基于动态 IP 地址去重和 NAT 识别的 IRC 僵尸网络大小度量

李润恒，甘亮，贾焰
（国防科学技术大学 计算机学院，湖南 长沙 410073）

摘 要：由于 IRC 僵尸网络的动态性，以及动态 IP 地址和 NAT IP 地址的影响，给僵尸网络的大小度量带来很大的难度，采用动态 IP 地址去重算法和基于通信频繁比对的 NAT IP 识别算法，给出僵尸网络大小准确的度量，实验验证了所提方法的有效性。

关键词：僵尸网络；通信；去重；NAT；大小度量
中图分类号：TP393.08 **文献标识码**：A **文章编号**：1000-436X(2010)9A-0183-07

IRC botnets' size measure based on duplicated removal of dynamic IP and NAT identifing

LI Run-heng, GAN Liang, JIA Yan
(School of Computer Science , National University of Defense Technology , Changsha 410073, China)

Abstract: For botnet's dynamic and the impact of dynamic IP and NAT IP,the measure of botnet's size was difficult. Removed duplicated and dynamic IP and identified NAT based on communication feature to estimate the size of botnet.Experiments were carried out for validation purposes.

Key words: botnet; communication; duplicated removal; NAT; size measure

1 引言

僵尸网络(botnet)是一种从传统恶意代码形态进化而来的新型攻击方式，为攻击者提供了隐匿、灵活且高效的一对多命令与控制机制，可以控制大量僵尸主机(bot)实现信息窃取、分布式拒绝服务攻击和垃圾邮件发送等攻击目的。

僵尸网络主要分为 IRC 僵尸网络、HTTP 僵尸网络和 P2P 僵尸网络。IRC 僵尸网络是最早产生而目前仍然大量存在的一类僵尸网络，基于标准 IRC 协议在 IRC 聊天服务器上构建其命令与控制信道，控制者通过命令与控制信道实现对大量受控主机的僵尸程序版本更新、恶意攻击等行为的控制。

毫无疑问，僵尸网络是当今互联网安全最大的威胁之一这一事实得到了业界人士和研究者的一致认同。但是僵尸网络的威胁到底有多大，是否能够定量地描述僵尸网络的大小以及危害等问题在相关领域一直存在一定争议。很多论文介绍僵尸网络的规模，动辄上百万、千万。但是僵尸网络是否有如此大规模，或者说是否有如此多的大规模的僵尸网络存在？黑客是否有必要维护如此大规模的僵尸网络？首先是出于安全的考虑，随着僵尸网络的规模增大，其被监测与摧毁的可能性也大大增加；其次，随着僵尸网络肉机数量的增加，对其控制服务器自身的负荷是个考验；最后，大约需要仅仅 1Gbit/s 的攻击强度就足够让雅虎(yahoo.com)和

收稿日期：2010-08-10
基金项目：国家高技术研究发展计划基金资助项目（“863”计划）（2007AA01Z474，2006AA01Z451）
Foundation Item: The National High Technology Research and Development Program of China(863 Program)(2007AA01Z474, 2006AA01Z451)

亚马逊(amazon.com)这样的大型网站倒塌[1]，并不需要太大数目的肉机就可以完成。那么准确度量僵尸网络的大小的难点在于以下几方面。

僵尸网络大小的定义目前在学术界还不明确。目前关于僵尸网络大小的定义主要有：在给定时间周期内通过某种方法所监测到的僵尸网络肉机总数；僵尸网络实时在线并保持和僵尸网络控制服务器通信的肉机数；受僵尸网络控制者控制的肉机数。

动态 IP 地址的影响。由于互联网许多用户是 IDSL 上网，使用的是动态 IP，即每次上线，肉机所在 ISP 动态分配一个 IP 给这个肉机，每次上线所使用的 IP 一般是不同的，因此要准确度量僵尸网络大小，必须对动态 IP 地址进行去重，但是由于不同 ISP 在不同时间关于动态 IP 分配规则可能会有很大差异，并且出于客户数量保密的因素，各 ISP 动态 IP 分配的规则不对外公布，造成了要对动态 IP 地址进行准确的去重相当困难。更重要的是，动态 IP 地址在整个互联网 IP 地址中占有相当大的比重，文献[2]实验获取的 IP 地址中，超过半数(61.7%)是动态 IP 地址。

NAT IP 地址的影响。由于互联网 IP 地址紧缺，一些局域网内部网络采用 NAT(network address translation)技术，使多台计算机使用一个 IP 共享 Internet 连接，在局域网内部网络中使用内部地址，而当内部节点要与外部网络进行通信时，就在网关将内部地址替换成公用地址。如果不准确识别 NAT IP，并估算其包含的肉机数，对于僵尸网络的大小度量也会造成很大的误差。

IRC 僵尸网络的动态性。IRC 僵尸网络健壮性差，可通过摧毁单个 IRC 服务器来切断僵尸网络控制者与 bot 的联系。因此，僵尸网络控制者使用动态域名服务将僵尸程序连接的域名映射到其控制的多台 IRC 服务器上，一旦正在工作的 IRC 服务器失效，僵尸网络的受控主机会连接到其他的 IRC 服务器。这种行为称为僵尸网络的迁移。此外，出于安全及自身负荷的考虑，某些大型僵尸网络采用分层管理模式。准确度量僵尸网络大小必须考虑僵尸网络的迁移以及采用分层管理模式的情况。

本文的工作基于国家网络安全监测平台监测到的 IRC 僵尸网络服务器与 bot 间的命令与控制(C&C)通信数据。针对以上度量僵尸网络大小的难点，采用动态 IP 地址聚集将动态 IP 地址集合映射为肉机集合，和基于概率的动态 IP 地址去重方法消除动态 IP 地址的影响；采用基于通信频繁比对的 NAT IP 识别方法并估算 NAT 所含肉机数量消除 NAT IP 地址的影响，从而给出僵尸网络大小准确的度量。最后实验验证了各方法的有效性，并分析了僵尸网络各类大小的关系以及僵尸网络大小的分布情况。

2 相关研究工作

关于僵尸网络的大小度量，文献[3, 4]采用蜜罐技术模拟成肉机，进入 IRC 僵尸网络频道，从频道信息对僵尸网络进行深度分析，包括根据出现在频道上的肉机 ID 数推断任意时刻僵尸网络的在线肉机数和一定时间范围的僵尸网络足迹数；根据僵尸网络频道的返回信息直接获取僵尸网络大小信息。本方法的局限在于：针对这个方法，先进的 IRC 僵尸网络不向肉机返回其他肉机的信息，从而不能获取其他僵尸网络肉机信息。文献[5]采用 DNS 重定向的方法，即修改 DNS 配置，让肉机重定向到蜜罐主机，这样能获得监测时间内的不同肉机 IP 数，但是由于动态 IP 地址的影响，实际的肉机数远小于 IP 计数，并且这个方法不能获得在任意一个时刻在线的肉机数。Rajab[6]等人对“僵尸网络的规模监测和估计”这一重要问题进行了细致的探讨。将僵尸网络生命周期内任意时间点上感染的全部僵尸主机数量定义为僵尸网络的全局足迹，而将特定时间点接受僵尸网络命令与控制信道控制的在线僵尸主机数量定义为僵尸网络的实时规模。David Dagon[7]等人从僵尸网络的效力、效率、鲁棒性等方面度量了僵尸网络。文献[8]提取僵尸网络恶意攻击行为（如垃圾邮件发送）的特征，根据其特征识别僵尸网络，并且度量其大小。文献[9]通过邮件 IP 处理了动态 IP 并度量了大小。文献[2]专门讨论了动态 IP 地址问题。目前关于僵尸网络大小度量的研究主要存在的问题是，侧重点在于提出僵尸网络大小度量所面临的问题，而非实际解决问题；很难通过实验和理论证明验证所提大小度量方法的有效性。

3 僵尸网络大小度量

对于僵尸网络的大小，目前学术界还没有严格的定义，本文仍然借鉴文献[6]给出带时间戳的僵尸网络大小 3 个定义如下。

足迹(footprint)大小 $Fbot_i$，它是在给定时间周

期内，通过某种方法所监测到的僵尸网络肉机总数。其中 i 为僵尸网络标号，在不引起歧义的情况下，省掉 i。考虑到动态 IP 的影响，这里定义的足迹大小是指对动态 IP 地址去重之后的肉机数量。

在线(online)肉机数量 $Obot_i(t)$，即实时在线，并与僵尸网络控制服务器通信的肉机数量。

受控(control)肉机数量 $Cbot_i(t)$，即受僵尸网络 i 控制的肉机数量。随着新的主机被感染，肉机被打上补丁，以及网络抓机，受控肉机数量也是随时间变化的函数。

3.1　国家网络安全监测平台

863-917 网络安全监测平台[10]是国家“863”计划设立的网络安全应急项目（917 工程）建设的网络安全监控平台，是保障国家网络安全和网上重要信息系统安全的重要监测平台，由 CNCERT/CC 负责建设并运行。

863-917 网络安全监测平台底层为网络型 IDS 系统，实时监测我国互联网中僵尸网络、木马通信事件。采用协议与结构相关的僵尸网络检测方法，利用蜜网蜜罐获取僵尸网络信息提取僵尸网络报文级通信特征，在国家重要路由器节点部署网络型 IDS，对路由报文使用特征匹配检测僵尸网络 C&C 通信，检测到的僵尸网络 C&C 通信包括僵尸程序登录 IRC 频道后为与控制服务器保持连接而定期与控制服务器之间的 PING、PONG 命令[11]、IRC 服务器与 bot 间的控制命令与通告消息等。

3.2　动态 IP 地址去重

通过国家网络安全监测平台监测到的僵尸网络控制服务器与肉机间的通信数据，可以初步计算僵尸网络的 3 种大小。但是，互联网上众多 ADSL 上网的主机没有一个固定的 IP，当主机联网，互联网服务提供商(ISP)从一个 IP 库中对其随意分配一个未经使用的 IP 地址。这一 IP 地址只会在该主机上网的时间段中保留，下一次上线可能分配不同的 IP 地址。动态 IP 地址问题导致度量僵尸网络大小时有很大误差。因此，要准确度量僵尸网络的大小，必须对动态 IP 地址进行去重。

文献[12]对僵尸网络进行同源判别时，对僵尸网络的肉机 IP 进行了动态 IP 地址聚集，不适为一种动态 IP 地址去重的方法，动态 IP 地址聚集方法描述如下：

动态 IP 地址聚集即将 bot 的 IP 地址集合映射为 bot 集合。bot IP 聚集理想的结果是每一个 bot 使用过的 IP 聚集到同一个集合，不同 bot 对应聚集后的集合不同，即聚集后的集合与 bot 集合一一对应。对于给定的僵尸网络，设其 bot 集合为 B，$B=\{b_1,\cdots,b_n\}$，bot 数量为 n，即 $|B|=n$。这些 bot 使用过的 IP 地址集合为 I，$|I|=m, m\geqslant n$，$f(B)=I$，f 为 B 到 I 的 1 对多映射。

IP 地址是 4 个小数点隔开的十进制整数，考虑到 ISP 给 bot 主机动态分配的 IP 地址集合具有局部性，对 bot IP 地址进行聚集操作，去掉 IP 地址的小数点间隔的第 4 部分，这样的操作记作映射 g。

考虑到僵尸网络肉机分布很稀疏，有 2 个假设：对于给定的僵尸网络，它在任何一个 ISP 中最多有一个肉机；含有该僵尸网络肉机的任意 2 个 ISP 不属于同一网段，若某个 ISP 拥有的 IP 超过一个网段，则它每次分配给某个肉机的 IP 都属于同一网段。这 2 个假设在实际中可能并不严格成立，聚集的结果可能也有误差，但把属于同一网段的动态分配的 IP 聚成一个集合，聚集后的 IP 计数会更接近肉机数。基于这 2 个假设，对 IP 进行同一网段聚集。容易证明以下定理：

定理 1　若 $\forall IP_i, IP_j \in f(b_k), (k=1,\cdots,n)$，$g(IP_i)=g(IP_j)$，则 $|g(I)|\leqslant|B|$；

若 $\forall IP_i \in f(b_k), \forall IP_j \in f(b_l)(k,l=1,\cdots,n, k\neq l), g(IP_i)\neq g(IP_j)$，则 $|g(I)|\geqslant|B|$。

由定理 1 得到如下定理。

定理 2　若 $\forall IP_i, IP_j \in f(b_k), (k=1,\cdots,n)$，$g(IP_i)=g(IP_j)$，$\forall IP_i \in f(b_k), \forall IP_j \in f(b_l)(k,l=1,\cdots,n,k\neq l)$，$g(IP_i)\neq g(IP_j)$，则 $|g(I)|=|B|$。

根据定理 2 的假设，对僵尸网络的足迹 (给定监测时间内所监测到的 bot IP)即集合 I 进行聚集操作，得到 $g(I)$，它与 bot 集合一一对应，计算僵尸网络间 bot 的重叠率以此来度量僵尸网络的相似性。记僵尸网络 A，B 的重叠率为 $S(A,B)$，则

$$S(A,B)=\max\left(\frac{|g(I_A)\cap g(I_B)|}{|g(I_A)|},\frac{|g(I_A)\cap g(I_B)|}{|g(I_B)|}\right)$$

3.3　NAT IP 识别及其所含肉机数估算

由于互联网 IP 地址紧缺，一些局域网内部网络采用 NAT 技术，使多台计算机使用一个 IP 共享 Internet 连接，在局域网内部网络中使用内部地址，而当内部节点要与外部网络进行通信时，就在网关

将内部地址替换成公用地址。NAT IP 地址的存在，进一步影响了肉机和 IP 地址的一一对应关系，如果不准确识别 NAT IP，并估算其包含的肉机数，对于僵尸网络的大小度量也会造成很大的误差。为识别 NAT IP，首先提取僵尸网络通信频率特征日周期曲线[12]如下：

统计 n 天的数据，计算僵尸网络通信频率日周期函数 $CF'(t)$ $(0 \leqslant t \leqslant 24h)$ 如下：

1) 把每天的通信数据分成 24h/w 份（w 为统计时间间隔大小，它的含义是：认为在 w 间隔内有通信的 IP 数为该时间跨度内在线肉机数 $Obot(t)$，根据僵尸网络 IRC 服务器与 bot 通信数据的特点，本文 w 取 10min），每一份时间跨度为 w，计算每一份数据中不同 IP 个数，得到在线肉机函数 $Obot(t)$ 的统计值。

2) 选定时间窗口大小 w 为 10min，把每天的数据分成 24h/w 份，每一份时间跨度为 w，统计每一份数据的不同 IP 数（由于时间间隔小，消除了动态 IP 的影响），得到在线肉机函数 $Obot(t)$ 的统计值，认为在 w 间隔内有通信的 IP 数为该时间跨度的在线肉机数 $Obot(t)$。

3) 计算通信量 $CC(t)$，时间间隔取 10min。

4) 计算通信频率函数 $CF(t)=CC(t)/Obot(t)$，即单位肉机的通信量函数。

5) 对于 $CF(t)$，需要进行一些预处理：

$CC(t)$ 和 $Obot(t)$ 同时为 0：$CF(t)$ 不应该设为 0 和 1，2 个值都不能真实反映这个僵尸网络正常的通信特征。自然而然地应该使用插值的方法，本文使用线性插值。

$CF(t)$ 异常：分析发现，$CF(t)$ 会在某些时间段异常大（异常小还没发现），本文认为异常值不能反映这个僵尸网络正常的通信特征，应该剔除掉，同时记录下这些时间段以进一步的分析。分析这些异常点是否代表在这段时间该僵尸网络有特别的活动。$CF(t)$ 异常检测详细见后面论述。判断 $CF(t)$ 异常大的方法为：计算 $CF(t)$ 的平均值 $AVE(CF(t))$，若 $CF(t) > \beta \times AVE(CF(t))$。则认为 $CF(t)$ 异常（β 为阈值，本文设定为 2）。

6) 平均 n 天的数据，得到 $CF'(t)(0 \leqslant t \leqslant 24\text{h})$。

通过理论分析与实验易知，僵尸网络肉机中 NAT IP 地址的通信频率明显大于所属僵尸网络的通信频率。如图 1 所示，图中 2 曲线分别为僵尸网络通信频率和该僵尸网络中某 NAT IP 的通信频率变化曲线。

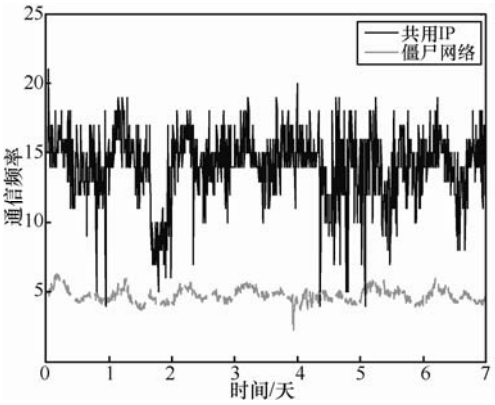

图 1 共用 IP 与其所属僵尸网络通信频率曲线对比图

3.3.1 NAT 识别算法

对于僵尸网络 botnet 和其中的某一 IP，要识别该 IP 是否是 NAT IP。记 IP 的平均通信频率为 $AVE(CF'_{\text{IP}}(t))$，该僵尸网络的平均通信频率为 $AVE(CF'_{\text{botnet}}(t))$。首先选取该僵尸网络通信时间跨度短，即动态 IP 若干个，因为它们代表了这个僵尸网络的正常通信频率。计算它们各自的平均通信频率，实验结果发现，其值近似服从正态分布 $N(\mu,\delta^2)$，计算这些值与 $AVE(CF'_{\text{botnet}}(t))$ 差值的累积密度函数（CCDF）。判别阈值 $\eta = P(\left|x - AVE(CF'_{\text{botnet1}}(t))\right| > \eta) < \delta$，若 $AVE(CF'_{\text{IP}}(t)) > \eta$，则该 IP 为 NAT IP。

3.3.2 NAT 包含肉机数估算方法

若某 IP 是 NAT IP，则估算其所含肉机数。NAT 所含肉机数量分析由于受很多因素的影响，是件很难的工作。笔者把问题简化，建立数学模型如下。

不妨假设每台肉机 Bot_i 的通信频率 $CF_i(t)$ 服从分布：$CFS(t)\alpha_i(t)$，其中 $CFS(t)$ 为通信频率日周期函数，$\alpha_i(t)$ 为该肉机在线判断二值函数，$\alpha_i(t)=0$ 表示该肉机不在线，$\alpha_i(t)=1$ 表示肉机在线。

再假设 NAT 的通信频率 $CF_{NET}(t)$ 服从分布：

$$\frac{E(CFS_{\text{NET}}(t))}{E(CFS_{\text{normal}}(t))}CFS(t)$$

假设该 NAT 含肉机数为 N，则有：

$$\sum_{i=1}^{N} CFS(t)\alpha_i(t) \sim \frac{E(CFS_{\text{NAT}}(t))}{E(CFS_{\text{normal}}(t))}CFS(t)$$

在上式中，要计算 N，可以简化为：

$$N = \frac{E(CFS_{\text{NET}}(t))}{E(CFS_{\text{normal}}(t))} \times \frac{1}{ODbot}$$

其中，*ODbot* 为肉机在线密度（online density）。

4　实验与分析

4.1　实验设置

为了验证所提方法的有效性，使用国家网络安全监测平台 2009 年 4 月 1 日至 2009 年 5 月 30 日 60 天内监测到的 723 个僵尸网络通控制服务器与肉机间的通信数据进行了实验分析，实验环境为 Inter Core2 Duo CPU T9550 (2.66GHz)、2GB 内存、Window XP，算法在 C++下实现。

4.2　动态 IP 地址去重实验结果

表 1 为国家网络安全监测平台 2009 年 4 月 1 日至 2009 年 5 月 30 日 60 天内监测到的部分僵尸网络的足迹大小，经过动态 IP 地址去重方法去重前后的估计值对比。

表 1　botnet1~botnet4 的动态 IP 聚集结果

僵尸网络	聚集前 IP 数	聚集后 IP 数
Botnet1	42778	8532
Botnet2	5124	1105
Botnet3	35127	7108
Botnet4	31883	7024

由于 60 天内所监测到的僵尸网络部分只含有很少的肉机通信数据，为了更好地统计效果，选取其中 700 个通信数据较多的僵尸网络进行进一步的实验。图 2 为选取的 700 个僵尸网络动态 IP 地址聚集前后足迹大小比例曲线图。该比例平均值为 4.724，并且 700 个数据的方差为 0.297。

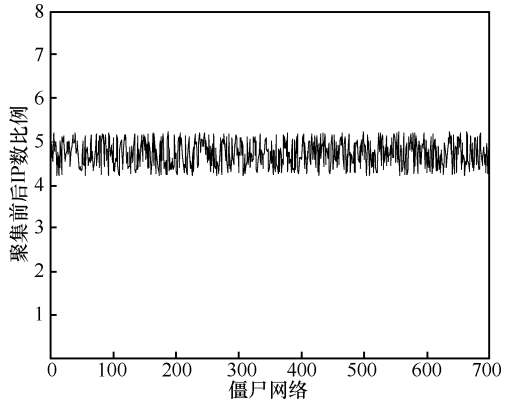

图 2　僵尸网络动态 IP 地址去重前后 IP 数比例曲线

僵尸网络受控肉机数估算：以 botnet1 举例计算其受控肉机数 $Cbot_1(t)$，计算 $Cbot_1(t)$ 的难点是区分肉机是失去控制或者不在线，由于肉机和互联网一般用户本质上没有区别，本文利用互联网用户上线统计数据得到互联网用户不上线间隔时间分布函数 $\Gamma(x)$。设置信率为 η，取这样的 n，使 $p(\Gamma(x)>n)=1-\eta$，这个公式的意义是肉机（互联网用户）有 η 的可能在 n 天内没有上线，反过来说，若一个肉机 n 天没有通信，则有 η 的脱离僵尸网络控制了，可能是被用户打上补丁了、查杀了僵尸程序，可能是断网，可能是受其他僵尸网络所控制。

本文试验取 n=3，得到 $Cbot_1(t)$ 的平均值为 2 054，通过以上方法，计算得到该僵尸网络的平均在线肉机数 $AVE(Obot_1(t))$ 为 170。

记肉机平均在线密度为 *ODbot*，即肉机平均有多少比例时间在线，同样，可以通过互联网用户统计数据得到这个值。则有 $AVE(Obot(t))=AVE(Cbot(t))\times ODbot$。通过上式反算得到 *ODbot* 为 8.28%，即得到 botnet1 的肉机平均每天在线约 2h。

4.3　僵尸网络各大小关系及其分布

关于僵尸网络 3 个大小的关系：图 3 为 700 个僵尸网络足迹（时间范围 60 天）与最大肉机数比例曲线。

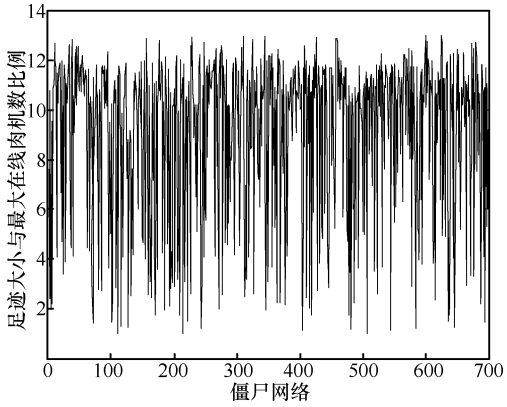

图 3　700 个僵尸网络足迹与最大在线肉机数比例

从图 3 可以看出，部分僵尸网络足迹与最大肉机数比例接近于 1，显然误差较大。通过对应的僵尸网络通信数据分析可以知道，这些僵尸网络的通信数据较少，由于样本数量的关系造成了误差。因此，进一步的减小僵尸网络数量为 500 个。图 4 为选定的 500 个僵尸网络足迹与最大肉机数比例曲线。其数据均值为 10.904，方差为 1.052。

图 5 为 700 个僵尸网络足迹（时间范围 60 天）与最大受控肉机数比例曲线。

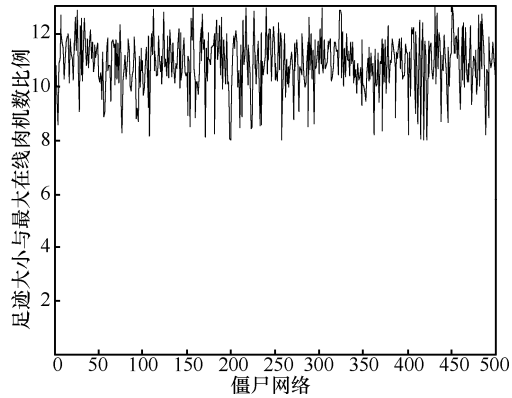

图 4 500 个僵尸网络足迹与最大在线肉机数比例

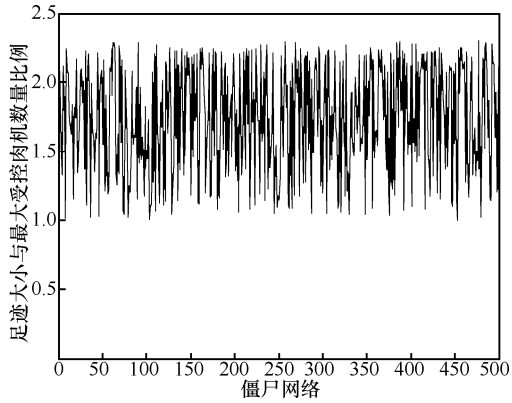

图 5 700 个僵尸网络足迹与最大受控肉机数比例

同样的考虑，选取其中 500 个通信数据量较大的僵尸网络进行考虑，其足迹与最大受控肉机数比例如图 6 所示。其数据均值为 1.908，方差为 0.2689。

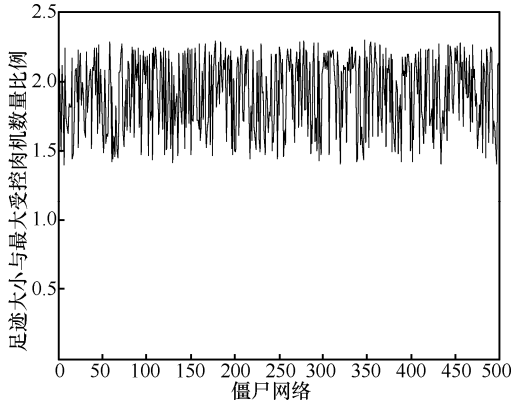

图 6 500 个僵尸网络足迹与最大受控肉机数比例

关于僵尸网络大小分布，图 7 和图 8 是文献[6]关于僵尸网络大小度量的结果。图 7 是采用分布式蜜罐收集系统收集的 9 个月的数据，蜜罐是模似成肉机，进入 IRC 僵尸网络频道，从频道信息得到的不同肉机 ID 数分布互补累积分布函数 CCDF。图 8 是用同样的方法而得到的最大在线肉机 ID 数分布 CCDF。

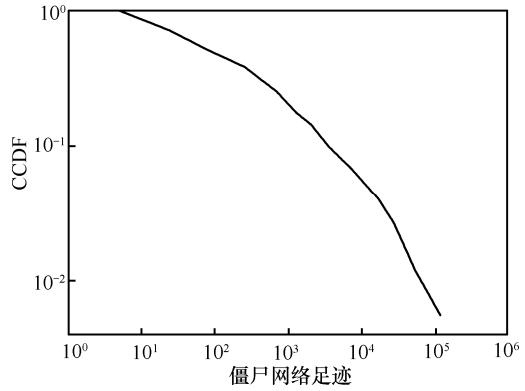

图 7 僵尸网络足迹大小分布

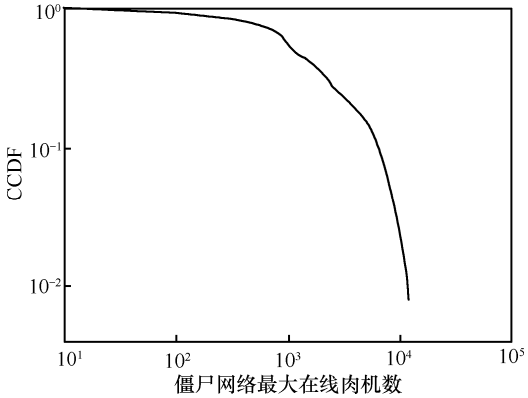

图 8 僵尸网络最大在线肉机数量分布

图 9 是本文实验数据集不进行动态 IP 去重而得到的僵尸网络足迹大小分布 CCDF。

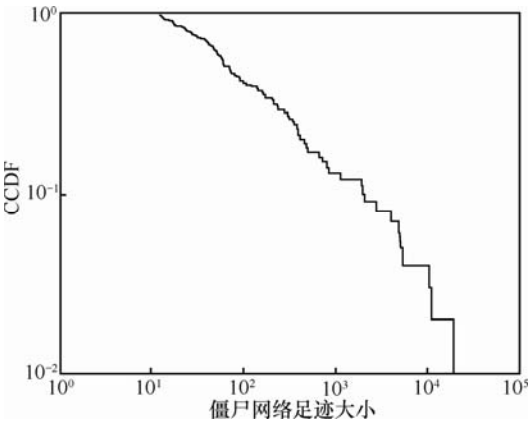

图 9 不进行动态 IP 去重后僵尸网络足迹大小分布

图 10 是采用动态 IP 去重方法所得到的僵尸网络足迹大小分布 CCDF。

5 结束语

本文采用动态 IP 地址去重算法和基于通信频繁比对的 NAT IP 识别算法，对僵尸网络的大小进行准确的度量。下一步的工作主要有：

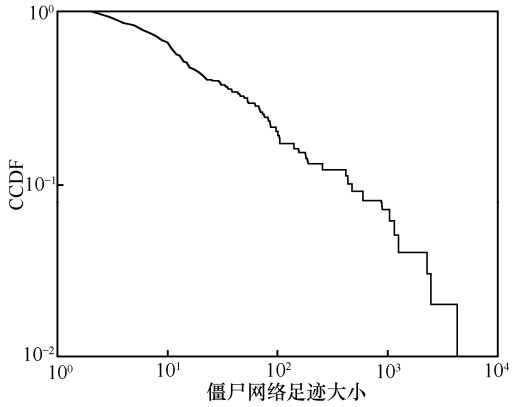

图 10　动态 IP 去重后僵尸网络足迹大小分布

1) 研究僵尸网络足迹，包括时间、在线肉机数、受控肉机数之间的关系，试图用解析表达式表达，从而根据某一项可以推导出其他项的值。

2) 在相应的有效的监测方法条件下，研究 HTTP，P2P 僵尸网络的大小度量模型。

参考文献：

[1] Basis for denial of service attacks. SKY_Server[EB/OL]. http:// wenku. baidu.com/view/b2544d37ee06eff9aef80797.html.

[2] XIE Y, YU F, ACHAN K, *et al*. Wobber, how dynamic are IP addresses?[A]. ACM SIGCOMM Computer Communication Review[C]. 2007. 312-322.

[3] FREILING F C, HOLZ T, WICHERSKI G. Botnet tracking: exploring a root-cause methodology to prevent distributed denial-of-service attacks[J]. Lecture Notes in Computer Science, 2005, 3679: 319-335.

[4] RAJAB A. A multifaceted approach to understanding the botnet phenomenon[A]. Proc of the 6th ACM Internet Measurement Conf (IMC 2006)[C]. Riode Janeriro: ACM, 2006.41-52.

[5] DAGON D, ZOU C, LEE W. Modeling botnet propagation using time zones[A]. Proc of the 13th Annual Network and Distributed System Security Symp (NDSS 2006)[C].2006. 234-242.

[6] FABIAN M, TERZIS M A. My botnet is bigger than yours (maybe, better than yours): why size estimates remain challenging[A]. Proc of the 1st Workshop on Hot Topics in Understanding Botnets (HotBots 2007)[C]. 2007. 111-119.

[7] DAGON D, GU G, LEE C, *et al*, A taxonomy of botnet structures[A]. Proc of the 5th ACM Internet Measurement Conf. (IMC 2006)[C]. Riode Janeriro: ACM, 2005. 51-60.

[8] LI Z, HU J, HU Z, *et al*. Measuring the botnet using the second character of bots[J]. Journal of Networks, 2010, (5): 98-105.

[9] ZHUANG L, DUNAGAN J, SIMON D R, *et al*. Characterizing botnets from email spam records[A]. Proc of the 13th Annual Network and Distributed System Security Symp (NDSS 2008)[C]. 2008. 189-199.

[10] 国家计算机网络应急技术处理协调中心(CNCERT/CC),全国网络与信息技术培训项目管理中心(NTC-MC).网络安全应急实践指[M]. 电子工业出版社, 2008.
National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), National Certificate of Network and Information Technology-Management Center of China(NTC-MC)[M]. Emergency and Practice Guideline of Network Security(NTC-MC)[M].Publishing House of Electronics Industry, 2008.

[11] CHEN Y. IRC-based botnet detection on high-speed routers[A]. Proceedings of the ARO-DARPA-DHS Special Workshop on Botnets Arlington, VA, 2006. 105-117.

[12] 李润恒, 王明华, 贾焰. 基于通信特征提取和 IP 聚集的僵尸网络相似性度量模型[J]. 计算机学报, 2010, 33: 45-54.
LI R H, WANG M H, JIA Y. Modeling botnets' similarity based on communication feature extraction and IP assembly[J]. Chinese Journal of compute, 2010, 33: 45-54.

作者简介：

李润恒（1982-），男，四川乐山人，国防科学技术大学博士生，主要研究方向为网络安全和数据挖掘。

甘亮（1977-），男，江西南昌人，国防科学技术大学博士生，主要研究方向为网络安全和数据分析。

贾焰（1960-），女，四川成都人，博士，国防科学技术大学教授、博士生导师，主要研究方向为信息安全和数据库。