

## 假设检验模型网络异常监控算法的研究和实现

高 强<sup>1</sup>, 丁岳伟<sup>2</sup>, 何 璐<sup>3</sup>

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

(kyo-gao@163.com)

**摘 要:**针对传统安全监控方式对网络异常判断的不足,提出一种基于假设检验的网络异常分析算法。该算法提出一个新的概念“网络性能值”来描述网络状态。计算在网络正常情况下该网络主机的“网络性能值”分布参数,采集在检查的时间段内一定数量网络性能值样本,通过假设检验方法判断在该时间段内网络是否异常。对该算法进行程序测试,可以得出该算法与传统的监控方法相比降低了网络负荷,并提高了时间段内网络安全判断的正确率。

**关键词:**网络监控;假设检验;网络性能;网络安全;网络负载

**中图分类号:** TP309.1 **文献标志码:** A

## Research and realization of network security monitoring algorithm based on hypothesis verification model

GAO Qiang<sup>1</sup>, DING Yue-wei<sup>2</sup>, HE Lu<sup>3</sup>

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** To overcome the disadvantage of traditional network security monitor system, this paper introduced a new definition—network performance (NP) which can describe the status of network. Calculate the distributing parameter in the security network environment was calculated, and the amount of NP during the monitoring time was collected, then whether the network is secure was estimated by using hypothesis verification. At last, the result of experiment shows the new algorithm can reduce the network burthen, and increase the accuracy of network security estimation.

**Key words:** network monitoring; hypothesis verification; network performance; network security; network load

### 0 引言

网络安全监控是国内外研究的一个热点,提出了很多的网络监控模式,而这些网络监控大多采用集中式<sup>[1]</sup>,即采用 SNMP 协议或者其他一些网络协议即时获取网络流量<sup>[2]</sup>、主机状况等信息<sup>[3]</sup>,从而对即时的网络是否安全做出判断<sup>[4]</sup>。但是随着网络规模的成倍增加,该模式使得网络流量负荷急剧增长,而且更重要的是传统模式获取的是网络即时的性能状态,也许正常与异常在某一时刻是偶然事件,从而导致对网络安全判断的偏差。

为弥补上述的不足,本文提出一种基于假设检验的网络异常分析算法,该算法采用假设检验来判断一段时间内网络是否安全正常。在该算法中主要有如下步骤:

- 1) 在正常网络情况下大量获取并记录各种参数值,从而得到网络性能值的分布情况;
- 2) 在一定时间段内获取一定数量的网络参数值,采用假设检验的方法对在该时间段内网络异常进行判断;
- 3) 当网络出现异常时候进行安全处理,当未出现异常的时候将数据加入 1) 的数据中,更新网络性能分布情况。

通过程序的实现和实验测实验证,本文方法可以有效提高网络监控中网络异常判断的正确性。

### 1 网络异常检验算法设计

针对传统网络监控系统网络流量负荷大、网络判断能力差的缺点,设计一种基于假设检验的网络安全异常检验算法,

该算法结合网络正常状态和监控期间的状态来判断在该时间段内网络是否异常,这既能减少即时的流量负荷,也可以提高网络异常判断的偶然性。

#### 1.1 算法参数定义

**定义 1** 网络连接时间  $CT$ 。重要的网络状况属性,代表连接网络主机所需要的时间。具体的说是指 TCP/IP 三次握手所需要的时间,即包括源主机发送连接请求所需时间  $t_1$ ,目的主机发送确认应答时间  $t_2$ ,源主机再次发送确认应答时间  $t_3$ ,而  $CT = t_1 + t_2 + t_3$ 。该值取值范围为  $(0, +\infty)$ 。

**定义 2** 数据吞吐率  $DTP$ 。重要的网络状态属性,代表在一定时间  $T$  内不同主机间数据传输处理的大小  $D$  的速率,  $DTP = D/T$ 。该值主要取决于主机性能、网络接口卡以及其所处的网络。该值单位取为 Mbps,取值范围为  $[0, D_{\max})$ ,其中  $D_{\max}$  取决于硬件和网络协议。

**定义 3** 数据完整性  $DI$ 。重要的网络状态属性,代表数据传输中正确处理报文数  $n$  占总共传输报文  $m$  的比例,即  $DI = (n/m) \times 100\%$ ,其取值范围为  $[0, 1]$ 。

**定义 4** 数据安全性  $DS$ 。重要的网络安全属性,代表在该网络中关键信息是否加密传输。 $DS$  的取值范围只有 0 和 1,  $DS = \begin{cases} 0, & \text{未加密传输} \\ 1, & \text{加密传输} \end{cases}$ 。

**定义 5** 网络性能值  $NP$ 。网络状况的综合属性,算法中主要通过该值来分析判断网络安全异常情况,该值是综合上述 4 个网络参数值所得。其计算公式如下:

收稿日期:2009-04-16。 基金项目:上海市研究生创新基金项目(JWCXSL0902)。

作者简介:高强(1987-),男,江西抚州人,硕士研究生,主要研究方向:软件工程、信息安全; 丁岳伟(1961-),男,上海人,教授,主要研究方向:信息安全、软件工程、知识挖掘; 何璐(1987-),女,江西抚州人,硕士研究生,主要研究方向:并行处理、网络计算。

$$NP = \frac{\lambda_1 \times \frac{t_2 \times DTP - t_1 \times CT}{\sum_{i=1}^2 t_i} \times DI + \lambda_2 \times DS}{\sum_{i=1}^2 \lambda_i} \quad (1)$$

在一般情况下  $DTP$ 、 $CT$  和  $DI$  三者的值会互相制约,而  $DS$  相对来说是独立的。所以该公式主要分为两部分:一部分为网络性能,由  $DTP$ 、 $CT$  和  $DI$  组成;另一部分为网络安全性,由  $DS$  组成。 $\lambda_1$ 、 $\lambda_2$  分别为这两大部分的权值,最后求加权平均得到网络性能值  $NP$ 。而在网络性能部分中  $DTP$ 、 $(-CT)$  和  $DI$  成正比, $t_1$  和  $t_2$  为  $CT$  和  $DTP$  的权值,求加权平均后乘以  $DI$  便得到网络性能部分的值。在不同的网络监控中有不同的侧重,可以通过改变权值  $t_1$  和  $t_2$ , $\lambda_1$  和  $\lambda_2$  以适应。

## 1.2 算法详细设计

引入上述 5 个定义后,可以通过这些关键参数值详细设计基于假设检验的网络异常监控算法。图 1 概括地表示了该算法的一些基本步骤。

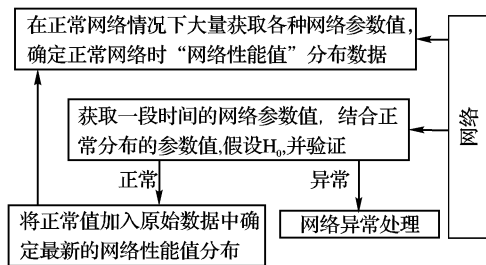


图 1 假设检验网络安全异常监控算法

如图 1 所示,该算法大致包括三个步骤:采集、监控和处理。

### 1.2.1 采集

环境为正常网络环境,即网络性能部分正常,网络安全部分无安全问题。在这个步骤中,获取足够多的网络参数值  $DTP$ 、 $CT$ 、 $DI$  和  $DS$ ,通过式(1) 计算得到大量的网络性能值  $NP$ 。可以发现, $NP$  值的分布一般都是属于  $(\xi_0, \sigma^2)$  正态分布。其中:

$$\text{期望 } \xi_0 = \sum_{i=1}^n NP_i / n \quad (2)$$

$$\text{方差 } \sigma^2 = \sum_{i=1}^n (\xi_i - \xi_0)^2 / n \quad (3)$$

在正常网络环境下获得的  $NP$  值的近似正态分布图如图 2 所示,以  $\xi_0$  为期望, $\sigma^2$  为方差。

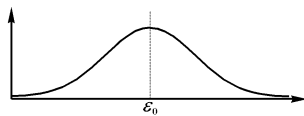


图 2  $NP$  值近似正态分布图

### 1.2.2 监控

监控步骤是该基于假设检验的网络异常算法的核心部分,主要采用概率统计的假设检验方法,再结合采集得到的数据来判断。

按照数理统计的思维,需要获取监控期间样本容量为  $j$  的  $NP$  样本值  $NP_j$ ,通过这些样本值来判断有关总体的一个假设是否成立,而该总体的假设即是整个一段监控时间内网络是否安全,这也是该算法最终所需的结论。

1) 该算法通过样本值  $NP_j$  计算得到样本均值  $\bar{X} =$

$$\sum_{i=1}^j NP_j / n_0$$

2) 确立假设命题,在该算法中,可以把需要解决的问题转化为一个标准的数理统计的问题模型:

一个分布属于  $(\xi, \sigma^2)$  正态分布,抽取一些样本值,判断其  $\xi$  相比于  $\xi_0$  是否有显著差异。

通过该问题模型,可以提出两个假设:原假设  $H_0$ :网络安全,  $\xi = \xi_0$ ;备择假设  $H_1$ :网络不安全,  $\xi \neq \xi_0$ 。原假设与备择假设相互对立,两者有且只有一个正确,所谓假设检验问题就是要判断原假设  $H_0$  是否正确,决定接受还是拒绝原假设,若拒绝原假设,就接受备择假设。

3) 对假设进行推断,奈曼(Neyman)和皮尔生(Pearson)提出一个原则:原假设要受到维护,不轻易被否定;这种只控制弃真而不考虑取伪的假设检验,即是显著性检验。本算法就是采用该检验方式。 $\alpha$  即是弃真的概率,称为显著性水平,最常用的  $\alpha$  值为 0.01、0.05、0.10 等。一般情况下,根据研究的问题,如果犯弃真错误损失大,为减少这类错误, $\alpha$  取值小些,反之, $\alpha$  取值大些。

给定显著水平,当原假设  $H_0$  为真时,临界值  $C$  应满足:

$$P(|\bar{X} - \xi_0| \geq C) = \alpha \quad (4)$$

由于网络性能值  $NP \sim N(\xi_0, \sigma^2)$ ,所以样本容量为  $j$  的平均网络性能值  $\bar{X}$  服从  $N(\xi_0, \sigma^2/j)$ ,令:

$$Z = \frac{\bar{X} - \xi_0}{\frac{\sigma}{\sqrt{j}}} \quad (5)$$

于是有:

$$P(|Z| \geq \frac{C}{\frac{\sigma}{\sqrt{j}}}) = \alpha \quad (6)$$

$\therefore Z \sim N(0, 1)$ ,并解式(4) ~ (6)

$$\therefore C = z_{\alpha/2} \times \sigma / \sqrt{j}$$

得到以下结论:当  $|\bar{X} - \xi_0| \geq C$  时,原假设处于拒绝域,否定原假设,取备择假设  $H_1$ ,即网络异常,不安全;相反地,当  $|\bar{X} - \xi_0| < C$  时,原假设成立,取原假设  $H_0$ ,即网络正常,安全。

### 1.2.3 处理

该步骤用于更新原分布,获得更为准确的正态分布参数。即当原假设成立,即网络正常时,将监控时获取的容量为  $j$  的样本加入采集的数据列中,重新计算其分布参数,如果样本中存在偶尔异常,进行处理(比如数据安全性);当原假设不成立,即网络异常时,对该异常进行处理。

## 2 算法的实现及性能分析

### 2.1 算法实现

为了测试该算法在实际中的性能情况,通过 C++ 语言对该算法程序实现。首先使用 SOCKET 编程获取如下 4 个关键参数:网络连接时间  $CT$ 、数据吞吐率  $DTP$ 、数据完整性  $DI$  和数据安全性  $DS$ 。其中前三个参数比较容易获得,而数据安全性是采用混杂网卡模式来截取网络数据包,并分析是否存在敏感字符串。关键代码如下:

```

//设置 SOCK_RAW, 接收所有的 IP 包
int ErrorCode = WSAIoctl( SockRaw, SIO_RCVALL, &dwBufferInLen,
    sizeof( dwBufferInLen ), &dwBufferLen, sizeof( dwBufferLen ),
    &dwBytesReturned, NULL, NULL);
//解析 IP 包, 省略号代表省略部分非关键代码
int DecodeIpPack(char * buf, int iBufSize) { ...DecodeTcpPack
    ( buf + iPhLen, iBufSize); ...}
  
```

```
//解析 TCP 包,并判断是否含有敏感字符,如有就代表网络传
//输非安全
int DecodeTcpPack ( char * TcpBuf, int iBufSize) { ... char *
TcpData = TcpBuf + TcpHeaderLen; ...
if ((ParamDecode) && (iBufSize > 40))
{
    if ((!strSensitive) || (strstr(TcpData, strSensitive)))
        { stop_Anlys = true; m_ds = 1; }
}
}
```

算法核心假设检验部分的关键实现代码如下:

```
//采集部分...
np[i] = (t1 * ((k1 * m_dtp[i] - k2 * m_ct[i]) * m_di[i] / (k1 +
k2)) + t2 * m_ds[i]) / (t1 + t2);
for(i=0; i <= num; i++)
{
    np[i] = (t1 * ((k1 * m_dtp[i] - k2 * m_ct[i]) * m_di[i] / (k1 +
k2)) + t2 * m_ds[i]) / (t1 + t2);
    S_np += np[i];
}
S_EXP = S_np/n;
for(i=0; i <= num; i++)
    { S_vnp += (np[i] - S_EXP) * (np[i] - S_EXP); }
S_VAR = S_vnp/n;
...
//监控部分,如采集部分一样获取样本均值
...C = Value_Z * sqrt(S_VAL) / sqrt(Sample_num);
if (fabs(v_exp - S_EXP) >= C)
    { m_security = false; }
else { m_security = true; }
...
```

## 2.2 算法性能分析

测试中,主要就该算法的网络负荷以及监控时间段内网络异常判断的正确率来进行分析。在测试环境中,给定显著性水平  $\alpha$  为 0.05,并采用一般的情况,即各个权值均为平均分配,均为 5。而测试的网络环境是主机均处于同一局域网内,该局域网是主机、交换机和路由器等构成的三层网络体系架构。网络异常判断均只对一台主机进行监控,异常判断正确率测试中,通过一个小程序随机产生异常情况,每一个样本值均做 30 次的测试,计算其正确率。

在网络负荷的实验测试中,首先假定传统模式是固定一小时 60 次即时检验,其网络负荷是一个固定值,所以如图 1 所示,假设检验模式在每小时样本少于 60 个的时候网络负荷远小于传统模式。

在网络异常判断的实验测试中,对比三个测试值,分别是采用非固定次数即时检验的传统模式、非固定样本值的假设检验模式和采用固定次数即时检验的传统模式。值得提出的是传统模式并没有直接进行一段时间内的网络异常判断,所以假定其判断方法为判断样本中异常所占的比例,若大于正常的比例,则判定网络异常,若小于则安全。从图 2 的实验结

果可以明显看出:假设检验模式的正确率远大于非固定传统模式,而只有在样本个数很少的时候假设检验模式才略小于固定传统模式,但此时固定的传统模式占用了相对来说非常大的网络负荷为代价。而且还可以发现在样本取 35 附近的时候正确率几乎达到最高值,再增加样本容量已经不能提升太多正确率。

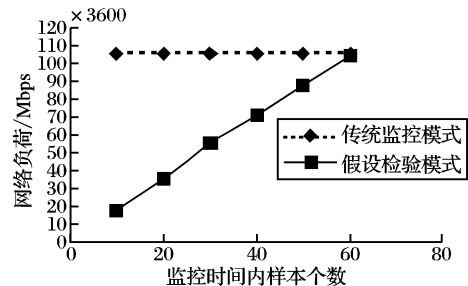


图1 两种模式网络负荷比较

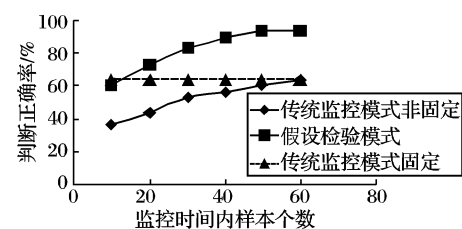


图2 两种模式判断正确率比较

综上测试可以得到结论:基于假设检验模式的网络异常判断比传统模式具有更低的网络负荷,并具能对一段网络时间内网络异常情况进行更为有效正确的判断。

## 3 结语

本文针对传统网络监控模式的缺点设计了基于假设检验的网络异常算法。该算法引入一个网络性能值  $NP$ ,通过采集正常网络情况和监控时间段内的该  $NP$  值样本,运用假设检验的算法来进行网络异常的判断。使用 C++ 语言实现了该算法,并与传统的网络监控模式进行了测试对比。可以得出基于假设检验的网络异常检验能够有效减少网络负荷,提升监控时间段内网络异常判断的能力。

### 参考文献:

- [1] 王新昌,杨艳,刘育楠,等.一种基于局域网络监控日志的安全审计系统[J].计算机应用,2007,27(2):292-294.
- [2] 张小川,陈庄,向勇.基于网络质量监控系统的实现[J].计算机应用,2006,26(S1):201-203.
- [3] 王旭仁,毕学尧,许榕,等.实时网络安全监控系统的设计和实现[J].计算机工程,2005,31(4):209-211.
- [4] 温研,王怀民,胡华平,等.分布式网络行为监控系统的研究与实现[J].计算机工程与科学,2005,27(10):13-15.
- [5] 夏新涛,王中宇.非统计假设检验原理及其应用[J].计量学报,2006,27(2):190-195.

(上接第 2643 页)

- [15] CHOI J, PARK K, KIM C. Analysis of cross-layer interaction in multirate 802.11 WLANs [J]. IEEE Transactions on Mobile Computing, 2009, 8(5): 682-693.
- [16] YUN J. Throughput analysis of IEEE 802.11 WLANs with automatic rate fallback in a lossy channel [J]. IEEE Transactions on Wireless Communications, 2009, 8(2): 689-693.
- [17] RAMACHANDRAN K, KREMO H, GRUTESER M, et al. Scalability

- analysis of rate adaptation techniques in congested IEEE 802.11 networks: An orbit testbed comparative study [C]//IEEE WOWMOM 2007. Espoo, Finland, Piscataway: IEEE, 2007: 1-12.
- [18] CHEN D R, ZHANG Y J. On throughput limit of multi-rate IEEE 802.11 WLANs: Basic access vs. RTS/CTS access [C]// IEEE WCNC 2008. New York: IEEE, 2008: 1414-1419.