ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.20, No.4, April 2009, pp.845−857
doi: 10.3724/SP.J.1001.2009.03337

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

有效扫描监测系统建模与部署*

马莉波[1+], 李 星[1,2], 张 亮[3]

[1](清华大学 电子工程系,北京 100084)

[2](清华大学 网络与信息工程研究中心,北京 100084)

[3](中国科学院 计算技术研究所,北京 100190)

On Modeling and Deploying an Effective Scan Monitoring System

MA Li-Bo[1+], LI Xing[1,2], ZHANG Liang[3]

[1](Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

[2](Research Center of Network Information and Engineering, Tsinghua University, Beijing 100084, China)

[3](Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: malibo@ccert.edu.cn

Ma LB, Li X, Zhang L. On modeling and deploying an effective scan monitoring system. *Journal of Software*, 2009,20(4):845−857. http://www.jos.org.cn/1000-9825/3337.htm

Abstract: Constructing an effective scan monitoring system is a necessary step for early detection and warning of unknown threats. Scan monitoring systems constructed by routable unused IP addresses will be more effective than those deployed in active networks for their special advantages in identifying threats precisely which results in low false alarm rate. Nowadays systematic researches on how to deploy such an effective monitoring system are still missing. This paper presents a novel scan monitoring model based on BGP route distribution to answer two practical deployment questions. One is how to design and deploy an ideal target-specified scan monitoring system and the other is how to evaluate the detecting effectiveness of actual limited deploying resources. On the basis of the model, this paper puts forward a new concept of deployment threshold which describes the most economical matching value between the monitoring system's scale and the scanner's scanning width on the same detection probability demand. According to the model and the deployment threshold, an effective monitoring system can be designed and appropriate detecting targets can be proposed which match the practical deploying resources to avoid blind deployment as before. Simulation results are coincident with the theretical analyses.

Key words: scan detection; scan monitoring system deployment; scan monitoring model; deployment threshold

摘 要: 构建有效的扫描监测系统是早期检测和预警未知威胁的必要措施.利用网络中未使用IP地址空间构建扫描监测系统,具有检测准确、虚警率低等活动网络监测不可实现的优势,是一种非常有效的实现方式.针对利用未使用IP地址实际部署有效扫描监测系统缺乏理论指导这一现状,提出一种新的基于路由分布的扫描监测模型,用于解决针对特定目标的有效扫描监测系统设计部署以及实际有限部署资源检测效用评估问题.基于模型提出部署阈值的概念,描述相同检测率要求下扫描监测系统规模与扫描源扫描宽度之间最经济的匹配阈值.基于路由分布的扫描

* Supported by the National Natural Science Foundation of China under Grant No.90412010 (国家自然科学基金)
Received 2007-11-19; Accepted 2008-04-07

监测模型和部署阈值,可为设计与实际部署资源相匹配的监测系统部署方案以及制定合理的检测目标提供理论参考,避免原有凭经验的盲目部署.仿真实验结果与理论分析结论相一致.

关键词: 扫描检测;扫描监测系统部署;扫描监测模型;部署阈值

中图法分类号: TP393 **文献标识码:** A

网络扫描是 0-day 攻击、僵尸网络和蠕虫等恶意行为寻找漏洞主机、扩大入侵规模的必备手段,因此,构建有效的扫描监测系统是早期检测和预警未知威胁的必要措施,也是主动安全防护体系[1−3]的有机组成部分.近年来,利用可路由网络中未使用 IP 地址空间进行网络安全事件监测已得到认可和应用[4−11].由于网络扫描不可避免地要访问到网络中未使用的 IP 地址空间,而任何进入未使用地址空间的行为均可认为是恶意或未经授权的[4],因此,利用网络中未使用 IP 地址空间构建扫描监测系统,具有检测准确、虚警率低等活动网络监测不可实现的优势,是一种非常有效的实现方式.

在利用未使用IP地址实际部署扫描监测系统时,遇到两个问题:一是不考虑部署地址资源限制时,针对特定检测目标的理想监测系统如何构建?二是由于实际可控的部署地址资源是有限的,那么如何事先评估这些有限部署资源的检测效用以决定是否值得构建一个监测系统?这些问题实质上是监测系统构建中监测点数目、大小、部署位置与扫描源以及监测系统检测效用之间的关系问题,但目前对此问题的系统研究还很少.在扫描检测方面,研究工作避开了底层监测系统的部署问题,研究重点在于系统构建之上的扫描源判定方法研究[11−14];在监测系统部署方面,文献[7–9]均侧重实际已构建监测系统的数据分析,并未解释监测系统为什么如此构建.与本文研究最相近的工作是文献[10],率先从理论上系统地分析了上述关系问题,但该研究基于的假设是已知或者部分已知全球漏洞主机密度,这在实际监测系统构建中根本无法获取,研究结果难于应用解决实际所面临的问题.国内尚未见到对此问题的深入研究.为此,本文提出一种新的基于路由分布的扫描监测模型,细致刻画和量化监测点数目、大小、部署位置与扫描源和监测系统检测率之间的关系,作为解决实际部署问题的理论依据.基于路由分布对扫描监测系统建模的原因在于:1) 为了保证监测系统监测有效,必须保证扫描源与监测点之间路由可达;实际网络设备中,如路由器,自行过滤扫描源发往可路由地址空间之外的分组包,因此无须在可路由地址空间之外考虑部署监测点;2) 全球路由分布可根据公开发布的 BGP 路由表[15,16]获得,较全球漏洞主机分布,其获取方式简便、可行;3) 全球路由分布的非均匀性为监测点的非均匀部署提供了部署索引.全球漏洞主机分布是全球路由分布的子集,虽然基于路由分布研究监测系统部署问题以部署地址资源的增加为代价,但与研究结果的可行性相比,这种代价是值得考虑的.

根据模型,首次提出部署阈值概念,描述相同检测率要求下监测系统规模与扫描源扫描宽度属性相匹配的最经济阈值,该阈值表明,监测系统到达一定规模后,无须继续扩大系统规模便可满足检测需求,因此,该部署阈值以及对应的扫描宽度阈值可作为理想监测系统规模和检测目标制定的选择依据;其次,针对应用最广泛的全网随机和本地优先扫描策略,详细分析了构建理想扫描监测系统的部署需求以及可以达到的检测目标;最后以现有规模的实际 Leurre’Com Honeypot Project[5](简称 LHP)全球分布式监测系统为样例,评估其对单点扫描源的检测效用,考察 LHP 扫描监测系统是否适于本文的应用检测目标.通过以上系统分析和评估,可以很好地解决实际监测系统部署所面临的如何构建和能否构建的问题,为设计与实际部署资源相匹配的监测系统部署方案以及制定合理的检测目标提供理论参考,避免了原有凭经验的盲目部署.实验仿真结果与理论分析结论相一致,验证了基于路由分布扫描监测模型的正确合理性以及对监测系统部署问题的指导作用.

本文第 1 节是建模预备知识.第 2 节是基于路由分布的扫描监测模型描述.第 3 节是模型在理想监测系统构建和监测系统效用评估方面的应用.第 4 节是仿真实验验证.第 5 节是总结与展望.

1 预备知识

1.1 扫描源属性描述

用扫描宽度和扫描频度作为扫描源扫描行为的属性描述.扫描宽度是扫描源访问不同目标地址的连接数

目,刻画了扫描范围;扫描频度是扫描源对每一个目标 IP 地址发出的扫描连接包数,刻画了扫描频次.如图 1 所示.用平均扫描速率描述扫描源在单位时间内访问不同目标地址的连接数目,它与扫描宽度的关系是

扫描宽度=平均扫描速率×时间.

监测系统的检测效用与扫描源的扫描宽度密切相关.

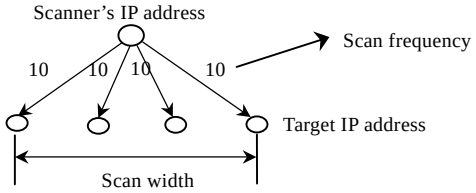


Fig.1 Attribution description of a scan source

图 1 扫描源的属性描述

1.2 扫描源的扫描策略

扫描源采用不同的扫描策略寻找入侵主机[17].本文关注两种实际应用最为广泛的扫描策略:全网随机和本地优先扫描策略,而采用相应策略的扫描源称为随机扫描源和本地优先扫描源.根据实际出现的扫描源样例,本地优先扫描策略描述如下:

- 以 $P_{16}(0\le P_{16}\le 1)$的概率选取与扫描源自身 IP 地址前/16 比特相同的目标地址;
- 以 $P_8(0\le P_8\le 1)$的概率选取与扫描源自身 IP 地址前/8 比特相同的目标地址;
- 以 $1-P_{16}-P_8$ 的概率全网随机选取目标地址.

全网随机扫描策略是本地优先扫描策略在 $P_{16}=P_8=0$ 时的特例.表 1 列出一些实际已知典型蠕虫扫描源的扫描策略.需要说明的是,本文虽然使用典型蠕虫扫描源作为分析样例,但并不表示本文研究结论只适用于蠕虫扫描源,而是普适于广泛采用上述扫描策略的扫描源,如僵尸网络的 Bot 主机、寻找目标的 0-day 攻击主机等.

Table 1 Known typical worm scanners

表 1 已知典型蠕虫扫描源

Scanning policy	Scanner's name	P_{16}	P_8	$1-P_{16}-P_8$	Average scan rate (scan width per minute)
Random scanning	Codered[18]	0	0	1	358
	Slammer[19]	0	0	1	2.4×10^5
Localized scanning	Sasser[20]	0.25	0.25	0.5	128/1024
	Codered II[21]	0.375	0.5	0.125	300/600
	Blaster[22]	0.4	0	0.6	20
	Nimda[23]	0.5	0.25	0.25	128

1.3 扫描目标网络分类

根据扫描源采取的扫描策略将整个 IPv4 地址空间划分成不同的扫描目标网络,以方便估算监测点分别部署于不同目标网络时需要的监测点数目.见表 2.

Table 2 Classification of the scanning target network

表 2 扫描目标网络分类

Scanning policy	Number of scanning target networks	Description of scanning target networks	Size of scanning target networks	Scanning probability
Random scanning	1	The global network	2^{32}	1
Localized scanning	3	Subnets having the same first two octets with a scanner	2^{16}	$P_{16}+P_8\times1/2^8+(1-P_{16}-P_8)\times1/2^{16}$
		Subnets having the same first octet with a scanner excluding above /16 subnets	$2^{24}-2^{16}$	$P_8+(1-P_{16}-P_8)\times1/2^8$
		Other /16 subnets	$2^{32}-2^{24}$	$1-P_{16}-P_8$

1.4 监测点部署网络单元

根据 Route Views Project[15]和亚太网络信息中心(APNIC)[16]发布的全球 BGP 路由表统计信息,2006-08~2007-08 这 12 个月内,网络数目在不同网络前缀长度下的平均分布如图 2 所示.可知全球可路由网络前缀长度集中于 16~24 之间,前缀长度为 24 的网络数目占网络总数目的 53%,前缀长度大于 24 的网络数目仅占网络总数目的 1.5%.任何大于/24 的网络均可按照/24 网络大小进行划分,因此选取/24 网络作为监测点部署网络单元.

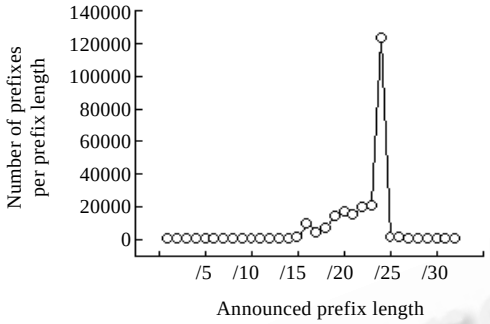


Fig.2 Distribution of the number of prefixes per prefix length

图 2 不同网络前缀长度下的网络数目分布

1.5 路由分布

Route Views Project 和亚太网络信息中心(APNIC)发布的全球 BGP 路由表统计信息表明,并非整个 IPv4 地址空间均是可路由的.截止到 2008 年 4 月,Internet 上公告的可路由地址总数目为 1 857 581 472,约占整个 IPv4 地址空间的 43%[24].按照 IP 地址所在的/8 网络前缀编号进行索引,每个/8 网络中包含的可路由地址空间大小不同.图 3、图 4 给出根据 Route View Project 2007-09-29 BGP 路由表统计得出的每个/8 网络内公告的/16、/24 网络数目分布.图示表明,每个/8 网络内可路由网络数目分布是不均匀的;即使某个/8 网络内所有/16 网络均路由可达,也并不表明/8 网络内所有/24 网络路由可达.结合第 1.4 节分析以及本文的研究重点在于监测点在哪里部署、部署多少的问题,我们认为能够到达/24 网络的路由统计信息足以支持扫描监测系统部署单元位置的选择和数目估算,因此在建模中,采用的路由分布是指每个/8 网络中包含的可路由/24 网络数目分布,以此作为监测点部署单元的选取指示.而在本文实验验证部分,将 Route View Project 公布的 BGP 路由表具体路由信息写入数据库中模拟实际可路由地址空间,通过仿真实验对模型部署估算结果进行验证.

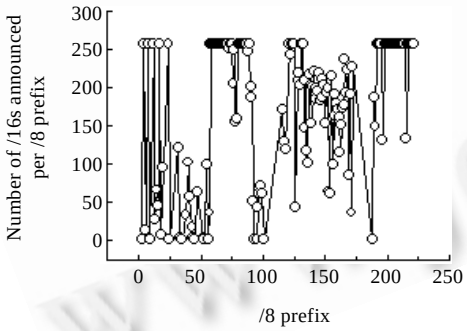


Fig.3 Number of /16s announced per /8 prefix

图 3 每个/8 网络中公布的/16 网络数目

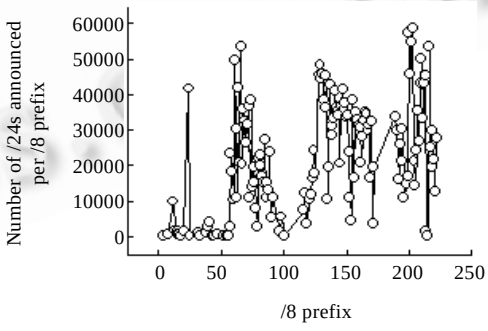


Fig.4 Number of /24s announced per /8 prefix

图 4 每个/8 网络中公布的/24 网络数目

1.6 模型符号和说明

表 3 给出基于路由分布的扫描监测模型使用符号和说明.

Table 3 Model notions and descriptions

表 3 模型符号和说明

Notion	Name	Description
Ω_s	Scan space	Size of the IP address space accessed by a scanner
Ω_{rm}	Monitoring system scale	Sum of the product of monitors' number and size in a monitoring system
P_d	Detection probability	The probability of a scanner detected by a monitoring system
P_{Ω_s}	Scanning probability	The probability of a scanner accessing a scan space
m	Prefix number of /8 subnets	Decimal prefix number of /8 subnets
$P_{r8}(m)$	Routing probability	The probability of the mth /8 subnet is routed in a scan space
$N_{r8}(m)$	Number of /24 subnets	The number of public routable /24 subnets in the mth /8 subnet
k_m	Number of monitors	The number of /24 subnets allocated from the mth /8 subnet for building a monitoring system
n_{mi}	Size of a monitor	Number of unused IP addresses from the ith /24 subnet within the mth /8 subnet for building a monitoring system
N_d	Scan width	Different target IP addresses accessed by a scanner
s	Average scan rate	The average number of connections which a scanner accesses different targets in a time tick
t_d	Detection time	The elapsed time from the instant a scan source sends its first scan up to the point where at least one scan from that host is detected by any monitor in a monitoring system

2 基于路由分布的扫描监测模型描述

设全球可路由 IPv4 地址空间大小为Ω,首先按照/8 网络前缀对其进行划分,其次将每个/8 网络空间按照/24 网络大小进一步划分.设第 m 个/8 网络内包含的/24 网络数目为 $N_{r8}(m)$,m=0,1,2,…,2^8–1,$N_{r8}(m)$由事先根据 BGP 路由表中得到的统计信息确定,则

$$\Omega = \sum_{m=0}^{2^8-1} N_{r8}(m)\cdot 2^8 \tag{1}$$

设扫描源的扫描空间大小为Ω_s,扫描空间访问概率为 P_{Ω_s},则当扫描宽度为 N_d 时,至少有一个访问连接被监测系统检测到的概率为

$$P_d = 1-\left[1-\frac{P_{\Omega_s}}{\Omega_s}\cdot\sum_{m=0}^{2^8-1}P_{r8}(m)\cdot\sum_{i=1}^{k_m}n_{m_i}\right]^{N_d},\ k_m \le N_{r8}(m) \tag{2}$$

其中,$P_{r8}(m)$的取值由以下判决函数决定:

$$P_{r8}(m)=\begin{cases}1, \text{ if } m\text{th /8 network exists in route table}\\ 0, \text{ else}\end{cases} \tag{3}$$

设监测系统规模表示为

$$\Omega_{rm} = \sum_{m=0}^{2^8-1}P_{r8}(m)\cdot\sum_{i=1}^{k_m}n_{m_i},\ k_m \le N_{r8}(m) \tag{4}$$

则公式(3)表示为

$$P_d = 1-\left[1-P_{\Omega_s}\cdot\frac{\Omega_{rm}}{\Omega_s}\right]^{N_d} \tag{5}$$

根据扫描源平均扫描速率和扫描宽度的关系,公式(2)、公式(5)可表示为

$$P_d = 1-\left[1-\frac{P_{\Omega_s}}{\Omega_s}\cdot\sum_{m=0}^{2^8-1}P_{r8}(m)\cdot\sum_{i=1}^{k_m}n_{m_i}\right]^{s\cdot t_d} = 1-\left[1-P_{\Omega_s}\cdot\frac{\Omega_{rm}}{\Omega_s}\right]^{s\cdot t_d},\ k_m \le N_{r8}(m) \tag{6}$$

公式(2)称为基于路由分布的扫描监测模型.模型中,P_d描述整个监测系统对某个扫描源的检测率; P_{Ω_s},Ω_s描述扫描源的扫描策略;$P_{r8}(m)$,$N_{r8}(m)$将监测点限制于扫描空间中的可路由地址空间,从而限定了监测点所在的网络位置和可选数目;k_m 描述监测点数目,n_{mi} 描述监测点大小;N_d 描述扫描源扫描宽度属性.公式(5)是模型的宏观表达,其中,Ω_{rm} 描述了整个监测系统的规模大小,而Ω_{rm}/Ω_s 描述了监测系统规模在整个扫描空间中所占有的比例.整个模型细致刻画和量化了监测系统部署中监测点网络位置、监测点数目和大小、监测系统规模以及

扫描源属性、扫描策略等因素对监测系统检测率的影响.公式(6)是模型的时间表达方式,建立了检测时间与监测系统检测率之间的关系.由公式(5),可以估算对于特定扫描源的监测系统整体规模大小,再根据公式(4)可将监测点具体部署于可路由的网络中.表 4 列出基于路由分布扫描监测模型与文献[10]提出的监测模型情况对比.

Table 4 Comparison between route based scan monitoring model and model proposed by Ref.[10]

表 4 基于路由分布的扫描监测模型和文献[10]提出的监测模型比较

Model name	Research assumption	Difficulty to get	Attainability of research results
Scan monitoring model based on the route distribution	Global BGP route distribution	Easy	Attainable
Monitoring model proposed by Ref.[14]	Global vulnerable population distribution	Difficult	Unattainable

3 模型应用

实际监测系统部署面临两个主要问题:一是在不考虑部署地址资源限制时,理论上针对特定类型扫描策略理想监测系统如何构建的问题;二是在实际可控部署地址资源有限时,有限部署资源检测效用评估的问题.为此,依据本文所提出的模型,首先提出部署阈值概念并理论推导确定监测系统部署阈值;其次,针对应用最广泛的全网随机和本地优先扫描策略构建理想扫描监测系统;第三,对现有规模的实际 Leurre'Com Honeypot Project 全球分布式监测系统检测效用进行评估;最后对模型应用研究进行小结.

3.1 理想扫描监测系统构建

3.1.1 部署阈值的概念

设 $r=\Omega_{rm}/\Omega_s$ 表示监测系统规模在扫描空间中所占比例,由公式(6)可得

$$N_d=\frac{\log(1-P_d)}{\log(1-P_{\Omega_s}\cdot r)} \tag{7}$$

当监测系统检测效用 P_d 确定时,考察扫描宽度 N_d 与监测系统规模占比 r 之间的关系.由于 $\frac{dN_d}{dr}<0$,则 $\frac{d^2N_d}{dr^2}>0$,N_d 是 r 的单调减凹函数.如图 5 所示(P_d=0.95, $P_{\Omega_s}=1$).

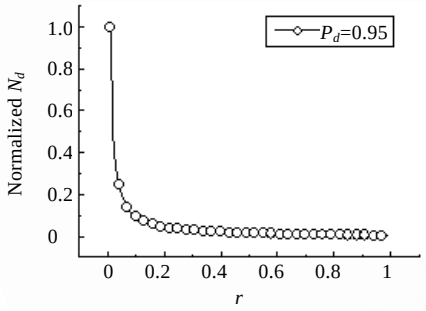


Fig.5 Relationship between N_d and r

图 5 N_d~r 关系曲线

如图 5 所示,N_d 值随 r 值的变化可以分成两个阶段:一是 N_d 值随 r 值的增大急速降低阶段,这表明 N_d 值敏感于 r 值的变化;二是 N_d 值随 r 值的增大缓慢减小阶段,这表明 N_d 值不敏感于 r 值的变化.两个阶段之间存在着变化阈值,称为部署阈值,用 r_{thd} 表示,与其对应的扫描宽度用 N_{d_thd} 表示.部署阈值的物理意义在于监测系统的检测率要求相同的条件下,监测系统规模在扫描空间中的占比达到 r_{thd} 时,能够最经济而有效地检测到扫描宽度≥N_{d_thd} 的扫描源.当监测系统规模在扫描空间中的占比超过 r_{thd} 时,监测系统规模的增大既不能有效降低检测所需的扫描宽度,也无益于提高整个监测系统的检测效用,因而无须进一步扩大系统规模,造成部署地址资源的浪费.因此,通过估算部署阈值可以得到经济适用的监测系统规模和扫描源扫描宽度值,为实际监测系统构建和检测目标制定提供理论参考.

3.1.2 部署阈值的理论估算

由于 N_d 是 r 的单调减凹函数,因此,部署阈值应出现在整条曲线最凹处.数学上用曲率表示一条曲线的弯曲程度.设 $y=f(x)$,则曲线的曲率为

$$z=\frac{|y''|}{(1+y'^2)^{3/2}} \tag{8}$$

按照曲率公式(8)计算公式(7)表示曲线的最大曲率值,设为 r_q.由于公式(7)表示曲线是单调减凹函数,因此曲率极值点唯一,且当 $0<r<r_q$ 时,随着 r 值减小,$|dN_d/dr|$值增大,表明监测系统的规模变化对扫描宽度的影响明显;而当$r>r_q$时,随着r值的增大,$|dN_d/dr|$值减小,表明监测系统的规模变化对扫描宽度的影响降低.其物理意义与部署阈值意义相符,因此选取公式(7)所示曲线的曲率最大值对应的 r 值作为部署阈值.

3.1.3 理想全网随机扫描监测系统构建

全网随机扫描源参数取值为 $P_{\Omega_s}=1$,$\Omega_s=2^{32}$.取 P_d=0.99 作为监测系统检测需求,计算得出部署阈值 r_{thd}=0.16,对应的扫描宽度阈值 N_{d_thd}=27.具体部署与检测作用见后面的表 7.需要说明的是,理想全网随机扫描监测系统基于对扫描源完全未知的假设前提下构建,它给出了监测系统规模上限和扫描源属性要求下限的理论参考值.对于实际已知典型全网随机扫描源,如 Codered 蠕虫、Slammer 蠕虫,平均扫描速率大于或远大于 N_{d_thd},因此一方面,使用理想构建的全网随机扫描监测系统完全可以 99%的概率检测到这两种已知类型的扫描源;另一方面,如果特定针对 Codered 蠕虫和 Slammer 蠕虫设计监测系统,则根据公式(6),随着预设检测时间的延长,监测系统规模会大为缩减.如 P_d=0.99,当检测时间为 30 分钟时,检测一个类似 Codered 蠕虫扫描源所需监测系统规模仅占可路由地址空间的 0.099%.

3.1.4 理想本地优先扫描监测系统构建

存在两种本地扫描源监测系统构建形式:一种是只依靠自身可控的本地网络资源构建监测系统,称为独立本地扫描监测系统;另一种是独立本地扫描监测系统彼此合作,联合构建基于全网的本地扫描监测系统,称为全网联合本地扫描监测系统.这样既可以缩减每个本地网络监测系统规模、降低部署资源成本,又可以兼顾全网随机扫描源的检测.

1) 独立本地扫描监测系统构建

以典型本地扫描源 Sasser 蠕虫采用的扫描策略作为独立本地扫描监测系统构建设计参考,原因在于,据文献[24]的研究结果表明,现有实际的本地扫描策略,其传播速率已接近最优本地扫描策略.而 Sasser 蠕虫采用的是访问本地网络概率最小的一种策略,因此,研究 Sasser 蠕虫的监测系统构建,对于采用更大本地网络访问概率的扫描源也是适用和有效的.

对于本地优先扫描源,监测点部署于不同的扫描目标网络空间所需的监测点数目不同,与网络位置有关.用 Ω_{s_i} (i=1,2,3)表示 3 类本地优先扫描源的扫描目标网络空间,则 $P_{\Omega_{s_1}}\approx P_{16}, P_{\Omega_{s_2}}\approx P_8, P_{\Omega_{s_3}}=1-P_{16}-P_8$.设 Ω_{rm_i} (i=1,2,3)分别表示 3 类目标网络空间中监测系统规模大小,则根据公式(6)可得:

$$P_d=1-\left[1-\sum_{i=1}^{3}P_{\Omega_{s_i}}\cdot\frac{\Omega_{rm_i}}{\Omega_{s_i}}\right]^{N_d} \tag{9}$$

根据公式(9)可得:

$$\frac{\partial P_d}{\partial \Omega_{rm_i}}=N_d\cdot\frac{P_{\Omega_{s_i}}}{\Omega_{s_i}}\cdot\left[1-\sum_{i=1}^{3}P_{\Omega_{s_i}}\cdot\frac{\Omega_{rm_i}}{\Omega_{s_i}}\right]^{N_d-1} \tag{10}$$

由于 $\Omega_{s_2}>>\Omega_{s_1}$, $\Omega_{s_3}>>\Omega_{s_1}$,因此 $\frac{\partial P_d}{\partial \Omega_{rm_1}}>>\frac{\partial P_d}{\partial \Omega_{rm_2}}$, $\frac{\partial P_d}{\partial \Omega_{rm_1}}>>\frac{\partial P_d}{\partial \Omega_{rm_3}}$.这表明部署于扫描目标网络空间 Ω_{s_1} 内的监测点对整个系统检测率的影响要远大于位于其他扫描目标网络空间内的监测点,因此,对于本地扫描源,要重点考虑在扫描源所在的 Ω_{s_1} 内部署监测点.当 $P_{\Omega_{s_1}}=0.25$,P_d=0.99 时,只考虑在 Ω_{s_1} 空间内部署监测点,得到部署阈值 r_{thd} 为 0.037 6,N_{d_thd} 为 486.具体部署与检测作用见后面的表 7.

2) 全网联合本地扫描监测系统构建

路由地址空间中的每个/16 网络均可存在类似 Sasser 蠕虫的本地扫描源,因此考虑全网联合建立本地扫描监测系统.设可路由地址空间共有 N_8 个/8 网络,每个/8 网络公布的/16 网络数目不同,设为 N_{16_i}(i=1,…,N_8),平均值为 $\bar{N}_{16}$.公式(11)给出理想全网监测系统对源自第 i 个/8 网络中任意一个/16 网络本地优先扫描源的检测效用:

$$P_{di}=1-\left[1-\left(\frac{\Omega_{m_1}}{2^{16}}\cdot P_{\Omega_{s_1}}+P_{\Omega_{s_2}}\cdot\frac{\Omega_{m_2}}{2^{16}}\cdot\frac{N_{16\_i}-1}{2^8-1}+P_{\Omega_{s_3}}\cdot\frac{\Omega_{m_3}}{2^{16}}\cdot\frac{N_8-1}{2^8-1}\right)\right]^{N_d} \tag{11}$$

为简便估算,设每个/16 网络内监测系统规模相同,大小为Ω_m;设每/8 网络平均包括 $\bar{N}_{16}$ 个/16 网络,则监测系统对源自任意一个/16 网络的本地扫描源的检测概率近似为

$$P_d\approx 1-\left[1-\frac{\Omega_m}{2^{16}}\left(P_{\Omega_{s_1}}+P_{\Omega_{s_2}}\cdot\frac{\bar{N}_{16}-1}{2^8-1}+P_{\Omega_{s_3}}\cdot\frac{N_8-1}{2^8-1}\right)\right]^{N_d} \tag{12}$$

根据 2007 年 9 月 29 日公布的 BGP 路由表统计得出, $\bar{N}_{16}=141$,N_8=158.当 P_d=99%时,根据式(12)可得部署阈值 r_{thd} 为 0.007,N_{d_thd} 为 926,将该阈值调整为 r_{thd}=0.0135,N_{d_thd}=486,以便与独立本地扫描监测系统比较.可知全球可路由地址空间中,平均每/16 网络只需提供 887 个未使用 IP 地址,平均每/24 子网提供 4 个,如此构建的联合分布式监测系统就能以 99%的概率检测到扫描宽度 $N_d\geq 486$ 的本地扫描源.与独立本地扫描监测系统相比,全网联合监测系统对每/16 网络中的未使用 IP 地址数目需求降低了 36%.对于本地优先扫描源,如果独立本地扫描检测系统能够联合合作,那么每个本地网络可以更小的部署地址资源代价获取更大的监测收益.具体部署与检测作用见后面的表 7.

研究结论为实际 Leurre’Com Honeypot Project 全球分布式监测系统部署方案提供了理论佐证.该项目要求每个自愿加入的监测点提供 4 个可路由未使用 IP 地址,但相关发表文献中一直未看到为何如此选择的原因解释.根据本节研究结论,如果该全球监测系统能够发展到全网每/24 网络提供 4 个可路由未使用 IP 地址,那么这种监测系统部署方案能够对各类未知扫描源发挥有效的检测作用.

3.2 监测系统效用评估

截止到 2007 年 11 月,实际 Leurre’Com Honeypot Project(简称 LHP)全球分布式监测系统已建立 50 个监测点,分布于全球 30 多个国家[5],每个监测点占用 4 个未使用 IP 地址(属于同一个/24 网络).LHP 的目标之一是利用未使用 IP 地址构建全球早期预警系统[25],包括对网络已知和未知恶意扫描源进行早期预警.清华大学 CCERT 小组作为该项目成员之一所关心的是,能否借助已有的监测系统实现自己尽早定位每个恶意扫描源的检测目标.为此,对现有规模 LHP 监测系统对单点扫描源的检测效用进行评估,为未来检测目标的制定提供参考.本文以检测时间作为检测效用的衡量值.尽管 IPv4 源地址不可信,但目前尚无更好的扫描源表示方法,因此认为 24 小时之内出现的同一个扫描源具有相同扫描行为.如果检测时间超过 24 小时,则认为监测系统检测效用低下;否则认为监测系统检测有效.

3.2.1 全网随机扫描源检测效用评估

LHP 监测系统现有部署规模约为 200 个 IP 地址,以类似 Codered 蠕虫和 Slammer 蠕虫的扫描源作为单点全网随机扫描源的样例.根据公式(6)得到不同检测率要求下,监测系统对类似上述两种扫描源扫描宽度的要求以及到达该扫描宽度所需的时间,见表 5.

Table 5 Random scan width requirements for the current LHP monitoring system and reaching time required by real scanners on different detection probability demands

表 5 不同检测率要求下,实际 LHP 监测系统对随机扫描源扫描宽度要求以及实际扫描源所需到达时间

Detection probability P_c (%)	99	95	90	85	80	50	40
Scan width N_d	1.9×10^8	6.4×10^7	4.9×10^7	4.1×10^7	3.4×10^7	1.5×10^7	1.1×10^7
Reaching time required by a Codered worm scanner (min)	5.3×10^5	1.7×10^5	1.36×10^5	1.14×10^5	9.5×10^4	4.2×10^4	3.1×10^4
Reaching time required by a Slammer worm scanner (min)	791	267	204	171	141	62	46

表 5 估算结果显示,现有规模 LHP 监测系统对于高速随机扫描源检测有效,但对于类似 Codered 蠕虫的中速以及更低速的扫描源,检测效用低下.

3.2.2 本地优先扫描源检测效用评估

实际监测点分布于不同的国家,两个监测点位于同一个/8 网络的数目很少,清华大学在同一个/16 网络部署了两个监测点,因此不妨假设 $\Omega_{rm_1}=8,\Omega_{rm_2}=0,\Omega_{rm_3}=192$.以已知典型本地优先扫描源作为检测样例.表 6 给出不同检测率要求下,监测系统对扫描源扫描宽度的要求以及扫描源到达该扫描宽度所需的时间.

Table 6 Localized scan width requirements for the current LHP monitoring system and reaching time required by localized scanners on different detection probability demands

表 6 不同检测率要求下,实际 LHP 监测系统对本地优先扫描源扫描宽度要求以及实际扫描源所需到达时间

Detection probability P_d (%)	99	95	90	85	80	50
Scan width N_d	1.5×10^5	9.8×10^4	7.5×10^4	6.2×10^4	5.2×10^4	2.2×10^4
Reaching time required by a Sasser worm scanner (min)	1 172	765	586	484	406	172
Reaching time required by a CoderedII worm scanner (min)	500	326	250	206	173	73
Reaching time required by a Blaster worm scanner (min)	7 500	4 900	3 750	3 100	2 600	1 100
Reaching time required by a Nimda worm scanner (min)	791	267	204	171	141	62

表 6 估算结果表明,现有规模 LHP 监测系统对于平均扫描速率大于 300/分钟的本地优先扫描源检测有效,对低速扫描源检测效用低下.

综合以上评估结果,现有 LHP 监测系统虽然规模较小,但在高速全网随机扫描源和中速本地优先扫描源检测上能够发挥有效检测作用;对低速扫描源检测效用低下.根据此评估结果可以进一步认识 LHP 监测系统的作用,从而制定更符合自身需求的检测目标和监测系统部署策略.

3.3 模型应用小结

表 7 对基于路由分布扫描监测模型应用研究进行小结(P_d=99%).

Table 7 Summary of model application

表 7 模型应用小结

Name of a monitoring system	Deployment policy	Detection of a random scanner		Detection of a localized scanner	
		Demand for a unknown scanner	If or not it can effectively detect known scanners like Codered and Slammer	Demand for a unknown scanner	If or not it can effectively detect known scanners like Sasser, CoderedII and Nimda
Ideal random scanning monitoring system	Scale of the monitoring system is 39% of the size of global routable IP address space. Monitors are preferentially deployed in the /8 subnet which owns more /24 subnets according to the route distribution. Monitors are not sensitive to the network position and the size of a monitor is not less than 128.	Scan width ≥27	Yes	Scan width ≥27	Yes
Ideal independent localized scanning monitoring system	Scale of the monitoring system is 3.76% of the size of the /16 subnet in which a scanner locates and monitors are deployed. Monitors are sensitive to the network position. The size of a monitor is average 10.	Scan width $\geq8\times10^6$	Yes	Scan width ≥486	Yes, but it should be a scanner comes from localized subnets
Ideal allied localized scanning monitoring system	Scale of the monitoring system is 3.76% of the size of the global routable IP address space. Monitors are deployed in each /16 subnet and sensitive to the network position. The size of a monitor is 4.	Scan width ≥562	Yes	Scan width ≥486	Yes
Real LHP global distributed monitoring system	The ratio of the scale of the monitoring system to the size of global routable IP address space is 1.12×10^{-7}. Monitors are deployed in 50 different /24 subnets. The size of a monitor is 4.	Scan width $\geq1.9\times10^8$	In 24 detection hours, Slammer Yes, Codered No	Scan width $\geq1.5\times10^5$	In 24 detection hours, Blaster No, others Yes

4 仿真实验验证

首先对基于路由分布扫描监测模型构建的理想监测系统进行仿真验证.由于理想全网随机扫描监测系统规模巨大,难于仿真,因此主要对本地优先扫描监测系统进行仿真验证.如果实验结果与理论分析吻合,则证明本文提出模型正确、合理,基于模型构建的描监测系统也合理.其次,验证实际 LHP 分布式监测系统的检测效用.

4.1 理想本地优先监测系统仿真实验

4.1.1 实验设计

实验采用随机离散事件发生器模拟扫描源行为.事件发生器根据实际 Blaster 蠕虫源代码修改而成,添加 3 个扫描策略控制参数,分别是扫描概率 P_{16},P_8 以及平均扫描速率 s.使用随机数产生器以纯概率方式生成扫描目标,每次实验随机种子不同,以保证事件发生器与实际恶意代码采用的扫描实现相吻合.监测系统根据 Route View 项目 2007 年 9 月 29 日公布的全球 BGP 路由表用 Mysql 数据库构建.独立本地优先扫描监测系统在与测试扫描源相同的/16 网络内部署 256 个监测点,每个监测点任选 10 个 IP 地址;全网联合本地优先扫描监测系统在可路由地址空间内部署 5 837 250 个监测点,每个监测点任选 4 个 IP 地址.测试扫描源名称及参数见表 8.实验验证 60 分钟内,时间间隔 1 分钟,两类理想本地优先扫描监测系统对测试扫描源的检测率.每时间间隔的实验次数为 100,检测率是指测试扫描源被监测系统检测到的次数占总实验次数的比例.测试扫描源参数见表 8.

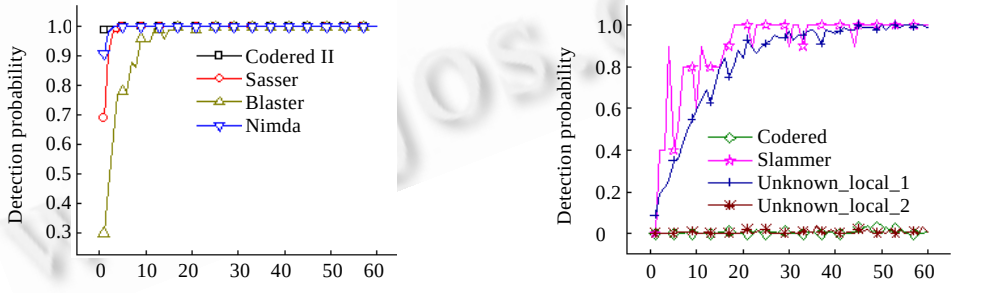
Table 8 Parameters of testing scanners

表 8 测试扫描源参数

Scanner type		Scanner name	P_{16}	P_8	$1-P_{16}-P_8$	Average scan rate
Known	Random scanning	Codered	0	0	1	358
		Slammer	0	0	1	2.4×10^5
	Localized scanning	Codered II	0.375	0.5	0.125	300
		Sasser	0.25	0.25	0.5	128
		Blaster	0.4	0	0.6	20
		Nimda	0.5	0.25	0.25	128
Unknown	A localized scanner comes from local subnets	Unknown_local_1	0.25	0.25	0.5	10
	A localized scanner comes from nonlocal subnets	Unknown_local_2	0.25	0.25	0.5	1 024

4.1.2 实验结果与分析

图 6 示出独立本地扫描监测系统对测试扫描源的检测结果.图 6(a)中的 Blaster 蠕虫扫描源以及图(b)中的 Unknown_local 扫描源清晰地表明,当扫描宽度趋于 500 时,检测概率趋于 1,这与扫描宽度≥486 的理论分析结果相吻合.图 6(b)示出监测系统对于类似 Codered 蠕虫的随机扫描源以及非源自本地的本地优先扫描源检测效用很差.图 6 示出,现有已知本地优先扫描源和类似 Slammer 蠕虫的高速扫描源均可被有效检测;对于源自本地网络的低速扫描源不能被有效检测;对于非源自本地的本地优先扫描源,即使扫描速度是源自本地网络低速扫描源的 100 倍,也不能被有效检测.这与理论分析结论相吻合.



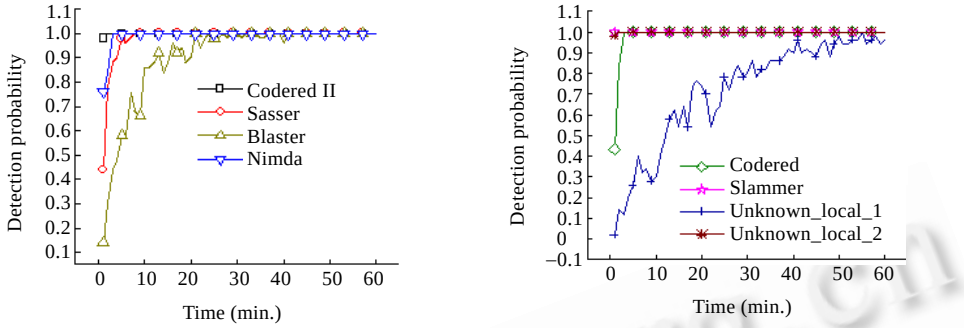
(a) Known local preference scanners (a) 已知本地优先扫描源

(b) Random scanners and unknown local preference scanners (b) 随机和未知本地优先扫描源

Fig.6 Detection results of local network based ideal local preference scan monitoring system

图 6 独立本地优先扫描监测系统检测结果

图7示出全网联合本地扫描监测系统对测试扫描源的检测结果.图7(a)中Blaster蠕虫扫描源的检测曲线在25分钟左右达到检测率1,这与扫描宽度≥486的理论分析结果完全吻合.图7(a)与图6(a)相比,两类监测系统对已知本地优先扫描源的检测结果一致,但是全网联合本地优先扫描监测系统与独立监测系统相比,本地网络需提供的监测地址数目减少了60%.图7(b)与图7(a)相比,全网联合本地扫描监测系统对于类似Codered蠕虫的随机扫描源以及非源自本地的本地优先扫描源能够以高检测率有效检测.这与理论分析结论也相一致.



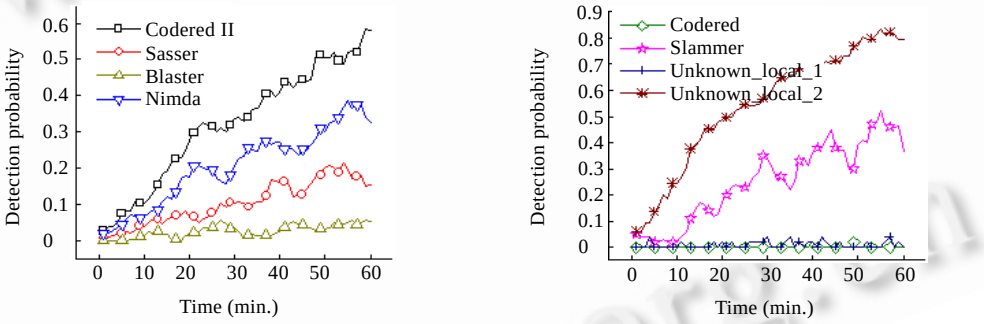
(a) Known local preference scanners (b) Random scanners and unknown local preference scanners
(a) 已知本地优先扫描源 (b) 随机和未知本地优先扫描源

Fig.7 Detection results of global network based ideal local preference scan monitoring system

图7 全网联合本地优先扫描监测系统检测结果

4.2 LHP实际分布式监测系统评估验证

扫描源仿真与理想本地优先监测系统仿真实验完全相同,监测系统根据LHP实际已部署IP地址列表用Mysql数据库实现.实验结果如图8所示.



(a) Known local preference scanners (b) Random scanners and unknown local preference scanners
(a) 已知本地优先扫描源 (b) 随机和未知本地优先扫描源

Fig.8 Detection results of real LHP distributed monitoring system

图8 实际LHP分布式扫描监测系统检测结果

图8给出实际LHP分布式扫描监测系统检测结果.图8(a)与图7(a)相比,检测曲线明显振荡,表明对于已知本地优先扫描源,该监测系统很难获得稳定的检测率.虽然随着时间的增加,检测曲线整体呈现增长趋势,检测时间达到60分钟时,检测率仍不超过60%,这与表8理论分析结果一致.图8(a)中,Sasser与Nimda蠕虫扫描源平均扫描速率相同,但相同时间下监测系统获取的检测率不同,原因在于在监测系统规模非常有限的情形下,扫描源对本地网络的访问概率对检测率影响更为显著.上述实例中,Nimda蠕虫对本地/8网络的访问概率高于Sasser蠕虫,因此,监测系统在同一时间对Nimda蠕虫的检测率要高于Sasser蠕虫.这表明本文基于Sasser蠕虫扫描策略设计本地优先扫描理想监测系统是可行的,至少确定了监测系统规模的上限.图8(b)给出LHP监测系

统对类似 Codered 蠕虫和低速本地优先扫描源的检测效用低下;对类似 Slammer 蠕虫的高速随机扫描源系统检测率非常不稳定,具有一定的检测效用.这些都与表 7、表 8 的理论分析结果一致.

5 总结与展望

本文针对目前利用未使用 IP 地址构建扫描监测系统缺乏理论指导的现状,提出基于路由分布的扫描监测模型.该模型与文献[10]提出的基于全球漏洞主机分布的扫描监测模型相比,全球可路由主机分布根据 BGP 公告可实际获取,从而研究结论对于实际扫描监测系统部署具有指导作用.基于该模型,提出监测系统部署阈值概念,描述相同检测率要求下监测系统规模与扫描源扫描宽度属性之间最经济匹配取值.将模型应用于实际扫描监测系统部署的两个问题:1) 不考虑部署地址资源限制时,理论上针对特定类型扫描策略理想监测系统如何构建问题;2) 实际可控部署地址资源有限时,有限部署资源检测效用评估问题.本文以应用最广泛的全网随机扫描源和本地优先扫描源理想扫描监测系统构建作为第 1 个问题的应用样例,而以实际 Leurre'Com Honeypot Project 全球分布式监测平台在单点扫描源检测上的效用评估作为第 2 个问题的应用样例.仿真实验结果表明,基于该模型对监测系统部署阈值的估计是准确的,理论估计结果与仿真结果一致,表明基于路由分布扫描检测模型对于实际扫描监测系统部署具有有效的指导作用.这样在实际构建扫描监测系统之前,可根据基于路由分布的扫描监测模型估算实现检测目标所需要的理论部署资源,或者评估可控部署资源的检测能力,从而为制定和设计适合自身资源情况的检测目标和监测系统部署方案提供理论指导,避免原来凭经验的盲目部署.

目前,本文的研究工作主要考虑单目标扫描监测系统部署的相关问题,并未考虑具有自相似特性的多扫描源检测和扫描监测系统部署问题;本文的研究工作基于 IPv4 地址空间展开,在 IPv6 网络环境下,地址空间的巨大增长必然促使扫描策略发生变化,如何跟随扫描策略变化构建 IPv6 环境下有效扫描监测系统等都是需要我们进一步深入思考的问题.

致谢 衷心感谢清华大学信息学院薛一波研究员对论文初稿的细致评阅和意见反馈,感谢清华大学下一代互联网实验室吕国涵博士对论文的修改建议,感谢清华大学 CCERT 小组郑先伟工程师一直以来对研究工作的大力支持以及小组其他成员对论文的讨论和建议.

References:

[1] Guo XP. Setup active security protection system. 2005-08-01/2007-10-29 (in Chinese with English abstract). htttp://www.cert.org.cn/upload/ 2005AnnualConferenceCNCERT/1MainConference/11.GuoXunping-SASPS.pdf

[2] Symantec deep sight early warning services. 2006-07/2007-10-29. http://eval.symantec.com/.../enterprise/brochures/ent-brochure_symc_deepsight_early_warning_services_07-2006.pdf

[3] Ruixing network security early warning center (SDS-1000T). 2006-01-01/2007-11-05 (in Chinese with English abstract). http://it.rising.com.cn/product/ wlaqyj.html

[4] Honeynet project. 2003-01-01/2007-05-15. http://www.honeynet.org

[5] Leurrecom.org honeypot project. 2003-06-01/2007-05-15. http://www.leurrecom.org

[6] Moore D, Shannon C, Voelker GM, Savage S. Network telescopes. Technical Report, 2004-04/2007-11-05. http://www.caida.org/outreach/papers/2004/tr-2004-04/tr-2004-04.pdf

[7] Cooke E, Bailey M, Mao ZM, McPherson D. Toward understanding distributed blackhole placement. In: Proc. of the ACM CCS Workshop on Rapid Malcode. Washington: ACM Press, 2004. 54−64. http://portal.acm.org/citation.cfm?id=1029618.1029627

[8] Bailey M, Cooke E, Jahanian F, Nazario J, Watson D. The Internet motion sensor: A distributed blackhole monitoring system. In: Proc. of the ISOC Network and Distributed Systems Security Symp. San Diego, 2005. http://www.isoc.org/isoc/conferences/ndss/05/proceedings/papers/ims-ndss05.pdf

[9] Yegneswaran V, Barford P, Jha S. Global intrusion detection in the DOMINO overlay system. In: Proc. of the ISOC Network and Distributed Systems Security Symp. (NDSS). San Diego, 2004. http://www.isoc.org/isoc/conferences/ndss/04/proceedings/Papers/Yegneswaran.pdf

[10] Rajab MA, Monrose F, Terzis A. On the effectiveness of distributed worm monitoring. In: Proc. of the 14th USENIX Security Symp. Baltimore, 2005. 225−237. http://portal.acm.org/citation.cfm?id=1251413

[11] Zou CC, Gong WB, Towsley D, Gao LX. The monitoring and early detection of Internet worms. IEEE/ACM Trans. on Networking, 2005,13(5):961−974.

[12] Leckie C, Kotagiri R. A probabilistic approach to detecting network scans. In: Proc. of the 8th IEEE Network Operations and Management Symp. (NOMS 2002). 2002. 359−372. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/7927/21855/01015594.pdf?arnumber=1015594

[13] Jung J, Paxson V, Berger AW, Balakrishnan H. Fast portscan detection using sequential hypothesis testing. In: Proc. of the 2004 IEEE Symp. on Security and Privacy. 2004. 211−225. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1301325

[14] Venkataraman S, Song D, Gibbonsy PB, Blum A. New streaming algorithms for fast detection of superspreaders. In: Proc. of the ISOC Network and Distributed Systems Security Symp. San Diego, 2005. 200−210. http://www.isoc.org/isoc/conferences/ndss/05/proceedings/papers/superspreader.pdf

[15] University of oregen route views project. 1998-01-01/2007-10-29. http://www.routeviews.org

[16] APNIC. http://www.apnic.net/mailing-lists/bgp-stats//archive/2008/04

[17] Wu J, Vanagala S, Gao L, Kwiat K. An effective architecture and algorithm for detecting worms with various scan techniques. In: Proc. of the ISOC Network and Distributed Systems Security Symp. San Diego, 2004. 143−156. http://www.isoc.org/isoc/conferences/ndss/04/proceedings/Papers/Wu.pdf

[18] Codered worm. 2001-07-19/2007-09-10. http://www.symantec.com/security_response/writeup.jsp?docid=2001-071911-5755-99

[19] W32.SQLExp.worm. 2003-01-25/2007-09-10. http://www.symantec.com/security_response/writeup.jsp?docid=2003-012502-3306-99

[20] W32.sasser.worm. 2004-05-01/2007-09-10. http://www.symantec.com/security_response/writeup.jsp?docid=2004-050116-1831-99

[21] Codered II. 2001-08-04/2007-09-10. http://www.symantec.com/security_response/writeup.jsp?docid=2001-080421-3353-99

[22] W32.blaster.worm. 2003-08-11/2007-09-10. http://www.symantec.com/security_response/writeup.jsp?docid=2003-081113-0229-99

[23] W32.nimda.A@mm. 2001-09-18/2007-09-10. http://www.symantec.com/security_response/writeup.jsp?docid=2001-091816-3508-99

[24] Chen ZS. Modeling and defending against internet worm attacks [Ph.D. Thesis]. Atlanta: Georgia Institute of Technology, 2007.

[25] Pouget F, Dacier M, Pham, VH. Leurre.com: On the advantages of deploying large scale distributed honeypot platform. In: Proc. of the E-Crime and Computer Conf. Monaco, 2005. http://www.eurecom.fr/util/publidownload.en.htm?file=/homesdocs/publications/htdocs/ce/pougfa-050329.pdf

附中文参考文献:

[1] 郭训平.建立主动安全防护体系.2005-08-01/2007-10-29. http://www.cert.org.cn/upload/2005AnnualConferenceCNCERT/1MainConference/11.GuoXunping-SASPS.pdf

[3] 瑞星网络安全预警中心(SDS-1000T).2006-01-01/2007-11-05. http://it.rising.com.cn/product/wlaqyj.html

马莉波(1971－),女,山东临沂人,博士生,高级工程师,主要研究领域为入侵检测,早期预警.

张亮(1969－),男,博士,研究员,主要研究领域为多媒体通信信号处理,SOC.

李星(1956－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为多媒体通信,计算机网络.