大数据动态安全 SAT 双向防御模型的研究*

罗恩韬[1]，胡志刚[2]，杨 杰[1]
(1. 湖南科技学院 计算机与通信工程系，湖南 永州 425199；2. 中南大学 软件学院，长沙 410073)

摘 要：为了保障大数据计算的安全，结合大数据安全的安全研究和可信云的概念，提出一个大数据计算下 SAT 的双向防御系统模型。该模型通过动态安全因子对用户数据证据进行规范；逐步确定各行为证据的安全权限，实现行为的可靠性测试，为用户提供最大限度的安全防御。实验结果表明，该系统模型可以有效地提升大数据的数据安全，降低黑客的恶意攻击行为，从而提高大数据分析效率。

关键词：大数据；SAT 防御系统模型；双向防御
中图分类号：TP391 **文献标志码**：A **文章编号**：1001-3695(2014)05-1470-05
doi:10.3969/j.issn.1001-3695.2014.05.045

Research on bi-directional defense SAT model of big data dynamic safety

LUO En-tao[1], HU Zhi-gang[2], YANG Jie[1]
(1. *Dept. of Computer & Communication Engineering, Hunan University of Science & Engineering, Yongzhou Hunan* 425199, *China*; 2. *School of Software, Central South University, Changsha* 410073, *China*)

Abstract: In order to guarantee the safety of big data calculation, and based on the concepts of security research and trustworthy clouds of big data safety, this paper put forward a bi-directional defense system SAT model under big data calculation. The model regulated users' data evidence through dynamic safety factors, gradually determined the security permission of various behavior evidences, achieved reliable test of behaviors, as well as offered safety defense to users to the maximum limit. Experimental result shows that the system model can effectively improve the data safety of big data and reduce the hostile attack of hackers so as to improve the analysis efficiency of big data.

Key words: big data; SAT model; bi-directional defense

在大数据的迅速发展下，在数据安全方面，单独依靠病毒探测、安全登录认证和入侵检测技术已经不能满足云端面对日益增多的病毒攻击的需要，各种破坏行为的出现，直接促进了大数据安全和可信云[1]理念的诞生。大数据安全的核心思想是通过一系列的安全解决方案来规避风险；而可信云思想是结合可信网络[2]的思想，对来自 Web 终端的信息请求予以验证和判断，从而保证它的合法性和安全性，对于恶意请求予以隔断，从而达到安全防护的目的，进而增加云端的安全性。

关于信任机制的风险判断，很多文献都给出了自己的意见，从而解决以往安全机制所遗漏的可信行为漏洞。文献[3]提出了云计算下基于信任的防御模型，是云计算时代的动态信任模式，它主要从引入行为证据进行判断，这个模式在对病毒的探测和防御方面有一定的作用，但是它也有不足之处，例如在如何获取行为证据的动态安全信任因子时没有很详细地说明，同时在处理数据任务的安全性和考虑信任值的变动情况的论证也不够充分。文献[4]主要引入 QoS 和信任函数的思想，利用引进直接信任函数和间接信任函数来保证大数据安全，该方法在一定程度上能够提高系统的可信性，降低任务失败和被恶意行为欺骗的概率，但是在恶意节点的处理上执行效率一般。文献[5]主要引入信任因子和向量的数学模型来建立安全机制，利用各因子的关系来判断网络中的恶意行为，但是太过复杂，也没有系统的风险规避解决方案。文献[6]提出了引入风险函数和多个信任属性来进行判断，有较强的新颖点，但是量化程度不够。本文依据大数据安全的特点，提出了动态安全信任因子的计算方法，充分考虑了大数据安全的信任区间，提出了大数据基于动态双向 SAT 安全防御模型的安全防御策略，并利用了 SmartBits 网络分析仪的测试，验证了其可行性。

1 大数据动态安全 SAT 双向防御架构

1.1 大数据安全动态 SAT 双向防御架构分层

本文在构建系统模型时，参考了服务网格环境下基于行为的双层信任模型的研究[7]思想和大数据环境下 DPSO 资源负载均衡算法[8]。同时在大数据的安全鉴权上借鉴了粒子群的优化算法思想[9]，利用经过优化的动态因子来规范大数据端和终端的安全权限，从而设计云端网络应用、网络服务、网络资源和云客户端之间的协同依赖，确定它们彼此的安全关系，达到双向的安全防御。双向防御架构可以分为：

a)行为数据拆分中心(S 层)，负责大数据的分布式处理，在数据拆分算法方面，首先对数据集进行预处理，利用 MapReduce 框架把终端提交的任务划分成无逻辑关联的子任务，将

收稿日期：2013-07-03；**修回日期**：2013-09-29 **基金项目**：国家自然科学基金资助项目(61272148)；湖南省科技厅一般项目(2011FJ3069)；湖南省科技学院计算机应用技术重点学科资助项目；湖南科技学院 2013 年科研资助项目

作者简介：罗恩韬(1978-)，男，湖南永州人，副教授，硕士，主要研究方向为大数据、大数据安全、大数据汇聚计算、多维分析(1462914111@qq.com)；胡志刚(1963-)，男，教授，博导，主要研究方向为并行计算、虚拟资源的能效模型计算；杨杰(1976-)，男，副教授，主要研究方向为图形计算、数字模糊处理、人脸识别.

正负类分别汇聚成若干子簇;然后对两两组合的正负子簇用 SMO 算法进行交叉学习,得到多个基本分类器;最后对这些基本分类器进行集成学习[10],从而找出不同的数据分割类,完成数据的拆分。

b)行为数据分析中心(A 层),负责大数据的分析,也就是大数据的并发处理操作,在该层面向大数据深度分析键值对应关系[11]。利用 map 函数产生一组中间键值对,然后传递 reduce 函数,由 reduce 函数将这组值进行合并产生一组规模更小的值进行分析。整个过程中,由底层分布式文件系统(GFS)和本地文件系统的中间数据进行交互,从而完成对文件的分析和对应。在处理过程中,大数据端服务中心涵盖大型的病毒数据中心,储存终端分析报告,对终端进行风险评估。当终端提交非法任务时,该终端负责主动向云端报送分析报告,并且将自身的非法任务进行隔离,由云端的集群服务器进行进一步分析。

c)行为数据包装中心(T 层),完成数据的加工整合、数据转换和封装存储。在执行过程中,充分考虑数据的加密存储,利用 IMA 度量列表结果以及数据签名策略来保证数据的准确包装,规避因为平台复杂性造成的值不确定性和数据精度的问题[12]。同时对终端提交的分析报告进行分析和监测,对可疑文件进行综合处理,最后将安全的数据和合理的处理意见下发给云终端进行使用。

d)云客户端。云终端是整个系统的底层,它不仅要上报自己的分析报告,同时也要负责对云端下发的处理意见进行自身的可选择操作。

1.2　动态 SAT 双向防御架构的相关描述

动态 SAT 双向防御机制是一个有别于其他数据安全防御模型的研究概念,研究重点主要包括两个方面,即大数据端动态安全因子的获取以及大数据端数据和终端数据之间的数据安全性鉴权和合法性鉴权。这主要体现为云端服务商随机计算动态安全因子来提高安全级别,而终端安全主要依靠终端客观实体自身及相互之间的安全性和合法性的信任关系。

1)安全性　衡量大数据中动态 SAT 双向防御机制对客观实体任务的真实性验证程度。本文定义了云端大数据和终端之间动态的信任因子和强弱关系函数模型来验证数据报文交互的安全性。

2)合法性　用户任务的合法性是指大数据的提供方在接收提交的任务时对提交任务的合法性能监测。同时长时间数据端和服务器端信息的交互可能因为数据报被第三方获取从而导致数据失密,造成安全隐患。本文的 SAT 双向安全机制中主要使用保密性能函数及合法级别量化指标来定义用户的合法性。

2　动态 SAT 双向防御机制的行为

2.1　大数据端动态安全信任因子的获取

动态安全信任因子对于定义大数据交互安全权限具有重要的意义,因为终端的安全行为在单位时间内的变化是动态的,所以有必要在大数据端设置动态安全信任权限。在本节中关于动态安全信任因子的取值算法借鉴了传统的线性递减惯性权重粒子群算法,并加以改进,提出了一种动态安全因子的粒子群优化算法。它由大数据种群粒子的进化度和耦合度决定,使数据粒子群在迭代过程中动态因子随粒子进化度和耦合度的变化而改变,解决了传统大数据粒子算法收敛精度不高、易陷入局部最优值、安全性较弱的缺陷。

定义 1　假设动态信任因子的初值是等于单位个体最优值 $T_{\text{old}}(m-1)$,那么在迭代过程中大数据端的动态信任因子应该优于或者等于上一次的个体最优值,如果等于上一次的最优值,则值不进行改变;如果值优于上一次,则值进行交换 $T_{\text{old}}(m-1)=T_{\text{new}}(m)$。在这个过程中信任权限不断进行变化,信任级别也不断进行调整,从而更大程度合理地保证数据交互的安全。

定义 2　为了适应大数据动态数据粒子群的变化情况,并克服在多维数据情况下的寻优能力不足,有必要设置动态因子进化度 θ 来保证大数据服务提供商的数据寻优能力。

$$\theta=\frac{T_{\text{old}}(m-1)}{T_{\text{new}}(m)}\qquad 0<\theta\leqslant 1 \tag{1}$$

该参数考虑到大数据数的据粒子的运行状况,反映了大数据粒子的运动和进化程度。在数据分析前期,θ 值的波动较大,说明大数据的数据变化较大,算法需要分析的数据也相对较多,因此搜寻最优解的时间也较长。此时要适当地减少数据量进行运算以便获得区间内的最优值,而经过算法若干次迭代后,θ 值接近于 1 的水平,则表明动态因子获得了单位时间内最佳寻优能力。

定义 3　T_{size} 表示分析的数据粒子数量,$\sum_{i=1}^{T_{\text{size}}}T_{i(m)}$ 表示取得最优解的大数据粒子群总和,动态因子的耦合程度为

$$\partial=\frac{T_{\text{size}}\times T_{\text{new}}(m)}{\sum_{i=1}^{T_{\text{size}}}T_{i(m)}} \tag{2}$$

当 ∂ 值越大说明数据越分散,而如果 ∂ 趋向 1,说明数据有可能趋于求得局部最优解。因此可以适当增加数据分析范围。

定义 4　根据定义 1 ~ 3,可以得到动态信任因子的函数 $u_{ij}^{(k+1)}$ 和权重因子 q。

$$u_{ij}^{(k+1)}=q\times u_{ij}^{k}+m_1\times r_1\times(T_{ij}^{\text{old}}-T_{ij}^{k})+m_2\times r_2\times(T_j^{\text{new}}-T_{ij}^{k}) \tag{3}$$

$$q=w_{\max}-\text{iterator}\times\frac{w_{\max}-w_{\min}}{\text{iterator}_{\max}} \tag{4}$$

其中:k 代表级数;i、j 分别代表大数据分析粒子和优化参数;m_1 和 m_2 代表大数据的训练因子,通常为一个常数;r_1 和 r_2 为[0,1]闭区间里利用 random 函数产生的随机数;T_{ij}^{old} 和 T_j^{new} 分别表示第 i 个大数据粒子群在第 j 个位置的最优解和第 j 个参数的最优位置。T_{ij}^{k} 和 V_{ij}^{k} 分别表示第 i 个粒子群第 j 个参数经过 k 次迭代后的位置和速度;q 表示动态权重因子,q 取值较大时则有较强的全局搜索能力,而取值较小时则倾向于局部搜索,并且 q 的活动范围为[0,1];$w_{\max}$ 表示最高安全和合法水平。

2.2　动态 SAT 双向防御机制的行为定义

定义 5　大数据端实体之间进行信息数据交互时,如果大数据实体之间的安全水平处于大数据的最大动态信任因子安全权限之内,而对于其他的情况大数据端统一设定它的访问权限为 0,则认为是强安全。

$$\text{info}_{\text{super}}(i,j)=\begin{cases}1 & \text{if}\quad \text{level}_{\max}<us_i\leqslant w_{\max}\\ 0 & \text{others}\end{cases} \tag{5}$$

其中:$w_{\max}$ 表示拟定的大数据端的最高安全水平,见定义 4;us_i 代表大数据端的动态信任因子,见式(3);$\text{level}_{\max}$ 代表平均安全水平。

定义 6 如果大数据实体之间的安全水平处于大数据的最大动态信任因子安全权限之内,而对于其他的情况大数据端统一设定它的访问权限为可选择,则认为是弱安全。

$$\text{info}_{\text{weak}}(i,j)=\begin{cases}1 & \text{if } us_i \leqslant w_{\max}\\ (1-(w_{\max}-us_i))/\text{level}_{\max}-w_{\max} & \text{others}\end{cases} \tag{6}$$

如果大数据端的安全水平低于大数据端设立的最小动态信任因子标准,则认为是非安全。

$$\text{info}_{\text{no}}(i,j)=us_i/\text{level}_{\max} \tag{7}$$

定义 7 对于客户任务实体(MT)和大数据端合法性能的函数关系为:如果大数据端的数据合法水平处于大数据端的最高数据合法验证级别范围之内,对于其他的情况大数据端统一设定它的访问权限为 0,则认为大数据端的数据符合强合法。

$$\text{law}_{\text{super}}(i,j)=\begin{cases}1 & \text{if } \text{ul}_i \leqslant w_{\max}\\ 0 & \text{others}\end{cases} \tag{8}$$

定义 8 如果大数据端的数据合法水平处于大数据端的最高数据合法验证级别范围之内,对于其他的情况满足指数函数关系,则认为大数据端的数据符合弱合法。应该对于该类数据进行可选择执行。

$$\text{law}_{\text{weak}}(i,j)=\begin{cases}1 & \text{if } ul_i < w_{\max}\\ 1-\exp(-w_{\max}+1-\text{ul}_i))/ & \text{others}\\ \exp(w_{\max}-1)\end{cases} \tag{9}$$

定义 9 如果大数据端的数据合法水平处于指数函数关系之下,则认为大数据端的数据不合法,应该不执行该类数据。

$$\text{law}_{\text{no}}(i,j)=1+\exp(ws_i)/1-\exp(-1) \tag{10}$$

其中:客户任务实体的合法性验证 ul_i 动态因子属于[0,1]的平滑区间,而安全性验证由大数据服务提供方的动态信任因子决定。在开始对大数据进行资源整理时,假设大数据不安全信息的识别率为 Tj_i,那么随着时间的增加,新的不确定的数据增加,因此整个大数据服务端对不安全数据识别可靠性能应该降低。假设客户任务实体的数据信息单元在大数据服务端被验证的时间为 t,则不安全报文的识别率为 $SS_i=\exp(MTj_i\times t)$,同样对于客户任务实体和大数据保密性能的函数关系为

$$\text{info\_trust}(i,j)=n_1\times\text{law}_{\text{no}}(i,j)+n_2\times\text{info}(i,j) \tag{11}$$

其中:n_1、n_2 为大数据服务端对大数据合法性和安全性的判断的权值。当权值越高,则数据的保密性也越强。

2.3 不安全样本证据的获取与分析整理

大数据服务端和客户终端的分析样本非常重要,因为只有拥有足够大的样本,才能证明该方案的合理性。

根据 2.2 节定义设定大数据端的集群服务器为 m 个,其中 n 个节点服务器($1\leqslant n<m$)组成一个样本信息获取运算(MIU)单元,假设进入样本信息获取运算单元的数据报 α 服从泊松分布,那么在数据报进入云端的节点服务器时,就开始排队等待,等待的时间层参数 $1/\alpha$(即等待时间为 α)的指数分布,大数据端服务器的选择服务时间满足参数为 δ 的指数分布。当某数据报的等待时间超过了设定时间时,则数据报丢失,并产生一个反馈报文,要求重新发送进行新一轮等待。

设样本信息获取运算(MIU)单元的服务状态复合马尔可夫理论随机决策,MIU 的工作状态可以用一个二维数组表示为 $[J_{(t)},k_{(t)}]$,其中 $J_{(t)}$ 为数据报文的请求队列的样本长度,而大数据服务器集群对于终端提交上来的数据报文的识别率与报文的包头有关。根据马尔可夫理论,不安全文件的识别率具有遍历性,其概率分布表示为

$$A_{I,J}=\lim A[J_{(t)}=i,K_{(t)}=j] \tag{12}$$

其中:$i=0,1,2$; $j=0,1,2,\cdots$。

而对于不安全文件的识别率,应该随着大数据集群服务器的病毒库实时更新而变化。

2.4 动态 SAT 双向防御机制的算法执行流程

本文的系统架构在架设的时候,主要从大数据的数据存储量和运算能力进行考虑,同时也借鉴了文献[13,14]的大数据的资源分配算法和负载均衡思想,充分考虑了服务器的运行能力,各个作业类似多线程进行工作互不影响。当探测到有需要处理的海量数据时,大型服务集群就会自动进行运算处理,并由分析中心进行分析,然后将数据进行整合加工和数据转换,最后下发给连接到大数据端的终端,由客户端进行选择操作工作。而相关安全的健全保证则依赖 2.2 节的行为定义。

动态 SAT 双向防御机制的算法步骤如下:

a)初始化,将一个等待处理的任务进行排序,排序队列为 $E_i(i=1,2,3,4,5,\cdots,n)$。

当需要对信息数据进行加工处理时,海量数据池进行数据传输,S 层进行数据的内部拆分和任务调度。任务调度主要根据粒子算法来排斥局部最优解和收敛速度慢的问题。同时引入逆转变异策略进行初级检测,如果是可信有效数据,则进行处理;如果是非有效数据,则退出作下一轮等待。任务 $T(n_i)$ 满足:

$$T(n_i)=\frac{\sum n_i t_i(n_i)}{M}\quad i=1,2,3,4,5,\cdots,n \tag{13}$$

其中:$T(n_i)$ 代表可拆分的有效可信数据;$\sum_i n_i=E$,E 代表单位时间内处理大数据环境下的总的任务数;n_i 表示在资源 i 上分配的 n_i 个任务。

b)有效数据进行输入,登录成功,数据进行排序等待,在进入 S 层拆分数据前,考虑搭建一个判定模块,判定模块会对数据的操作行为由云端进行实时监控并随时存储信息。如果队列不为空,则进行拆分;如果队列为空,则进行标记,并递归进入到下一个等待队列。

```
{method()
for(i=0;i<queue.length;i++){
    …… if(queue!==null){
        continue;
        }else{
        mark++;
        this.method();
        }
    }
}
```
(14)

c)S 层进行数据拆分,信息整理模块将输入的有效信息进行整理和有序拆分;拆分成功的数据进行下一分析层模块处理。

如果数据的总任务数 E 可以分成 $E_i(i=1,2,3,\cdots,n)$ 个子任务,其分别对应 $m_i(i=1,2,3,\cdots,n)$ 条可信记录,则此分割的函数满足:

$$\text{GiNi}(E)=\frac{m_1}{m}\text{GiNi}(E_1)+\frac{m_2}{m}\text{GiNi}(E_2)+\cdots+\frac{m_n}{m}\text{GiNi}(E_n) \tag{15}$$

d)A 层进行数据分析,数据分析层模块将整理的信息进行分析并同以往数据库进行比对,如果确定有重复信息则进行筛除,并实时更新数据库,以降低信息冗余度。经过筛选后的信息将会输送到 T 层处理。寻找冗余信息点,对于一个离散型的

冗余数据,假设有 2^n 种划分方法,因此将云端的信息点进行时间排序 $V_1, V_2, V_3, V_4, \cdots, V_n$,从小到大依次计算所有信息点的 GiNi 值,从而利用二分法找到最合理的分析对象。

e)T 层进行数据的包装和传输,根据计算任务 E 中各属性列的 GiNi 指数及数据样本索引号,传送至数据包装中心;数据包装中心根据索引号生成记录所在节点的哈希表,进行数据的筛选。经过筛选的数据重新包装处理后,将信息输出给需要接收的信息单元。

f)数据库同步进行信息更新。

g)数据输出判定,数据信息经过一个 SAT 模型进行处理分析后决定是否进行下一轮操作;同时 SAT 中的安全判定模块分析是否将这个包装的数据信息通知所有云终端。

h)如果不需要进行下一轮信息的处理则退出;如果需要下一轮处理,则重新执行。动态 SAT 双向防御架构算法流程如图 1 所示。

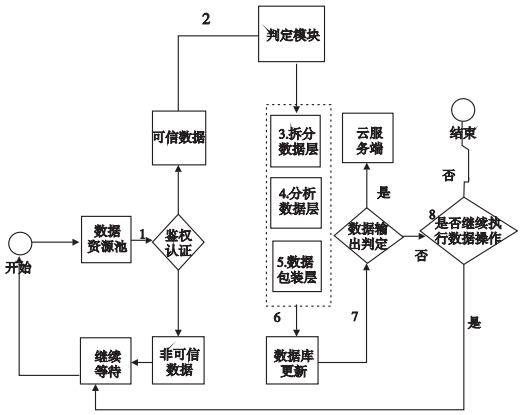

图1　SAT双向安全防御模型流程图

3　测试实验

3.1　实验分析

为验证上述 SAT 双向大数据安全防御系统模型和相关算法的正确性和有效性,利用 Smartbits 公司的网络吞吐量平台进行了仿真测试,并参考了文献[15]的阶段调度算法。算法参数作如下设置:大数据安全任务测试总群为 100 个,最大回归测试次数为 1 000 次(10 h 内的性能测试),报文长度为可选长度(64 ~ 65534 Byte),测试了五种云平台数据报文的测试类型和 500 个数据任务的测试用例、测试结果都取每 100 组测试结果的平均值、测试方面主要包括新建连接速率测试、并发连接数测试、应用层网络流量测试、检测率测试。测试代码运用 VML 脚本语言编写,使用 Tomcat 作为服务器,实验监测外设配 置为英特尔 Core i3-2100 CPU 2.1 GHz,4. 00 GB 内存,Windows7 操作系统。实验环境如图 2 所示。

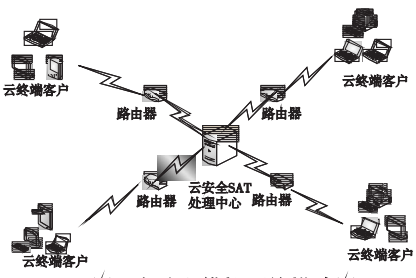

图2 实验室模拟环境搭建图

3.2　实验模拟和描述

1)新建连接速率测试　在 Smartbits600 上构建分别从 WAN 侧和 LAN 侧的大流量数据访问报文,测试业务为用户访问测试,经过测试,不带端口扫描的报文连接数可以稳定在 25 万 HTTP 连接/s 以上,而模拟加载端口扫描的攻击报文,其连接数目也可以连接在 20 万 HTTP 连接/s 以上,充分显示了该模型的超强应用处理能力。连接速率测试图如图 3 所示。

2)并发连接数测试是在 Smartbits 的主控板上进行并发连接数测试,并逐渐增加模拟用户并发连接,测试业务为下载测试。经过测试,即使在极不稳定的情况下,模拟用户种子的连接数量仍然可以维持在 350 万端点的连接水平,由此可知连接性能符合要求。并发连接数测试图如图 4 所示。

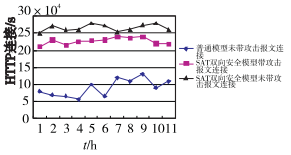

图3　连接速率测试图

3)应用层网络流量测试　在 Smartbits 的主控板上进行网络流量测试,测试业务为组播测试,经过多种报文的组合发包测试,在应用层报文的吞吐量可以稳定在 10 Gbps 左右,并且不因为业务的改变而影响流量。其性能完全可以满足用户万兆光网络数据传输的应用需求。应用层网络流量测试图如图 5 所示。

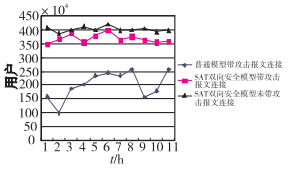

图4 并发连接数测试图

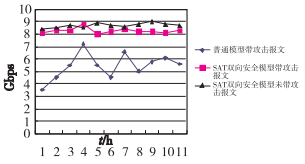

图5　应用层网络流量测试图

4)检测率测试　在 Smartbits 业务板的 WAN 侧的端口上分别构造长度不等的 TCP 和 UDP 连接报文,测试业务为报文识别测试;在发包过程中,同时利用其他报文进行干扰和调整压力,并记录识别情况。

5)丢包测试　在 Smartbits 业务板上的 LAN 侧对 WAN 侧进行发包,然后在 WAN 对包进行捕获,LAN 逐渐调整发包数量和压力,测试业务为 ICMP。经过测试未出现报文的丢失情况,请求均得到了响应,数据包无漏检情况。

3.3　实验分析

仿真测试将 SAT 双向安全模型带攻击报文的状态和不带攻击报文的状态与普通模型带攻击报文状态进行了对比实验。从图 3 中的对比数据可以看出,不管是否带攻击报文,本文所提出的模型曲线最平稳,振荡最小,表明该算法模型稳定性最好;而普通模型受到数据报文的攻击时稳定性能明显较低,即从任务的执行角度分析,本文提出的算法模型要优于传统的大数据安全的安全防御模型。

在图 4 中,普通模型连接端点的变化幅度较大,端点连接数目会随着攻击报文的增加而降低,而本文提出的算法模型端点连接稳定,平均掉线数值在 20 个端点左右,端点之间的稳定连接数差值不超 30 个,说明该算法在端点的连接数最能保证大数据安全的稳定性。

在图 5 中,当云端对终端进行组播业务测试时,普通模型的

业务码流会因为视频业务的增大而互相产生干扰,而本模型提出的算法因为很好地节约了队列的等待排序问题,同时终端对于大数据端的组播业务进行了可选执行,对于超过业务承载量的则进行排队等待。所以无丢包现象发生,视频业务正常。

实验结果表明,经过基于 SAT 双向安全评估模型下大数据动态安全 SAT 双向防御方案的实验测试,当某客户在与云端交互的过程中,逐渐加大如非连接、扫描重要端口,入侵防御测试,检测率测试,应用层网络流量测试等恶意行为比例时,总体信任度的变化情况。随着恶意行为比例的增大,SAT 安全模型的信任值与传统的安全防护方案相比,在合法性、安全性认证鉴权上,以及终端对云端下发业务的可选执行上均体现了优势,SAT 双向安全模型能更快地监测出有风险的用户,从而有利于大数据端下发授权,降低大数据的数据交互风险。

4 结束语

通过本次测试可以看出,这项技术方案的优势在于超越了以前拦截 Web 威胁的传统方法和以前的通过把大多数特征码文件保存到数据库中检测的技术方案,得以在 Web 威胁到达用户终端或大数据端之前即可对其予以拦截。同时该方案降低了大数据端的计算时间开销,使云端对用户的信任级别和权限作出更快的判断。实验表明,本模型适合作为云数据时代的海量信息处理的防御系统模型,它为客户实体终端和大数据服务端的数据交互保障提供了新的理论支撑,同时对以后云数据冗余去重、提供代码的容错性有了新的借鉴意义。该方案不但在功能架构上可以满足当前大数据网络安全防御的应用需求,并且对目前危害性最大的未知威胁也有很高的防范能力,因此可以在万兆网络的核心网及骨干网进行部署和商用。但是本文依然有不足之处,例如大数据安全信任因子求解的算法,求解的质量有待于进一步提升;同时本文尚未对不同粒子种群情况下算法的比较进行分析,将在后续研究中继续进行。

参考文献:

[1] 李虹,李昊. 可信云安全的关键技术与实现[M]. 北京:人民邮电出版社,2010:5-15.

[2] 林闯,彭雪海. 可信网络研究[J]. 计算机学报,2005,28(5):751-758.

[3] 周茜,于炯. 云计算下基于信任的防御系统模型[J]. 计算机应用,2011,31(6):1531-1535.

[4] 吴慧,于炯,于斐然. 云计算环境下基于信任模型的动态级访问控制[J]. 计算机工程与应用,2012,48(23):102-106.

[5] JAMEEL H. A trust model for ubiquitous systems based on vectors of trust values[C]//Proc of the 7th IEEE International Symposium on Multimedia. Washington DC: IEEE Computer Society, 2005: 674-679.

[6] 李小勇,桂小林. 可信网络中基于多维决策属性的信任量化模型[J]. 计算机学报,2009,32(3):405-415.

[7] 王珊,高迎,程涛远,等. 网格环境下基于行为的双层信任模型的研究[J]. 计算机工程与应用, 2005,25(9):1974-1977.

[8] 冯小靖,潘郁. 云计算环境下的 DPSO 资源负载均衡算法[J]. 计算机工程与应用,2013,6(6):105-108.

[9] 陈小全, 张继红. 基于改进粒子群算法的聚类算法[J]. 计算机研究与发展,2012,49(S1):287-291.

[10] 韩凌波,王强,蒋正锋,等. 一种改进的 K-means 初始聚类中心选取算法[J]. 计算机工程与应用,2010,46(17):150-152.

[11] 覃雄派,王会举,杜小勇,等. 大数据分析——RDBMS 与 MapReduce 的竞争与共生[J]. 软件学报,2012,23(1):32-45.

[12] 沈晴霓,杜虹,文汉,等. 一种基于完整性度量架构的数据封装方法[J]. 计算机研究与发展, 2012,49(1):210-216.

[13] 王珊,高迎,程涛远,等. 遗传算法与蚂蚁算法的融合[J]. 计算机研究与发展, 2003,40(9):1352-1356.

[14] 华夏渝,郑骏,胡文心. 基于云计算环境的蚁群优化计算资源分配算法[J]. 华东师范大学学报:自然科学版,2010(1):127-134.

[15] ZHANG Ze-hua,ZHANG Xue-jie. A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation[C]//Proc of the 2nd International Conference on Industrial Mechatronics and Automation. 2010:240-243.

(上接第 1469 页)

[3] 周剑, 张明新. 软件可信评估综述[J]. 计算机应用研究, 2012, 29(10):3609-3613.

[4] Trusted Computing Group. TCG architecture overview[EB/OL]. (2006). https://www.trustedcomputinggroup.org.

[5] 孟琳琳, 赵伟男, 孙海龙,等. Web 服务可信证据收集与评估机制研究[J]. 计算机科学与探索, 2011, 5(7):642-651.

[6] 林闯. 可信网络研究[J]. 计算机学报, 2005, 28(5): 751-758

[7] 王怀民, 唐扬斌, 尹刚, 等. 互联网软件的可信机理[J]. 中国科学:E 辑, 2006,36(10): 1156-1169.

[8] 梅宏, 曹东刚. 软件可信性:互联网带来的挑战[J]. 中国计算机学会通讯, 2010,6(2):20-27.

[9] ABDUL R A, HAILES S. A distributed trust model[C]//Proc of New Security Paradigms Workshop. New York:ACM Press,1997: 48-60.

[10] BETH T, BORCHERDING M, KLEIN B. Valuation of trust in open network[C]// Proc of the European Symposium on Research in Security. Berlin:Springer-Verlag, 1994: 3-18.

[11] JøSANG A J. An algebra for assessing trust in certification chains [C]// Proc of Network and Distributed System Security Symposium. San Diego:The Internet Society,1999.

[12] 徐锋, 吕建, 郑玮, 等. 一个软件服务协同中信任评估模型的设计[J]. 软件学报,2003,14(6): 1043-1051.

[13] 王远, 吕健, 徐锋, 等. 一个适用于网构软件的信任度量及演化模型[J]. 软件学报, 2006, 17(4): 682-690.

[14] GRANDISON T,SLOMAN M. A survey of trust in Internet applications[J]. IEEE Communications Surveys and Tutorials,2003,3(4):2-16.

[15] ZADEH L A. Fuzzy sets[J]. Information and Control, 1965, 8(3):338-353.

[16] 丁学雷, 王怀民, 王元元, 等. 面向验证的软件可信证据与可信评估[J]. 计算机科学与探索, 2010, 4(1):46-53.

[17] 刘旭东, 郎波, 谢冰,等. 软件可信分级规范:版本 2.0,国家高技术研究发展计划(863)重点项目"高可信软件生产工具与集成环境"技术文档 TRUSTIE-STC V2[R]. 2009.

[18] 赵俊峰, 王亚沙, 谢冰, 等. 一种支持构件服务质量的构件管理框架[J]. 电子学报, 2004, 32(12A): 165-168.

[19] 梅宏, 谢涛, 袁望洪, 等. 青鸟构件库的构件度量[J]. 软件学报, 2000, 11(5):634-641.

[20] 吴晓娜, 李必信, 李超. 一种个性化的可信服务选择方法[J]. 东南大学学报:英文版,2013,29(1):16-21.

[21] 石双元,陈朝德. 业务构件可信性的度量研究[J]. 计算机工程与科学, 2009, 31(5):84-86.

[22] 毛国蓉, 李雪静, 葛孝, 等. 基于软件构件质量模型的度量及应用[J]. 计算机应用与软件, 2005, 22(5):1-4.