基于智能流量预测的入侵检测方法

于宝华[1],巩林明[1],邓洪涛[1],李 伟[1],张振国[2]

(1. 石河子大学,新疆 石河子 832000;2. 陕西科技大学,陕西 西安 710021)

摘 要:为了提高入侵检测系统检测的实时性,提出了一种基于智能流量预测的入侵检测方法。该方法拟合了智能 Agent 的智能性、自主性和自适应性的优点以及灰色预测对不确定资源的科学预测的优点,用流量预测智能 Agent 预测得到的流量序列来代替未来一段时间段内的实际流量,并把这个预测序列作为检测对象集的一部分。然后用人工方法模拟了流量处理 Agent 与预测智能 Agent 的活动,并通过对实际的采集数据进行仿真,实验证明了预测智能 Agent 的预测活动的科学性。

关键词:灰色预测;小波变换;数据处理智能 Agent;预测智能 Agent

中图分类号:TP301;TP393.01 **文献标识码**:A **文章编号**:1674-7720(2011)01-0050-03

A method for intrusion detection based on intelligent prediction to network traffic

Yu Baohua[1], Gong Linming[1], Deng Hongtao[1], Li Wei[1], Zhang Zhenguo[2]

(1.Shihezi University, Shihezi 832000, China; 2.Shanxi University of Science Technology, Xi'an 710021, China)

Abstract: In order to improve the real-time detection of intrusion detection system, an intrusion detection method based on intelligent prediction to network traffic is proposed. The merits of intelligence, autonomy and adaptability to the intelligent Agent, and the benefits of science on grey prediction to the uncertain resource are fitted in this method. The method uses the prediction sequence of the intelligent Agent for forecasting to substitute the actual network traffic in a future time period, and view this prediction sequence as part of the detection set to the intrusion detection system. Then the activities of intelligent Agents for data processing and forecasting to network traffic are simulated in an artificial way. And the scientific nature of the prediction method employed by intelligent prediction Agent is proved through a simulation experiment to the actual acquisition data.

Key words: grey prediction; wavelet transform; intelligent Agent for data processing; intelligent Agent for prediction

由于分布式入侵检测系统和被监听流量的分布性,将智能 Agent 引入到网络流量预测单元,可以增强系统的智能性、自主性和自适应性[1-2]。且网络流量的预测是入侵检测系统预警机制的基础,当有大流量冲击时,即攻击者向被保护网络发送大量的数据,超过 NIDS 的处理能力,将会发生丢包的情况,从而可能导致入侵行为漏报。为此,在 IDS 中增加智能网流预测处理功能,即将预测处理后所得数据传送到 IDS 的智能决策器,智能决策器根据事先设定的规则来判断是否有入侵即将发生,进而降低入侵行为漏报率。分布式入侵检测系统(DIDS)虽然利用分布计算技术克服了集中处理问题的缺点,但是还不能消除从收集攻击信息到发出响应这一段时延对系统实时性所带来的不良影响。

正是基于以上原因,本文将智能 Agent 和基于灰色小波的网络流量组合预测[3-4]方法引入到分布式入侵检测系统,提出了基于智能流量预测的入侵检测方法。

1 入侵检测中的智能 Agent

由于网络流量时间序列逐渐呈现出非线性和多尺度变换的特性,加之受到多种复杂随机因素的影响,加剧了网络流量行为的复杂多变。如果以非平稳性的流量序列作为初始训练输入可以降低模型的训练精度和延长训练周期,但非平稳的流量序列经小波变换后将被分解成为若干个平稳性较好的分量[5-8],灰色资源预测就是对不确定资源的预测[9-10],其能够从系统的一个或几个离散数列中找出系统的变化关系,建立系统的连续变化关系。所以本设计中数据处理智能 Agent 采用了小波

技术，而预测智能 Agent 采用了灰色预测方法。

1.1 数据处理智能 Agent

数据处理智能 Agent 的工作是预测前对原始流量的处理，以平稳化数据。具体工作如下：

(1)平稳化原始流量序列，即用 Mallat 算法对$\{x(k)\}$进行小波分解，并且对分解过程中各层产生的分量序列分别进行重构，便可得到某层的计算信号量 x：

$$x=\sum_{j}^{J} d_j+a_J \tag{1}$$

式中，$d_j:\{d_{j,1},d_{j,2},\cdots,d_{j,N}\}$为第 j 层分解得到的细节信号，$a_J:\{a_{J,1},a_{J,2},\cdots a_{J,N}\}$为第 J 层分解得到的逼近信号[11-13]。

(2)分别给各层的细节分量(即小波系数 d_j 和逼近信号 a_J)选择恰当的平移值 $Q(Q>0)$进行平移处理，获得某时刻的非负的平滑网络流量序列，完成该时刻下网络流量序列的重构。

1.2 预测智能 Agent

预测智能 Agent 负责的是在不确定的网络环境中对流量资源做出科学合理的估计，其具体工作如下：

(1) 对 $t+1$ 时刻各层小波系数 $d_{1,t+1},d_{2,t+1},\cdots,d_{J,t+1}$ 与尺度系数 $a_{J,t+1}$ 进行预测，其步骤如下：

①对各层小波系数 $d_j(j=1,2,\cdots,J)$和尺度系数 a_J 进行平移处理；

②用经步骤(1)的各层序列对 GM(1,1)模型进行参数估计；

③用 GM(1,1)模型对平移后的 $d_{1,t+1},d_{2,t+1},\cdots,d_{J,t+1},a_{J,t+1}$ 进行预测计算，得到预测值：$\hat{d}_{1,t+1},\hat{d}_{2,t+1},\cdots,\hat{d}_{J,t+1},\hat{a}_{J,t+1}$；

④用步骤③中得到的$\hat{d}_{1,t+1},\hat{d}_{2,t+1},\cdots,\hat{d}_{J,t+1},\hat{a}_{J,t+1}$计算得到原始流量序列 x 的预测值：

$$\hat{x}_{t+1}=\sum_{j=1}^{J}\hat{d}_{j,t+1},\hat{a}_{J,t+1} \tag{2}$$

(2)合成预测信号，其合成方法采用的是小波重构方法，用形式化语言表述为：

$$x_t=\sum_{j}^{J} d_{j,t}+a_{J,t} \tag{3}$$

式中，时刻 $t:\{t|t<N\}$的 x_t 表示 t 时刻具体的网络流量，而 $d_{j,t}$ 与 $a_{J,t}$ 分别是经数据处理智能 Agent 处理后的第 j 层分解得到的细节信号和第 J 层分解得到的逼近信号[14]。所以要预测 x_t 的下一步的计算信号量 x_{t+1}，就可以借鉴此种方法，进而有：

$$x_{t+1}=d_{j,t+1}+a_{j,t+1} \tag{4}$$

1.3 智能流量预测单元涉及的因素

智能流量预测单元各类型的智能 Agent 所涉及的因素如表 1 所示。

2 智能预测预报模块总体构架

在智能预测预报模块中，由信息搜集智能 Agent、数据处理智能 Agent、预测智能 Agent、数据分析智能 Agent、决策智能 Agent 五种类型的智能 Agent 相互协作，共同完成对网络流量的预测。智能预测预报模块的总框架如图 1 所示。

表 1 智能流量预测单元涉及的因素

智能 Agent 类型	感知对象	执行动作	目标	环境
数据处理智能 Agent	原始流量	处理	准确	
数据分析智能 Agent	网络流量预测流量	分析	直接发现异常	网络
决策智能 Agent	其他智能 Agent	决策	迅速决策	病毒
流量预测智能 Agent	流量序列	预测	发现异常报警	

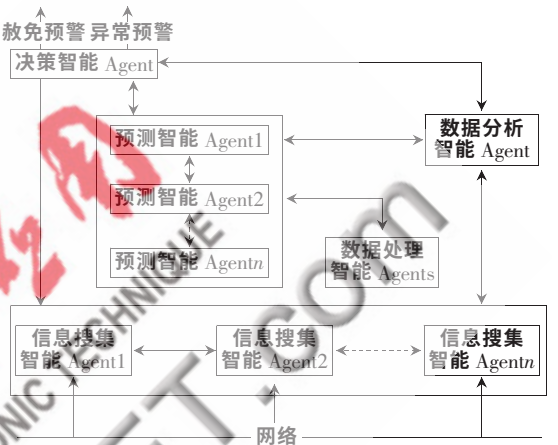

图 1 智能流量预测预报模块

3 基于流量预测预报的入侵检测模型的构建

本模型由流量分析机制、流量预测预报、数据库(database)和决策模块组成，其结构如图 2 所示。

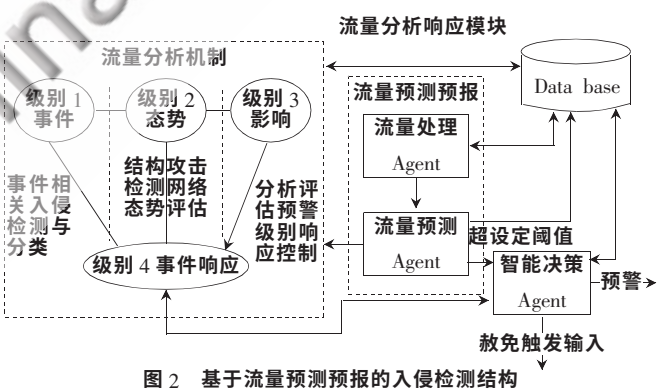

图 2 基于流量预测预报的入侵检测结构

各部分功能说明如下：

(1)流量分析机制：其功能类似于普通的入侵检测系统功能，具有四级响应机制，只是检测对象有所不同，其检测对象不仅包含原始流量序列，还包含基于原始流量预测得到的预测序列。

(2)流量预测预报模块：本模块设计的目的为弥补检测捕捉数据实时性的不足与单点失效 (将信息捕获、预处理和简单的分析判断分散在各个探测节点上进行，极大地提高了入侵检测的效率，但由于所有探测节点需要以中央处理器为核心协同工作，容易出现单点失效的问题)。其主要思想是利用小波技术与数据平移对原始流

量序列进行处理，然后进行较为精确的组合预测，并将此预测序列(代替未来某一段时间内的流量序列)与前面提及的原始流量序列共同作为检测对象。需要强调的是，基于分布入侵检测系统的数据捕获与处理机制——在网络中的相关位置安装探测节点，按照一定的规则捕获原始信息，进行简单预处理和分析判断，再统一提交给中央处理器，由中央处理器进行检测判断。

(3)数据库：由检测对象集(原始流量序列与基于原始流量预测得到的预测序列的合集)、特征集两部分组成。需要强调的是特征集也是可变的，随着对新型入侵的深入学习与识别，其值不断变大。

(4)决策模块：根据流量预测送来的预测值与事先设定的预警触发阈值决定是否进行入侵预警。

4 人工办法对流量预测智能 Agent 采用的技术及方法进行仿真实验

从上面设计的入侵检测系统模型可得出：如果预测流量能够准确地替代原始流量，那么本设计将会改善入侵检测系统的时延性。为此，有必要对流量处理智能 Agent 与流量预测智能 Agent 采用的技术及方法的科学性进行论证。

4.1 数据处理

按照数据处理 Agent 与流量预测智能 Agent 构建的算法，对实际的采集数据进行仿真实验。

其中，作为信号数据 S 的网络流量来源于参考文献[15]，从 2006 年 7 月 1 日到 7 月 11 日，11 天中的网络每小时通过的流量，即 264 个实验数据，把这 264 个实验数据表述成了一个初始网络流量时间序列 $\{s(t), t=1,2,\cdots,264\}$，以前 10 天的数据共 240 个数据建立小波灰色无偏模型的信号输入 S，当进行预测检验时，后 24 个数据与前面的 240 个数据共同作为信号输入 S[16-17]。

由上面的信号分解程序对信号数据 S 进行多尺度分解[18]，可得到流量信号经小波分解系数处理的序列为 $\{d_1(k), d_2(k), d_3(k), a_3(k)\}$，如图 3 所示。

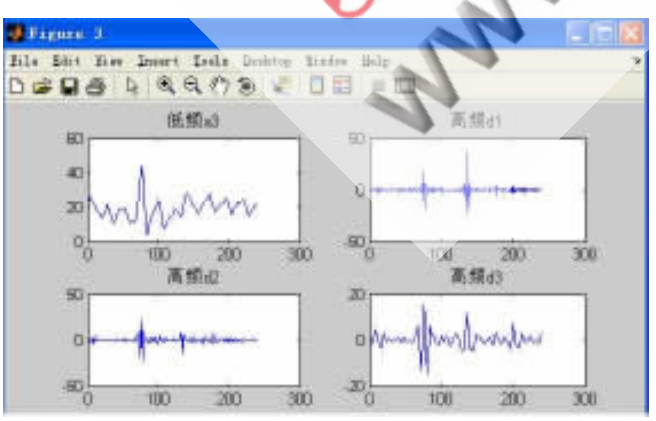

图 3 流量信号的分解系数

需要指出的是：为了防止流量处理智能 Agent 在对原始流量处理的过程中对输入数据的裁剪，在 Matlab 程序中，用 4 个“save'*.txt'*-ascii;”语句将分解的数据直接进行保存。这样保证了后续的平移处理的数据个数与原始网络流量序列的长度是一样的。

4.2 仿真效果分析

针对以上信号数据 S 按照小波预测模型和组合模型对网络流量序列预测分别进行仿真，得到如图 4 和图 5 所示的预测效果。

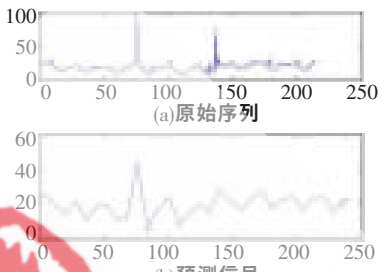

图 4 用小波预测方法预测的流量信号

图 5 用灰色小波组合方法预测的流量信号

对比以上预测结果，充分表明该方法具有较好的预测效果，能够较准确地预测出下一时间段的网络流量。

把预测智能 Agent 送来的预测序列值作为检测对象集的一部分来实现对入侵的提前预警的方法，应用到分布式入侵检测领域，将会极大地改善现有分布式入侵检测系统的时延性。

参考文献

[1] 王帅.基于移动 Agent 的分布式网络入侵免疫系统的设计与实现[D].武汉：武汉大学，2004.

[2] 王汝传，徐小龙，黄海平.智能 Agent 及其在信息网络中的应用[M].北京：北京邮电大学出版社，2006.

[3] 刘渊. 基于递归神经网络的网络流量组合预测模型[J].计算机工程与设计，2008，29(3)：700-703.

[4] 陈华友.组合预测方法有效性理论及其应用[M].北京：科学出版社，2008.

[5] 吕林涛，李军怀.时间序列模式及其预测模型算法应用[J].计算机工程，2004，17(4)：3-4.

[6] 杨晓波，胡黎伟.时间序列理论在电信行业预测决策系统中的应用[J].计算机工程与应用，2004，28(4)：2-3.

[7] RETCARI G, CLINKLER T. Practical OSPF traffic engineering [J]. IEEE Communication Letters, 2004(11):

689-691.

[8] BROCKWELL P J, DAVIS R A.时间序列的理论与方法[M].北京:高等教育出版社,2001.

[9] 邓聚龙.灰色数理资源科学导论[M].武汉:华中科技大学出版社,2007.

[10] 刘思峰.灰色系统理论及应用[M].北京:科学出版社,2008.

[11] 洪飞, 吴志美. 基于小波的多尺度网络流量预测模型[J].计算机学报,2006,29(1):166-170.

[12] 程光,龚俭,丁伟.基于小波的网络流量分解模型[J].小型微型计算机系统,2005,26(3):400-401.

[13] 葛哲学.小波分析理论与MATLAB2007实现[M].北京:电子工业出版社,2007.

[14] 赵洋.一种基于小波分析理论的灰色预测方法[J].西南民族大学学报,2005,31(4):448-501.

[15] 流量文库.http://newsfeed.ntcu.net/~news/2006/,2010-09-01.

[16] 陈淑燕.交通量的灰色神经网络预测方法[J].东南大学学报:自然科学版,2004,34(4):541-544.

[17] 郑成兴.网络流量预测方法和实际预测分析[J].计算机工程与应用,2006(23):129-130.

[18] QIAO Y, SKICEWICZ J, DINDAP. An empirical study of the multiscale predictability of network traffic [C]. IEEE Proceedings of High Performance Distributed Computing, 2003.

(收稿日期:2010-09-011)

作者简介:

于宝华,男,1979年生,硕士,讲题,主要研究方向:网络安全与数据库。

巩林明,男,1979年生,讲师,主要研究方向:网络安全。