RFC2330: Framework for IP Performance Metrics [DB/OL]. http://rfc.net/rfc2330.html

[8]DUFFIELD Nick. Sampling for Passive internet Measurement: A Review [DB/OL]. http://www.research.att.com/~duffield/papers/STS102.pdf

[9]ESTAN Cristian. Presentation: New Directions in Traffic Measurement and Accounting [DB/OL]. http://www.imconf.net/imw-2001/slides/71.Estan.ppt

作者简介：黄以卫(1980-)，男(汉族)，江苏泗阳人，鞍山科技大学网络信息中心工程师，硕士研究生，主要研究方向：网络流量监测，网络安全。

Biography: Huang Yi-wei (1980-), male (The han nationality), from siyang city of jiangsu province, network information center's engineer of anshan university of science and technology, postgraduate, main research fields: network traffic monitoring and network security.

(114044 辽宁 鞍山科技大学网络信息中心) 黄以卫 战学刚

通讯地址：(114044 辽宁 鞍山科技大学网络信息中心) 黄以卫

(收稿日期：2007.4.03)(修稿日期：2007.5.05)