基于欧氏空间距离计算的SynFlood攻击检测方法进一步讨论

刘辉宇[1] 陈 凯[1] 彭 涛[1,2] 陈晓苏[1]

(华中科技大学计算机科学与技术学院 武汉 430074)[1] (武汉纺织大学计算机科学学院 武汉 430073)[2]

摘 要 基于TCP协议中Syn,Fin和Rst 3种报文段的关系,提出了一种新的SynFlood攻击检测方法:将Syn,Fin和Rst 3者之间的关系映射到欧氏空间中,将某一时间段内的Syn,Fin和Rst的关系映射为一个点,将无攻击行为存在时的Syn,Fin和Rst之间的关系映射为一条线,分析点与线之间的距离来检测SynFlood攻击,同时使用移动平均技术对上述距离进行平滑处理,以提高检测效率和准确度。实验结果表明,该方法对直接式SynFlood攻击和反射式SynFlood攻击均具有较好的检测准确度,并且产生的误报率较低,数据报文处理能力较高,能够部署于大中型网络的骨干路由器上。

关键词 Syn洪泛攻击,欧氏空间距离,偏离度,移动平均,攻击判别值

中图法分类号 TP393.08 **文献标识码** A

Further Discussion on SynFlood Attack Detection Based on Distance Computation in Space Geometry

LIU Hui-yu[1] CHEN Kai[1] PENG Tao[1,2] CHEN Xiao-su[1]

(School of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan 430074, China)[1]

(School of Computer Science, Wuhan Textile University, Wuhan 430073, China)[2]

Abstract This paper gave a new method to detect the SynFlood attack by analyzing the relationship between Syn segment, Fin segment and Rst segment in TCP protocol. Firstly, the relationship between Syn segment, Fin segment and Rst segment is mapped to Space Geo-metry: the relationship in a given time frame is mapped to one point in Space Geometry while that when no attack behavior exists is mapped to a line in Space Geometry. The distance between the point to the line can hence be used to detect and determine the SynFlood attack. Furthermore, the efficiency and accuracy are improved by using moving average technology which can anti-aliasing the distance discribed above. The experimental result shows that the method can detect the direct SynFlood attack and the reflect SynFlood attack accurately and have low rate of false alarm. Also the method can be deployed to mid-large scale networks because of its high performance for processing data packets.

Keywords SynFlood attack, Distance in space geometry, Deviation, Moving average, Attack discriminant

1 引言

随着计算机网络的发展及其应用范围的不断扩大,安全问题显得日益尖锐和突出,各种攻击手段层出不穷,其中SynFlood攻击是一种比较常见且较难检测和防御的攻击方法。该攻击手段利用了TCP协议的三次握手机制,众多攻击机以一定的速率向目标机发送建立连接的Syn报文段,在收到目标机反馈的Syn+Ack报文段后,不执行三次握手第三阶段,直接丢弃目标机所反馈的Syn+Ack报文段。目标机在收到Syn报文段并反馈Syn+Ack报文段后,会为即将建立的连接预分配资源。这种预分配由超时机制进行资源管理,以防资源不当使用,但只要发送速率超过一定值,就能够使得目标机的资源被迅速消耗,从而无法接受新的连接建立请求,从现象上意味着该目标机瘫痪[1]。

目前,抵御SynFlood攻击的方法主要有Syn Cache[2],Syn Cookies[3],Syn Proxying[4],Syn Defender[5],Syn Kill[6],这些方法大都部署于需保护的服务器上或者服务器所在网络的出口防火墙上,这种防护机制仅限于保护确定的某些服务器,而无法检测和防御全网的SynFlood攻击。

针对SynFlood攻击的检测方法主要有基于规则的方法、统计学方法、有限状态机方法和模式匹配方法[7]。本文前期工作提出了一种基于欧氏空间距离计算的SynFlood攻击检测方法[8],该方法具有较高的数据处理能力,可部署在互联网的主干路由器上,并具有较好的直接式SynFlood攻击检测能力,能够有效地在早期进行直接式SynFlood攻击检测,使得后续进行直接式SynFlood攻击早期阻断成为可能。但由于评价特征值的选取具有一定的局限性,无法有效地检测反射式SynFlood攻击,此外检测算法参数的选择不是特别恰当,

到稿日期:2011-01-13 返修日期:2011-03-30 本文受国家自然科学基金(60873030)资助。

刘辉宇(1978—),男,博士生,讲师,主要研究方向为计算机网络安全技术、计算机网络应用技术,E-mail:liuhuiyu@mail.hust.edu.cn;陈 凯(1977—),男,博士生,讲师,主要研究方向为计算机网络安全技术;彭 涛(1981—),男,博士生,讲师,主要研究方向为计算机网络安全技术;陈晓苏(1953—),男,硕士,教授,主要研究方向为计算机网络安全技术、计算机网络应用技术。

且后期对偏离度的平滑处理不够周全，导致误报率较高。本文将对该方法做进一步探讨，提出一组新的评价特征值，以支持反射式 SynFlood 攻击检测，并引入一种新的偏离度平滑技术，以期有效降低误报率。

第 2 节描述检测方法中使用的数学工具；第 3 节详细阐述检测方法；第 4 节将在一个实际的 SynFlood 攻击环境中对方法进行检验，并进一步讨论算法中几个主要参数对检测结果的影响；最后是结论。

2 数学模型及攻击检测思想

2.1 基本术语

定义 1(Syn 报文段数目，N_{syn}^{T}) 某一时间段 T 内网络中 Syn 字段为 1 的报文段数目。

定义 2(Fin 报文段数目，N_{Fin}^{T}) 某一时间段 T 内网络中 Fin 字段为 1 的报文段数目。

定义 3(Rst 报文段数目，N_{Rst}^{T}) 某一时间段 T 内网络中 Rst 字段为 1 的报文段数目。

根据 TCP 连接建立过程的三次握手机制，在一段时间内如果所有的 TCP 连接都是正常的，不存在攻击行为，也不存在任何异常的 TCP 连接(该状态后文称为理想状态)，那么 N_{syn}^{T}，N_{Fin}^{T} 和 N_{Rst}^{T} 应满足式(1)给出的关系：

$$\begin{cases} N_{syn}^{T}=N_{Fin}^{T} \\ N_{Rst}^{T}=0 \end{cases} \tag{1}$$

在实际网络中，由于存在链路故障、失效连接、病态路由、空连接等情况以及时间段 T 的持续时间很难“恰当”确定，式(1)难以被严格满足，N_{syn}^{T}、N_{Fin}^{T} 和 N_{Rst}^{T} 之间的关系实际如式(2)所示：

$$\begin{cases} N_{syn}^{T}\neq N_{Fin}^{T} \\ N_{Rst}^{T}>0 \end{cases} \tag{2}$$

当网络中存在直接式 SynFlood 攻击时，N_{syn}^{T}，N_{Fin}^{T} 和 N_{Rst}^{T} 之间的关系如式(3)所示：

$$\begin{cases} N_{syn}^{T}\to\infty \\ N_{Fin}^{T}\to 0 \\ N_{Rst}^{T}\to 0 \end{cases} \tag{3}$$

当网络中存在反射式 SynFlood 攻击时，N_{syn}^{T}，N_{Fin}^{T} 和 N_{Rst}^{T} 之间的关系如式(4)所示：

$$\begin{cases} N_{syn}^{T}\to\infty \\ N_{Fin}^{T}\to 0 \\ N_{Rst}^{T}\to\infty \end{cases} \tag{4}$$

定义 4(同步判别因子，SDF_T) 某一时间段 T 内的 Syn 报文段数(N_{syn}^{T})与 Fin 报文段数(N_{Fin}^{T})的比值：

$$SDF_T=\frac{N_{syn}^{T}}{N_{Fin}^{T}} \tag{5}$$

称之为同步判别因子。

定义 5(复位判别因子，RDF_T) 某一时间段 T 内的 Rst 报文段数(N_{Rst}^{T})与 Fin 报文段数(N_{Fin}^{T})的比值：

$$RDF_T=\frac{N_{Rst}^{T}}{N_{Fin}^{T}} \tag{6}$$

称之为复位判别因子。

根据式(1)，理想状态下的同步判别因子和复位判别因子取值为：

$$\begin{cases} SDF_T=\dfrac{N_{syn}^{T}}{N_{Fin}^{T}}=1 \\ RDF_T=\dfrac{N_{Rst}^{T}}{N_{Fin}^{T}}=0 \end{cases} \tag{7}$$

根据式(2)，在实际网络中，如果不存在攻击行为，同步判别因子和复位判别因子取值为：

$$\begin{cases} SDF_T=\dfrac{N_{syn}^{T}}{N_{Fin}^{T}}\approx 1 \\ RDF_T=\dfrac{N_{Rst}^{T}}{N_{Fin}^{T}}\approx 0 \end{cases} \tag{8}$$

根据式(3)，网络中存在直接式 SynFlood 攻击时，同步判别因子和复位判别因子取值为：

$$\begin{cases} SDF_T=\dfrac{N_{syn}^{T}}{N_{Fin}^{T}}\to\infty \\ RDF_T=\dfrac{N_{Rst}^{T}}{N_{Fin}^{T}}\to 0 \end{cases} \tag{9}$$

根据式(4)，网络中存在反射式 SynFlood 攻击时，同步判别因子和复位判别因子取值为：

$$\begin{cases} SDF_T=\dfrac{N_{syn}^{T}}{N_{Fin}^{T}}\to\infty \\ RDF_T=\dfrac{N_{Rst}^{T}}{N_{Fin}^{T}}\to\infty \end{cases} \tag{10}$$

2.2 判别元组

定义 6(判别元组，DTuple) 某一起始时刻为 t、持续时间为 T 的时间段内同步判别因子、复位判别因子及 t 构成的三元组：

$$(SDF_T, RDF_T, t)$$

称之为判别元组。

根据判别元组的定义，在理想状态下，判别元组应满足式(7)给定的关系。在实际网络中，如果不存在攻击行为，则判别元组应满足式(8)给定的关系。如果存在直接式 SynFlood 攻击，则判别元组应满足式(9)给定的关系。如果存在反射式 SynFlood 攻击，则判别元组应满足式(10)给定的关系。因此可以使用判别元组来标识起始时刻为 t、持续时间为 T 的时间段内 TCP 的连接状态。

2.3 偏离度

假定在检测攻击时，所有检测周期持续时间 T 等长，那么基于欧氏空间，可以将 SDF_T，RDF_T 及检测周期起始时刻 t 分别映射为 x 轴、y 轴和 z 轴。这样，根据式(7)，在欧氏空间中可用直线 $\begin{cases} x=1 \\ y=0 \end{cases}$ 标识理想状态。

根据判别元组的定义，在理想状态下，如果能够恰当地选择检测周期持续时间 T，则所有判别元组对应的点均应分布在欧氏空间的直线 $\begin{cases} x=1 \\ y=0 \end{cases}$ 上。但考虑到网络的实际运行规律，检测周期持续时间 T 实际上很难被“恰当”确定，式(7)难以被严格满足，即样本点不可能严格分布在直线 $\begin{cases} x=1 \\ y=0 \end{cases}$ 上，将存在一定的偏离。为更好地描述此种偏离情况，引入偏离度的概念。

定义 7(偏离度，D_T^t) 设 $P(x_t, y_t, t)$ 为 t 时刻起始持续时间为 T 的时间段所对应的判别元组在欧氏空间中所对应的点，定义点 $P(x_t, y_t, t)$ 与直线 $\begin{cases} x=1 \\ y=0 \end{cases}$ 之间的欧氏距离为该时间段的网络连接状态相对于理想状态的偏离度，由文献[9]可知其计算公式如式(11)：

$$D_T^t=\sqrt{(x_t-1)^2+(y_t-0)^2} \tag{11}$$

由偏离度 D_T^t 定义可知，D_T^t 越小，表示当前网路连接状态越接近理想状态；D_T^t 越大，表示当前网络连接状态越背离

理想状态，甚至当 D_T 超过某一阈值时，可判定存在 SynFlood 攻击。因此可以借助偏离度 D_T^t 判别 SynFlood 攻击的存在性。

2.4 攻击判别值

考虑到实际网络环境的复杂性，Syn，Fin 和 Rst 报文段的分布存在明显的不均匀性。即使网络中不存在任何攻击行为，单个时间段的偏离度检测也难以满足式(8)，可能会存在较大的偏离，从而导致误报。当然，这种不均匀性带来的负面影响可以随着时间段持续时间 T 取值的增加而被削弱。另外，当存在瞬间建立大量 TCP 连接的极端情形时，同样可能导致偏离度超过检测阈值，造成误判。考虑到 SynFlood 攻击的持续性特征，为减少上述不均匀性与不规则变动所引发的误报，可对检测结果及最近的若干次历史检测结果进行移动平均。为此引入攻击判别值的概念。

定义 8(攻击判别值，Δ_T^t) 对某个以 t 时刻起始、持续时间为 T 的时间段而言，该时间段的偏离度 D_T^t 及其最近的 $n-1$个时间段的偏离度可构成一个偏离度序列$\{D_T^t, D_T^{t-1}, \cdots, D_T^{t-n+1}\}$，对该序列进行移动平均，其值定义为该时间段的攻击判别值 Δ_T^t。

由文献[10]可知，攻击判别值 Δ_T^t 的计算公式如式(12)：

$$\Delta_T^t=\omega_0 D_T^t+\omega_1 D_T^{t-1}+\cdots+\omega_{n-1} D_T^{t-n+1} \quad t\geqslant n \tag{12}$$

$$\begin{cases}\omega_0\geqslant 0,\omega_1\geqslant 0,\cdots,\omega_{n-1}\geqslant 0\\ \omega_0+\omega_1+\cdots+\omega_{n-1}=1\end{cases}$$

式中，D_T^{t-i} 为$[t-iT, t-iT+T)$时间段的偏离度，ω_i 为$[t-iT, t-iT+T)$时间段对应的偏离度所占权重，n 为平均项数即移动步长。

令 D_{MAX}为无攻击行为的实际网络中的最大偏离度，根据攻击判别值定义 Δ_T^t 可知：

(1)在理想状态下，攻击判别值 Δ_T^t 取值为 0；

(2)在实际网络中，如果不存在攻击行为，则攻击判别值 Δ_T^t 取值在$(0, D_{MAX}]$范围内变动；

(3)网络中如果存在 SynFlood 攻击，随着不完整 TCP 连接在全部 TCP 连接中所占的比例增大，攻击判别值取值 Δ_T^t 将随之逐渐增大，并随着攻击的持续进行，其取值将超过 D_{MAX}，并趋向于无穷大。

综述所述，可以通过如下两个步骤判别某个时间段是否存在 SynFlood 攻击：

(1)对该时间段计算偏离度；

(2)将该时间段及其最近连续若干个时间段对应的偏离度视为一个序列，进行移动平均，计算攻击判别值。如果该攻击判别值大于事先设定的阈值，则可判定该时间段内存在 SynFlood 攻击。

3 检测算法

3.1 设定

(1)检测周期持续时间为 T；

(2)计算偏离度的时刻用 t 表示，$t=T, 2T, 3T, \cdots$；

(3) 移动步长为 n；

(4) 攻击判别值检测报警阈值为 D_{MAX}。

3.2 算法描述

从时刻 0 开始，对每个以 $t(t=0, T, 2T, \cdots)$时刻为起始、持续时间为 T 的检测周期执行如下操作：

(1) 分别统计$[t, t+T)$时间段内的 Syn 报文段数(N_{SYN}^T)，Fin 报文段数(N_{FIN}^T)和 Rst 报文段数(N_{RST}^T)；

(2) 计算$[t, t+T)$时间段的同步判别因子 SDF_T 和复位判别因子 RDF_T，构建相应的判别元组(SDF_T, RDF_T, t)；

(3) 计算当前检测周期网络连接状态与理想状态的偏离度 D_T^t，即欧氏空间中点 $P(SDF_T, RDF_T, t)$与直线$\begin{cases}x=1\\ y=0\end{cases}$之间的距离；

(4) 计算攻击判别值 Δ_T^t，则有：

① 如果 $\Delta_T^t<D_{MAX}$，可判定$[t, t+T)$时间段内不存在 SynFlood 攻击，$t=t+T$，返回(1)，进入下一个检测周期；

② 如果 $\Delta_T^t\geqslant D_{MAX}$，可判定时$[t, t+T)$间段内存在 SynFlood 攻击，$t=t+T$，返回(1)，进入下一个检测周期。

3.3 算法复杂性分析

3.3.1 空间复杂度分析

算法在分析 TCP 数据流时实际需要存储的数据主要包括：

(1)当前正在分析的报文段；

(2)检测周期内 Syn 报文段数(N_{SYN}^T)；

(3)检测周期内 Fin 报文段数(N_{FIN}^T)；

(4)检测周期内 Rst 报文段数(N_{RST}^T)；

(5)检测周期内的同步判别因子 SDF_T；

(6)检测周期内的复位判别因子 RDF_T；

(7)检测周期内的偏离度 D_T^t；

(8)最近 $n-1$ 个检测周期的偏离度；

(9)检测周期内的攻击判别值 Δ_T^t。

其中 n 作为移动步长是预设常量，不随检测算法运行而改变。因此对于确定的 n 而言，算法最终消耗的存储空间是固定的。综上分析可知，检测算法的空间复杂度为 O(1)。

3.3.2 时间复杂度分析

检测算法涉及的主要计算环节及其耗时分别为：

(1)分析每一个 TCP 报文段，判断该报文段类别，假设对每个报文段进行分析和判别耗时为 a 秒；

(2)根据对每个报文段进行分析和判别的结果，更新 Syn 报文段数目(N_{SYN}^T)、Fin 报文段数目(N_{FIN}^T)和 Rst 报文段数目(N_{RST}^T)，假设该更新过程耗时为 b 秒；

(3)在每个检测周期末，根据该检测周期内的 Syn 报文段数目(N_{SYN}^T)、Fin 报文段数目(N_{FIN}^T)和 Rst 报文段数目(N_{RST}^T)计算该检测周期内的同步判别因子 SDF_T 和复位判别因子 RDF_T，假设该计算过程耗时为 c 秒；

(4)在每个检测周期末，根据该检测周期内的同步判别因子 SDF_T 和复位判别因子 RDF_T，计算该检测周期内的偏离度 D_T^t，假设该计算过程耗时为 d 秒；

(5)在每个检测周期末，根据该检测周期内的偏离度和最近 $n-1$ 个检测周期的偏离度 D_T^t 计算攻击判别值 Δ_T^t，并与报警阈值 D_{MAX}进行比较，进而判别该检测周期内是否存在 SynFlood 攻击，假设该计算及判别过程耗时为 e 秒。

假设每个检测周期内有 N 个 TCP 报文段，则根据上述分析过程可知，一个检测周期内算法的总耗时为$(a+b)\cdot N+c+d+e$秒，其中 a、b、c、d、e 均为常量。由此可知，算法的时间消耗与需分析的 TCP 流中 TCP 报文段个数 N 呈线性关系，即检测算法的时间复杂度为 O(N)。

4 实验及结果分析

4.1 实验平台

实验采用的平台如表 1 所列。

表 1　实验环境

CPU	Intel Core2 Duo T5600
Memory	2GB DDRII 533
OS	Windows XP Pro SP3

4.2　实验数据集

本文采用 The Internet Traffic Archive 中的 dec-pkt-1 数据集作为实验背景流量，如表 2 所列。

表 2　实验数据集

数据集	采集时间	包数量
dec-pkt-1	1995/03/08 22:00－23:00	330 万

在该数据集上第 560～1160s 时间段内插入反射式 SynFlood 攻击数据流，在第 2180～2780s 时间段内插入直接式 SynFlood 攻击数据流。

4.3　相关参数设定

检测算法中涉及到的参数分别有 T，n，w_i 和 D_{MAX}。

实验中分别考察了 T，n，w_i 和 D_{MAX} 的不同选择对实验结果的影响，具体见 4.5 节。考虑到检测算法期待获得较高检测率、较低的误报率和漏报率，并且由于算法将部署于骨干网的路由器上，为减少检测所带来的额外开销，最后确定 T，n，w_i 和 D_{MAX} 4 个参数的取值，如表 3 所列。

表 3　参数设定

检测周期长度(T)	20s
移动平均步长(n)	5
移动平均权重(w_i)	$\frac{(i+1)^2}{\sum_{0}^{n-1}(i+1)^2}$
攻击判别值阈值(D_{MAX})	1.5

4.4　实验结果及分析

以 20s 为检测周期长度，分别计算最终实验数据流中各检测周期的 Syn，Fin 和 Rst 报文段数目，所得结果如图 1 所示。其中横轴表示时间，纵轴表示各类型报文段数目。

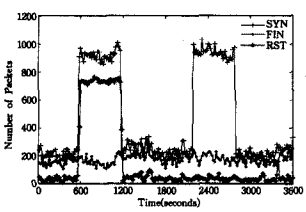


图 1　Syn，Fin 和 Rst 报文段分布情况

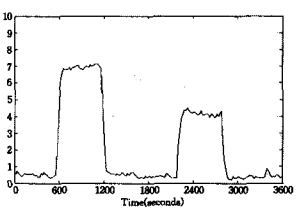


图 2　攻击判别值的变化情况

从图 1 可发现，约 600～1200s 时间段内 Syn 和 Rst 报文段数目急剧增加，约 2200～2800s 时间段内 Syn 报文段数目急剧增加，Syn 报文段、Fin 报文段和 Rst 报文段的数目分布规律明显不满足式(8)所给出的关系。进一步考察该段时间内 Syn 报文段、Fin 报文段和 Rst 报文段的数目分布偏离情况，针对每个检测周期计算偏离度和攻击判别值，其中攻击判别值的计算结果如图 2 所示。其中横轴表示时间，纵轴表示攻击判别值。

从图 2 可以看出，在不同时间段，攻击判别值发生较大变化，具体如表 4 所列。

表 4　不同检测时间段下攻击判别值的变化情况

检测时间段	攻击判别值取值
0～560s	0.3039～0.7473
560～1200s	6.0830～7.1419
1200～2180s	0.2645～1.3625
2180～2800s	3.8640～4.5066
2780～3600s	0.2003～1.4135

从表 4 可知：

(1) 0s≤t<560s 时，攻击判别值 Δ_T^t 稳定在 0.3039～0.7473 之间，小于设定阈值 D_{MAX}，算法判定该段时间内不存在攻击行为；

(2) 560s≤t<1200s 时，攻击判别值 Δ_T^t 迅速上升到 6.0 以上，大于设定阈值 D_{MAX}，算法判定该段时间内存在 SynFlood 攻击；

(3) 1200s≤t<2180s 时，攻击判别值 Δ_T^t 又回到 0.2645～1.3625 之间，小于设定阈值 D_{MAX}，算法判定该段时间内不存在攻击行为；

(4) 2180s≤t<2800s 时，攻击判别值 Δ_T^t 迅速上升到 3.8 以上，大于设定阈值 D_{MAX}，算法判定该段时间内存在 SynFlood 攻击；

(5) 2800s≤t≤3600s 时，攻击判别值 Δ_T^t 又回到 0.2003～1.4135 之间，小于设定阈值 D_{MAX}，算法判定该段时间内不存在攻击行为。

对上述结果进行归纳，如表 5 所列。

表 5　SynFlood 攻击检测结果

总检测周期数	180
应被检测到的周期数	60
实际检测到的周期数	63
实际检测到的有效周期数	60

结合表 4 和表 5 可知，原数据流中存在 SynFlood 攻击的时间段包含在算法检测认定存在 SynFlood 攻击的时间段内，即算法的检测准确率达到了 100%，漏报率为 0%。

另外，1160～1180s，1180～1200s，2780～2800s 这 3 个检测周期并不存在 SynFlood 攻击，但算法判定这 3 个检测周期存在 SynFlood 攻击，与实际情形不相符合。导致误判的原因在于攻击停止后，进行攻击判别值计算时，这 3 个检测周期的攻击判别值仍然受到前几个存在 SynFlood 攻击的检测周期偏离度的影响，即受到移动平均的滞后效应所造成的负面影响。因此算法在本场景中的误报率经计算可为 4.76%。通过上述分析可以认为，算法可以有效地检测直接式 SynFlood 攻击和反射式 SynFlood 攻击。

4.5　参数对检测结果影响的进一步研究

算法中涉及的参数有检测周期长度 T、报警阈值 D_{MAX}、移动步长 n 和移动平均各数据点权重 w_i，本小节将基于 4.2 节所构造的网络数据流进一步研究上述参数对检测结果的影响。

4.5.1　检测周期长度对检测结果的影响

在固定移动步长 n、移动平均数据点权重 w_i 和攻击判别值阈值 D_{MAX} 等 3 个参数的同时，对检测周期长度 T 取不同的值，实验结果如表 6 所列。

表 6　n=5，D_{MAX}=1.5，w_i 取平方平均时 T 的变化

T	应检测周期数	实际检测周期数	误报周期数	漏报周期数	检测率	误报率	漏报率
5s	240	244	4	0	100%	1.64%	0
10s	120	123	3	0	100%	2.44%	0
20s	60	63	3	0	100%	4.76%	0
30s	40	43	3	0	100%	6.98%	0
40s	30	33	4	1	96.67%	12.12%	3.33%
50s	24	27	3	0	100%	11.11%	0
60s	20	24	4	0	100%	16.67%	0

由表 6 可知，在不同检测周期长度下检测率和漏报率几乎不受影响，而误报率随着检测周期长度的增加而增加。究

其原因在于随着检测周期长度的增加，实际检测出的周期数绝对值在减少，而移动平均的滞后效应并没有得到削弱，从而造成误报率增大的现象。若考虑应用到实际的网络环境中，检测周期数趋向于无穷，此时检测周期长度对误报率造成的影响将会被削弱。另外，从表 6 中发现 $T=40s$ 时，存在漏报情形，这是由于攻击结束时间位于一个检测周期的中间位置，即该检测周期内仅存在少量的攻击流量。在进行计算时，这些攻击流量被正常流量湮没，从而导致漏判。总体而言，可以认为检测周期长度的设置对算法结果的影响比较微弱。

4.5.2 *移动步长 n 对检测结果的影响*

在固定检测周期长度 T、移动平均数据点权重 w_i 和攻击判别值阈值 D_{MAX} 等 3 个参数的同时，对移动步长 n 取不同的值，实验结果如表 7 所列。

表 7 $n=5, D_{MAX}=1.5, w_i$ 取平方平均时 n 的变化

n	应检测周期数	实际检测周期数	误报周期数	漏报周期数	检测率	误报率	漏报率
1	60	64	4	0	100%	6.25%	0
2	60	64	4	0	100%	6.25%	0
3	60	63	3	0	100%	4.76%	0
4	60	63	3	0	100%	4.76%	0
5	60	63	3	0	100%	4.76%	0
10	60	67	8	1	98.33%	11.94%	1.67%
20	60	72	15	3	95.0%	20.83%	5.0%
30	60	79	23	4	93.33%	29.11%	6.67%

由表 7 可知，移动步长 n 取值对检测算法的结果影响较大。在一定范围内，检测率随移动步长 n 的增大而减小，而漏报率和误报率则随移动步长 n 的增大而增大。这主要是由于移动步长的增加，会增强移动平均的滞后效应，从而导致漏判和误判均会增加，进而降低检测率，提高误报率和漏报率。当然，移动步长的选择也不是越小越好。从表 7 可以看出，虽然随着移动步长的减小，检测率在增加，漏报率在减少，但当移动步长减少到一定程度时，误报率开始增加。这主要是由于随着移动步长的减小，移动平均对偏离度的平滑作用在削弱，攻击判别值的计算更易受突发正常流量的影响，从而增加误判。因此，选择一个恰当的移动步长有助于提升 SynFlood 攻击检测的效果。

4.5.3 *攻击判别值阈值对检测结果的影响*

在固定检测周期长度 T、移动平均数据点权重 w_i 和移动步长 n 等 3 个参数的同时，对攻击判别值阈值 D_{MAX} 取不同的值，实验结果如表 8 所列。

表 8 $T=20s, n=5, w_i$ 取平方平均时 D_{MAX} 的变化

D_{MAX}	应检测周期数	实际检测周期数	误报周期数	漏报周期数	检测率	误报率	漏报率
≤0.2	60	180	120	0	100%	66.67%	0
0.5	60	101	41	0	100%	40.59%	0
1.0	60	65	5	0	100%	7.69%	0
1.5	60	63	3	0	100%	4.67%	0
4.2	60	41	0	19	68.33%	0	31.67%
4.6	60	30	0	30	50.0%	0	50.0%
6.8	60	22	0	38	36.67%	0	63.33%
≥7.2	60	0	0	60	0	0	100%

由表 8 可知，阈值 D_{MAX} 决定了算法的检测灵敏度。D_{MAX} 取值越小，检测算法对攻击的判别越灵敏，可以有效地提高检测的准确率，但同样可能将正常的突发流量误判为攻击，从而增加误判率。相反，D_{MAX} 取值越大，检测算法对攻击的判别越迟钝，虽然可以有效地减少随机噪音的干扰，降低误报率，但也可能将攻击数据流视为正常数据流而增加漏报率。这个阈值的设置通常受当时的网络流量影响很大，难以找到一个适用于各种网络流量情形的通用阈值。

4.5.4 *移动平均数据点权重设置对检测结果的影响*

在固定检测周期长度 T、攻击判别值阈值 D_{MAX} 和移动步长 n 等 3 个参数的同时，对移动平均数据点权重 w_i 取不同的值，实验结果如表 9 所列。

表 9 $T=20s, D_t=1.5, D_{MAX}=1.5$ 时 w_i 的变化

w_i	应检测周期数	实际检测周期数	误报周期数	漏报周期数	检测率	误报率	漏报率
简单平均	60	65	7	2	96.67%	10.77%	3.33%
线性平均	60	65	5	0	100%	7.69%	0
平方平均	60	63	3	0	100%	4.76%	0

由表 9 可知，当权重取简单平均时，检测结果最差，这是因为简单平均没有考虑各数据点对变化趋势的影响，将当前数据点与历史数据点视为同等重要，而事实上当前数据点比历史数据点对未来的影响更重要，因此有必要增大当前数据点的权重。权重采用线性平均和平方平均时，均提高了当前数据点在数据点序列中的权重。从表 9 可以发现，此时检测效果明显改善，其中权重采用平方平均时检测效果较之采用线性平均时要好，这是因为平方平均比线性平方更能反映当前变化。

结束语 本文基于 TCP 协议中 Syn，Fin 和 Rst 3 种报文段的关系，引入了判别元组、偏离度和攻击判别值，对于不存在攻击的正常数据流，其偏离度应趋于 0。而对于存在 SynFlood 攻击的异常数据流，将存在较大的偏离，因而可以使用偏离度来检测网络数据流中是否存在 SynFlood 攻击。

考虑到 Syn，Fin 和 Rst 3 种报文段的分布不均匀性以及 SynFlood 攻击具有连续性特征，本文采用移动平均方法对连续若干个检测周期的偏离度进行平滑处理，以提高攻击检测的准确性，降低误报率。同时检测算法具有较高的数据处理能力，时间复杂度为 $O(N)$，可部署在大中型网络的骨干路由器上，不会对路由器的数据包转发性能产生大的影响。

必须指出，尽管基于 dec-pkt-1 数据集的实验表明检测算法的准确率达到了 100%，但检测算法中的报警阈值与移动步长对检测结果影响较大，其设定与实际网络流量模型有关。下一步将研究这两个参数如何在一个具体的网络中能够自适应地获得较优的取值。此外，检测算法可以较好地检测出 SynFlood 攻击的存在性，但无法区别直接式 SynFlood 攻击和反射式 SynFlood 攻击。下一步将探讨如何区分上述两种不同的 SynFlood 攻击，以期算法检测结果更细致，具有更强的实用性。

参 考 文 献

[1] 陈波. SYN Flood 攻击的原理、实现与防范[J]. 计算机应用研究，2003，12：80

[2] Jonathan L. Resisting SYN Flooding DoS Attacks with a SYN Cache[C]// Proceedings of USENIX BSDCon on File and Storage Technologies. San Francisco, California, USA: USENIX, 2002：89-98

[3] Bermtan D J, Schenk E. Linux Kernel SYN Cookies Firewall Project[EB/OL]. http://www.bronzesoft.org/projects/scfw

[4] Netscreen 100 Firewall Appliance[EB/OL]. http://www.netscreen.com/

[5] Check Point Software Technologies Ltd. SynDefender[EB/OL]. http://www.checkpoint.com/products/firewall-1

[6] Schuba C L, Krsul Ivan V K, Markus G, et al. Analysis of a Denial of Service Attack on TCP[C]// Proceedings of IEEE Symposium on Security and Privacy. Oakland, CA, USA: 1997 IEEE Symposium on Security and Privacy, 1997: 208-223

[7] Shaikh R A, Iqbal A A, Samad K. Review over Anomaly Detection Algorithms for Detecting SYN Flooding Attacks[C]// Proceedings of Student Conference on Engineering Sciences and Technology. Karachi, Pakistan: CIS Department, NED UET, 2005: 1-5

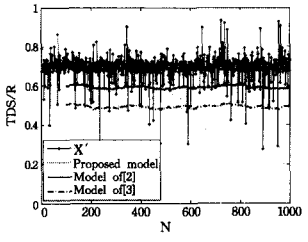
[8] Liu Hui-yu, Chen Kai, Chen Xiao-su. SynFlood Attack Detection Based on Distance Computation in Space Geometry[C]// Proceedings of 2010 International Conference on Computer Application and System Modeling. Taiyuan, Shanxi, China: IACSIT, 2010: V4-585-V4-591

[9] 严绍宗，童裕孙. 实变函数论与泛函分析[M]. 北京：经济科学出版社，1990：39-40

[10] 于善奇. 应用统计技术[M]. 北京：中国标准出版社，2001：267-269

（上接第 76 页）

考虑位于信任链第一层的直接反馈，则模型蜕变为等权值的情况。文献[3]的模型则基于衰减系数为时效期内各个信誉度评价设置权值，其时效期相当于本文的计算窗口 H。在仿真实验中，为便于比较，在本文模型和文献[3]的模型中，统一按照式(7)的 $a=1/H$ 设计衰减系数，并按照 H 设置计算窗口。图 5 表明，尽管样本集 X' 的值在某些时刻存在较明显的波动，但本文的信誉度模型所获得的信誉度计算结果较为稳定。而文献[2,3]的模型所获得的信誉度计算结果则随满意度评价的变化波动明显。



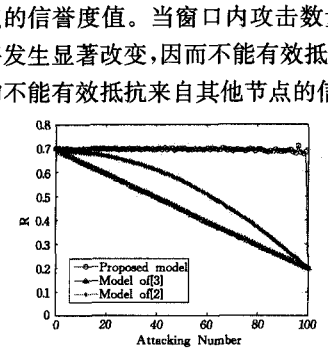
为便于观察，文献[2,3]模型的计算结果分别沿纵轴向下平移了 0.1 和 0.2 个单位

图 5 不同信誉度量化模型的计算结果

4.3 本文模型与其它模型抗攻击效果的对比

对有序样本集 X'，以窗口 H 截取其中的 100 个连续样本，分别采用本文模型、文献[2,3]的模型计算节点的信誉度值。衰减系数和窗口设置同 4.2 节。以 0.2 为攻击值，从左至右依次替换信誉度评价值，对被评价节点的信誉进行诋毁攻击。

图 6 给出了 3 种模型随攻击数增加的信誉度计算结果的变化情况。可以看出，当攻击数为 0 时，3 种模型取得了近似的计算结果。随着攻击数的增加，文献[2,3]的模型计算结果迅速降低。当攻击数达到 100(即攻击值充斥整个计算窗口)时，这两种模型的信誉度计算结果即为攻击值 0.2。文献[3]的模型虽然也基于云理论，但它将计算窗口内云的期望的估计值作为节点的信誉度值。当窗口内攻击数量较多时，云的期望估计值将发生显著改变，因而不能有效抵抗攻击。因此，这两种模型均不能有效抵抗来自其他节点的信誉攻击。



攻击值为 0.2

图 6 不同信誉度量化模型的抗攻击效果

而本文模型则表现出极强的抗攻击性。随着攻击数量的增加，本文模型的计算结果与不存在攻击时相比，几乎不发生明显变化。即使是在攻击数接近 100% 也只发生了轻微的下降。这是由于攻击节点很难改变按照被评价节点整个生命期内的满意度样本集 X 获得的其信誉云数字特征值。依照此数字特征值指导窗口 H 内局部的信誉度计算，偏离信誉云期望的满意度评价将获得趋低倾向的确定度值，进而获得很小的权值。因而，即使攻击数量很多，模型也表现出极强的抗攻击性。只是在攻击数达到 100% 时才失效。因此，与其他模型相比，本文的信誉度计算模型在抗攻击性方面表现出明显的优越性。

结束语 可信网络中节点信誉是影响访问控制、服务授权、交易决策等信任决策的一个关键指标。节点信誉及其聚合过程具有模糊性和随机性特征。采用云理论，能够在广度和深度上科学地描述节点在其生命期内的信誉，实现信誉这一定性概念的合理量化，并揭示节点信誉聚合过程中的模糊性和随机性规律。以节点在其生命期内的信誉云数字特征，指导在局部窗口内其信誉度值的计算，能兼顾信誉的稳定性和随其服务行为以及其它节点评价行为变化的动态适应性。仿真实验验证了利用满意度评价值的确定度和衰减系数设计信誉量化模型权值的有效性。与其它信誉量化模型相比，本文模型计算结果稳定，且表现出较强的抗攻击性。

参 考 文 献

[1] Zhou Rong-fang, Kai H. Power-Trust: a robust and scalable reputation system for trusted Peer-to-Peer computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2007, 18(4): 450-473

[2] 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)：405-416

[3] 王守信，张莉，李鹤松. 一种基于云模型的主观信任评价方法[J]. 软件学报，2010，21(6)：1341-1352

[4] 田春岐，邹仕洪，王文东，等. 一种基于推荐证据的有效抗攻击 P2P 网络信任模型[J]. 计算机学报，2008，31(2)：270-281

[5] 李明楚，杨彬，钟炜，等. 基于反馈机制的网格动态授权新模型[J]. 计算机学报，2009，32(11)：2187-2199

[6] 李德毅，孟海军，史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展，1995，32(6)：15-20

[7] 李德毅，刘常昱，杜鹢，等. 不确定性人工智能[J]. 软件学报，2004，15(11)：1583-1594

[8] 李德毅，刘常昱. 论正态云模型的普适性[J]. 中国工程科学，2004，6(8)：28-34