基于免疫主机的蠕虫非线性传播新模型优化

佟晓筠 李巧军
(哈尔滨工业大学(威海)计算机科学与技术学院 威海 264209)

摘 要 基于Two-Factor传播模型提出了一种新的QSIRV传播模型,该模型更合理地考虑了被免疫主机的失效性。通过仿真得出,QSIRV模型较Two-Factor模型能够更好地描述蠕虫的传播规律以及传播过程中网络流量和蠕虫流量之间的相互影响,尤其是对免疫后的主机数目变化的仿真更是符合实际情况,同时考虑了已隔离、免疫及被感染主机的数量的影响以及人们对蠕虫传播的警惕性的提高。对QSIRV模型进行了进一步的改进。仿真结果验证,改进后的模型可以更快地遏制蠕虫传播。

关键词 蠕虫,蠕虫传播模型,模型优化,QSIRV模型,免疫主机
中图法分类号 TP393.08 **文献标识码** A

Optimization of New Nonlinear Propagation Model of Worm Based on Immune Hosts

TONG Xiao-jun LI Qiao-jun
(School of Computer Science, Harbin Institute of Technology, Weihai 264209, China)

Abstract The thesis analyzed the propagation model of worm in detail and the model was optimized. We presented a new propagation model of worm named as the QSIRV propagation model based on the Two-Factor propagation model. The model reasonably considers the failure of the immunity. The experimental results show the QSIRV model can better descript the worm propagation and the interaction between the worms spread of network traffic and worms flow, and especially the simulation of the changed number of the host immune is consistent with the actual situation, and the model also considers the influence for isolation, immune, number of infected hosts and the people vigilance to worm. The paper further improved the QSIRV model. The improved model can faster resist the worms spread.

Keywords Worm, Propagation model of worm, Optimization of model, QSIRV model, Distributed worm detection

1 前言

在恶意代码传播模型的研究中,病毒传播模型较多,而针对网络蠕虫传播的模型较少。现有的蠕虫的传播模型多基于生物学的传染病模型。目前,国内外提出的蠕虫传播模型有如下几种:经典的简单传染病模型(SEM)[1]、Kermack-Mckendrick模型(KM)[2]、TWO-FACTOR模型(TW)[3]、Worm-Anti-Worm模型[4]、针对各类网络应用的蠕虫传播模型P2P应用的蠕虫传播模型[5]。

文献[6]中提出IPv6网络中的路由蠕虫传播模型,并提出了一种新型路由蠕虫RoutingWorm-v6,即良性蠕虫对抗恶性蠕虫的传播模型[7]。这表明,如果对良性蠕虫采取一些控制策略,将达到较好的抑制恶性蠕虫的效果。文献[8]在分析蠕虫传播特点的基础上提出了一种使用本地网协同检测蠕虫的算法CWDMLN。文献[9]提出了一种蠕虫检测机制,但这种检测机制必须部署在路由器上,对环境要求太高,不适用于一般的中小型网络。

针对已有蠕虫检测模型的不足,本研究详细分析和优化蠕虫传播模型,对于发现大范围内蠕虫传播,减少蠕虫对网络的破坏及预警,有着很大的现实意义。

2 Two-Factor模型研究

在Two-Factor模型中,主机同样有3种状态:"易感染"状态、"被感染"状态和"免疫"状态。与KM模型不同,主机还可能处于"易感染→免疫"的状态转化之中,也就是说Two-Factor模型考虑了"易感染"主机的免疫。

Two-Factor模型存在几个待定的动态参数:$\beta(t)$、$R(t)$和$Q(t)$。$\beta(t)$是感染率,随时间变化而变化;$R(t)$表示t时刻从"被感染"群体中被免疫的主机数量;$Q(t)$表示t时刻从"易感染"群体中被免疫的主机数量。

因此Two-Factor模型把易感染主机的免疫过程建模为

$$\frac{dQ(t)}{dt}=\mu S(t)J(t) \tag{1}$$

基于Two-Factor模型给出动态特性的假设,得到Two-Factor模型完整的微分方程组:

到稿日期:2011-06-12 返修日期:2011-10-03 本文受国家自然科学基金项目(60973162),山东省自然科学基金项目(ZR2009GM037),山东省科技攻关项目(2010GGX10132),哈尔滨工业大学(威海)校科学研究基金(HIT(WH)2009)资助。
佟晓筠(1963—),女,教授,博士生导师,主要研究方向为网络与信息安全、混沌密码学,E-mail:tong_xiaojun@163.com;李巧军 硕士生,主要研究方向为网络与信息安全。

$$\begin{cases} dS(t)/dt=-\beta(t)S(t)I(t)-dQ(t)/dt \\ dR(t)/dt=\gamma I(t) \\ dQ(t)/dt=\mu S(t)J(t) \\ \beta(t)=\beta_0[1-I(t)/N]^{\eta} \\ N=S(t)+R(t)+I(t)+Q(t) \\ I(0)=I_0<<N;S(0)=N-I_0;R(0)=Q(0)=0 \end{cases} \tag{2}$$

式中，γ 为被感染主机的免疫率，$J(t)=I(t)+R(t)$ 表示被感染过的主机，μ 是常数，$\mu J(t)$ 表示 t 时刻易感染主机的免疫率。

选取 $\gamma=0.03$，$N=100$，$I(0)=1$，$R(0)=0$，$Q(0)=0$，$\mu=0.01$，$\sigma=3$，β_0 分别为 0.02，0.04，0.06。根据式(1)和式(2)绘制出被感染节点数 $I(t)$ 关于时间 t 的函数图像，如图 1 所示。

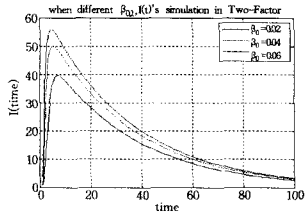

图 1　Two-Factor 模型中不同参数得到的蠕虫传播趋势

既然 Two-Factor 模型考虑到易感染主机的免疫率会随时间递增，那么被感染主机的免疫率 γ 也应该随时间递增，所以将 γ 作为常数不太合适。

3　蠕虫传播模型的优化

3.1　QSIRV 模型

上面提到的 Two-Factor 模型很好地描述了蠕虫的传播模型，但是进一步探索会发现，Two-Factor 模型并没有把传播过程中的某些因素考虑进去。

第一，免疫状态的主机群体中的某些主机又进入到易感染状态的主机群体。现在我们考虑一个实际存在的情况，被感染状态的主机群体中的主机 A 经过免疫一段时间后，因为失去免疫力回到了易感染主机群体，而变为易感染主机。

第二，隔离状态的主机群体中的某些主机又进入到易感染状态的主机群体。考虑另外一个实际存在的情况，主机 B 从易感染状态的主机群体中被隔离出来，进入到隔离状态的主机群体中，经过一段时间，人们认为网络中的蠕虫可能已经不存在了，于是将主机 B 再次接入网络，但是蠕虫还没有消失，因此主机 B 再次进入到了易感染状态的主机群体中。

在 QSIRV 模型中，主机有 4 种状态：易感染状态、被感染状态、免疫状态和隔离状态。其中免疫状态包括 $Q(t)$ 和 $R(t)$。$Q(t)$ 是从易感染的主机群体进行免疫得到的，而 $R(t)$ 是从被感染的主机群体进行免疫得到的。与 Two-Factor 模型不同，主机还可能处于“免疫→易感染”的状态转化和“隔离→易感染”的状态转化之中，QSIRV 模型如图 2 的数学表达。

QSIRV 模型存在几个待定的动态参数：$\mu(t)$、$\alpha(t)$、$\beta(t)$、$\gamma(t)$、$Q(t)$、$S(t)$、$I(t)$、$R(t)$、$V(t)$。$N=Q(t)+S(t)+I(t)+R(t)+V(t)$ 在任何 t 时刻都成立，是整个主机群体的总主机数目。其中 $\mu(t)$ 是从易感染状态到免疫状态的免疫率。$\alpha(t)$ 是从被感染状态到隔离状态的隔离率，随时间而变化。$\beta(t)$ 是感染率，也随时间而变化。$\gamma(t)$ 是从被感染状态到免疫状态的免疫率。$Q(t)$ 表示 t 时刻免疫状态 1 的主机数。$V(t)$ 表示 t 时刻隔离状态的主机数。$R(t)$ 表示 t 时刻免疫状态 2 的主机数。$I(t)$ 表示 t 时刻被感染状态的主机数。那么 t 时刻到 $t+\Delta t$ 时刻内，“易感染”主机数量变化值为：

$$S(t+\Delta t)-S(t)=-\beta(t)S(t)I(t)\Delta t-\frac{dQ(t)}{dt}\Delta t+\omega V(t)\Delta t+\kappa R(t)\Delta t \tag{3}$$

式中，$S(t)$ 为 t 时刻易感染主机数目，因此

$$\frac{dS(t)}{dt}=-\beta(t)S(t)I(t)-\frac{dQ(t)}{dt}+\omega V(t)+\kappa R(t) \tag{4}$$

QSIRV 模型同 Two-Factor 模型一样，也把易感染主机的免疫过程建模为：

$$\frac{dQ(t)}{dt}=\mu(t)S(t)J(t) \tag{5}$$

基于 QSIRV 模型考虑的两种情况，得到 QSIRV 模型完整的微分方程组：

$$\begin{cases} dQ(t)/dt=\mu(t)S(t)J(t) \\ \dfrac{dS(t)}{dt}=-\beta(t)S(t)I(t)-\mu(t)S(t)J(t)+\omega V(t)+\kappa R(t) \\ \dfrac{dI(t)}{dt}=\beta(t)S(t)I(t)-\alpha(t)I(t)-\gamma(t)I(t) \\ dR(t)/dt=\gamma(t)I(t)-\kappa R(t) \\ dV(t)/dt=\alpha(t)I(t)-\omega V(t) \\ \beta(t)=\beta_0[1-I(t)/N]^{\eta} \\ \alpha(t)=\alpha_0[1-V(t)/N]^{\varphi} \\ \mu(t)=\mu_0[1+Q(t)/N]^{\sigma} \\ \gamma(t)=\mu_0[1+R(t)/N]^{\sigma} \\ J(t)=I(t)+R(t)+V(t) \\ N=Q(t)+S(t)+I(t)+R(t)+V(t) \\ I(0)=I_0<<N;S(0)=N-I_0 \\ Q(0)=R(0)=V(0)=0 \end{cases} \tag{6}$$

式中，$\alpha(t)$ 为被感染主机的动态免疫率，随时间变化而变化。而 Two-Factor 模型中被感染主机的免疫率是静态的。动态免疫率更符合现实。α_0 为动态免疫率的初始值，指数 φ 是常数，用于调节免疫率对“被免疫”主机数量的灵敏度。μ 是常数，$J(t)$ 表示 t 时刻被感染过的主机总数，$\mu J(t)$ 表示 t 时刻易感染主机的免疫率。β_0 为感染率的初始值，指数 η 是常数，用来调节感染率对“被感染”主机数量的灵敏度。

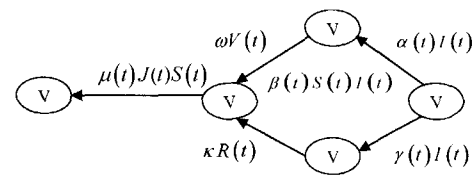

图 2　QSIRV 模型图

在 QSIRV 数学模型中，设置 $\sigma=0$、$\alpha_0=0$、$\omega=0$、$\kappa=0$，可以得到 Two-Factor 模型，所以 QSIRV 模型也可以得到 SEM 模型和 KM 模型。

选取 $\gamma_0=0.03$、$\beta_0=0.02$、$\mu_0=0.01$、$\kappa=0.01$、$\alpha_0=0.04$、$\omega=0.02$、$\varphi=3$、$\sigma=3$、$\eta=3$ 和 $N=100$。在 QSIRV 模型的仿真中，也可以让 ω 分别取值 0.02，0.04，0.06，来观察“免疫→易感染”这个新的状态转换对蠕虫传播模型的影响。

根据 QSIRV 模型完整的微分方程组，可以仿真出 QSIRV 模型 5 种主机群体的变化趋势，如图 3 所示。

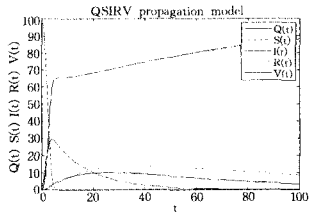

图 3　QSIRV 模型的总体趋势

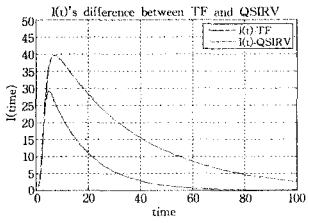

图 4　Two-Factor 和 QSIRV 两个模型中 $I(t)$的比较

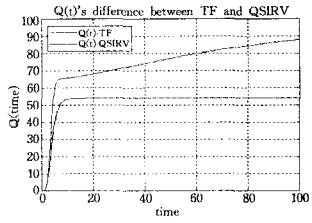

图 5　Two-Factor 和 QSIRV 两个模型中 $Q(t)$的比较

(1)通过图 4 Two-Factor 和 QSIRV 两个模型中 $I(t)$的比较，可见 Two Factor 模型中峰值为 39，接近 40，百分比是 39%；QSIRV 模型中峰值为 29，百分比是 29%。在 QSIRV 考虑了更多的因素后，被感染的主机群体的主机总数峰值比 Two Factor 模型中的要少。

(2)通过图 5 Two-Factor 和 QSIRV 两个模型中 $Q(t)$的比较，可见 Two-Factor 模型中 $Q(t)$的趋势在达到峰值之后几乎处于不变状态，而 QSIRV 模型中 $Q(t)$的趋势在没有达到峰值前前期和 Two-Factor 模型中 $Q(t)$的趋势是一样的，都是在前 10 个单位时间内陡然增大，从 10 个单位时间后边开始缓慢增加。在实际中，$Q(t)$不可能恒定不变，所以 QSIRV 可更好地描述实际情况。

(3)另外，通过图 3 QSIRV 模型的总体趋势，发现在 50 个单位时间后，因为人们对蠕虫的发现和抵御，实际上主机几乎处于免疫状态和有一部分是隔离免疫状态。通过仿真实验可以看出来，QSIRV 模型无论是与优化前的 Two-Factor 模型还是与优化后的 Two-Factor 模型进行比较，QSIRV 模型的总体趋势能够更加描述现实中的蠕虫传播模型。

QSIRV 传播模型是 Two-Factor 模型的改进，考虑了“免疫→易感染”和“隔离→易感染”这两种情况，一定程度上来说，QSIRV 传播模型更加完善地描述了现实当中的蠕虫传播模型。但是 QSIRV 传播模型将“被感染→隔离免疫”的隔离率 κ 看成常数是不合适的。还有，就是在总的群体 N 中，存在主机的“死亡”而推出群体 N 以及新主机进入群体 N 的情况，所以 QSIRV 传播模型也需进一步改进。

3.2　QSIRV 模型的改进

(1)考虑 $R(t)$和 $I(t)$综合数量对 $V(t)$的影响

$R(t)+I(t)$的数值变化对 $V(t)$是存在影响的。当 $R(t)+I(t)$数值变化时，人们对自己主机的隔离意识会受到影响，从而影响到 $V(t)$，故将$(R(t)+I(t))$作为“被感染→隔离”状态转换的一个因素。$V(t)$的微分方程是：

$$dV(t)/dt=\alpha(t)(R(t)+I(t))I(t)-\omega V(t) \tag{7}$$

(2)考虑 $V(t)$和 $I(t)$综合数量对 $R(t)$的影响

同理，$V(t)+I(t)$的数值对 $R(t)$也是存在影响的。当 $V(t)+I(t)$数值变化时，人们对自己主机的免疫意识会受到影响，从而影响到 $R(t)$，故将 $V(t)+I(t)$作为“被感染→免疫”状态转换的一个因素。$R(t)$的微分方程是：

$$dR(t)/dt=r(t)(V(t)+I(t))I(t)-\kappa R(t) \tag{8}$$

改进的 QSIRV 状态图如图 6 所示，数学模型如式(9)所示。

$$\begin{cases} dQ(t)/dt=\mu(t)S(t)J(t) \\ \dfrac{dS(t)}{dt}=-\beta(t)S(t)I(t)-\mu(t)S(t)J(t)+\omega V(t)+\kappa R(t) \\ \dfrac{dI(t)}{dt}=\beta(t)S(t)I(t)-\alpha(t)I(t)-\gamma(t)I(t) \\ dR(t)/dt=\gamma(t)(V(t)+I(t))I(t)-\kappa R(t) \\ dV(t)/dt=\alpha(t)(R(t)+I(t))I(t)-\omega V(t) \\ \beta(t)=\beta_0[1-I(t)/N]^{\eta} \\ \alpha(t)=\alpha_0[1-V(t)/N]^{\varphi} \\ \mu(t)=\mu_0[1+Q(t)/N]^{\sigma} \\ \gamma(t)=\mu_0[1+R(t)/N]^{\sigma} \\ J(t)=I(t)+R(t)+V(t) \\ N=Q(t)+S(t)+I(t)+R(t)+V(t) \\ I(0)=I_0<<N;S(0)=N-I_0 \\ Q(0)=R(0)=V(0)=0 \end{cases} \tag{9}$$

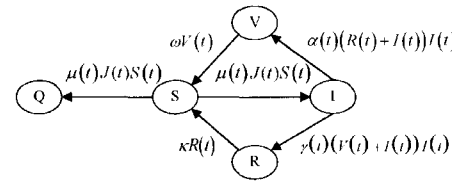

图 6　QSIRV 模型改进后的状态图

通过观察比较图 3 和图 7 发现，当考虑到周围主机感染蠕虫的情况时，易感染主机和被感染主机将在很短时间内被隔离免疫。通过仿真发现，如果人们能够及时从周围环境(如朋友、同事等)获取主机被感染情况，则可以很好地遏制蠕虫的传播。

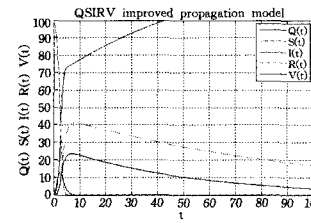

图 7　QSIRV 模型改进后的总体趋势

结束语　本文针对 Two-Factor 传播模型的不足，提出了新的传播模型优化模型，并对优化后的 QSIRV 模型进行了进一步改进。新优化后模型 QSIRV，尤其是对免疫后的主机数目变化的仿真更是符合实际情况。考虑到周围用户主机感染情况对自己隔离免疫主机产生的影响，对 QSIRV 模型进行了改进。改进后的 QSIRV 模型很好地证实和描述了对蠕虫传播的警惕性的提高，可以更快地遏制蠕虫传播。

参 考 文 献

[1]　Streftaris G, Gibson G J. Statistical Inference for Stochastic Epidemic Models [A] // Proceedings of the 17th International Workshop on Statistical Modeling [C]. Chain, 2002: 609-616

(下转第 140 页)

由于篇幅限制，本文仅对 Behavior MCM 进行描述。实体能够执行某种行为，说明其具有相应的功能。某种行为往往不是孤立地被执行的，而是从属于一定的任务。在行为的执行过程中，往往受到行为规则的约束，其结果可能产生一些信息。因此，与行为相关联的概念还包括功能、任务、规则、事件、信息。Behavior MCM 描述如图 4 所示。

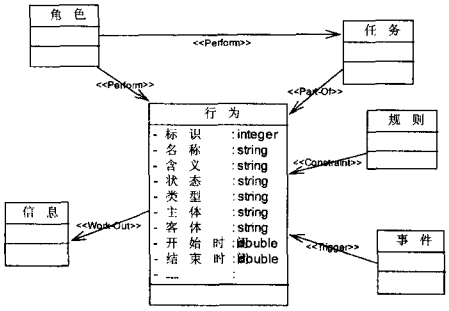

图 4　基于本体的 Behavior MCM 描述

4.3　装备保障仿真的 CM 构建

在建立装备保障系统的 OMCM 后，需要实现 OMCM 到概念模型的映射。本文以机动行为模型的构建为例，描述 Behavior MCM 向机动行为概念模型映射的过程。

在 UML 中，UML 活动图是 UML 用于对系统的动态行为建模的图形工具之一，描述了从活动到活动的流。其建模技术思想主要来源于事件图、SDL 状态建模技术和 Petri-Net 技术，非常适合于描述工作流和并发处理行为。UML 活动图的主要元素包括初始节点(Initial Node)、结束节点(Final Node)、活动(Activity)、动作(Action)、泳道(Swimlane)、对象节点(Object Node)、注释(Note)、控制流(Control Flow)、对象流(Object Flow)、分叉或结合(Fork/Join)、分支或合并(Branch/Merge)。

为了实现 Behavior MCM 向行为 CM 的映射，需要将 Behavior MCM 中的元素与 UML 活动图的基本元素进行关联。由图 4 可知，Behavior MCM 的元素包括行为、角色、任务、规则、事件和信息。其中，行为对应活动图中的动作；角色对应活动图中的泳道，每一个被泳道分开的部分表示不同的角色；任务可以认为是一种复杂的行为，因此也可以用活动图中的动作描述；规则是对行为逻辑的规范和约束，在 UML 活动图中，主要通过控制流、对象流、分叉或结合、分支或合并来描述；事件是驱动行为执行的激励，在活动图中用标有触发事件名的边表示；信息表达了实施行为过程中的交互情况，在活动图中用注释来表示。

本文以 Behavior MCM 向机动行为 CM 映射为例，用 UML 活动图来描述保障分队机动行为模型，其过程如图 5 所示。

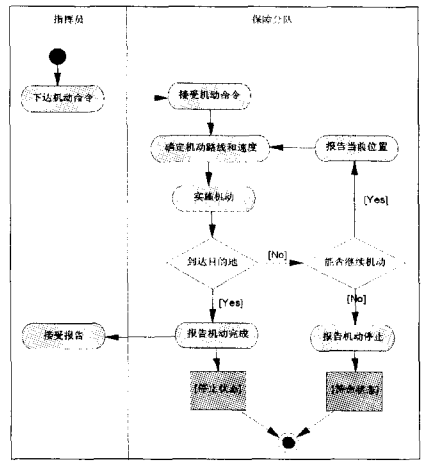

图 5　装备保障机动行为流程图

结束语　本文针对概念模型开发中存在的问题，引入本体思想和元建模思想，提出了基于本体的元概念模型(OMCM)的概念。通过将 OMCM 和概念模型进行映射，实现了概念模型的建模，体现了良好的重用性。通过将基于 OMCM 的概念建模方法应用于装备保障仿真系统概念模型的构建，证实了本方法的可行性和有效性。

参 考 文 献

[1] Balci O, Arthur J D, Nance R E. Accomplishing Reuse with a Simulation Conceptual Model[C] // Mason S J, Hill R R, Mönch L, et al. , eds. Proceedings of the 2008 Winter Simulation Conference. Miami, FL: IEEE, 2008: 959-965

[2] 何克清，何扬帆，王翀，等. 本体元建模理论与方法及其应用[M]. 北京：科学出版社，2008

[3] Heaton L. OMG Meta Object Facility(MOF) 1. 4 Specification [EB/OL]. http://www. omg. org/mda/specs. htm, 2010-03-05

[4] Guarino N, Giaretta P. Ontologies and Knowledge Bases: Towards a Terminological Clarification[C] // Mars I, ed. Towards very Large Knowl-edge Bases-Knowledge Building and Knowledge Sharing. N J: IOS Press, 1995: 25-32

[5] Evermann J. A UML and OWL description of Bunge's upper-level ontology model [J]. Software and Systems Modeling, 2009, 8(2): 235-249

[6] 苗东升. 系统科学大学讲稿[M]. 北京：中国人民大学出版社，2007

[7] 国刚，周峰，孙更新. UML 与 Rational Rose2003 软件工程统一原理与实践教程[M]. 北京：电子工业出版社，2007

(上接第 101 页)

[2] Wang Y, Wang C X. Modeling The Effects of Timing Parameters on Virus Propagation [A] // Proceedings of the ACM CCS Workshop on Rapid Malcode(WORM 2003) [C]. Washington D C: ACM Press, 2003: 61-66

[3] Dantu R, Cangussu J, Yelimeli A. Dynamic control of worm propagation[J]. Information Technology, 2004, 1(3): 419-423

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

[5] Zhou L, Zhang L, McSherry F. A first look at peer-to-peer worms: Threats and defenses[A] // Proceedings of 4th International Workshop on Peer-to-Peer Systems(IPTPS)[C]. 2005: 24-35

[6] 徐延贵，钱焕延，李华峰. IPv6 网络中的路由蠕虫传播模型[J]. 计算机应用研究，2009: 3920

[7] 张殿旭，彭军，何虹. Internet 蠕虫传播模型研究[J]. 网络通讯与安全，2007: 1244-1246

[8] 张新宇，卿斯汉，李琦，等. 一种基于本地网络的蠕虫协同检测方法[J]. Journal of Software, 2007: 412-421

[9] 赵广松，张涛. 基于蠕虫传播特性的蠕虫检测系统设计[J]. Computer Security, 2009: 114-118