基于危险理论的自动入侵响应系统模型

彭凌西[1,2]，谢冬青[1]，付颖芳[3]，熊伟[1]，沈玉利[4]

（1. 广州大学 计算机科学与教育软件学院，广东 广州 510006；2. 网络与数据安全四川省重点实验室，四川 成都 611731；
3. 北京工业大学 计算机学院，北京 100124；4. 仲恺农业工程学院 计算机科学与工程学院，广东 广州 510225）

摘 要：提出了一种基于危险理论的自动入侵响应系统模型(AIRSDT)，对网络活动中自体、非自体、免疫细胞、记忆检测器、成熟检测器和未成熟检测器进行了形式化描述，建立了主机和网络实时危险定量计算方程，并根据主机和网络当前所面临攻击的各类攻击和总体网络危险强度，自动调整入侵响应策略。理论分析和实验结果充分表明，模型有助于解决自动入侵响应研究中难以判断真正“危险”的入侵或者攻击行为的问题，降低入侵响应次数和响应综合代价。

关键词：危险理论；自动入侵响应系统；网络实时危险评估；人工免疫

中图分类号：TP393.08 **文献标识码：**A **文章编号：**1000-436X(2012)01-0136-09

Automated intrusion response system model based on danger theory

PENG Ling-xi[1,2], XIE Dong-qing[1], FU Ying-fang[3], XIONG Wei[1], SHEN Yu-li[4]

(1. Department of Computer and Education Software, Guangzhou Univ., Guangzhou 510006, China;
2. Network and Data Security Key Laboratory of Sichuan Province, Chengdu 611731, China;
3. College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China;
4. College of Computer Science and Engineering, Zhongkai University of Agriculture, Guangzhou 510225, China)

Abstract: A novel automated response system model based on the danger theory (AIRSDT) was given. With the descriptions of self, non-self, immunocyte, memory detector, mature detector and immature detector of the network transactions, network danger evaluation equations for host and network were built up. Then, the automated response actions were taken or adjusted according to the danger of each network attack, including holistic risk degrees of the host and network. Both the theory analysis and experimental results prove that AIRSDT not only helps to solve the problem that the current automated response models could not detect the ‘true’ intrusion or attack action, but also greatly reduces the response times and response cost.

Key words: danger theory; automated intrusion response system; real-time network risk evaluation; artificial immune

1 引言

入侵响应系统的作用是在入侵检测系统 (IDS, intrusion detection system)对网络活动进行监视，发现各种攻击企图、攻击行为或者攻击结果后，进行适当的入侵响应[1]。入侵响应技术是基于主动防御思想的新一代的网络安全技术，在保护网络安全上具有十分突出的意义，是目前网络安全技术发展的重要方向。

现有对自动入侵响应的研究资料或文献相对

收稿日期：2010-08-15；**修回日期：**2010-11-30

基金项目：国家自然科学基金资助项目（61100150）；广东省自然科学基金资助项目（S2011040004528, 10451009101004574）

Foundation Items: The National Natural Science Foundation of China (61100150); The Natural Science Foundation of Guangdong Province (S2011040004528, 10451009101004574)

较少，在这些研究中，Fisch 最早对入侵响应分类进行了研究，提出了以入侵被检测到的时机以及响应目标进行分类，该方法的缺点是考虑的因素不够全面[2]。Curtis 在此基础上提出了更为全面的入侵响应方式分类方法，该方法综合考虑了攻击的进度、攻击的类型、代价作为响应决策的依据[3]。Wenke L 提出了一种基于响应代价作为响应决策的依据，从而出现了基于成本敏感的响应决策模型，但方法中响应代价的量化方法过于粗糙[4]。Wu Y S 等设计并实现了一种名为 ADEPTS 的自动响应系统模型，利用入侵有向图对系统传播风险进行建模，并根据具体的攻击效果来进行适当的响应[5]。Mu C P 等提出了一种基于层次任务网络计划的入侵响应决策模型，包括响应决策制定过程以及决策响应时间[6]。Cuppens 等提出了一种基于上下文和本体方法入侵响应方法，并定义了一套规则来执行这种基于策略的入侵响应[7]。张剑等提出了一种可回卷的自动入侵响应系统，该方法根据 IDS 检测到的攻击行为进行响应，待判定攻击会话结束后进行响应回卷[8]。穆成坡等提出了一种基于模糊认知图的自动入侵响应决策推理机制[9]。吴姚睿等提出了一种通过关系图建立攻击群模型的方法，通过判断攻击序列重构协同入侵行为的攻击过程，并对攻击行为做出响应，从而达到最大程度地减少响应成本的目的[10]。

从总体上分析，当前自动入侵响应研究主要存在 2 个问题，首先是由于目前 IDS 普遍存在的高误报率和漏报率，造成自动入侵响应系统难以判断真正“危险”的入侵或者攻击行为，因而影响了入侵响应决策的判断。另外，目前国内外提出的一些入侵响应模型或方法，大多缺乏严格、全面、定量的数学模型，进行动态响应策略调整，因而在具体应用中存在很大的局限性。

计算机安全问题与生物免疫系统(BIS, biological immune system)所遇到的问题具有惊人的相似性，两者都要在不断变化的环境中维持系统的稳定性。基于人工免疫(AIS, artificial immune system)的网络安全技术具有多样性、自适应、顽健性等特点，被认为是一条非常重要且有意义的研究方向[11,12]。目前，在人工免疫应用于网络安全技术的研究中，Hofmeyr 和 Forrest 等基于“自体—非自体”理论[13]，提出了一种通用的人工免疫系统实现架构，并设计了一个计算机免疫系统(CIS, computer immune system)LISYS[14]，LISYS 对后来 CIS 的研究产生了深远的影响，现今基本上所有的 CIS 模型都是基于 LISYS 的结构。著名免疫学家 Matzinger 提出了著名的危险理论[15]，危险理论指出，免疫系统中不只是进行“自体—非自体”的识别，而是对受侵害组织的“危险”发出危险信号。由于现有计算机免疫系统模型存在的因自体空间巨大而效率低下的问题[16]，因此在相关的网络安全技术研究中，有必要进一步引入危险理论，将识别原理从对“自体—非自体”的识别转变为对“危险”的识别。

为使网络信息系统能对真正具有“危险”的入侵攻击进行响应，降低入侵响应系统响应代价和响应次数，基于危险理论，本文在建立网络安全环境下人工免疫系统的自体、非自体、免疫细胞等集合定义后，提出了一种新的基于网络实时危险的自动入侵响应系统模型(AIRSDT, automated intrusion response system model based on danger theory)，该模型首先对网络入侵或攻击进行实时危险评估，实时定量计算各类攻击和总体网络危险强度，然后据此自动调整入侵响应策略，解决了目前自动入侵响应研究中难以判断真正“危险”的入侵或者攻击行为的问题，从而降低了入侵响应次数和响应综合代价，理论分析和实验结果均表明，AIRSDT 为自动入侵响应系统研究提供了一种有效的新方法。

2　模型理论

定义论域 $D=\{0, 1\}^l$，l 为正整数，抗原集合 $Ag \subset D$，自体集合 $self \subset Ag$，非自体集合 $Nonself \subset Ag$，有 $Self \cup Nonself=Ag$，$Self \cap Nonself=\varnothing$，其中 Ag 表示通过从网络 IP 数据分组中提取的 IP 地址，端口号或协议类型等网络事务特征的二进制表示，$Self$ 集为正常网络服务事务，$Nonself$ 集为来自网络的攻击。

用四元组$<d, age, count, s>$来描述免疫检测器集合 B，其中 d 为抗体基因，$d \in D$，age 为抗体年龄，$age \in N$，$count$ 为匹配数，$count \in N$，s 为模拟“危险理论”，检测器受到攻击后的“危险性”，$s \in R$，N 为自然数集合，R 为实数集合。免疫检测器由记忆检测器和成熟检测器组成，即 $B=M_b \cup T_b$，定义匹配 $Match=\{<x, y>|x, y \in D, f_{match}(x, y)=1\}$ 为 D 中的匹配关系，$f_{match}(x, y)$ 为 x 和 y 之间的匹配，可采用海明距离、欧式距离、r-连续比特(r-contiguous bites)匹配函数[13]等计算方式。记忆检测器集 M_b 为

不与自体匹配且与抗原匹配数不小于 ß 的免疫检测器组成。定义成熟免疫检测器集合 T_b，由不与自体匹配且与抗原匹配数未超过匹配数阈值 ß 的免疫检测器组成，ß ∈N。类似地，用二元组<d, age>定义未成熟免疫检测器集合 I_b，I_b 由抗体基因库产生以及随机生成。I_b 在通过自体耐受，通过 a 个耐受周期并删除那些识别自体抗原的未成熟细胞后，成为成熟免疫细胞，a ∈N。

AIRSDT 首先对入侵或攻击行为进行检测并评估，实时定量计算出主机和网络所面临的总体危险和各类攻击危险，然后根据具体危险数值对网络入侵或攻击行为进行入侵自动响应或进行响应回滚。

2.1 入侵实时危险评估

对一个输入的抗原集合 Ag，分 d 代（d 为正数）进行训练，每代选出一定数量的抗原组成 sAg 抗原集合，通过 B 集合的检测把它分类为自体和非自体。具体过程分为以下 3 个阶段。第一阶段为 0 时刻到一个耐受期 a 结束的时刻，需要定义初始的自体集合 $Self(0)$和未成熟细胞集合 $I_b(0)$，后者经前者耐受后成为成熟细胞。第二阶段从 a+1 时刻到记忆细胞产生的时刻，为自学习阶段，成熟细胞通过克隆选择产生能识别大量不同非自体抗原的记忆细胞，而通过检测被分类为自体的抗原最后送给未成熟细胞进行耐受。第三阶段从记忆细胞产生到系统终止，免疫系统各部件产生完毕，进行实际环境中的检测：记忆细胞检测，成熟细胞对剩下抗原的检测，最后未成熟细胞以剩余抗原为自体进行耐受[12]。

每个记忆细胞对应检测某种攻击，主机或网络中的记忆细胞对检测到的入侵或者攻击抗原进行实时危险评估。从时间 t-1 到 t 时刻的单位时间内，对每个记忆细胞 x，如果与某个抗原 g 相匹配，即满足 $f_{match}(x.d, g.d)=1$，则认为记忆细胞检测到一个入侵或攻击抗原，其危险值 s 按式(1)增加。如果检测到多个同样抗原，则按式(1)对其危险数值将进行累计计算，表明该类攻击威胁在持续增加，其中 $?_1$ (>0 的常数)为初始的危险数值，$?_2$ (>0 的常数)模拟奖励因子。

$$x.s(t) = h_1 + h_2 x.s(t-1) \tag{1}$$

反之，如果该记忆检测器在该时间间隔内没有检测到入侵或攻击抗原，则其危险性按式(2)进行计算。

$$x.s(t) = x.s(t-1)e^{-1} \tag{2}$$

如果连续 $?$($?$∈N)个时间间隔均未检测到攻击抗原，不难得出 $x.s(t)= x_0s(t_0)e^{-?}$，$?$ 越大，危险值越小，表明危险在衰减，当 $?$→8 时，$x.s(t)$→0，该类威胁将被清除，攻击警报将解除。

对于记忆细胞 x 和 y，如果存在 $f_{match}(x.d, y.d)=1$，即记忆细胞 x 和 y 满足一定的相似性，则认为这 2 种记忆细胞所检测到的攻击为同一类攻击。

设危险指标 $0 \le r_k(t) \le 1$ 为主机 k 在 t 时刻所面临的危险：$r_k(t)=1$，表明当前系统极度危险；$r_k(t)=0$，表明当前系统没有危险。$r_k(t)$值越大，表明当前系统面临的危险越高。考虑到各种主机的资产权重以及各类攻击的危险性不一样，设定 μ_i 表示第 i 类攻击的危险性，对于主机 k，t 时刻面临的第 i ($1 \le i \le I$) 类网络攻击的网络安全危险 $r_{k,i}$ 由式(3)进行计算：

$$r_{k,i}(t) = 1 - \frac{1}{1 + Ln\left(m_i \sum_{x \in A_{k,i}(t)} x.s + 1\right)} \tag{3}$$

对于某个主机 k，t 时刻整个主机的整体网络危险 $r_k(t)$通过式(4)进行计算：

$$r_k(t) = 1 - \frac{1}{1 + Ln\left(\sum_{i=1}^{I} m_i \sum_{x \in A_{k,i}(t)} xs(t) + 1\right)} \tag{4}$$

对整体网络，其面临的第 i 类攻击的整体网络危险 $R_i(t)$，由所包含主机 k($1 \le k \le K$)的分类危险及其资产权重 $?_k$ 以及所包含的子网 n($1 \le n \le N$)的分类子网危险 $R_{n,i}$ 和资产权重 ξ_n 按式(5)进行加权计算，而其中的子网还可以保护下一级子网，依此类推。

$$R_i(t) = 1 - \frac{1}{1 + Ln\left(\sum_{k=1}^{K} v_k r_{k,i} + \sum_{n=1}^{N} x_n R_{n,i}(t) + 1\right)} \tag{5}$$

类似地，整个网络系统的整体安全危险 $R(t)$，由主机的整体危险 r_k、资产权重 $?_k$ 以及所包含子网的整体网络危险 R_n($1 \le n \le N$)和资产权重 ξ_n 进行计算，如式(6)所示。

$$R(t) = 1 - \frac{1}{1 + Ln\left(\sum_{k=1}^{K} v_k r_k(t) + \sum_{n=1}^{N} x_n R_n(t) + 1\right)} \tag{6}$$

2.2 自动入侵响应与回滚

自动入侵响应取决于 2 个条件，即网络实时危

险与攻击强度。当网络实时危险低于给定阈值时，检测到的一些无关报警信息和虚假警报被忽略，不进行响应，对于检测的一些攻击信息，也通过网络实时危险评估关联起来，根据具体的网络实时危险进行响应；而攻击强度条件大于 0，则是为了避免该时间段没有检测到攻击，而由于前一时间段遭受攻击后，网络实时危险数值较高，但现在攻击已经停止，这种情形则不需要响应。

式(7)用于描述主机中入侵响应的产生，主要来自 2 个方面：对主机 k，当主机的整体危险 $r_k(t)$大于 q_k $(0<q_k<1)$，并且主机遭遇所有的攻击（假设主机中包含了 i 类攻击）的攻击强度大于 M_k；当主机遭遇的第 i 类攻击的网络实时危险 $r_{k,i}(t)$大于 $d_{k,i}(0<d_{k,i}<1)$，且主机遭遇的该类攻击的攻击强度大于 $N_{k,i}$。

$$Response(t)=\begin{cases}1, \ \text{if } r_k(t)>q_k \wedge \\ \sum_{i=1}^{I}\sum_{x\in A_{k,i}}(x.count(t)-x.count(t-\Delta t))>M_k \\ \vee\, r_{k,i}(t)>d_{k,i} \wedge \\ \sum_{x\in A_{k,i}}(x.count(t)-x.count(t-\Delta t))>N_{k,i} \\ 0, \text{其他}\end{cases} \tag{7}$$

这 2 个方面所对应的具体意义是，如果主机的整体危险 $r_k(t)$及被攻击强度分别超过了给定阈值，表明主机遭遇所有攻击的危险程度已经影响了主机的安全运行；而主机遭遇的某类攻击的网络危险 $r_{k,i}(t)$且被攻击强度分别超过了给定阈值，表明主机检测到同类及变种网络攻击已经具有“危险”。

$$Response'(t)=\begin{cases}1, \ \text{if } R(t)>?' \wedge \\ \sum_{i=1}^{I}\sum_{x\in A_i}(x.count(t)-x.count(t-?t))>M' \\ \vee\, R_i(t)>d_i' \wedge \\ \sum_{x\in A_i}(x.count(t)-x.count(t-?t))>N_i' \\ 0, \text{其他}\end{cases} \tag{8}$$

对于整个网络，响应动作来自 2 个方面：当网络的整体危险 $R(t)$大于 $?'(0<?'<1)$，并且网络中遭遇所有的攻击（假设网络中包含了 i 类攻击）的攻击强度大于 M；当网络遭遇的某类攻击的网络危险 $R_i(t)$大于 $d_i'(0<d_i'<1)$，并且网络中遭遇的该类攻击（第 i 类攻击）的攻击强度大于 N_i，具体定义如式(8)所示。

类似地，这 2 个方面所对应的具体意义是，如果网络的整体危险 $R(t)$且被攻击强度分别超过了给定阈值，表明网络遭遇所有攻击的危险程度已经影响了网络的安全运行；而网络的遭遇的某类攻击的网络危险 $R_i(t)$及被攻击强度分别超过了给定阈值，表明网络检测到同类及变种网络攻击已经具有“危险”。

在 AIRSDT 模型中，对于主机，只有在主机遭遇的整体网络危险和某类攻击危险，及被攻击强度分别大于给定的阈值的时候才进行入侵响应；而对于整个网络，只有在网络遭遇的整体网络危险和某类网络攻击危险，以及遭遇攻击的攻击强度大于给定的阈值的时候才进行入侵响应。由于模型通过计算网络、主机以及分别的分类网络攻击的总体网络危险和攻击强度来判断是否进行自动响应，因此模型能较好地解决 Wenke L 等提出的成本敏感模型[4]中难以应对协同攻击的问题。

Curtis 对所有响应方式进行了概括，并将响应方式分为基于主机的和基于网络的响应方式[3]，Curtis 列出的 11 种基于网络的响应方式，结合其他文献提出来的 6 种响应方式，另外本文提出 6 种响应方式，较全面的自动入侵响应方式如表 1 所示。

在全部响应方式中，1~4，12~13，17 的响应措施比较温和，8~11，14，16，20 的响应措施比较严厉，其余的响应措施介于两者之间。其中 1~4，12~13，21~22 的响应措施属于被动方式，其他属于主动方式。8~11，14~15 的响应方式一般受到法律等因素的约束，而 5~7，14~16，20 的方式能有效地阻断攻击。

AIRSDT 根据网络或主机所面临的定量总体网络危险进行自动响应，网络危险程度越高，响应策略就越严厉，反之响应策略相对温和。在本文实验中，根据具体网络危险所采取的响应策略如表 2 所示，从表中同时可看出，对所有的攻击，AIRSDT 模型均采取了记录该安全事件的响应方式，另外在网络危险高于 0.1 时，AIRSDT 均产生报警信息。在实际过程中，可根据网络安全需要来设定具体的响应策略。

表 1 基于网络的入侵响应方式

编号	响应方式	具体解释
1[3]	记录安全事件	对入侵事件记录日志
2[3]	产生报警信息	向控制台告警，或通过 E-mail、短信等通知管理员
3[3]	记录附加日志	怀疑阶段，记录附加日志以帮助收集信息
4[3]	激活附加的入侵检测工具	用占用资源较多的检测工具
5[3]	隔离入侵者 IP	阻断入侵者 IP 的数据分组
6[3]	禁止被攻击对象的特定服务	禁止被攻击对象的端口或服务
7[3]	隔离被攻击对象	关闭已受攻击的特定端口或服务
8[3]	警告攻击者	向入侵者发警告威慑信息
9[3]	跟踪攻击者(攻击源回溯)	追踪入侵者是采用追踪技术定位攻击者的位置
10[3]	断开危险连接	发送 TCP 的 RESET 分组
11[3]	攻击攻击者	对攻击者进行攻击
12[17]	蜜罐技术	用来观测黑客如何探测并最终入侵系统的技术
13[18]	蜜网技术	实质是一类研究型的高交互蜜罐技术
14[17]	动态修改防火墙策略	与防火墙联动，调整防火墙的策略
15[17]	攻击源吸收和转移技术	攻击源吸收和转移技术能在秒级时间将攻击分组吸收到诱骗系统，保护主机服务，同时不切断与攻击者的连接
16[17]	黑名单	将攻击者名称加入黑名单，禁止访问
17	取证技术	对入侵行为进行动态取证
18	服务切换(服务漂移)	将服务切换到另外一个服务器上
19	热备份	不停止服务对数据备份
20	冷备份	暂停服务并对数据备份
21	全恢复	通过备份数据恢复全部数据
22	差异恢复	只恢复变化了的数据

表 2 网络危险及建议采取的响应策略

危险数值	响应策略
0~0.1	1
0.1~0.2	1,2
0.2~0.3	1,2,3
0.3~0.4	1,2,4,12,13,17
0.4~0.5	1,2,5,16,17,18
0.5~0.6	1,2,6,7,14,17,19
0.6~0.7	1,2,8,15,17,20
0.7~0.9	1,2,9,10,17,22
0.9~1.0	1,2,11,17,21

自动响应模型根据网络危险情况作出响应策略，并采用消息机制的方式，将具体的自动响应策略发送给执行响应策略的主机或者网络，其中发送的响应消息 *Message* 为一个 5 元组，如式(9)所示。对上述 22 种响应方式，对应操作 *Action* 和相应的回滚操作 *Action* 如表 3 所示，其中的字符“*F*”表示该操作为空操作。

$$Message: =< Sender >< Receiver > < SendTime >< ValidTim >< Action > \quad (9)$$

在 AIRSDT 模型执行自动入侵响应策略后，如果网络或者主机的网络危险呈现上升的趋势，表明网络或者主机所遭遇网络危险越来越大，AIRSDT 将采取更加严厉的响应措施，避免信息系统进入更加“危险”的状态，从而确保信息系统运行安全。

表 3 响应操作及相应回滚操作

编号	响应操作	意义	回滚操作	意义
1	*Log*	记录安全事件	*F*	空操作
2	*Alert*	产生报警信息	*F*	空操作
3	*Enable*	激活附加的入侵检测工具	*Disable*	失效附加的入侵检测工具
4	*Add*	记录附加日志	*F*	空操作
5	*Lock*	隔离入侵者 IP	*Unlock*	将入侵者 IP 释放
6	*Stop*	禁止被攻击对象的特定服务	*Start*	重启服务
7	*Shutdown*	隔离被攻击对象	*Restart*	重新启动被攻击对象
8	*Warn*	警告攻击者	*F*	空操作
9	*Track*	跟踪攻击者	*F*	空操作
10	*Reset*	断开危险连接	*F*	空操作
11	*Attack*	攻击攻击者	*F*	停止攻击攻击者
12	*Honeypot*	空操作	*F*	空操作
13	*Honeynet*	空操作	*F*	空操作
14	*Link*	与防火墙联动	*UnLink*	撤销防火墙策略
15	*Absorb*	攻击吸收	*Release*	停止吸收
16	*Blacklist*	黑名单	*UnBlacklist*	从黑名单去掉
17	*Forensics*	动态取证	*F*	空操作
18	*Migrate*	服务漂移(切换)	*MigrateBack*	服务漂回
19	*HotBackup*	热备份	*StopHot-Backup*	停止备份
20	*ColdBackup*	停止服务冷备份	*StopCold-Backup*	恢复服务
21	*AllRecovery*	全恢复	*F*	空操作
22	*DiffRecovery*	快速恢复	*F*	空操作

对于主机，式(10)对主机进行再次响应的过程进行了描述，其中 *Response* (*t*+? *t*)表示主机在 *t*+? *t* 时刻再次进行响应，表明由于主机遭遇到更加严重的攻击，系统将采取更加严厉的响应措施。

$$Response(t+?\,t)=\begin{cases}1, & \text{if } r_k(t+\Delta t)>r_k(t)\vee \\ & \quad r_{k,i}(t+\Delta t)>r_{k,i}(t) \\ 0, & \text{其他}\end{cases} \quad (10)$$

类似地，对于网络，式(11)对网络进行再次响应的过程进行了描述，其中 $Response'(t+?\,t)$表示网络在 $t+?\,t$ 时刻再次进行响应，表明由于网络遭遇到更加严重的攻击，系统将采取更加严厉的响应措施。

$$Response'(t+?\,t)=\begin{cases}1, & \text{if } R(t+?\,t)>R(t)\vee \\ & \quad R_i(t+?\,t)>R_i(t) \\ 0, & \text{其他}\end{cases} \quad (11)$$

当网络危险低于给定的危险阈值后，AIRSDT 则自动撤销所采取的入侵响应策略，以免影响正常的网络服务。

对于主机，式(12)对主机进行自动响应回滚的过程进行了描述，其中 $Rollback(t+?\,t')$表示主机在 $t+?\,t'$时刻进行响应回滚[8]，μ_k 表示主机所面临整体危险进行自动响应回滚的阈值，而 $?_{k,i}$ 表示主机对第 i 类攻击危险进行自动响应回滚的阈值。

$$Rollback(t+?\,t')=\begin{cases}1, & \text{if } r_k(t)\le \mu_k\wedge \\ & \quad r_{k,i}(t)\le ?_{k,i} \\ 0, & \text{其他}\end{cases} \quad (12)$$

对于网络，式(13)对网络进行自动响应回滚的过程进行了描述，其中 $Rollback'(t+?\,t')$表示网络在 $t+?\,t'$时刻进行响应回滚，μ'表示网络或者子网所面临整体危险进行自动响应回滚的阈值，而 $?_i'$ 表示网络对第 i 类攻击危险进行自动响应回滚的阈值。

$$Rollback'(t+?\,t')=\begin{cases}1, & \text{if } R(t)\le \mu'\wedge \\ & \quad R_i(t)\le ?_i' \\ 0, & \text{其他}\end{cases} \quad (13)$$

对于自动响应回滚策略，同样采用消息机制的方式，如式(9)所示，将要执行的自动响应回滚策略发送给具体执行的主机或者网络。

由于 AIRSDT 模型通过计算出实时网络危险和被攻击判断是否进行响应，因此具有响应速度较快的优点，对拒绝式服务具有一定的抵抗能力；另外，可以与防火墙很便捷实现联动，具有协同性较强的优点；最后，在网络实时危险低时，可以自动进行响应回滚，因此具有资源利用率高等特点。

3　仿真实验

实验在广州大学信息安全研究所的网络安全实验室进行，实验环境为 100M 的局域网，其中有计算机 20 台，通过一个 C 类 IP 地址 202.192.87.* 连接到 Internet，服务器的操作系统为 Red Hat 9.0，服务器提供 WWW，E-mail 以及 FTP 服务，抗原定义为定长为从 IP 分组中提取的包含 IP 地址、端口号、协议类型、数据分组内容等网络事务特征的 128 位二进制字符串。

3.1　自动入侵响应实验

为验证入侵响应子模型的有效性，对网络进行了模拟的 smurf 攻击实验，其中网络整体危险响应阈值和分类攻击响应阈值 $?'$和 d_i'均分别设定为 0.4。图 1 和图 2 分别给出了正常服务情况、通过 AIRSDT 进行入侵响应以及没有响应情况下 3 个状态下的网络流量和 CPU 利用率的对比情况。

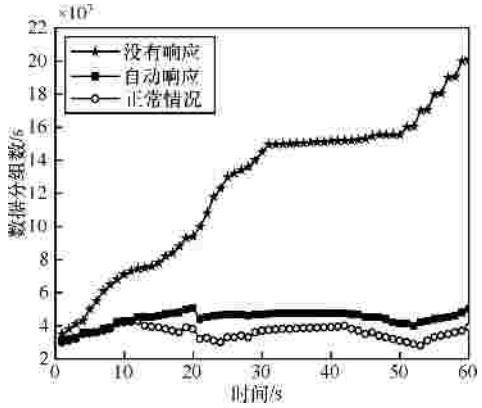

图 1　3 个状态下的网络流量对比

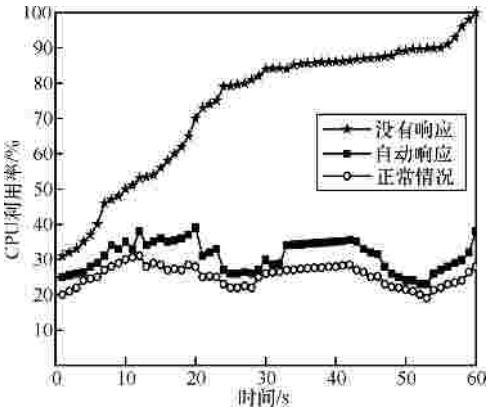

图 2　3 个状态下的 CPU 利用率对比

从图 1 可看出在，从 0s 到 12s 之间，在没有进行入侵响应情况下，受网络攻击的影响，网络流量持续增加，相应图 2 中 CPU 利用率也随之增高，

而从 12s 到 20s 之间，随着攻击强度的继续增加，网络流量持续增高，相应图 2 中 CPU 利用率也持续增高。而在进行入侵响应的情况下，网络流量和 CPU 利用率虽然都有所增加，但都比较稳定。

对照图 1 和图 2 看出，对 AIRSDT 进行自动响应与不进行响应情况进行比较，进行自动响应后的网络流量要比不进行响应流量小很多，CPU 利用率要低出很多，但均比正常情况下略大，这是因为进行响应后，发送响应消息增加了网络通信量，而进行响应处理也略微提高了 CPU 利用率。

实验结果显示，AIRSDT 通过自动入侵响应，有效减小了网络流量，并降低了 CPU 利用率。

3.2 同类响应模型对比实验

为证明 AIRSDT 有效降低了入侵响应代价，将 AIRSDT 与 RARS 模型[8]在入侵响应次数，以及响应代价等进行综合比较，RARS 模型与自动入侵响应模型相比，已较好地降低了响应代价。

安排网络安全专业人员对网络进行一天的模拟攻击，图 3 给出了检测到的网络攻击强度与 AIRSDT 的网络危险曲线，其中的危险曲线是传统 IDS 和已有入侵响应系统均不具有的功能，而 A、B、C 3 个点为系统的自动入侵响应点。对图 3 进行分析，从 8:30~9:30 之间，系统遭遇了 portscan 攻击，尽管攻击强度很大，但整体危险数值并不高，只有 0.098，但检测到的 portscan 攻击危险数值为 0.825，高于 0.4，因此系统在 A 点进行了入侵响应。在 11:00~11:30 之间，系统遭遇了 sshtrojan 攻击，尽管攻击强度比较小，但由于该攻击运行攻击者登录到受害主机，系统危险数值迅速增大到 0.488，故系统在 B 点进行了入侵响应。而在 16:30~17:00 之间，系统遭遇了 appache2 和 smurf 攻击，这些攻击属于资源耗尽型攻击，系统实时危险增到 0.7，因此系统在 C 点进行了入侵响应。

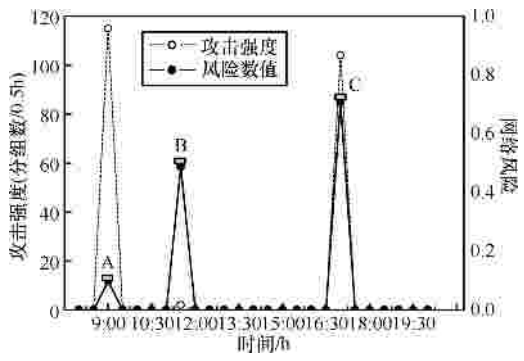

图 3 网络攻击强度与危险曲线

入侵响应系统中的代价类型包括响应撤销代价 *RRCost*、操作代价 *OCost*、响应代价 *RCost*、损失代价为 *Dcost*，设入侵响应系统 *R* 的期望综合代价为 *TCost*(*R*)，*TCost*(*R*)为所有代价的总和，入侵事件类型集合为 *E*，则得到：

$$TCost(R)=\sum_{e\in E} RRCost(e)+OCost(e)+RCost(e)+DCost(e) \tag{14}$$

各类响应代价的量化是一个难点，本实验中结合了 Wenke L 提出[4]的方法，另外将 AIRSDT 模型与 RARS 进行响应次数 *n*，以及各类响应代价的比较，得到的结果如表 4 所示。

表 4 入侵响应次数及响应代价比较

模型	*n*	*OCost*	*RRCost*	*RCost*	*DCost*	*TCost*
AIRSDT	3	100	100	40	85	325
RARS	221	100	22 100	2 175	3 450	27 825

表 4 的实验结果表明，AIRSDT 与 RARS 模型进行比较，由于 AIRSDT 能够实时准确计算出系统面临的“真正”危险，并进行自动响应，这样不仅降低入侵响应的次数，提高了入侵响应的质量，而且大大降低了入侵响应的综合代价。

4 同类研究对比

在自动入侵响应研究中，如何判断整个网络系统所面临的总体危险和各类攻击的危险情况，判断真正具有“危险”行为的网络入侵或攻击行为，进行自动入侵响应一直是研究中一个比较难以解决的问题，本文在网络危险评估实时定量计算研究基础上，提出了一种根据“危险”的自动入侵响应系统模型 AIRSDT，表 5 给出了 AIRSDT 模型与引言中一些模型进行的比较。

从表 5 可看出，由于当前已有的入侵响应模型无法对系统面临的整体危险进行计算，所指定的响应策略只能被预先静态指定，而 AIRSDT 结合了对系统整体网络危险进行实时定量计算方法，从而能根据“危险”情况进行响应决策，自动调整入侵响应策略进行响应回滚策略，这样不但能够降低响应次数，而且达到了降低响应代价的目的，因此，AIRSDT 模型具有良好的自适应性。

基于网络危险的自动入侵响应技术属于积极主动的网络安全防御技术，当系统面临较高的网络危险时候，系统可以采用比较严厉的响应策略，例

表 5　　自动入侵响应同类研究对比

模型	实时危险评估能力	自动响应能力说明
Fisch[2], Curtis[4]	不具备	属于基于分类的响应决策模型，缺乏对未知攻击分类能力
Wenke[3]	不具备	基于代价评估和优化入侵检测和响应系统的思想，但量化方法过于粗糙，难以应对协同攻击
WuYS[5], MU[6], Cuppens[7], MU[9], Wu[10]	不具备	分别基于入侵有向图、层次任务网络计划、攻击上下文和本体、模糊认知图、攻击群关系图等被系统攻击情况判定入侵者攻击意图，然后对入侵行为进行响应，对未知攻击无法判断
RARS[8]	不具备	判定攻击是否属于同一入侵会话，然后进行自动响应，响应后在一定时间后进行回卷，响应代价有一定程度减小
AIRSDT	具备	能够实时定量准确计算网络整体危险和各类攻击危险，并根据具体危险数值调整入侵响应策略进行响应回卷，响应代价和次数得到减小，具有良好自适应性，能应对协同攻击，响应速度快，协同性强，资源利用率高

如断开入侵者连接或者攻击入侵者，与已有自动入侵响应技术相比较，由于 AIRSDT 能准确实时定量计算系统所面临的危险并自动调整响应策略，所以 AIRSDT 是对已有自动入侵响应技术的突破，对克服传统网络安全信息系统的技术缺陷，具有一定的理论意义和实际应用价值。

5　结束语

本文首次将生物免疫系统中的危险理论应用到自动入侵响应系统的研究中，提出了一种新的自动入侵响应系统模型 AIRSDT，AIRSDT 根据网络危险程度自动调整响应策略，使得网络信息系统只对具有“危险”的网络入侵或者攻击行为进行响应，这样既保证了系统的运行性能，又大大降低了自动入侵响应系统的综合代价，从而提高了系统运行的自适应性，但是，根据网络信息系统重要性程度，所采取的自动入侵响应策略有待进一步的完善。

参考文献：

[1] 段雪涛，贾春福，刘春波. 基于层次隐马尔科夫模型和变长语义模式的入侵检测方法[J]. 通信学报，2010, 31(3): 109-114.
DUAN X T, JIA C F, LIU C B. Intrusion detection method based on hierarchical hidden Markov model and variable-length semantic pattern[J]. Journal on Communications, 2010, 31(3): 109-114.

[2] FISCH E A. Intrusion Damage Control and Assessment: a Taxonomy and Implementation of Automated Responses to Intrusive Behavior[D]. Texas A&M University, College Station, TX, 1996.

[3] CURTIS A C. A methodology for using intelligent agents to provide automated intrusion response[A]. Proceedings of the IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop[C]. New York, USA, 2000. 110-116.

[4] WENKE L, WEI F, MATTHEW M, *et al*. Toward cost-sensitive modeling for intrusion detection and response[J]. Journal of Computer Security, 2002, 10(1/2): 5-22.

[5] WU Y S, FOO B, MAO Y C, *et al*. Automated adaptive intrusion containment in systems of interacting services[J]. Computer Networks, 2007, 51(5): 1334-1360.

[6] MU C P, LI Y. An intrusion response decision-making model based on hierarchical task network planning[J]. Expert Systems Applications, 2010, 37(3): 2465-2472.

[7] CUPPENGS B N, CUPPENS F, AUTREL F, *et al*. An ontology-based approach to react to network attacks[J]. International Journal of Information and Computer Security, 2009, 3(3): 280-305.

[8] 张剑，龚俭.可回卷的自动人侵响应系统[J]. 电子学报,2004, 32(5): 769-771.
ZHANG J, GONG J. Rollbackable automated intrusion response system[J]. Acta Electronica Sinica, 2004, 32(5): 769-771.

[9] 穆成坡，黄厚宽，田盛丰. 基于模糊认知图的自动入侵响应决策推理机制[J]. 北京交通大学学报, 2005, 29(2): 12-16.
MU C P, HUANG H K, TIAN S F. Fuzzy cognitive maps for decision supporting automatic intrusion response mechanism[J]. Journal of Beijing Jiaotong University, 2005, 29(2): 12-16.

[10] 吴姚睿，刘淑芬. 基于攻击群模型的协同入侵的响应方法[J]. 电子学报, 2009, 37(11): 2416-2419.
WU Y R, LIU S F. A response method for cooperative intrusions based on the attack group model[J]. Acta Electronica Sinica, 2009, 37(11): 2416-2419.

[11] LI T. Dynamic detection for computer virus based on immune system[J]. Science in China, Series F: Information Science, 2008, 51(10):1475-1486.

[12] 李涛. 计算机免疫学[M]. 北京：电子工业出版社, 2004.
LI T. Computer Immunology[M]. Beijing: Publishing House of Electronic Industry, 2004.

[13] FORREST S, PERELSON A, ALLEN L, *et al*. Self-nonself discrimi-

nation in a computer[A]. IEEE Computer Society Symposium on Research in Security and Privacy, Proceedings[C]. Los Alamitos, USA, 1994. 202-212.

[14] HOFMEYR S, FORREST S. Architecture for an artificial immune system[J]. Evolutionary Computation, 2000, 8(4): 443-473.

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

[16] ZENG J Q, LIU X J, LI T, *et al*. A self-adaptive negative selection algorithm used for anomaly detection[J]. Progress in Natural Science, 2009, 19(2): 261-266.

[17] **张峰**. **基于策略树的网络安全主动防御模型研究**[D]. **成都**: **电子科技大学**, 2004.
ZHANG F. Policy Tree Based Proactive Defense Model for Network Security[D]. Chengdu: University of Electronic Science and Technology, 2004.

[18] **王璐**, **秦志光**. **业务蜜网技术与应用**[J]. **计算机应用**, 2004, 24(3): 43-45.
WANG L, QIN Z G. Technology and application of production honeynet[J]. Computer Applications, 2004, 24(3): 43-45.

作者简介：

彭凌西（1978-），男，湖南岳阳人，博士，广州大学副教授，主要研究方向为网络安全技术、人工免疫等。

谢冬青（1965-），男，湖南益阳人，广州大学教授、博士生导师，主要研究方向为算法分析与设计、信息安全等。

付颖芳（1976-），女，湖南洞口人，北京工业大学博士后、讲师，主要研究方向为无线 Mesh 网安全、可信计算等。

熊伟(1977-)，男，江西丰城人，博士，广州大学副教授，主要研究方向为分布式计算。

沈玉利（1955-），男，山东费县人，博士，仲恺农业工程学院教授，主要研究方向为模式识别与智能系统、无线传感器网络等。