浙江理工大学学报(自然科学版),第 39 卷,第 1 期,2018 年 1 月
Journal of Zhejiang Sci-Tech University (Natural Sciences)
Vol. 39, No. 1, Jan. 2018

DOI:10.3969/j.issn.1673-3851(n).2018.01.016

基于 Arnold 变换的网络安全态势感知扫描策略研究

岳　巍,张华熊

(浙江理工大学信息学院,杭州 310018)

摘　要:在网络安全态势感知报文总体收发速度不变的情况下,为降低对同一 IP 或连续 IP 地址段内目标的高频、高强度的感知行为,提出了一个网络安全态势感知的优化策略,该策略在经典 Arnold 置乱算法的基础上,结合公网连续 IP 地址段的特点,利用广义 Arnold 系数矩阵的灵活输出特性,对全网域 IP 地址进行混沌置乱的安全态势感知。与主流的随机扫描、段内扫描策略的感知方式相比,该策略不仅提高了对网络安全态势的整体感知效率,而且降低了对同一 IP 或连续 IP 地址段内目标的感知频率和强度。

关键词:Arnold 变换;IP 地址;置乱扫描;网络安全态势感知

中图分类号:TP393.08　　**文献标志码**:A　　**文章编号**:1673-3851 (2018) 01-0098-05

0　引　言

随着网络安全威胁变得日趋复杂,单一或独立的网络安全防护措施无法从整体上对网络当前运行状态和未知安全威胁作出有效的感知与预测。为增强各种网络防护措施的关联性,对网络数据进行融合分析和协同管理的网络安全态势感知(Network security situation awareness)研究已成为新的研究热点[1-3]。当前提出的众多网络安全感知模型与方法,基本概念均源于 1988 年 Endsley 提出的态势感知模型概念[4],主要分为态势要素觉察、态势理解和态势投射三个阶段。在这三个阶段中,完整的态势要素觉察阶段需要获取特定的特征信息,此阶段主要分为主动信息探知和被动融合分析[5-7]。在主动信息探知中,脆弱性扫描是最为常见的感知信息提取手段,主要通过端口扫描获得目标主机的开放端口、提供服务的状态、协议类型等状态信息来提取态势信息,为下一阶段的态势信息融合做准备。比较常用的脆弱性扫描器有 Nmap、Nessus、X-scan、Zmap 和 Masscan 等[8]。

在网络安全态势感知系统中,为保证感知信息的实效性,感知源会较为频繁地对目标进行脆弱性扫描,而脆弱性扫描过程中扫描策略的选取关系着感知信息获取的完整性、实效性等多项指标。脆弱性扫描器中主流的扫描策略主要有随机性扫描策略、IP 地址段内扫描策略。因安全感知式的脆弱性检测与具有入侵性质的扫描存在类似之处,而一般同一机构下连续的 IP 地址段内主机或节点基本共享一套检测机制,段内的集中突发式的脆弱性扫描容易将具有积极反馈的感知信用连接行为判定为非信用连接的异常行为类[9-11],所以针对目前大规模网络下高速的脆弱性检测感知探测行为,上述感知策略存在各自的不足:随机性扫描策略存在重复扫描、无法动态调整扫描范围等问题,影响整体的感知效率;IP 地址段内扫描策略存在易造成对目标的高频扫描问题,影响后续获取感知信息的完整性。

为保留随机扫描策略输出的混沌特性与 IP 地址段内扫描输出结果采样特性,同时弥补各自的不足,本文在收发感知报文速率不变的情况下,提出了一种利用(广义)Arnold 变换进行脆弱性扫描的输出策略。该策略不仅能提供易于动态调整的实时混沌性输出,而且在避免对同一 IP 或连续 IP 地址段内目标的高频感知行为的同时,降低了对目标的感知强度。本文的研究可为大规模网络下高速的安全

收稿日期:2016-10-02　　网络出版日期:2017-12-11

作者简介:岳　巍(1991-),男,河南禹州人,硕士研究生,主要从事计算机网络方面的研究。

通信作者:张华熊,E-mail:zhxhz@zist.edu.cn

态势感知探测提供了更为有效的脆弱性扫描策略。

1　策略构建

1.1　置乱模型

为避免对同一 IP 或连续 IP 地址段内目标的高频感知行为,本文以连续 IP 地址段为感知收发单位,进行段间置乱的感知报文收发。

对于一般的服务器等网络设备,其所占 IP 地址资源有限,即连续 IP 地址首尾跳变间隔较小。而相比于归属较为稳定且同一 IP 地址段首尾 IP 地址跳变间隔较大、划分种类丰富多变的高校 IP 地址段,后者更适宜作为感知收发单位的参考对象。

首先,本文把 IP 地址分为四个字段(以 IPv4 版本为例),整体记为 A. B. C. D(IPv6 可分为八个字段,置乱原理与下文基本相同)。表 1 仅显示数据库(纯真 IP 数据库[12])中提取的 927 个地址段中部分高校 IP 地址段及其 C 字段宽度值。

表 1　四省高校部分 IP 地址段及 C 字段宽度

四省高校	IP 地址(段)	C 字段宽度
高校 1	58.206.192.0—58.206.223.255	32
	60.12.143.0—60.12.143.255	1
高校 2	58.195.248.0—58.195.255.255	8
	210.32.24.0—210.32.27.255	4
高校 3	60.176.36.0—60.176.48.255	13
	60.177.49.0—60.177.51.255	3
高校 4	59.69.208.0—59.69.223.255	16
	125.46.17.0—125.46.17.127	1
高校 5	122.206.128.0—122.206.143.255	16
	218.28.30.198	0
高校 6	125.40.138.105	0
	211.84.80.0—211.84.87.255	8
高校 7	59.66.0.0—59.66.63.255	64
	166.111.228.0—166.111.231.255	4
高校 8	115.27.0.0—115.27.70.255	71
	124.17.17.0—124.17.18.255	2
高校 9	202.106.22.0—202.106.22.255	1
	202.112.112.0—202.112.127.255	16
高校 10	59.75.128.0—59.75.135.255	8
	111.115.64.0—111.115.79.255	16
高校 11	49.209.0.0—49.209.15.255	16
	60.13.177.0—60.13.177.255	1
高校 12	59.75.224.0—59.75.239.255	16
	59.75.242.0—59.75.247.255	6

如表 1 示,所有 IP 地址段跳变间隔的阈值大小均能用 C 字段宽度值表示。

其次,对 927 个 IP 地址段进行加权均值预处理,来确定适宜 IP 地址段置乱的感知收发单位。本文对 IP 地址段中 C 字段的加权预处理提出如下定义:

定义 1:在 IP 地址中,将 C 字段宽度值(即连续 IP 地址段首尾跳变间隔)的阈值需大于等于 R 的特性称为 IP 地址置乱特性。

本文利用向上取整加权平均数计算公式(1)计算 C 字段宽度值阈值的均值 R:

$$R=\left\lceil\frac{\sum_{i=1}^{n}X_i\times N_i}{N_s}\right\rceil \tag{1}$$

式(1)中:X_i 表示第 i 网段下 C 字段宽度值,N_i 对应 X_i 下具有相同宽度值的字段个数,N_s 表示表内所示所计算范围内网段的总个数。

对所有 927 个地址段使用式(1),向上取整得 $R=4$。即需置乱后输出 IP 地址字段中 C 字段的阈值大于等于 4。

1.2　置乱算法

常用的置乱算法包括 Arnold 变换、骑士巡游变换、Hilbert 曲线变换、幻方变换等[13]。上述变换多用于二维的图像处理,由 1.1 节可知,本文将 IP 地址分为四个维度,故需将 A、B、C、D 四个字段进行分组预处理,即每组的一对值分别对应图像置乱中的横纵坐标值。最后对两组结果组合输出 IP 地址。

1.2.1　广义 Arnold 变换

广义的 Arnold 变换[14]是在 V. J. Arnold 提出的经典 Arnold 变换基础上的改进,其变换矩阵定义如下:

$$\begin{bmatrix}X_{n+1}\\Y_{n+1}\end{bmatrix}=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}X_n\\Y_n\end{bmatrix}\bmod N(N\geqslant 2),$$

其中:X_n、Y_n 表示图像中的一个原坐标点;X_{n+1}、Y_{n+1} 表示经一次矩阵变换后的坐标点;N 表示矩阵的阶数;系数矩阵中 a、b、c、d 为正整数,需满足 $\begin{vmatrix}a&b\\c&d\end{vmatrix}=\pm 1$。本文记经典 Arnold 变换系数矩阵为 $\boldsymbol{A}=\begin{bmatrix}1&1\\1&2\end{bmatrix}$,广义 Arnold 变换系数矩阵为 $\boldsymbol{G}=\begin{bmatrix}a&b\\c&d\end{bmatrix}$。其中矩阵 $\boldsymbol{G}$ 的一般性书写格式 $\boldsymbol{P}=\begin{bmatrix}n&mn+1\\1&m\end{bmatrix}$ 或 $\boldsymbol{P}=\begin{bmatrix}mn+1&n\\m&1\end{bmatrix}$,其中 m、n 为正整数。可知经典

Arnold变换下任何置乱次数后的横纵坐标变化均可由广义Arnold的某一系数矩阵在$O(1)$时间复杂度下实现。特别的，矩阵$\boldsymbol{P}=\begin{bmatrix}1 & \frac{256}{R}+1\\ 1 & \frac{256}{R}\end{bmatrix}$或$\boldsymbol{P}=\begin{bmatrix}\frac{256}{R}+1 & 1\\ \frac{256}{R} & 1\end{bmatrix}$适用于$R=2^n$且C字段数值变化为升序的情况。降序需调整$a$、$b$、$c$、$d$位的正负号且行列变化时取绝对值。

矩阵$\boldsymbol{P}$最大的Lyapunov指数$\lambda=1+\frac{mn+\sqrt{(mn)^2+4mn}}{2}>1$，即对于任何$m>0$，$n>0$的正整数映射总是混沌的。对系数矩阵为$\boldsymbol{A}$的典型Arnold映射的Lyapunov指数分别为$\lambda_1=\frac{3+\sqrt{5}}{2}>1$、$\lambda_2=\frac{3-\sqrt{5}}{2}<1$，也称为混沌映射[15]。

相较于其他置乱算法的时间复杂度，广义Arnold变换两个不同系数矩阵组合，不仅可按$O(1)$时间复杂度实现典型Arnold变换N次置乱后反向映射的取值效果，而且可灵活适应R值的变化。置乱算法在IP地址变换输出中的比较情况见表2。

表2 置乱算法比较

置乱方式	时间复杂度	空间复杂度	邻差L
骑士巡游	$O(n)$	$O(1)$	$L\leqslant 2$
Hilbert曲线变换	$O\left(\log_2\left(\frac{n}{2}\right)\right)$	$O(1)$	$L\leqslant 3$
幻方变换	$O(n^2)$	$O(1)$	$L=1$(多数)
普通/广义Arnold变换	$O(n^2)/O(1)$	$O(1)$	$1\leqslant L\leqslant 255$

表2中，邻差表示各置乱算法在初次置乱后矩阵内原相邻自然数点对应的横纵坐标值的差值。其中幻方变换在256阶矩阵下为双偶变换[16]，置乱后矩阵内每行均有128对相邻自然序数对，故其邻差多数为1。

综上所述，(广义)Arnold变换在保证输出IP地址序列混沌性、灵活适应R值变化特性的同时，又可以较为快速地输出符合IP地址置乱特性的置乱结果。

1.3 输出方案

1.3.1 最佳输出周期

按自然数序存储的初始矩阵经置乱后矩阵内的有序点会变得“杂乱无章”，取横(纵)坐标值的顺序方式也就影响了最终输出的结果，存在以下两种反向映射取数方式：

a) 置乱后按初始矩阵内的自然数序顺序取对应的横纵坐标值。该取值方式简称自然序取值。

b) 按置乱后矩阵内的整数序取对应初始矩阵下同一点对应的横纵坐标值。该取值方式简称置乱序取值。

表3—表5为3×3方阵下以矩阵$\boldsymbol{A}$为系数矩阵，初次置乱后的不同取数方式及二维矩阵输出序列说明。

表3 IP地址对应的初始矩阵

坐标	0	1	2
0	1	2	3
1	4	5	6
2	7	8	9

表4 自然序取值

输出序列行号	取值序列	横坐标值	纵坐标值
1	1	0	0
2	2	1	1
3	3	2	2
4	4	1	2

表5 置乱序取值

输出序列行号	取值序列	横坐标值	纵坐标值
1	1	0	0
2	9	2	2
3	5	1	1
4	6	2	1

每种映射取数方法、不同的系数矩阵下的输出序列具有不同的周期特性。以$\boldsymbol{A}$矩阵为系数矩阵，置乱次数为68次时，自然序取值下的周期为例，周期数统计结果见表6。

表6 周期数统计

统计变量	行	列
跳变输出周期Tc	474(1—38)，218(39—256)	361(1—151)，105(152—256)
固定输出周期Ts	256	256
跳变输出差值X	68	22

表6中跳变输出周期Tc，括号前面的数值表示括号内输出序列行中数值的周期，其数值关系是Arnold变换中点间拉伸和折叠纹理特性的映射表现；固定输出周期Ts为变换矩阵的阶数；跳变差值X为跳变周期下横(纵)坐标值变化至相同时的输出间隔差，对变换矩阵阶数进行取模运算。

1.3.2 最佳置乱次数

图像置乱中最佳置乱度一般与置乱分布均匀性和随机性有关[17-18]，而非与本文应用中提出的与IP地址置乱特性有关。因此对IP地址最佳置乱次数提出如下定义：

定义 2:

$$D_m=(|V_n-V_2|<R)\quad(2<n\leqslant 256)\qquad(2)$$

式(2)中:D_m 表示第 m 次置乱下,输出序列中横(纵)坐标差值大于等于 R 值的行(列)间距值;V_n 表示输出序列中第 n 行的横(纵)坐标值;n 为大于 2 且小于等于 256 的自然数;其中置乱周期内最大的 D_m 对应的置乱次数 m 即为最佳置乱次数,而求出最佳置乱次数时使用的横(纵)坐标序列值(即 V 值)即作为 IP 地址里面的 C 字段输出。

1.3.3　输出格式

依置乱算法所确定的分组特性,本文将最终输出格式定义如下:

定义 3(置乱 IP 地址输出格式):输出序列的行中对应的横纵坐标值为一组,记为 $A_n\circ B_n$,最后输出的 IP 地址格式记为 $A_m\circ B_m\circ A_n\circ B_n$。

定义 3 中输出格式的组合可分为同系数矩阵同置乱次数的组合、同系数矩阵不同置乱次数的组合和不同系数矩阵置乱组合三类。为保证输出序列采样感知的混沌性,即继承随机扫描策略的输出优点,生成 $A_n\circ B_n$ 输出序列的系数矩阵宜为 $\boldsymbol{A}$ 或 $\boldsymbol{P}$($a>0$,$b>0$)的组合输出。

2　结果及分析

为计算全网域下以 R 值为 IP 地址段跳变阈值的平均跳变时间(即计算相邻的对同 IP 地址段扫描行为的时间间隔)、同一 IP 地址段内单位时间感知强度等,本文在百兆快速以太网实验环境下,以最小 64 字节的发包长度、线速每秒最多与约 2.3 万个 IP 地址完成基本握手和挥手动作的 7 次报文交互 TCP 态势感知探测行为为实验条件,对本文提出的感知策略进行实验。实验以矩阵$\begin{bmatrix}1 & 65\\ 1 & 64\end{bmatrix}$和矩阵$\begin{bmatrix}1 & 8\\ 1 & 7\end{bmatrix}$作为广义 Arnold 的系数矩阵,模拟输出系数为$\begin{bmatrix}1 & 7\\ 1 & 8\end{bmatrix}$的经典 Arnold 变换下,置乱次数为 50 次时的横纵坐标值的情况输出感知 IP 序列为例,为避免对同一 IP 或连续 IP 地址段内目标的高频感知行为,置乱序取值下输出序列按 $Tc=512$ 且横坐标值作为 C 字段值循环组合输出,采样结果如图 1 所示。

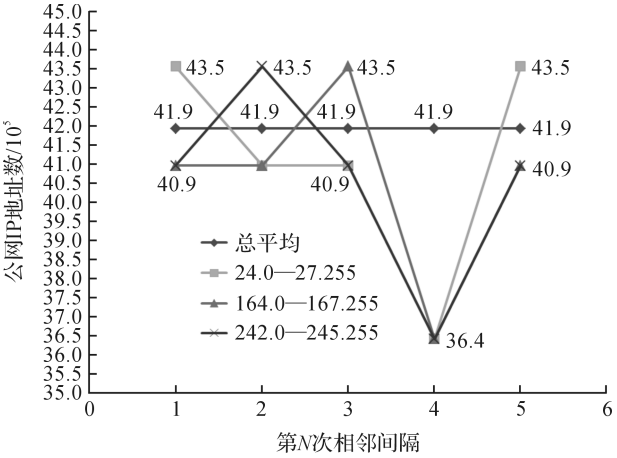

图 1　输出序列中相应 IP 地址间隔

图 1 中横坐标标目表示 IP 地址置乱特性阈值 $R=4$ 的情况下,同一 IP 地址段内不同 IP 地址再次被感知的记录数;纵坐标标目表示在同种情况下,同一 IP 地址段内不同 IP 地址再次被感知时的公网 IP 地址间隔数;标值线表示三个高校内 IP 地址段(以 C.D 位表示)同段内不同 IP 地址再次被感知时的 IP 地址间隔数变化情况,其中总平均表示每个 IP 地址段的间隔数均值。

本文扫描策略与随机扫描策略和段内扫描策略的均值比较情况见表 7。

表 7　感知策略比较

策略	本文策略	随机策略	段内策略
单位时间段感知强度 $I/(n\cdot s^{-1})$	0.006	$0<I\leqslant 1024.000$	1024.000
整体段间采样时间 Ta/s	177.758	$177.758\leqslant Ta\leqslant 176342.654$	176342.654
同地址段扫描间隔 Si/s	177.758	$2.777\leqslant Si$	2.777

注:表中数据均以 $R=4$ 即一个段内 1024 个 IP 地址为标准。

表 7 中本文策略下的实验数据是在图 1 总平均间隔数下的统计数据,策略列中单位时间段感知强度为本实验环境下单位时间内扫描同一网段 IP 的平均次数;整体段间采样时间是指在同一机构下连续的 IP 地址段内主机或节点基本共享一套检测机制的情况下,以 IP 地址段为单位,完整感知一次网域的平均最短时间;同地址段扫描间隔为完成基本完整的态势感知探测行为下,相同地址段内不同 IP 地址相邻感知平均最短时间间隔。因随机策略的感知随机性,该策略的统计数值极不稳定,仅列出本实验环境下的边界值,且边界值仅为理论情况。通过比较表 7 中的数据,本文策略与随机策略和段内策略相比,I 与 Ta 可以维持低水平稳定状态,Si 可以维持较高水平的稳定状态。

3 结 论

本文主要提出了一种安全态势感知下态势要素主动提取阶段的一种IP地址段间随机跳跃式的感知扫描策略,实验结果表明,该方法具有以下优势:

a) 在不影响网络安全态势感知整体速率的情况下能根据R值动态调整扫描间隔,较好地降低对同一连续IP地址段的段内感知强度,提高整体的感知效率。

b) 本文提出的输出策略在继承随机扫描策略输出结果混沌性的特点的同时,解决了单位时间内IP地址段内扫描策略下同IP地址段内IP感知行为过于频繁的问题。

本文的研究仅在单一的安全感知信源节点处进行,然而一个感知系统内的安全感知的信源节点通常不止一处,如何有效在不同信源节点分布式的协同此策略将会是后续的工作。

参考文献:

[1] 席荣荣,云晓春,金舒原,等. 网络安全态势感知研究综述[J]. 计算机应用,2012,32(1):1-4.

[2] Singh M, Bhandari P. Building a framework for network security situation awareness[C]//International Conference on Computing for Sustainable Global Development. IEEE,2016:2578-2583.

[3] Li C Q, Qiu G W. Research on network security threat management base on situation awareness platform[J]. Cyberspace Security,2017,8(1):19-23.

[4] 龚俭,臧小东,苏琪,等. 网络安全态势感知综述[J]. 软件学报,2017,28(4):1010-1026.

[5] 李林. 网络安全态势感知系统设计与关键模块实现[D]. 北京:北京邮电大学,2015:5-15.

[6] Mathews M, Halvorsen P, Joshi A, et al. A collaborative approach to situational awareness for cybersecurity[C]// International Conference on Collaborative Computing: Networking, Applications and Worksharing. IEEE,2017: 216-222.

[7] Cheng W Z, Chen Z G, Deng X H, et al. Network security situation awareness method based on multi-source and multi-level information fusion[J]. Journal of Shanghai Jiaotong University,2015,49(8):1144-1152.

[8] 张勇. 网络安全态势感知模型研究与系统实现[D]. 合肥:中国科学技术大学,2010:2-6.

[9] 马超,程力,孔玲玲. 云环境下SDN的流量异常检测性能分析[J]. 计算机与现代化,2015(10):92-97.

[10] Kuze N, Shu I, Yagi T, et al. Detection of vulnerability scanning using features of collective accesses based on information collected from multiple honeypots[C]// 2016 IEEE/IFIP Network Operations and Management Symposium. IEEE,2016:1067-1072.

[11] Leau Y B, Khudher A A, Manickam S, et al. An adaptive assessment and prediction mechanism in network security situation awareness[J]. Journal of Computer Science, 2017,13(5):114-129.

[12] 纯真. IP地址数据库[DB/OL]. [2017-07-30]. http://www.cz88.net/.

[13] 文昌辞,王沁,苗晓宁,等. 数字图像加密综述[J]. 计算机科学,2012,39(12):6-9.

[14] 李用江. 数字图像置乱算法的研究[D]. 西安:西安电子科技大学,2011:62-68.

[15] 谢国波,丁煜明. 基于Logistic映射的可变置乱参数的图像加密算法[J]. 微电子学与计算机,2015(4):111-115.

[16] Jacob G, Murugan A. On the construction of doubly even order magic squares[EB/OL]. (2014/02/11)[2017-07-30]. https://hal.archives-ouvertes.fr/hal-00945129.

[17] 曹光辉,胡凯,张兴. 图像置乱度评估的层次分析法[J]. 中国图象图形学报,2014,19(6):868-874.

[18] 郭琳琴. 图像置乱及置乱度评价方法综述[J]. 西安文理学院学报:自然科学版,2013,16(3):49-52.

Research on network security situation awareness scanning strategy based on Arnold transformation

YUE Wei, ZHANG Huaxiong

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310000, China)

Abstract: Under the condition where the overall receiving and transmitting speed of network security situation awareness message remains unchanged, an optimizing strategy for network security situation awareness is proposed to reduce high-frequency and high-intensity perception behavior of the targets in the same IP or continuous IP address field. Based on the classical Arnold scrambling algorithm, this strategy combines the characteristics of the continuous IP address field of the public network and takes advantage of flexible output characteristic of generalized Arnold coefficient matrix to perceive network security situation in full domain IP address by chaotic scrambling algorithm. This method improves overall perception efficiency and reduces perception frequency and intensity of targets in the same IP or continuous IP address field, compared with the perception mode of mainstream random scanning and in-field scanning strategy.

Key words: Arnold transformation; IP address; scrambling scan; network security situation awareness

(责任编辑:康 锋)