第 33 卷 第 5 期 计 算 机 工 程 2007 年 3 月
Vol.33 No.5 Computer Engineering March 2007
·安全技术· 文章编号：1000—3428(2007)05—0143—03 文献标识码：A 中图分类号：TP393.08

基于 Honeypot 的可疑度模型

汪 洁，王建新
(中南大学信息科学与工程学院，长沙 410083)

摘 要：提出了基于 Honeypot 系统的可疑度模型，它可以从外部访问主机中判断入侵主机。这个模型通过分析 Honeypot 系统里一段时间内发生的所有事件，对访问 Honeypot 系统的主机赋予一个可疑度的值。如果可疑度的值超过某一个阈值，则此访问者被认为是入侵者。采用了大量的模拟试验对模型进行了测试和分析，在特定的 Honeypot 系统内测试了模型的误判率，结果证明可疑度模型对于检测 Honeypot 系统里的入侵者是一个有效的方法。

关键词：蜜灌；可疑度模型；入侵检测；网络安全

Anomaly Degree Model Based on Honeypot

WANG Jie, WANG Jianxin
(College of Information Science & Engineering, Central South University, Changsha 410083)

【Abstract】An anomaly degree model based on honeypot is proposed and it can distinguish intruders from hosts without hostility. This model gives anomaly values to the hosts who visit honeypot by analyzing all events of honeypot system in a period of time. If the anomaly values exceed a threshold, the visitor can be marked as an intruder. A great lot of simulative test cases are used to validate the model. False alarm rate of model is tested in a honeypot system. The results indicate that anomaly degree model is an effective method to detect intruders in honeypot system.

【Key words】 Honeypot; Anomaly degree model; Intrusion detection; Network security

Honeypot是一种资源，它的目的就是监视、记录那些探测、攻击和威胁其安全的活动[1]。目前，Honeypot概念已经获得普遍的认同，而且也实现了它更加深远的价值。如Honeynets工程使用Honeypot学习了许多关于黑客集团的工具、策略以及攻击的动机[2]。Nathalie Weiler研究开发Honeypot用来保护网络不受拒绝服务攻击的影响[3]。

Honeypot也被认为是一种保护局域网的积极防御手 段[4]。Lawrence Teo提出了一种称为Japonica的安全框架，它把Honeypot作为一种积极的实体[5]。Miyoung Kim讨论过设计和实现一个积极的Honeypot系统[6,7]。目前很少有研究关于怎样在Honeypot系统里面进行入侵检测，仅仅是认为任何与Honeypot交互的行为最有可能是未授权或恶意行为，任意与Honeypot的连接都可能是一个探测、攻击或破坏行为。但事实上，合法的用户也可能偶尔会访问Honeypot。本文提出应用于Honeypot系统的可疑度模型，来解决这个问题。

1 可疑度模型

1.1 可疑度定义

定义 1 可疑度

可疑度是指在一段时间内主机试图或者正在进行攻击的可疑程度，它是对某台具体主机而言，一般是指某台主机的可疑度。

定义 2 可疑度模型

可疑度模型是设计用来对外部可疑主机进行攻击检测的模型，它是通过对 Honeypot 系统产生的警报数据的综合分析来得出外部主机的可疑度，从而判断外部主机是发起攻击的主机还是仅仅进行了误访问的主机。

为了与其它一般的 Honeypot 进行区别，本文定义诱饵来作为网络中的虚拟主机。如果一个外部主机与任意一个诱饵进行交互，那么就称它是在访问诱饵。外部主机被称为访问者，当它访问 Honeypot 系统时，会产生一个事件。一个网段内有大量的诱饵，通过分析在一段时间内 Honeypot 系统内的事件来分析主机的可疑度，从而发现入侵者和可疑主机。

可疑度模型是应用于 Honeypot 系统，这里的 Honeypot 系统是基于以下 2 个基本假设：

假设 1 入侵者都可能会访问 Honeypot 系统。对这个假设存在可能性进行解释如下：设网络里有 n 个诱饵和 m 台真实主机，那么入侵者以 $\frac{n}{n+m}$ (n > m)的概率访问 Honeypot 系统。因为在局域网内配置了大量的诱饵，入侵者有可能掉入 Honeypot 系统里面。而且诱饵一般都会配置成比较脆弱的主机系统，所以认为这个假设有存在的可能性。

假设 2 存在外部主机对 Honeypot 系统的误访问。由于某些偶然的因素(例如某主机 A 修改了 IP 地址，而 Honeypot 系统使用了主机 A 原有的 IP 地址，从而使得原本想访问 A 的主机却对 Honeypot 进行了误访问)，或某些用户的一些错误的操作及操作系统错误的配置都可能会使外部主机误访问 Honeypot 系统，所以外部主机对 Honeypot 系统的误访问是存在的。

1.2 可疑度计算涉及的主要因素

根据可疑度及可疑度模型的具体定义，对于外部主机可

基金项目：国家自然科学基金资助项目(60403032)
作者简介：汪 洁(1980－)，女，硕士、助教，主研方向：网络信息安全；王建新，博士、教授
收稿日期：2006-04-02 **E-mail：**wang_jie@mail.csu.edu.cn

疑度的计算需要考虑很多因素，具体如下：

(1)主机访问 Honeypot 的次数

访问诱饵的次数反映了主机可疑活动的活跃性。显然，主机访问 Honeypot 的次数越多，越有可能是入侵者。利用这个特征对于检测网络拓扑发现、主机探测、针对主机的端口扫描等几乎所有攻击行为都有帮助。

(2)主机访问 Honeypot 的范围

由于入侵者不了解所处网络的拓扑结构，特别是关于真实主机的信息，因此一般情况下入侵者会大范围地访问诱饵。而正常的没有恶意的主机，即使偶尔访问了诱饵，也是在很小的范围内。

这个因素主要针对入侵者的下列行为：网络拓扑发现，主机探测，慢速的网络试探。这些行为的特点是需要大量访问不同的主机。

(3)事件的频率

入侵者的行为，往往在时间上比较集中(在一段时间内频率高)。所以在 Honeypot 内部，一段较短的时间内，很可能会发生大量的入侵者访问诱饵的行为。针对这种特点，可以考虑事件的频率来评价主机的可疑度。访问 Honeypot 的行为频率越高的主机，越有可能是入侵者。

本文将一段时间内入侵者对Honeypot的访问(如探测、攻击等)称为一次活动，即活动是入侵者行为的序列，这一段时间就是活动的持续时间。不同的黑客、使用不同的工具、在攻击的不同阶段，一次攻击的持续时间是不同的，然而可以给出一个假定的活动平均持续时间 N_{PAAD}(presumptive-average-action-duration)。

(4)事件的属性

当访问者访问 Honeypot 系统时，他每次的访问行为，具有下列 3 个基本的属性：

1)访问的长度：访问长度是指访问者发送报文的负载长度，或 TCP/UDP 连接中的报文长度；

2)攻击工具；

3)访问端口的性质。

2 可疑度的计算

2.1 可疑度计算中的几个定义

本文在可疑度计算过程中会涉及以下几个概念：

(1)诱饵的集合

设 Honeypot 中诱饵的集合为 B，$B=\{b_1,b_2,...,b_{N_{bait}}\}$，表示在 Honeypot 系统里有 N_{bait} 台诱饵主机。

(2)主机的集合

设网络中主机的集合为 H，$H=\{h_1,h_2,...,h_{N_{host}}\}$，表示在网络中有 N_{host} 台主机系统。

(3)事件

Honeypot 系统发生的事件是诱饵收到的来自可疑主机的一条报文。

在 Honeypot 系统中设所有事件的集合为 E，$E=\{e_1,e_2,...,e_{N_{event}}\}$，表示在 Honeypot 系统内发生了 N_{event} 个事件。

事件具有以下几个属性：

1)Time(e)：表示事件 e 发生的时间；

2)Visitor(e)：表示事件的访问者；

3)Bait(e)：表示发生事件的诱饵；

4)Protocol(e)：表示访问端口的协议类型；

5)VisitPort(e)：表示事件所访问的端口，如果事件没有访问端口，则 VisitPort(e)＝0；

6)Payload(e)：表示事件的有效载荷长度。

为了进行可疑度计算，本文有以下两个定义：

定义 3 时间粒度 Δ_t

时间粒度 Δ_t 是一个时间间隔，在可疑度计算过程中每隔 Δ_t 时间计算外部主机的可疑度。在本文中 Δ_t 取 1s。因此，可以得到计算可疑度的时间点的集合：

$T=\{t_i\in time \mid \forall i, t_{i+1}=t_i+\Delta_t, i\geq 0\}$，其中 t_0 是系统初始时间。那么 $T_{[i,j)}$ 是在 T 上定义的一段时间，具体表示为

$$T_{[i,j)}=\{t\in \text{time} \mid t_i\leq t<t_j, \text{ where } t_i,t_j\in T,\ i<j\}$$

定义 4 观察窗口 WW_i

观察窗口 WW_i 是指一段时间，Honeypot 系统使用 WW_i 时间内产生的事件来计算访问者的可疑度。本文假定观察窗口必须大于假定的活动平均持续时间，即 $WW_i > N_{PAAD}$。

本文中 WW_i 取 3 600s，N_{PAAD}＝60s，即观察窗口为 1h，假定的攻击者平均活动持续时间是 1min。

本文使用 E^{WW_i} 表示观察窗口内的事件集合。

$$E^{WW_i}=\{e_j\in E \mid \text{Time}(e_j)\in WW_i\}$$

2.2 可疑度的计算方法

处于观察窗口 WW_i 时，某台主机 h 的可疑度计算公式如下：

$$AS^{WW_i}(h)=w_1\times AS_{times}{}^{WW_i}(h)+w_2\times AS_{range}{}^{WW_i}(h)+w_3\times AS_{frequency}{}^{WW_i}(h)+w_4\times AS_{event\text{-}properties}{}^{WW_i}(h)$$

式中 w_i 指加权值。

(1) $AS_{times}{}^{WW_i}(h)$

$AS_{times}{}^{WW_i}(h)$ 是根据主机 h 访问诱饵的次数所得出的值。在观察窗口 WW_i 内主机 h 的事件集合：

$$E^{WW_i}{}_h=\{e_j\in E^{WW_i} \mid \text{Visitor}(e_j)=h\}$$

根据测试可以得出观察窗口 WW 内的事件的最大值 $\max|E^{WW}|$。令

$$AS_{times}{}^{WW_i}(h)=\frac{|E^{WW_i}{}_h|}{\max|E^{WW}|}\times 10$$

(2) $AS_{range}{}^{WW_i}(h)$

$AS_{range}{}^{WW_i}(h)$ 是指主机 h 访问 Honeypot 系统内诱饵的范围所得出的值。在观察窗口 WW_i 内主机 h 访问的诱饵集合如下：

$$B^{WW_i}{}_h=\{b_j\in B \mid \exists e_1\in E^{WW_i}{}_h \text{ let } \text{Bait}(e_1)=b_j\}$$

$|B^{WW_i}{}_h|$ 即主机 h 在观察窗口中访问的不同诱饵的个数，显然 $|B^{T_i}{}_h|\leq N_{bait}$，令

$$AS_{range}{}^{WW_i}(h)=\frac{|B^{WW_i}{}_h|}{N_{bait}}\times 10$$

(3) $AS_{frequency}{}^{WW_i}(h)$

$AS_{frequency}{}^{WW_i}(h)$ 是根据外部主机 h 在 Honeypot 系统内产生事件的频率而得出的值。因为规定观察窗口大于(远大于)活动平均持续时间，所以在整个观察窗口直接计算频率不能准确反映入侵者活动的特征。考虑观察窗口内每个大小为 N_{PAAD} 的小区间，在这个小区间内计算事件的频率，观察窗口的事件频率是各个子区间事件频率的最大值。根据测试可以得出观察窗口 WW 内某外部主机产生事件的最大频率值

$\max(mf)$。下面给出 $AS_{frequency}^{WW_i}(h)$ 的计算方法。

令 $E^{T_{[u,v)}}{}_h = \{ e_j \in E \mid Visitor(e_j) = h, Time(e_j) \in T_{[u,v)} \}$

$$frequency(T_{[u,v)}, h) = \frac{\left| E^{T_{[u,v)}}{}_h \right|}{N_{PAAD}}$$

令 $$mf = \max_{T_{[u,u+PAAD)} \subset WW_i} frequency(T_{[u,u+PAAD)}, h)$$

则

$$AS_{frequency}^{WW_i}(h) = \frac{mf}{\max(mf)} \times 10$$

(4) $AS_{event\text{-}properties}^{WW_i}(h)$

$AS_{event\text{-}properties}^{WW_i}(h)$ 是指与主机 h 相关的事件的可疑度值。本文利用有效载荷长度、端口的危险度来计算每次访问(即事件)的可疑度值，然后取平均值作为观察窗口内事件的可疑度值。

1)访问的长度(payload length)

通过学习样本，计算出没有恶意的主机误访问诱饵后的有效载荷长度的平均值 μ 和方差 σ。然后由 Chebyshev 多项式的原理

$$PL(e) = 5 \times \left(1 - \frac{\sigma^2}{(Payload(e) - \mu)^2} \right)$$

可以得出访问的长度值 $PL(e)$ 的值小于 5。

2)访问端口的性质

每一个事件都与一个具体的目的端口相联系，对每一个目的端口可能存在的潜在危险程度以一个具体的数值表示，该数值称为端口风险值。由 PRD_{TCP} 和 PRD_{UDP} 分别表示如下：

$PRD_{TCP} : Port \mapsto \{1,2,...,10\}$, where $Port = \{ n \in N \mid 0 \le n \le 25560 \}$

$PRD_{UDP} : Port \mapsto \{1,2,...,10\}$, where $Port = \{ n \in N \mid 0 \le n \le 25560 \}$

事件的端口风险值 $PR(e)$ 为

$$PR(e) = \begin{cases} PRD_{TCP}(Port(e)) & if\ Protocol(e) = TCP \\ PRD_{UDP}(Port(e)) & if\ Protocol(e) = UDP \\ 0 & others \end{cases}$$

则事件的可疑度值为

$$AS_{event\text{-}properties}^{WW_i}(h) = \frac{1}{\left| E^{WW_i}{}_h \right|} \sum_{e \in E^{WW_i}{}_h} (PR(e)) + PL(e)$$

3 模型的测试

3.1 测试方案

可疑度模型的测试是在一个分布式的 Honeypot 系统里面进行的。Honeypot 系统由代理端和总控制中心两部分组成，代理端设置大量的诱饵主机，它能够将诱饵上产生的所有数据发往总控制中心，模型设置在总控制中心。

测试包括外部主机对诱饵进行攻击和外部主机对诱饵进行正常访问。测试中外部主机对诱饵进行攻击采用的攻击工具如表 1 所示。

表 1 攻击工具

工具类型	工具名称
扫描类攻击工具	Nmap 扫描器、SuperScan 扫描器、X-Scan 扫描器
远程溢出类攻击工具	Snake IIS、Webdavx3、0x333samba
口令猜测类攻击工具	BrutusAET2
后门类攻击工具	Back office2000、nodoom、netbus
拒绝服务类攻击工具	ms04007dos

测试中外部主机对诱饵进行正常访问选用的正常服务类型为 Ftp、Telnet、Web。

3.2 测试结果

本文记录了 8 条可疑度变化曲线。图 1 的 5 条可疑度变化曲线为攻击工具产生的可疑度变化曲线。纵坐标表示可疑度值，横坐标表示时间。以远程溢出攻击 Webdavx3 为例，在攻击开始后一段时间内可疑度曲线上升，即可疑度值从 0 开始增加，当攻击结束后可疑度曲线逐渐降低，其值最后变为 0。不同的攻击，攻击过程所花费的时间不同，所以不同攻击的可疑度曲线在时间刻度上表现的长短也不一样，如 Nmap 主机扫描的可疑度变化曲线最短。

图 2 中的 3 条曲线也是代表 3 台外部主机对 Honeypot 进行 3 类正常服务访问时的可疑度变化情况，当正常访问开始时可疑度曲线缓慢从 0 开始上升，当访问结束后可疑度曲线逐渐降为 0。

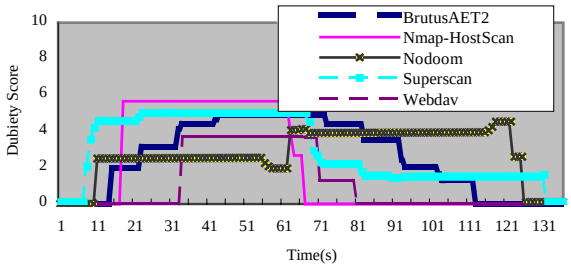

图 1 进行攻击的外部主机可疑度曲线

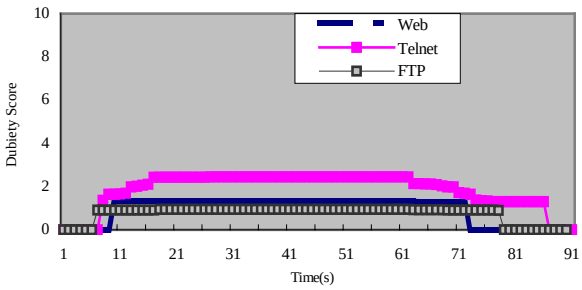

图 2 进行正常访问的外部主机可疑度曲线

将图 1 与图 2 进行比较，可以很明显地看出图 1 中可疑度的最高值偏高，而图 2 的可疑度值偏低。

3.3 模型的误判率及漏判率的计算

本文静态设定阈值为 3，并定义可疑度高于 3 的外部主机为入侵者。在上面测试方案所使用的 Honeypot 系统内，对可疑度模型的误判率及漏判率进行了计算。

本文选取常见的正常服务对 Honepot 系统进行访问，误判率为 0。同时采用 Nussus 的脚本来作为攻击方使用的攻击脚本，对可疑度模型的漏判率的进行了计算。计算分为以下两部分来进行：

(1)使用 50 个攻击脚本，对 Honeypot 系统已经配置的漏洞进行攻击，漏判率为 24%。

(2)使用 100 个攻击脚本，对 Honeypot 系统没有配置的漏洞，误判率为 83%。

由此可见，可疑度模型对外部主机的漏判率很大程度上取决于 Honeypot 系统的模拟程度及其具体的配置。

4 结束语

本文在 Honeypot 技术的基础上，提出了可疑度模型。该模型综合考虑了主机访问诱饵的次数、访问的范围、事件发生的频率、访问的数据长度以及访问端口的性质等因素，对外部主机的可疑度进行了计算，较好地解决了网络攻击判定

问题。 （下转第 148 页）