基于 P2P 的僵尸网络及其防御

应凌云[1], 冯登国[1,2], 苏璞睿[2]
(1. 中国科学院研究生院, 信息安全国家重点实验室, 北京 100080;
2. 中国科学院软件研究所, 信息安全国家重点实验室, 北京 100190)

摘 要: 僵尸网络作为网络犯罪活动的平台, 正朝着 P2P 等分布式结构发展. 研究僵尸网络的发展方向以及构建技术, 有助于我们全面地了解僵尸网络活动的特点, 从而更好地开展僵尸网络的检测和防范研究. 本文分析了攻击者的需求, 提出了一种基于层次化 P2P 网络技术的新型僵尸网络结构, 并对这种僵尸网络的可行性和具体的传播、通讯、控制等各个方面进行了深入分析和探讨. 在此基础上, 我们通过模拟实验对各种防御策略的有效性进行了分析和评估, 实验数据表明, 在考虑实际可操作性条件下, 现有的防御策略难以有效摧毁 P2P 结构僵尸网络. 最后, 我们讨论了这种新型僵尸网络可能的防御方法.

关键词: 僵尸网络; P2P; 恶意代码; 网络安全
中图分类号: TP393 文献标识码: A 文章编号: 0372-2112 (2009) 01-031-07

P2P-Based Super Botnet: Threats and Defenses

YING Ling-yun[1], FENG Deng-guo[1,2], SU Pu-rui[2]
(1. *State Key Laboratory of Information Security, Graduate University of Chinese Academy of Sciences, Beijing* 100080, *China*
2. *State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences, Beijing* 100190, *China*)

Abstract: As a cyber crime platform, botnet is one of the biggest network security threats. Researching the evolvement of botnets, as well as possible botnets construction methods, can improve us in-depth understanding of details of botnets, and also guide us in the botnets defense research. In this paper, we proposed a P2P-based super botnet structure, analyzed the feasibility and discussed all aspects of this new type botnet. Secondly, we evaluated effectiveness of various botnets defense strategies, our simulation results show that, when taking the actual operational conditions into account, it is difficult to completely destroy P2P-based botnets. Finally, we discussed some guidelines for defending against such new botnets.

Key words: botnet; peer to peer (P2P); malware; network security

1 引言

僵尸网络(Botnet)作为网络犯罪活动的平台, 是当前网络安全领域面临的主要问题之一, 它具有一定的组织结构, 蠕虫似的传播特征, 木马似的后门特征, 并往往采用变形、Rootkit 等病毒技术, 是恶意代码技术的综合. 大量的 DDoS 攻击, 垃圾邮件发送都与僵尸网络有关. 早期僵尸网络的命令控制通道(Command & Control)建立在 IRC 协议上, 攻击者借助 IRC 服务器对整个僵尸网络进行控制. 近年来出现了多种新型的僵尸网络, 如基于 IM、HTTP 等不同协议的僵尸网络, 并出现了采用树形结构、随机网络拓扑以及具有部分 P2P 特征的僵尸网络.

由于僵尸网络技术的快速发展和危害的日益加剧, 僵尸网络相关研究也逐渐受到了国内外学者的重视. 如 D. Dagon 等人利用蠕虫传播模型对僵尸网络传播过程进行了建模分析, 指出了僵尸网络规模随时区变化的特点[1]. J. Zhuge 等人对 IRC 僵尸网络活动进行了统计分析[2]. G. Gu 和 A. Karasaridis 等人提出了基于 IDS 的僵尸网络检测方法[3~5], P. Barford 等人则指出僵尸网络流量存在随机性, 利用网络流量指纹难以有效检测的问题[6]. 然而, 这些研究集中于 IRC 僵尸网络, 新型僵尸网络尤其是 P2P 僵尸网络是目前研究的前沿和热点问题, J. B. Grizzard 等人通过对僵尸程序 Peacomm 的分析, 指出了僵尸网络朝 P2P 结构发展的趋势[7], P. Wang 和 R. Vogt 等人则提出了分布式结构僵尸网络的构建方案[8,9].

由于僵尸网络技术的快速发展, 探讨新型僵尸网络的可行性和构建模式, 对于今后研究工作的开展具有重要意义. 本文通过分析现有各种僵尸网络结构的特点, 提出了一种全新的基于 P2P 的僵尸网络结构, 分析了这种僵尸网络在构建, 交互和控制等各个阶段的特点, 并通过模拟实验讨论了各种防御策略的有效性和可行性,

收稿日期: 2008-02-27; 修回日期: 2008-09-01
基金项目: 国家 863 高技术研究发展计划(No. 2006AA01Z412, 2007AA01Z451); 国家自然科学基金(No. 60703076)

最后提出了若干可能的检测和防御方案. 作者相信, 通过研究僵尸网络的攻击模式, 可以帮助研究人员更有效地评估潜在的威胁, 分析现有僵尸网络防治措施的局限性, 为进一步研究有效的僵尸网络检测和摧毁技术提供指导.

本文的主要贡献如下:

①论证了结构化 P2P 僵尸网络的可行性. 本文详细分析了不同僵尸网络结构的特点和局限性, 指出了 P2P 结构的特性与僵尸网络的需求之间密切重合, 并对构建结构化 P2P 僵尸网络的可行性进行了论证;

②提出了一套结构化 P2P 僵尸网络构建方案. 通过分析僵尸网络需求, 本文给出了一种 P2P 僵尸网络构建方法, 对其中的引导, 控制等重要环节进行了分析, 并对其中重要参数的选取进行了讨论;

③评估了不同防御策略的有效性. 模拟实验数据表明, 针对高连接度节点的摧毁策略能有效破坏僵尸网络结构, 但在缺乏足够的僵尸网络信息的情况下, 各种防御策略都难以彻底摧毁 P2P 结构僵尸网络.

本文的组织结构如下: 第二节分析各种僵尸网络的特点和 P2P 僵尸网络的可行性. 第三节介绍我们提出的基于 P2P 的新型僵尸网络模型. 第四部分讨论各种僵尸网络摧毁方案的有效性. 最后对本文的工作进行了总结, 并简要介绍了下一步的工作重点.

2 P2P 僵尸网络的可行性

僵尸网络的命令控制通道是连接僵尸主机和攻击者的纽带, 决定了僵尸网络的生命力和攻击性. 以命令控制结构中是否存在中心控制点划分, 可以将僵尸网络分为中心化结构和分布式结构两种. J. Zhuge 等人的研究表明, 非 IRC 结构尤其是分布式结构僵尸网络的数目正迅速增加, 并且就单个僵尸网络的规模而言, 分布式结构僵尸网络要远大于中心化结构僵尸网络[2].

2.1 IRC 僵尸网络的局限性

IRC 僵尸网络的泛滥源于它相对简单的构建技术. 然而由于 IRC 结构本身的特点, 限制了这种僵尸网络的攻击力:

①单点失败: 整个僵尸网络依赖于 IRC 服务器的中心控制, 网络健壮性很差, 关闭 IRC 服务器即可摧毁整个僵尸网络. 虽然可用动态域名解析等技术增强网络的可生存性, 但是结构上的单点依赖决定了网络的健壮性无法进一步提高.

②规模受限: 受到 IRC 服务器软硬件资源限制, 中心控制点无法承载大规模的并发网络连接, 限制了僵尸网络的有效规模(< 5,000) 和攻击力.

③明文传播: 由于 IRC 协议通过明文传输, 流量比较容易检测, 容易暴露中心控制服务器位置和网络活动信息, IRC 僵尸网络生存期普遍较短的现象印证了这一点[2].

通过以上分析我们可以看到, 组织结构上的缺陷限制了 IRC 僵尸网络的攻击力, 攻击者要想发动大规模的攻击, 必须通过同时控制许多个僵尸网络, 这显然大大提高了攻击者的操作难度和攻击成本.

2.2 结构化 P2P 僵尸网络的特点

在已知的具有 P2P 特性的僵尸网络中, Sinit 依赖随机扫描, Nugache 通过内置的引导节点, 组成随机结构的 P2P 网络. Phatbot 采用的 WASTE 协议只支持小规模 P2P 网络, 导致网络不可扩展. SpamThru 仍然采用中心控制结构, P2P 协议只作为备用. Slapper 着重在简单的 P2P 协议上实现复杂的通讯机制, 网络没有结构, 且存在网络分割问题. Peacomm 采用 Overnet 协议, 僵尸节点与正常节点混合, 存在 index poisoning 和隐蔽性问题. 这些僵尸网络由于没有良好的组织结构, 导致网络规模受限, 并且容易出现网络分割问题, 可控性较差.

结构化 P2P 网络具有分布式, 无中心的特点, 同时具有良好的可扩展性和自组织能力, 能很好地适应异构网络, Skype 的成功表明, 这种结构能够支持> 10^6 用户同时在线. 利用结构化 P2P 技术构建僵尸网络能极大地扩大僵尸网络的规模, 同时无中心的网络结构, 使得防御者检测和摧毁僵尸网络非常困难. 并且, 结构化 P2P 网络还能确保消息延迟在合理范围内, 使得发动 DDoS 等需要协调大规模僵尸主机同时参与的攻击活动更容易. 这种庞大规模、高可生存性和隐蔽性的僵尸网络, 对于攻击者具有很强的吸引力.

在技术上, 已经有 JXTA, FreePastry 等各种开源 P2P 实现可以利用, 降低了攻击者构建 P2P 僵尸网络的技术难度. 我们相信, 攻击者已经具有足够的动力和技术能力构建基于结构化 P2P 的僵尸网络.

3 基于 P2P 的僵尸网络

下面, 我们将从攻击者的角度来设计和构造一个结构化 P2P 僵尸网络. 我们提出如下三个设计目标:

①隐蔽性: 隐蔽性是僵尸网络的首要特征, 也是僵尸网络威胁性的重要衡量标准. 隐蔽性不仅包括僵尸网络整体信息和网络通讯内容的隐蔽, 也包括攻击者信息的隐蔽.

②可控性: 攻击者构建僵尸网络是为了能够利用僵尸网络达到一定的目的, 因此保持对僵尸网络的控制权, 防止僵尸网络被他人劫持对攻击者而言至关重要.

③健壮性: 在确保僵尸网络对攻击者安全、可用的同时, 攻击者必然会尽最大努力提高僵尸网络的健壮性. 这对于最大化僵尸网络价值而言十分关键.

下面, 我们从网络架构, 信息交互和网络控制等各个方面, 阐述在不同阶段攻击者为了实现上述设计目标可能采取的策略, 以及与一般 P2P 应用的差异, 并对这些设计原则的依据和影响进行讨论.

3.1 架构

如何充分利用已有信息实现对大量僵尸主机的有效管理是僵尸网络构建的关键. 由于 IP 地址占用和漏洞主机分布呈现聚簇现象[10,11], 僵尸主机分布也将具有类似的聚簇现象, 因此, 我们采用基于超级节点(Super Node)的层次化 P2P 结构充分利用这一特点. 其中超级节点 SuperBot 构成整个僵尸网络的核心 SuperNet; 普通节点 PeerBot 通过 SuperBot 接入网络, 构成 PeerNet. 整体结构如图 1 所示.

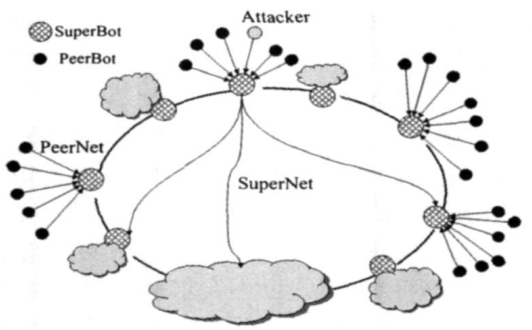


图1　层次化P2P结构的僵尸网络示意图

为了保证网络的健壮性, SuperBot 的选择需要满足一定的条件. Y. Xie 等人的研究表明, 因特网中约有 33% 的主机具有公网 IP 地址且固定时间超过 3 天[12]. 并且统计数据显示, 有超过 50% 的因特网用户采用宽带接入[13]. 考虑到主机在线时间服从 Pareto 分布, 即当前在线时间越长, 继续在线的可能性越高[14]. SuperBot 选择准则可定为:

①具有公网 IP 并且 IP 固定 3 天以上;

②带宽大于 1mbps;

③相同条件下按照在线时长优先选取.

上述准则可以避免 SuperBot 节点的抖动, 有利于提高整个网络的健壮性.

同时, SuperBot 维护一个大小为 K_{sv} 的 SuperView 列表用于缓存网络局部视图, 保存已知的其他 SuperBot 信息, 包括 IP, 端口, RTT 和心跳(Heartbeat)记数. 此外, 为了减少网络半径, 根据 Small World 理论, 我们以 RTT 作为 SuperBot 之间距离衡量的准则, S_i 按照 $P(S_j) \propto 1/RTT_{ij}^2$ 的原则选择 S_j 作为 SuperView 中的项[15]. 每个 SuperBot 还维护一个大小为 K_{pv} 的 PeerView, 保存与它相连 PeerBot 的信息, 包括 IP, 端口, RTT 估计, 在线时长和心跳记数, 用于评估节点性能. 与普通的 P2P 应用不同, 为了提高隐蔽性, SuperBot 离线后不保存这些信息.

PeerBot 则维护一个大小为 K_{bl} 的 Bootlist, 内容包括已知 SuperBot 的 IP, 端口, RTT 估计和心跳记数, 用于确保自身在网络抖动情况下仍然能够接入僵尸网络.

3.2 传播和引导

在僵尸网络初始构建阶段, 攻击者可以预先建立若干个 SuperBot 作为种子节点. 当主机 A 感染主机 B 时, 发送 Bootlist 给 B, 而当 B 成功接入网络时, SuperBot 主机 S 随机挑选一部分 SuperBot 信息发送给 B. B 通过如下原则替换自身 Bootlist 信息:

①替换失效节点;

②保留与 B 在同一 B 类网段内的 SuperBot;

③其余进行随机替换.

通过上述的替换策略, 可以确保节点引导列表的有效性, 同时使节点倾向于连接距离更近的 SuperBot, 形成与英特网类似的聚簇结构. 与蠕虫等的传播感染不同, 这里必须引入随机因素以消除 Bootlist 中的感染路径传递关系, 提高网络的隐蔽性.

节点接入时采用如下算法选择接入点:

算法 1: 接入点选择算法

输入: Bootlist[1 … n]

输出: 可作为接入点的有效 SuperBot

如果 PeerBot 不是第一次接入网络并且上一次连接的 SuperBot 有效

返回上一次连接 SuperBot 作为接入点;

对于第 i = {1, 2, 3 …, n} 条引导项 Bootlist[i]

如果 Bootlist[i]和 PeerBot 在同一个 C 网段内

返回 Bootlist[i]作为接入点;

对于第 j = {1, 2, 3, …, n} 条引导项 Bootlist[j]

如果 Bootlis[j]和 PeerBot 在同一个 B 网段内

返回 Bootlist[j]作为接入点;

从 Bootlis 中随机选取 b 条选引导项;

对于第 t = {1, 2, 3 …k} 条选中项, 估算 RTT 值

返回 RTT 最小的项作为接入点;

返回 Bootlist 中第一务有效项作为接入点.

同时, 为了避免大量节点同时连接对僵尸网络稳定性的影响, 在节点接入时采用指数退避算法. 节点第一次连接失败时, 等待时间 T 后重试, 若再失败, 则再等待 $2T$, 以此类推, 第 n 次重试前, 需要等待 2^nT. 这可以避免类似 Skype 网络在发生大规模节点重连时网络性能急剧下降的情况.

通过如上所述的 Bootlist 替换机制和接入点选择算法, 确保同一区域内 PeerBot 包含的可用接入点信息的分散性, 避免网络出现热点和单点失败问题.

3.3 处理抖动

为了适应主机故障或网络问题造成的节点抖动, 互联节点间隔 t 时间发送心跳包, 相互检测对方的活

性,这里的 t 必须要有足够的随机性以免触发 *IDS* 等检测系统. 通过随机化心跳包的载荷数据长度 L 还可以进一步降低流量特征. 关于 t 和 L 的选取,我们将在3.5节中进行专门讨论.

当 SuperBot 检测到 PeerBot 离线时只需调整 PeerView,而当 PeerBot 检测到 SuperBot 离线时,则需要利用 Bootlist 重新引导加入网络. 当 SuperBot 检测到 SuperView 中邻居节点离线时,我们采取延迟替代策略,只有当 $K_{sv}/2$ 的邻居失效时 SuperBot 才触发邻居更新功能,获取其他在线节点信息替换失效项目.

当网络规模变化时,网络通过 SuperBot 的进化和退化过程实现动态自调整. 设定每个 PeerNet 规模上限为 C_{max},该参数取决于 SuperBot 的性能,当前规模记为 C,同时限定 PeerNet 规模下限为 C_{min},该参数决定了网络对节点抖动的敏感程度.

当某个 SuperBot(记为 S)满足 $C \leqslant C_{min}$ 时触发 SuperBot 到 PeerBot 的退化过程:

步骤一:S 根据 SuperView 中 RTT 信息选择 3 个距离最近的邻居,将自身 PeerNet 中的 PeerBot 随机派往这些邻居节点;

步骤二:S 自身也作为 PeerBot 加入到某个邻居节点的 PeerNet.

而当 S 满足 $C \geqslant C_{max}$ 时触发 PeerBot 到 SuperBot 的进化过程:

步骤一:S 根据 PeerView 信息和 SuperBot 选择准则挑选出候选 PeerBot(记为 P),并向 P 发送进化通知;

步骤二:P 随机选择 SuperBot-SuperBot 之间的通讯端口 p_s 和 SuperBot-PeerBot 之间的通讯端口 p_b,返回给 S;

步骤三:S 将 PeerNet 中节点随机分成大小为 C/2 的两部分,一部分保留,并通知另一部分 PeerBot 向 P (ip, p_b) 连接;

步骤四:S 将自身的 SupeView 发给 P,使 P 通过 (ip, p_s) 连入 SuperNet;

若 P 进化失败,则 S 重复上述过程.

采用上述的进化-退化机制实现网络的动态自调整,不仅保证了网络的健壮性,也提高了整个僵尸网络的隐蔽性. 由于 SuperBot-SuperBot 之间通讯端口和 SuperBot-PeerBot 之间的通讯端口不同,所以 PeerBot 无法伪装成 SuperBot 加入 SuperNet,使得防御者难以渗透进僵尸网络,无法准确掌握僵尸网络活动情况.

3.4 交互与控制

为了消除通讯流量中可能存在的特征模式,隐藏流量,攻击者根据 RSA 算法产生一对密钥,在僵尸程序内嵌入公钥 KU,同时保留对应的私钥 KR. 引入公钥有两重目的,一是用于控制者认证,二是作为产生加密密钥的参数. 鉴于公钥加密的开销,可利用通讯双方的 IP,在 A 欲与 B 通讯时根据公式(1)产生会话密钥(Session Key, Ks),对流量进行加密.

$$Ks = HASH(KU + IP_A + IP_B) \tag{1}$$

由于会话密钥 Ks 与通讯双方的 IP 相关,通讯端口随机选择,使得网络通讯不存在可以被用于检测的特征. 虽然对手可以通过分析僵尸程序取得 KU,并根据通讯双方 IP 在线产生 Ks 并解密相关数据,但由于实时处理能力和通讯延时的限制,这种方式不具有实际可行性.

为了保持对僵尸网络的控制,攻击者用 KR 对所有命令进行加密,僵尸程序里则用 KU 对命令进行解密和来源认证. 同时利用命令序号和命令缓存转发机制,确保命令能够覆盖到所有节点. 由于 P2P 网络的自治性,使得整个僵尸网络具有很好的可控性.

同时,攻击者可通过可加载模块动态地更新僵尸程序. 为了避免更新机制被对手利用,攻击者用 KR 对更新模块做签名认证,确保只有他能够触发更新过程. 这种更新方式实现简单但非常灵活,不要求整个网络中所有节点都得到更新,也不要求所有节点都得到同样的更新. 受控更新确保了攻击者可以按需改变僵尸程序的行为和网络结构,甚至是认证模式和 KU 本身,极大地提高了僵尸网络的灵活性和可控性.

3.5 参数选择

由于 SuperNet 是整个僵尸网络的核心,因此,是否存在足够多的满足 SuperBot 条件的备选主机是个问题. 由于因特网中位于 NAT 后的主机约占 19%[16],记存在目标漏洞的主机数为 r,根据 3.1 节的讨论可知,满足 IP 和带宽条件的节点比率为 33% 和 50%,备选 SuperBot 主机数 $N_{candidate}$ 为:

$$N_{candidate} = r * 33\% * (1 - 19\%) * 50\% = 0.13r$$

即在僵尸程序充分传播后,平均每个 SuperBot 需要承载的 PeerBot 数为 6.7 个. 为了确保在 SuperBot 分布不均,节点抖动较为严重的情况下网络健壮性能够得到保证,可取 C_{min} 为 3, C_{max} 为 20,则最少只需理论计算值的 36.6% 的 SuperBot 数即可满足网络完整性条件. 值得注意的是,由于位于 NAT 前和具有公网 IP 两个概率实际上不完全独立,导致 $N_{candidate}$ 实际值比理论计算值偏小,平均 PeerNet 规模比理论值偏大,模拟实验也证实了这一点.

由于无法事先确定漏洞主机的准确分布情况,我们假定漏洞主机均匀分布. 若僵尸网络规模为 $N = 10^6$,则平均每个 PeerNet 规模 C_{avg} 和所需 SuperBot 数 N_s 分别为:

$$C_{avg}=(C_{min}+C_{max})/2=11.5,$$

$$N_s=N/(1+C_{avg})=80,000$$

根据 3.1 节的邻居节点选择策略, 在 $K_{sv}=2*\log_2(80000/2)=30$ 时即可以达到 O(logN) 级别的性能[15]. 同时, 根据节点进化过程, K_{pv} 由 C_{max} 限定.

对于 PeerBot, 在不进行 Bootlist 中失效接入点替换的情况下, 保持至少一个接入点有效的概率如公式(2)所示:

$$P(bl)=K_{bl}/(K_{bl}+1) \quad (2)$$

即取 $K_{bl}=20$ 就可以保证在最坏情况下仍能达到 95% 的引导成功率[17].

在心跳检测机制中, 考虑到网络延时, 取 t 服从[60 ~ 300]的均匀分布, 同时根据互联网流量的数据包长度分布特点, 取 L 服从[40~1400]的均匀分布[18], 则有:

$$Exp(t)=180, Exp(L)=720$$

$$(20+20+720)*(C_{max}+K_{sv})/Exp(t)=1.69kbps$$

根据 SuperBot 选择准则可知, 心跳流量所占带宽不超过可用带宽的 0.2%, 这足以保证僵尸网络的常规流量能够混杂于背景流量之中.

与普通 P2P 应用不同, 这里的参数选择必须考虑僵尸网络的特殊性. 其中 K_{sv} 越大, SuperNet 互联程度越高, 网络越健壮, 但同时一个 SuperBot 掌握的网络信息越多, 发生信息泄漏时对僵尸网络的威胁也越大. K_{bl} 越大, 备选的接入点越多, PeerBot 接入网络的成功率越高, 当对手渗透进网络时, 获取的信息也越多. 攻击者可以通过观察网络的运行情况, 动态调整相关参数, 达到隐蔽性、可控性和健壮性三个目标的均衡.

4　防御策略分析与评估

研究僵尸网络的攻击过程和发展趋势, 是为了更好地研究相应的检测和防御方法. 下面我们对如何检测和防御这种新型僵尸网络做进一步的探讨.

现有的僵尸网络检测方法依赖于利用 IRC 协议的特点, 容易被绕过. 如 J. R. Binkley 等人提出的通过解析 IRC 通讯流量来检测僵尸网络的技术[19], 对本文提出的这种新型僵尸网络就无能为力. 同时, 流量加密技术的采用, 导致 IDS 等基于流量指纹的检测方法也不能有效检测到类似的新型僵尸网络. 利用 Honeypot 等技术检测僵尸网络, 结合恶意代码分析技术和网络渗透技术收集僵尸网络信息, 通过破坏 P2P 僵尸网络的关键节点摧毁僵尸网络是比较可行的防御方法.

为了深入分析各种防御策略的有效性, 我们利用离散事件模拟器对不同的防御策略进行了评估. 模拟的僵尸网络包括 100,000 台僵尸主机, 僵尸主机被移除后不可重复感染. 各种参数参照第三节的讨论. 为了符合僵尸网络检测普遍滞后于僵尸网络传播的事实, 各种防御策略在僵尸网络构建完成后开始实施. 同时为了尽可能消除随机因素的影响, 我们对每种策略都重复进行 5 次. 具体的防御策略如表 1 所示. 其中 Region 策略模拟了某一网段发现并清除了僵尸程序的情况, SuperView 策略和 PeerView 策略则代表了复杂网络中基于节点中心化程度(Centrality)的攻击[20] Region 策略每轮摧毁僵尸主机范围为 300, 其他防御策略每轮摧毁的僵尸主机数量为 100.

表 1　防御策略含义

防御策略	含义
Plain	不采取任何系统的防御措施, 对照基准
Random	随机摧毁 SuperBot 和 PeerBot
Region	随机摧毁某一区域内的 SuperBot 和 PeerBot
PeerOnly	随机摧毁 PeerBot
SuperOnly	随机摧毁 SuperBot
SuperView	按照 K_{sv} 从大到小的顺序摧毁 SuperBot
PeerView	按照 K_{pv} 从大到小的顺序摧毁 SuperBot

为了能够准确评估各种策略在不同条件下的效果, 我们分完全信息条件和不完全信息条件两种情况进行分析. 完全信息条件是指防御者事先拥有僵尸网络的各种信息, 如僵尸主机 IP, SuperBot 分布等信息, 是理想状态下的防御效果评估. 不完全信息条件是指防御者只能通过各种实际可行的手段, 如 Honeypot, 网络渗透等收集僵尸网络信息, 利用这些不完整的信息开展僵尸网络防御的情况, 是实际状态下的防御效果评估.

4.1　完全信息条件下的防御效果评估

我们首先假定防御者掌握整个僵尸网络的活动信息, 即防御者不仅知道僵尸网络的结构, 并且清楚各个僵尸节点的属性, 分布和具体地址. 同时, 为了能够比较不同防御策略的效果, 我们取僵尸网络的连接率 α 为评价标准, 具体定义如下:

$$\alpha=N_{connected}/N_{infected} \quad (3)$$

其中, $N_{connected}$ 为已连入僵尸网络的僵尸节点的数目, 即 PeerBot 与 SuperBot 数目之和, $N_{infected}$ 为感染节点的数目, 即已连入和未连入僵尸网络的僵尸节点数目之和. 各种不同防御策略的模拟结果如图 2 所示.

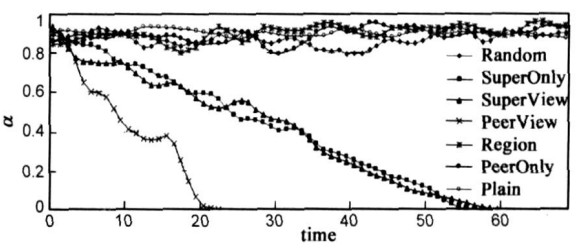


图2　各种不同防御策略下僵尸网络的连接率

从图 2 中我们可以看到, 在 Ramdom, Region 和

PeerOnly 三种防御策略下, 随着防御措施的不断施行, 失效节点逐渐增加, 连接率的抖动也不断增大, 但对于整个僵尸网络结构基本没有影响, 连接率始终保持在 90% 左右, 表明这种新型僵尸网络结构对于这三种防御策略具有很强的抗攻击能力.

同时, 我们也可以看到针对 SuperBot 的三种防御策略都能完全摧毁僵尸网络. 随着防御措施的实施, 大量 SuperBot 下线, 从而引起其他 SuperBot 的 PeerView 和 SuperView 变动加剧. 由于 SuperBot 的连接数等于 PeerView 与 SuperView 之和, 而 PeerView 变化比 SuperView 变化更剧烈, 因此各个 SuperBot 总连接度的差异主要受 PeerView 的大小差异决定. 图 2 表明 PeerView 策略最有效, 相当于首先摧毁网络中连接度最高的节点, 这一结论与复杂网络的特点相符合[20].

4.2 不完全信息条件下的防御效果评估

在实际的防御中, 由于僵尸网络的隐蔽性和网络结构的复杂性, 防御者不可能完全掌握僵尸网络结构和节点信息. 考虑到各种防御策略依赖的条件, 只有 Random, PeerOnly, SuperOnly 和 Region 四种策略具有实际可操作性. 由于 Random 和 Region 策略并不需要任何预先知识, 防御效果基本如图 2 所示, 这里不再重复. 而 PeerOnly 防御策略只针对普通节点, 也无法有效摧毁僵尸网络, 限于篇幅限制也不再单独讨论. 接下来我们重点分析 SuperOnly 策略.

SuperOnly 策略的有效性依赖于防御者收集到的 SuperBot 信息量. 为此, 我们以防御者能够收集到的 SuperBot 比例 β 作为衡量标准, 模拟分析了不同强度的 SuperOnly 策略的防御效果, β 的定义如下:

$$\beta = N_{collected} / N_{active} \quad (4)$$

其中, $N_{collected}$为防御者能够收集到的 SuperBot 节点数目, N_{active}为所有在线的 SuperBot 节点数目. 同时, 我们用 z 表示防御强度, 其值为防御者收集 SuperBot 信息的总时间与僵尸网络自我调整时间的比值, 具体数据如图 3 所示.

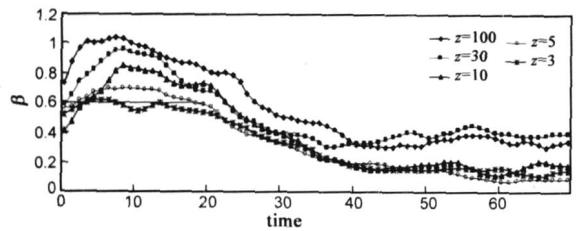


图3 不同强度SuperOnly防御策略下防御者能够收集到的SuperBot比例

从图 3 我们可以看到, 随着防御措施的不断开展, 僵尸网络的结构受到的破坏不断加大, SuperView 中失效条目增多, 导致僵尸网络抖动加剧. 然而, 这也导致了防御者能够收集到的 SuperBot 比例不断下降, 防御效果逐渐减弱. 最后, 僵尸网络通过动态调整修复网络的能力与防御措施对网络的破坏能力达到一个平衡状态. 同时, 从图中我们也可以看到, 提高防御强度能够在一定程度上提升防御效果, 但是网络的动态性限制了防御者能够获取的信息数量, 导致额外的资源投入并不能换来对应的回报. 由于只能收集并摧毁部分 SuperBot, SuperOnly 防御策略也不能完全摧毁僵尸网络.

值得注意的是, 在 $z = 100$ 的情况下, 我们看到图中存在 β 值超过 100% 的情况. 这是由于信息收集时间过长, 在收集过程中部分 SuperBot 下线或退化了, 导致收集的部分信息过时. 在总的收集时间较长的情况下, 过时信息的绝对数量也相应增加. 信息收集时间越长, 收集的数据过时就越严重, 这也是导致提高防御强度不能明显提升防御效果的原因.

4.3 讨论

通过上面的分析我们看到, 由于 P2P 系统本身特有的信息分散性, 任何单一防御点都难以获得开展有效防御所需的信息, 加强信息收集能力是防御 P2P 僵尸网络的关键. 这可以通过两个方面进行努力, 一是提高单一防御点的僵尸网络信息收集能力, 可以通过提高信息收集过程的自动化程度实现, 这依赖于僵尸程序自动化分析, 僵尸网络通讯协议自动化解析以及僵尸网络渗透工具等方面的研究. 二是通过在不同防御点间进行充分的信息共享和交换, 通过大量相互合作的防御点进行协同防御, 能够将不同防御点获得的僵尸网络局部信息组合成全局信息, 从而更有利于防御措施的开展. 这涉及到不同管理域之间的协调问题, 其中, 合作机制以及相关支持工具的研究是关键.

此外, 对于 P2P 僵尸网络这样的复杂攻击, 我们看到基于流量的检测技术存在诸多缺陷. 利用僵尸程序必然存在网络通讯和参与攻击活动这两个特点, 通过行为分析, 从主机端进行僵尸程序检测, 通过网络防御与主机防御结合的方式可以提高防御效果. 这种主动防御与深层防御结合的方式将是僵尸网络防御的重点.

5 结束语

僵尸网络作为网络攻击活动的平台, 正朝着以 P2P 为代表的分布式结构发展. 本文分析了结构化 P2P 僵尸网络的可行性, 详细阐述了 P2P 僵尸网络的组织结构, 通讯方式和控制机制等问题, 并给出了关键的系统参数及其选择标准. 通过对各种防御策略的模拟分析表明, 在考虑实际可操作性条件下, 现有各种防御策略都难以有效摧毁 P2P 结构僵尸网络.

下一步, 我们将把研究重点放在 P2P 僵尸网络的检测和防御技术上, 包括利用复杂网络理论分析 P2P

僵尸网络的网络分割临界点, 研究利用程序动态分析技术进行自动化僵尸程序分析的方法和利用分布式协同网络进行僵尸网络信息收集的技术.

参考文献:

[1] D Dagon, C Zou and W Lee. Modeling Botnet Propagation Using Time Zones[A]. In Proc. NDSS' 06[C], 2006.

[2] J Zhuge, T Holz, X Han, J Guo and W Zou. Characterizing the IRC-based Botnet Phenomenon[R], TR-2007-010, 2007.

[3] G Gu, P Porras, V Yegneswaran, M Fong and W Lee. BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation[A]. In Proc. USENIX Security' 07[C]. 2007. 167-182.

[4] G Gu, J Zhang and W Lee. BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic[A]. In Proc. NDSS' 08[C], 2008.

[5] A Karasaridis, B Rexroad and D Hoeflin. Wide-scale Botnet Detection and Characterization[A]. In Proc. USENIX HotBots' 07[C]. 2007. 7-7.

[6] P Barford and M Blodgett. Toward Botnet Mesocosms[A]. In Proc. USENIX HotBots' 07[C]. 6-6.

[7] J B Grizzard, V Sharma, C Nunnery, B B Kang and D Dagon. Peer-to-Peer Botnets: Overview and Case Study[A]. In Proc. USENIX HotBots' 07[C]. 1-1.

[8] P Wang, S Sparks and C Zou. An Advanced Hybrid Peer-to-Peer Botnet[A]. In Proc. USENIX HotBots' 07[C]. 2-2.

[9] R Vogt, J Aycock and M Jacobson. Army of Botnets[A]. In Proc. NDSS' 07[C]. 2007. 111-123.

[10] Y Pryadkin, R Lindell, J Bannister and R Govindan. An Empirical Evaluation of IP Address Space Occupancy[R]. ISI-TR-2004-598, 2004.

[11] Z Chen and C Ji. Optimal worm-scanning method using vulnerable-host distributions[J]. International Journal of Security and Networks, 2007, 2(1/2): 71-80.

[12] Y Xie, F Yu, K Achan, E Gillum, M Goldszmidt and T Wobber. How Dynamic are IP Addresses? [J]. Comput. Commun. Rev., 2007, 37(4): 301-312.

[13] J B Horrigan and A Smith. Home Broadband Adoption 2007 [OL]. http://www.pewinternet.org/pdfs/PIP_Broadband2007.pdf, 2007.

[14] D Stutzbach and R Rejaie. Understanding churn in peer-to-peer networks[A]. In Proc. ACM IMC' 06[C]. 2006. 189-202.

[15] J M Kleinberg. Navigation in a small world[J]. Nature, 2000, 406(6798): 845.

[16] M A Rajab, F Monrose and A Terzis. On the impact of dynamic addressing on malware propagation[A]. In Proc. ACM WORM' 06[C]. 2006. 51-56.

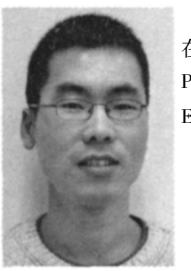
[17] D Leonard, V Rai and D Loguinov. On lifetime-based node failure and stochastic resilience of decentralized peer-to-peer networks[J]. SIGMETRICS Perform. Eval. Rev, 2005, 33(1): 26-37.

[18] C Shannon, D Moore and K C Claffy. Beyond folklore: observations on fragmented traffic[J]. IEEE/ACM Trans. Netw., 2002, 10(6): 709-720.

[19] J R Binkley and S Singh. An Algorithm for Anomaly-based Botnet Detection[A]. In Proc. USENIX SRUTI' 06[C]. 2006. 43-48.

[20] P Holme, B J Kim, C N Yoon and S K Han. Attack vulnerability of complex networks[J]. Physical Review E, 2002, 65(5): 056-109.

作者简介:



应凌云　男, 1982 年出生于浙江永康, 博士在读, 主要研究方向为网络安全、恶意代码分析、P2P 网络.

E-mail: yly@is.iscas.ac.cn.

冯登国　男, 1965 年出生于陕西靖边, 博士、研究员、博士生导师, 主要研究方向为密码学与信息安全.

苏璞睿　男, 1976 年出生于湖北宜昌, 博士、副研究员, 主要研究方向为恶意代码分析与防范.