

U 盘病毒随机传播模型

左黎明, 刘二根, 汤鹏志

(华东交通大学 基础科学学院, 江西 南昌 330013)

摘 要: 分析了 U 盘病毒传播的特征和已有研究方法的不足, 建立了一个随机传播模型, 最后对模型进行了验证, 通过模型预测结果和实际数据的比较, 实验数据表明, 该模型能够比较好地刻画 U 盘病毒传播过程和发展趋势, 对指导反病毒具有一定的现实意义.

关键词: 恶意代码; U 盘病毒; 传播模型

中图分类号: TP309; TP393

文献标识码: A

文章编号: 1000-7180(2010)03-0106-03

A Stochastic Propagation Model of USB-Flash Storage Virus

ZUO Li-ming, LIU Er-gen, TANG Peng-zhi

(School of Natural Science, East China Jiaotong University, Nanchang 330013, China)

Abstract: A concrete analysis about the characteristic of propagation of USB-Flash storage virus and the shortcoming and weakness of the traditional models is given, a new stochastic propagation model is proposed. At last, the mathematic model is verified. Comparison between the predicted result made by this model and the real propagation data shows that the model gives a better description of USB-Flash storage virus's propagation. The model is of some significance in anti-virus.

Key words: malicious code; USB-Flash storage virus; propagation model

1 引言

计算机已经成为人们日常生活不可缺少的部分. 但病毒、蠕虫、木马等恶意代码的种类数量、产生速度和攻击频率与日俱增, 严重威胁了计算机系统和网络的安全. 目前与此相关的理论研究主要集中在病毒和蠕虫的机理、传播、扫描、免疫等方面^[1-5].

基于传染病毒的模型虽然能在一定程度上描述病毒的大致传播规律, 但是这些模型的一个假设前提是病毒传播是依靠网络迅速的传播进行, 短时间内被感染主机能够通过网络不间断地攻击并感染新的主机, 这是使用微分方程建模的一个必要的前提. 而实际情况是极少数 U 盘病毒能够通过网络传播, 其传播过程主要是依靠移动介质进行, 可以直接绕过防火墙和内外网关, 而且是随机的, 这是传染病模

型所无法解释的^[1-6]. 大部分 U 盘病毒传播的速度要比这些模型预测的慢得多, 并且也只是感染整个网络中特定的一小部分电脑, 并呈现感染地域的随机性. 这一矛盾表明, 现有的理论模型并不能如实地反映实际网络中 U 盘病毒传播的真正规律. 分析产生这一矛盾的原因主要是影响 U 盘病毒传播的关键因素是主机的接触率 (即两台主机依靠 U 盘传递数据的而间接接触的次數), 它是动态的而且随机性较大. 正是这种差距很大的特征, 使得 U 盘病毒的传播过程与一般病毒明显不同.

2 U 盘病毒传播的随机模型

2.1 模型分析与假设

一个公用计算机网络中有感染 U 盘病毒的计算机 (以下简称带毒主机) 和没有感染 U 盘病毒的

收稿日期: 2009-01-20; 修回日期: 2009-03-20

基金项目: 国家自然科学基金项目 (10661007); 江西省自然基金项目 (2007GZS1054); 华东交通大学校立科研基金 (07JC03)

计算机(以下简称健康主机),由于使用者经常使用移动存储设备 U 盘在计算机间拷贝数据,可以假设任何两台计算机之间的通过 U 盘进行数据交换的行为是随机的,当健康主机与带毒主机间通过 U 盘进行数据交换时,由于主机间的操作系统、保护措施以及用户操作习惯存在差异,可以认为健康主机是否被感染也是随机的. 以下建立一个随机数学模型来描述这种随机规律,这对于采取措施控制 U 盘病毒的传播是有益的.

为简化,提出如下的一般化假设.

(假设 A) 主机只包括两类主机:带毒主机和健康主机,不考虑具有免疫功能的主机. 设主机总数为 n ,健康主机数为 s ,带毒主机数为 i ,即 $n = s + i$;

(假设 B) 主机间通过 U 盘进行一次数据拷贝交换的过程称为一次接触,任何两台主机通过 U 盘的接触是相互独立的,具有相同的概率,假设每台主机每天平均通过 U 盘与 m 台主机接触;

(假设 C) 当健康主机与带毒主机进行一次接触时,健康主机被感染的概率为 λ .

2.2 模型的建立

模型的目标是寻找健康主机中每天平均新被感染的主机数与已知参数 n, i, m, λ 的关系,为此只需知道一台健康主机每天被感染的概率,这就只要求出一台健康主机被一台指定的带毒主机接触并感染的概率,这个概率可由一台健康主机被一台指定带毒主机接触的概率乘以在接触时感染的概率得到. 设任意二台主机通过 U 盘接触的概率为 p (即一台健康主机与一台指定带毒主机接触的概率)^[7-8].

由接触的独立性,一台健康主机每天通过 U 盘接触的主机数服从二项分布,根据假设 B,这个分布的均值是 m . 利用二项分布的性质并注意到主机总数是 n (除了本身还有 $n - 1$ 台),有

$$m = (n - 1)p \tag{1}$$

所以

$$p = \frac{m}{n - 1} \tag{2}$$

假设 p_1 是一台健康主机被一台指定带毒主机接触并感染的概率,则

$$p_1 = \lambda p = \frac{\lambda m}{n - 1} \tag{3}$$

设 p_2 是一台健康主机每天被感染的概率,则由对立事件概率的计算方法:

$$p_2 = 1 - (1 - p_1)^i = 1 - (1 - \frac{\lambda m}{n - 1})^i \tag{4}$$

健康主机中被感染的台数也服从二项分布,假

设数学期望为 μ (概率意义下健康主机中每天平均新被感染台数),显然有

$$\mu = sp_2 = (n - i)p_2 \tag{5}$$

而其分布的均方差为

$$\sigma = \sqrt{sp_2(1 - p_2)} = \sqrt{p_2(1 - p_2)(n - i)} \tag{6}$$

下面考虑将结果简化,主要针对式(4)来处理. 因为通常 $n \geq m, n \geq 1$,取式(4)右端展开级数的前两项,即

$$p_2 \approx 1 - (1 - \frac{\lambda mi}{n} + \dots) \approx \frac{\lambda mi}{n} \tag{7}$$

最后得到

$$\mu \approx \frac{\lambda mi(n - i)}{n}, \tag{8}$$

$$\frac{\sigma}{\mu} = \sqrt{\frac{1 - p_2}{(n - i)p_2}} = \sqrt{\frac{n - \lambda mi}{\lambda mi(n - i)}} \tag{9}$$

式(8)给出了健康主机中平均每天新被感染的台数 μ 与 n, i, m, λ 的关系,式(9)就看作对数学期望 μ 的相对误差的度量.

3 实验验证与结果分析

为了排除已有 U 盘病毒和反病毒软件的对实验结果的干扰,实验中用 C 语言实现了一个不带任何破坏功能的全新的 U 盘种植者病毒(拷贝到 U 盘上为 curse.exe 和 autorun.inf,为增强迷惑性,图标为标准文件夹的图标,病毒体在主机上释放后主文件为驱动程序 udisk.sys),该病毒只具有传播复制并发送被感染主机的操作系统版本和主机 IP 到指定邮箱的功能,占用系统资源极少,并在 30 天后自动卸载,而且能够控制在一定网络地址范围内,因此对计算机系统和网络无任何安全威胁. 通过实验室环境下的多次模拟测试,再进行统计分析得到 $\lambda = 0.5$. 为了评估该 U 盘病毒的真实传播效果,我们将学校内三处打印点用于打印的 3 台主机作为最初传染源. 剔除非 windows 平台主机、服务器和无法联网主机,全校共有 934 台有效连网主机(含可以探测到静态 IP 地址的笔记本). 实验从 2008 年 9 月 1 日开始实施,9 月 30 日结束. 表 1 中给出了不同系列 windows 操作系统有效感染台数,表 2 为每天感染台数记录.

表 1 操作系统与有效感染台数的关系表

操作系统类型	感染台数(单位:台)
Windows Vista	0
Windows XP/2003	141
Windows 2000	55
Windows 98/ME	0

表 2 实际每天新增感染台数与时间的关系

时间区间	实际每天新增感染台数向量(排列按时间先后顺序记录,单位:台)
第 1 天到第 10 天	[3,7,11,18,15,23,47,21,15,19]
第 11 天到第 20 天	[8,2,3,0,2,0,0,1,0,0]
第 21 天到第 30 天	[0,0,1,0,0,0,0,0,0,0]

从表 1 来看,被记录下 IP 和操作系统类型的有效感染主机的总数占 21%,由于实际环境的复杂性,有不少被感染主机可能无法被记录,因此这个比例应该比实际感染的比例小.从操作系统类型看,有效感染主机主要是 NT 内核操作系统,但由于 Vista 操作系统采用驱动签名技术,实验病毒驱动不能正常运行,造成了无法感染,这也间接证明了 Vista 操作系统的安全性.从实验结果表 2 来看,在最初的 13 天是高感染期,随后几乎没有新的有效主机被感染.结合实际情况,一台主机通过 U 盘平均一天之内最多与两台其他主机接触.初始 $i=3$,取参数如下:

$$n=934, m=2, \lambda=0.5$$

通过式(8),估算出理论有效感染台数与时间的关系如表 3 所示.

表 3 理论每天新增感染台数与时间的关系

时间区间	理论每天新增感染台数向量(排列按时间先后顺序记录,单位:台)
第 1 天到第 10 天	[3,6,12,23,45,82,141,209,230,145]
第 11 天到第 20 天	[33,1,0,0,0,0,0,0,0,0]
第 21 天到第 30 天	[0,0,0,0,0,0,0,0,0,0]

由表 2 和表 3 可以发现,理论预估与实际情况的感染曲线形态特征是一致的,预估最初的 12 天是高感染期,这也与实际情况 13 天的高感染期是比较接近的.这说明文中给出的模型是符合 U 盘病毒传播实际发展趋势的.但是,理论上预估的值在高峰段比实际值要高出许多,分析其主要原因是实际情况要远比文中的假设复杂,例如有许多有效的被感染主机无法正常发送电子邮件到指定邮箱,造成有不少有效被感染主机无法被记录.实际中并不是所有的主机之间相互接触的概率都是一样的,有些主机之间几乎是根本没机会接触,另外不少主机并非天天能够正常运行或者安装使用了类似于 360safe 的

主动 U 盘病毒保护系统,最后还有网络状态的好坏和其他病毒对实验也有一定的影响.

4 结束语

通过分析 U 盘病毒的传播特征,建立了一个随机传播模型,可以对 U 盘病毒传播过程进行解释.从实际环境的实验结果来看,模型能够在一定程度上反映真实的传播过程.目前基于移动介质传播的病毒发展速度非常快,具有网络蠕虫、木马、广告软件多种特征的 U 盘病毒已经出现,由于复合类型病毒传播途径多样化,导致分析过程越来越复杂,建立一个精致的数学模型描述该病毒传播方式和如何更有效防范病毒^[7-8]将是一个非常具有挑战性的课题.

参考文献:

[1] 郭祥昊,钟义信. 计算机病毒传播的两种模型[J]. 北京邮电大学学报,1999,22(1):92-94.

[2] 文伟平,卿斯汉,蒋建春. 网络蠕虫研究与进展[J]. 软件学报,2004,15(8):1208-1219.

[3] 郝向东,王开云,张春瑞,等. 蠕虫在 P2P 网络中的传播研究[J]. 计算机工程,2007,33(8):139-141.

[4] 王勇,黄国兴,彭道刚. 计算机网络病毒传播模型[J]. 计算机工程与应用,2006,30(1):139-141.

[5] 韩兰胜,洪帆,韩淑霞. 邮件病毒传播的迭代模型[J]. 计算机工程,2007,33(6):1-3.

[6] 陆丽华,罗鹏飞,李星,等. 网络恶意移动代码扩散模型综述[J]. 计算机应用,2003,23(6):17-22.

[7] 于璐,吴振强. 基于免疫的自适应安全模型[J]. 微电子学与计算机,2009,26(1):113-117.

[8] 石艳荣,贺永强. 一种基于关联的 IDS 告警分析模型[J]. 微电子学与计算机,2008,25(12):122-124.

作者简介:

左黎明 男,(1981-),硕士,讲师.研究方向为信息安全、非线性系统.

刘二根 男,(1964-),硕士,教授.研究方向为图论及其应用.

汤鹏志 男,(1961-),硕士,教授.研究方向为信息系统及其安全.