DCA 算法在僵尸网络检测中的应用

杨智兴,方贤进
(安徽理工大学 计算机科学与工程学院,安徽 淮南 232000)

摘 要:Greensmith 根据树突状细胞机理设计实现了树突状细胞算法DCA(Dendritic Cell Algorithm),并将其用于入侵检测系统中。实验结果表明,该算法具有较高的效率,并得到了广泛的应用。本文对 DCA 算法进行了描述并简要介绍了其在僵尸网络检测中的应用,具有实际意义。

关键词:僵尸网络;DCA;入侵检测
中图分类号:TP391 **文献标识码**:A **文章编号**:1674-7720(2014)06-0087-02

Application of DCA algorithm in botnet detection

Yang Zhixing, Fang Xianjin
(College of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232000, China)

Abstract: According to the mechanism of dendritic cells, Greensmith has designed to achieve a dendritic cell algorithm for the intrusion detection system. Experimental results show that the algorithm has higher efficiency, and has be widely used. This paper describes the DCA algorithm and briefly depicts its application in brief botnet detection, and it is considered to have practical significance.

Key words: botnet; DCA; intrusion detection

僵尸网络是在网络蠕虫、特洛伊木马、后门工具等传统恶意代码形态的基础上发展、融合而产生的一种新型攻击方式。目前一个具有通用性的定义是:僵尸网络(botnet)是攻击者出于恶意目的,传播僵尸程序控制大量主机,并通过一对多的命令与控制信道所组成的网络。

根据国家互联网应急中心 2013 年 10 月[1]公布,CN-CERT 监测发现境内近 740 000 个 IP 地址对应的主机被木马或僵尸程序控制,木马或僵尸网络控制服务器 IP 总数为 12 275 个。其中,境内木马或僵尸网络控制服务器 IP 数量为 6 763 个,按地区分布数量排名前三位的分别为广东省、江苏省、云南省。境外木马或僵尸网络控制服务器 IP 数量为 5 512 个,主要分布于美国、韩国、中国台湾。其中,位于美国的控制服务器控制了境内 303 588 个主机 IP,控制境内主机 IP 数量居首位,其次是位于葡萄牙和荷兰的 IP 地址,分别控制了境内 135 178 个和 109 893 个主机 IP。

僵尸网络的肆虐给网络安全带来了巨大的威胁,现阶段针对僵尸网络的检测研究[2]也层出不穷,本文简要介绍 DCA 算法在僵尸网络检测中的应用。

1 DCA 算法

人类免疫系统 HIS(Human Immune System)中免疫应答是从 DC 开始的复杂过程。DC 是一种抗原提呈细胞(APC),它从淋巴系统迁移到机体组织(Tissue),摄取抗原和蛋白质碎片,同时采集抗原所处环境中的分子作为危险信号,摄取抗原并采集信号之后从机体组织返回淋巴结(Lymph Node),并将抗原提呈给 T 细胞以识别抗原。另外,DC 能够处理环境分子,并释放特定的细胞因子(cytokines)以影响 T 细胞分化过程。DC 进行决策并驱动 T 细胞进行免疫应答。

Greensmith 通过对 DC 生理功能和角色的研究,对 DC 行为进行建模,设计实现了 DCA 算法。

DCA 是基于 DC 群体(population)的算法[3-4],对抗原信号形式的数据流进行处理。DC 群体不断更新,更新频率和种类控制与算法实现细节有关。群体中每个 DC 执行抗原和信号的采集。DC 存储采集的抗原,并将输入信号转换为输出信号。

DCA 输入信号包括 PAMP、DS (Danger Signal)、SS(Safe Signal)和 IS(Inflammation Signal)。DC 对输入信号进行处理,产生 3 种输出信号——CSM、半成熟信号(semi)和成熟信号(mat)。为了避免对复杂的实际生物信号转换机制建模,iDC 信号处理使用加权求和公式来模拟,以减

少计算开销，使 DCA 适用于实时异常检测。DC 每次更新累积输出信号之后，比较 CSM 和迁移阈值(Migration Threshold)，若 CSM 超过迁移阈值，则从组织删除此 DC，采样周期结束，DC 迁移到淋巴结进行结果分析。

DC 迁移之后进行累积输出信号评估，semi 和 mat 浓度较大者成为细胞环境。用于对 DC 采集的所有抗原进行标记，标记成环境 0 或者 1，最终用于产生 MCAV，代表抗原异常程度。用户可以将其与阈值进行比较，判断抗原是否异常。

DCA 算法流程如图 1 所示[5]。

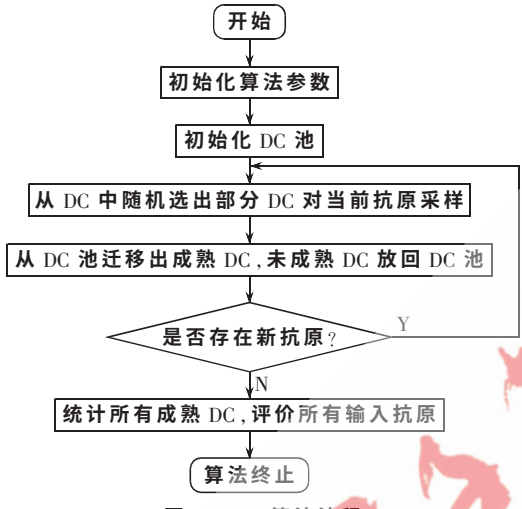

图 1 DCA 算法流程

2 DCA 算法在僵尸网络检测中的应用

基于群体的 DCA 算法的输入为时间序列数据，是由信号与抗原组成的数据流。要将 DCA 算法应用到僵尸网络的检测中[6]，最重要的是要解决算法中的时间序列输入数据的获取以及信号与抗原的映射，下面给出解决方案。

2.1 算法输入数据的获取方法

检测算法的输入数据应能反映受害主机(即受僵尸程序或其他恶意程序感染的主机)状态。通过对僵尸程序的分析得出，僵尸网络要完成一些列的功能必须通过调用相关的系统函数才能够实现，如通信函数 C、文件存取函数 F 和键盘状态函数 K。其中通信函数包括 send、sendto、recv、recvfrom、socket、connect 和 IcmpSendEcho；文件存取函数包括 CreateFile、OpenFile、ReadFile 和 WriteFile；键盘状态函数包括 GetAsyncKeyState、GetKeyboardState、GetKeyNameText 和 keybd_event。通过 API 调用追踪工具能很简单地得到这些函数的调用数据，作为算法的时间序列输入数据。

2.2 算法输入信号的选择、映射及归一化方法

现阶段，僵尸网络的命令与控制信道的构建方式趋于多样化，基于不同协议构建的僵尸网络信号的选择、映射及归一化方法不尽相同，这里以基于 IRC 协议的僵尸网络为例，给出输入信号的选择、映射及归一化方法。

PAMP 信号的映射：把 bot 执行的键盘拦截活动映射为 PAMP 信号，该信号值来自于键盘拦截活动需调用的相关 API 函数的调用变化率，这些 API 函数包括 GetAsyncKeyState、GetKeyboardState、GetKeyNameText 和 keybd_event。通过初级试验定义“Maxps=1 s 内击键所产生的 API 函数调用的最大次数”，然后通过线性变换将 Maxps 映射到 100 作为 PAMP 信号的最大值。那么设 $PAMP_t$ 为在时间窗 t 内所产生的键盘状态 API 函数调用的数量。则在 t 时刻，PAMP 信号的变化率可定义为：

$$PAMP_t=\frac{PAMP_t-PAMP_{t-1}}{Max_{ps}}\times 100,\quad \forall t$$

(1)Danger 信号的映射：由于 bots 直接对 botmaster 的命令响应，因此发送和接收数据的微小时间差都可观测到。因此将 Danger 信号定义为对每个进程拦截 send 和 recv 函数调用的网络发送和接收数据的时间差。设定一个临界范围(0，Max_{ds})表示异常响应时间，如果响应时间落在临界范围则表示很快的响应时间，并被认为是危险的。Danger 信号的计算公式如下：

$$Danger=100\times\left(1-\frac{T_{recv,send}}{Max_{ds}}\right)$$

其中 $T_{recv,send}$ 为执行 recv 和 send 函数调用的时间差。

(2)Safe 信号的映射：由于 bots 发送给 botmaster 信息使用 send 函数调用，或者 bots 发起 SYN/UDP Flooding 攻击使用 sendto 和 socket 函数调用，这都会在短时间内产生大量的函数调用。因此定义($Range_{ss1}$，$Range_{ss2}$)为调用两个连续通信函数的时间差，例如(send，send)、(sendto，sendto)、(socket，socket)。然后将这些时间差归一化转换成范围[0，10]，再划分成 3 个子范围。Safe 信号的计算公式如下：

$$SAFE=\begin{cases}\Delta T\times 0.2, \text{if } \Delta T\in[0,5]\\ \Delta T\times 0.5, \text{if } \Delta T\in(5,20]\\ \Delta T, \text{if } \Delta T\in(20,\infty]\end{cases}$$

其中 ΔT 就是调用两个连续通信函数的时间差。

2.3 算法输入数据中抗原的映射方法

抗原有可能就是潜在的恶意进程，是系统状态的反映，它有可能就是造成系统状态改变的因素，因此将产生 API 函数调用的进程 PID 映射成抗原。

本文简单阐述了 DCA 算法的机理，描述了将 DCA 算法应用到僵尸网络检测中各个信号量的定义。

针对 DCA 算法在僵尸网络检测中应用的下一步研究，包括基于 P2P 协议的僵尸网络、基于 HTTP 协议的僵尸网络及无协议特征的僵尸网络的信号量的定义，并对实验结果进行分析。

参考文献

[1] 国家互联网应急中心.CNCERT/CC.CNCERT 互联网安全威胁报告[R].[2013-10-17].http://www.cert.org.cn.

[2] 王海龙，龚正虎，侯婕.僵尸网络检测技术研究进展[J].计算机研究与发展，2010，47(12)：2037-2048.

[3] 陈岳兵,冯超,张权,等.基于 DCA 的数据融合方法研究[J].信号处理,2011,27(1):102-105.
[4] 陈岳兵,冯超,张权,等.树突状细胞算法原理及其应用[J].计算机工程,2010,36(8):173-176.
[5] 邓小武,李森林,胡萍.树突状细胞算法形式化及其在入侵检测中的应用[J].青岛科技大学学报,2013,34(4):88-92,96.
[6] AI H Y,AICKELIN U,GREENSMITH J.DCA for bot detection[C].Proc.of CEC'08,2008.

(收稿日期:2013-12-23)

作者简介:

杨智兴,男,1988 年生,硕士研究生,主要研究方向:信息安全。

方贤进,男,1970 年生,博士研究生,主要研究方向:信息安全、智能计算。