基于时态知识模型的网络入侵检测方法研究

凌 军[1] 曹 阳[1),2)] 尹建华[1] 黄天锡[1]

[1](武汉大学电子信息学学院 武汉 430072)

[2](武汉大学软件工程国家重点实验室 武汉 430072)

摘 要 在分析国内外现有入侵检测技术和系统的基础上,提出了一种基于时态知识模型和可变滑动窗口的实时模式提取算法,并在此基础上,实现了基于规则的、层次化的智能入侵检测原型系统(RIDES).实验结果表明:该系统不仅能快速检测网络入侵,而且具有一定的学习能力,能够适应不同的网络应用环境.

关键词 入侵检测;时态知识模型;可变滑动窗口;模式提取算法

中图法分类号 TP393

Study on Method of Network Intrusion Detection Based on Temporal Knowledge Model

LING Jun[1] CAO Yang[1),2)] YIN Jian-Hua[1] HUANG Tian-Xi[1]

[1] (*College of Electronic and Information, Wuhan University, Wuhan* 430072)

[2] (*State Key Laboratory of Software Engineering, Wuhan University, Wuhan* 430072)

Abstract On the basis of analyzing of existing method and system, this paper presents a real time intrusion detection expert system (RIDES). A new layered structure which can easily deal with huge amount of data is applied to RIDES. In order to effectively detect intrusion in real time, a pattern extraction algorithm based on temporal knowledge model and varying glide time window is provided. The temporal knowledge model we suggested takes account into time factor, which is helpful for extracting temporal knowledge. To demonstrate the validity of the algorithm, RIDES is developed on Linux operating system, and tested in real network environment. The experimental results reveal that the system can detect and report variety of intrusions and the algorithm is viable. The system can be easily integrated into network security products.

Keywords intrusion detection; temporal knowledge model; varying glide window; algorithm of pattern extraction

1 引 言

近年来,网络已应用到社会经济、人民生活的各个领域,人们在得益于信息革命所带来的巨大机遇的同时,不得不面对信息安全问题的严峻考验:近年来网上攻击事件层出不穷,给计算机网络安全造成极大威胁.因此,确保网络和计算机系统免受攻击、

收稿日期:2002-04-12;修改稿收到日期:2003-03-26.本课题得到国家自然科学基金(60132030,69983005)和国家教育部博士点基金(RFDP1999048602)资助.**凌 军**,男,1976 年生,博士研究生,主要研究方向为智能网络管理、网络安全.E-mail: lingjun1976@163.com. **曹 阳**,男,1943 年生,教授,博士生导师,主要研究领域为智能网络管理、网络安全、网络性能评价. **尹建华**,男,1976 年生,博士研究生,主要研究方向为智能网络管理、网络仿真. **黄天锡**,男,1935 年生,教授,博士生导师,主要研究方向为无线电物理、网络管理.

安全而稳定地运行已成为一个重要的研究课题.

一种通过监测系统和网络内部数据活动,发现入侵并能及时做出反应的主动防御的入侵检测技术[1]越来越受到重视,而智能化是入侵检测未来重要的发展方向[2]. AI(人工智能)技术(特别是基于规则的专家系统)对入侵检测极有帮助,例如 SRI International 公司研究开发的 IDES/NIDES 及其后继项目 MERALD[2,3] 以及美国国家安全中心(NCSC)开发的 Multics 入侵检测和报警系统(MIDAS)[4],均将基于统计的异常检测和基于规则的专家系统方法结合起来,规则部件采用 P-BEST 专家系统语言[5]设计,其关键技术是将入侵知识编码,通过分析引擎确定发生的网络或系统事件是否为入侵.

这些系统取得了较大的成功,但随着计算机系统和网络的发展,也面临着一些实际问题:(1)对大批量数据进行连续不断的实时监控,容易丢包,造成漏警;(2)自适应程度不高的,不能随着环境的改变自动改变规则;(3)智能化程度不高,没有考虑数据间的时间关系.

因此,针对以上问题,本文提出了一种新的时态知识模型和基于该模型的滑动窗口模式提取算法,并在此基础上实现了一个实时环境下基于知识的智能入侵检测系统(RIDES). 本文主要内容包括系统结构模型、实时数据和时态知识的表示、实时模式提取算法以及系统评价和实验结果分析.

2 系统结构

在 RIDES 中,由于实时性的要求,系统的实现目标是尽可能快地捕获数据包进行分析,同时还要减轻网络的负担,以提高检测效率. 根据上述要求,可以将 IDS 设计为分层体系结构(图 1),主要由四大模块组成:数据采集、事件引擎、专家系统和人机界面模块. 其基本设计思想是尽快地处理数据,以达到监控高速、大流量网络而不丢包的要求,而分层体系结构为实现上述目标提供可能. 在分层体系结构中,最下层数据处理模块处理大量的数据,并根据安全策略进行过滤,减轻上层的处理压力. 各相邻模块间将互不影响,使专家系统模块将主要精力用于入侵推理工作,从而使检测效率大大提高.

如图 1 所示,系统的基本组成为:

(1) 数据采集. 系统的数据源为网络,利用 libpcap 进行数据采集. libpcap 通过将网卡设置为混杂模式,捕获网络上每一数据包做进一步的处理. Libpcap 具有解码功能,能读出数据包的内容,并根据规则对数据包进行过滤,确定哪些数据包保留,哪些数据包丢弃. 从而减轻上层的处理压力,达到处理速度高、包丢失率低的效果.

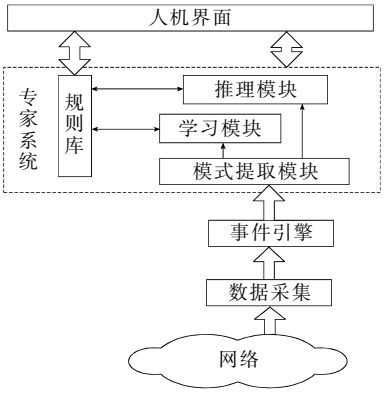

图 1 系统结构

(2) 事件引擎. 经 libpcap 过滤的包传递给事件引擎. 事件引擎通过检查包标志、源地址/目的地址以及源端口/目的端口等信息产生事件,传递给上层专家系统模块进行处理.

(3) 专家系统. 主要由模式提取模块,规则库、推理模块、学习模块组成. 在系统运行过程中,模式提取模块根据事件引擎生成的事件,采用可变时间窗口算法(第 4 节具体介绍)挖掘事件之间的内在联系. 规则库,用于存储关于入侵特征的知识,知识模型在第 3 节讨论. 推理模块即专家系统的逻辑推理部分,是专家系统的核心. 它根据获取的模式,与规则库中的规则,按照一定的策略进行推理,以确定是否有入侵发生. 学习模块根据当前状况不断修正规则库中的规则.

(4) 人机界面. 为安全管理人员提供生动形象的界面,使管理人员与系统的交互更加方便简洁.

3 知识模型

知识模型定义专家系统所使用的知识. 在 RIDES 中,监控对象的数据经常随时间变化. 为了准确地描述网络的当前状态和分析入侵,必须记录过去的信息. 因此和离线分析系统相比,需要引入时间概念来处理涉及时间的信息,这也引发了新的问题:实时数据与时态知识的表示.

定义 1. 可测度集. 设 F 为一组属性值集合,包含所有审计数据或连接记录中可能出现的属性值,例如源/目的地址、连接标识等,称之为可测度

集.设 $V=\{filed_1, field_2, \cdots, filed_n\}$ 为 F 的子集,称之为记录属性.

可测度集表明一个数据包或一条连接的属性集合,例如网络连接的可测度集为($timestamp$, $service$, src_host, dst_host, src_port, dst_port, $flag$),分别表示连接的时间戳、服务类型、源主机、目的主机、源端口、目的端口,标志等属性.其中有一些属性对于准确描述连接的性质很重要,例如源主机,还有一些属性只是辅助信息.网络连接可以由一个五元组组成:〈$timestamp$, src_host, src_port, dst_host, dst_port〉,我们将重要的信息作为记录属性.

定义 2.　事件 $E.V$ 的实例化.形式为 $E=\{(field_1, value_1), (field_2, value_2), \cdots, (field_n, value_n)\}$.其中 $value_i(1\leq i\leq n)$ 为 $field_i$ 类型的值.记录属性的实例化称之为事件,表 1 给出一些事件的实例.

表 1　网络事件

Timestamp	*Src_host*	*Src_port*	*Dst_host*	*Dst_port*
11:41:40	192.168.0.1	4561	192.168.0.10	80
11:41:45	192.168.0.7	4442	192.168.0.1	23
11:41:46	192.168.0.1	3212	192.168.0.80	21
11:41:56	192.168.0.4	5634	192.168.0.10	80

定义 3.　时间点描述事件开始和结束的时刻,分别表示为 $Begin(E)$和 $End(E)$.时区表示事件持续的时间区间,表示为 $Inteval(E)$.时距表示事件持续的时间长度,表示为 $Duration(E)$.

定义 4.　时间约束 TC 描述时间点、时区、时距之间的相互关系.主要有时间点约束 $PC(T_i, T_j)$,基本的关系组合是{>,=,<},分别表示时间点 T_i 早于、等于、迟于时间点 T_j;时间点与时区的约束 $PIC(T, I)$,基本的关系组合是{>,=,<},分别表示时刻 T 处于时区之前、之间、之后.时间点与时距的约束 $PDC(T_i, T_j, D)$,基本关系组合是{>,=,<},分别表示 T_i 和 T_j 之间的时间差值大于、等于、小于时距.其它一些约束以及时间约束之间的推导参见文献[6].

定义 5.　关系约束表示为 $RC(E_i, E_j)=E_i \oplus E_j$,表明 E_i 和 E_j 属性值的定量关系,⊕为约束算子.

根据上述定义,可将含有时间信息的对象形式化表示为

对象::=〈对象名,状态特征值,时间约束集〉.

由此可以表示实时系统中的时态知识,规则可简单地形式化为

〈规则〉::=(〈模式〉〈后件〉),

〈模式〉::=(〈对象表〉〈前件〉),

〈对象表〉::=(〈对象〉|〈对象表〉),

〈前件〉::=(*And* 〈时间约束〉〈关系约束〉),

〈时间约束〉::=(*PC PIC PDC*),

〈关系约束〉::=(*RC*),

〈后件〉::=(结论).

4　基于可变滑动窗口的实时模式提取算法

实时入侵检测系统需要不断地采集网络数据,这些数据往往表现出以下特性:(1)随机性.由于网络业务量的复杂多变以及用户访问网络的随意性,使得这些数据呈现很强的随机性;(2)相关性.单纯的网络事件(例如 telnet,ftp,www 访问等)往往不能完整地反映网络状况,但在较长的时间范围内表现出较强的相关性.

因此,对网络数据的处理,不能仅仅孤立地对网络事件进行处理,而必须综合考虑一个时间段范围内的特征,才能真正反映它们的具体属性.本节提出一种基于可变时间窗口的模式提取算法.

4.1　基本原理

在数据采集过程中,设第 i 个时间窗口的输入记录序列为

$$S_i = \{F_{ij} \mid j = 1, 2, \cdots, n\},$$

其中 F_{ij} 为可测度集,含有源/目的地址、端口等参量.n 为在第 i 时间窗口内采集的记录数,它在 0 以及某一上限之间随机变化,故而 S_i 为一随机过程.

网络数据包形成连接记录后,连接信息主要由时戳、源/目的地址(主机)、源/目的端口(服务类型)组成,这些是网络数据的必要属性.为了挖掘属性之间的关联规则,并尽可能地减少规则冗余,引入参考量集的概念,用以表示网络数据中重要的属性.参考量集一般由源/目的地址(主机)、源/目的端口(服务类型)组成.例如对于一种拒绝攻击,可以设定端口为参考量,然后计算其它属性(例如源地址)与参考量的关系,如果这种关系满足一定的规则(例如源地址数多于最大值),则可认为是发生拒绝攻击.由此可见,参考量集的选择十分重要.

定义 6.　参考量集,设 R 为可测度集 F 的子集,其中包含的元素为可测度集的重要特征,称之为参考量集.

定义 7.　时间窗口上下文 TWC,结构定义为 $(R, F, Rel(R, F), Length)$,其中 R 为参考量集,$R=$

$\{r_1, r_2, \cdots, r_n\}$，$F$ 为可测度集，$F=\{f_1, f_2, \cdots, f_m\}$. $Rel(R,F)=\{R(r_i, f_j) \mid 1\leq i\leq n, r_i\in R, 1\leq j\leq n, f_j\in F\}$描述 r_i和 f_j之间的定量关系.

在实际的处理过程中，我们视前 k 个时间窗作为一个完整的相关窗，计算当前窗口上下文 TWC_i 中的每一个 Rel 项与前 k 个窗口上下文的相应 Rel 项的关系约束. 由于相关窗总是随着新事件的产生不断后移，这样就能实现当前窗口上下文 TWC_i 与前 k 个窗口上下文的模式提取，这个过程称之为可变滑动窗口模式提取. 这种技术的优点如下：

(1)把当前窗口上下文 TWC_i 与前 k 个窗口上下文的关系约束计算视为一个完整的模式提取过程，避免了漏掉相关信息；(2)每一窗口特征为可测度集与参考量集的关系，将该特征提交推理模块和学习模块，可以减轻推理模块和学习模块的压力；(3)任意时间窗口内事件数随网络环境的变化而变化，是一个随机数，因此，能够处理剧烈变化的网络，而不显著增加推理模块的压力；(4)网络事件具有较强相关性，因此相邻时间窗具有相似的模式特征.

4.2 算法描述

算法描述如下：

```
输入：参考量集 R，时间窗大小 t，时间窗口数 k
输出：模式集 P
Begin
  While(TRUE) do begin
  等待当前事件 E_i，
  if BEGIN(E_i)在当前窗口中
  then
    将 E_i存入数组 A
  else
    forall A do begin
      更新当前时间窗上下文
      end for
      保存当前时间窗上下文至数组 B
      if B 元素个数等于 k
      then
        forall B do begin
      计算时间窗口之间的关系约束 RC
      计算时间窗口之间的时间约束 TC
          形成模式，存入模式集 P
        end for
      else
      重构新的时间窗口上下文
      end if
    end if
  end while
end begin
```

算法主要分两步：

(1) 时间窗口上下文信息的获取，在时间窗上下文中参考量集 R 由函数输入，根据实际需要选取，可测度集 F 由获取的事件属性组成，$Rel(R,F)$ 确定两者之间的关系，计算方法为累计求和. 假设参考量集 R 为$\langle dst_host\rangle$，F 为$\langle type, src_host, src_port, dst_port\rangle$，$Rel(R,F)$形式如表 2 所示.

表 2 $Rel(R,F)$ 形式

可测度集 F				参考量集
$Type$	src_host	src_port	dst_port	dst_host
123	23	45	11	主机 1
23	32	46	67	主机 2
…	…	…	…	…
23	33	3	3	主机 n

表中每一项表示：在该事件窗口内的所有事件，针对 dst_host 中 $type$，src_host，src_port，dst_port 的统计数量，用以表示每一时间窗的特征.

(2) 各时间窗口关系的确定，在取得相邻 k 个时间窗上下文后，计算关系约束和事件约束，由定义 5，关系约束表示各时间窗口上下文中 $Rel(R,F)$ 之间的定量联系，采用最大差值方法计算，即取相邻时间窗口上下文中变化量最大的值作为关系约束.

4.3 模式应用

在 RIDES 系统中，模式提取算法所获取的模式主要应用于两个模块：

(1) 学习模块

在学习模块中，根据上述算法获取行为模式后，将形成的新规则存入规则库，形成规则的基本原理是计算时间窗口关系约束的取值范围，例如，在正常行为下：$\langle RC_1, RC_2, \cdots, RCn, min\leq RC_i\leq max\rangle$表示正常情况下 n 个时间窗口的关系约束取值范围为 (min, max). 限于篇幅，学习算法另外撰文详加阐述，本文只介绍基本步骤：

1. 遍历模式集 P；

2. 计算模式集中关系约束各项的取值范围和时间约束，作为形成规则的关系约束和时间约束；

3. 依据规则的形式化描述，形成规则，并插入到规则库中，或对规则库中已有的规则进行修改.

在正常情况下运行学习模块获得正常行为模式，形成规则存入规则库，这样可以根据实时获取的模式是否偏离正常模式来判断是否有入侵发生，从而可以检测到一些未知攻击. 同时，也可以在各种攻击情况下运行学习模块获取各种攻击模式，形成规则存入规则库，可以根据实时获取的模式是否和这

些异常模式相同,判断是否有入侵发生,从而可以高效检测到各种已知攻击类型.因此,RIDES 系统同时可以进行特征检测(signature-based detection)和异常检测(anomaly detection).

(2) 推理模块

在推理模块中,实时模式提取算法获取网络的行为模式后,该模式与规则库中的规则匹配,以确定是否为入侵.例如,假设行为模式为$\langle RC_1, RC_2, \cdots, RC_n \rangle$,如果 $RCi > max$,则可认为发生入侵.基本步骤为

1. 遍历规则库;

2. 当前模式关系约束是否和规则库中的攻击行为关系约束相符,如是则报警;如否,继续第 3 步;

3. 当前模式关系约束是否不在正常行为规则关系约束范围内,如是则报警;如否,退出推理模块.

5　实验结果分析

RIDES 系统在 RedHat 7.2 上用 perl 语言实现.为了对 RIDES 系统以及上述算法做进一步的分析,我们在实验室网络环境下,运行 RIDES 进行各项测试.下面以端口扫描和 ping to death 攻击为例,介绍系统的运行结果.同时给出正常情况下的实验结果.

(1) 端口扫描

扫描的基本原理是通过连接远程 TCP/IP 不同的端口服务,并记录目标主机给予的回答,通过这种方法,可以搜集到很多关于目标主机的各种有用的信息(比如:是否能用匿名登录,是否有可写的 FTP 目录,是否能用 TELNET 等等).

以 dst_host 作为参考变量,时间窗口分别为 1 和 4,在 TCP 端口扫描时,时间窗口上下文中的 $Rel(R,F)$ 结果分别如图 2 和图 3 所示.其中 $Rel(R,F) = \{tcpNum, udpNum, icmpNum, dstportNum\}$,分别表示在时间窗口内 TCP 连接、UDP 包、ICMP 包以及目的端口的累计和.由图 2 可知,在一些连续的时间窗口内,$tcpNum$ 出现较大的增长,说明在这段时间里 TCP 连接的数量急剧增加.同时,$dstportNum$ 与 $tcpNum$ 具有相同的增加趋势,表明有大量针对不同端口的 TCP 连接,与 TCP 端口扫描的特征吻合,说明在这段时间内攻击者利用端口扫描获知目标系统的端口信息.图 3 是将窗口大小调为 4s 时的实验结果,与图 2 的变化趋势基本一致,表明窗口的大小不会影响模式提取结果.

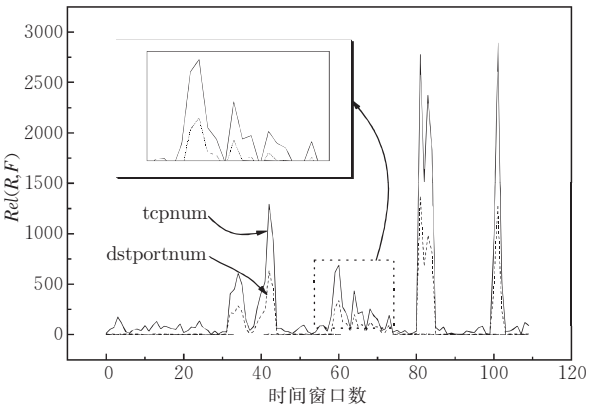

图 2　TCP 端口扫描时间窗口 $Rel(windowsize=1s)$

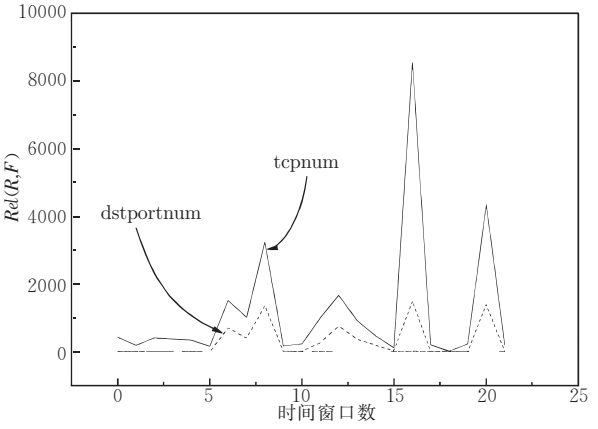

图 3　TCP 端口扫描时间窗口 $Rel(windowsize=4s)$

(2) ping to death

ping to death(致命 ping 攻击)是一种拒绝攻击服务,如果系统很脆弱,那它就会崩溃,即便不崩溃,也会浪费大量网络资源.

我们也以 dst_host 作为参考变量,时间窗口分别为 1 和 4,在发生 ping to death 攻击时,时间窗口上下文中的 $Rel(R,F)$ 结果分别如图 4 和图 5 所示.在图 4 和图 5 中,$icmpNum$ 在一段时间窗口内达到峰值并持续一段时间,说明在这段时间里有大量

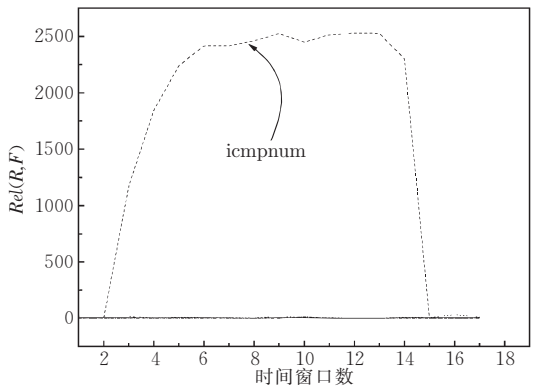

图 4　ping to death 时间窗口 $Rel\ (windowsize=1s)$

ICMP 包,与 ping to death 攻击特征吻合. 同时,窗口大小没有对模式提取结果产生影响.

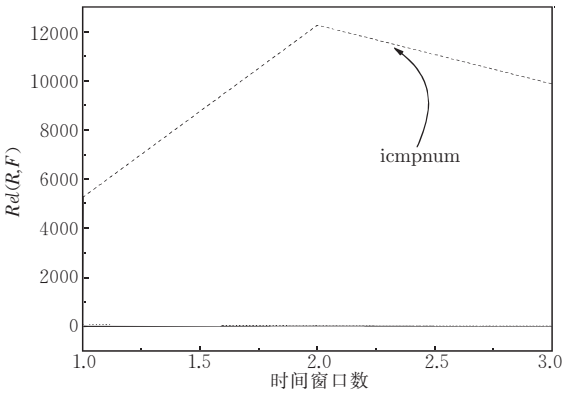

图 5 ping to death 时间窗口 *Rel* (*windowsize*=4s)

(3) 正常模式

作为对比,我们也提供正常情况下的实验数据,时间窗口大小分别为 1 和 4,时间窗口上下文中的 *Rel*(*R*,*F*)结果分别如图 6 和图 7 所示. 由图可知,*tcpNum*,*udpNum*,*icmpNum* 变化趋势基本平稳,同时没有其它关联特征(例如 *dstportNum*)出现,与网络正常情况下的行为特征基本吻合,表明网络状况正常. 同时,窗口大小没有对模式提取结果产生影响.

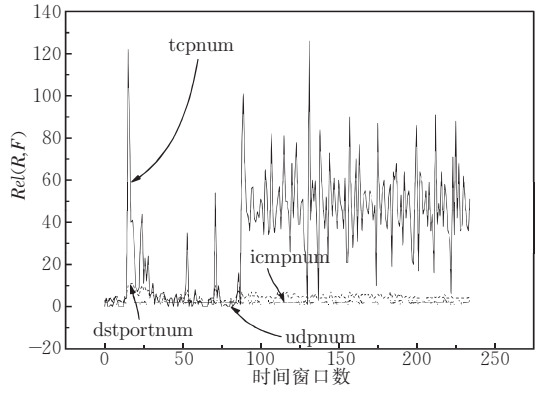

图 6 正常情况下时间窗口 *Rel* (*windowsize*=1s)

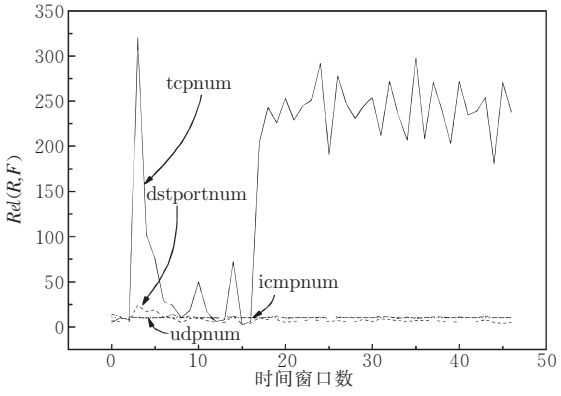

图 7 正常情况下时间窗口 *Rel* (*windowsize*=1s)

由以上各图可以看出,在不同的攻击情况下,各时间窗口表现出不同的特征,因此根据可变滑窗模式提取算法很容易获得各种攻击模式和正常模式.

6 结 论

RIDES 采用分层的体系结构,该结构能够尽快地处理数据,具有监控高速网络而不丢包的能力. 由于使用时态知识模型,可将各种时间概念加入到传统知识中,为实时入侵检测专家系统奠定了基础. 在此基础上提出了一种基于可变时间窗口的模式提取算法,该算法能有效地获取各种攻击模式,以便应用到随后的实时学习和匹配. 实验结果表明:RIDES 不仅能够快速实时处理网络数据包,而且还具有学习能力,能够根据网络环境的不同而改变规则. 该系统能够快速发现各种入侵,为网络管理人员及时发现入侵、维护网络安全提供极大的帮助.

同目前通用的入侵检测系统(例如 snort)相比,RIDES 充分考虑了网络流量的时间特性,采用新的模式提取算法挖掘网络一段时间内的特征,为确定网络是否遭到攻击提供新的思路. 实验结果证明该算法有效可行. 此外,RIDES 还具有实时学习能力,能够适应不同的网络环境.

同时,在系统的扩充和完善方面还有许多工作要做,例如开发更加完善的模式提取算法、学习算法和匹配算法,提高系统的自适应性,将分布式技术应用于 RIDES,增强监控大规模网络的能力等. 这些都是我们今后工作的重点.

参 考 文 献

1 Cannady J, Harrell J. A Comparative analysis of current intrusion detection technologies. In: Proceedings of the 4th Technologies for Information Security Conference, Houston, 1996. 50～57

2 Anderson D, Frivold T, Valdes A. Next-generation intrusion detection expert system (NIDES): A summary. SRI International Technical Report SRI-CSL-95-07, 1995

3 Porras P A, Neumann P G. EMERALD: Event monitoring enabling responses to anomalous live disturbances. In: Proceedings of the 20th National Information Systems Security Conference, Maryland, 1997. 353～365

4 Sebring M, Shellhouse E, Hamma M E. Expert system in intrusion detection: A case study. In: Proceedings of the 11th National Computer Security Conference, Houston, 1988. 74～81

5 Lindqvist U, Porras P A. Detecting computer and network misuse through the production-based expert system toolset (P-BEST). In: Proceedings of the 1999 IEEE Symposium on Security and Privacy, Oakland, 1999. 146～161

6 Guo Hong-Fei, Zhou Jian-Chang. The representation of real-time data and temporal knowleged. Joural of Software, 1997, 8(1):45～50(in Chinese)
(郭宏飞, 周建常. 实时数据及时态知识的表示. 软件学报, 1997, 8(1): 45～50)

LING Jun, born in 1976, Ph. D. candidate. His main research interests include intelligent network management, network security.

CAO Yang, born in 1943, professor and Ph. D. supervisor. His main research interests include intelligent network management, network security and network performance evaluation.

YIN Jian-Hua, born in 1976, Ph. D. candidate. His main research interests include intelligent network management, network simulation.

HUANG Tian-Xi, born in 1935, professor and Ph. D. supervisor. His main research interests include radio physics and network management.