BBS用户回复网络的抗毁性分析

吴 敏 李 慧 张 柯 秦丽娟

(首都师范大学教育技术系 北京 100048)

摘 要 结合复杂网络理论,对网络的抗毁性进行了初步的分析。首先基于节点度和介数概念提出了5种攻击策略,并阐述了随机网络、无标度网络以及BBS用户回复网络的构建方法;其次描述了网络抗毁性的定义及其测度;最后采用不同的攻击方法对不同的网络进行攻击,对攻击结果作了详细的分析,结果表明蓄意攻击能在短时间内使网络崩溃,尤其是BBS用户回复网络,抗毁能力远不及随机网络和无标度网络;但进行随机攻击时,BBS用户回复网络体现了良好的容错能力。为了使BBS用户回复网络遭受蓄意攻击时受到的损害程度最低,必须对网络中的重要节点施以保护,使其体现一定的现实意义。

关键词 BBS网络,抗毁性,复杂网络

中图法分类号 TP399 **文献标识码** A

Analysis on Error and Attack Tolerance of Reply Network BBS

WU Min LI Hui ZHANG Ke QIN Li-juan

(Educational Technology Department, Capital Normal University, Beijing 100048, China)

Abstract With the theories of complex network, the paper researched error and attack tolerance of reply network on the Bulletin Board System. Firstly, the authors proposed five methods based on the degree and betweenness of nodes, including failure and attacks, and described the construction model of random network, scale-free network and reply network on BBS. Then, the definition and measure of error and attack tolerance were presented. Finally, the authors attacked networks with five methods. The results show that attacks can breakdown the network in seconds, while failure has little influence on networks, especially the reply network on BBS, which means the reply network owned weaker attack tolerance and stronger error tolerance. Aimed to reduce the damage by attacked, we must protect the important nodes in the network, which have strong practical significance.

Keywords BBS network, Error and attack tolerance, Complex network

1 引言

复杂网络在现实世界中普遍存在,已成功应用于社会、政治、医药、经济、管理等领域,如社会关系网络、科学家合著者网络、细胞神经网络、城市公路交通网络、因特网、万维网等,呈现出广阔的应用前景。作为复杂性科学有力的研究工具,复杂网络已成为一门崭新的交叉科学,为研究复杂系统相关研究提供了全新的思想方法和视角[1],逐渐成为网络时代复杂性科学研究中一个极重要的挑战性课题。

复杂网络本质上的非同质拓扑结构,决定了网络中每个节点的重要程度不同。一旦网络受到蓄意攻击致使某些关键节点或者链路发生故障,将会导致网络的某些功能不能实现,给网络用户带来不便,有时甚至使网络崩溃而失去所有的服务功能[2],而网络的抗毁性就是衡量网络受到蓄意攻击后仍能继续提供一定服务的能力。随着人类社会日益网络化以及人们网络安全意识的提高,复杂网络抗毁性研究的重大理论意义和应用价值已日益凸显,获得了广大的关注,成为复杂网络研究的热点之一。

本文以全局效率作为衡量抗毁性的测度,对随机网络、无标度网络以及真实BBS网络采取了5种攻击策略,并分析了不同网络经过不同攻击后全局效率的变化。

2 复杂网络模型

随机网络、无标度网络是两个经典的网络模型,是复杂网络演化研究的里程碑;BBS用户回复网络因其涉及面广、用户量庞大、信息量丰富,可作为现实复杂网络的代表。随机网络、无标度网络以及BBS用户回复网络的构建为进一步实现抗毁性的测量奠定了基础。

2.1 随机网络模型

随机网络理论由匈牙利数学家Erdős和Rényi提出[3],他们提出的模型称为经典的E-R模型。E-R模型的定义为:在由N个顶点、$C_N^2=N(N-1)/2$条边构成的图中,随机连接g条边形成一随机网络,记为$G_{N,g}$。

另一种与E-R模型等价的随机网络模型是二项式模型,

本文受国家社会科学基金(10CTQ012),北京市属高等学校人才强教计划项目(PHR201108137)资助。

吴 敏(1988-),女,硕士生,主要研究方向为复杂网络,E-mail:wuminmosquito@sina.com;李 慧(1977-),副教授,硕士生导师,主要研究方向为网络与计算智能;张 柯(1979-),硕士生,主要研究方向为复杂网络;秦丽娟(1987-),硕士生,主要研究方向为复杂网络。

其定义如下[4]：给定的节点数目 N 固定不变，假定任意节点对之间有条边连接的概率为 p，形成的网络记为 $G_{N,p}$。这样，整个网络中边的数目是一个随机变量，其期望值为 $\frac{pN(N-1)}{2}$。设 G_0 是节点为 $V_1,V_2,\cdots,V_N$ 且有 g 条边连接组成的一个随机网络，则按上述构造过程，得到 G_0 的概率为 $P(G_0)=p^g(1-p)^{N(N-1)/2-g}$。如果令 $g=pC_N^2$，则两个模型 $G_{N,g}$ 和 $G_{N,p}$ 互相等价，由任一个模型得出的结果可以非常容易地推广到另一模型。

许多学者对随机图进行了非常好的研究，通过严格的数学证明，得到了大量近似和精确的结果[5]，如随机图的节点度服从泊松分布，具有较大的平均路径长度和较小的聚类系数[6]等。

2.2 无标度网络模型

1999 年，Barabási 和 Albert 通过追踪万维网的动态演化过程，发现了许多复杂网络具有大规模的高度自组织特性，即多数复杂网络的节点度服从幂律分布，并把具有幂律度分布的网络称为无尺度网络[4]。Barabási 认为增长和择优选择是无标度网络形成的两个必不可少的生成机制，这一观点已被学术界普遍接受。

BA 模型的增长模式为：初始网络中有 m_0 个节点，在每个时间步上增加一个新节点，并将该节点连接到 $m\leqslant m_0$ 个节点上；择优连接表明新的节点选择连接节点具有偏好性，即它选择节点 i 的概率正比于节点 i 的度。

BA 网络最终演化成为标度不变状态，即当网络规模 $N\to\infty$ 时，幂指数 $\gamma\to3$[4]。BA 模型的平均路径长度很大，聚类系数也很小，但比同规模的随机图的聚类系数要大，不过当网络趋于无穷大时，聚类系数近似为零。

2.3 BBS 用户回复网络的构建

BBS 是电子布告栏系统的英文缩写，已成为网络信息传播的重要途径，为广大用户提供发布信息及进行交互讨论的 Web 应用[7]。通常 BBS 指含有不同版块的整个讨论区，每个版块作为论坛的子区域用于更为细分的话题讨论。论坛中的话题称为帖子，根据帖子发布时间及其回复关系的不同可细分为原帖和回帖。BBS 用户具有唯一的 ID，发表原帖的用户称为发帖者，发表回帖的用户称为回帖者，通过他们之间的回复行为建立用户 ID 之间的连接关系。

本文选取了国内某著名 BBS 论坛中的一个热门版块作为研究对象，此论坛涉及面广、用户量庞大、信息量丰富，具有一定代表性。利用网络爬虫工具抓取了该版块 6 个月的帖子信息，包括帖子标题、发帖者 ID、帖子内容、回帖者 ID 以及回帖时间等。初步统计分析表明：该版块有 125628 个帖子信息、4174 名用户 ID 信息。

为了深入探讨 BBS 用户之间的回复关系，定义有向 BBS 用户回复网络的结构如下：将版块中的用户抽象成有向 BBS 用户回复网络中的节点，用户之间的回复关系对应网络中的有向边，有向边的方向由回帖者指向发帖者。在真实 BBS 论坛中，两个用户之间通常存在多次回复关系，用户也可能对自己的帖子进行回复。为了简化问题，建立更加清晰的网络，本文规定任意两个用户之间最多只存在两条有向边，且不考虑用户回复自己帖子形成的自环。根据上述原则，对挖掘的版块数据进行预处理，删除重复边、自环，构建有向 BBS 用户回复网络[8]。

通过对构建的 BBS 用户回复网络的演化分析，发现网络的非同质性由最初的“无序”状态向“有序”状态靠拢，直到最后的“稳定”状态。网络的幂指数介于 1～3 之间，与同规模的随机网络和无标度网络相比，BBS 网络的聚类系数要大得多，平均路径长度较小。

3 复杂网络抗毁性定义及其测度

与复杂网络抗毁性相关的概念包括节点介数、攻击策略、全局效率等。

3.1 节点介数

网络中不相邻的节点 j 和 k 之间的路径主要依赖于连接节点 j 和 k 的路径上所经过的节点，如果某个节点被其他许多路径经过，则表示该节点在网络中很重要。定量地描述某个节点在网络中的影响力或重要性可以用顶点的介数来衡量，这一定义最早由 Freeman 提出[9]。顶点 i 的介数 B_i 定义为：

$$B_i=\sum_{j,k\in v}\frac{n_{jk}(i)}{n_{jk}} \tag{1}$$

式中，n_{jk} 表示节点 j、k 之间的最短路径的个数，$n_{jk}(i)$ 表示节点 j、k 之间的最短路径中经过节点 i 的个数。研究表明，顶点的最大介数与网络的同步能力密切相关，顶点的最大介数越大，网络的同步能力越弱。

3.2 攻击策略

Holme 等人就攻击策略对复杂网络抗毁性的影响作了较全面的研究，认为一般网络的破坏方式分为去点和去边两种方式，选择方式有随机攻击和选择性攻击两种类型，分别称为网络的容错能力与抗攻击能力。研究选择性攻击时，节点（或边）的移除顺序会很大程度地影响网络结构，故攻击策略按移除顺序的不同可分为以下 4 类[10]：

1）基于原始的度值移除策略（ID）

使用原始网络的信息，按初始网络度降序选择节点，从度值最大的节点开始一个一个移除节点。

2）基于原始的介数移除策略（IB）

使用原始网络的信息，按初始网络介数降序选择节点，从介数最大的节点开始一个一个移除节点。

3）基于重新计算的度值移除策略（RD）

每一步移除一个节点后，重新计算节点的度，选择度值最大（即连接边数最多）的节点作为待删除的节点，以使之对网络的完整性造成最大的破坏。

4）基于重新计算的介数移除策略（RB）

每一步移除一个节点后，重新计算节点的介数，选取介数最大的节点作为待删除的节点，以使之对网络的完整性造成最大的破坏。

3.3 全局效率

一般认为“抗毁性”是指网络在人为破坏作用下的可靠性，它假定破坏者具有关于网络结构的全部资料，并采用一种确定的破坏策略。一般用最大簇大小、孤立簇和平均路径长度来衡量网络遭到攻击后的破坏程度。但一些网络随着破坏程度的加大，最大簇大小逐渐变小，但平均路径长度却是先变大后变小，这种差异性给抗毁性的研究带来了很多不便[11]。

为此，Holme 等用全局效率 E 来衡量攻击后的网络性能[12]，全局效率的定义如下：

$$E=\frac{1}{N(N-1)}\sum_{i\neq j}\frac{1}{t_{ij}} \tag{2}$$

式中，N 为网络节点总数目；t_{ij} 为节点 i 和节点 j 之间的最短路径长度。E 值越小，说明信息传递的速度越慢，如果 E 值为零，意味着信息不能传递，网络处于崩溃状态。

4 攻击效果分析

本文分别构建了 BBS 用户回复网络、无标度网络以及随机网络，并对 3 个网络进行攻击，观察网络全局效率的变化规律，以下从两个角度分析攻击结果。

4.1 不同策略攻击同一网络的结果分析

1)随机网络

对同一随机网络进行 5 种不同的攻击，全局效率随删除节点比例的变化如图 1 所示。由图可知，随机攻击对随机网络的影响明显小于蓄意攻击的影响，4 种蓄意攻击策略的作用效果基本相似。采取任意一种蓄意攻击，随机网络在删除节点的比例达到 0.25 左右时崩溃。

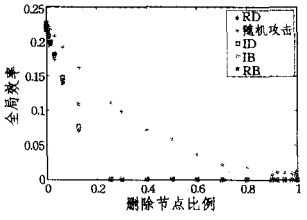

图 1 不同策略攻击随机网络的结果

2)BA 无标度网络

对同一无标度网络进行 5 种不同的攻击，全局效率随删除节点比例的变化如图 2 所示。同理，随机攻击对无标度网络的影响依然较小，4 种蓄意攻击策略的效果曲线基本相似。由图可知其对无标度网络的影响比对随机网络的影响稍微大一些，因其倾斜度相对增大。采取蓄意攻击时，在攻击初期，对网络的影响明显，网络的非同质结构发生巨大的变化，但网络达到崩溃时删除节点的比例与随机网络的接近。

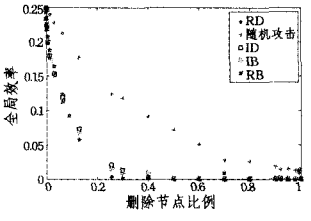

图 2 不同策略攻击无标度网络的结果

3)BBS 用户回复网络

对同一 BBS 用户回复网络进行 5 种不同的攻击，全局效率随删除节点比例的变化如图 3 所示。随机攻击对无标度网络的影响依然很小，4 种蓄意攻击策略的效果曲线基本相似，对 BBS 网络的影响都是巨大的。当删除节点比例在 0.05 左右时，网络就接近崩溃，即只要删除少数的节点，就能完全破坏网络的功能，说明 BBS 用户回复网络的抗攻击能力相当弱。若要维护网络正常运行，使其不受外界的蓄意攻击，可以重点关注节点度大的节点或介数大的节点。

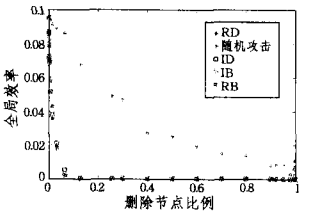

图 3 不同策略攻击 BBS 用户回复网络的结果

4.2 同一策略攻击不同网络的结果分析

比较图 1、图 2 和图 3 的初始全局效率值，可知三者并不相等，BBS 网络的全局效率最小而无标度网络和随机网络的全局效率相当，说明 BBS 网络传播信息的速度最快，也说明无标度网络和随机网络与真实网络存在差距，不能充分地描述真实网络的性质。为了更科学地分析相同攻击策略对不同网络的影响，现将各个网络的攻击结果进行归一化，即将每次的攻击结果比上初始的全局效率值。

1)蓄意攻击策略

对同样规模的 BBS 网络、无标度网络和随机网络进行 4 种蓄意攻击，3 个网络的全局效率随删除节点比例的变化如图 4—图 7 所示。在受到攻击时，3 个网络的全局效率受到不同的影响，其中对 BBS 网络的影响最大，因其曲线的倾斜度最大，且最早达到网络崩溃；无标度网络和随机网络的曲线相似，受到的影响都相对较小，说明在被蓄意攻击时，BBS 网络的抗击能力明显不如无标度网络和随机网络。

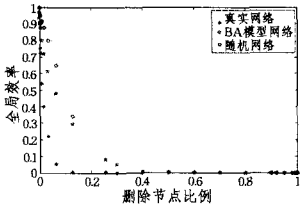

图 4 ID 策略攻击不同网络的结果

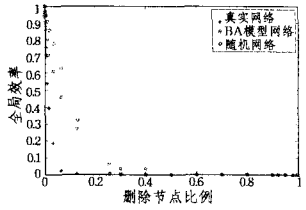

图 5 IB 策略攻击不同网络的结果

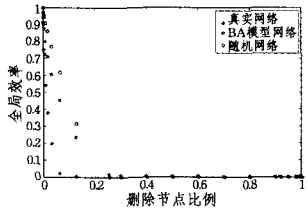

图 6 RD 策略攻击不同网络的结果

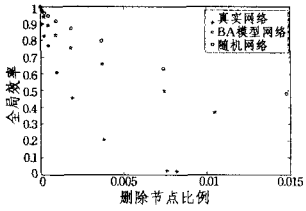

图 7 RB 策略攻击不同网络的结果

2)随机攻击策略

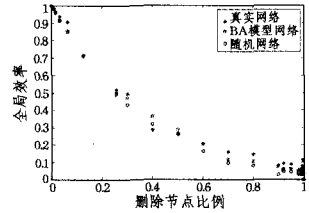

图 8 随机策略攻击不同网络的结果

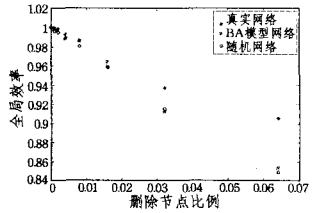

图 9 随机策略攻击不同网络的前期结果

对同样规模的 BBS 网络、无标度网络和随机网络进行随

（下转第 34 页）

这个例子中由于 p 的大小超过了 buf 的限制，在调用 strcpy 函数之后，肯定会发生缓冲区溢出。表 3 中列出了 MCBuffer 转化生成的 Promela 代码，可以直接利用 SPIN 进行验证。表中黑斜体部分表示在 SPIN 中需要进行确认和验证的属性。如果这些性质不满足，则说明源程序中相关的缓冲区操作存在缓冲区溢出。

图 2 是例 3 通过 MCBuffer 检测的执行结果。可以看出，MCBuffer 成功地检测到了源代码中存在的缓冲区溢出问题。

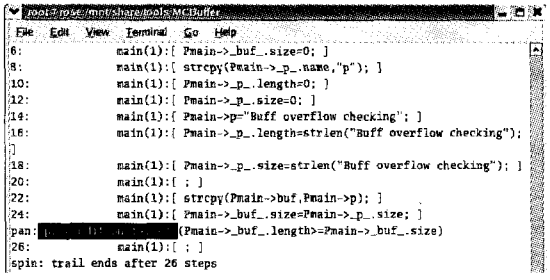

图 2　MCBuffer 检测结果

结束语　传统的静态方法误报较多，所有的可能漏洞都需要人工进行分析[5]；基于模型检验的缓冲区溢出检测方法属于静态方法的范畴，但若能够充分利用模型检验产生的反例，反例验证技术和模型自动精化技术则会有较低的误报率，可减少人工的参与。相对于动态检测方法而言，此方法又具有静态方法的特点，能够在系统运行之前进行检测，从而不会引入性能损失。因此基于模型检验进行缓冲区溢出检测应该是一个非常值得研究的方法。

但是本方法目前仍然处于研究的初级阶段，仅仅实现了一个简单的原型工具，整个工具中作用十分重要的反例验证器和模型自动精化技术的研究还没有深入开展，这也是后续工作的重点。

参 考 文 献

[1] Wagner D, Foster J S. A First Step Towards Automated Detection of Buffer Overrun Vulnerabilities[EB/OL]. http://www.cs.berkeley.edu/~daw/papers/overruns-ndss00.ps, 2000

[2] 张明军. 缓冲区溢出分析中的指针分析技术研究[D]. 长沙：国防科技大学，2004

[3] Holzmann G J, Smith M H. Software Model Checking: Extracting verification models from source code[J]. Software Testing Verification and Reliability, 2001, 11(6): 65-79

[4] Holzmann G J. The Model Checker Spin [J]. IEEE Trans on Software Engineering, 1997, 23(5): 279-295

[5] Larochelle D, Evans D. Statically Detecting Likely Buffer Overflow Vulnerabilities[C]// Proceedings of the USENIX Security Symposium. 2001

(上接第 30 页)

机攻击，3 个网络的全局效率随删除节点比例的变化如图 8、图 9 所示。由图 8 易知随机攻击策略对 3 个网络的影响较蓄意攻击更小，因 3 条曲线达到崩溃的速度明显减缓；但 3 条曲线的整体趋势类似，这说明当采用随机攻击策略，对 3 个网络的影响相当。由图 9 可知，在攻击初期，随机攻击对真实网络的影响略小于无标度网络和随机网络，这说明真实网络有一定的容错能力。

由以上两部分的结果分析可知，随机攻击对不同网络的影响不大，但蓄意攻击却能破坏网络的非同质结构，直接导致网络崩溃，对 BBS 用户回复网络的影响尤其明显，这说明在随机攻击下，无标度网络和随机网络有更强的稳定性。在蓄意攻击下，BBS 用户回复网络崩溃得更早，只要少数“核心节点”被移除，整个网络就会陷入瘫痪。这表明 BBS 用户回复网络面对蓄意攻击显得异常脆弱。

结束语　结合复杂网络理论，首先基于节点度和介数概念提出了 5 种攻击策略，并阐述了随机网络、无标度网络以及 BBS 用户回复网络的构建方法；其次描述了网络抗毁性的定义及其测度；最后采用不同的攻击方法对不同的网络进行攻击，对攻击结果作了详细的分析，结果表明蓄意攻击能在短时间内破坏网络结构，使网络崩溃，效果显著，其中 BBS 用户回复网络最不堪一击，抗毁能力远不及随机网络和无标度网络；但进行随机攻击时，BBS 用户回复网络体现了较好的容错能力。为了使 BBS 用户回复网络遭受蓄意攻击时受到的损害程度最低，必须对网络中的重要节点施以保护，使其体现一定的现实意义。

本文对网络的抗毁性作了初步的研究，但仍存在不足：采用的抗毁性测度即全局效率不能很好地区分 4 种蓄意攻击策略的攻击效果，这也将是下一步的研究方向，即优化抗毁性测度，使之更精确地描述每次攻击的结果。

参 考 文 献

[1] 章忠志，周水庚，方锦清. 复杂网络确定性模型研究的最新进展[J]. 复杂系统与复杂性科学，2008，5(4)：29-46

[2] 丁琳，谭敏生，肖炜. 复杂网络抗毁性研究综述 [J]. 电脑知识与技术，2009，5(1)：51-61

[3] Barabási A L, Albert R. Emergence of Scaling in Random Networks [J]. Science, 1999, 286(5439): 286-309

[4] Barabási A L, Albert R. Statistical Mechanics of Complex Networks [J]. Reviews of Modern Physics, 2002, 74(1): 47-97

[5] 章忠志. 复杂网络的演化模型研究 [D]. 大连：大连理工大学，2006

[6] 孟微. 复杂网络分析方法在情报学科研合著网络分析中得应用[D]. 北京：中国科学技术信息研究所，2008

[7] 吴渝，肖开洲，刘洪涛，唐红. BBS 虚拟社区的演化规律探索及仿真[J]. 系统工程理论与实践，2010，30(10)：1883-1890

[8] Wu Min, Li Hui, Zhang Ke, et al. An Evolutionary Model of Reply Network on Bulletin Board System [C]// International Conference of Information Technology, Computer Engineering and Management Science: ICM 2011. Nanjing, 2011

[9] Freeman L. Sociometry. 1977, 40: 35-41

[10] Holme P, Kim B J, Yoon C N, et al. Attack vulnerability of complex networks[J]. Phys. Rev. E, 2002, 65(5): 056109

[11] 丁琳，谭敏生，肖炜. 复杂网络抗毁性研究综述[J]. 电脑知识与技术，2009，5(1)：51-53

[12] Crucitti P, Latora V, Marchiori M, et al. Error and Attack Tolerance of Complex Networks[J]. Physica, 2004, 340: 388-394