树突状细胞分化模型在人工免疫系统中的应用研究

倪建成[1,2],李志蜀[2],孙继荣[2],周利平[2]
(1. 曲阜师范大学计算机学院,山东日照 276826;2. 四川大学计算机学院,四川成都 610065)

摘 要: 树突状细胞(Dendritic Cell,DC)是先天性免疫系统的重要组件,其分化机制是正确引发与调节适应性免疫响应的关键.首先,在描述 DC 分化的生物机理基础上,抽象出了 DC 的信息处理过程.其次,在阐释 DAMP 等四类外部信号的含义与功能、信号融合过程的基础上,定义了未成熟、完全成熟与半成熟 DC Agent,刻画了它们的分化数学模型与演化过程.最后,论证了各类 DC Agent 数量与生存周期之间的关系.实验结果表明 DC 分化机制对降低入侵检测误报率、实现自我调节和进一步增强计算机系统安全具有重要的理论意义与应用价值.

关键词: 人工免疫; 危险理论; 树突状细胞; 分化模型; 信号融合
中图分类号: TP393 **文献标识码**: A **文章编号**: 0372-2112 (2008) 11-2210-06

Research on Differentiation Model and Application of Dendritic Cells in Artificial Immune System

NI Jian-cheng[1,2], LI Zhi-shu[2], SUN Ji-rong[2], ZHOU Li-ping[2]
(1. *College of Computer Science, Qufu Normal University, Rizhao, Shandong* 276826, *China*;
2. *College of Computer Science, Sichuan University, Chengdu, Sichuan* 610065, *China*)

Abstract: The most one critical component in innate immune system is dentritic cell (DC), which differentiation mechanism is also the key to initiate and control adaptive immune response correctly. Firstly, based on describing biological principles of DCs differentiation, the information processing procedure for DCs is abstracted. Secondly, laying the foundation of illustrating four-category behavioral signals such as danger-associate molecular patterns et al., and informational fusion procedure, agents including immature, fully mature and semi-mature DCs are defined, and detailed mathematical differentiation model are formulized and deducted. Lastly, several relations between quantity and lifecycle of different Agent category are proved. Simulation testing results demonstrate that DCs differentiation mechanism has theoretical significance and practical value on computer security besides decreasing false positive rate as well as achieving homeostasis for intrusion detection systems.

Key words: artificial immunity; danger theory; dendritic cells; differentiation model; signal fusion

1 引言

生物免疫系统[1,2]包括先天性和适应性免疫两个紧密相关子系统,但半个世纪以来,无论在生物学还是计算机安全领域,免疫理论多聚焦于具有抗原特异性特征的适应性免疫系统的研究与应用.然而,众多研究[3,4,6]表明,具有快速及抗原非特异性特征的先天性免疫具有激发、调节与控制适应性免疫响应的作用.因此,处于先天性免疫系统核心地位的树突状细胞分化机制日益引起了学者的广泛关注.

目前,基于经典自体-非自体(SNS)模型的免疫理论及其应用取得了一定的研究成果[2,14],但是,它们普遍存在误报率高、缩放性差、漏报率高等严重缺陷.免疫淋巴细胞在胸腺中进行中央耐受时,非完备自体集导致了异常检测时的高误报率;覆盖非自体空间对检测器集的规模需求制约了系统的可缩放性;同时,检测器集对非自体空间覆盖时漏洞的存在及自体的动态变异则是漏报产生的根源.危险模型[6~10]利用外围耐受机制将抗原与信号关联,生成了与时间相关的抗原上下文,克服了传统免疫响应仅利用中央耐受机制的弊端,有效解除了误报率、缩放性和漏报率问题产生的根源.鉴于 DC 分化机制及信号融合过程的复杂性,相关数学模型的建立是促进生物免疫理论进展,奠定人工免疫理论基础的前提.

2 树突状免疫细胞

DC[7,8]是先天性免疫系统中具有结构和行为特征的抗原提呈细胞.结构特征反映在抗原的捕获、处理和提呈;行为特征表现于信号的监视、融合以及趋化因子、协同刺激分子的分泌.下面,在探究 DC 分化机理基础

收稿日期:2007-09-26;修回日期:2008-04-23
基金项目:国家自然科学基金(No. 60072014);四川省科技公关项目(No. 05GG021-003-2);山东省自然基金(No. Q99G03)

上,我们抽象出用于人工免疫系统的 DC 信息融合过程.

2.1　树突状细胞的生物分化机理

DC 主要散布于外围组织和第二淋巴器官,分为未成熟 DC(ImDC)、完全成熟 DC(FmDC)和半成熟 DC(SmDC)[9].它们的区别见表 1.

表 1　未成熟、完全成熟及半成熟 DC 的主要区别

要素名称	ImDC	FmDC	SmDC
MHC-Ⅱ浓度	低	高	高
CD80/86 浓度	低	高	高
趋化因子	IL-12⁻,IL-10⁻,IL-6⁻,TNF-α⁻	IL-12⁺,IL-6⁺,TNF-α⁺	IL-12⁻,IL-10⁻/⁺,IL-6⁻,TNF-α⁻
位置	外围组织	二级淋巴器官	二级淋巴器官
状态	休眠态	免疫态	稳态
T 细胞响应类型	无反应力	Th1 免疫响应	Th2 免疫耐受

ImDC 在休眠状态下具有较低的 MHC-Ⅱ分子与协同刺激分子浓度,因而使 T 细胞没有反应力.但是,通过调节表面富含的抗原受体(如 C 型外源凝集素受体和 DEC-205)浓度,ImDC 利用模式识别受体(Pathogen Recognition Receptor,PRR)甄别由进化压力形成的病原体相关分子模式(Pathogen Associating Molecule Pattern,PAMP),感知局部组织内的多层次免疫刺激信息,并依据免疫刺激的严重程度而在细胞表面分别形成不同的 MHC 分子、标志 DC 成熟度的协同刺激分子(如 CD80,CD86 等)和决定分化状态的趋化因子(如 IL-12,IL-10 等),进而分化为具有不同形态和表型特征的 FmDC 与 SmDC,并以免疫态或稳态迁移至脾和淋巴节点等组织.

免疫刺激信息包括抗原与各层次危险信号.抗原是病原体、细菌等的缩氨酸片段;依据信号源,危险信号分为外部和内部信号:外部信号由入侵实体及被监视系统产生,反映了 DC 的行为特征,是 DC 分化的决定因素;内部信号由免疫细胞本身产生,反映了细胞间的制约关系.

具有高浓度 MHC-Ⅱ和协同刺激分子的 FmDC 在 IL-12 等的刺激作用下与 CD4 + Th1 淋巴细胞结合,进而激活 CTL 细胞以杀死细胞内病原体.SmDC 是一种对 Th2 细胞具有抑制和钝化作用的特殊成熟 DC,富含的趋化因子 IL-10 将钝化 MHC-Ⅱ分子与受体(TCR)匹配的 T 细胞并抑制其繁殖,从而产生适应性免疫的外围组织耐受.

2.2　树突状细胞与人工免疫

人工免疫是对生物免疫系统的细胞功能模型进行抽象而衍生的算法集合[11].依据 DC 的结构和行为功能,虚拟 DC 的信息处理过程与组织结构可抽象为图 1 所示的数据流图.

在图 1 中,人工免疫系统包括组织、DC 和淋巴节点等三层区室.组织区室包含多个独立 DC 区室,主要用于存储和预处理系统监视 Agent 收集的多层次内外部信号和抗原相关信息;DC 区室是 ImDC 对抗原和信息采集、处理、融合与实现分化的区域,分化形成的 FmDC/SmDC 携带上下文相关抗原迁移至淋巴区室;淋巴节点区室则是 FmDC/SmDC 与 T 淋巴细胞互相识别、引发与调节适应性免疫响应,最终移除或容忍已识别的入侵病原体的区域.

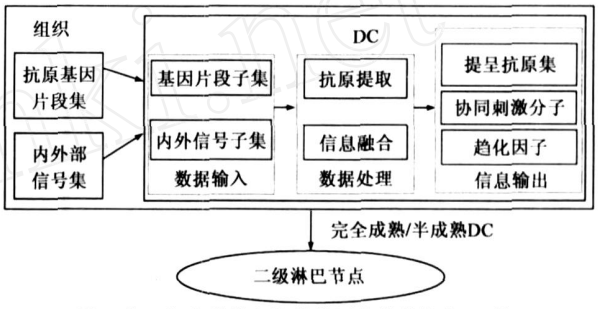


图1　人工免疫系统中组织的区室化结构与DC的抽象信息处理过程

3　树突状细胞的分化理论模型

3.1　抗原与信号

抗原是人工免疫系统应用问题域解空间的结构化特征向量.对于入侵检测问题,依据 Polly Matzinger 于 1994 年提出的危险理论[6],我们将主机系统定义为组织,进程 ID(PID)及系统调用 ID(SID)前向序列对[5]作为抗原基因片段,因而,抗原可表示为序列对 $< pid, < sid_i, sid_j >_k >$.其中,$pid$ 为进程内部标识符,$< sid_i, sid_j >_k$ 为系统调用前向序列对.

信号是人工免疫系统所解决问题的行为特征.内部

表 2　四类外部信号的区别与联系

名称	功能	关系	示例	含义
DAMP	促进 *FMC* 及 *CSM* 的分泌	受 *S* 抑制	连接错误;ICMP PING 协议产生的"目标不可达"错误等	建立在既定的广义(指它不能单独断言病原体是有害的)攻击特征之上的属性度量
D	促进 *FMC* 及 *CSM* 的分泌	受 *S* 抑制	急剧变化的网络流量、进程占用的 CPU 时间及主存容量等	建立在病原体产生的广义危害结果之上,是衡量异常程度的属性度量
S	促进 *SMC* 及 *CSM* 的分泌,抑制 *FMC* 分泌	抑制 *DAMP* 与 *D* 信号	稳定的网络流量、进程占用的 CPU 时间及主存容量等	是系统行为处于稳定状态的指示器和用户正常行为的属性度量
I	促进 *CSM*、*FMC* 及 *SMC* 的分泌	放大或缩小其它三类信号作用	与监视主机相连且处于安全或异常状态的网络主机数;攻击与系统/程序漏洞的关联度等	是放大其它信号效能,反映系统整体状态的属性度量

信号包括协同刺激分子(*CSM*),完全成熟细胞因子(*FMC*)和半成熟细胞因子(*SMC*):*CSM* 决定了 *DC* 的迁移;*FMC* 表示抗原所处组织环境的危险程度;*SMC* 表示抗原所处组织环境的安全程度.外部信号包括与 *PAMP* 对应的危险相关分子模式(*DAMP*)、危险信号 *D*、安全信号 *S* 及致炎信号 *I*.它们的区别与联系见表 2.

3.2 信号融合

危险信号的融合及内部信号的生成是非常复杂的过程,至今仍有许多未知领域.本文借鉴 DCA 算法[11]采用的加权度量函数进行信号融合处理.

假设给定的四类外部信号向量分别包含 *M* 项元素.表 3 描述了输入信号 *DAMP*, *D*, *S* 分别影响内部信号 *CSM*, *SMC*, *FMC* 的权重,则对 $p \in \{CSM, FMC, SMC\}$,有式(1)成立.

表 3 输入信号类对输出信号类的影响权重

W_{ijp}	$j=1$(*DAMP*)	$j=2$(*D*)	$j=3$(*S*)
$P=1$(*CSM*)	w_1	$0.5\ w_1$	$1.5\ w_1$
$P=2$(*SMC*)	0	0	w_2
$P=3$(*FMC*)	w_3	$0.5\ w_3$	$-1.5\ w_3$

$$CC_P(SV) = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3} W_{ijp} SV_{ij}}{\sum_{i}\sum_{j=1}^{3} |W_{ijp}|} \cdot \frac{\sum_{i} W_{i4p}(SV_{i4}+1)}{\sum_{i} |W_{i4p}|} \quad (1)$$

其中,W_{ijp} 为输入信号矢量项 i 对应信号类 j 针对输出信号 p 的权重.$i, j \in N$,且 $i \in [1, M]$, $j \in [1, 3]$. SV_{ij} 表示 SV 中信号类 j 对应该类输入信号向量域 i 的值,SV 表示四类输入信号对应 *M* 项向量域的采样值矩阵.

文献[1,4]指出了致炎信号 *I* 的作用,并给出了等值权重.事实上,相同致炎信号源对不同信号应该具有不同权重,即不同的刺激和抑制作用.因此,致炎信号的权重可由式(2)描述.

$$W_{i4p} = K_1 * I^{OS}_{i4p} + K_2 * I^{IS}_{i4p} \quad (2)$$

其中 $K_1, K_2 \in R$,为组织外部与内部致炎信号的影响系数,且 $K_1, K_2 \in [-1, 1]$; I^{OS}_{i4p} 与 I^{IS}_{i4p}(取值 0 或 1)为外部及内部致炎信号与 p 的相关性.则有,$W_{i4p} \in [-2, 2]$.

3.3 树突状细胞 Agent 分化模型

定义 1 树突状细胞 Agent(*DC_Agent*):指具有非特异性抗原识别与提呈功能以及信息处理能力的免疫细胞 Agent,表示为六元有序组 $<SV, AV, CC, T, L, age>$.其中,SV 为输入信号向量矩阵,AV 为抗原基因片段集合,CC 为输出信号向量,*T* 为 DC Agent 迁移阈值,*L* 为生命周期,*age* 为年龄.为方便描述,我们用符号"."引用向量的域值.

定义 2 未成熟 DC Agent:指尚未迁出组织区室且细胞年龄在生命周期 L_{im} 内的 *DC_Agent*,表示为:

$imDC\_Agent = \{ <SV, AV, CC, T, L_{im}, age> \mid CC.CSM < T$ 且 $age \leq L_{im}\}$

定义 3 完全成熟 DC Agent:指由组织区室迁移至二级淋巴节点的能够激活相应 T 细胞产生免疫响应,且年龄在生命周期 L_{fm} 内的 *DC_Agent*,表示为:

$fmDC\_Agent = \{ <SV, AV, CC, T, L_{fm}, age> \mid CC.CSM \geq T$ and $CC.FMC > CC.SMC$ and $age \leq L_{fm}\}$

定义 4 半成熟 DC Agent:指由组织区室迁移至二级淋巴节点的能够抑制相应 T 细胞产生免疫响应,且年龄在生命周期 L_{sm} 内的 *DC_Agent*,表示为:

$smDC\_Agent = \{ <SV, AV, CC, T, L_{sm}, age> \mid CC.CSM \geq T$ and $CC.FMC \leq CC.SMC$ and $age \leq L_{sm}\}$

假设未成熟/完全成熟/半成熟 DC Agent 集合分别用符号"∪"表示为 $\cup xxDC\_Agent$,则 *DC Agent* 集合 $\cup DC\_Agent$ 可由式(3)表示,且有式(4)成立.

$$\cup DC\_Agent = \cup imDC\_Agent \cup \cup fmDC\_Agent \cup \cup smDC\_Agent \quad (3)$$

$$\cup imDC\_Agent \cap \cup fmDC\_Agent \cap \cup smDC\_Agent = \varnothing \quad (4)$$

3.3.1 未成熟 DC Agent 模型

在生命周期 L_{im} 内,式(5)表示了每代未成熟 DC Agent 的演化过程.初始状态($t=0$)时,*imDC_Agent* 初始化向量值并被赋予固定的迁移阈值和生命周期,年龄初始化为 0.在每代进化过程中,*imDC_Agent* 在持续输入信号的刺激作用下,依据式(1)改变协同刺激分子和完全成熟/半成熟细胞因子值,并且递增其年龄.

$$imDC\_Agent(t) = \begin{cases} \{ <SV^0, AV^0, CC^0, T^0, L_{im}, 0> \} & , t=0 \\ \{ <SV^t, AV^t, CC^t, T, L_{im}, age> \mid CC^t_p(SV^t) = CC^{t-1}_p(SV^{t-1}) + CC_p(SV^t),\ age = age+1\} & , t \geq 1 \end{cases} \quad (5)$$

当未成熟 DC Agent 的协同刺激分子值超过迁移阈值时,可以依据完全成熟/半成熟因子值断言 Agent 的分化路径,即完全成熟或半成熟 DC Agent,分别由式(6)与式(7)描述.

$$fmDC\_Agent_{New}(t) = \{ x \mid x \in \cup imDC\_Agent(t), x.CC.CSM \geq x.T \text{ and } x.CC.FMC > x.CC.SMC, x.SV = 0, x.T = 0, x.L = L_{fm}, x.age = 0\} \quad (6)$$

$$smDC\_Agent_{New}(t) = \{ x \mid x \in \cup imDC\_Agent(t), x.CC.CSM \geq x.T \text{ and } x.CC.FMC > x.CC.SMC, x.SV = 0, x.T = 0, x.L = L_{sm}, x.age = 0\} \quad (7)$$

然而,当 $imDC\_Agent$ 的年龄超过其生命周期但协同刺激分子值没有达到迁移阈值或抗原子集为空集时,该 Agent 正常死亡. 式(8)描述了死亡 $imDC\_Agent$ 集合.

$$imDC\_Agent_{Dead}(t)=\{x\mid x\in \cup imDC\_Agent(t),\ x.age>x.L_{im} \text{ and } (x.CC.CSM<x.T \text{ or } x.AV=\varnothing)\} \quad (8)$$

同时,当部分未成熟 DC Agent 在 t 时刻分化或死亡后,为维持其结构和行为功能,系统需要增补相应数目的 $imDC\_Agent$,如式(9).

$$imDC\_Agent_{New}(t)=\{<SV,AV,CC,T,L_{im},age>\mid SV=0,AV=\varnothing,CC=0,T \text{ 为迁移阈值},L_{im} \text{为生命周期},age=0\} \quad (9)$$

由此,我们可用式(10)描述未成熟 DC Agent 集合的演化趋势.

$$\cup imDC\_Agent(t)=\cup imDC\_Agent(t-1)-fmDC\_Agent_{New}(t)-smDC\_Agent_{New}(t)-imDC\_Agent_{Dead}(t)\cup\cup imDC\_Agent_{New}(t) \quad (10)$$

3.3.2　完全成熟 DC Agent 模型

在生命周期 L_{fm} 内,式(11)反映了每代 $fmDC\_Agent$ 的演化过程. 初始状态($t=0$)时,$fmDC\_Agent$ 集合为空集. 在每代进化过程中,迁移到淋巴节点但暂时未被 Th1 淋巴细胞识别的 $fmDC\_Agent$ 仅递增其年龄.

$$fmDC\_Agent(t)=\begin{cases}\varnothing & ,t=0\\ \{<SV,AV,CC,T,L_{fm},age>\mid age=age+1\} & ,t\geq 1\end{cases} \quad (11)$$

若 $fmDC\_Agent$ 在生命周期内完成与 Th1 细胞的交互识别,则由 Th1 抑制其活性并导致其被动死亡,其被动死亡集合表示为 $fmDC\_Agent^{Th1}_{Dead}(t)$;若超过生命周期仍然未被 Th1 细胞识别则令其自然死亡,式(12)表示了自然死亡 $fmDC\_Agent$ 集合. 因此,正常死亡的完全成熟 DC Agent 集合包括两部分,如式(13)所示.

$$fmDC\_Agent^{age}_{Dead}(t)=\{x\mid x\in \cup fmDC\_Agent(t),\ x.age>x.L_{fm}\} \quad (12)$$

$$fmDC\_Agent_{Dead}(t)=fmDC\_Agent^{Th1}_{Dead}(t)\cup fmDC\_Agent^{age}_{Dead}(t) \quad (13)$$

所以,完全成熟 DC Agent 集合在 t 时刻的变化趋势可由式(14)描述.

$$\cup fmDC\_Agent(t)=\cup fmDC\_Agent(t-1)\cup fmDC\_Agent_{New}(t)-fmDC\_Agent_{Dead}(t) \quad (14)$$

3.3.3　半成熟 DC Agent 模型

在生命周期 L_{sm} 内,式(15)反映了每代 $smDC\_Agent$ 的演化过程. 初始状态($t=0$)时,$smDC\_Agent$ 集合为空集. 在每代进化过程中,迁移到淋巴节点但暂时未被 Th2 淋巴细胞识别的 $smDC\_Agent$ 仅递增其年龄.

$$smDC\_Agent(t)=\begin{cases}\varnothing & ,t=0\\ \{<SV,AV,CC,T,L_{sm},age>\mid age=age+1\} & ,t\geq 1\end{cases} \quad (15)$$

半成熟 DC Agent 若在生命周期内完成与 Th2 细胞的交互识别则由 Th2 抑制其活性并导致其被动死亡,其被动死亡集合表示为 $smDC\_Agent^{Th2}_{Dead}(t)$;若超过 L_{sm} 而未被 Th2 细胞识别则自然死亡,如式(16)示. 因此,死亡的半成熟 DC Agent 集合包括两部分,如式(17)所示.

$$smDC\_Agent^{age}_{Dead}(t)=\{x\mid x\in \cup smDC\_Agent(t),\ x.age>x.L_{sm}\} \quad (16)$$

$$smDC\_Agent_{Dead}(t)=smDC\_Agent^{Th2}_{Dead}(t)\cup smDC\_Agent^{age}_{Dead}(t) \quad (17)$$

因而,半成熟 DC Agent 集合在 t 时刻的变化趋势可由式(18)描述.

$$\cup smDC\_Agent(t)=\cup smDC\_Agent(t-1)\cup smDC\_Agent_{New}(t)-smDC\_Agent_{Dead}(t) \quad (18)$$

4　树突状细胞 Agent 生命周期与数量间的关系

在计算机安全领域,基于危险理论的 DC 分化机制主要应用于入侵检测问题. U Aickelin[3]指出:应用 APC 激活机制的扩展危险理论在检测快速扩散的病毒和早期入侵扫描方面具有优势. 在建立先天性免疫模型[4]、DCA 算法[11]及系统实现[13]的基础上,针对静态分类、端口扫描和 SYN[12]攻击等的系列实验验证了 DC 分化机制的有效性和可用性. 然而,上述文献皆没有从理论上给出生命周期参数对 DC 分化的影响. 但是,生命周期的设置直接影响了 Agent 数量,进而影响了系统的入侵检测率. 下面,我们给出并证明相关的定理,并假设除生命周期和各类 Agent 数量之外的系统参数及条件保持固定.

定理 1　在固定 L_{im} 情况下,完全成熟和半成熟 DC Agent 的数量之和与未成熟 DC Agent 的数量成正比.

证明:由式(10),一代未成熟 DC Agent 分化后只存在三个子集,即死亡、完全成熟和半成熟 DC Agent. 由式(8)知,未成熟 DC Agent 在既定迁移阈值条件下,其死亡数量仅与 L_{im} 和采样的抗原集合 AV 相关. 然而,在固定 L_{im} 及攻击强度时,未成熟 DC Agent 所能采集的抗原不变,因而,其死亡数量保持不变. 故命题成立.

事实上,各类 DC Agent 的死亡数量与生命周期具有一定的比例关系.

定理 2　未成熟/完全成熟/半成熟 DC Agent 的自

然死亡数量与各自生命周期成反比.

证明:由式(8)及定理1的证明,未成熟DC Agent获取抗原基因片段子集并累积其协同刺激分子值的能力显然与 L_{im} 成正比,因而,其死亡数量与 L_{im} 成反比.在纯真T细胞数目不变的情况下,由式(12)及(16),完全成熟/半成熟DC Agent的自然死亡数量仅与各生命周期 L_{fm}、L_{sm} 相关;由式(11)及(15),纯真 T 细胞与完全成熟/半成熟DC Agent结合的可能性与各自生命周期成正比,亦即其被动死亡数量与各自生命周期成正比.因而,其自然死亡数量与各自生命周期成反比.故命题成立.

推论1 完全成熟/半成熟DC Agent的被动死亡数量与各自生命周期成正比.

定理2及推论1对入侵检测问题具有重要的物理意义.被动死亡数量不仅反映了入侵检测的效率,而且在一定程度上反映了攻击的强度.然而,由式(14)和式(18)知,完全成熟/半成熟DC Agent的死亡数量局限于各自生命周期内存在的活性DC Agent数量.下面,我们分析完全成熟/半成熟DC Agent数量的决定因素.

定理3 如果未成熟DC Agent的数量固定,那么其生命周期与半成熟DC Agent的数量成正比,而与完全成熟DC Agent的数量无关.

证明:由式(1)(5)(6)知,完全成熟DC Agent的生成仅与外部信号刺激强度相关,即只有当存在严重威胁或攻击时,才导致未成熟DC到完全成熟DC的分化.因而,在刺激强度不变的情况下,未成熟DC Agent的生命周期与完全成熟DC Agent的数量无关.因而,当未成熟DC Agent的数量固定时,由定理2和定理1可知,半成熟DC Agent的数量与未成熟DC Agent的生命周期成正比.命题得证.

5 仿真实验

5.1 实验环境及参数设置

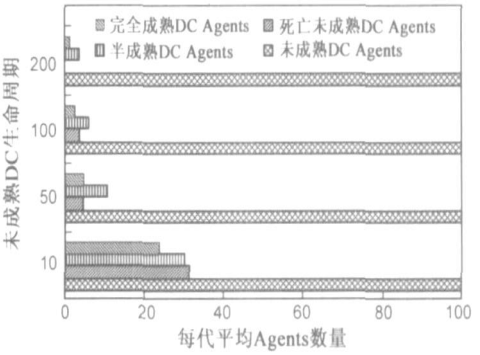
为验证未成熟DC生命周期对半成熟与成熟DC数量的影响及DC分化机制应用于入侵检测时的有效性,我们在运行Red Hat LINUX 7.2(内核版本2.4.16)操作系统、1G主存3.06GHz主频的机器上进行了多次实验.实验采用广泛部署且提供FTP服务的wu-fpt2.6.0作为异常检测应用程序,应用inetd监督程序管理,使用命令strace("strace -p pid -f -o output.file")收集进程的系统调用序列,改进的程序process_monitor(http://www.cs.nott.ac.uk/~jpt/software/)采集有关外部信号.实验共运行10次,取平均值为最终结果.

主要实验参数设置为:$w_1 = 4$, $w_2 = 1$, $w_3 = 4$;每类输入信号仅包含一个信号域,且规律性文件的数量(*num_reg*)为 *DAMP* 信号,进程使用的内存比例(*rss*)为 *D* 信号,lsof 命令报告的文件总数(*num_files*)为 *S* 信号,并将其分别规格化为[0,100]区间内实数值;组织最多容纳1500个系统调用;每DC_Agent最多存储50个抗原,含有6个系统调用受体;*imDC_Agent* 数量初始化为100,$L_{im} = (10, 50, 100, 200)$, $T = 600$,致炎信号 $I = 0$.

5.2 实验结果

在训练阶段,共收集5组正常数据且有52738个系统调用,36107个信号向量.在实验阶段,分别采用一个文件名匹配漏洞(http://www.cert.org/advisories/CA-2001-33.html)脚本、两个SITE EXEC漏洞(http://www.cert.org/advisories/CA-2000-13.html)脚本进行攻击,分别收集1916、6956、7271个系统调用以及860、2593、2750个信号向量.

图2给出了不同 L_{im} 取值对各类DC Agents数量的影响.我们可以看出:在每代未成熟DC Agents数量保持不变的情况下,伴随 L_{im} 的延长,*smDC_Agent* 总量在递增,*imDC_Agent* 死亡总量在递减,但 *fmDC_Agent* 总量并未变化.这不仅充分说明组织环境状态信息对未成熟DC Agent分化的影响,而且反映了完全成熟DC Agents数量仅决定于攻击产生的突发性信号与抗原因素,从而进一步验证了定理1与3的正确性.此外,*imDC_Agent* 死亡数量的递减不仅提高了系统的效率,而且降低了系统的误报率.



针对系统的检测效能,当 $L_{im} = L_{fm} = L_{sm} = 100$ 时,在低于4.3%的CPU占用率和小于2MB主存占用量的基础上,系统能获得83%的入侵检测率和15%的误报率,而且其响应时间小于0.2秒.检测结果表明:虽然系统因单一DC Agent断言产生了一些误报,但其资源利用率和响应时间都较传统基于适应性免疫理论的入侵检测系统采用人工响应方式时有了突出的改善,能够满足实时入侵检测系统的需求.限于篇幅,我们将在另外文章中给出群体DC Agents决策算法及其具体的演化过程描述.

6 结论

DC分化机制一方面关注攻击行为和系统状态的融合,另一方面通过不同分化路径引发并调节适应性

免疫响应.因而,集成先天性免疫的入侵检测系统可以有效克服误用检测方法存在的高漏报率、维护滞后性及工作量大等缺陷,避免异常检测方法存在的高误报率和人工响应延迟等缺点.本文应用 Agent 的智能与移动等特性,在抽象 DC 信息处理过程的基础上,首次给出了未成熟/完全成熟/半成熟 DC Agent 的定义及其之间的集合关系,分别描述了它们的数学分化模型和演化过程,证明了生命周期与 DC Agent 数量之间所存比例关系的相关定理.

复杂的 DC 分化机制至今仍存在大量的未知理论,但其与适应性免疫子系统的相互协调与制约关系正受到广大研究者的关注.因而,我们未来的工作将集中于群体 DC Agents 决策算法及其与适应性免疫细胞 Agents 的协同工作机制研究.

参考文献:

[1] K C Mcculiough, A Summerfield. Basic concepts of immune response and defense development[J]. ILAR, 2005, 46(3): 230 - 240.

[2] J Kim, P J Bentley, U Aickelin et al. Immune system approaches to intrusion detection　a review[J]. Natural Computing, 2007, 6(4): 413 - 466.

[3] U Aickelin, P Bentley, et al. Danger theory: the link between AIS and IDS[A]. Proc of the 2nd International Conference on Artificial Immune Systems [C]. Edinburgh, U. K.: LNCS, 2003. 147 - 155.

[4] J Twycross, U Aickelin. Towards a conceptual framework for innate immunity[A]. Proc of the 4th International Conference on Artificial Immune Systems [C]. Berlin/Heidelberg: Springer, 2005. 112 - 125.

[5] A Somayaji, S Forrest. Automated response using system-call delays[A]. Proc of the 9th USENIX Security Symposium[C]. Berkeley, USA: USENIX Association, 2000. 185 - 198.

[6] P Matzinger. The danger model: a renewed sense of self[J]. Science, 2002, 296(5566): 301 - 305.

[7] B D Brown, D Lillicrap. Dangerous liaisons: the role of "danger" signals in the immune response to gene therapy[J]. Blood, 2002, 100(4): 1133 - 1140.

[8] J Banchereau, R M Steinman. Dendritic cells and the control of immunity[J]. Nature, 1998, 392(6673): 245 - 252.

[9] M B Lutz, G Schuler. Immature, semi-mature and fully mature dendritic cells: which signals induce tolerance or immunity? [J]. Trends in Immunology, 2002, 23(9): 445 - 449.

[10] R Medzhitov, C A Janeway. Decoding the patterns of self and nonself by the innate immune system[J]. Science, 2002, 296(5566): 298 - 300.

[11] J Greensmith, U Aickelin et al. Articulation and clarification of the dendritic cell algorithm[A]. Proc of the 5th International Conference on Artificial Immune Systems[C]. Oeiras, Portugal: LNCS, 2006. 404 - 417.

[12] J Greensmith, U Aickelin. Dendritic cells for SYN scan detection[A]. Proc of the 9th Annual Conference on Genetic and Evolutionary Computation[C]. NY, USA: ACM Press, 2007. 49 - 56.

[13] J Twycross, U Aickelin. Libtissue-implementing innate immunity[A]. Proc of the IEEE World Congress on Computational Intelligence[C]. Vancouver, Canada: IEEE, 2006. 499 - 506.

[14] 王益丰,李涛,胡晓勤,等.一种基于人工免疫的网络安全实时风险检测方法[J].电子学报,2005,33(5):945 - 949.
Wang Yi-feng, Li Tao, Hu Xiao-qin, et al. A real-time method of risk evaluation based on artificial immune system for network security[J]. Acta Electronica Sinica, 2005, 33(5): 945 - 949. (in Chinese)

作者简介:

倪建成　男,1971 年生于山东曲阜.四川大学计算机科学学院博士研究生.主要研究方向为网格计算、人工免疫.
E-mail: nijch@163.com

李志蜀　男,1946 年生于重庆.四川大学计算机科学学院教授、博士生导师.主要研究方向为计算机网络、智能控制与多媒体技术等.
E-mail: lzshu5@yahoo.com.cn