具有免疫特征的僵尸网络传播模型

黄 彪[1] 谭 良[1,2]

(四川师范大学计算机学院 四川省可视化计算与虚拟现实重点实验室 成都 610068)[1]

(中国科学院计算技术研究所 北京 100080)[2]

摘 要 僵尸网络是一种从传统恶意代码进化而来的新型攻击方式,已成为Internet安全的一个重大威胁。僵尸网络传播模型是研究僵尸程序传播特性最常用的一种方法,当前僵尸网络的主流传播模型并没有考虑到部分主机的免疫特性,因而目前的这些主流传播模型对僵尸程序在Internet上的传播特性反应得不够准确。提出了一种新的具有免疫特征的僵尸网络传播模型。该模型基于Internet的实际情况,重点考虑了Internet中部分脆弱主机由于提前从易感染的网络中移除而具有免疫特征。仿真结果表明,基于免疫特征的僵尸网络传播模型更符合真实Interne网络中僵尸程序的传播规律和感染特性。

关键词 僵尸网络,僵尸程序,传播模型,免疫

中图法分类号 TP393 **文献标识码** A

New Botnet Propagation Model with Immunity

HUANG Biao[1] TAN Liang[1,2]

(College of Computer, Sichuan Normal University, Chengdu 610068, China)[1]

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)[2]

Abstract Botnet is a novel attack mode evolved from the traditional malicious softwares, currently it has become a major threat to Internet. The propagation model of botnet is the most common method to study the spreading features of bots, but current botnets' propagation models don't consider characteristics of some hosts' immunity, so these propagation models can not exactly describe how the bots spread on the Internet. In this paper, a new botnet propagation model with the immunity was proposed. This model carefully considerates about the real situation of the Internet, especially immunity of the host which removed from the susceptible network in advance. Simulation result shows that the botnet propagation model with immunity more exactly satisfies the practical propagation laws and infection characteristics of bot on Internet.

Keywords Botnet, Bot, Propagation model, Immunity

1 引言

僵尸网络是攻击者利用用户主机的安全漏洞以及社会工程学等方法传播僵尸程序,令其感染互联网上的大量主机,而这些被感染的主机通过一定的控制信道与攻击者建立连接并接收命令[1]。在僵尸网络初步建立以后,攻击者就可以控制受害主机在接收其命令的情况下继续传播僵尸程序,当受害主机的数量足够大时,攻击者就可以通过受害主机的网络资源,在互联网上建立一个攻击平台,获得想要的利益。

僵尸网络是近几年来网络上恶意软件最普遍的攻击方式,也是目前网络攻击中危害最大的攻击手段之一,攻击者通过控制僵尸网络,可以对网络以及其中的大量用户造成严重的安全威胁。自2005年以来,僵尸网络数量一直呈现上升趋势。CipherTrust公司[2,3]2005年通过研究发现,每天约有15万左右新僵尸程序出现,而在中国出现的僵尸程序就占了15%~20%。国内某安全实验室在2005年6月3日至9月19日之间平均每天发现约3万个僵尸网络客户端,尤其是在2005年8月19日至9月19日期间,他们通过观测和检验发现了一个僵尸主机高达15万的僵尸网络[4]。文献[5]提供了两组数据。第一组是CNCERT对部分木马和僵尸程序的抽样监测结果。数据显示,2009年中国境内被僵尸程序控制的主机IP数量为83.7万个,境外有1.9万个主机地址参与控制,来自美国的占22.34%,排名第一。第二组数据来自美国Symantec公司(全球最大的网络安全公司)2008年的因特网安全威胁报告。报告指出,网络攻击源的数量,美国居世界第一,占总量的25%;位于美国的僵尸控制服务器数量居世界首位,占世界总量的33%。与此相反,中国被植入僵尸程序的计算机数量居世界首位,占世界总量的13%。可以看出,僵尸网络在中国的威胁已经日益增大,中国已经成为世界上最大的僵尸网络受害国。目前的僵尸网络不再是一种单纯的网络攻击行为,而已经成为攻击者攻击的一个平台,在这个平台上,攻击者可以发动分布式拒绝服务攻击、发送垃圾邮件攻

本文受国家自然科学基金面上项目(60970113)资助。

黄 彪(1988—),男,硕士,主要研究方向为网络安全;谭 良(1973—),男,博士,教授,主要研究方向为信息安全、网络计算。

击、获取受害主机内的敏感信息或牟取经济利益等。

由于僵尸网络的普遍性以及其危险性，越来越多的专家致力于研究它的检测和防御[6]。检测和防御技术依赖于僵尸程序的传播特性，而僵尸网络传播模型是研究僵尸程序传播特性最常用的一种方法。因此，针对僵尸程序的传播特性建立其传播模型对检测和防御僵尸网络的攻击有重要意义。

目前，已有的僵尸网络传播模型均是继承自蠕虫的传播模型，但是与蠕虫的自动传播不同，僵尸程序在传播过程中是受控的[7]。通过对一些经典蠕虫程序的分析，研究其传播模型对于僵尸网络的构建与传播具有一定的参考价值。现今研究中，已经提出的网络蠕虫传播模型包括基本的 SEM 模型[8]、KM 模型[9]，以及考虑蠕虫反制机制的双因素模型[10]等。但是由于僵尸网络传播时受攻击者控制，蠕虫的传播模型并不适用于对僵尸网络发展趋势的预测，因此，Dagon 等人在文献[11]中提出了一个基于时区的僵尸网络传播模型，相较于蠕虫的传播，这种传播模型更适用于僵尸网络，但是该模型一方面对当前融合了多种传播形式的僵尸网络没有进行系统和全面的描述；另一方面并没有考虑到部分主机的免疫特性，因而对僵尸程序在 Internet 上的传播特性反应得不够准确。

本文结合蠕虫传播模型与 Dagon 提出的僵尸网络传播模型，提出一种具有免疫特征的传播模型。该模型重点考虑了僵尸网络传播过程中，人们通过检测手段预知到网络中可能存在恶意程序而进行预先免疫处理，令部分安全措施较弱的主机从危险网络中移除的情况。从易感染的网络中移除的主机即处于免疫状态。这种具有免疫特征的传播模型对于僵尸网络传播的刻画更加准确。

2 相关工作

目前，在僵尸网络传播模型上所进行的研究多数是基于蠕虫的传播模型。Streftaris 在文献[8]中阐述了网络蠕虫的传播和生物病毒的传播在自我复制和传播行为上有着相似的特点，同时提出了一种简单传染病模型—SEM 模型，这种模型的基本假设为：①蠕虫传播的网络环境为一个全连通的结构，蠕虫代码可以从被感染主机直接传送到目标主机；②网络中的每台主机有两种状态：易感染状态和被感染状态；③一旦一台主机被感染后，就始终保持被感染的状态。可是 SEM 模型太过于理想，对于蠕虫传播过程中受到的各种影响，如网络拥塞、主机从感染状态恢复到正常状态等都未进行考虑和分析，所以并不符合网络中蠕虫的传播。Frauenthal 等人在文献[9]中提出了 KM 模型。这种模型考虑了受害主机的 3 种情况，即易感染、被感染和免疫状态，但由于没有考虑受害主机的补丁更新和升级，因此依然不能很好地反应蠕虫传播的真实情况。文献[12]在传统的蠕虫传播模型 SEM 模型和 KM 模型的基础上，提出了一种双因素蠕虫传播模型，这种模型考虑了蠕虫传播过程中的两个因素：①蠕虫的传播过程中，存在人为的应对措施，那些特别脆弱的易感染主机会在未被感染前就从蠕虫传播环境中移除；②蠕虫传播过程中引起的网络拥塞。但僵尸程序不同于蠕虫，所以只能借鉴蠕虫的传播模型来研究僵尸网络的传播，而不能直接套用。

文献[7]介绍了僵尸网络的演化过程和基本定义，深入剖析了僵尸网络的功能结构与工作机制，讨论了僵尸网络的命令与控制机制和传播模型，并归纳总结了目前跟踪、检测和防御僵尸网络的最新研究成果。虽然对僵尸网络的功能结构和工作机制等方面进行了探讨，参考了蠕虫病毒的结构，进一步研究分析了当前主流僵尸程序，但是没有提出一种具体的僵尸网络传播模型。Lin Yong 在文献[13]中从僵尸网络的定义出发，探讨了僵尸网络的组成、分类和典型代表，分析了僵尸网络的工作过程，以及可能产生的危害，并提出相关问题的解决方法思路。在描述僵尸网络的分类以及其工作过程时，将僵尸网络的工作过程分为四个阶段：传播、加入、等待和控制。研究发现，有些僵尸程序具有自动扫描大规模网段，利用系统漏洞和弱口令等手法查找脆弱主机进行传播，传播方式与蠕虫相类似。因此研究蠕虫的传播对僵尸网络的研究有重要意义。Dagon 在文献[11]中提出了一种基于时区的僵尸网络传播模型，这种模型考虑了计算机在夜间下线，导致部分僵尸节点失效的情况，但是该模型只关注了远程攻击漏洞的传播方式，而当前的僵尸网络也融合了邮件病毒、即时通信和 P2P 文件共享软件等其他方式，对这些类型的僵尸网络还没有很好的刻画与描述。Wang 在文献[14]中提出了一种先进的混合型 P2P 僵尸网络设计框架，在这种框架中，有两种僵尸主机，一种是 servent 节点，它们拥有非私有的静态 IP 地址。一种是 client 节点，它们的特点是有动态分配的 IP 地址或者私有的 IP 地址隐藏在防火墙后，并且不能从 Internet 接入。在这种设计框架下，Wang 等人还进一步提出通过命令认证、节点对加密机制、个性化服务端口等机制保证僵尸网络的健壮性和韧性，但其对僵尸网络的传播模型并没有进行具体的分析和研究。

总之，目前这些模型虽然都对僵尸网络的传播进行了描述，但均没有较好考虑其免疫特征。要对僵尸网络进行更好的刻画，就需要从主机的免疫特征入手。

3 具有免疫特征的僵尸网络传播模型

3.1 免疫特征

僵尸程序是网络蠕虫的进一步发展，根据一些经典的蠕虫传播可以得出僵尸网络的传播模型。虽然 Dagon 等人[11]提出的僵尸网络传播模型能反应僵尸网络的上下线状态，并且也考虑了网络中主机从感染状态恢复到正常状态的情况，但是文献[12]中提到的双因素蠕虫传播模型中考虑的免疫状态是两种情况，而这两种情况都符合僵尸网络的传播特性。因此本文提出的免疫特征指：(1)部分被感染主机通过打补丁等措施从感染状态恢复到正常状态；(2)僵尸程序在网络中传播一段时间以后，人们会意识到该网络中存在恶意程序，而预先移除一些尚未被感染但是安全措施较低且有安全漏洞的主机。这种免疫特征不仅在蠕虫网络中存在，在僵尸网络的传播中同样存在。

3.2 传播模型

定义 $I(t)$ 是在 t 时刻被感染了的主机数，$S(t)$ 为在 t 时刻

易感染的主机数，N 表示网络中最初的易感染主机数。由于易感染主机数是随着时间 t 变化的[11]，在每个时间段的 N 数量都不同，N 随着 t 的增加和网络规模的增大而逐渐增加，因此定义 η 为每个时间段 N 的增长率，即 N 的数量为（N_0 为传播开始时的初始易感染主机数）：

$$N=(1+\eta)N_0 \tag{1}$$

考虑到受害主机上下线的状态[11]，定义 $I'(t)$ 为当前在线的感染主机数；$S'(t)$ 为当前在线的易感染主机数；$\alpha(t)$ 为 t 时刻的主机在线比率，一般情况是白天达到峰值，在深夜由于大部分计算机下线，所以此值到达低谷。因此在 t 时刻已经被感染和易感染主机数量分别为：

$$I'(t)=\alpha(t)I(t) \tag{2}$$

$$S'(t)=\alpha(t)S(t) \tag{3}$$

为了描述部分主机从感染状态变为免疫状态而从网络中移除的情况，根据最基本的 Kermack-Mckendrick 传播模型[9]，假设被感染主机在持续发送蠕虫数据包一段时间后，可能被关闭或者计算资源被耗尽，无法继续发送蠕虫数据包，或者被感染主机在被感染一段时间后，经过用户处理，从感染状态恢复，并对此蠕虫免疫，不会被再次感染。定义 $R(t)$ 为 t 时刻从感染状态恢复到免疫状态的主机数；γ 为恢复率，从 KM 模型可以得出：

$$dR(t)/dt=\gamma I'(t) \tag{4}$$

KM 模型讨论的是把从感染状态恢复的主机从网络中移除的情况，但是从易感染状态到免疫状态的变化过程较为复杂。本文提出的具有免疫特征的僵尸网络传播模型就重点考虑了这种复杂的变化情况。根据文献[12]提出的双因素蠕虫传播模型，在蠕虫最开始传播的时候，用户并不会知道网络中存在蠕虫，相应的防御措施采用较少。随着蠕虫数量的增长，被感染的主机越来越多，人们对蠕虫传播的关注程度逐渐增加，因此对易感染主机进行预先免疫处理，令其从蠕虫环境中移除。定义这种进行了提前免疫处理的易感染主机为 $Q(t)$；$J(t)$ 为在 t 时刻已经被感染了的主机数，即 $I(t)+R(t)$；μ 为修正系数，具体根据恶意程序的传播情况进行调整：

$$dQ(t)/dt=\mu S'(t)J(t) \tag{5}$$

可以看出，$Q(t)$ 的变化与 $J(t)$ 变化关系密切，并且与 $S'(t)$ 相关。因为考虑了 $Q(t)$，根据双因素蠕虫传播模型可知在 t 时刻的易感染主机数为：

$$S(t)=N-I(t)-R(t)-Q(t) \tag{6}$$

用 β 表示感染率，且 $\beta=\eta/\Omega$[15]，η 为恶意代码的扫描率，Ω 为恶意代码扫描的主机地址空间大小，再考虑到感染主机上下线的状态可得 $S(t)$ 的变化情况为：

$$dS(t)/dt=-\beta S'(t)I'(t)-dQ(t)/dt \tag{7}$$

综上所述，根据式(1)－式(7)得到僵尸网络增长的方程组：

$$\begin{cases} N=(1+\eta)N_0 \\ I'(t)=\alpha(t)I(t) \\ S'(t)=\alpha(t)S(t) \\ dR(t)/dt=\gamma I'(t) \\ dQ(t)/dt=\mu S'(t)J(t) \\ S(t)=N-I(t)-R(t)-Q(t) \\ dS(t)/dt=-\beta S'(t)I'(t)-dQ(t)/dt \end{cases} \tag{8}$$

化简式(8)，可以知道这种具有免疫特征的僵尸网络被感染主机数 $I(t)$ 的变化情况为：

$$dI(t)/dt=\beta\alpha^2(t)I(t)[(1+\eta)N_0-I(t)-R(t)-Q(t)]-\gamma\alpha(t)I(t)$$

4 模型仿真与分析

根据上述提出的微分方程，假设 $I(0)=10$，$S(0)=N-I(0)$，$R(0)=Q(0)=0$，$N=1000000$，$\beta=0.0000008$，$\mu=0.00000006$，$\alpha=0.6$，$\gamma=0.05$，$\eta=0.2$，用 Matlab 解微分方程可得感染主机的变化情况，如图 1 所示。

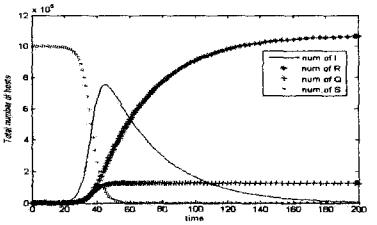

图 1 具有免疫特征的僵尸网络感染主机变化情况

图中的 I 反映被感染主机的变化情况。可以看出，I 的变化曲线是随着时间的增加而逐渐增大的，感染主机的数量在僵尸程序传播了 30 分钟以后迅速增加，僵尸程序开始大规模传播，并且在 50 分钟左右达到最大值，被感染主机数接近 780000 台。随着僵尸程序的传播，不断有被感染主机恢复到免疫状态或者从易感染网络中移除（R 和 Q 的增加），当被感染主机的恢复速度超过其增长速度后，随着时间推进，I 值逐渐减少，在 200 分钟左右减少到 0 台，至此僵尸程序传播结束。

R 的曲线表示主机从感染状态恢复至免疫状态的变化情况。在 I 增长了 10 分钟以后，用户通过各种处理手段，令被感染的主机恢复正常，达到免疫状态，R 的数量大幅增加。并且在被感染主机达到最大值以后，依然保持原有的速度增长，直到 I 趋近于 0 时才保持稳定。由于在僵尸程序传播过程不断有主机加入易感染的网络中，因此随着 N 的增大，R 的值在传播结束时超过了初始的 N 值。

Q 表示未被感染的主机提前进行了预防措施，在 I 增长了一段时间以后，人们意识到网络中可能存在恶意程序，从而采取了免疫措施让易感染的主机从网络中移除。在僵尸程序十分活跃的时候，被移除的主机数也相应增多，因此 Q 的数量随着 I 的增长而增加，并且在僵尸程序传播了 30 分钟以后 Q 的增加速度加快。当 I 增大到最大值时，僵尸程序的传播减慢，僵尸网络的活跃程度降低，所需从网络中移除的主机数量就停止增长，保持 150000 台的数量。

S 为网络中存在的易感染主机。S 的数量随着感染主机的增加而减小，在 I 增长到最大值后的一小段时间，S 的数量减小到 0 台，意味着网络中所有主机在 1 个小时左右都被感染过一次或者已经从易感染网络中移除。

在相同的条件下，仿真双因素蠕虫的传播情况[12]可以得出图 2。

从图 2 可以看出，双因素蠕虫传播模型的感染主机数增长得更快，在 t 到 10 分钟的时候就开始了增长，而图 1 中的 I

是在30分钟左右才开始增长，说明蠕虫传播因为不受控，是自动传播，所以恶意程序传播的效率更高，在不到30分钟的时间就达到了感染主机的最大值。但是由于传播得快，因此相应的感染主机被免疫和被移除的速度更快，因此图2中的最大感染主机数只有500000台，比图1少了近300000台。而I在100分钟时就趋近于0，传播结束，因此图2中的蠕虫传播的生命周期比图1的僵尸程序更短。

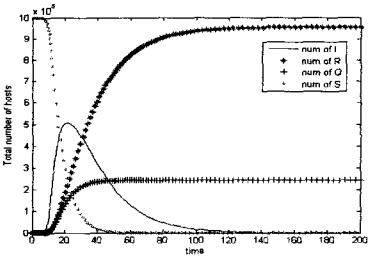

图2　双因素蠕虫模型的感染主机变化情况

图2中的蠕虫并没有考虑网络中主机上下线的状态，因此感染主机数S在40分钟降为0，即在这个时刻网络中所有主机已经被感染过一次，较图1中的被感染速度更快。由于蠕虫的快速传播，让用户无法很快地对已经被感染的主机进行免疫措施，因此R的值在传播结束时并未超过初始的N值。而Q的数量由于主机的在线情况不同，图2中的Q比图1中的被移除数更多，所以并不符合僵尸程序受控的特点。

同样对Dagon提出的基于时区的僵尸网络传播模型[11]进行模拟，可以得到图3的情况。

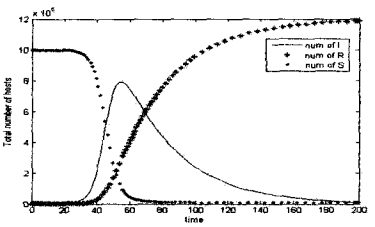

图3　基于时区的僵尸网络感染主机变化情况

把图3与图1进行比较，由于只考虑了主机从感染到免疫的状态，而未考虑主机没被感染前就从网络中移除的情况，因此感染主机的数量比图1中的更多，达到了近800000台。虽然感染开始的时间同样都是30分钟，达到感染主机最大值的时间也同样是在50分钟左右，图3与图1中的僵尸程序生存周期也相似，但是因为考虑的情况不够全面，所以感染主机I的数量太大，网络中被感染主机数过多，并不符合真实环境中僵尸程序的传播情况。

综上所述，由于在基于时区的僵尸网络传播模型的基础上加入了Q，即考虑了部分主机提前从网络中移除而达到免疫状态的情况，因此具有免疫特征的僵尸网络中感染主机数I的增长更加符合僵尸程序传播的真实情况。

结束语　随着技术的进步，僵尸网络也会继续发展，关于僵尸网络的研究也需要不断进行。分析僵尸网络的传播模型对于防御僵尸网络有着重要意义，而当前主流的僵尸程序都是借鉴蠕虫的传播，因此研究僵尸程序的传播模型就需要从蠕虫的传播技术着手，传统的蠕虫模型并不能很好地反应出僵尸网络的特征，部分主机提前进入免疫状态的情况在僵尸网络的构建中有着重要的地位。本文提出的具有免疫特征的僵尸网络传播模型，在基于时区的僵尸网络传播模型基础上，加入了部分主机未被感染前就已经采取了免疫措施，从脆弱状态成为免疫状态的情况。虽然这种模型进一步描述了僵尸网络传播的特性，但是它对传播过程中的网络拥塞等导致感染率变化的因素尚未进行进一步的考虑。

参 考 文 献

[1] Geer D. Malicious bots threaten network security [J]. IEEE Computer, 2005, 38(1): 18-20

[2] 警惕僵尸网络"还魂"[R]. 计算机世界, 2005

[3] CipherTrust, CipherTrust ' s Zombie Stats [J/OL]. http://www.ciphetrust.com/resources/statistics/zombie.php, 2006

[4] 陈明奇. 狙击僵尸网络[R]. 国家计算机网络应急技术处理协调中心运行一部, 2005

[5] 新华社. http://pdf.wenweipo.com/2010/01/25/a11-06-0125m.pdf, 2010

[6] Freiling F, Holz T, Wicherski G. Botnet tracking: Exploring aroot-cause methodology to prevent distributed denial-of-service attacks[R]. AIB-2005-07. CS Dept. of RWTH Aachen University, April 2005

[7] Zhou Jian-wei, Han Xin-hui, Zhou Yong-lin, et al. Research and Development of Botnets[J]. Journal of Software, 2008, 19(3): 702-715

[8] Streftaris G, Gibson G J. Statistical Inference for Stochastic Epidemic Models[C]// Proc. of the 17th IWSM. Chania: [s. n.], 2002: 609-616

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

[10] Zou C C, Gong W, Towsley D. Worm propagation modeling and analysis under dynamic quarantine defense[C]// Proceedings of the ACM CCS Workshop on Rapid Malcode. 2003: 51-56

[11] Dagon D, Zou C C, Lee W. Modeling botnet propagation using time zones[C]// Proc. of the 13th Annual Network and Distributed System Security Symp. (NDSS 2006). 2006

[12] Zou C C, Gong W, Towsley D. Code RedWorm Propagation Modeling and Analysis[C]// Proceedings of the 9th ACM Conference on Computer and Communication Security (CCS 2002). Washington, DC, 2002

[13] Lin Yong. Discussion of The BotNet [D]. Kunming: Yunnan Jiaotong College

[14] Wang P, Sparks S, Zou C C. An advanced hybrid peer-to-peer botnet[C]// Proc. of the 1st Workshop on Hot Topics in Understanding Botnets (HotBots 2007). Boston, 2007

[15] Zou C C, Towsley D, Gong W. On the performance of Internet worm scanning strategies[J]. Elsevier Journal of Performance Evaluation, 2005