动态演化的互联网病毒建模与分析*

杨春霞，胡丹婷，胡　森
（南京信息工程大学 信息与控制学院，南京 210044）

摘　要：互联网是一个不断生长的无标度网络，在其生长过程中伴随着计算机病毒的传播。基于此，兼顾静态和动态网络上的病毒传播过程，建立了含有拓扑结构演化的计算机病毒传播模型，研究了网络增长速度、网络平均度以及计算机连接度对病毒传播的影响。实验结果表明，网络演化速度越快，病毒传播也越快。同时平均度较高的网络更有助于病毒的传播。此外，病毒在爆发期间主要集中在连接度较大的计算机上。所得结论对控制病毒传播意义重大。

关键词：无标度网络；网络拓扑演化；病毒传播模型
中图分类号：TP393　　**文献标志码**：A　　**文章编号**：1001-3695(2012)07-2678-03
doi:10.3969/j.issn.1001-3695.2012.07.075

Modeling and analyzing virus propagation on dynamic Internet

YANG Chun-xia, HU Dan-ting, HU Sen
(*School of Information & Control, Nanjing University of Information Science & Technology, Nanjing* 210044, *China*)

Abstract: Internet is a growing scale-free network, accompanied by the propagation of computer virus. So its security has been a critical problem. This paper constructed a virus model where viruses spread on a increasingly evolving Internet and analyzed the dependence of the density for infected computers on its topology. All the results obtained here indicate that the larger the rate of network evolvement is, the faster virus spreads. Besides the higher the average degree of dynamic network is, the faster virus spreads. It also find that the higher connectivity the computers have, the higher the probability virus break out. All these findings will help for controlling Internet virus propagation.

Key words: scale-free networks; dynamic network topology; virus propagation model

0　引言

互联网的迅速发展改善了人们的生活，但因网络病毒的层出不穷，互联网的安全已受到极大威胁。为了增强互联网抵御病毒袭击的能力，有必要深刻理解计算机病毒在互联网中的传播特性。

近年来，已有专家注意到网络拓扑的特性在很大程度上影响着网络病毒的传播。如文献[1]在不同网络拓扑上对计算机病毒传播进行了仿真，发现无标度网络相对于小世界网络和随机网络更利于病毒传播。文献[2]从无标度网络的角度建立了含有拓扑特性的病毒传播数学模型，揭示了在无标度网络上不存在传染率的临界值，这与规则和随机网络上得到的结论恰恰相反。文献[3]指出互联网的无标度特性，并利用传统的SIR模型研究互联网上计算机位置以及抗病毒能力对病毒传播的影响，得出接连数较多的计算机在病毒的传播及抑制方面起着关键作用。

尽管有关网络拓扑影响病毒传播的研究已取得重要进展，但这方面的研究仍存在一些不足。大部分模型未能体现互联网规模的动态增长特性。现实中每天都有大量个人计算机、服务器和路由器接入互联网，同时会有部分的计算机与互联网断开，但总的来说互联网呈现出一个不断增长的趋势。为简单起见，本文仅从网络动态增长的角度去研究病毒的传播过程。另外，传统的SIR模型中设定了免疫计算机只能由染病计算机恢复而来[4]，却没有考虑到计算机在未感染病毒之前也可以通过安装补丁或升级杀毒软件的方式直接成为免疫计算机。针对以上不足，本文将建立含网络拓扑演化的病毒传播模型并进行相应仿真，研究网络增长速度、网络拓扑平均度以及计算机连接度对病毒传播的影响，这样的研究对于抵御互联网上计算机病毒的传播来说意义重大。

1　无标度互联网的建立

网络中的节点 v_i 表示互联网中的服务器 $i(i=1,2,\cdots,N)$，节点之间的连边 E_{ij} 表示计算机 i 与 j 之间建立的物理连接。计算机的连接数对应于节点的度。因此，互联网上所有的计算机就构成了一个复杂网络，并且不断演化增长。新加入的用户总是倾向于与连接数多的服务器或路由器相连，以至于互联网的度分布表现出很强的幂律分布特性。尽管随着时间的推移，网络中的节点和边在不断增加，但网络拓扑特性却不会发生很大变化，一致地呈现出无标度性[5,6]。因此，本文采用Barabási等人[7]所提出的BA模型来描述互联网的这一特性，得到了节点总数 $N_0=1\ 000$ 的无标度网络，其度分布如图1所示。在随后的病毒传播模型中，互联网的节点和边将不断增

收稿日期：2011-12-15；**修回日期**：2012-01-31　　**基金项目**：国家自然科学基金资助项目(60874111)；江苏省青蓝工程资助项目；江苏省高校自然科学基金资助项目(07KJD120128)

作者简介：杨春霞(1969-)，女，四川成都人，教授，硕导，博士，主要研究方向为复杂网络与复杂系统(y.cx@163.com)；胡丹婷(1987-)，女，硕士，主要研究方向为复杂网络；胡森(1986-)，男，硕士，主要研究方向为复杂网络.

长,病毒也在不断传播。

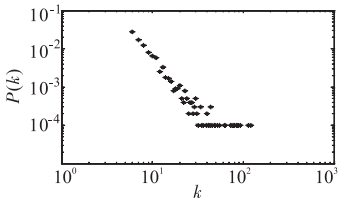

图 1　无标度网络度分布
(N_0 = 1 000, m = 5)

2　病毒传播模型的建立

模型中将计算机分为三类:a)易感染计算机(S),它们不会感染其他计算机,但有可能被感染或进行自我升级成免疫计算机;b)染病计算机(I),它们已染有病毒,具有传播病毒的能力;c)免疫计算机(R),它们已经升过级或查杀病毒而获得了免疫能力,因而不具有传染性,也不会再次被感染。假设单位时间内每个染病计算机独立感染一个易感染个体的概率为 α,并称之为感染率,它与具体的病毒传播机理有关。每个染病计算机升级或是被治愈的概率为 β,并称之为治愈率,它与用户升级计算机及查杀病毒的行为有关。计算机三种状态的转换如图 2 所示。

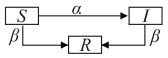

图2　病毒传播过程中计算机三种状态的转换

本文分别用 $s(t)$、$i(t)$、$r(t)$ 表示 t 时刻群体中 S、I、R 三类计算机所占比例,则 $\alpha i(t)s(t)$ 为单位时间内新增的感染计算机比例,$\beta s(t)$ 为单位时间内易感染计算机经过升级系统而变成免疫计算机的比例,$\beta i(t)$ 为单位时间内感染计算机经杀毒后成为免疫计算机的比例。则在改进的 SIR 模型中,计算机病毒传播可用下列微分方程组描述:

$$\begin{cases} \dfrac{ds}{dt} = -\alpha i(t)s(t) - \beta s(t) \\ \dfrac{di}{dt} = \alpha i(t)s(t) - \beta i(t) \\ \dfrac{dr}{dt} = \beta s(t) + \beta i(t) \end{cases}$$

值得注意的是现实中新加入互联网的计算机、路由器、服务器等倾向于与度高的路由器、服务器相连,它们往往都不含有计算机病毒。因此,文中网络的增长演化机制仍然采用 BA 模型中的增长与择优原则,且新加入网络中的计算机均为易感染状态。考虑到病毒的传播伴随着网络的动态演化,也就是说病毒在传播过程中除了感染网络中原有的计算机外,它也会感染新加入的节点,那么,新加入的计算机将是病毒蔓延的新方向。为此,将病毒传播分为两种情况:静态网络传播和动态网络传播。静态网络传播是指染病计算机将病毒传播给网络中已存在的邻居计算机。而动态网络传播是指染病计算机以一定概率与新加入的易感染计算机相连,同时以一定的概率感染新加入的计算机。

初始时,已有的静态无标度网络中有 I_0 个感染计算机,$N - I_0$ 个易感染计算机。单位时间 Δt 内,动态网络演化下的病毒传播过程如下:

a)感染计算机会以概率 α 感染网络中原有的邻居计算机(对应于静态网络传播)。

b)网络中新增 n 个易感染计算机。每个新加入的计算机逐个连接到已存于网络中的 m 个计算机上,且选择连接某个计算机的概率取决于被选计算机的连接度 k,即 $\prod(ki) = k_i / \sum_j k_j$,其中分子 k_i 表示被选择连接的计算机的度,分母为网络中现有计算机连接度之和。同时,感染计算机会以概率 α 感染与它新建连接的计算机(对应于动态网络传播)。

c)易感染计算机和感染计算机会以概率 β 升级为免疫计算机,并且不再被感染。

伴随着网络的演化,病毒将不断传播,图 3 描述了上述网络动态演化中病毒的传播过程。在 t = 0 时,编号为 4 的计算机处于感染状态(黑色圆),其他计算机均处于易感染状态(白色圆)。在 t = 1 时,编号为 1 和 3 的易感染计算机被编号为 4 的邻居计算机感染而变成感染计算机。在 t = 2 时,网络中新增的编号为 9 和 10 的两个易感染计算机分别选择了与度较高的编号为 4、7 的计算机相连,同时编号为 9 的计算机以一定的概率被编号为 4 的计算机感染而变成感染计算机。在 t = 3 时,编号为 4 的感染计算机经过查杀病毒变成了免疫计算机,编号为 10 的易感染计算机经过升级系统也可能变成免疫计算机(灰色圆)。

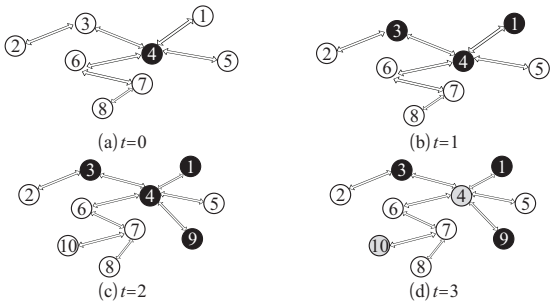

图 3　病毒在动态演化网络上的传播过程

3　实验分析

3.1　感染率α、治愈率β的仿真

由微分方程可知,感染率 α 和治愈率 β 两因素共同影响计算机病毒在演化网络中的传播,图 4 和 5 分别给出了相应的结果。

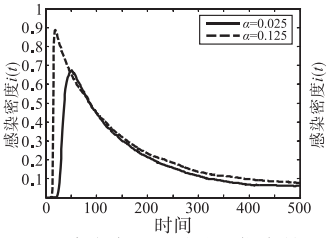

图4 感染率α不同时的仿真结果
(β=0.015,n/t=50,m=50)

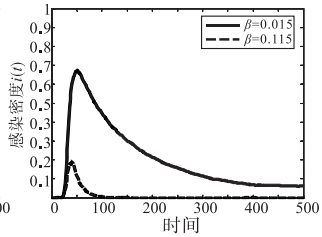

图5 治愈率β不同时的仿真
(α=0.125,n/t=50,m=50)

图 4 给出了相同治愈率 β 而感染率 α 不同时感染密度的变化趋势,其中实线和虚线分别对应感染率 α = 0.025 和 α = 0.125 时的情况。从图中可以看出,当 α = 0.025 时,感染密度上升速度相对较慢,且感染密度所达峰值在 0.65 左右。而当 α = 0.125 时,感染密度上升速度相对较快,且感染密度所达峰值在 0.9 左右。显然,α 越小,感染密度能够达到的最高值也越低,且病毒传播越慢。

图 5 给出了相同感染率 α 而治愈率 β 不同时感染密度的变化趋势,其中实线和虚线分别对应感染率 β = 0.015 和 β = 0.115 时的情况。从图 5 中可以看出,当 β = 0.015 时,感染密度所达峰值在 0.65 左右。当 β = 0.115 时,感染密度所达峰值在 0.2 左右。显然,治愈率 β 越大,感染密度能够达到的最大

值就越小。

上述两个实验表明,感染率 α、治愈率 β 两因素共同影响病毒在演化网络中的传播。如果 α 越大、β 越小,则 $i(t)$ 越大,反之 $i(t)$ 越小。该结论表明,其他条件一定,感染率越大将加速病毒蔓延,相反,治愈率越大将抑制病毒的蔓延。

3.2 网络演化速度 n/t 的仿真

病毒传播除了受感染率 α 和治愈率 β 的影响外,动态演化网络的拓扑变化将怎样影响病毒传播呢?这是一个值得研究的问题。接下来将考察网络在不同演化速度 n/t 情况下病毒的传播特性。

图 6 给出了相同治愈率 β 和感染率 α 而网络演化速度 n/t 不同时的易感染密度 $s(t)$、感染密度 $i(t)$ 以及免疫密度 $r(t)$ 的变化趋势,其中图 6(a)与(b)分别对应 $n/t=10$ 和 $n/t=50$ 两种情况。从图 6(a)可以看出,当单位时间内网络新增 10 台计算机时,感染密度 $i(t)$ 会在 $t=100$ 时到达峰值,与此同时易感染密度 $s(t)$ 趋于最低值,免疫密度 $r(t)$ 开始逐渐上升。从图 6(b)可以看出,当单位时间内网络新增 50 台计算机时,$i(t)$ 会在 $t=50$ 时到达峰值,与此同时易感染密度 $s(t)$ 趋于最低值,免疫密度 $r(t)$ 开始逐渐上升。显然,网络演化速度越快,病毒传播越快,免疫开始时间也相对提前。

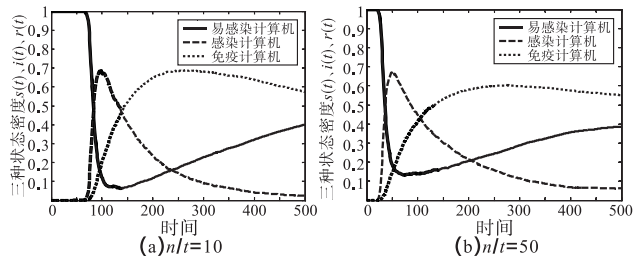

图 6 网络不同演化速度时的病毒感染情况
($\alpha=0.025$, $\beta=0.015$, $m=5$)

3.3 网络不同平均度的影响

网络演化过程中其平均度不断变化,因此考察了平均度不同的动态网络上病毒的传播特性。

在 BA 模型中每次新加入 m 个节点,其网络平均度 $\langle k\rangle=2m$[8]。即设置不同的 m 值可以改变网络平均度的大小。图 7 给出了网络平均度不同时感染密度的变化趋势,其中实线、虚线和点状虚线分别对应 $m=4$、$m=5$、$m=6$ 时的情况,即网络平均度为 8、10 和 12 时的情况。从图中可以看出,当动态网络平均度为 8 时,感染密度在 $t=100$ 时达到 0.6 左右的峰值。当动态网络平均度为 10 时,感染密度在 $t=50$ 时达到 0.65 左右的峰值。当动态网络平均度为 12 时,感染密度在 $t=25$ 时达到 0.7 左右的峰值。显然,在动态网络演化过程中,平均度越大,病毒传播越快,同时计算机的感染密度也越大,也就是说,平均度较高的动态网络更利于病毒传播。

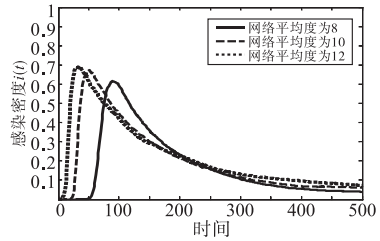

图 7 网络平均度不同情况下的感染情况
($\alpha=0.025$, $\beta=0.015$, $n/t=50$)

3.4 计算机连接度不同的影响

观察连接度不同的计算机的感染密度的变化趋势,如图 8 所示。图中给出了动态网络演化过程中不同连接度的计算机的感染密度情况(这里的感染密度是指在 t 时刻度为 k 的感染计算机数与度为 k 的计算机总数之比[9]),其中实线、虚线、点状虚线分别对应度为 6、10 和 15 的计算机的感染密度。从图中可以看出,网络中度为 6 的计算机感染密度所达峰值在 0.7 附近,度为 10 的计算机感染密度所达峰值在 0.8 左右,度为 15 的计算机感染密度所达峰值在 0.9 附近。显然,在动态网络演化过程中,病毒爆发时,连接度越大的计算机,其感染密度也越大,即动态网络中度较大的计算机更易被感染。

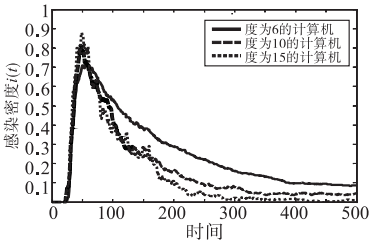

图 8 相同网络环境下连接度不同的计算机的感染情况
($\alpha=0.025$, $\beta=0.015$, $n/t=50$, $m=5$)

4 结束语

如今,人们的生活离不开互联网,然而各类病毒在互联网上肆无忌惮地蔓延给国家和个人带来了巨大的经济损失。因此,进行网络病毒传播的研究将意义重大。针对互联网是一个不断生长的无标度网络,本文兼顾静态和动态网络上的病毒传播过程,建立了含有拓扑结构演化的计算机病毒传播模型,并进行了相应的仿真,得到的有意义的实验结果如下:

a)网络演化速度越快,病毒传播速度也随之加快,同时计算机免疫初始时间也相对提前。

b)网络平均度越大更有助于病毒的传播。

c)动态网络中度较大的计算机更易被感染(如知名网站)。

这些实验结果对于抵制互联网上的病毒传播意义重大。

参考文献:

[1] ZOU C C, TOWSLEY D, GONG Wei-bo. Modeling and simulation study of the propagation and defense of Internet e-mail worms[J]. IEEE Trans on Dependable Secure Computing, 2007, 4(2): 105-118.

[2] PASTOR-SATORRAS R, VESPINGNANI A. Epidemic spreading in scale-free networks[J]. Physical Review Letters, 2001, 86(14): 3200-3203.

[3] 李涛,关治洪,吴正平. 病毒在无标度网络上的传播及控制仿真研究[J]. 计算机应用研究,2011,24(12):177-178,182.

[4] 韩兰胜. 计算机病毒的传播模型及其求源问题研究[D]. 武汉:华中科技大学,2006.

[5] 许丹, 李翔, 汪小帆. 复杂网络理论在互联网病毒传播研究中的应用[J]. 复杂系统与复杂性科学, 2004, 1(3):10-26.

[6] TRAJKOVI L. Analysis of Internet topologies[J]. IEEE Circuits and Systems Magazine, 2010, 10(3):48-54.

[7] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(509):509-512.

[8] 郭雷, 许晓鸣. 复杂网络[M]. 上海:上海科技教育出版社,2006.

[9] 刘俊,金聪,邓清华. 无标度网络环境下 E-mail 病毒的传播模型[J]. 计算机工程,2009,35(21):131-134,137.