树突细胞算法在线分析组件的研究

安丽丽,方贤进
(安徽理工大学 计算机学院,安徽 淮南 232001)

摘 要:原始树突状细胞算法(DCA)的离线分析过程,将会导致时间差异,从而产生假警报,增加了虚警率,也会导致攻击的成功发生,这对一个入侵检测系统来说是致命的。因此,文中的目的就是在不影响检测精度的前提下提高检测速度。于是文中提出了分片思想的在线分析组件与 DCA 相集成的方法,即根据抗原采样数量或者时间将一系列已处理的信息分割成为更小的部分,使得每个分片独立地进行实时的、周期性的分析,这样在每个分片内的入侵攻击就能及时地被识别出来。文中给出了 DCA 在线分析模块的伪代码描述,并且将其应用于 SYN 端口扫描的检测实验中。结果表明,DCA 在线分析模块在不影响检测精度的前提下有效地提高了检测速度。

关键词:在线分析组件;危险理论;树突细胞算法;分片
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2014)01-0147-04
doi:10.3969/j.issn.1673-629X.2014.01.038

Study of Dendritic Cell Algorithm Online Analysis Module

AN Li-li, FANG Xian-jin
(School of Computer, Anhui University of Science & Technology, Huainan 232001, China)

Abstract: The analysis process of original dendritic cells algorithm (DCA) is offline, which results in time difference, producing false alarms and increasing false negative rate, and leading to the success of the attack, which is fatal for an intrusion detection system. Propose integrating online analysis with the DCA using segmentation idea, that is, segmentation involves partitioning a sequence of processed information into relative smaller segments, in terms of the number of data items or time. The analysis is performed within each individual segment. Intrusions appeared within the duration of this segment can be identified. The pseudo code of DCA with online analysis module is also presented, and it is applied to experiments of detection of SYN port scan. The experimental results indicate that the DCA with online analysis module can effectively improve detection speed without compromising detection accuracy.

Key words: online analysis module; danger theory; dendritic cells algorithm; segmentation

0 引 言

2005 年第 4 届国际人工免疫学会议上由 Julie Greensmith 等人提出一种基于“危险理论”的全新算法-树突状细胞算法(Dendritic Cell Algorithm, DCA)[1],随后成功应用于计算机良好的性能安全中的 SYN 端口扫描检测[2]、僵尸网络(botnet)检测[3]以及无线传感网络和机器等方面。DCA 在这些方面的成功应用显示其良好的算法性能,并且与其他方法(如支持向量机[4]、NSA[5]、C4.5 决策树[6]等)相比较 DCA 降低了误报率,显示了其在检测方面的良好的性能。

然而到目前为止,DCA 的分析过程只是离线的,不能及时地鉴别入侵,更不用说及时地对入侵行为产生进一步的响应,这也将导致攻击的成功发生,而这对于 IDS 来说是致命的问题。因此 DCA 在线实时分析是迫在眉睫的,文中主要是进行 DCA 在线实时分析组件的研究。

1 原 DCA 离线分析组件原理及实现异常分析的方法

原 DCA 算法包括系统初始化,数据流程和离线分析[7]。系统初始化包括生成初始化值。在初始化阶段后,输入数据(信号与抗原)被送到数据处理阶段。该阶段有三项子功能模块:数据分配模块、信号变换模块和时间相关模块。首先数据赋值将输入数据中的抗原和信号分离出来,然后将信号交给信号变换模块,将抗原分配给特定 DCs。然后信号转换模块将输入信号转

收稿日期:2013-03-25 **修回日期**:2013-06-27 **网络出版时间**:2013-11-12
基金项目:国家自然科学基金资助项目(61240023)
作者简介:安丽丽(1986-),女,硕士研究生,CCF 会员,研究方向为网络与信息安全;方贤进,教授,博士,研究方向为计算智能与信息安全。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20131112.1636.026.html

换为输出信号。在文中用户自定义的输入流中使用 PAMP、Danger、Safe 三类信号。PAMP 是指每秒接收到错误信息的数量;Danger 是指异常信号,指示值高则表示可能异常,指示值低则表示其可能正常;Safe 是指正常信号,是所观测到的较低指示值的信号。

“CSM” 和“K” 这两个的信号转换得到了输出信号[8],其数值由式(1) 得到:

$$O_j = \sum_{i=0}^{n} (W_{ij} \times S_i) \quad \forall j \tag{1}$$

式中,O_j 表示输出信号;n 表示输出信号类别的数量值减去1;S_i 是指输入信号;W_{ij} 是指从 S_i 到 O_j 的转换权值。其输出信号可以被所有的 DCs 访问。每个 DC 在信号和抗原内部之间执行一个时间相关函数,每个独立的 DC 通过迁移阈值生成一个时间窗,信号和抗原在该时间窗内相互关联。为了相关的正确性,假定信号是在抗原出现之后出现,并且时间间隔非常小,小于前面所生成的时间窗[9]。与此同时,系统对输出信号进行求和。一旦一个 DC 的累加值 CMS 超过了迁移阈值(如式(2) 所示),则该 DC 的状态将会转变为一个成熟 DC。同时该 DC 将停止信号转换和时间关联。K 的累加值和抗原一起组成“已处理信息” 被呈递给分析阶段。一旦一个成熟 DC 呈递了“已处理信息” 它将被重置为半成熟 DC,因此 DCA 初始化规模不会改变。

为了简化表示省略了信号转换中的连接权值,根据 DCs 三种形态有以下两个公式:

$$\mathrm{CSM} = P + S + D \tag{2}$$

$$\begin{aligned} k &= \mathrm{Mature} - \mathrm{Semi} - \mathrm{mature} \\ &= P + D - S - S \\ &= P + D - 2S \end{aligned} \tag{3}$$

在分析阶段,所有的被成熟的 DC 呈递的“已处理信息” 是为分析处理阶段做准备的。分析结果是为了衡量一个抗原类型是正常还是异常,在此方法中引入了一个中间值 K 来减少信号转换的输出(从三个输出减少为二个输出),计算公式如下所示[10]:

$$K_\alpha = \frac{\sum K_i}{\sum \alpha_i} \quad \forall i \tag{4}$$

式中,α_i 表示通过 DC_i 采样的抗原类型 α 的数量;K_i 表示的是 K 的累积值;K_α 的值越大表示抗原类型 α 异常的可能性越大,反之则其可能性越小。

如上所述,在原来 DCA 中分析过程是在执行数据处理阶段后离线进行的,是在将所有的已处理信息呈递完成之后才会进行分析,这将产生时间差,将会导致攻击的成功发生,这对一个有效的入侵检测系统来说是不够有效的。因此设想把分析阶段移动到数据处理阶段,这样在数据处理阶段同时进行分析以达到实时分析的目的来提高系统的检测效率。

2 DCA 在线分析的研究

2.1 DCA 中在线分析的实现方法、原理

DCA 中的一个实时分析组件主要应用于一个有效的入侵检测系统中。一个有效的、功能齐全的入侵检测系统应当尽可能快地、准确地识别入侵。为了改进原离线 DCA 存在的问题,文中对实时分析方法进行研究,将分析阶段移动到数据处理阶段,实现边处理数据边进行分析的实时、周期性分析的方法(如图 1)。

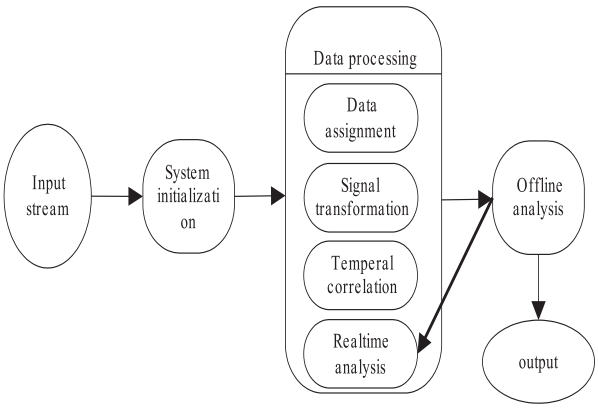

图 1 实时分析方案

文中 DCA 所采用“分片”(segmentation)思想是根据抗原采样数量进行分片的。该思想是将 DCs 的输出值与它所取样的抗原序列分割成相对小的片,处理一个 segment,然后呈递并分析检测该 segment,与此同时处理呈递另一个小分片,从而增强了 DCA 的检测速度,提高了其检测效率(如图 2)。

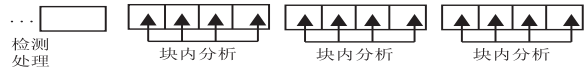

图 2 分割成块并各自进行块内分析

文中只介绍静态固定分片的大小对系统的影响,并为以后动态分片的研究打下基础。对每一个分片 segment 进行分析都会产生一个抗原类型 α 的检测结果 K_α,这样产生的就是多个检测结果集合,比离线分析得到的一个检测结果集合具有更高的检测速度和检测效率,如表 1 所示。表中基于采样抗原数量,根据分片思想将 processed message 分成四个 segment。

表 1 在 DCA 中拟采用的实时分析原理

Segment1	Segment2	Segment3	Segment4
$(K_1(\alpha_1),\alpha_1)$	$(K_2(\alpha_3),\alpha_3)$	$(K_3(\alpha_2),\alpha_2)$	$(K_4(\alpha_1),\alpha_1)$
$(K_1(\alpha_4),\alpha_4)$	$(K_2(\alpha_4),\alpha_4)$	$(K_3(\alpha_5),\alpha_5)$	$(K_4(\alpha_2),\alpha_2)$
$(K_1(\alpha_7),\alpha_7)$	$(K_2(\alpha_7),\alpha_7)$	$(K_3(\alpha_3),\alpha_3)$	$(K_4(\alpha_6),\alpha_6)$
$(K_1(\alpha_{10}),\alpha_{10})$	$(K_2(\alpha_{10}),\alpha_{10})$	$(K_3(\alpha_8),\alpha_8)$	$(K_4(\alpha_9),\alpha_9)$

注:设定一个阈值 θ,只要任何一个 $K_m(\alpha_i) > \theta$,则判断抗原 α_i 为异常,$m = 1,\cdots,4$,$i = 1,\cdots,10$。

基于抗原数量的分片方法是指抗原抽样数量达到分片大小时,就将其呈递、分析。通过对该分片方法的研究分析,可以从不同方面对入侵系统进行探测研究。从而可以提供更多的,对未来进一步发展动态分片有帮助的算法见解。

以下给出实时分析的伪代码:

```
Input:antigens and signals //由抗原和信号组成的输入流,根据时间戳生成排列顺序
output:antigen types+ K_α
set DC population size;
initialize DCs;
Set segmentation size segment_length;//设置分片大小为 segment_lenght
Set anomaly metric threshold θ ; //设置异常迁移阈值
length_of_lst=0;
While data do
switch input do
Case antigen
agCounter++;
cellIndex = agCounter % populationSize;
DC of cellIndex assigned antigen;//将抗原指派给 DC 细胞的索引号为 cellindex
update DC's antigen prole;
End
Case signa
calculate csm and k ;
foreach DC do
DC. lifespan -=csm;
DC. sum K + = k ;
if DC. lifespan <= 0 then//当 csm 超过迁移阈值时
Append((DC. sumK, antigen), lst);//记录该 DC 收集的抗原与信号累计值并存储到表 lst 中
length_of_lst++;// 将由 DC 呈递的 processed informationlst 存储在 lst 中
reset DC;
if (length_of_lst % segment_length==0)//processed information 达到分片大小
draw segment data from lst[length_of_lst] to lst[length_of_list-segment_length+1];
for each antigen type do//对当前片的信息进行实时的分析
calculate K_α;
if K_α > θ// 其值超过阈值就会发出异常警报
raise alarm;
End
End
End
End
End
End
End
```

2.2　DCA 实时分析的实验方法

表 2 是参数设置,表 3 是信号转换权值。

表 2　参数设置

DC Population size	Migration thresholds	Segments size
100	$12*x, x \in [1,100]$	$1*10^n, n \in \{2,3,4,5,6\}$

表 3　信号转换权值

PAMP	Danger	Safe
4	2	6
8	4	-13

(表中数据第一行为输出值 CSM 的转换权值,第二行为输出值 K 的转换权值)

文中对基于 segmentation 在线分析组件的 DCA 进行测试,所有的实验都运行在 Intel Pentium 4 CPU 3.2 GHz (OS RedHat fedora core release 1, kernel 2.4.22-1)环境中,C 编程采用的编译器为 GNU gcc 4.0.1。实验中所用数据集是在实际网络环境中利用 SYN scan 扫描[11]接收的,并将其作为系统的原始输入数据,由于数据集是通过 ssh 连接得到的,因此其中既有正常进程也有异常进程。

(1)在文中的 DCA 在线分析方法中输入数据同样采用三类信号:PAMP、Safe 和 Danger。

①PAMP 表示每秒接收到的 ICMP Destination Unreachable (DU)错误信息的数量。

②Danger 表示的是扫描主机网卡所处理的 TCP 包与所有其他数据包的比率,它的出现一定意味着某些异常的发生。

③Safe 则表示在 SYN 扫描过程中网络数据包的平均大小降低到 40 bytes 时所观测到的,扫描攻击一般倾向于发送大量的小的数据包,所以大数据包一般表示网络的正常行为。

由于采用每一秒钟捕获一个信号集的方法,因此所有信号的大小被规格化为[0,100]之间。这些通过 SYN scan 扫描的信号(如图 3)所示。

(2)对 antigens 类型的定义和映射如下:

将主机上系统调用的高频率出现的进程 ID(PID)作为抗原个体,这些抗原类型有 Nmap、Firefox 和 pts 三类,其扫描数据如图 4 所示。Nmap 指的是对受害者主机进行 SYN 端口扫描的进程。而 pts 则是作为一个辅助伪终端守护进程存在,Nmap 是其子进程。Firefox 表示的是一个 Web 浏览器在实验中被看作是正常进程,而另外两个则与之相反。Nmap 和 pts 被认为是入侵攻击行为,在系统扫描中对于进程 Nmap 和 pts 期望得到较高的 K_α 值。而被认为是一个正常行为的 Firefox 进程,则会得到较低的 K_α 值。在实验中收集的数据包

括 1 300 万多个抗原实例和超过了 4 800 个信号集。

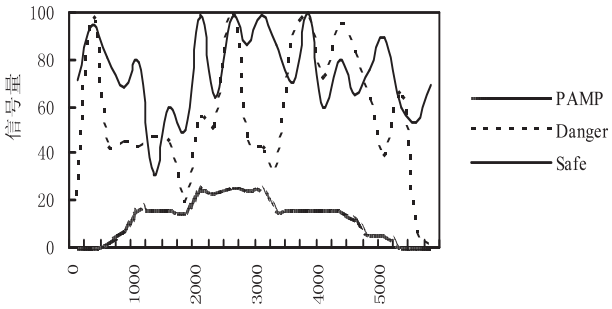

图 3 随时间推移的信号输入值

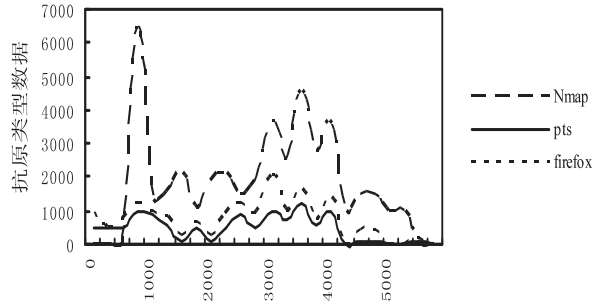

图 4 随时间推移的抗原类型数量

原 DCA 没有进行分片或者说即为一片,而现在的将分片思想引入其中的在线分析 DCA 中,分别将其分片为 $1*10^2$、$1*10^3$、$1*10^4$、$1*10^5$ 和 $1*10^6$,计算 K_α 值,然后将单样本单侧 t-test($\alpha=0.05$)[12] 原 DCA 与将其分为以上各个分片的双样本双侧测试的现 DCA 得到的结果进行比较,比较结果如表 4 所示。表中第一行表示分片大小,第二,三,四行分别表示进程 Nmap,Firefox 和 pts 的 K_α 值。

表 4 两种方法进行 t-test 检测结果 K_α 值的比较(☆表示有较大差异)

1	$1*10^2$	$1*10^3$	$1*10^4$	$1*10^5$	$1*10^6$
-970.34	< 0.05 ☆	< 0.05 ☆	< 0.05 ☆	0.16	0.38
-1 390.23	< 0.05 ☆	< 0.05 ☆	< 0.05 ☆	< 0.05 ☆	< 0.05 ☆
-1 002.12	< 0.05 ☆	< 0.05 ☆	< 0.05 ☆	0.25	0.41

从表 4 中可以看出当分片大小为 10^2、10^3 和 10^4 时,对于 Nmap 进程和 pts 进程两种算法结果有很大的不同,可以说后者有更好的检测结果。而当分片大小为 10^5 和 10^6 时,对于进程 Firefox 的检测结果两者有着明显的不同。因此将分片应用到 DCA 中可以快速地识别入侵以提高检测系统的性能。分片能够实时分析处理呈递的 DCs,在不改变检测精度的前提下提高检测速度,从而提高系统的性能,多次分片的反复实验比较,降低了虚警率,进一步提高系统的检测精度。因此将分片思想应用到 DCA 的在线实时分析的方法是可行并且有效的。

3 结束语

文中是在 DCA 离线分析的基础上,为了避免因时间问题而导致攻击的成功发生提出了在线实时分析的方法。在该方法中应用“分片”思想以实现实时分析的目的,提高了入侵检测的检测精度和效率。文中采用的是固定的、静态的分片的方法,下一步要进行对动态自适应分片方法的研究。

参考文献:

[1] Greensmith J, Feyereisl J, Aickelin U. The DCA: Some comparison a comparative study between two biologically-inspired algorithms[J]. Evolutionary intelligence, 2008, 1(2): 85-112.

[2] Greensmith J. Dendritic cells for SYN scan detection[C]// Proceedings of the genetic and evolutionary computation conference. [s. l.]: Elsevier Science Limited, 2007: 49-56.

[3] Al-Hammadi A, Greensmith J. DCA for bot detection[C]// Proceedings of the IEEE world congress on computational intelligence (WCCI 2008). [s. l.]: [s. n.], 2008.

[4] Greensmith J, Aickelin U. The deterministic dendritic cell algorithm[C]//Proc of 7th international conference on artificial immune systems (ICARIS). [s. l.]: [s. n.], 2008: 291-303.

[5] Zhou Ji, Dasgupta D. V-detector: An efficient negative selection algorithm with “Probably Adequate” detector coverage [J]. Information sciences, 2009, 179(10): 1390-1406.

[6] Liao Jun, Jiang Haitao, Zhang Hong. Artificial immunity-based misbehavior detection architecture for mobile ad hoc networks [J]. Journal of Nanjing university of science and technology, 2011, 35(5): 652-658.

[7] Aickelin U, Greensmith J. Sensing danger: Innate immunology for intrusion detection[R]. [s. l.]: [s. n.], 2007.

[8] Shi Yuhui, Eberhart R. A modified particle swarm optimization [C]//Proc of ICEC’98. Anchorage, AK, USA: [s. n.], 1998.

[9] Amaral J L M. Fault detection in analog circuits using a fuzzy dendritic cell algorithm[C]//Proc of ICARIS 2011, LNCS 6825. Berlin: Springer-Verlag, 2011.

[10] Greesmith J, Twycross J, Aickelin U. Dendritic cells for anomaly detection[C]//Proc of the congress on evolutionary computation. [s. l.]: [s. n.], 2006: 664-671.

[11] Fang Xianjin, Li Jian. Investigation of DCA and its application to nmap portscan detection[J]. China communications, 2012, 2012(3): 145-152.

[12] Greensmith J, Aickelin U, Tedesco G. Information fusion for anomaly detection with the dendritic cell algorithm[J]. Information fusion, 2010, 11(1): 21-34.

树突细胞算法在线分析组件的研究

作者： 安丽丽， 方贤进， AN Li-li， FANG Xian-jin
作者单位： 安徽理工大学 计算机学院,安徽 淮南,232001
刊名： 计算机技术与发展
ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2014(1)