一种基于社会网络分析的P2P僵尸网络反制策略

陈端兵 万 英 田军伟 傅 彦
(电子科技大学计算机科学与工程学院 成都 610054)

摘 要 设计并实现了一种基于社会网络分析的P2P僵尸网络反制策略。该策略包括两个方面:第一,挖掘网络中的关键节点和桥梁节点,重点防护这两类节点;第二,挖掘网络中的社区结构,重点监控社区间的关键通讯。模拟实验表明,该策略能准确挖掘网络中的关键节点、桥梁节点和关键通讯边,挖掘桥梁节点和关键边的准确率分别达到了95%和93%。使用提出的策略可有效控制僵尸网络病毒和黑客攻击指令的传播,从而达到反制P2P僵尸网络的目的。

关键词 社会网络分析,P2P僵尸网络,社区结构,反制策略

P2P Botnet Control Strategy Based on Social Network Analysis

CHEN Duan-bing WAN Ying TIAN Jun-wei FU Yan
(School of Computer Science, University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract A P2P botnet control strategy based on social network analysis was designed. The strategy includes two aspects: the first is to mine and protect the key nodes and bridge nodes in the P2P botnet; the other is to detect communities in the network and survey the key edges between communities. Experimental results show that the strategy proposed in this paper could mine the key nodes, bridge nodes and key edges precisely. Precision of bridge nodes and key edges mining is 95% and 93%, respectively. The propagation of botnet virus and hackers' attacking commands could be controlled effectively by the strategy proposed in this paper.

Keywords Social network analysis, P2P botnet, Community structure, Control strategy

1 引言

近年来,僵尸网络已对网络安全和用户数据安全带来了极大的威胁,已成为安全领域研究和讨论的热点问题,引起了国内外安全业界的高度重视。为了让僵尸网络更具隐蔽性和韧性,黑客不断地对僵尸网络组织形式进行创新和发展,出现了基于P2P协议构建命令与控制信道的僵尸程序,著名的案例包括通过构建P2P网络支持DDOS攻击的Slapper[1]、使用随机扫描策略寻找邻居节点的Sinit[2]、基于WASTE协议构建控制信道的Phatbot[3]等。Slapper, Sinit, Phatbot, Spam Thru, Nugache和Peacomm等P2P僵尸网络实现了各种不同的P2P控制机制[4]。Wang等人[5]提出了一种更加先进的混合型(即半分布式)P2P僵尸网络命令与控制机制的设计框架,Vogt等人[6]则提出了一种层叠化的"super-botnets"僵尸网络群构建方式,其通过小型僵尸网络间邻居节点关系和基于公钥加密的通信机制构造僵尸网络群。

在僵尸网络的防御和反制方面,Overton等人给出了防御僵尸程序感染的方法[7],其中包括遵循基本的安全策略以及使用防火墙、DNS阻断、补丁管理等技术手段。对于集中式僵尸网络,在发现僵尸网络控制点的基础上,最直接的反制方法是通过CERT(Computer Emergency Response Team)部门协调、处理、关闭控制点,防御者还可以使用DNS劫持技术[8]来获取僵尸主机的IP列表,从而及时通知僵尸主机用户及时移除僵尸程序。然而,P2P僵尸网络不存在集中的控制点,因此对P2P僵尸网络的反制更加困难。

目前P2P网络比较常见的是半分布式拓扑结构。半分布式P2P网络具有复杂网络的一些典型特征:(1)无标度特征,即网络中节点度的分布服从幂律分布;(2)社区结构特征,即整个网络由若干个社区构成。社区内部节点之间存在较多的连接,而各个社区之间的连接相对较少。

本文利用社会网络分析技术[9],针对P2P僵尸网络的特点,提出了一种P2P僵尸网络反制策略。其主要思想是:找到网络中的关键僵尸机和起"桥梁"作用的僵尸机,或者僵尸机之间通讯的关键链路,对这些僵尸机节点或者链路进行严密监控,在必要时将其从网络中移除,从而把整个僵尸网络截断为几个小的部分,有效阻止攻击指令到达目的主机,将攻击的破坏范围降到最小。通过实验验证了算法的有效性,挖掘桥梁节点和关键边的准确率分别达到了95%和93%,从而达

到稿日期:2008-07-04 返修日期:2008-12-23 本文受国家高技术研究发展计划(编号:06AA01Z414,07AA01Z440),国家242信息安全计划项目(编号:2007B27),四川省应用技术研究与开发项目支撑计划(编号:2008GZ0009)资助。

陈端兵 博士,讲师,主要研究方向为数据挖掘、社会计算、NP难问题高效求解等,E-mail:dbchen@uestc.edu.cn;万 英 硕士研究生,主要研究方向为数据挖掘;田军伟 硕士研究生,主要研究方向为数据挖掘;傅 彦 教授,博士生导师,主要研究方向为数据挖掘、信息安全、模式识别等。

到有效破坏僵尸网络的目的。

2 相关概念

2.1 社会网络分析

社会网络分析[9](Social Network Analysis,SNA)是数据挖掘的一个重要分支。它是一种链接分析(Link Analysis)技术,通过研究社会网络,理解社会网络结构和行动者的行为过程。目前,社会网络分析技术已渗透到各种应用领域,如经济领域、广告传播与市场营销、计算机网络病毒与犯罪网络、恐怖袭击预测以及 IT 与通信行业等[10]。

(1)中心性

中心性用来评价个体在网络中的重要程度[9]。中心性通常包括程度中心性、亲近中心性和中介中心性 3 种常见的形式。程度中心性与中介中心性是计算个体在网络中重要程度的最主要的两项指标,根据这两个概念可提出一种 P2P 僵尸网络反制策略。

程度中心性可用于衡量哪些节点是团体中的主要节点。这样的个体,在社会学的意义上就是社会地位很高的人,在组织行为学上则是权力很大的人,在社会网络中则是很有影响力的成员。程度中心性高的个体,在团体中拥有高的地位。在社会网络中,程度中心性即是一个节点的关系数量的总和。采用下面的公式计算个体的程度中心性:

$$C_D(n_i) = d(n_i) = \sum_j X_{ij} \tag{1}$$

式(1)体现的是个体和外界的关系总和。其中,X_{ij} 取 0 或 1,1 代表行动者 j 与行动者 i 有关系,否则表示没有关系。为了便于在不同网络间进行比较,对式(1)进行标准化,如式(2)所示。

$$C_D'(n_i) = \frac{d(n_i)}{g-1} \tag{2}$$

其中 g 是网络中的节点数。

(2)关键节点

网络中度很大的节点,即此类节点和其他很多节点有联系,具有较高的程度中心性。图 1 中的节点“A”,“B”和“C”就是此网络中的关键节点。

(3)桥梁节点

网络中流量很大的节点,许多节点间通信都必须通过此节点。这类节点在网络中扮演着桥梁角色,对于整个网络的连通性有着至关重要的作用。如图 1 中的节点“G”和“M”即是桥梁节点。

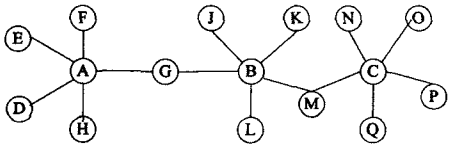

图 1 关键节点、桥梁节点示意图

2.2 社区结构

复杂网络的一个重要特征就是网络所呈现出的社区结构(Community Structure)。社区是指网络中同一类型的节点以及这些节点之间的连边所构成的子图。社区内部的节点连接非常紧密,各个社区之间的连接相对比较稀疏。复杂网络中社区结构挖掘的研究起源于社会学的研究工作,Wu 和 Huberman[11]以及 Newman 和 Girvan[12-14]的研究成果,使得复杂网络中的社区结构挖掘成为近几年复杂网络研究领域的一个热点并形成了一个重要的研究方向。

(1)关键边

社会网络图中各个社区之间的连边称为关键边。社区间信息传递必定会经过关键边。

(2)模块度

“模块度”是社区结构挖掘技术中一个非常重要的概念,它是 Newman 和 Girvan 定义的一个用来评价网络分解满意度的指标[12]。假定网络已经被分裂成 g 个社区,定义一个 $g\times g$ 的矩阵 e,其元素 e_{ij} 表示原网络中连接社区 i 和社区 j 的节点的边数与网络总边数的比值。该矩阵的迹 Tre 表示网络中连接同一社区中节点的边与网络总边数的比值,该矩阵行元素的和 $a_i=\sum_j e_{ij}$ 表示连接社区 i 中的节点的边与网络总边数的比值(该数值只依赖于网络中节点的度,不依赖于其它网络特征,如社区结构)。如果将网络看成一个节点度给定前提下所生成的随机网络,那么连接社区 i 和社区 j 中的节点的边数与网络总边数的比值为 a_ia_j,特别地,连接社区 i 中的节点的边数与网络总边数的比值为 a_i^2。因此,若将网络看成一个节点度给定的随机网络,则连接同一社区中节点的边的数量与网络总边数的比值为

$$\sum_i a_i^2 = \sum_i \sum_{jk} e_{ij}e_{jk} = ||e^2|| \tag{3}$$

其中,$||x||$表示向量 x 中所有元素绝对值之和。于是网络的模块度定义为

$$Q = Tre - ||e^2|| \tag{4}$$

3 僵尸网络反制策略

3.1 P2P 僵尸网络的特征

根据 Stutzbach 等人[15]对 Gnutella 网络拓扑分析和我们对 P2P 网络的模拟分析,得出目前的 P2P 网络具有以下几个典型的特征:

(1)无标度特性。即网络中节点度的分布服从幂律分布规律。

(2)网络中存在一部分关键节点,关键节点具有很高的程度中心性。一旦这些关键节点感染病毒信息,那么僵尸网络病毒将会以非常快的速度在网络中散播。如果我们对这部分关键节点进行保护,则能有效阻碍僵尸网络病毒的传播。

(3)网络中的关键节点彼此之间存在联系。该联系可能是直接联系,也可能是间接联系,即关键节点通过其他节点相连。

(4)网络中两节点之间一般是选择最短路径来相互通信的。

(5)网络中存在一部分桥梁用户,这些用户有较高的中介中心性,在网络中起着桥梁的作用,这类用户本身可能就是关键用户。如果找到并对这些节点进行防护,那么黑客发送的攻击指令就不能顺利地传播,DDOS 等攻击就不会造成大的破坏,僵尸网络的危害就可以降到较低的层次。

(6)网络中存在社区结构,即整个网络由若干个社区构成。

3.2 断点反制算法

断点反制算法的基本思想是挖掘网络中的关键节点。通过关键节点找到桥梁节点,重点监控这些桥梁节点,在必要时删除这些桥梁节点,以达到阻碍僵尸网络病毒或者黑客攻击指令传播,从而无法攻击目标主机。算法的具体流程如下:

step 1　遍历整个网络，挖掘网络中的关键节点，关键节点的数目可取网络总节点数的 10%～15%；

step 2　计算每两个关键节点之间的最短路径，记录最短路径经过的节点；

step 3　对步骤 2 得到的节点序列，统计其中每个节点出现的次数，出现次数最多的节点即是网络的桥梁节点；

step 4　删除步骤 3 中找出的桥梁节点及与该节点相连的边；

step 5　重复步骤 1－4，直至网络分成若干个不连通的子网络。子网络的数目取决于僵尸网络本身的规模和该僵尸网络的危害程度。

逐个删除桥梁节点后，单个连通网络的规模就控制在较小范围内，从而有效地控制僵尸病毒攻击指令向外传播。

3.3　断边反制算法

断边反制算法的基本思想是挖掘网络中社区间通信的关键边，在监控到网络异常时断开这些关键边，从而可以及时防止异常信息向整个网络散播。

3.3.1　基本概念和符号

为了准确找到网络中的关键边，本文设计并实现了一种基于分裂思想的社区挖掘算法。算法中涉及到如下几个基本概念和符号。

(1)边介数：网络中节点间通信通过边的信息流量，与网络中经过边的最短路径的数目有关。本文中边介数的计算采用了 Newman 和 Girvan 提出的边介数计算方法[12]。

(2)Q：社区的模块度，是社区划分优劣的度量标准，文献[16]给出了模块度 Q 的详细定义。

(3)L 层：是指按照宽度优先搜索，从起始节点 i 到达目标节点 j 的路径长度。起始节点记为第 0 层。

(4)$k_i^l(j)$：表示第 l 层节点 j 对当前社区的贡献度，其值等于那些与节点 j 邻接但还没有加入到当前社区且不在 l 层的节点的个数。

(5)Δk_i^l：$\Delta k_i^l=\frac{k_i^l}{k_i^{l-1}}$ 表示第 l 层节点的贡献度与第 $l-1$ 层节点的贡献度的比值。

(6)α：为一个预先设定的阈值，是一个经验值，一般取为 1.2[17]。

3.3.2　局部社区挖掘

本过程挖掘网络中某个节点 i 所属的局部社区，此过程的伪代码描述如下：

step 1　从节点 i 出发，$l=0$，初始化社区 $C=\{i\}$，计算 k_i^0；

step 2　$l++$，将 l 层的节点的所有邻居节点添加进当前社区 C，计算 k_i^l；

step 3　计算 $\Delta k_i^l=\frac{k_i^l}{k_i^{l-1}}$，如果 $\Delta k_i^l<\alpha$，转到步骤 4；否则返回到步骤 2；

step 4　考察最后一轮加入社区的每个节点，剔除那些与社区外部联系更多的节点。

3.3.3　网络社区挖掘算法

网络社区挖掘算法伪代码描述如下：

step 1　计算网络 G_0 中所有边的边介数，找到边介数最大的边 e_{ij} 并删除(i,j 分别是这条边连接的两个顶点)，得到网络 G_1；

step 2　在网络 G_1 中，用 3.3.2 节的方法挖掘节点 i 所在的局部社区 C_1；

step 3　计算社区 C_1 模块度 Q_1；

step 4　在网络 G_1 中，用 3.3.2 节的方法挖掘节点 j 所在的局部社区 C_2；

step 5　计算社区 C_2 模块度 Q_2；

step 6　选取 Q_1，Q_2 中较大者所对应的网络划分方式。

社区划分过程结束时，关键边自然也就找到了，根据定义，关键边就是社区之间的连边。得到网络中的关键边后，可在监控到网络异常时断开这些边，从而破坏网络的连接，阻止僵尸程序或者攻击命令的顺利传播。

4　实验结果与分析

对本文提出的算法，在奔腾 4 计算机(1.8GHz 主频，256MB 内存)上用 Java 语言编程实现，分别用两个网络：(1)含有 55 个节点的仿真 P2P 网络；(2)空手道俱乐部内部成员的关系网络，对算法的有效性进行了验证。

4.1　实验一

利用本文提出的断点反制算法挖掘出了仿真 P2P 网络(如图 2 所示)中的 4 个关键节点，如表 1 所列。然后根据关键节点之间的最短路径经过节点的次数挖掘出了网络中的两个桥梁节点，分别是 18 号节点和 11 号节点。将这两个桥梁节点从网络中删除后，原始的网络被分成了互不连通的 3 个部分，网络得到了有效的破坏，如图 3 所示。

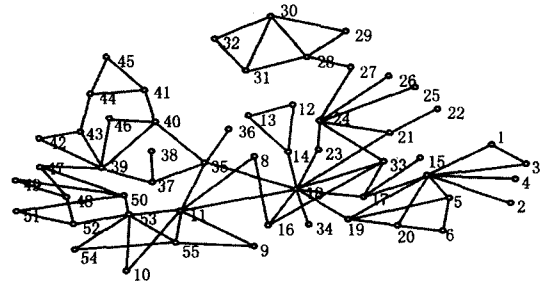

图 2　含有 55 个节点的 P2P 网络

表 1　首次执行时得到的关键节点列表

节点编号	节点的度
18	10
7	8
11	7
24	6

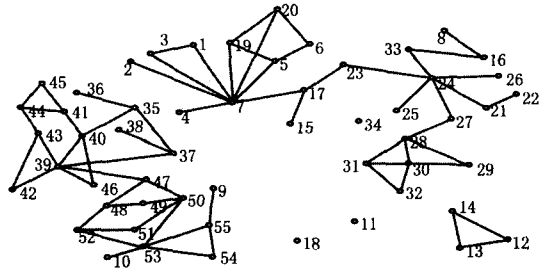

图 3　去除 2 个桥梁节点后的网络

从实验结果可以看出，本文提出的断点反制算法可有效地破坏网络，对僵尸网络的控制有较好的效果。

利用本文提出的断边反制算法将网络分成了两个较大的

社区，删除两个社区相连接的 3 条边（如表 2 所列）之后的网络如图 4 所示。

表 2　社区间的边

社区间的边	端点编号
e_1	35,18
e_2	11,18
e_3	8,16

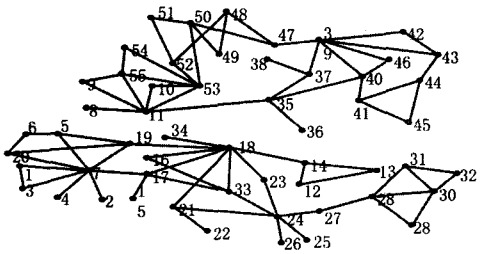

图 4　去除 3 条边后的网络

从图 4 可以看出，删除社区间的连边后，原始的网络被分成了互不连通的 2 个部分，网络也得到了有效的破坏。若需要进一步破坏网络的结构，则重复执行断边反制算法，即可达到目的。

4.2　实验二

本文还利用 Zachary 研究的空手道俱乐部内部成员的关系网络（如图 5 所示）[18] 对算法性能进行了测试。

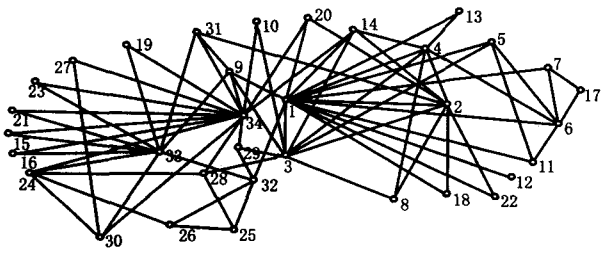

图 5　Zachary 研究的空手道俱乐部内部成员的关系网络

利用本文提出的断点反制算法挖掘出了 Zachary 网络中的 5 个关键节点，如表 3 所列。利用本文提出的桥梁节点发现方法，依次挖掘出了其中 3 个桥梁节点，分别是 1 号节点、3 号节点和 34 号节点。将这 3 个桥梁节点从网络中删除后，原始的 Zachary 网络被分成了互不连通的 6 个部分，其中还有 4 个部分被孤立起来了，网络得到了有效的破坏，如图 6 所示。

表 3　首次执行时得到的关键节点列表

节点编号	节点的度
34	17
1	16
33	12
3	10
2	9

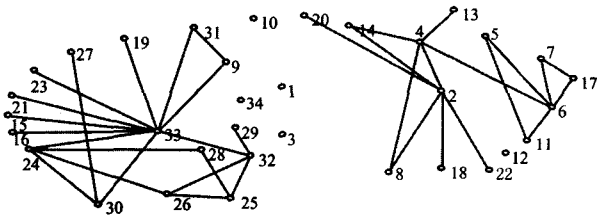

图 6　去除 3 个桥梁节点后的 Zachary 网络

利用本文提出的断边反制算法将 Zachary 网络分成了两个较大的社区，删除连接两个社区间的 9 条边之后的网络如图 7 所示。

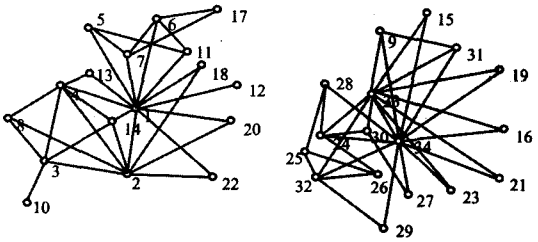

图 7　去除关键边后的 Zachary 网络

从实验结果可以看出，本文提出的算法能有效地对网络进行破坏，对僵尸网络的控制有较好的效果。我们在实验中发现关键节点的数目对最终桥梁节点的获取有一定影响。本文对于关键节点的数目的选择是用经验值来确定的，关键节点的选取是今后进一步研究的课题。

结束语　本文基于社会网络分析方法提出了一种僵尸网络反制策略，利用模拟的 P2P 网络和 Zachary 网络对算法的有效性进行了验证。实验表明：该策略可以比较准确地发现网络中的关键节点、桥梁节点和关键边。通过对关键节点的防护，可以有效地减缓僵尸网络的传播速率；通过对桥梁节点进行防护，能够有效地控制僵尸网络的传播；通过对关键边进行监控，能够以较小代价破坏僵尸网络的拓扑结构。未来的工作将致力于算法中关键节点的研究和算法效率的进一步提高。

参 考 文 献

[1] Arce I, Levy E. An analysis of the slapper worm[J]. IEEE Security & Privacy, 2003, 1(1): 82-87

[2] Sinit P2P Trojan analysis[OL]. http://www.lurhq.com/sinit.html

[3] Phatbot Trojan analysis[OL]. http://www.lurhq.com/phatbot.html

[4] Grizzard J B, Sharma V, Nunnery C. Peer-to-Peer botnets: Overview and case study[C]// Proceedings of the 1st Workshop on Hot Topics in Understanding Botnets. Cambridge, MA, USA, Apr. 2007

[5] Wang P, Sparks S, Zou C C. An advanced hybrid peer-to-peer botnet[C]// Proceedings of the 1st Workshop on Hot Topics in Understanding Botnets. Cambridge, MA, USA, Apr. 2007

[6] Vogt R, Aycock J, Jacobson M J. Army of botnets[C]// Proceedings of the 14th Annual Network & Distributed System Security Conference. San Diego, California, USA, Mar. 2007: 111-123

[7] Overton M. Bots and botnets: Risks, issues and prevention[C]// Proceedings of the 15th Virus Bulletin Conference. Dublin, Ireland, Oct. 2005. http://momusings.com/papers/VB2005-Bots_and_Botnets-1.0.2.pdf

[8] Dagon D, Zou C C, Lee W. Modeling botnet propagation using time zones[C]// Proceedings of the 13th Annual Network and Distributed System Security Symposium. San Diego, Feb. 2006. http://www.isoc.org/isoc/conferences/ndss/06/proceedings/papers/modeling_botnet_propagation.pdf

（下转第 111 页）

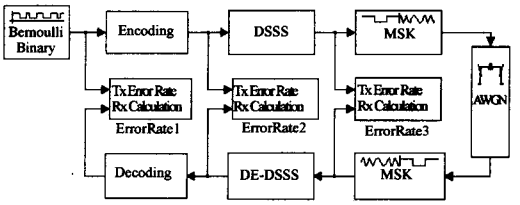

图 4　基于 SIMULINK 的 JTIDS 通信仿真系统结构

图 5 中的 a,b,c 分别表示未编码的调制系统、带有 RS 编码的调制系统以及带有 RS 编码与交织的调制系统 3 种抗干扰体制下，符号误码率与信噪比的关系，可见 RS 编码与交织能够极大的改善数据链系统的抗干扰性能，根据图 c 可知，当信噪比$\frac{E_s}{n_0}\geqslant 3$时，符号误码率低于 10^{-5}，满足数据链通信的要求。

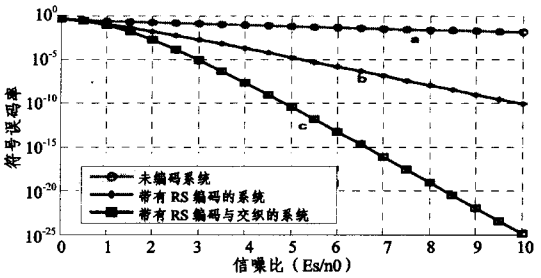

图 5　符号误码率与信噪比的关系

假设信道为高斯白噪声信道且信号只受到部分频带干扰的影响，通常情况下 $j_1'>>n_0$，所以对信道进行抗干扰性能分析时，忽略高斯白噪声对系统性能的影响，只存留部分频带干扰，此时 $p_s=\mu Q\left(\sqrt{\frac{\mu E_s k}{j_1' n}}\right)$。如图 6 所示，图 a,b,c 分别表示信干比$\frac{E_s}{j_1'}$等于 10dB,15dB,20dB 时，符号误码率与 μ 的关系，随着 μ 的变化系统符号误码率出现最大值，此时为最大部分频带干扰(干扰效果最佳)，也就是说，这时的 μ 值造成系统性能大大降低，继而符号误码率不断下降。

如图 7 所示，图 a,b,c 分别表示 μ 等于 1,0.8,0.6 时，符号误码率与信干比$\frac{E_s}{j_1'}$的关系，随着信干比的增大，符号误码率与信干比近乎成反比例关系。另外，当信干比不是很大时，μ 越大，符号误码率越大；当信干比达到一定程度时，μ 越大，符号误码率越小，这是因为干扰噪声的功率不够大时，增加干扰的覆盖范围，只会降低部分频带干扰的效果。

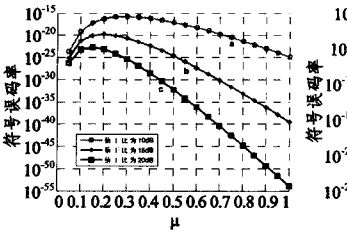

图 6　符号误码率与 μ 的关系

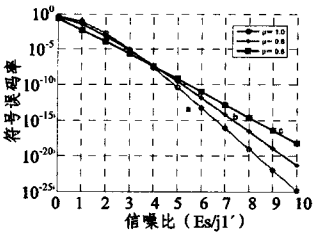

图 7　符号误码率与信干比的关系

结束语　本文以 JTIDS 为背景，研究了数据链通信系统的结构以及采用的抗干扰体制，分析了该系统在不同信道环境和人为干扰方式下的符号误码率，并运用 SIMULINK 通信仿真平台对该系统在高斯白噪声及部分频带干扰情况下的抗干扰性能进行了仿真，对仿真结果进行了分析。分析结果显示，由于 JTIDS 系统采用了 RS 编码、交织等多层纠检错编码体系以及直扩、MSK 调制和跳频的组合调制体制等一系列抗干扰技术，使其在有高斯白噪声以及部分频带干扰情况下，误码率也较低，其抗干扰的性能较好，满足战术数据传输要求。

参 考 文 献

[1] 刘徐德. 战术通信、导航定位和识别综合系统文集(第一集)[M]. 北京：电子工业出版社，1991

[2] Sklar B. 数字通信基础与应用(第二版)[M]. 徐平平，宋铁成，等译. 北京：电子工业出版社，2002：335-337

[3] Proakis J G. Digital Communications[M]. 3rd ed. McGraw-Hill, 1995

[4] 曲伟华. JTIDS 系统的 RS 码性能分析[J]. 电讯技术，2007，47(2)

[5] Zheng H, Zhang N. Performance Analysis of Hybrid DS-SFH/MSK Spread-Spectrum System under Multione Jamming[C]// IEEE MILCOM'99. 1999：567-570

[6] Ryu Heung-Gyoon. Effects of Partial-Band Interference on the Hybrid DS/SFH MSK System in Rayleigh Fading Channel. IEEE MILCOM

[7] Tan Z, Blake I F. Performance analysis of noncoherent DS-SFH spread spectrum multiple access for indoor wireless communications[C]// IEEE MILCOM '92. 1992, 3：851-855

[8] 杨光，周经伦，罗鹏程. 基于 SIMULINK 的数据链通信系统仿真研究[J]. 系统仿真学报，2008

[9] 侯印鸣. 综合电子战[M]. 北京：国防工业出版社，2000

(上接第 104 页)

[9] 罗家德. 社会网络分析讲义[M]. 北京：社会科学文献出版社，2005：150-156

[10] 温粉莲，唐常杰，乔少杰，等. 基于社会网络最短路径挖掘犯罪集团核心[J]. 计算机科学，2006，33(11)：266-268

[11] Wu F, Huberman B A. Finding communities in linear time: a physics approach[J]. European Physical Journal B, 2004, 38(2)：331-338

[12] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69：026113

[13] Newman M E J. Detecting community structure in networks[J]. European Physical Journal B, 2004, 38(2)：321-330

[14] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 69 (6)：066133

[15] Stutzbach D, Rejaie R. Characterizing today's Gnutella topology [R]. CIS-TR-04-02. University of Oregon, 2004

[16] Newman M E J. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences, 2006, 103(23)：8577-8582

[17] Bagrow J P, Bollt E M. Local method for detecting communities [J]. Physical Review E, 2005, 72：046108

[18] Zachary W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4)：452-473