第32卷 第23期 计 算 机 工 程 2006年12月
Vol.32 No.23 Computer Engineering December 2006
·博士论文· 文章编号：1000—3428(2006)23—0001—03 文献标识码：A 中图分类号：TP393.08

基于无尺度易感应用网络的拓扑蠕虫传播模型

彭雪娜，赵 宏
(东北大学软件中心，沈阳 110004)

摘 要：分析了基于无尺度易感应用网络的拓扑蠕虫的传播特性，包括其感染整个应用网络所需要的传播时间和其在传播过程中对相关主机和网络资源的占用情况等。通过与扫描蠕虫相比较，分析出该类拓扑蠕虫传播时间更短，并且在传播过程中具有更好的隐蔽性，在实施最终攻击前很难被检测，从而使其对网络和主机具有更大威胁。针对这种威胁，文章提出了几种用于检测和防御基于无尺度网络应用拓扑蠕虫的可能方法。

关键词：拓扑蠕虫；传播模型；无尺度网络；蜜罐；网络连通性

Topology Worm Propagation Model Based on Scale-free Susceptible Application Network

PENG Xuena, ZHAO Hong
(Software Center of Northeast University, Shenyang 110004)

【Abstract】 A scale-free susceptible application network based topology worm propagation model is introduced, and its several features are analyzed, including its full infection time and resource occupation feature and etc. Compared with other worms, the worm is both faster and stealthier, and harder to be detected before final malicious attack. All these special features make it an even more risky worm type. According to its latent threat, several possible defense mechanisms are also introduced.

【Key words】 Topology worm; Propagation model; Scale-free network; Honeypot; Network connectivity

随着互联网的普及，网络应用的日益发展，网络攻击尤其是网络蠕虫已经成为人们面临的巨大威胁。

目前互联网上流行的蠕虫是扫描蠕虫，这类蠕虫的特点是通过网络扫描和探测获取易感节点信息，并对其进行感染。

随着蠕虫技术的发展，网络应用环境的日趋复杂，具有更强传播能力且对网络威胁更大的蠕虫将会不断涌现。本文针对扫描蠕虫的弱点，提出了一种基于无尺度易感应用网络的拓扑蠕虫的模型，此类蠕虫能够克服扫描蠕虫传播行为明显的弱点，实现快速而且轻量负载的传播，将会给对此类蠕虫的识别、拦截、控制等提出新的挑战。

1 相关工作

过去几年中，研究者在蠕虫传播模型方面的研究成果对本文产生了一定影响，如Nicholas Weaver，Vern Paxon所做的根据蠕虫生命周期中的不同行为阶段特征对蠕虫的分类研究[1]，Stuart Staniford等对扫描蠕虫传播模型进行的研究[2]。

其次，复杂网络领域的最新研究成果对本文也产生了一定的影响。这种成果主要是指A·L·Barabasi在 1999 年提出的无尺度网络模型[3]，这种无尺度网络模型突破了传统的E-R网络模型，使得人们对现实世界中存在的大量网络结构及其特性的认识发生了重大的改变。无尺度网络具有无尺度特性和小世界特性。这种特性使得网络中任意节点只需经过很少的几个中间节点就可以到达其它任意节点。无尺度网络特性给蠕虫传播带来极大便利。科学研究证实，目前包括e-mail、gnutella在内的很多应用网络都符合无尺度网络模型[4,5]。

研究发现，如果蠕虫在无尺度易感应用网络上利用网络拓扑信息进行传播，将会得到良好传播特性(传播时间短、传播过程负载低、隐蔽性强等)，这些特性会使对基于拓扑信息的蠕虫的防御和检测变得更为困难，从而给网络带来很大的潜在威胁。

2 基于无尺度易感应用网络的拓扑蠕虫传播模型

2.1 基本定义

定义 1 在判定某一网络节点能否被特定蠕虫感染时所遵循的标准称为该蠕虫的感染标准，记作C_{infect}。

定义 2 一个网络节点v是某蠕虫的易感节点v_s，当且仅当$sc(v, C_{infect}) == true$，其中$sc(v, C_{infect})$判断v是否满足相应$C_{infect}$。

定义 3 某蠕虫的易感集合$V_s= \{v \mid v.ip \in [0, 2^{32}] \text{ and } sc(v, C_{infect})==true\}$。

定义 4 任取特定蠕虫的易感节点v_s，如果存在应用信息E，能够揭示易感集合中其它节点的地址相关信息e，那么将易感集合中的节点按照应用信息E关联起来，就形成该蠕虫易感网络$G(V_s, E)$。

定义 5 如果在特定蠕虫的易感网络$G(V_s, E)$中，$\forall v_1, v_2 \in V_s$，都存在一条v_1到v_2的路径，那么就称该易感网络$G(V_s, E)$是连通易感网络。

定义 6 if $\exists G'(V_s', E')$，其中$V'_s \subseteq V_s, E' \subseteq E$，且$\forall v_1, v_2 \in V_s'$，都存在一条$v_1$到$v_2$的路径，那么称$G'(V_s', E')$是$G(V_s, E)$的连通易感子网，记作$G'(V_s', E') \xrightarrow{SubCon} G(V_s, E)$。

定义 7 如果$\exists$ $G'(V_s', E'), G'(V_s', E') \xrightarrow{SubCon} G(V_s,$

基金项目：国家信息安全管理中心资助项目(2001-研 2-A-005)
作者简介：彭雪娜(1979－)，女，博士生，主研方向：计算机网络安全，网络管理；赵 宏，教授、博导
收稿日期：2006-02-10 **E-mail：**pengxn@neusoft.com

E)，∀G“(V_s”, E”)，G“(V_s”, E”) $\xrightarrow{SubCon}$ G(V_s, E)能够推出G“(V_s”, E”) $\xrightarrow{SubCon}$ G’(V_s’, E’)，则称G’(V_s’, E’)为G(V_s, E)的最大连通易感子网，记作G’(V_s’, E’) $\xrightarrow{MaxSubCon}$ G(V_s, E)。

2.2 基于无尺度易感应用网络的拓扑蠕虫传播模型

AN(V, E)表示应用网络，G(V_s, E)表示无尺度易感应用网络，且G(V_s, E)是AN(V, E)的子网，设v_0是攻击者最初释放蠕虫的应用节点。算法 1 描述了蠕虫在任意节点v(包括v_0)上的传播模型，图 1 描述了该算法在特定网络中的传播实例。

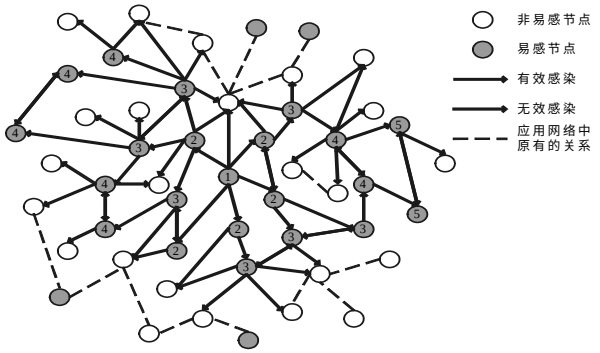

图 1 拓扑蠕虫在应用网络中的传播模型

算法 1 蠕虫传播模型

(1)蠕虫在节点 v 上搜集其所了解的应用网络上相邻节点的拓扑信息；

(2)蠕虫根据这些拓扑信息来感染应用网络上的相邻节点，如果相邻节点不是易感网络中的节点，那么放弃感染，如果相邻节点是易感网络中的节点且已经被感染，那么也放弃感染，否则，感染该节点，新感染节点重复步骤(1)；

(3)当蠕虫尝试向所有相邻节点进行感染以后，即停止传播，进入潜伏状态。

按照上面的传播模型，蠕虫的传播过程可以被抽象为树型结构。其中以初始节点v_0为这棵树的根，蠕虫传播过程中所感染的节点为树上的节点，每个树上节点与其第 1 感染源之间建立的关系E对应树中的亲子关系。这样构建起来的一棵树称之为蠕虫的传播树T(V_s’, E_0, v_0)，简记为T。当蠕虫传播结束时，该蠕虫传播树的节点集合与初始节点所在的最大连通易感子网的节点集合相同。如果易感网络本身就是连通易感网络，那么蠕虫传播树的节点集合与该易感网络的节点集合相同。

2.3 容易被拓扑蠕虫利用的应用网络

拓扑蠕虫的传播是依赖于应用网络中节点上应用拓扑信息的，因此保证在节点上就可以获取到其他节点的拓扑信息，是拓扑蠕虫传播的基础。拓扑蠕虫的传播过程，是由一个感染节点逐步向其所在的易感连通子网其它节点扩散的过程，因此该易感连通子网的网络规模决定了拓扑蠕虫的影响范围。另外如果易感连通网络是无尺度网络，能够表现出小世界特性，那么这种结构将为拓扑蠕虫的高性能传播提供有利条件。因此，易感节点本身具有易感网络的部分拓扑信息、易感网络是具有一定规模的连通网络且为无尺度网络时，拓扑蠕虫在其上是很容易传播的。而能够衍生出这种易感网络的应用网络都是容易被拓扑蠕虫所利用的。

在现实网络中，这样的应用网络比比皆是，如 e-mail 应用网络，各种即时通信应用网络、p2p 应用网络。在这些应用网络中，其节点都具有网络的部分拓扑信息，能够符合拓扑蠕虫传播的基本要求。在这样的应用网络上，如果有严重漏洞被发现，那么黑客就能够很容易地构造拓扑蠕虫，在其易感网络上进行传播。同时，由于目前这些应用网络是无尺度网络，且网络节点的异构性并不太强，有的甚至很弱，因此，基于此的易感应用网络很可能形成具有一定规模的无尺度网络，而这种网络所表现的小世界特性为拓扑蠕虫的传播提供了有利条件。

3 传播特性分析

3.1 拓扑蠕虫传播过程的时间分析

图 2 是一棵蠕虫传播树T，其中V_0是该传播树的根，即蠕虫传播的初始节点。该传播树中，用$p(c)$表示c的父节点，且$p(NULL)=NULL$，$p(V_0)=NULL$，则节点v的所在层次(辈分)为

$$level(v)=\begin{cases}1, v=v0\\ i+1, level(p(v))=i\end{cases}$$

对于节点c(level(c)=n)，如果取i∈N，存在 $\underbrace{p(...(p(p}_{i\text{个}p}(c)...)\neq NULL$，则称 $\underbrace{p(...(p(p}_{i\text{个}p}(c)...)$为c的第n-i代祖先，简记$p^i(c)$。当$i$=0 时，$p^0(c)$表示节点c本身。$p^{n-1}(c)$表示$v_{0}$。

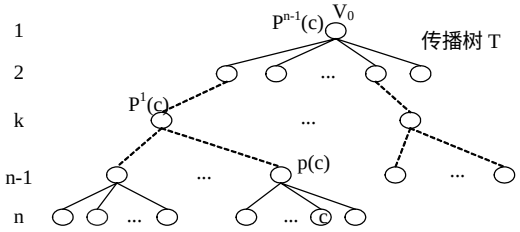

图 2 传播树

假定把第一代节点 c 首次感染的时间作为参照点——时刻 0，则第 n 代子孙 c 首次被感染的相对时间递规定义如下：

$$t_{[n\mid c]}=\begin{cases}t_{[(n-1)\mid p(c)]}+t_{info}+(I_{[(n-1)\mid p(c)],[n\mid c]}-1)*t_0, & \text{若} n>1\\ 0, & \text{若} n=1\end{cases} \quad (1)$$

其中，$t_{[n\mid c]}$表示第n层节点c被感染的相对时间，$t_{[(n-1)\mid p(c)]}$表示其父亲即第n-1 层节点$p(c)$被感染的相对时间，t_{info}表示被感染获取其他易感节点信息所需的时间，$I_{[(n-1)\mid p(c)],[n\mid c]}$表示第n层节点c是第n-1 层节点$p(c)$的第I个孩子，$t_0$表示在父亲$p(c)$上感染一个孩子所需要花的时间(简单起见，假设任何一个父节点感染其任一孩子的时间均相同)。

设$t_{max}(n)$ 表示传播树T中，最后一个第n层节点得到传播的时间，则

$$t_{\max}(n)=\max(\{t_{[n\mid c]}\mid isposterity(c,n)==true\}) \quad (2)$$

其中 $isposterity(c, n) = true$，如果 c 是为传播树 T 中的第 n 层节点，否则 $isposterity(c, n) = false$。

设$d(c)$表示任意节点c的出度，$d_{max}(n)$表示T树中第n层节点的最大出度，d_{max}表示整个传播网络G中节点的最大出度。结合式(1)、式(2)，不难看出：

$t_{max}(n)= max(\{t_{[n|c]}|isposterity(c, n)==true\})\leq t_{max}(n-1)+t_{info}+max(\{(d(n-1)-1)*t_0|isposterity(c, n)==true\})$

推导可得

$t_{max}(n)\leq n*d_{max}*t_0+(n-1)*(t_{info}-t_0)$，且$n\geq 2$ (3)

用N_T表示传播树T的最大层数，且$N_T\geq 2$。

带入式(3)，有$t_{max}(N_T)\leq N_T*d_{max}*t_0+(N_T-1)*(t_{info}-t_0)$

用D表示应用网络AN的直径，显然，$N_T<D$, 因此：

$t_{max}(N_T)\leq D*d_{max}*t_0+(D-1)*(t_{info}-t_0)$ (4)

由前面定义可知传播树$T(V_s', E_0, v_0)$的感染节点集合为$T(V_s', E_0, v_0)$的根v_0所在最大连通易感子网$G(V_s', E)$中的V_s'；当$G(V_s, E) == G(V_s', E)$时，$T(V_s', E_0, v_0)$的感染节点集合即为V_s，即易感网络$G(V_s, E)$中的全部节点都被蠕虫感染了。当$G(V_s, E) == G(V_s', E)$时，结合上面对$t_{max}(N_T)$的计算，可以得出拓扑蠕虫在目标易感网络上感染全部易感节点所需要的最长时间为$D * d_{max} * t_0 + (D-1) * (t_{info} - t_0)$；如果$d_{max} * t_0 << t_{info}$，则感染最长时间近似$(D-1) * t_{info}$；如果$d_{max} * t_0 >> t_{info}$，则感染最长时间近似$D * d_{max} * t_0$。由于无尺度网络具有小世界特性(即D值较小)，且对于任意给定网络，d_{max}有限，因此当$d_{max} * t_0 >> t_{info}$，且设置较短的单节点感染时间$t_0$，基于无尺度应用网络的拓扑蠕虫可以在很短的时间内完成传播。

按照此公式计算，假设某拓扑蠕虫中参数t_{info}=0.01s，感染一台机器的时间为 0.1s，且易感应用网络是一个具有 300 000 台主机的无尺度网络，其中d_{max} = 30，D = 10，那么由于这里$d_{max} * t_0 >> t_{info}$，因此该拓扑蠕虫将易感网络上的所有主机全部感染所需要的最长时间仅为 30s。

3.2 拓扑蠕虫传播过程的资源占用分析

3.2.1 主机资源占用分析

蠕虫传播过程中，主机资源占用情况主要表现为：(1)易感节点被感染时的资源占用；(2)被感染后，获取其它节点拓扑信息时的资源占用；(3)继续感染拓扑相邻已感节点的资源占用。按照被感节点上蠕虫所处生命期不同，蠕虫给所在主机带来的即时主机资源占用情况可以描述为

$$p_{ins\tan t} = \begin{cases} p_{\inf o} / t_{\inf o}, \text{蠕虫获取相关拓扑信息 期间} \\ L / t_0, \text{蠕虫传播期间} \end{cases}$$

其中，p_{info}表示获取其他节点拓扑信息所占用的资源总量，t_{info}表示所耗费的时间，L表示蠕虫样本的长度，t_0为传播一个蠕虫样本所花时间。

在现实网络中广泛传播蠕虫的大小一般在几KB到十几KB左右，比如冲击波蠕虫(I-worm/Blaster)样本大小为 6KB，红色代码蠕虫(CodeRed)样本大小为 4KB。基于这种蠕虫长度数量级，结合 3.1 节中所举的易感网络例子，设L=10KB，t_0=0.1s，则蠕虫传播期间即时主机网络资源占用仅为100Kbps，这在一台P4 2.5GHz的CPU，512MB内存的主机和100MBps接入的网络环境来说，其负载基本表现不出来。

3.2.2 网络资源占用分析

蠕虫传播过程中的网络资源占用可以简单理解为传播期间的带宽占用。对于特定物理网络(局域网络)而言，拓扑蠕虫在网络中传播造成的带宽影响 TA 可以如此计算：

$$TA = M * L / t_0$$

其中，M 为特定物理网络中当前正在产生流量的已感染节点个数。

假设特定物理网络中有M=100，蠕虫大小L=10KB，感染一个节点所需时间t_0=10s，则TA=100×10KB/10s= 100KBps，即 800Kbps。可见，在适当的蠕虫传播速度下，其对网络资源的占用很少。

3.3 拓扑蠕虫与扫描蠕虫的比较

结合 Stuart Staniford 等对扫描蠕虫的分析成果，将拓扑蠕虫与扫描蠕虫比较，可以发现以下特点：

(1)拓扑蠕虫的传播时间更短。按照 3.1 节中提出的拓扑蠕虫的传播时间特性的计算方法$D * d_{max} * t_0 + (D-1) * (t_{info} - t_0)$，对于 3.1 节中给出的蠕虫示例，计算其感染整个网络所需要时间应在 30s以内。而在文献[6]中，相同条件下扫描蠕虫所需的时间大约为几个小时。

(2)拓扑蠕虫传播时资源占用更少。通过 3.2 节的分析发现，拓扑蠕虫传播过程中所占用的主机资源和网络资源都很少，在主机和网络上的行为表现都不明显。

(3)拓扑蠕虫的监测和防御更为困难。由于拓扑蠕虫的传播时间和其占用的资源都更少，并且可能伪装成正常的应用流量，因此这类蠕虫传播时相对扫描蠕虫而言隐蔽性更强，更加难于监测和防御。

(4)拓扑蠕虫的潜在威胁更大。由于扫描蠕虫具有十分明显的网络流量特征，因此，虽然在其传播过程中能够造成很大的影响，但是容易被发现，对网络的威胁仅限于传播时即时产生的危害；与之相对，拓扑蠕虫由于自身的传播特性，使其具有长期潜伏在网络中的条件，进而给网络造成更深远的潜在威胁。

4 基于无尺度易感应用网络的拓扑蠕虫的防御机制

通过第 3 节的分析发现，基于无尺度应用网络的拓扑蠕虫传播特性与传统蠕虫有所不同，拓扑蠕虫传播时间更短，传播负载更不明显，不容易采用传统的方法进行监测和防御。因此，本文针对此类蠕虫的特点提出了几种检测/防御方法。

4.1 用蜜罐/蜜网进行检测－响应

这种方法的出发点在于希望通过对蠕虫进行及时、有效的监测和响应，把已经被蠕虫感染的网络所面临的可能损失降到最低点。在组织网络中部署蜜罐/蜜网，并将其与真实应用网络建立一定联系，可以起到吸引蠕虫的作用。通过对蜜罐/蜜网进行在线或周期性的离线强审计，可以有效识别潜伏的蠕虫。蜜罐/蜜网中覆盖了常用的网络应用和常用版本，当拓扑蠕虫传播时，蜜罐很有可能被感染，即有机会记录蠕虫传播、潜伏行为，因此通过蜜罐能够检测蠕虫的存在并了解其动态。同时对蜜罐采集到的蠕虫行为信息进行特征提取，可以发现其传播和潜伏特征。基于这些特征，通过网络安全管理及防御部件，就有可能对感染此类蠕虫的节点进行清除和治理，从而达到降低蠕虫风险的目的。

4.2 重点保护无尺度应用网络中的 HUB 节点

这种方法充分利用无尺度易感应用网络的结构特征，实现对拓扑蠕虫有效的监测与防御。无尺度网络中网络节点连接呈幂律分布，这意味着大多数节点具有较少连接，而只有少数节点拥有大量连接。对于基于无尺度易感应用网络的拓扑蠕虫传播模型而言，这意味着，蠕虫在无尺度网络中的部分随机节点上传播失败，不会影响到拓扑蠕虫的有效传播，而在拥有大量连接的几个核心节点上传播失败，则会严重影响到拓扑蠕虫的传播。因此，对具有大量连接的核心节点进行重点保护将有助于抑制拓扑蠕虫的传播。对于具有大量连接的核心节点实施的保护可以是多重的，具体包括系统加固、防火墙、入侵检测、一致性校验等。这些措施都可以有针对性地降低拓扑蠕虫可能带来的损失。

4.3 增大应用网络节点的异构程度

这种方法的出发点在于如何从根本上防止拓扑蠕虫在应用网络中的广泛传播，从而使应用网络免于遭受拓扑蠕虫可能造成的损失。通过对应用网络中节点的异构特征进行分析后发现，对于具有较好连通性的相同应用网络而言，网络节点的异构程度也将严重影响拓扑蠕虫的传播。以图 3 中的网络为例，如果网络中节点是高度同构的，即所有节点上安装相同厂商、相同版本号的软件发行版本，那么，一旦拓扑蠕虫在这样的网络上爆发，整个应用网络都将成为易感网络，考虑到网络自身具有较好的连通性，因此蠕虫在这样的同构网络中十分容易传播。与之相反，如果应用节点是高度异构的，即应用网络中，应用软件的发行厂商多样化，发行版本多样化，那么，当蠕虫在这样的应用网络上爆发时，易感网

（下转第 23 页）