一种基于 Hurst 指数的异常检测软件*)

任勋益[1] 王汝传[1,2] 蔡小华[1] 傅雷杨[1]

(南京邮电大学计算机学院 南京 210003)[1] (南京大学计算机软件新技术国家重点实验室 南京 210093)[2]

摘 要 根据 Hurst 指数变化发现异常是进行入侵检测的一种新思路。基于这样的思路,实现了一个异常检测软件。该软件通过数据包捕获,对数据进行特征提取,对特征进行时间序列划分,采用 R/S 和小波分析两种方法进行 Hurst 求解,最后根据求解的 Hurst 值判断异常。实际使用表明,基于 Hurst 指数的异常检测软件具有无需学习,检测快速的优点。

关键词 Hurst 指数,异常检测,软件

Software for Anomaly Detection Based on Hurst

REN Xun-yi[1] WANG Ru-chuan[1,2] Cai Xiao-hua[1] Fu Lei-yang[1]

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)[1]

(State Key Laboratory for Novel Software Technology Nanjing University, Nanjing 210093, China)[2]

Abstract Anomaly detection based on Hurst provids one new idea, based on which, this paper implemented one software for anomaly. The software captures data packets, extracts features, and implements time series division, adopts two methods--R/S and wavelet analysis for computing Hurst value, lastly, judgs whether anomaly happens based on the valve. The actual use of software proves that the anomaly detection based on Hurst does not need to learn and has advantage of efficient detection.

Keywords Hurst, Anomaly detection, Software

1 引言

异常检测关键问题在于如何定义一个精确判定模型,对此,很多学者研究神经网络和支持向量的推广性。这些研究期望获得一个良好的主观检测模型,然而有些学者从另一个角度研究入侵检测,他们希望通过发现系统本身蕴含的客观规律构造一种客观模型,这种模型的好处是无须训练,可以直接应用入侵检测之中。这种客观模型的基础是网络流量自相似理论,该理论认为:局域网和因特网流量呈现分形特点,又称自相似性[1-4],其自相似性特征是一个客观模型,利用自相似特征的变化就可以判定异常发生与否。自相似特征有一些指数可以描述,如 Hausdorff 维数、Lyapunov 指数、Hurst 指数等。其中,Hurst 指数最为常用。本文的目的在于实现一种基于 Hurst 指数进行异常检测的软件。在第 2 节介绍异常检测原理,第 3 节对软件进行设计,第 4 节是软件实现。

2 基于 Hurst 指数检测异常的原理

1994－1995 年,Leland, Beran, A. Veres 等人对 Bellcore 的局域网、视频流数据和 WAN,FASTPAC 等网络的测量结果显示,实际网络流量模型具有统计自相似性。网络流量的这种特性可以用 Hurst 指数来衡量,Hurst 值的范围在[0.5,1],Hurst 值越大,自相似程度越高,典型自相似性 Hurst 值为 0.72。

在网络终端用户及多种业务的共同作用下,网络业务流量表现出的突发性是造成自相似性的主要原因,例如,网络用户的个体行为的突发性与随意性,文件的重尾分布等等原因。而 A. Veres 等人通过模拟产生单个 TCP 对话流量的自相似现象,得出 TCP 拥塞控制机制也是一个能产生自相似现象的确定性因素。正常网络流量具有自相似性,而当发生网络攻击时,攻击数据包将阻塞网络中正常的 TCP 数据包的传输,自相似性会降低,当攻击使得网络几乎完全阻塞时,网络流量将趋向于泊松分布过程,即 Hurst 值趋向于 0.5。可见,当发生网络攻击时,Hurst 值将有较为明显的变化。从 Hurst 值的变化,就可以检测到是否发生了网络攻击,这就是基于 Hurst 指数检测异常的原理。

目前,Hurst 求解有多种方法[5],如聚集方差法、周期图法、R/S 法、Whittle 法、小波方差法[6-8]等。本文所设计的软件实现两种方法:一种方法是 R/S 法,另一个是小波方差法。

3 软件体系结构

软件设计了几个功能模块,包括网络数据捕获器、特征提取器、时间序列划分、R/S 求解 Hurst 值。小波求解 Hurst 值、异常判断、图形用户接口,如图 1 所示。

该体系结构的使用过程如下:

Step1 通过网络数据包接收器从网络层接收网络数据

*)本课题得到国家自然科学基金(60573141 和 60773041),江苏省高技术研究计划(BG2006001),国家高科技 863 项目(2006AA01Z201、2006AA01Z219、2006AA01Z439、2007AA01Z404、2007AA01Z478),2006 江苏省软件专项,现代通信国家重点实验室基金(9140C1105040805)和江苏省高校自然科学研究计划(07KJB520083)资助。**任勋益** 讲师,主要研究方向为计算机网络和网格计算、移动代理和信息安全技术;**王汝传** 教授,博士生导师,主要研究方向是计算机软件、计算机网络和网格、信息安全、移动代理和虚拟现实技术等;**蔡小华** 硕士研究生,主要研究方向为计算机软件技术、信息安全;**傅雷杨** 硕士研究生,主要研究方向是计算机网络、信息安全和计算机在通信中的应用等。

包。为了获得局域网内的数据包，必须将网卡设置为混杂模型；

Step2　从获得的数据包中，提取求解 Hurst 值所需要的特征量。在此需要两个特征量，一个是时间，一个是数据包大小，将这些特征存储于特征库；

Step3　得到这两个特征后，对数据包进行时间序列划分；

Step4　采用 R/S[1,8,9] 或者小波分析法求解 Hurst 值[10]；

Step5　根据经验阀值判断当前的 Hurst 值是否正常。

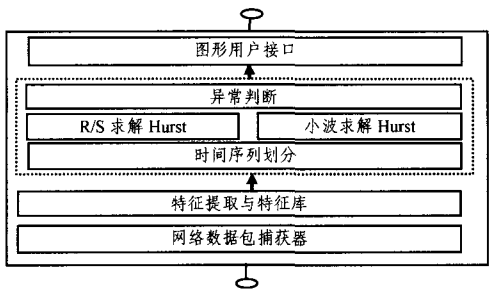

图 1　软件体系结构

4　软件设计与实现

4.1　网络数据包捕获器

数据包捕获可以采用 3 种方式：第 1 种采用 Netfilter 框架；第 2 种采用现成的 Libpcap 库为基础进行编程；第 3 种为自己开发 Socket 方式的数据包捕获。本软件基于 Socket 编程数据包捕获器线程函数，主要处理流程如下：

```
while(true)
{
  // 调用 pDS 数据源的 GetPacket()函数获取数据包
  int nResult = pDS->GetPacket(pktHdr, packetContent, BUFF_SIZE);
  if(nResult == -1)
    break; // 文件读取结束
  if(nResult == 0)
    continue; // 其它错误
  // 遍历 list 链，发送数据包到达消息
  std::set(CDevice *)::iterator iter;
  for(iter= deviceList. begin(); iter! = deviceList. end(); ++iter)
    (*iter)->OnPacketArrival(pktHdr, packetContent);
}
```

捕获器在进行数据包捕获时，设置好数据源格式，初始化后打开，调用 GetPacket()循环捕获数据包。

4.2　特征提取器与特征库

数据包特征提取器负责将数据包中的流量特征提取出来，并存储到特征库中，如图 2 所示。

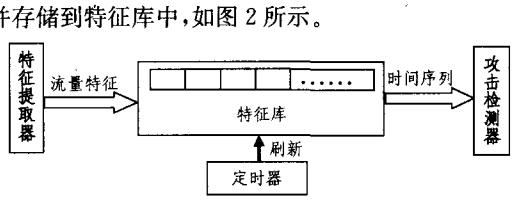

图 2　特征处理总体框架图

左边特征提取器提取出的流量特征传递到特征库中保存。当 OnPacketArrival()消息到达时，对协议进行简单分析，分析结果为以下格式：

数据包到达本机的时间	数据包长度	数据包协议类型	源 IP 地址	目的 IP 地址

因此，数据包到达的时间及数据包长度可以从数据包的伪包头中获取。特征库内部采用链表形式。右边攻击检测器从特征库中提取一定长度的特征值序列，经过预处理后送入攻击检测器进行攻击检测分析计算，由于特征库中只保存一定长度的数据信息，因此需要定时刷新库中的数据，去除超时的数据。在此，定时器的设计采用 setitimer 结合 SIGNAL 的方式。

int setitimer(int which, const struct itimerval * value, struct itimerval * ovalue);

由 setitimer 设置系统定时，并将系统 SIGALRM 信号与相关处理函数绑定，当 setitimer 函数超时发出 SIGALRM 信号时，调用绑定的函数。这样就实现了定时刷新特征库中的数据。

4.3　时间序列划分

使用 Hurst 指数检测异常对数据的计算有一定的格式要求，因此必须将原始数据经过处理才能使用。输入数据要求主要就是将原始数据按照时间等间隔划分，每一时间间隔作为一个时隙，将该时隙内所接收到数据包的长度总和作为该时隙的抽样值。这里我们设计一个函数 GetNumericSeq，算法如下：

```
bool CPktChrtLib::GetNumericSeq(double array[], int arraylen, int interval)
{
  if(共享队列有数据)
    pthread_mutex_lock(&mutexSignal); // 对共享数据队列加锁;
    指针 p 指向队列第一个结点，部分和 subsum = 0;
  while(没有到达队列尾)
  {
    获取 p 结点时间信息
    if(p 结点时间小于当前时隙的右限)
      subsum += p->数据包长度;
    else
    { array[arrayIndex++] = subsum; // 部分和存入数组中，数组指示器下移一位
      if(arrayIndex >= arraylen) // 数据满了，跳出循环
      break;
      subsum = 0;
        更新时间下限，准备计算下一时隙
      }
      p 指向队列下一元素;
  }// 循环结束
  pthread_mutex_unlock(&mutexSignal); // 对共享数据队列解锁;
  if (读取数据长度小于所要求的数据序列长度)
    进行数据填充
}// 算法结束
```

在该算法中，我们使用了数据填充技术，攻击检测器从特征库中读取数据时，如果特征库中存储的数据不够多，不足以填满整个给定的 array 数组，这时候就需要进行数据填充。数据填充的目的在于使攻击检测器不至于因为数据量不足而无法进行计算。数据填充也需要有一个前提，即填充的数据

不能破坏或尽可能少影响原数据的相关性。在进行数据填充时，一般采用方法零填充、周期延拓法填充、对称延拓法填充、光滑常数延拓、平滑延拓。

4.4 R/S与小波分析求解 Hurst

R/S求解步骤主要是重新划分时间序列，计算标准差和极差，对数坐标拟合[8,9]。R/S求解 Hurst 值的实现流程如图3所示。

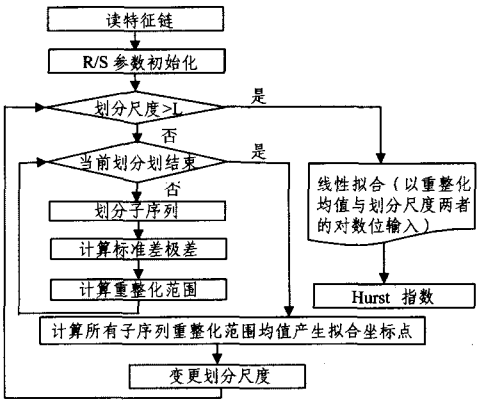

图3 R/S求解 Hurst 值流程

具体的实现，可以定义一个结构体类型，保存分析序列的每一个特征，并将特征加入一循环链表。在包特征提取模块中，定义这样一个结构：

```
struct packetcharacter{
        int time;
        int size;
}
```

在这个结构体中，time 表示数据包到达的时间，size 表示到达的数据包的长度。在这里采用一个链表元素的定义为：

```
struct list{
        struct packetcharacter p;
        struct list next;
}
```

改变时间尺度，对链表进行遍历，循环得到子序列；对所有子序列计算结果进行拟合，近似得到一条直线；通过直线的斜率得到 Hurst 值。小波分析求解 Hurst 值是对时间序列进行小波分解，最后再进行线性拟合求解 Hurst 值，小波求解 Hurst 参数详细过程见文献[10]。

5 软件性能

本系统界面开发使用 Gtk+图形编程、Glade 界面辅助设计、多线程技术等，整个系统是在 Anjuta 集成开发环境下完成的。

使用R/S检测器检测攻击情况如图4所示。“流量-时间”曲线记录了每一秒中到达的数据包个数，从该曲线可以直观地看出网络流量发生了剧变，从 100packet/s 变化到 12800packet/s，此时，“Hurst-时间”曲线则记录了每一秒检测器计算出的 Hurst 值，可以看出 Hurst 值立即下降。大约在攻击持续一半的时候，Hurst 值下降到最低点，约 0.28。攻击结束时，Hurst 值很快回升到 0.7 以上。

同一时刻，软件只有一种检测器在运行，通过参数设置来切换检测器。切换到小波检测器，一次攻击检测结果如图5，其检测效果也非常好。本软件经过反复实验，对典型的 DDoS 攻击具有良好的检测性能[11]，检测到攻击的时间不超过 5s，准确率在 95%以上。小波分析检测的效果较好于 R/S 方法。

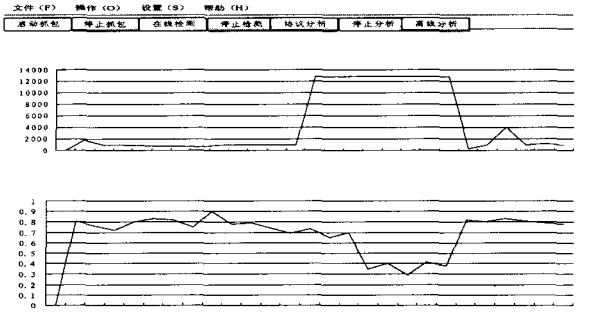

图4 基于R/S检测异常

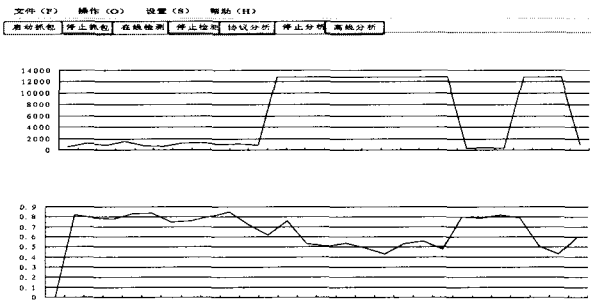

图5 基于小波检测异常

结束语 本文设计和实现的软件表明基于 Hurst 指数的异常检测具有很好的效果，相比神经网络和支持向量机为代表的智能学习方法，它不需要主观学习，只需要能较快地求解出 Hurst 值，根据 Hurst 值的变化就可以发现异常。下一步工作是将该软件扩展到路由器上，研究在核心网上该软件的改进及优化。

参 考 文 献

[1] Leland W, Taqqu M, Willinger W. On the self-similar nature of Ethernet traffic (Extended Version)[J]. IEEE/ACM Trans on Networking, 1994, 2(1): 1-15

[2] Paxson V, Flord S. Wide area traffic: the failure of poisson modeling[J]. IEEE/ACM Trans on Networking, 1995, 3(3): 226-244

[3] Dang T D, Molnar S. On the Effects of Non-Staionarity in Long Range Dependent Tests[R]. Budapest Univ Technology and Economics Tech. Rep. Budapest, Hungary, 1999

[4] Abry P , Veitch D . Wavelet analysis of long range dependent traffic[J]. IEEE Trans on Infor Theory, 1998, 44(1): 2-15

[5] 第文军，薛丽军，蒋士奇. 运用网络流量自相似分析的网络流量异常检测[J]. 兵工自动化，2003，22(6)：28-31

[6] 李永利，刘贵忠，王海军. 自相似数据流的 Hurst 参数小波求解法分析[J]. 电子与信息学报，2003，25(1)：100-105

[7] 李弼程，罗建书. 小波分析及其应用[M]. 北京：电子工业出版社，2003

[8] 任勋益，王汝传，张登银. R/S 和小波分析法检测 DDoS 攻击的研究与比较[J]. 南京邮电大学学报，2006，26(6)：48-51

[9] 傅雷扬，王汝传，任勋益. R/S 方法求解网络流量自相似参数的实现与应用[J]. 南京航空航天大学学报，2007，39(3)：358-362

[10] 任勋益，王汝传，王海艳. 基于自相似检测 DDoS 攻击的小波分析方法[J]. 通信学报，2006，27(5)：6-11

[11] 蒋凌云，王汝传. 基于流量自相似模型的 SYN2Flood DDoS 攻击防范[J]. 南京邮电大学学报，2007，27(2)：91-94