第 38 卷 第 4 期
Vol.38 No.4

计 算 机 工 程
Computer Engineering

2012 年 2 月
February 2012

• 安全技术 •

文章编号：1000—3428(2012)04—0140—03 文献标识码：A 中图分类号：TP18

具有排列扫描特征的多态蠕虫传播模型

汪 洁，王建新，何小贤

(中南大学信息科学与工程学院，长沙 410083)

摘 要：多数蠕虫传播模型都是基于简单的随机扫描，蠕虫形态相对固定。为此，研究排列扫描技术，结合自然生物的取食繁殖规则，提出一种多态蠕虫动态传播的数学模型。通过一系列相互独立的方程表现蠕虫的整体行为，计算传播过程中各类被感染蠕虫的数目。仿真实验结果表明，该模型能准确描述多态蠕虫的传播过程。

关键词：排列扫描；多态蠕虫；传播模型；多态变形技术

Propagation Model of Polymorphic Worm with Permutation Scanning Characteristic

WANG Jie, WANG Jian-xin, HE Xiao-xian

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

【Abstract】Most of models are based on single random scanning, and worm has relatively fixed morphology. In view of this problem, this paper researches on permutation scanning technique, combines natural biological feeding and breeding rules, and proposes a mathematical model to characterize the dynamic propagation of polymorphic worms. The overall behavior of the worm is presented and the number of different type of worms in the process of propagation is analyzed by a series of inter-dependent equations. Experimental simulation result shows that this model can describe polymorphic worm propagation exactly.

【Key words】permutation scanning; polymorphic worm; propagation model; polymorphic and metamorphic technology

DOI: 10.3969/j.issn.1000-3428.2012.04.045

1 概述

近年来，随着多态变形技术的出现，多态蠕虫病毒频繁爆发，对互联网的安全造成威胁。目前多数研究都是集中在检测和防御多态蠕虫的攻击[1]以及它们所使用的用来规避入侵检测系统(Intrusion Detection Systems, IDS)检测的技术[2]。为有效防御和对抗多态蠕虫，理解多态蠕虫的传播特性是关键因素。

多数传统的蠕虫建模研究都是集中在相对简单地随机扫描蠕虫或者是形态不发生改变的蠕虫[3]。但在蠕虫的传播过程中，许多随机的扫描是无效的。因此，为了能更快速地进行传播，出现了许多新的蠕虫扫描传播方式，如 hit-list 扫描和排列(permutation)扫描等[4]，这使蠕虫的传播更为迅速和隐蔽，针对这些扫描方式的建模方法较少。文献[5]对排列扫描蠕虫进行了精确的建模。该模型通过一系列内部独立的不同方程来求得所有被感染主机之间的交互，从而刻画出蠕虫的全部行为。然而，该模型仅对形态不发生变化的简单蠕虫进行了建模，未考虑多态蠕虫的情况。

针对多态蠕虫的传播，文献[6]设计了一个多态蠕虫传播模型，但该模型主要是针对蠕虫在随机扫描方式下的传播，而且建模时假设蠕虫扫描地址空间无限大，未考虑多态蠕虫在复杂扫描方式，如排列扫描方式下的传播。为更多地了解多态蠕虫的传播特性，本文基于生物方法提出一个数学模型，描述了结合排列扫描技术的多态蠕虫传播，并通过一系列方程计算传播过程中各类蠕虫的数目。

2 概念描述

排列扫描[4]通过加密将 IP 地址空间映射到一个地址连续的虚拟环，称其为排列环(permutation ring)。环中每一个初始感染的主机从它自己在环中的位置开始沿着排列环依次开始扫描。当它感染了一台新的主机时，它会继续扫描这台主机的下一个地址。一个新被感染的主机会随机地在环上选择一个位置开始连续的扫描。当一个扫描的主机 h_1 碰到一个被感染的主机 h_2 时，h_1 知道在环中排在 h_2 后的主机都已经被感染，所以，它将随机从环上重新选择一个地址开始扫描。当 h_1 第 k 次碰到被感染主机时停止继续进行扫描。本文仅考虑最简单的一种情况，即当 $k=1$ 时的排列扫描。复杂的情况将在以后进行更深入地研究。排列扫描类似于某些自然生物的取食-繁殖规则。排列扫描时一台主机不能进入另外一台主机的扫描空间，即每台主机均有自己的扫描空间，这类似于自然生物的取食过程中的独占性。

本文主要是对具有排列扫描特征的多态蠕虫传播过程进行建模。在某一多态蠕虫的所有样本中，将具有相同特征的蠕虫分为一类。这里的特征是指可以被 IDS 用来对蠕虫进行检测的特征。

首先对进行排列扫描的多态蠕虫的传播过程进行描述。

(1)环境

网络包含 n 台脆弱主机，通过加密将 n 台主机的 IP 地址映射到一个地址连续的排列环。n 台主机作为蠕虫的食物而存在。环境中包含一个蠕虫巢穴，巢穴中包含新生的蠕虫和

基金项目：中央高校基本科研业务费专项基金资助项目(2010122000 59)；中南大学自由探索计划基金资助项目(2011QNZT035)

作者简介：汪 洁(1980－)，女，讲师、博士、CCF 会员，主研方向：网络与信息安全；王建新，教授、博士；何小贤，讲师、博士

收稿日期：2011-09-05 **E-mail：**jwang@mail.csu.edu.cn

已经停止觅食的蠕虫。

(2)个体

环境中的个体为蠕虫。蠕虫包含的类型数目为 s。蠕虫有 2 种状态：active 和 retired。active 是指具有活动能力的蠕虫；retired 是指已经停止进食的蠕虫。蠕虫觅食后，并产生新的蠕虫的事件称为觅食成功事件。如果蠕虫觅食时，发现所处位置的食物是其他蠕虫觅食后的残骸，则称为觅食不成功事件。active 又分为 ineffective 蠕虫和 effective 蠕虫，ineffective 是指未来不再有存活能力的蠕虫；effective 是指未来有存活能力的蠕虫。effective 又包括 nascent 蠕虫和 non-nascent 蠕虫，nascent 是指新生的从巢穴中出来的蠕虫，并且未来具有存活能力；non-nascent 蠕虫是指 nascent 蠕虫完成了第 1 次进食，转变成为了 non-nascent 蠕虫。

(3)个体行为规则与相互作用规则

每个 active 蠕虫都有它的取食空间，取食空间覆盖了所有已经成为 active 蠕虫食物的主机所在的位置。由于蠕虫是沿着环依次进食的，因此，第 1 次进食的主机所在的位置为取食空间的上边界，目前正在进食的位置为取食空间的下边界。新生的蠕虫从巢穴中出来，在环境中随机选择一个位置开始觅食，如果它所在的位置属于其他蠕虫的取食空间，则该新生的蠕虫就成为了 ineffective 状态，并在开始进食时会发现食物已经被取走，所以，会进一步转变成为 retired 蠕虫。如果新生的蠕虫爬出来的位置不属于其他蠕虫的取食空间，则新生的蠕虫成为 effective 状态，即是 nascent 蠕虫。当 nascent 进食一次之后，会成为 non-nascent 蠕虫，并继续沿着环寻找食物。每进食一次后会繁衍一个新的蠕虫，并将其放回巢穴，该新的蠕虫以一个固定的概率 ρ 转化为另一个类型，以 $1-\rho$ 的概率保持不变。当 non-nascent 蠕虫沿着环寻找食物，进入其他蠕虫的取食空间时，会发现所处的位置的食物是其他蠕虫觅食后的残骸，则该蠕虫会转变为 retired 蠕虫，处于 retired 状态的蠕虫会回到巢穴。

(4)系统运行过程

在初始状态，巢穴中有少量新生蠕虫，其中各个类型的数量相同。此时，新生的蠕虫均会转变为 nascent 蠕虫。其他状态的蠕虫的数量均为 0。新生的蠕虫随机的选择环中的任意位置，由于此时所有位置都可以发生进食成功事件，因此这些蠕虫都属于 effective 状态，成为 nascent 蠕虫。它们在第 1 次进食后，nascent 蠕虫会转变成为 non-active 蠕虫，同时繁衍出新的属于某一类型的蠕虫并放入巢穴中。所有属于 non-active 蠕虫均会沿着环的下一个位置继续觅食，觅食的同时会繁衍出新的蠕虫并放入巢穴中。当它们进入其他蠕虫的取食空间时，会发现所处的位置食物已经只剩下残骸，则它们将转变为 retired 蠕虫回到巢穴。

3 具有排列扫描特征的多态蠕虫传播建模

3.1 数学模型

对蠕虫进行建模主要是获得环境中的食物残骸数量，以及 active 状态的各类蠕虫和处于 retired 状态的各类蠕虫随时间的变化情况。如果蠕虫不觅食，则所有的数量均不会发生变化，因此，蠕虫建模是建立在蠕虫觅食的基础上。蠕虫觅食会产生进食成功事件或进食失败事件，则环境中食物的数量、active 蠕虫的数量和 retired 蠕虫的数量会随之发生改变。

假设环境中总的生物数量为 N，其中，V 代表其中可作为食物的生物总数(即环境中脆弱主机的总数)；r 是蠕虫觅食的速度，即在单位时间内觅食的空间；v 是初始状态巢穴中新生蠕虫的数目。由于蠕虫每进食一次会产生一个蠕虫，因此蠕虫的数目和被变成残骸的食物数量是相同的。使用 $u(t)$, $i(t)$, $a(t)$, $r(t)$, $x(t)$, $y(t)$, $\alpha(t)$ 分别表示剩下食物的数量、已被吃掉食物的数量、active 蠕虫的数量、retired 蠕虫的数量、effective 蠕虫的数量、ineffective 蠕虫的数量、nascent 蠕虫的数量与食物总量的比值，可以得到：

$$u(t)+i(t)=1\ ,\ i(t)=a(t)+s(t)\ ,\ a(t)=x(t)+y(t)$$

此外，用 f_{hit} 表示蠕虫能够寻觅到的食物数量。食物均匀的分布在排列的地址空间中，所以，环中每一个位置都有 $\frac{V}{N}$ 的概率包含食物。在 dt 时间内，活动主机觅食的空间共有 $r\times dt$ 个地址。因此，$f_{\text{hit}}=r\times dt\times\frac{V}{N}$，此处食物的数量为食物和食物残骸的数量之和。

当一个处于 effective 状态的蠕虫觅食时，用 $f_{\text{new}}(t)$、$f_{\text{old}}(t)$ 分别表示进食成功的概率和进食失败的概率。在 t 时刻，剩下的食物数量为 $V(1-i(t))$。位于蠕虫取食空间下边界的食物残骸数量为 $V(x(t)-\alpha(t))$。所以，effective 进食成功的概率为：

$$f_{\text{new}}(t)=\frac{V(1-i(t))}{V(1-i(t))+V(x(t)-\alpha(t))}=\frac{(1-i(t))}{(1-i(t))+(x(t)-\alpha(t))}$$

进食失败的概率为：

$$f_{\text{old}}(t)=1-f_{\text{new}}(t)=\frac{(x(t)-\alpha(t))}{(1-i(t))+(x(t)-\alpha(t))}$$

当新生的蠕虫从巢穴中出来时，转变为 effective 蠕虫的概率用 $f_{\text{eff}}(t)$ 表示，转变为 ineffective 蠕虫的概率用 $f_{\text{ineff}}(t)$ 表示，$f_{\text{ineff}}(t)=i(t)$, $f_{\text{eff}}(t)=1-f_{\text{ineff}}(t)$。

3.2 模型描述

推断 $i(t)$、$x(t)$、$\alpha(t)$、$y(t)$ 和 $r(t)$ 随着时间改变的变化情况，用 $di(t)$、$dx(t)$、$d\alpha(t)$、$dy(t)$ 和 $dr(t)$ 表示在时间 t 后的 dt 时间内这些变量的变化，得到一系列方程，来描述蠕虫的传播模型。

(1) $di(t)$ 表示在 dt 时间内被进食掉的食物总数。

$$di(t)=x(t)f_{\text{hit}}f_{\text{new}}(t)$$

(2) $dx(t)$ 表示 dt 时间内 effective 蠕虫数目的变化。

$$dx(t)=x(t)f_{\text{hit}}f_{\text{new}}(t)f_{\text{eff}}(t)-x(t)f_{\text{hit}}f_{\text{old}}(t)$$

分析在增加的 effective 蠕虫中，各类蠕虫所占的比例。用 $x_1(t)$, $x_2(t)$,…, $x_s(t)$ 分别表示第 1 类，第 2 类，…，第 s 类 effective 蠕虫所占的比例，该数目是所有被各代蠕虫感染的主机数与整个环境中脆弱主机数目的比。每 i 类蠕虫进食之后繁殖出第 i 类蠕虫的概率为 $1-\rho$，繁殖出第 j 类蠕虫的概率为 $\frac{\rho}{s-1}$。所以，在 dt 时间内新产生的第 i 类蠕虫从巢穴中出来转变为 effective 的比例为：

$$x_i(t)f_{\text{hit}}f_{\text{new}}(t)(1-\rho)f_{\text{eff}}(t)+\sum_{l\in[1,s],l\neq i}x_l(t)f_{\text{hit}}f_{\text{new}}(t)\rho f_{\text{eff}}(t)$$

同理，退休的第 i 类 effective 蠕虫比例为 $x_i(t)f_{\text{hit}}f_{\text{old}}(t)$。因此，

$$dx_i(t)=x_i(t)f_{\text{hit}}f_{\text{new}}(t)(1-\rho)f_{\text{eff}}(t)+\sum_{l\in[1,s],l\neq i}x_l(t)f_{\text{hit}}f_{\text{new}}(t)\rho f_{\text{eff}}(t)-x_i(t)f_{\text{hit}}f_{\text{old}}(t)$$

(3) $d\alpha(t)$ 表示 nascent 蠕虫在 dt 时间内数目的变化。

$$d\alpha(t)=x(t)f_{\text{hit}}f_{\text{new}}(t)f_{\text{eff}}(t)-\alpha(t)f_{\text{hit}}$$

在所有 nascent 中，各类 nascent 蠕虫的数目分别用 $\alpha_1(t)$,

$\alpha_2(t),\cdots,\alpha_s(t)$ 表示。由于每 i 类 effective 蠕虫进食后繁殖出第 i 类蠕虫的概率为 $1-\rho$，繁殖出第 j 类蠕虫的概率为 $\frac{\rho}{s-1}$，因此，

$$d\alpha_i(t)=x_i(t)f_{\text{hit}}f_{\text{new}}(t)(1-\rho)f_{\text{eff}}(t)+\sum_{l\in[1,s],l\neq i}x_l(t)f_{\text{hit}}f_{\text{new}}(t)\rho f_{\text{eff}}(t)-\alpha_i(t)f_{\text{hit}}$$

(4) $dy(t)$：当一个新生的蠕虫从巢穴中出来时，如果进入了其他蠕虫的取食空间，则它就成为了 ineffective 蠕虫。在 dt 时间内，新生蠕虫的数量为 $x(t)f_{\text{hit}}f_{\text{new}}(t)$，这些新生蠕虫成为 ineffective 蠕虫的概率为 f_{ineff}，所以新增加的 ineffective 蠕虫为 $x(t)f_{\text{hit}}f_{\text{new}}(t)f_{\text{ineff}}$。当 ineffective 蠕虫发生进食失败事件时，它就会转变为 retired 蠕虫。因为 ineffective 蠕虫处于其他蠕虫的取食空间，所以当它进食时，必然发生进食失败事件，而环境中蠕虫能够找到的食物在整个空间中的比例为 f_{hit}，则 ineffective 蠕虫发生进食失败事件的概率为 f_{hit}，因此，

$$dy(t)=x(t)f_{\text{hit}}f_{\text{new}}(t)f_{\text{ineff}}-y(t)f_{\text{hit}}$$

用 $y_1(t),y_2(t),\cdots,y_s(t)$ 分别表示各类 ineffective 蠕虫的数量，同理可得：

$$dy_i(t)=x_i(t)f_{\text{hit}}f_{\text{new}}(t)(1-\rho)f_{\text{ineff}}(t)+\sum_{l\in[1,s],l\neq i}x_l(t)f_{\text{hit}}f_{\text{new}}(t)\rho f_{\text{ineff}}(t)-y_i(t)f_{\text{hit}}$$

(5) $dr(t)$：$r_1(t),r_2(t),\cdots,r_s(t)$ 表示各类 retired 蠕虫数目。

$$dr_i(t)=x_i(t)f_{\text{hit}}f_{\text{old}}(t)-y_i(t)f_{\text{hit}}$$

4 仿真实验与分析

仿真实验设置网络是全连接的，即新生的蠕虫随机的从巢穴往外寻找食物，而且寻找食物所花的时间是相同的。每一个单位时间内，蠕虫进食 1 次。第 i 类 effective 蠕虫进食之后繁殖出第 i 类蠕虫的概率为 $1-\rho$，繁殖出第 j 类蠕虫的概率为 $\frac{\rho}{s-1}$。蠕虫觅食速度为 r，蠕虫类型的数目为 S，环境中食物的总数，即脆弱主机的总数为 V。

在初始时刻，环境中仅存在一个第 0 代蠕虫。当 $\rho=0.52$，$r=1.0$，$S=4$，$V=20\,000$ 时，蠕虫数目随时间的变化如图 1 所示。

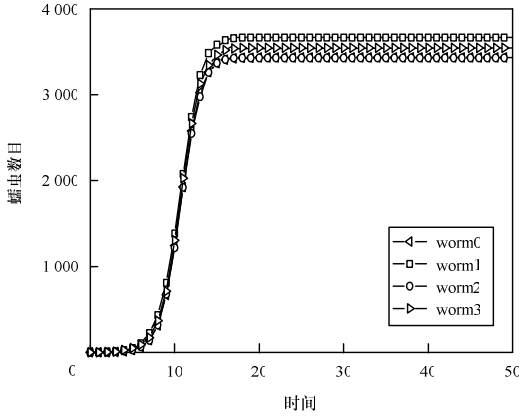

图 1 蠕虫数目随时间的变化

实验中蠕虫包含的类型数目 $S=4$。当初始时刻巢穴中的蠕虫数目为 m 时，可以首先设置各类蠕虫的数目均为 $\left\lfloor\frac{m}{S}\right\rfloor$。则剩下的 $m-S\times\left\lfloor\frac{m}{S}\right\rfloor$ 个蠕虫的数目可能为 $0,1,\cdots,S-1$。实验中剩下的蠕虫数目可能为 0,1,2,3。当这些蠕虫设置成某一类蠕虫时，会使某些类型的蠕虫数目比其他类型的蠕虫数目多 1。由于初始时刻蠕虫的数目与最终网络中被感染主机的数目无关，因此本文实验只考虑巢穴中蠕虫数目分别为 0,1,2,3 的 4 种情况。限于篇幅，本文给出第 1 种情况的实验结果，如图 2 所示，从中可见蠕虫繁殖过程中转变为另外一类蠕虫的概率 ρ 和最终环境中各类蠕虫数目之间的关系。

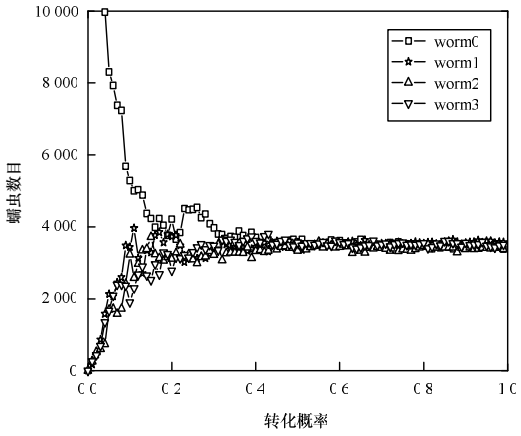

图 2 蠕虫数目随转化概率的变化

5 结束语

随着蠕虫技术的发展，出现了较为高级的传播方式，如排列扫描，同时出现了能改变蠕虫形态的多态变形技术，它们有能力使蠕虫进行更隐秘的扫描，并改变其形态以规避 IDS 的检测。本文在研究排列扫描技术的基础上，对具有排列扫描特征的多态蠕虫传播进行建模。模型应用生物方法对蠕虫的传播行为进行了动态刻画。下一步工作将研究如何快速检测和防御该类蠕虫的传播。

参考文献

[1] 汪 洁，王建新，陈建二. 基于彩色编码的多态蠕虫特征自动提取方法[J]. 软件学报, 2010, 21(10): 2599-2609.

[2] Perdisci R, Dagon D, Lee W, et al. Misleading Worm Signature Generators Using Deliberate Noise Injection[C]//Proc. of 2006 IEEE Symposium on Security and Privacy. Atlanta, USA: IEEE Press, 2006.

[3] 黄光球，刘秀平. 基于元胞自动机的网络蠕虫病毒传播仿真[J]. 计算机工程, 2009, 35(2): 167-169.

[4] Staniford S, Paxson V, Weaver N. How to Own the Internet in Your Spare Time[C]//Proc. of the 11th USENIX Security Symposium. San Francisco, USA: [s. n.], 2002.

[5] Manna P K, Chen Shigang, Ranka S. Inside the Permutation-scanning Worms: Propagation Modeling and Analysis[J]. IEEE/ACM Transactions on Networking, 2009, 18(3): 858-870.

[6] Stephenson B, Sikdar B. A Quasi-species Approach for Modeling the Dynamics of Polymorphic Worm[C]//Proc. of INFOCOM'06. Barcelona, Catalunya: IEEE Press, 2006.

编辑 金胡考