基于排列熵与决策级多传感器数据融合的P2P僵尸网络检测方法

宋元章
(中国科学院长春光学精密机械与物理研究所 长春 130033)

摘 要 提出了一种基于排列熵和决策级多传感器数据融合的P2P僵尸网络检测算法。首先分别构建流量异常检测传感器和异常原因区分传感器:前者利用排列熵刻画网络流量的复杂度特征(该特征并不依赖于特定类型的P2P僵尸网络),通过利用Kalman滤波器检测该特征是否存在异常;后者利用TCP流量特征在一定程度上减弱P2P应用等网络应用程序对P2P僵尸网络检测的误差影响。最后利用D-S证据理论对上述传感器的检测结果进行决策级数据融合以获得最终的检测结果。实验表明,提出的方法可有效检测新型P2P僵尸网络。

关键词 P2P僵尸网络,排列熵,多传感器数据融合,Kalman滤波器
中图法分类号 TP393.08 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.7.025

P2P Botnet Detection Based on Permutation Entropy and Multi-sensor Data Fusion on Decision Level

SONG Yuan-zhang
(Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China)

Abstract Aiming at the problems of the existing P2P botnet detection methods, a novel P2P botnet detection algorithm based on the permutation entropy and the multi-sensor data fusion on the decision level was proposed. Firstly, it builds the abnormalities detection sensor and the reasons of abnormalities distinguishing sensor. The former sensor uses the permutation entropy to describe accurately the complexity characteristics of network traffic, which does not vary with the structure of P2P network, the P2P protocol and the attack. And the Kalman filter is used to detect the abnormalities of the complexity characteristics of network traffic. Considering that the traffic flow of Web applications is likely to affect the detection result, the latter sensor utilizes the features of TCP flow to solve the problem. Finally, the final result was obtained by fusing the results of two above sensors with the D-S evidence theory. The experiments show that the algorithm proposed in the paper is able to detect P2P botnet with high accuracy.

Keywords P2P botnet, Permutation entropy, Multi-sensor data fusion, Kalman filter

1 引言

僵尸网络(botnet)是一种恶意主机群,攻击者可以利用二次注入对bot节点的负载进行改变,从而非常便捷、迅速地改变最终要发送攻击的种类,例如分布式拒绝服务攻击(Distributed Denial of Service, DDoS)、网络钓鱼(Phishing)、垃圾邮件攻击(Spamming)等。当前新型P2P僵尸网络采用P2P网络的分散式结构来构建其命令与控制机制(Command and Control, C&C),该结构因为没有控制中心,所以有效避免了单点失效,健壮性和可靠性更强。目前,关于P2P僵尸网络分析和检测方面的研究如下:S. Sarat等人[1]和M. Steggink等人[2]研究分析了Storm的生命周期过程、流量特征和躲避检测的机制等,提出了基于网络流量特征的检测方法,例如大量不可达IP地址、特定长度的IP包等;P. Porras等人[3]针对P2P僵尸网络的会话和交互方式进行了详细分析,在此基础上进行特征提取,并通过模式匹配进行僵尸网络的检测;王志等人[4]提出了一种僵尸网络控制命令发掘方法,通过分析bot程序执行轨迹对二进制代码块的覆盖率特征实现对僵尸网络控制命令空间的发掘,对Zeus、SdBot、AgoBot的执行轨迹进行了代码块覆盖率分析实验,结果表明该方法能够快速、准确发掘僵尸网络的控制命令集合。在僵尸网络协同检测方面,王海龙等人[5]在分析现有僵尸网络检测体系的基础上提出了一种可在信息、特性和决策3个级别进行协同的僵尸网络检测层次协同模型;臧天宁等人[6]基于通用图灵机思想,提出了处理大规模网络安全事件的协同联动模型,其可分析发生在不同时间段、地理位置的安全事件可能存在的潜在隐藏关系。文献[7-9]总结和分析了国内外僵尸网络的工作机制的发展过程,并对僵尸网络的检测、防御等方面进行了详尽细致的分类和分析,最后对当前研究存在的问题进行了讨论和建议。

分析可知,P2P僵尸网络在检测方面有以下问题:

(1)大多数检测方法从P2P僵尸网络的某几个特有的、细致的特征入手进行检测,未对网络流量的宏观特征进行足够深入的分析。当出现某种新型P2P僵尸网络,并且该僵尸

到稿日期:2015-06-02 返修日期:2015-08-31 本文受国家863高技术研究发展计划资助项目(2011AA7031024G),国家自然科学基金资助项目(90204014)资助。
宋元章(1986—),男,硕士,助理研究员,主要研究方向为网络安全、分布式计算,E-mail:songyuanzhang@163.com。

网络的网络结构、协议和攻击类型等与现有 P2P 僵尸网络不同时，将会导致 P2P 僵尸网络检测出现较大的漏报率。

(2)大多数检测方法未考虑 P2P 应用程序对 P2P 僵尸网络检测的影响。因为从网络架构和流量角度来分析 P2P 僵尸网络是一种用于发动恶意攻击的特殊类型的 P2P 网络，或者说它可看作是一种恶意 P2P 应用程序，所以两者的流量特征相似程度较高。如果未考虑 P2P 应用程序对 P2P 僵尸网络检测的影响，将会导致 P2P 僵尸网络检测出现较大的误报率。

(3)大多数检测方法通过数据挖掘、机器学习等方法进行 P2P 僵尸网络检测，这需要大量的历史数据、先验知识，并需要事先对分类器进行训练，检测效率不理想。

基于以上分析，本文提出一种基于排列熵和决策级多传感器数据融合的 P2P 僵尸网络检测方法：

(1)该方法主要关注网络流量的复杂度特征，对网络流量的宏观特性进行一定描述，当出现与现用 P2P 僵尸网络不同的新型僵尸网络时仍能保证一定的检测准确度。

(2)该方法所关注的网络流量的复杂度特征属于网络行为特征，无须对数据包内容进行检测，在 P2P 僵尸网络对其数据包采用加密处理时仍可以对其进行检测。

(3)该方法充分考虑了 P2P 应用对检测 P2P 僵尸网络产生的影响，利用 TCP 流量特征对上述影响进行削弱。

2 P2P 僵尸网络检测方法

2.1 方法概述

对 P2P 僵尸网络的典型代表 Storm 的生命周期过程和流量特征进行分析。经分析可知，在其生命周期过程中有如下流量特征：

(1)UDP 流主要用来构建和维持命令与控制机制，特别是当 bot 节点寻找发现其他 bot 节点、保持在线(keep alive)时会导致 UDP 流大量增加。

(2)感染 bot 程序的主机通过随机连接某些 bot 节点尝试加入僵尸网络，此时会出现较多的连接失败，从而导致 ICMP 流异常。

(3)当攻击者利用僵尸网络进行 Spamming 时，bot 节点会发送大量垃圾邮件，进而导致 SMTP 流异常。

本文针对 Storm 生命周期过程及其流量特征进行了详细分析，提出了一种基于排列熵和决策级多传感器数据融合的 P2P 僵尸网络检测算法，如图 1 所示。

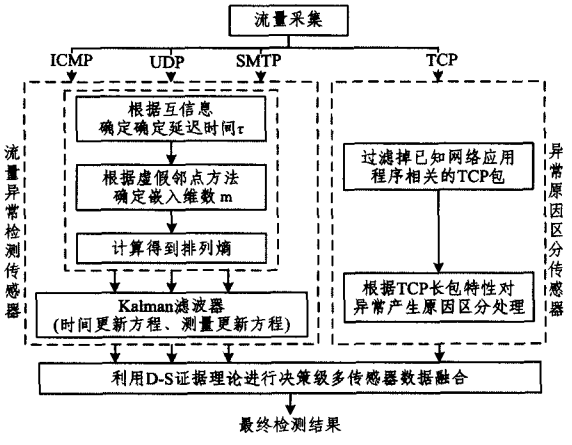

图 1 本文检测方法处理流程示意图

(1)构建流量异常检测传感器和异常原因区分传感器。

1)流量异常检测传感器：该传感器首先利用排列熵对 UDP 流、ICMP 流和 SMTP 流的时间序列复杂度进行刻画，该特征并不依赖于某种 P2P 僵尸网络特有的网络结构、协议和攻击类型等。在 UDP 流、ICMP 流和 SMTP 流的排列熵计算完成后，利用 Kalman 滤波器检测该特征是否存在异常。

2)异常原因区分传感器：该传感器利用 TCP 长包的比例特征对导致流量异常发生的原因进行区分处理，从而减弱 P2P 应用程序产生的网络流量对检测的影响。

(2)利用 D-S 证据理论对上述传感器的结果进行决策级数据融合以得到最终检测结果。

(3)对最终检测结果进行判断。

2.2 流量异常检测传感器

2.2.1 排列熵

排列熵(Permutation Entropy，PE)是一种信号复杂度的度量，已被广泛用于时间序列复杂度及动力学分析，它能够更好地对非线性、非平稳信号进行相关分析[10,11]，放大时间序列的内在微小变化。在反映一维时间序列复杂度的性能方面，与 Lyapunov 指数、分形维数等复杂度参数相比，排列熵具有计算简单快捷、抗噪声能力强等特点[12]，因为计算排列熵时无需使用时间序列的实际数据值，只需要识别时间序列中元素的相对大小关系即可。

已知某单一时间序列$\{x(i),i=1,2,\cdots,n\}$，设其排列熵为 H_p，具体计算方法如下：

(1)为得到该时间序列信号的潜在复杂性信息，需要首先将其重构成更高维的相空间。利用延时嵌入定理[13]对该时间序列进行相空间重构，将其重构为 $N=n-(m-1)\tau$ 个子序列 $X(i)$：

$$X(i)=[x(i)\ \ x(i+\tau)\ \cdots\ x(i+(m-1)\tau)] \tag{1}$$

其中，m 为嵌入维数，τ 为延迟时间。

通过选择合适的延迟时间 τ 和嵌入维数 m，就可使用单一的时间序列还原系统的动力特性[14]。

1)使用互信息法[15]确定延迟时间 τ：选择使得互信息出现局部极小值时所对应的 τ 作为延迟时间；

2)使用虚假邻点法[16]确定嵌入维数 m：选择使得伪近邻点百分比骤降至 0 或接近 0 且不随维数增大而变化时对应的 m 作为嵌入维数。

(2)将 N 个子序列 $X(i)$ 中的元素按照元素值非降的顺序进行排列，如果元素值相等，那么按照元素在原时间序列$\{x(i),i=1,2,\cdots,n\}$中的顺序进行排列。最终可以得到：

$$X'(i)=[x(i+(j_1-1)\tau)\ \ x(i+(j_2-1)\tau)\ \cdots\ x(i+(j_m-1)\tau) \tag{2}$$

其中，$j_1,j_2,\cdots,j_m$ 为 $X(i)$ 中元素的位置，$1\leqslant j\leqslant m$。对于每个子序列 $X(i)$，可以得到如下有序序列 $P(l)$：

$$P(l)=(j_1,j_2,\cdots,j_m) \tag{3}$$

其中，$l=1,2,\cdots,k$，且 $k\leqslant m!$。

(3)计算$\{x(i),i=1,2,\cdots,n\}$的排列熵 H_p。

针对 N 个子序列，设 $P(l)=(j_1,j_2,\cdots,j_m)$ 出现的概率为 P_h，则

$$P_h=\frac{P(l)\text{的个数}}{k} \tag{4}$$

根据 Shannon 熵的定义，时间序列$\{x(i),i=1,2,\cdots,n\}$的 k 种有序序列的排列熵 H_p 为：

$$H_p=\frac{P(l)\text{的熵}}{m\text{个元素全排列的熵}}=\frac{-\sum_{h=1}^{k}(P_h\ln P_h)}{-\sum_{h=1}^{h=m!}(\frac{1}{m!}\ln(\frac{1}{m!}))}$$

$$=\frac{-\sum_{h=1}^{k}(P_h\ln P_h)}{\ln(m!)} \tag{5}$$

经分析可知，$0\leqslant H_p\leqslant 1$。当 $H_p=0$ 时表示时间序列是增序列或减序列，当 $H_p=1$ 时表示系统是完全随机的，即 H_p 越小时间序列越规则，H_p 越大时间序列的复杂性越高。

由第 2.1 节可知，Storm 爆发时会导致网络流量中 UDP 包、ICMP 包和 SMTP 包数目增加，使得它们对应的时间序列表现出一定程度的递增趋势，进而导致对应的排列熵 H_p 减小。因为其变化程度可能不明显，为了进一步提高检测精度，利用 Kalman 滤波器对排列熵的变化进行检测。

此外，为便于利用 D-S 证据理论进行检测结果的数据融合，进行如下假设：设当前时刻为 k，此时时间序列$\{x(i),i=1,2,\cdots,n\}$的排列熵为 H_p。定义排列熵检测变量为：

$$F_k=1-H_p \tag{6}$$

2.2.2 Kalman 滤波器

Kalman 滤波器是一种时间序列预测方法，作为最优估计理论已被广泛用于信号处理等领域。在无法获得模型的确切性质时，Kalman 滤波器仍能够估计信号将来的状态，并且使估计均方差最小，另外它采用递归的形式计算，需要的存储空间小，实时性较好[17-19]。Kalman 滤波器由时间更新方程和测量更新方程组成。本文采用两点差分法对 Kalman 滤波器进行初始化。

(1)时间更新方程

1)向前推算状态变量：

$$X_{k|k-1}=AX_{k-1|k-1}+BU_{k-1} \tag{7}$$

其中，U_{k-1}为 $k-1$ 时刻系统的控制量，A 和 B 为系统参数，$X_{k-1|k-1}$为 $k-1$ 时刻的后验状态估计，$X_{k|k-1}$为 k 时刻的先验状态估计。

2)向前推算误差协方差：

$$P_{k|k-1}=AP_{k-1|k-1}A^{\mathrm{T}}+Q \tag{8}$$

其中，$P_{k-1|k-1}$为对应 $X_{k-1|k-1}$的后验估计误差协方差，Q 是系统过程噪声的协方差，$P_{k|k-1}$为对应 $X_{k|k-1}$的先验估计误差协方差。

(2)测量更新方程

1)结合 $X_{k|k-1}$和测量值 Z_k，计算得到 $X_{k|k}$：

$$X_{k|k}=X_{k|k-1}+Kg_k(Z_k-HX_{k|k-1}) \tag{9}$$

其中，Kg_k 为卡尔曼增益(Kalman Gain)：

$$Kg_k=P_{k|k-1}H^{\mathrm{T}}/(HP_{k|k-1}H^{\mathrm{T}}+R) \tag{10}$$

其中，H 是测量系统的参数，H^{T} 为 H 的转置矩阵，R 为测量噪声协方差。

2)更新后验估计误差协方差 $P_{k|k}$：

$$P_{k|k}=(I-Kg_kH)P_{k|k-1} \tag{11}$$

其中，I 为单位矩阵。

分别将第 2.2.1 节中 UDP 流、ICMP 流和 SMTP 流对应的排列熵检测变量 F_{UDP_k}、F_{ICMP_k}、F_{SMTP_k} 作为系统测量值输入到 Kalman 滤波器中，从而得到对应的后验状态估计，分别表示为 R_{UDP_k}、R_{ICMP_k}、R_{SMTP_k}。

2.3 异常原因区分传感器

P2P 僵尸网络可看作是一种由恶意 P2P 应用程序构建的 P2P 网络，所以 P2P 僵尸网络和 P2P 网络的流量特征相似程度较高。第 2.1 节中提到的流量异常可能是由 P2P 僵尸网络爆发引起的，也可能是由网络环境中某正常 P2P 应用引起的，因此应该对引起上述异常的原因进行相应区分处理。

经分析可知，P2P 僵尸网络主要通过 TCP 包进行二次注入来改变 bot 节点的负载，而 P2P 应用程序主要通过 TCP 进行数据传输，并且长度通常超过 1300 字节，因此可使用 TCP 长包的比例 Pr 来区分引起上述异常的原因。

设 TCP 包数为 N，TCP 长包数为 N_L，处理流程如图 2 所示，其中 DPI(Deep Packet Inspection)详见文献[20]。

定义函数：

$$R_{TCP_k}=\begin{cases}1, & Pr<T_{TCP}\\ 0, & Pr\geqslant T_{TCP}\end{cases} \tag{12}$$

当 $Pr<T_{TCP}$时，第 2.1 节中提到的流量异常是由 P2P 僵尸网络导致的可能性较大，该阈值可通过 Kaufman 算法[21]视不同网络场景进行动态修改。

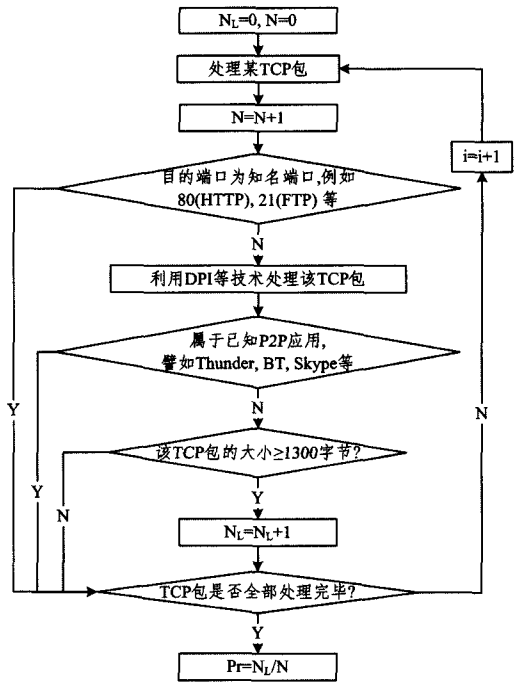

图 2 TCP 流处理流程示意图

2.4 决策级多传感器数据融合

由于僵尸网络具有复杂多变的特点，因此采用决策级多传感器数据融合的方法综合考虑流量异常检测传感器和异常原因区分传感器的结果，从而降低僵尸网络检测的漏报率和误报率。多传感器数据融合又称为多源数据融合，它将各种传感器的观测数据的互补信息和冗余信息利用某种准则进行优化和组合，以获得对被观测环境的一致性描述。根据数据抽象层次不同，融合可以划分为数据级融合、特征级融合和决策级融合。D-S 证据理论是决策级融合的经典方法，该方法通过积累证据缩小假设集，不需要事先得到判决的条件概率和先验概率[22,23]。本文采用 D-S 证据理论对上述传感器结果进行决策级融合：首先各传感器基于自身数据做出局部决策，然后对各局部决策进行融合的产生最终结果。

设随机变量 X 可能取值的论域为 U，称 U 为随机变量 X 的识别框架，如果 U 内所有元素互不相容。在本文中 $U=\{normal,abnormal\}$，$normal$ 为当前网络环境中监控的网络

流量无异常，*abnormal* 为当前网络环境中监控的网络流量存在异常。

设识别框架为 U，2^U 为 U 的幂集，称 $m(A)$ 为 A 的基本概率赋值，如果对于函数 $m:2^U\to[0,1]$ 满足如下条件：

$$(1)\ m(\varnothing)=0 \tag{13}$$

$$(2)\ \sum_{A\in 2^U} m(A)=1 \tag{14}$$

设函数 $m:2^U\to[0,1]$ 为识别框架 U 上的基本概率赋值，称函数 BEL 为 U 上的信任函数，如果函数 $BEL:2^U\to[0,1]$ 满足如下条件：

$$BEL(A)=\sum_{B\subset A} m(B),\ \forall A\subset U \tag{15}$$

称 A 为信任函数 BEL 的焦元，如果 $BEL(A)>0$。

设识别框架 U 上有信任函数 BEL_1 和 BEL_2，与之相应的基本概率赋值为 m_1 和 m_2，与之相应的焦元为 $A_1,\cdots,A_k$ 和 $B_1,\cdots,B_r$，设

$$K=\sum_{A_i\cap B_j=\varnothing} m_1(A_i)m_2(B_j) \tag{16}$$

则称式(16)为 Dempster 组合规则。

$$m(C)=\begin{cases}\dfrac{\sum_{A_i\cap B_j=C} m_1(A_i)m_2(B_j)}{1-K}, & C\neq\varnothing\\ 0, & C=\varnothing\end{cases} \tag{17}$$

当 $K\to 1$ 时，如果对高度冲突的证据直接使用 Dempster 组合规则进行融合，那么会导致出现有悖常理的结果。为了解决该问题，本文采用文献[24]中的方法对冲突证据进行相应处理后再使用 Dempster 组合规则进行融合。

为将上述检测结果 R_{UDP_k}、R_{ICMP_k}、R_{SMTP_k}、R_{TCP_k} 进行决策级数据融合，可采用 Dempster 组合规则对多个证据进行两两组合以得到最终检测结果，因为 Dempster 组合规则满足交换律和结合律。

2.5 具体处理过程

设当前时刻为 k，基于排列熵与决策级多传感器数据融合进行 P2P 僵尸网络检测的具体过程如下。

(1)构建流量异常检测传感器。

1)获得 UDP 流的比例值 C_{UDP}、ICMP 流的比例值 C_{ICMP} 和 SMTP 流的比例值 C_{SMTP}；

2)分别计算 UDP 流、ICMP 流和 SMTP 流的排列熵 H_{UDP_k}、H_{ICMP_k} 和 H_{SMTP_k}，以获得相关时间序列信号的潜在复杂性信息，在此基础上获得对应的排列熵检测变量 F_{UDP_k}、F_{ICMP_k} 和 F_{SMTP_k}；

3)分别将上述排列熵检测变量作为系统测量值输入到 Kalman 滤波器中，建立 Kalman 滤波器模型以检测 UDP 流、ICMP 流和 SMTP 流的异常，分别得到检测结果 R_{UDP_k}、R_{ICMP_k}、R_{SMTP_k}。

(2)构建异常原因区分传感器。

计算得到 TCP 流的 R_{TCP_k} 值(即为该传感器的输出结果)，以在一定程度上消除网络应用对 P2P 僵尸网络检测的误差影响。

(3)利用 D-S 证据理论进行决策级多传感器数据融合，得到最终检测结果 R_k。

(4)最终检测结果判断。

设判断 P2P 僵尸网络是否爆发的阈值为 T，当 $R_k\geqslant T$ 时表示 P2P 僵尸网络爆发，该阈值可通过 Kaufman 算法[21]视不同网络场景进行动态修改。

3 实验数据与分析

3.1 实验环境

实验数据分为正常网络数据和 P2P 僵尸网络流量数据。前者来自某研究所网络服务器，后者来自采用虚拟机技术参照文献[25]搭建的实验环境。

为了构建大规模网络环境，利用虚拟机工具在物理主机上安装多个虚拟机，选取某些虚拟机安装网络封包分析工具并配置为路由设备，这种虚拟机主要用来进行流量采集和分析。目前已知的 P2P 僵尸网络检测相关研究大多以 Storm 作为实验对象，因此在本文实验中亦采用 Storm 对本文方法进行测试。

3.2 网络流量实验

本实验主要观测网络流量的变化情况，如图 3 所示。当 bot 主机开始通信时，僵尸网络的 C&C 机制的活跃表现使得 UDP 包数目大幅度增加。同时，bot 节点的下述行为导致数据包数目出现了一定的波动：为获取攻击者发送的命令，bot 节点周期性地查询相关 key；为保持在线，bot 节点周期性地与其他 bot 节点联系，因此导致数据包数目出现一定的波动。ICMP 包数目从 70 增加到 800，主要因为 bot 主机在 bootstrap 过程中需要随机连接其他节点，此时会发生较高概率的连接失败。由于垃圾邮件攻击的延迟，该实验仅观测到较少的 SMTP 流，下面的实验暂不考虑 SMTP 流。

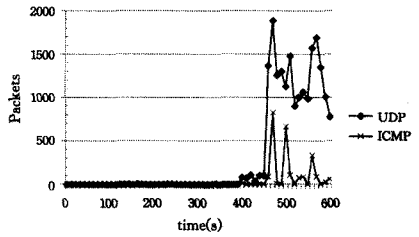

图 3 网络流量数据的变化情况

3.3 排列熵实验

本实验主要观测体现网络流量时间序列复杂度特征的排列熵的变化，如图 4 所示。计算排列熵时选择延迟时间 $\tau=2$，嵌入维数 $m=5$。正常网络场景中的排列熵比较大，存在小幅度的波动。注入 Storm bot 后，UDP 流和 ICMP 流持续增加，对应的时间序列表现出递增特征，导致表征复杂度的排列熵减小。出现一定波动的原因与 3.2 节中的原因相同。

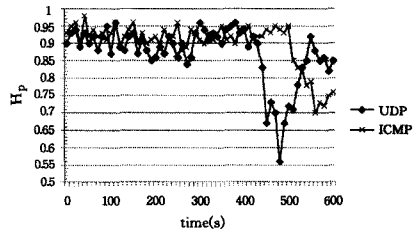

图 4 排列熵的变化情况

3.4 漏报率和误报率实验

本实验主要测试本文方法在不同情况下的检测精度，并将其与已有方法进行比较。本实验涉及到的方法如表 1 所列。在目前已知的僵尸网络检测相关文献中，实验环境不尽相同，不失一般性，本实验选择了如下 4 种数据，实验结果如表 2 所列。表 2 中第 1 组和第 2 组实验数据来自某研究所网

络服务器，第 3 组和第 4 组数据分别是由第 1 组和第 2 组实验数据与采集的 bot 流量数据经过文献[26]所述方法合并后得到的。

表 1　实验涉及检测方法概述

序号	方法名称	方法概述
1	Multi-chart CUSUM[27]	利用 Multi-chart CUSUM 算法检测各种类型数据包数量的异常检测 P2P 僵尸网络
2	本文方法	利用排列熵刻画网络流量的复杂度特征，同时充分考虑了 P2P 应用对检测 P2P 僵尸网络产生的影响，最后利用 D-S 证据理论对检测结果进行有效融合

表 2　漏报率和误报率比较

实验数据	数据说明	真实情况	检测结果	
			Multi-chart CUSUM	本文方法
第 1 组	正常流量	0	23	4
第 2 组	正常流量＋P2P 流量	0	121	10
第 3 组	正常流量＋bot 流量	1000	762	875
第 4 组	正常流量＋P2P 流量＋bot 流量	1000	1268(784)	1023(862)

注："1023(862)"表示本文方法在第 4 组数据中检测到了 1023 次攻击，其中 862 次是真正的攻击。

第 1 组数据中本文方法检测结果与真实情况比较接近。通过比较第 1 组和第 2 组数据，可分析得到对异常原因进行相应区分处理的必要性。第 2 组数据是在正常网络环境（其他条件与第 1 组数据相同）中加入了大量 P2P 应用程序流量，因为本文方法构建异常原因区分传感器，使用 TCP 长包的比例特征对导致流量异常发生的原因进行区分处理，从而减弱了 P2P 应用程序产生的网络流量对检测的影响，所以检测的误报率较低。第 3 组数据是在没有 P2P 应用的正常网络环境中注入 bot，背景数据与第 1 组数据相同，本文方法的漏报率较低。第 4 组数据是在含有 P2P 应用的正常网络环境中注入 bot 程序，背景数据与第 2 组数据相同，本文方法的检测结果比较理想。利用排列熵度量网络流量的复杂度可以较好地表征其内在动力学微小的变化，通过利用 Kalman 滤波器检测排列熵的变化进一步提高检测精度。同时本文方法考虑到了正常 P2P 应用程序对 P2P 僵尸网络检测的影响。

综上所述，本文提出的 P2P 僵尸网络检测方法的检测效率和检测精度较高。

结束语　本文提出了一种基于排列熵和决策级多传感器数据融合的 P2P 僵尸网络检测算法：流量异常检测传感器利用排列熵刻画网络流量的复杂度特征，通过利用 Kalman 滤波器检测该特征是否存在异常；异常原因区分传感器利用 TCP 流量特征在一定程度上减弱 P2P 应用等网络应用程序对 P2P 僵尸网络检测的误差影响，利用 D-S 证据理论对上述两种传感器的检测结果进行决策级数据融合以获得最终检测结果。实验表明，本文提出的方法可有效检测新型 P2P 僵尸网络。下一步工作重点：如何更精确地描述不依赖于特定某种 P2P 僵尸网络的网络流量特征。

参 考 文 献

[1] Sarat S, Terzis A. Measuring the Storm Worm Network: Technical Report 01-10-2007 [R]. HiNRG Johns Hopkins University, 2007

[2] Steggink M, Idziejczak I. Detection of peer-to-peer botnets[D]. University of Amsterdam, Netherlands, 2007

[3] Porras P, Saidi H, Yegneswaran V. A Multi-perspective Analysis of the Storm (Peacomm) Worm: SRI Technical Report 10-01 [R]. Computer Science Laboratory, SRI International, CA, 2007

[4] Wang Zhi, Cai Ya-yun, Liu Lu, et al. Using coverage analysis to extract Botnet command-and-control protocol [J]. Journal on Communications, 2014, 35(1): 156-166(in Chinese)
王志，蔡亚运，刘露，等. 基于覆盖率分析的僵尸网络控制命令发掘方法[J]. 通信学报，2014，35(1)：156-166

[5] Wang Hai-long, Hu Ning, Gong Zheng-hu. Bot_CODA: botnet collaborative detection architecture[J]. Journal on Communications, 2009, 30(10A): 15-22(in Chinese)
王海龙，胡宁，龚正虎. Bot_CODA：僵尸网络协同检测体系结构[J]. 通信学报，2009，30(10A)：15-22

[6] Zang Tian-ning, Yun Xiao-chun, Zhang Yong-zheng, et al. A Model of Network Device Coordinative Run[J]. Journal of Computers, 2011, 34(2): 216-228(in Chinese)
臧天宁，云晓春，张永铮，等. 网络设备协同联动模型[J]. 计算机学报，2011，34(2)：216-228

[7] Zhuge Jian-wei, Han Xin-hui, Zhou Yong-lin, et al. Research and Development of Botnets[J]. Journal of Software, 2008, 19(3): 702-715(in Chinese)
诸葛建伟，韩心慧，周勇林，等. 僵尸网络研究[J]. 软件学报，2008，19(3)：702-715

[8] Jiang Jian, Zhuge Jian-wei, Duan Hai-xin, et al. Research on Botnet Mechanisms and Defenses[J]. Journal of Software, 2012, 23(1): 82-96(in Chinese)
江健，诸葛建伟，段海新，等. 僵尸网络机理与防御技术[J]. 软件学报，2012，23(1)：82-96

[9] Karim A, et al. Review: Botnet detection techniques: review, future trends, and issues[J]. Journal of Zhejiang University-Science C (Computers & Electronics), 2014, 15(11): 943-983

[10] Jia Feng, Wu Bing, Xiong Xiao-yan, et al. Early fault diagnosis of bearing based on multi-dimension permutation entropy and SVM [J]. Computer Intergrated Manufacturing Systems, 2014, 20(9): 2275-2282(in Chinese)
贾峰，武兵，熊晓燕，等. 基于多维度排列熵与支持向量机的轴承早期故障诊断方法[J]. 计算机集成制造系统，2014，20(9)：2275-2282

[11] Liu Yong-bin, Long Qian, Feng Zhi-hua, et al. Detection Method for Nonlinear and Non-Stationary Signals[J]. Journal of Vibration and Sock, 2007, 26(12): 131-134(in Chinese)
刘永斌，龙潜，冯志华，等. 一种非平稳、非线性振动信号检测方法的研究[J]. 振动与冲击，2007，26(12)：131-134

[12] Feng Fu-guo, Rao Guo-qiang, Si Ai-wei, et al. Research and application of the arithmetic of PE in testing the sudden change of vibration signal[J]. Journal of Vibration Engineering, 2012, 25(2): 221-224(in Chinese)
冯辅国，饶国强，司爱威，等. 排列熵算法研究及其在振动信号突变检测中的应用[J]. 振动工程学报，2012，25(2)：221-224

[13] Lv Yong, Li You-rong, Xiao Han, et al. Gear fault classification based on weighted phase space reconstruction and sample entropy[J]. Journal of Vibration Engineering, 2009, 22(5): 462-466(in Chinese)
吕勇，李友荣，肖涵，等. 基于加权相空间重构降噪及样本熵的齿轮故障分类[J]. 振动工程学报，2009，22(5)：462-466

[14] Christoph B, Bernd P. Permutation entropy: a natural complexity measure for time series [J]. Physical Review Letters, 2002, 88(17): 174102-1-4

[15] Fraser A M, Swinney H L. Independent coordinates for strange attractors from mutual information [J]. Physical Review A, 1986, 33(2): 1134-1140

[16] Cao L Y. Practical method for determining the minimum embedding dimension of a scalar series [J]. Physical D: Nonlinear Phenomena, 1997, 110(1/2): 43-50

[17] Xu Guo-dong, Song Jia-ning, Li Peng-fei. Pulsar navigation adaptive filtering algorithm based on information quality[J]. Optics and Precision Engineering, 2015, 23(3): 827-837(in Chinese)
徐国栋,宋佳凝,李鹏飞. 基于信息质量的脉冲星导航自适应滤波算法[J]. 光学精密工程, 2015, 23(3): 827-837

[18] Ji Shu-jiao, Zhu Ming, Lei Yan-min, et al. Video stabilization with improved motion vector estimation[J]. Optics and Precision Engineering, 2015, 23(5): 1458-1465(in Chinese)
吉淑娇,朱明,雷艳敏,等. 基于改进运动矢量估计法的视频稳像[J]. 光学精密工程, 2015, 23(5): 1458-1465

[19] Yang Gong-liu, Guo Wei-lin, Yuan Er-kai. Compensation of time delay in ship deformation measured by attitude matching[J]. Optics and Precision Engineering, 2015, 23(5): 1409-1415(in Chinese)
杨功流,郭蔚林,袁二凯. 姿态匹配法测量船体变形角中时间延迟的补偿[J]. 光学精密工程, 2015, 23(5): 1409-1415

[20] Sen S, Spatscheck O, Wang Dong-mei. Accurate, scalable in-network identification of p2p traffic using application signatures[C]// Proceedings of the 13th international conference on World Wide Web. New York, NY, USA: ACM, 2004: 512-521

[21] Kasera S, Pinheiro J, Loader C. Fast and robust signaling overload control[C]// Proceedings of Ninth International Conference on Network Protocols. Riverside, USA: IEEE, 2001: 323-331

[22] Bauer M. Approximation Algorithms and Decision Making in the Dempster-Shafer Theory of Evidence—an empirical study[C]// 12th Conference on Uncertainty in Artificial Intelligence (UAI 96). 1997: 217-237

[23] Yager R, Liu L. Classic Works of the Dempster-Shafer Theory of Belief Functions [M]. Springer-Verlag, Berlin, 2008

[24] Mruphy C K. Combing belief function when evidence conflicts [J]. Decision Support System, 2000, 29(1): 1-9

[25] Steggink M, Idziejczak I. Detection Of Peer-To-Peer Botnets [R/OL]. http://staff. science. uva. nl/~ delaat/sne-2007-2008/p22/report. pdf

[26] Zhaoa D, Traorea I, Sayed B, et al. Botnet detection based on traffic behavior analysis and flow intervals[J]. Computers & Security, 2013, 39(4): 2-16

[27] Kang Jian, Zhang Jun-Yao, Li Qiang, et al. Detecting New P2P Botnet with Multi-chart CUSUM[C]// International Conference on Networks Security, Wireless Communications and Trusted Computing(NSWCTC 2009). Wuhan, China, 2009: 688-691

(上接第 140 页)

效提高网页信息隐藏的安全性。针对即将到来的 HTML 5.0时代,在仅识别小写字母的情况下,如何更好地定义起始有效标签和终止有效标签是未来算法必须进一步考虑的重点。同时,信息隐藏量与添加干扰不可见字符、改变标签属性顺序对数量之间的关系,即安全性、隐藏容量与干扰量三者之间的关系,也是本文算法实用化关注的方向。

参 考 文 献

[1] Singh P, Chaudhary R, Agarwal A. A Novel Approach of Text Steganography based on null spaces[J]. IOSR Journal of Computer Engineering, 2012, 3(4): 11-17

[2] Zhao Q, Lu H. PCA-based web page watermarking[J]. Pattern Recognition, 2007, 40(4): 1334-1341

[3] Garg M. A novel text steganography technique based on Html documents[J]. International Journal of Advanced Science and Technology, 2011, 35: 129-138

[4] Katzenbeisser S, et al. Information hiding techniques for steganography and digital watermarking[M]. Artech House, 2000

[5] Lee I S, Tsai W H. Secret communication through web pages using special space codes in HTML files[J]. International Journal of Applied Science and Engineering, 2008, 6(2): 141-149

[6] Zhang X, Zhao G, Niu P. A novel approach of secret hiding in webpage by bit grouping technology[J]. Journal of Software, 2012, 7(11): 2614-2621

[7] Sun P, Lu H. An efficient web page watermarking scheme[C]// 2nd IEEE International Conference on Computer Science and Information Technology, 2009(ICCSIT 2009). IEEE, 2009: 163-167

[8] Chou Y C, Liao H C. A Webpage Data Hiding Method by Using Tag and CSS Attribute Setting[C]// 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). IEEE, 2014: 122-125

[9] IIIII Yang Y, Yang Y. An efficient webpage information hiding method based on tag attributes[C]// 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). IEEE, 2010, 3: 1181-1184

[10] Huang H, Zhong S, Sun X. An algorithm of webpage information hiding based on attributes permutation[C]// IIHMSP08 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008. IEEE, 2008: 257-260

[11] Huang H, Sun X, Sun G, et al. Detection of hidden information in tags of webpage based on tag-mismatch[C]// Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2007(IIHMSP 2007). IEEE, 2007, 1: 257-260

[12] Huang H, Sun X, Li Z, et al. Detection of hidden information in webpage[C]// Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007). 2007, 4: 317-321

[13] Wang Jian-feng, Huang Liu-sheng, Tian Miao-miao, et al. Web pages hidden information detection based on statistics and the SVM classification [J]. Chinese Computer Systems, 2014, 35(6): 1221-1225(in Chinese)
王剑锋,黄刘生,田苗苗,等. 基于统计和 SVM 分类的网页隐秘信息检测[J]. 小型微型计算机系统, 2014, 35(6): 1221-1225

[14] Channalli S, Jadhav A. Steganography an art of hiding data[J]. International Journal on Computer Science and Engineering, 2009, 1(3): 137-141

[15] Rafat K F, Sher M. Innocuous Communication via HTML Hiding Data in Plain Sight[J]. Arabian Journal for Science and Engineering, 2014, 39(2): 783-798

[16] Tang Guang-ming, Wang Ya-di. Information hiding safety study [J]. Computer Engineering, 2008, 34(16): 183-185(in Chinese)
汤光明,王亚弟. 信息隐藏安全性研究[J]. 计算机工程, 2008, 34(16): 183-185

[17] Huang Hua-jun, Wang Bao-wei, Sun Xing-ming, et al. An Algorithm of Webpage Information Hiding Based on Repeated Importing of the CSS Class Selectors[J]. Computer Research and Development, 2009, 46(z1): 138-142(in Chinese)
黄华军,王保卫,孙星明,等. 基于 CSS 类选择符重复引入的网页信息隐藏算法[J]. 计算机研究与发展, 2009, 46(z1): 138-142