

基于扫描流量统计的本地网蠕虫检测方法

巩永旺

- (1. 盐城工学院 信息工程学院, 江苏 盐城 224051;
2. 南京邮电大学 计算机学院, 江苏 南京 210003)

摘 要:为了准确检测外网蠕虫对本地网的传播,在研究蠕虫扫描行为模式的基础上,提出一种基于扫描流量统计的本地网蠕虫检测方法,并给出蠕虫检测方法实现的总体思路、关键算法和检测过程。该检测方法分为异常流量检测和扫描包特征匹配检测两个步骤,即首先使用马尔科夫和坎泰利不等式在网络边界检测进入本地网的扫描流量,提取异常流量中的可疑扫描包的特征;然后监控本地网,检测与可疑扫描包特征相匹配的本地网扫描活动,进而判定本地网是否感染外网蠕虫。分析与初步实验证明,该方法能够检测准确检测外网蠕虫对本地网的传播。

关键词:蠕虫检测;扫描流量统计;马尔科夫不等式;坎泰利不等式;扫描包特征

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2011)07-0145-04

Local Network Worm Detection Method Based on Scan Traffic Statistics

GONG Yong-wang

- (1. School of Information Engineering, Yancheng Institute of Technology, Yancheng 224051, China;
2. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract:In order to detect propagation of network worm from exterior network to local network accurately, a worm detection method based on scan traffic statistics was proposed after researching worm scan patterns, and the general ideal, key algorithms and worm detection process of which was discussed. The worm detection method consisted abnormal traffic detection and scan packets profile matching detection, which firstly detected abnormal scan traffic from exterior network to local network by the Markov's and Cantelli's inequalities, abstracting doubtful scan packets profiles from the abnormal traffic, and then through monitoring the local network and detecting the scan activities that matched the doubtful scan packets profiles, warned the propagation of worms from exterior network. The analysis and preliminary experimental results proved that the method can detect network worm from exterior network accurately.

Key words:worm detection; scan traffic statistical; Markov's inequality; Cantelli's inequality; scan packets profile

0 引言

网络蠕虫是一种可以在网络中自我复制传播的智能化、自动化的可执行的攻击程序或代码,它会扫描和探测网络上存在系统漏洞的节点主机,通过局域网或者国际互联网从一个节点传播到另外一个节点^[1]。与传统计算机病毒相比,网络蠕虫具有传播速度快和繁殖能力强的特点,其传播已对网络系统造成了很大的威胁,给人们带来了巨大的财产损失。比如,2001年7月19日爆发的 Code Red 蠕虫,14小时就感染主机超过35.9万台之多,造成26亿美元的损失;2003年1月

25日爆发的 Slammer 蠕虫,在10分钟之内感染了 Internet 上90%的脆弱性主机。蠕虫在传播过程中,为了发现可感染的目标,其传播的第一步通常是按一定的策略对网络实施扫描,致使网络充斥大量的扫描数据包,数据流量表现一定程度的异常^[2],因此,根据网络流量异常可以预警蠕虫的传播。例如,文献[3,4]提出的基于统计分析建立流量动态临界线或基线的蠕虫检测机制;文献[5]提出了使用马尔科夫和坎泰利不等式检测网络蠕虫,但是这些算法只考虑总的扫描流量的相关统计信息,流量检测方法受背景噪音的影响较大,通常只适用于监控大型网络^[6]。

文中将扫描流量检测与扫描包特征检测相结合,针对本地网络,提出了一种基于扫描流量统计的蠕虫检测方法。该方法在网络边界监控进入本地网的扫描流量,提取疑似蠕虫扫描包的特征,基于这些特征检测本地网络,预警本地网蠕虫的传播。分析与实验证明,

收稿日期:2010-12-27;修回日期:2011-03-09

基金项目:国家自然科学基金(60874091);江苏省"六大人才高峰"高层次人才项目(SJ209006)

作者简介:巩永旺(1976-),男,山东曹县人,讲师,博士研究生,研究方向为计算机网络、复杂网络及其信息安全技术。

该方法对于准确检测外网蠕虫对本地网的传播是有效的。

1 基于扫描流量统计的本地网蠕虫检测方法

1.1 相关定义

定义 1 扫描包: 主要指基于 TCP 协议的初始请求包 TCP-SYN 和基于 UDP 协议的扫描数据包以及 ICMP 扫描数据包等^[7]。

定义 2 扫描包基本特征 (B_profile): 用 4 元组 (dstPort, proto, length, data) 表示扫描数据包基本特征, dstPort 为目的端口, proto 为传输协议, length 为数据包负载长度, data 为数据包负载前 m 个字节数据^[8]。

定义 3 连接度: 在一个时间段 t 内, 若主机 i 向 m 个具有不同 IP 地址的主机发送了基本特征相同的扫描数据包 (设基本特征为 a), 则称 m 为主机 i 基于 a 在时间段 t 内的连接度。

定义 4 扫描包扩充特征 (E_profile): 除了具有扫描包 4 个基本特征外, 还具有 srcIp (源 IP 地址)、ci (连接度) 和 dstIpList (目的地址列表) 等 3 个扩充字段。

1.2 思路与总体流程

蠕虫的扫描探测是指被感染主机向网络中其他主机发送扫描包以发现脆弱性主机, 当成功发现和感染一个脆弱性主机后, 新的被感染主机作为一个新的感染源扫描其他主机, 因此, 蠕虫传播的扫描行为具有横向扩散的通信特征^[9,10], 如图 1 所示。

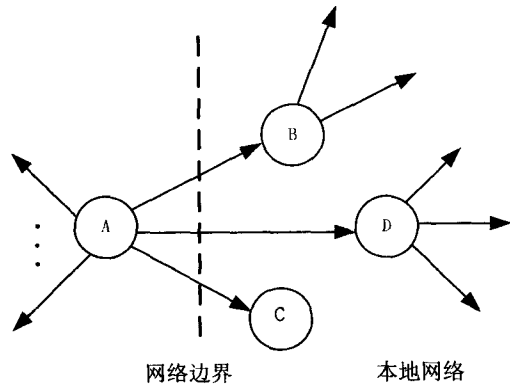


图 1 蠕虫扫描行为模式示意图

基于扫描流量统计的本地网蠕虫检测方法的基本思路为: 当外网有蠕虫向本地网传播时, 在网络边界会检测出异常的扫描流量, 提取类似于 A 节点发出的扫描包基本特征 (见定义 2) 和 A 节点基于这些扫描数据包连接度, 将基本特征与连接度作为在本地网检测蠕虫的特征码。如果在本地网检测出与特征码相匹配的扫描行为 (扫描包特征相同, 连接度大小相当), 比如来自 D 节点的扫描行为与特征码匹配, 则认为形成了一个传播链 $A \rightarrow D$, 判断为本地网感染外网蠕虫。

如果在本地网络中检测不出与特征码相匹配的扫描行为, 即不能形成外网到本地网络的传播链, 认为外网在对本地网络实施一种正常的扫描行为或本地网络主机对该类型蠕虫具有免疫能力, 判断为本地网未感染外网蠕虫。

检测方法总体流程如图 2 所示。首先在网络边界监控进入本地网的扫描流量, 当判断为异常流量时, 提取疑似蠕虫扫描的扫描包的特征 (基本特征和连接度的组合), 简称为可疑特征。蠕虫传播检测将可疑特征和作为蠕虫特征码检测本地网, 检测的步骤为: (1) 监视本地网中的扫描行为, 基于可疑特征中的基本特征进行扫描包的过滤, 只记录与可疑特征中基本特征相匹配的扫描包; (2) 按照一定的策略统计分析记录的扫描包; (3) 基于统计结果判断本地网的蠕虫传播, 定位被感染主机。

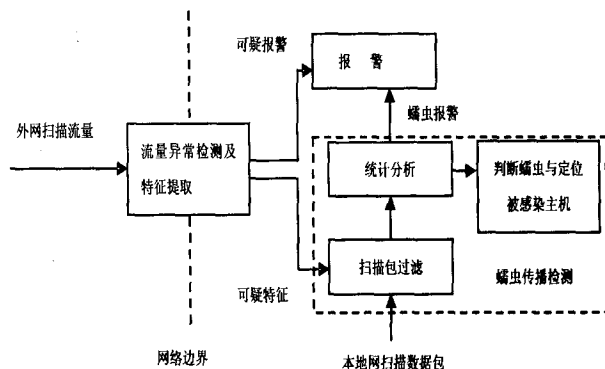


图 2 蠕虫检测总体流程

1.3 流量异常检测及特征提取

异常流量检测一般基于流量总量统计分析或流量比率统计分析的方式检测异常^[3,4,11,12]。为了尽快提取可疑特征, 本方法要求流量异常检测必须具有较低的复杂度, 较小的系统负载。文中借鉴文献[3], 采用基于马尔科夫不等式和坎泰利不等式判断异常扫描流量。在训练阶段计算出进入本地网的扫描流量异常阈值, 流量异常检测只需将实际的扫描流量统计值与阈值比较即可, 方法简单, 计算复杂度低。为减少背景噪音干扰, 在扫描流量统计时, 过滤掉非扫描数据包。

1.3.1 阈值的确定与流量异常判断

假设流量异常的概率为 p , 由马尔科夫不等式和坎泰利不等式, 流量异常阈值 V 的计算如公式(1)。

$$V = \min\left(\frac{E(X)}{p}, \left(\left(\frac{1}{p} - 1\right)D(X)\right)^{1/2} + E(X)\right) \quad (1)$$

$$D(X) \neq 0$$

其中, X 是随机变量, 表示扫描流量; $E(X)$ 、 $D(X)$ 分别表示其数学期望和方差。设 x_1, x_2, \dots, x_n 为正常流量时 X 的样本观测值, $E(X)$ 、 $D(X)$ 可分别用样本均值和样本方差近似表示, 如公式(2)、(3)。

$$E(X) \approx \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \quad (2)$$

$$D(X) \approx \frac{1}{n} \sum_{i=1}^n (x_i - E(x))^2 \quad (3)$$

采样样本值 x_i 是第 i 个采样周期内捕获的扫描流量的累加值,用公式(4)表示。

$$x_i = \int_{(i-1)T}^{iT} C(t) dt \quad (4)$$

其中, $C(t)$ 表示瞬间流量函数, T 为采样周期。

扫描流量异常判断函数表示为公式(5)。

$$F(x) = \begin{cases} 1 & x \geq V \\ 0 & x < V \end{cases} \quad (5)$$

其中, x 表示 T 周期内的实时扫描流量, 1 表示扫描流量异常, 0 表示扫描流量正常。

1.3.2 可疑特征的提取

可疑特征提取与扫描流量采样过程同步进行,一旦判断扫描流量异常,则同时生成可疑特征。

设用集合 A 存储扫描包扩充特征。对于每一个进入本地网的扫描包提取其基本特征 ($B_profile$), 判断该基本特征在集合 A 中有没有与其相匹配的扩充特征 (只匹配扩充特征的基本特征部分), 如果没有, 则对该基本特征增加扩充字段生成扩充特征, 作为一个新元素加入到集合 A ; 如果该基本特征在集合 A 中有匹配的扩充特征元素, 提取该扫描包的源 IP 地址, 判断目的 IP 地址是否在对应扩充特征元素的目的地址列表 $dstIpList$ 中, 如果不在, 则该扩充特征元素连接度字段 ci 加 1, 同时将扫描包的源地址 IP 加入到扩充特征元素的目的地址列表 $dstIpList$ 中; 否则, 丢弃该扫描包基本特征。扫描包扩充特征集合 A 的具体生成算法如图 3 所示。

```

Create_A () {
  for each 新收到扫描数据包 C (其基本特征(B_profile)为 a, 目的地址 IP)
    if ( $\forall v_i \in A) \cap (v_i.B\_profile \text{ not match } a)$ 
      特征 a 添加扩展字段并初始化;
       $a \rightarrow A$ ; //加入集合 A
    else if C 的目的 IP  $\notin v_i.dstIpList$ ;
       $v_i.ci + 1$ ;
      C 的目的 IP 地址加入到  $v_i.dstIpList$ ;
    Else
      丢弃特征 a;
  end for
}

```

图 3 扫描包扩充特征集合生成算法

根据蠕虫扫描的通信特征可知, 连接度大的主机发出的扫描包最有可能是蠕虫扫描包。因此, 当检测单类型蠕虫时, 选取连接度 ci 的扩充特征元素中的字

段组合 ($B_profile+ci$) 作为可疑特征; 当检测多种类型蠕虫时, 选取连接度 ci 最大的前 n ($n > 0$) 个扩充特征元素作为可疑特征。在选取多个可疑特征元素时, 如果存在仅仅是源 IP 地址不同的可疑特征元素, 则只选取连接度较大的一个作为可疑特征元素。

1.4 本地网蠕虫传播检测

本地网蠕虫检测采用类似误用检测的方法, 将可疑特征作为蠕虫特征码对本地网进行检测。在检测出感染蠕虫的情况下, 可定位被感染主机。

本地网蠕虫检测基于以下两点假设:

(1) 如果本地网存在相对于该类蠕虫的脆弱性主机, 则外网的蠕虫扫描一定会感染本地网中一台或多台脆弱性主机。即本地网中一定会检测到与可疑特征相匹配的扫描行为。

(2) 如果在网络边界采取了安全措施或本地网中没有相对于该类蠕虫的脆弱性主机, 即使有蠕虫扫描本地网, 等同于正常扫描行为。因为两者的行为都不会导致本地网蠕虫的传播, 即本地网对该类型蠕虫具有免疫能力。

本地网蠕虫传播检测中, 基于每个可疑特征生成一个基于源 IP 地址扫描包的统计集合, 集合中的元素为二元组 (s, c), 其中 s 表示源 IP 地址, c 表示来自 s 的与该可疑特征中基本特征相匹配的扫描包统计值。假设在网络边界提取了 n ($n \geq 1$) 个可疑特征 v_1, v_2, \dots, v_n , 在检测周期 T 内, 则会生成 n 个相应基于源 IP 地址的统计集合 B_1, B_2, \dots, B_n 。生成 B_i ($1 \leq i \leq n$) 的算法如图 4 所示。

```

 $B_i(1 \leq i \leq n)$  初始化为空
For 每个匹配可疑特征  $v_i$  的扫描数据包 a
  If a 的 srcIp 存在于  $B_i$  中的某个二元组元素中
    将此二元组的统计值加 1
  Else
    新增二元组( $s=srcIp, c=1$ )到  $B_i$ ;
End for

```

图 4 匹配可疑特征的扫描包统计算法

根据统计集合 B_i ($1 \leq i \leq n$) 判断本地网是否感染外网蠕虫, 满足公式(6)判定本地网感染蠕虫, 即本地网络中存在其扫描行为匹配可疑特征的主机 (扫描包特征匹配可疑特征的 $profile$, 连接度大于等于可疑特征中的 ci 值)。

$$\exists (s, c) \in B_i \cap (s, c). c \geq v_i.ci, 1 \leq i \leq n \quad (6)$$

根据统计集合 B_i ($1 \leq i \leq n$), 可以定位被感染主机, 满足公式(7)的源 IP 地址主机为被感染主机。满

足公式(7)的元素个数大致反映当前网络中感染蠕虫的主机数量。

$$\forall (s, c) \in B_i \cap (s, c). c \geq v_i \cdot c_i, 1 \leq i \leq n \quad (7)$$

1.5 算法分析

(1)算法能够降低误警率。当外网对本地网实施正常的网络扫描或类似网络活动时,虽然提取了可疑特征,由于这种扫描活动不具有感染性,因此,在本地网络不能检测出满足公式(6)的二元组 (s, c) ,系统就不会产生蠕虫报警。

(2)算法能够检测一种或多种蠕虫。算法可同时提取一个或多个可疑特征,因此,本地网检测时,可同时检测一种或多种蠕虫。

(3)算法能检测不同扫描率的蠕虫。算法将可疑特征各自的连接度值作为蠕虫检测中判断各自类型蠕虫的阈值,而不是采用统一的阈值,使蠕虫的判断更准确灵活,能够检测出不同扫描率的蠕虫。

另外,算法还可以定位本地网络中被感染的主机,便于进一步采取防护措施。但算法无法检测源自本地网内部的蠕虫传播。

2 模拟实验及分析

为了验证算法的有效性,搭建一个测试环境进行初步实验。用实验中心的计算机网络(200台主机)模拟本地网,与实验中心网络相连的其他网络模拟外网,外网与本地网通过路由器连接。在正常网络流量情况下,捕获一时间段的进入实验中心的扫描流量作为训练集和蠕虫检测实验时的背景流量,设 $p = 0.1, T = 20$ 秒,流量阈值 V 由训练集求得,另设 $m = 10$ 。

实验蠕虫样本为 RedCodeII 蠕虫样本和 Slammer 蠕虫样本。实验开始前,设置本地网络中2台主机为 RedCodeII 蠕虫脆弱性主机,3台主机为 Slammer 蠕虫脆弱性主机。实验时,将实验中心的网络与其他网络断开,用两台主机通过一个交换机与三路由器相连来模拟外网主机,一台主机(10.0.168.1/24)用于产生背景流量,另一台主机(10.0.168.2/24)释放蠕虫样本。实验过程中提取的可疑特征如表1所示。

表1 提取的可疑特征

可疑基本特征	连接度	来源主机
80, TCP, 3818, 4745 5420 2F64 6566 6175	108	10.0.168.2/24
1434, UDP, 376, 0401 0101 0101 0101 0101	15101	10.0.168.2/24

在得到可疑特征后,本地网中很快也检测出与这两种可疑特征相同的扫描数据包,而且这些数据包的源主机基于可疑特征的连接度远大于108和15101。说明本地网已感染 RedCodeII 蠕虫和 Slammer 蠕虫,这些扫描数据包的来源主机即为感染主机。

3 结束语

提出了一种基于扫描流量统计的本地网蠕虫检测方法,该方法将扫描流量异常检测和特征检测相结合,扫描流量异常检测是为了提取可疑的蠕虫扫描包特征,而特征检测是为了判断本地网是否感染外网蠕虫。通过实验与分析证明,该方法在检测外网蠕虫对本地网的传播是有效的。

由于条件的限制,实验时,文中采用的网络对外提供网络服务较少,因此正常情况下扫描本地网的连接较少,故而实验较少受噪音流量干扰。今后的工作包括在规模较大的网络上验证该方法的有效性、参数值的合理选取等。

参考文献:

- [1] 文伟平,卿斯汉,蒋建春,等.网络蠕虫研究与发展[J].软件学报,2004,15(8):1208-1219.
- [2] 杨新宇,史 巍,朱慧君.基于本地网络的蠕虫检测定位算法[J].中国科学 E 辑:信息科学,2008,38(12):2099-2111.
- [3] 王勇超,谢永凯,朱之平.基于统计分析建立流量动态临界线的蠕虫检测机制研究[J].计算机应用研究,2010,27(3):1032-1034.
- [4] 马艳春,肖创柏.基于动态基线分析方法的网络蠕虫检测机制研究[J].华北科技学院学报,2008,5(1):94-97.
- [5] Mowbray M. Network Worm Detection using Markov's and Cantelli's Inequalities[R]. U. K: HP Laboratories, 2009: 555-568.
- [6] 田俊峰,张 驰,刘 涛,等.基于本地网主机传播行为的蠕虫预警新方法[J].通信学报,2007,28(5):80-89.
- [7] 钱 旭,顾 巍,陈凌晖,等.网络蠕虫检测系统的设计与实现[J].现代图书情报技术,2007(1):44-48.
- [8] Akujobi F, Lambadaris I, Kranakis E. An Integrated Approach to Detection of Fast and Slow Scanning Worms[C]//ACM Symposium on Information, Computer and Communications Security (ASIACCS 2009). [s. l.]: [s. n.], 2009: 80-91.
- [9] 辛 毅,方滨兴,贺龙涛,等.基于通信特征分析的蠕虫检测和特征提取方法的研究[J].通信学报,2007,28(12):1-7.
- [10] 张新宇,卿斯汉,李 琦.一种基于本地网络的蠕虫协同检测方法[J].软件学报,2007,18(2):412-421.
- [11] Kim H J, Jung. C. Na, Jong S. Song. Network Traffic Anomaly Detection based on Ratio and Volume Analysis[J]. International Journal of Computer Science and Network Security, 2006, 6(5):190-193.
- [12] Lee S H, Kim H J, Na J C, et al. Abnormal Traffic Detection and Its Implementation[C]// The 7th International Conference On Advanced Communication Technology (ICACT2005). [s. l.]: [s. n.], 2005: 246-250.