文章编号: 1001- 2486(2010) 03- 0089- 05

基于两阶段感染过程分析的蠕虫传播模型 SSI*

刘 波[1], 刘 明[2], 肖 枫涛[1], 彭 磊[1]
(1. 国防科技大学 计算机学院, 湖南 长沙 410073; 2. 九江职业技术学院, 江西 九江 332007)

摘 要: 合理的蠕虫传播模型可以准确地描述蠕虫类恶意代码在网络中的传播行为, 有助于在开展蠕虫防护、检测和抑制技术的研究时更深入地分析蠕虫的传播机制。通过引入和分析蠕虫的感染时间这个在经典的传染病模型研究中被忽略的重要因素, 使用确定性建模分析方法推导得出 SSI 蠕虫传播模型, 合理地反映网络中的脆弱性主机在蠕虫传播时可能的状态转换。模拟仿真表明, SSI 模型可以相对准确地描述蠕虫类恶意代码在网络中的传播过程。

关键词: 传染病模型; 两阶段感染; 蠕虫感染时间; SSI 传播模型
中图分类号: TP309. 05; TP393. 08 文献标识码: A

SSI, a Worm Propagation Model Based on the Analysis of the Two-stage Infection

LIU Bo[1], LIU Ming[2], XIAO Feng-tao[1], PENG Lei[1]
(1. College of Computer, National Univ. of Defense Technology, Changsha 410073, China;
2. Jiujiang Vocational and Technical College, Jiujiang 332007, China)

Abstract: Logical worm propagation model can characterize how the worm-like malicious code spreads in network exactly, which contributes the analysis of worm propagation mechanism in depth to engaging in the study of worm defense, detection and containment. The significance of infection time is often neglected in worm propagation models based on classic epidemic model. By introducing and analyzing infection time, a worm propagation model called SSI is derived using deterministic modeling methods. This model reflects the possible state-transition of vulnerable hosts during worm propagation, and results in a better understanding and modeling of the propagation of Internet worms.

Key words: epidemic model; two-stage infection; worm infection time; SSI model

恶意代码, 尤其是兼具自我复制(Self-Replication) 和主动传播(Active-Propagation) 能力的蠕虫类恶意代码, 越来越成为影响网络安全的最严重威胁之一, 因此, 如果能够围绕着蠕虫类恶意代码的渗透传播机制开展深入研究, 准确剖析可能影响其传播速度的潜在因素, 将对于深入研究并且合理部署防护检测手段、遏制蠕虫类恶意代码的威胁具有重要的意义。而作为蠕虫类恶意代码渗透传播机制研究的重要内容, 围绕着传播模型的研究就是希望能够通过数学建模的方法尽量准确地描述蠕虫类恶意代码在网络中的渗透传播过程, 深入分析可能采用了不同扫描策略的蠕虫类恶意代码在网络中渗透传播时其数量规模随传播时间而发生的动态变化, 为更有针对性地开展防护检测技术的研究提供重要的理论依据。

1 蠕虫传播模型的相关研究

由于蠕虫类恶意代码在自我复制、主动传播等行为上都与生物界中的病毒存在着相似之处, 因此在病理学领域中研究病毒的传染过程时所使用的部分分析方法就很自然地被引入到蠕虫类恶意代码的传播模型研究中, 进而形成了 SEM 模型[1]、RCS 模型[2]、KM 模型[3] 和双因素模型[4] 等具有一定代表性的传播模型, 以帮助更好地分析蠕虫类恶意代码的传播过程。

* 收稿日期: 2010- 02- 26
基金项目: 国家 863 计划资助项目(2008AA017414)
作者简介: 刘波(1973—), 男, 副研究员, 博士生。

1.1 简单传染病模型 SEM(Simple Epidemic Model)

在简单传染病模型 SEM 中,每台脆弱性主机可能分别处于易被感染(Susceptible)和已被感染(Infected)两种不同的状态,并且随着蠕虫的传播,被感染的脆弱性主机在从 Susceptible 状态转变为 Infected 状态后将始终保持在 Infected 状态[1]。如果以 N 表示在蠕虫开始传播之前网络中脆弱性主机的总数,$I(t)$ 表示在某个时刻 t 网络中已经被蠕虫感染,处于 Infected 状态的脆弱性主机数量,那么 SEM 模型可以表示为式(1)所示的微分方程,其中 β 被称为双向感染率,代表了网络中已经被感染的脆弱性主机对易被感染的脆弱性主机的感染强度或者感染能力。

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \beta I(t)[N - I(t)] \tag{1}$$

1.2 随机传播模型 RCS(Random Constant Spread)

根据对 Code Red 蠕虫爆发后采集到的监测数据所进行的分析,Staniford 等提出了随机传播模型 RCS,用于描述随机扫描蠕虫在网络中的传播过程[2]:每个蠕虫实例将完全随机地选取潜在的脆弱性主机作为下一步的攻击目标;同一台脆弱性主机在蠕虫传播过程中不会被多个蠕虫实例重复感染;Infected 状态的脆弱性主机在传播过程中将始终处于 Infected 状态。

在RCS 模型中,每个蠕虫实例的平均感染能力以常数 K 来表示,即在每个单元时间内,每台 Infected 状态的脆弱性主机将不重复地感染其他 K 个脆弱性主机。如果以 $\alpha(t)$ 表示在某个时刻 t 网络中 Infected 状态的脆弱性主机在全部脆弱性主机中所占的比例,那么 RCS 模型可以用微分方程(2)表示为:

$$\frac{\mathrm{d}\alpha(t)}{\mathrm{d}t} = K\alpha(t)[1 - \alpha(t)] \tag{2}$$

1.3 基于传染病模型的其它蠕虫传播模型

以SEM 模型为基础,并对网络中的脆弱性主机可能由于修补了相应的漏洞,或者处于 Infected 状态的脆弱性主机由于清除了所感染的蠕虫实例而发生状态转换的情形加以考虑,又有 KM(Kermack-Mckendrik)模型、双因素(Two-Factor)模型等被用于研究蠕虫类恶意代码的传播过程。

KM 模型假设处于 Infected 状态的部分被感染主机在清除了感染的蠕虫实例后,将转变到 Removed 状态并对该蠕虫具有免疫能力,不会再被该蠕虫的其它实例所感染,即脆弱性主机可能的状态转换为"Susceptible→Infected→Removed"[3],因此KM 模型也称为 SIR 模型。

如果在SEM 模型的基础上再以 $R(t)$ 表示某个时刻 t 网络中处于Removed 状态的主机数量,即曾经处于 Infected 状态但是在时刻 t 之前清除了所感染的蠕虫实例后对该蠕虫免疫的主机数量,$J(t)$ 表示在时刻 t 网络中所有曾经被该蠕虫感染过的主机数量,$S(t)$ 表示在时刻 t 网络中仍然处于Susceptible 状态的脆弱性主机数量,β 表示蠕虫的感染强度,γ 表示被感染的脆弱性主机由 Infected 状态转变为 Removed 状态的速率,则 KM 模型可以表示为:

$$\begin{cases} \dfrac{\mathrm{d}J(t)}{\mathrm{d}t} = \beta I(t)[N - J(t)] \\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t} = \gamma I(t) \\ J(t) = I(t) + R(t) = N - S(t) \end{cases} \tag{3}$$

进一步地,Zou 等在研究中提出,不仅仅是处于 Infected 状态的脆弱性主机有可能转为 Removed 状态,Susceptible 状态的脆弱性主机也有可能在修补了漏洞后对该蠕虫免疫;同时,在传播过程中由于蠕虫的大量繁殖所导致网络的拥塞,也会影响到蠕虫的传播速度,其感染强度 β 将随着时间 t 的推移而逐渐递减[4]。因此,如果以 $\beta(t)$ 表示蠕虫在某时刻 t 的感染强度,$Q(t)$ 表示在时刻 t 网络中曾经处于 Susceptible 状态但是现已对该蠕虫免疫的主机数量,那么双因素模型可以表示为:

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \beta(t)I(t)[N - R(t) - I(t) - Q(t)] - \frac{\mathrm{d}R(t)}{\mathrm{d}t} \tag{4}$$

2　影响蠕虫传播过程的时间因素

可以看出, 虽然 SEM 模型、RCS 模型, 以及经过扩展的 KM 模型或者双因素模型等考虑到的可能影响蠕虫传播速度的部分潜在因素有所不同, 模型的具体表述也有所不同, 但都是以经典的传染病模型为基础, 在相近的前提假设下, 使用相似的数学分析方法推导出的时间连续的确定性模型。

但是, Chen 等在对采用随机扫描策略的 Code Red 蠕虫进行分析时提出, 以经典的传染病模型为基础所开展的蠕虫传播模型研究可能忽略了某些重要的时间因素[5]: 蠕虫的每个实例感染下一个处于 Susceptible 状态的脆弱性主机是需要时间来完成的。当某个原本处于 Susceptible 状态的脆弱性主机被蠕虫的某个实例完全感染之前, 这个脆弱性主机不会转变到 Infected 状态, 亦即没有传染性; 而传染病模型使用时间连续的微分方程来估算网络中被感染的脆弱性主机数量的动态增长, 这就意味着在某个时刻 t, 有若干"被部分感染"的脆弱性主机参与了对其它处于 Susceptible 状态的脆弱性主机的感染, 这就与实际的蠕虫传播过程存在着差异。因此, 在分析推导蠕虫传播模型时, 应该合理地反映出这个时间因素可能对蠕虫的传播过程所产生的影响。

3　SSI 蠕虫传播模型

由于病理学领域在对多种传染性疾病的传播过程进行建模时的研究也表明, 即使是时间连续的确定性分析方法, 同样可以很好地对大规模系统的动态特性变化进行建模[6], 因此下文在开展蠕虫传播模型的研究时将继续参考和借鉴经典的传染病模型所使用的时间连续的确定性分析方法及其部分前提假设, 重在分析蠕虫感染某个脆弱性主机并使其具有感染能力需要相应的时间延迟这个特性, 更准确地描述蠕虫在网络中的渗透传播过程。

3.1　蠕虫的感染时间与两阶段的感染过程

以TCP 扫描蠕虫为例, 蠕虫实例感染某个具体目标的过程可以进一步细分成两个阶段, 如图 1(a) 和图 1(b) 所示: 第一阶段, 蠕虫实例经过三次握手建立与目标的TCP 连接, 利用该连接发送漏洞利用代码对目标实施攻击; 第二阶段, 在成功地对目标实施攻击之后, 蠕虫实例将其副本通过不同的分发机制复制到目标。在这两个阶段完成之后, 被感染的目标才具有传染性。从感染源的视角, 第二阶段蠕虫的副本分发机制有 Push 和 Pull 两种方式, 分别与 Weaver 总结的 Self-Carried 和 Second Channel 分发机制[7]基本对应。

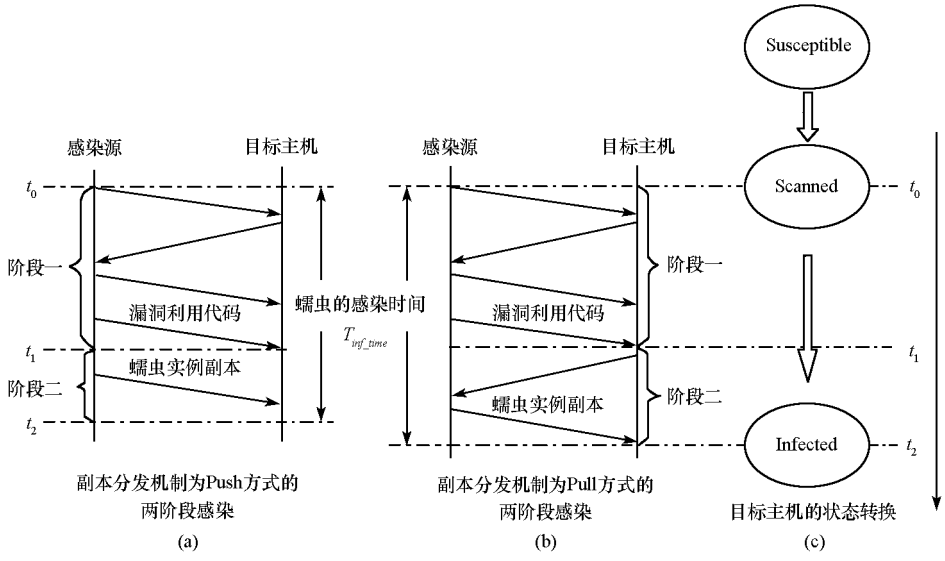

图 1　两阶段的脆弱性主机感染过程与脆弱性主机的状态转换

Fig. 1　Two-stage infection and state transition of vulnerable hosts

定义 蠕虫的感染时间: 每个蠕虫实例对目标完成两阶段感染过程所消耗的时间, 即从蠕虫实例发起对目标的攻击到被攻击的目标转变成具有传染性的 Infected 状态之间的时间间隔, 被称为蠕虫的感染时间, 记为 T_{inf_time}。

例如图1所示的TCP 扫描蠕虫, 如果蠕虫实例在时刻 t_0 开始启动对目标的攻击, 在时刻 t_1 完成对目标的攻击, 在时刻 t_2 完成蠕虫副本向目标的复制, 使得目标可以对网络中其它潜在的攻击目标重复上述的攻击过程, 那么蠕虫的感染时间 $T_{inf\_time} = t_2 - t_0$。

3.2 脆弱性主机的状态转换

如图 1(c) 所示, 当引入了两阶段感染和感染时间的概念之后, 网络中的脆弱性主机在蠕虫的传播过程中将可能分别处于易被感染(Susceptible)、已被扫描(Scanned)和已被感染(Infected)三种状态之一, 而可能的状态转换过程为"Susceptible → Scanned → Infected"。因此, 下文在分析这种状态转换过程的基础上推导出的蠕虫传播模型将被称为 SSI 传播模型。

3.3 SSI 模型

表1 SSI 模型中使用的符号及其含义

Tab.1 Notations used to derive SSI model

符号	符号的含义
Ω	蠕虫在传播过程中需要扫描的网络地址空间大小
$\mathcal{N}$	蠕虫开始传播前网络中由于存在漏洞而可能被感染的脆弱性主机的总数
η	平均扫描速率, 即在单位时间内, 每个蠕虫实例平均发出的扫描次数
τ	平均感染时间, 即每个蠕虫实例完成两阶段感染过程平均花费的时间
$S(t)$	在时刻 t, 网络中所有被蠕虫扫描过的脆弱性主机的数量
$I(t)$	在时刻 t, 网络中被蠕虫所感染的脆弱性主机的数量
I_0	初始状态下, 网络中处于 Infected 状态的脆弱性主机的数量

由于蠕虫在传播过程中所发出的每次扫描都将以$\frac{1}{\Omega}$的概率命中某个具体的 Susceptible 状态的脆弱性主机, 使其转换到 Scanned 状态, 因此在某个极小的时间间隔 δ 内, 如果以 p 表示每个蠕虫实例所发出的 $\eta\delta$ 次扫描至少命中网络中一个处于 Susceptible 状态的脆弱性主机的概率, 那么当 $\Omega \gg 1$ 且 $\Omega \gg \eta\delta$ 时, 有 $p = 1 - \left(1 - \frac{1}{\Omega}\right)^{\eta\delta} \approx 1 - e^{-\frac{\eta\delta}{\Omega}} \approx \frac{\eta\delta}{\Omega}$。

在某时刻 t, 网络中 Susceptible 状态的脆弱性主机数量为 $\mathcal{N} - S(t)$。当 δ 足够小时, 在 t 到 $t+\delta$ 的时间间隔内, 可以忽略每个蠕虫实例发出的 $\eta\delta$ 次扫描中有重复命中某个 Susceptible 状态的脆弱性主机的情形; 并且由于两个不同的蠕虫实例在时间间隔 δ 内发出的扫描同时命中某个潜在目标的概率是 δ 的2阶函数[8], 因此当 δ 足够小时, 也可以将不同的蠕虫实例在时间间隔 δ 内发出的扫描之间出现碰撞的情形忽略不计。如果暂不考虑在这个极小的时间间隔内由 Scanned 状态转换为 Infected 状态的脆弱性主机数量, 那么由于蠕虫的扫描而从 Susceptible 状态转换为 Scanned 状态的脆弱性主机数量为 $[\mathcal{N} - S(t)]p$。因此, 在时刻 $t+\delta$, 网络中所有曾经处于 Scanned 状态的脆弱性主机数量 $S(t+\delta) = S(t) + I(t)[\mathcal{N} - S(t)]p$。

由于某个处于 Scanned 状态的脆弱性主机需要经过感染时间 τ 之后才能完全转换为具有感染性的 Infected 状态, 因此当 $t \geqslant \tau$ 时, 网络中 Infected 状态的脆弱性主机数量 $I(t)$ 将等于时刻 $t-\tau$ 时, 网络中曾经处于 Scanned 状态的脆弱性主机数量, 即 $I(t) = S(t-\tau)$。因此有

$$S(t+\delta) = S(t) + S(t-\tau)[\mathcal{N} - S(t)]\frac{\eta\delta}{\Omega} \tag{5}$$

取 $\delta \to 0$, 那么对于 $\forall t \geqslant \tau$, 可以得出:

$$\frac{dS(t)}{dt} = \frac{\eta}{\Omega}S(t-\tau)[\mathcal{N} - S(t)] \tag{6}$$

另一方面, 由于初始状态下网络中处于 Infected 状态的脆弱性主机数量为 I_0, 那么在时刻 $t(0< t< \tau)$, 网络中将有 $I_0(\mathcal{N}- I_0)\left[1-\left(1-\frac{1}{\Omega}\right)^{\eta t}\right]$ 个 Susceptible 状态的脆弱性主机在被蠕虫扫描后转换到 Scanned 状态。由于 $\mathcal{N}\gg I_0$, $\Omega\gg 1$ 且 $\Omega\gg\eta t$, 因此对于 $\forall t< \tau$, 网络中所有处于 Scanned 状态的脆弱性主机数量为:

$$S(t)= I_0(\mathcal{N}- I_0)\left[1-\left(1-\frac{1}{\Omega}\right)^{\eta t}\right]\approx I_0\mathcal{N}\frac{\eta t}{\Omega} \tag{7}$$

同时, 由于当 $t< \tau$ 时, 蠕虫尚未完成对这些 Scanned 状态的脆弱性主机所实施的两阶段感染, 因此网络中不会新增加任何由 Scanned 状态转换为 Infected 状态的脆弱性主机。故此, 对于 $\forall t< \tau$, 有 $I(t)= I_0$。

综合上述分析过程, 并且借用 SEM 模型对于双向感染强度 β 的定义, 记 $\beta= \eta/\Omega$, 那么 SSI 模型可以综合表示为:

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}= \beta S(t-\tau)[\mathcal{N}- S(t)]\\ I(t)= S(t-\tau), \quad t\geqslant\tau\\ S(t)= \beta I_0\mathcal{N}t\\ I(t)= I_0, \quad 0< t< \tau\end{cases} \tag{8}$$

3.4　模拟仿真

以 Code Red 蠕虫为例, 在蠕虫的传播过程中大约有 359 000 台脆弱性主机被感染, CAIDA 的 Moore 等根据大规模网络监测的数据给出了在此期间被感染的脆弱性主机数量随时间而动态增长的曲线, 如图 2(a) 所示[9]。同时, 根据 Zou 等提供的分析数据, 每个 Code Red 蠕虫的实例平均每分钟发出大约 358 次扫描[8]。因此, 可以取 $\mathcal{N}= 359\ 000$, $\eta= 6$, $I_0= 1$, 分别对 SEM 模型和 SSI 模型进行模拟仿真, 结果如图 2(b) 所示。其中, 在对 SSI 模型进行模拟时取 $\tau= 5$。通过图 2 可以看出, SSI 模型可以相对比较准确地描述 Code Red 蠕虫在网络中的传播过程。

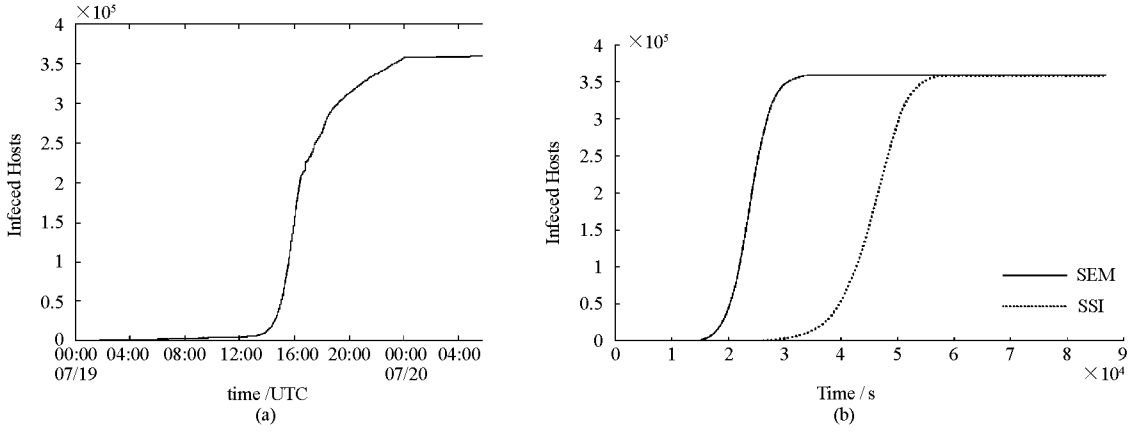

图 2　Code Red 蠕虫的传播过程

Fig. 2　Propagation of Code Red worm

4　结束语

SSI 模型使用时间连续的分析方法, 在简单传染病模型的基础上通过分析感染时间这个重要因素, 相对合理地反映出了蠕虫类恶意代码对脆弱性主机的两阶段感染过程, 有助于更加准确地理解蠕虫类恶意代码在网络中的渗透传播过程。

(下转第 108 页)

参考文献:

[1] Schmidl T M, Cox D C. Robust Frequency and Timing Synchronization for OFDM [J]. IEEE Transactions on Communications, 1997, 45(12): 1613-1621.

[2] 严春林. 正交频分复用系统中的同步技术研究[D]. 成都: 电子科技大学, 2004: 39-52.

[3] 李加生, 尹锁柱, 邓茜. 基于IEEE802.11a的OFDM同步算法及FPGA实现[J]. 计算机工程与科学, 2009, 31(7): 117-119.

[4] Zhang Z, Kayama H, Tellambura C. Joint Frame Synchronization and Carrier Frequency Offset Estimation in Multicarrier Systems [C]// IEEE Globecom, 2006: 1-6.

[5] Guo Y, Liu G, Ge J. A New Time and Frequency Synchronization Scheme for OFDM Systems [J]. IEEE Transactions on Consumer Electronics, 2008, 54(2): 321-325.

[6] Park B, Cheon H, Kang C, et al. A Novel Timing Estimation Method for OFDM Systems[J]. IEEE Communications Letters, 2003, 7(5): 239-241.

[7] Kim J J, Noh J H, Chang K H. Robust Timing& Frequency Synchronization Techniques for OFDM-FDMA Systems [C]// IEEE Workshop on Signal Processing Systems Design and Implementation, 2005: 716-719.

[8] Wu M, Zhu W P. A Preamble-aided Symbol and Frequency Synchronization Scheme for OFDM Systems [C]// IEEE ISCAS, 2005: 2627-2630.

[9] Manusani S K, Kshetrimayum R S, Bhattacharjee R. Robust Time and Frequency Synchronization in OFDM Based 802.1a WLAN Systems [C]// Annual IEEE India Conference, 2006.

[10] Zhou E, Hou X, Zhang Z, et al. A Preamble Structure and Synchronization Method Based on Central-symmetric Sequence for OFDM Systems [C]// IEEE VTC, 2008: 1478-1482.

[11] IEEE Std. 802.16a-2003. Part 16: Air Interface for Fixed Broadband Wireless Access Systems [S].

[12] 田野, 谈振辉, 冯永新, 等. 基于共轭对称结构训练符号的OFDM同步算法性能分析[J]. 小型微型计算机系统, 2009, 30(6): 1240-1243.

[13] GSM 05.05. Digital Cellular Telecommunications System (Phase 2+); Radio Transmission and Reception [S].

(上接第93页)

参考文献:

[1] Daley D J, Gani J. Epidemic Modeling: An Introduction[M]. Cambridge, UK: Cambridge University Press, 1999.

[2] Staniford S, Paxson V, Weaver N. How to Own the Internet in Your Spare Time[C]// Proceedings of the 11th USENIX Security Symposium, 2002.

[3] Frauenthal J C. Mathematical Modeling in Epidemiology[M]. New York: Springer-verlag, 1980.

[4] Zou C C, Gong W, Towsley D. Code Red Worm Propagation Modeling and Analysis[C]//Proceedings of the 9th ACM Conference on Computer and Communication Security, 2002.

[5] Chen Z, Gao L, Kwiat K. Modeling the Spread of Active Worms[C]// Proceedings of IEEE INFOCOM 2003, 2003.

[6] Andersson H, Britton T. Stochastic Epidemic Models and Their Statistical Analysis[M]. New York: Springer-verlag, 2000.

[7] Weaver N, Paxson V, Staniford S, et al. A Taxonomy of Computer Worms[C]// Proceedings of ACM CCS Workshop on Rapid Malcode (WORM'03), 2003.

[8] Zou C C, Gong W, Towsley D. On the Performance of Internet Worm Scanning Strategies[J]. Journal of Performance Evaluation, 2006, 63(7): 700-723.

[9] Moore D, Shannon C, Brown J. Code-red: A Case Study on the Spread and Victims of an Internet Worm[C]// Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement, 2002.

(上接第102页)

参考文献:

[1] Wehner D. High Resolution Radar[M]. Artech House, 1994.

[2] 刘华林, 杨万麟. 基于QR分解的广义辨别分析用于雷达目标识别[J]. 红外与毫米波学报, 2007, 26(3): 205-208.

[3] Kyung K, Dong S, Hyo K. Efficient Radar Target Recognition Using the MUSIC Algorithm and Invariant Features [J]. IEEE Trans. on Antennas and Propagation, 2002, 50(3): 325-337.

[4] Currie N, Brown C. Principles and Applications of Millimeter-wave Radar[M]. Artech House, 1987.

[5] 沈吉, 向锦武, 祁载康, 等. 目标运动对步进频率毫米波雷达锥扫测角的影响研究[J]. 电子学报, 2004, 32(6): 987-989.

[6] 王桂丽, 李兴国. 频率步进和脉冲多普勒复合测速研究[J]. 红外与毫米波学报, 2008, 27(3): 190-192.

[7] 李文臣, 王雪松, 王国玉. 机动目标一维距离像的展宽与补偿分析[J]. 宇航学报, 2008, 29(4): 1364-1368.