

通用有效的动态系统网络病毒传播模型方法研究^{*}

丁雪枫^{1,2}, 马良¹, 丁雪松³

(1. 上海理工大学 管理学院, 上海 200093; 2. 长春理工大学 计算机学院, 长春 130022; 3. 吉林大学 商学院, 长春 130000)

摘要: 研究了依存于网络拓扑的病毒传播特征, 并提出了一个通用的动态系统网络模型。该模型对爆发一种病毒后, 其是很快灭亡还是长期生存, 给出了一个简单易求的传播临界值来进行判断。最后通过对多个网络实际数据集进行计算机仿真, 证明了本模型及传播临界值的有效性和精确性。

关键词: 动态系统网络; 病毒传播模型; SIS 模型; 传播临界值

中图分类号: TP309 **文献标志码:** A **文章编号:** 1001-3695(2009)02-0696-03

Common and effective model of viral propagation with dynamical system network theory

DING Xue-feng^{1,2}, MA Liang¹, DING Xue-song³

(1. Business School, University of Shanghai for Science & Technology, Shanghai 200093, China; 2. School of Computer Science & Technology, Changchun University of Science & Technology, Changchun 130022, China; 3. Business School, Jilin University, Changchun 130000, China)

Abstract: This paper studied the virus propagation characteristics, which depended on the topology of the entire network. And then presented a mathematical model of virus propagation to derive the epidemic threshold value and the relative conditions. Finally, experiment demonstrates the effectiveness and the accuracy of the model.

Key words: dynamical system network; virus propagation model; SIS model; epidemic threshold

0 引言

随着计算机网络的飞速发展, 在计算机网络上传播的病毒也越来越具破坏性和威胁性。有些计算机病毒通过计算机网络的硬件进行传播, 有些是通过电子邮件、即时消息这类含有大量联系人地址簿或联系人列表的形式进行传播。研究表明, 无论哪种方式, 计算机网络的拓扑设计对于病毒传播具有重要的影响。在某些计算机网络拓扑中, 即使是从一个很微不足道的节点出现的病毒, 也有可能迅速在网络中蔓延开来, 造成巨大破坏。目前的反病毒技术大都是看网络中计算机某时刻是否已感染病毒, 针对已感染病毒的节点进行治疗及防御研究, 而不是从网络中的病毒传播能否存活的角度考虑以寻求控制及防御办法。当一种新的计算机网络病毒出现时, 有些网络情况下它会蔓延开来成为一种传播病毒, 而有些网络情况下它会自己逐渐自动消失。因此, 研究不同计算机网络拓扑下病毒能否大规模传播的规律的思想是十分可行的。理论上, 一个节点联系紧密的网络比一个节点松散的网络更容易为病毒的生存及传播提供可能。因此, 相同的病毒可能在一个网络中会很快灭亡; 而在另一个网络中则可能成为传播病毒。通过对不同网络拓扑图的特征研究发现, 病毒在传播过程中存在一个传播临界值, 低于这个值病毒会逐渐灭亡, 高于这个值病毒就会长

期生存并蔓延。当处于临界值之下时, 病毒以对数形式逐渐衰减, 对现实网络不会造成破坏影响。

本文基于依赖于网络拓扑形式的病毒能否存亡的思想, 利用图表示各种计算机网络拓扑, 通过研究病毒在任意网络拓扑中的传播行为特性, 探讨了病毒会最终灭亡以及病毒会流行传播的条件, 进而建立一个适用于任何一种网络拓扑形式的病毒传播模型, 并详细讨论了决定病毒存亡的传播临界值的取值条件。

1 知识背景

一直以来, 对于病毒的传播过程规律, 主要从两个方面进行研究, 即网络图中单节点的行为和不同网络拓扑形式对病毒传播的影响。大多数模型均针对某个特定的网络提出。

1.1 单节点行为的主要模型

目前, 为了研究网络中独立的单节点在病毒传播过程中的行为, 已经提出了很多模型。最重要的几个是:

a) SIS 模型。该模型中, 网络节点表示个体, 边代表个体之间的接触。节点在任意时刻处于两种状态之一: 易感染状态(S)和已感染状态(I)。一个已感染节点被治愈, 但马上就会转为易感染状态; 同时已感染节点还可以把病毒传染给网络中与之相连的邻居易感节点。

收稿日期: 2008-05-06; 修回日期: 2008-07-13 基金项目: 上海市重点学科建设资助项目(T0502)

作者简介: 丁雪枫(1980-), 女, 吉林长春人, 博士研究生, 主要研究方向为系统工程、智能优化、计算机应用(dxsf0310@sina.com); 马良(1964-), 男, 上海人, 教授, 博导, 博士, 主要研究方向为系统工程、智能优化; 丁雪松(1976-), 女, 吉林长春人, MBA 硕士, 主要研究方向为工商管理、金融财务。

b)SIR 模型。该模型中,节点可处于三种状态之一:易感染状态(S)、已感染状态(I)、移除状态(R)。一个已感染节点一旦被治愈,它就对此病毒具有免疫能力,可以从传播网络中移除。

c)基于 SIS、SIR 等模型衍生的扩展模型。大多数模型中,描述病毒传播都使用两个参数:(a)病毒的出生率 β ,表示一个已感染节点试图将病毒传播给与其相连接的易感邻居节点的概率;(b)病毒的治愈率 δ ,表示每个已感染节点可自身治愈的概率。如果 δ 很低,那么病毒传播就会存活较长时间;如果 β 很低,病毒则很快消亡。这说明,病毒能否在网络中进行传播是由某个值决定的,称该值为传播临界值。

1.2 节点关系的主要模型

由于受网络拓扑方式的限制,目前用来研究节点之间关系的模型方法大多需要引入很多假设条件。主要模型有:

a)KW 模型。由 Kephart 和 White 提出的有向图模型,他们给出了此模型下求解已感染节点个数 η_{KW} 的稳态方程以及传播临界值。例如随机图中, η_{KW} 的求解表达式为 $\eta_{KW} = N(1 - \delta/\beta \langle k \rangle)$;传播临界值为 $\tau_{KW} = 1/\langle k \rangle$ 。其中: N 是网络中所有节点的个数; $\langle k \rangle$ 是节点平均度数。该模型仅适用于 ER 随机图、树型图等节点分布遵循幂律的网络。该类网络中,大多数节点都是低连接度;少数具有高连接度的节点极容易被病毒感染并且传染其他节点,同时导致网络中病毒较难被根除。

b)MFA 模型。平均空间假设模型。该模型假设网络中具有相同度的节点都是平等的。该模型对于 BA 网络的已感染节点个数稳态求解式为 $\eta_{MFA} = 2Ne^{-\delta/m\beta}$, m 是网络中的最小度;传播临界值为 $\tau_{KFA} = \langle k \rangle / \langle k^2 \rangle$ 。其中: $\langle k \rangle$ 是网络平均度; $\langle k^2 \rangle$ 是度平方的均值。MFA 模型对于满足幂律分布的网络不具通用性。

c)关联网络模型。该网络模型假设网络中一个节点的关联性与其相邻节点的关联性有关。该模型包含一个度分布函数 $P(k)$, 以及一个用于确定度数为 k 的节点与度数为 k' 的节点关联的概率 $P(k|k')$ 函数。虽然现实网络中存在一些度相关的节点,但是只单纯地用 $P(k|k')$ 函数来表示其关联的关系过于笼统,多数情况下该公式求解的值与现实网络相差甚远。

2 病毒传播模型

2.1 病毒传播特性

在一个包含 N 个节点的网络图中,设每个节点只能处于两种状态之一,即已感染状态和易感状态。 $t+1$ 时的状态只与 t 时的状态有关,这样有 2^N 种可能的传播情形。病毒传播以马尔可夫链的形式进行蔓延。理论上不考虑实际因素,当概率为 1 时,经过一段相当长的时间可以达到所有节点无感染状态。当一种病毒逐渐衰退时,也同样是随时间以指数的形式衰退的。

2.2 建模思想

设任一离散时间步 Δt 的间隔非常小,即 $\Delta t \rightarrow 0$ 。在一个 Δt 内,一个已感染节点 i 试图感染其他邻居节点的概率为 β ,

同时,节点 i 被治愈的概率为 δ 。任一个节点处于两种状态之一,并可以进行快速转换,因此可以把病毒传播网络当做一个动态系统来进行研究。为了简化问题,对于节点 N 的个数很大,病毒以指数形式增长的网络,引入节点状态相互独立假设,即任一给定的节点所处的状态与其他节点所处的状态各自相互独立,将 N 代替 2^N 。实验证明引入该假设并没有对网络研究造成影响。

2.3 数学模型

设一个节点 i 在 t 时刻感染病毒的概率函数为 $p_i(t)$ 。当 i 的所有邻居都没有被感染,或者易感染节点没有成功将病毒传给 i ,用函数 $i_i(t)$ 表示 i 在 $t+1$ 时刻不会被其邻节点感染的概率:

$$i_i(t) = \prod_{j:i \text{ 的邻节点}} (p_j(t-1)(1-\beta) + (1-p_j(t-1))) = \prod_{j:i \text{ 的邻节点}} (1-\beta \times p_j(t-1))$$

假设每个节点的 $p_j(t-1)$ 是相互独立的,则有

$$1 - p_i(t) = (1 - p_i(t-1)) \times i_i(t) + \delta \times p_i(t-1) \times i_i(t) \quad i = 1, \dots, N \quad (1)$$

式(1)将马尔可夫链转换为一个非线性动态系统 (non-linear dynamical system, NLDS) 模型来表示,该等式可以有效且精确地描述网络上的病毒传播。图 1 显示了节点状态转换过程。



图1 节点状态转换过程图

模型中每个单一节点在时间步 t 时,或者处于易感状态(S)或者处于已感染状态(I)。一个易感节点 i 当前是健康的,但是会以 $1 - i_i$ 的概率被邻居节点传染上病毒;一个已感染节点以概率 δ 被治愈后,马上变为易感节点; i_i 的取值由病毒出生率 β 以及 i 节点周围的网络拓扑方式决定。为了确定数学模型的传播临界值求解表达式,首先给出动态系统模型下的传播临界值的定义。

综合 Kephart 等人的结论,非线性动态系统模型下的传播临界值 τ_{NLDS} 定义为

$$\beta/\delta < \tau_{NLDS} \Rightarrow \text{病毒感染随时间衰败}, p_i(t) \rightarrow 0, t \rightarrow \infty \quad \forall i$$

$$\beta/\delta > \tau_{NLDS} \Rightarrow \text{病毒会生存并蔓延}$$

3 传播临界值

动态系统网络可用无向图描述,一个无向图可表述为一个邻接矩阵。通过对现实数据集实验研究发现,非线性动态系统模型的传播临界值取值只与网络图邻接矩阵最大特征值有关。表达式为

$$\tau_{NLDS} = 1/\lambda_{1,A} \quad (2)$$

其中: $\lambda_{1,A}$ 是网络邻接矩阵 A 的最大特征值。要使得经过一段时间后,病毒逐渐趋于灭亡,即图中的每个节点感染概率变为 0,必须满足条件 $\beta/\delta < \tau_{NLDS} = 1/\lambda_{1,A}$;反之亦成立,如果 $\beta/\delta < \tau_{NLDS} = 1/\lambda_{1,A}$,无论最初爆发病毒感染节点的规模大小,病毒传播都会在一段时间后趋于灭亡,即感染概率将趋于 0。

病毒感染以任意一个随机路径以 $(\beta\lambda_{1,A})$ 速度增长,同时还以 δ 速率灭亡。因此,病毒传播的有效速率可以近似表达为 $\beta\lambda_{1,A}/\delta$ 。为了方便,引入一个刻度值 s ,使 $s = \beta/\delta \times \lambda_{1,A}$ 。动态

系统下,当 $s < 1$ 时,病毒最终灭绝; $s \geq 1$ 时,病毒将生存下来。只有 s 取值大于 1 时,病毒传播才会传播蔓延开来。

4 仿真分析

为验证模型的正确性,根据 Barabasi、Albert、密歇根大学等提供的数据,对以下实际数据集进行了仿真实验。

1) 随机网络 一个含有 256 个节点、982 条边、最大特征值为 8 691 的 ER 随机图。

2) 幂律网络 一个含有 3 000 个节点、5 980 条边、最大特征值为 11 541 的 BA 生成图。该生成图幂律度分布指数为 3。

3) 星型网络 该星型网络的中心中转节点与 10 000 星节点连接,最大特征值为 100。

4) 真实网络 一个美国俄勒冈州的自治系统之间相互连接的真实网络图。该网络包含 11 461 个节点、32 730 条边、最大特征值为 75 241。

对于每个数据集,所有节点都初始为已感染状态,然后通过计算机进行 10 000 个时间步,重复 100 次的仿真。

4.1 动态系统网络的精确性

图 2 显示了四个网络中非线性动态系统模型随时间推移, s 取不同值已感染节点数目的变化。

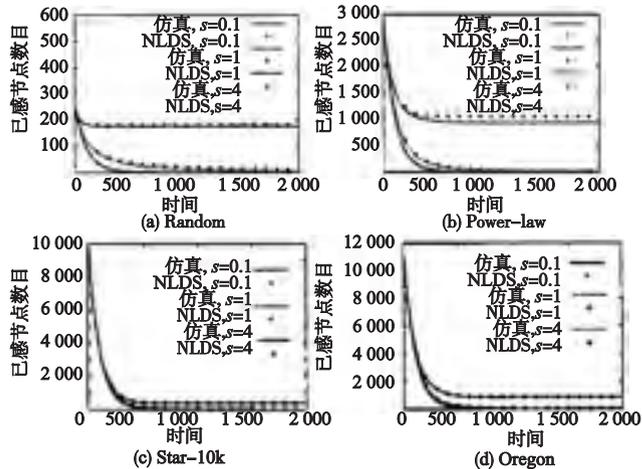


图2 病毒传播动态模型仿真结果图

图 2 表明了非线性动态系统模型仿真结果与实际结果十分接近;动态系统的数学模型可以很好地描述任意网络下的病毒传播行为。

4.2 传播临界值的精确性

图 3 显示了已感染节点个数取不同刻度值 s 情况下随时间变化的情况。

图 3 清晰地显示,当低于临界值($s < 1$)时,病毒感染会逐渐灭亡;高于临界值($s > 1$)时,病毒就会生存下来。传播临界值与由动态系统模型的数学公式求解值结果一致。这说明低于传播临界值时病毒将会逐渐灭亡,高于传播临界值时病毒将会生存并传播的判定条件是正确的。

4.3 低于临界值时病毒的指数衰退

图 4 显示了系统模型在处于临界值之下($s < 1$)时病毒以指数的形式衰退。

如图 4 所示,在四种不同的网络中,在刻度值 $s < 1$ 情况下取不同 s 值,已感染节点个数都是随时间以指数形式减少的。

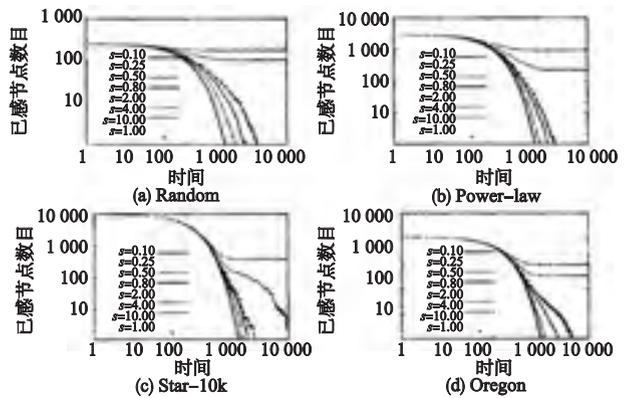


图3 传播临界值精确性仿真结果图

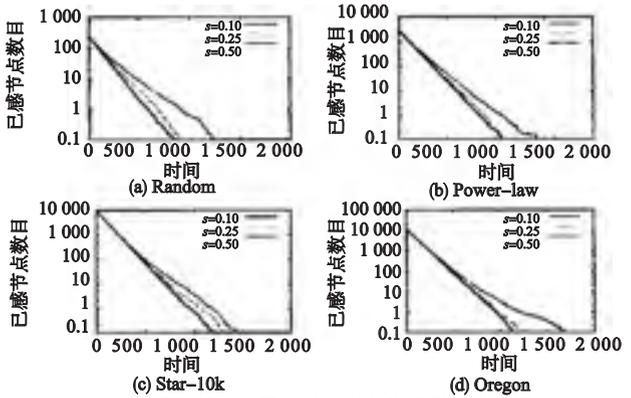


图4 低于临界值时病毒以指数形式衰退图

5 结束语

本文研究病毒传播的规律,提出一个通用的动态系统网络数学模型,确定了模型的传播界定值,通过仿真证明了有效精确性。传播临界值的取值由图的最大特征值 λ_1 决定, λ_1 值越大,图中节点易感性就越高,当满足 $\beta/\delta < 1/\lambda_1$ 时病毒就会彻底灭亡。由于模型中节点只有已感染或者易感染两种状态,不存在免疫节点,一个被治愈的节点会马上变成易感节点,这样就很难控制病毒的传播。未来的一个研究方向是临界值条件下,发现更好的节点免疫方法,最大限度地控制病毒的传播。

参考文献:

- [1] AIELLO W, CHUNG Fan, LU Lin-yuan. A random graph model for massive graphs[J]. STOC,2000,6(1):171-180.
- [2] ALBERT R, JEONG H, BARABASI A-L. Diameter of the World Wide Web[J]. Nature,1999,401(6749):130-131.
- [3] ALBERT R, BARABASI A-L. Topology of evolving networks:local events and universality[J]. Phys Rev Lett,2000,85(24):5234-5237.
- [4] RAVASZ E, BARABASI A-L. Hierarchical organization in complex networks[J]. Phys Rev E,2003,67(2):026112.
- [5] GOLDSTEIN M L, MORRIS S A, YEN G G. Problems with fitting to the power-law distribution[J]. Eur Phys J B,2004,41(2):255-258.
- [6] ZHOU T, WANG B H. Catastrophes in scale-free networks[J]. Chinese Phys Lett,2005,22(5):1072-1075.
- [7] CONDON P. R FC3580, IEEE 802.1x remote authentication dial in user service(RADIUS)usage guidelines[S]. 2003.
- [8] DONGEN M van. Graph clustering by flow simulation[D]. Netherlands: University of Utrecht,2000.
- [9] TANGMUNARUNKIT H,GOVINDAN R, JAMIN S, et al. Network topologies, power laws, and hierarchy[J]. ACM SIGCOMM Computer Communication Review,2002,32(1):76.