基于免疫原理的网络入侵检测算法改进*)

鲁云平 宋 军 姚雪梅

(重庆交通大学计算机与信息学院 重庆 400074)

摘 要 分析了基于免疫原理的网络入侵检测否定选择算法的不足,并对其进行了改进。通过增加排除匹配检测元过程,产生互不匹配的检测元,以提高检测集的整体检测能力,增强否定选择算法实用价值。理论分析和实验结果表明,改进算法的网络入侵检测效率更高。

关键词 网络入侵检测,否定选择算法,高效检测元集

Improvement on Network Intrusion Detection Algorithm Based on Immunological Principle

LU Yun-ping SONG Jun YAQ Xue-mei

(Institute of Computer & Information, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract This paper analyzed the insufficiency of negative selection algorithm based on immunological principle used in network intrusion detection, and made some improvement on it. A process eliminating the matched detectors is added in the negative selection algorithm to generate the unmatched detectors and increase the detecting ability of the detetion sets. Theory analysis and experimentation result indicat that the efficiency of network intrusion detection is enhanced by the improved algorithm.

Keywords Network intrusion detection, Negative selection algorithm, Efficient detection sets

网络入侵检测技术保护网络中的计算机免受入侵事件的危害和威胁,使其在不断变化的网络环境中维持系统稳定性。1994年,Forrest率先将人工免疫思想引入计算机入侵检测领域,提出了基于免疫特异性的否定选择算法,用于计算机病毒的检测和主机异常检测[1-3]。Hofmeyr等人则将基于免疫原理的否定选择算法推广到广播型局域网的入侵检测中,建立了具有分布性、多样性、容错性、动态保护、自适应性和可扩展性等特点的入侵检测系统[4]。本文针对基于免疫原理的网络入侵检测否定选择算法存在的不足进行了分析和改进,以提高算法的网络入侵检测效率,并对改进算法的网络入侵检测能力进行了理论和实验分析。

1 否定选择算法及其不足

基于免疫原理的网络入侵检测通过检测元与非我模式的匹配实现对非我模式的识别与检测,其研究重点为用于产生有效检测元的否定选择算法和用于事件识别的字符匹配规则。

在免疫系统中,免疫细胞的产生是一个随机过程,新产生的免疫细胞必须经过一个否定选择的审查过程,以防止免疫细胞与自我蛋白进行结合而引起自免疫反应。否定选择算法由免疫细胞的产生和耐受化过程中抽象而来[1-7],要求在给定自我模式集的情况下,产生只能与非我模式进行匹配并能检测异常模式的检测元集,其详细描述如下:

(1)定义一个自我模式集 S,作为生成有效检测元的训练集。

(2)产生候选检测元。通过一个随机过程产生一个长度为 l 的位串作为候选检测元。

(3)产生有效检测元集 R。将候选检测元与自我集 S 中的模式进行匹配试验,若匹配,则丢弃,返回(2);否则,该候选检测元为有效检测元,进入 R 集合,返回(2)。

(4)重复(2)、(3),直至产生一定数量的有效检测元。

否定选择算法所产生有效的检测元集 R 中的每一个检测元都与自我集 S 的任何模式不匹配。

否定选择算法产生有效检测元的匹配试验采用 k 连续位匹配规则[3],其定义为:

设两个长度为 l 的位串 x 和 y。若 x 和 y 至少有 k 个连续位相同,则 x 和 y 匹配,记为 $Match(x,y)$;否则,x 和 y 不匹配,记为 $\neg Match(x,y)$。其中,k 为匹配阈值,$0\leqslant k\leqslant l$。

检测系统对网络入侵的检测能力表现为检测元对非我模式空间的覆盖范围,检测元对非我模式的覆盖空间越大,能检测的网络入侵就越多。理想情况下,检测系统的所有检测元应完全覆盖非我空间;实际应用中,由于系统资源的限制,通常根据检测率的实际需要产生一定数量的检测元。

由于 k 连续位匹配规则的固有性质,使得有效检测元集 R 中的检测元在 k 连续位匹配规则下存在相互匹配的可能:

设 $s=b_1b_2b_3b_4\cdots b_{k-1}b_kb_{k+1}b_{k+2}\cdots b_l\ \forall b_i\in\{0,1\}$, $\forall r_1, r_2\in R$, r_1, r_2 与 s 匹配的最大字符块为 w,且 w 在 s 中的位置相同。

根据否定选择算法,R 集中任何检测元 r 与 s 在 k-连续位匹配规则下均不匹配,即 $\neg Match(r_1,s)$ 和 $\neg Match(r_2,s)$。

由于 r_1, r_2 与 s 匹配的最大字符块为 w,且 w 在 s 中的位置相同,则有 $|w|\in[0,k-1]$。同时,模式串定义在符号系统 $\{0,1\}$,则有 $Match(r_1,r_2)$。

R 集中的有效检测元 r_1, r_2 的覆盖空间存在一个非空的交集,使得该数量的有效检测元集 R 其网络入侵检测能力无

*)基金项目:重庆市自然科学基金(CSTC2006BB2413)。鲁云平 讲师,硕士,主要研究网络安全、综合业务数字网;宋 军 副教授,博士,主要研究宽带网络技术、计算机网络安全;姚雪梅 讲师,硕士研究生,主要研究入侵检测技术、编译技术。

法达到最大化。

2 否定选择算法的改进

为使否定选择算法产生的有效检测元其覆盖空间互不相交，达到检测元数量固定情况下有效检测元集整体 R 空间覆盖范围最大化，笔者对否定选择算法进行了改进，在否定选择算法中增加了一个排除匹配检测元的过程，改进算法描述如下：

(1)定义一个自我模式集作为生成有效检测元的训练集，且定义有效检测元集 R 为空集。

(2)产生候选检测元，通过一个随机过程来产生一个长度为 l 的位串作为候选检测元。

(3)产生有效检测元集 R。将产生的候选检测元与自我集中的模式进行匹配试验，若匹配，则丢弃该候选串，返回(2)；否则该候选检测元就是一个有效的检测元 r，进入下一步(4)。

(4)判断 r 是否已在 R 集中，若 R 中存在 r，就丢弃 r，并返回(2)；否则这个 r 进入 R。

(5)重复(2)(3)(4)三步，直到产生一定数量的检测元为止。

改进算法的具体实现如下：

变量定义—R 表示有效检测元；S 表示自我集；l 表示模式长度；k 表示匹配阈值；c 表示计数器；r 表示随机产生的长为 l 的候选检测元；d_num 表示所需检测元的数量。

```
PROC negative_selection(S);
    初始化,定义 l 和 k,d_num;=0,R;=∅,c;=0;match;=false;
    WHILE |R|<d_num+1 DO
    【while_num=0;//循环变量
    随机产生 r;
    WHILE while_num<|S| DO
        【从 Self 中依次取出模式 s;
        match;=matching(r,s);[判断 s 与 r 是否匹配?]
        IF match=true THEN exit;
            while_num;=while_num+1;】;
      IF match=false THEN
      【FOR i;=1 TO |R| DO
        【从 R 中依次取出检测元 detector;
        match;=matching(r,detector);[判断 r 与 detector 是否匹
配?]
        IF match=true THEN exit;
        r进入 R 中;】;
    】
ENDP;
FUNC matching(s;r);Boolean;
    c;=0;
    {s 与 r 根据 k 连续位匹配规则进行逐位比较;}
    从 s 与 r 中依次取对应位;
    IF s 与 r 的对应位相同 THEN c;=c+1
                ELSE c;=0;
    IF c=k THEN RETURN(true)
           ELSE RETURN(false)
ENDF;
```

3 改进算法检测能力分析

3.1 理论分析

设所有长度为 l 的模式空间为 U，匹配阈值为 $k(0<k\leqslant l)$，检测元 a_i，a_j 和 b 的覆盖空间分别为 A_i，A_j 和 B，其中 $Match(a_i,a_j)$，$\rightarrow Match(a_i,b)$，a_i 和 a_j 匹配的字符块为 w。

由于单个检测元的覆盖空间只与 k-连续位匹配规则的匹配阈值和检测元的长度有关系，因此在检测元的长度和匹配阈值确定的情况下，单个检测元的覆盖空间是相同的，即 $|A_i|=|A_j|\approx|B|$。

由于 $Match(a_i,a_j)$ 且 a_i 和 a_j 匹配的字符块为 w，根据 k-连续位匹配规则，$|w|\geqslant k$。$A_i\cap A_j=\{\forall p\mid p\in U$，且 p 中含有字符块 $w\}$，所有含有字符块 w 的模式 p 均与检测元 a_i 和 a_j 同时匹配，即检测元 a_i 和 a_j 能同时覆盖模式 p，模式 p 构成了 $A_i\cap A_j$。若 a_i 与 a_j 中满足匹配条件的子串有 m 个 $(m\geqslant 1)$，则 $A_i\cap A_j$ 就是这 m 个子串所得覆盖集合的并集，故有 $|A_i\cap A_j|_{min}=2^{l-k}$，即 $|A_i\cap A_j|\geqslant 2^{l-k}\geqslant 1$。可见 a_i 和 a_j 覆盖空间的交集不为 $\emptyset$。

由于 $\rightarrow Match(a_i,b)$，可设 a_i，b 匹配的最大字符块为 W，根据 k-连续位匹配规则，$|W|<k$。可将 a_i 和 b 表示为：

$a_i=p_1p_2\cdots p_iWp_{i+1}\cdots p_l$

$b=b_1b_2\cdots\overline{p}_iW\overline{p}_{i+1}\cdots p_l$

从 $p_1p_2\cdots p_i$ 或 $b_1b_2\cdots\overline{p}_i$ 或 $p_{i+1}\cdots p_l$ 或 $\overline{p}_{i+1}\cdots b_l$ 中紧靠 W 取出连续的 $k-|W|$ 位，分别与 W 构成子串，记为 W_1，W_2，W_3，W_4。显然，$|W_1|=|W_2|=|W_3|=|W_4|=k$，且 $W_1\neq W_2$，$W_3\neq W_4$，因此 $A_i\cap B$ 中的模式只能是同时含有 W_1 和 W_4 或同时含有 W_2 和 W_3 的模式，得：

$$|A_i\cap B|=2\times 2^{l-2k+|w|}=2^{l-2k+|w|+1}$$

W 在 a_i，b 中位置相同，若 $|p_1p_2\cdots p_i|<k-|W|$ 或 $|\overline{p}_{i+1}\cdots p_l|<k-|W|$，必有 $|b_1b_2\cdots\overline{p}_i|<k-|W|$ 或者 $|p_{i+1}\cdots p_l|<k-|W|$，此时 W_1 或 W_2 或 W_3 或 W_4 不存在，由于 $|W|\in[0,k-1]$，可得 $|A_i\cap B|_{max}=2^{l-k}$。

检测元 a_i，a_j 的整体覆盖空间为 $A_i\cup A_j-A_i\cap A_j$，检测元 a_i，b 的整体覆盖空间为 $A_i\cup B-A_i\cap B$，可见：

$$|A_i\cap A_j|_{min}=2^{l-k}\geqslant|A_i\cap B|_{max}$$

故 $|A_i|+|A_j|-|A_i\cap A_j|\leqslant|A_i|+|B|-|A_i\cap B|$。说明否定算法经过改进后，检测元的整体检测能力得到一定程度的提高。从另一个角度看，在系统资源有限的情况下，要达到相同的覆盖率，改进算法比原算法所需检测元的数量少，因此对于实际系统具有较好的实用价值。

3.2 实验分析

在理论分析的基础上，笔者通过实验对否定选择算法改进前后的整体覆盖空间进行比较。

由于检测元的覆盖空间只与 l 和 k 的值有关，不受自我集 S 的影响，实验过程中不考虑 S。为避免检测元产生过程的随机性给实验结果产生影响，在检测元生成阶段，对随机产生的同一个位串，一是根据原算法生成检测元集 R_1，二是根据改进后的算法生成检测元集 R_2。具体的实验步骤如下。

第一步　对模式长度 l，匹配阈值 k，检测元数量 n 等参数进行初始化。

第二步　构造长为 l 的模式空间。

第三步　随机产生一个长为 l 的位串。

第四步　该位串进入 R_1，若满足改进后的算法要求，就进入 R_2。返回到第三步，直到 $|R_1|=n$。

第五步　若 $|R_2|$ 小于 n，再随机产生多个位串，直到 $|R_2|=n$。

第六步　对 R_1 和 R_2 中所有检测元在相同阈值的匹配规则下进行整体覆盖空间的分析，产生分析结果。

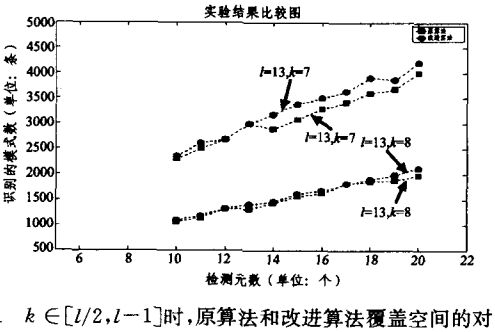

图 1　$k\in[l/2,l-1]$ 时，原算法和改进算法覆盖空间的对比

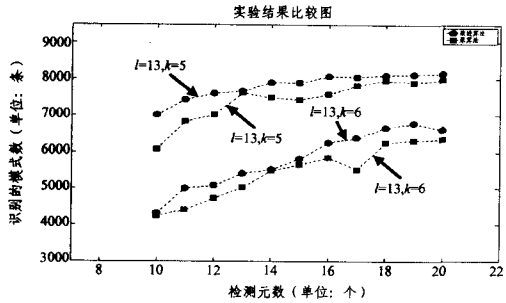

图 2 $k\in[0,l/2]$时原算法和改进算法覆盖空间的对比

实验结果如图 1、图 2 所示。实验结果表明，在模式长度 l 固定的情况下，无论匹配阈值 k 如何变化，改进算法能够识别的模式数量均高于原算法，即改进算法的检测能力优于原算法，特别是当 $k\in[0,l/2]$时，效果更加明显。图中算法改进前后识别的模式数量相同的点，说明存在原算法产生的检测元互不匹配的情况。

结束语 基于免疫原理的网络入侵检测技术中，产生有效检测元的否定选择算法采用 k 连续位匹配规则，使得有效检测元集 R 中的检测元存在相互匹配的可能，导致系统在资源有限的情况下对网络入侵的检测能力无法达到最大化。本文对否定选择算法进行了改进，提高了检测元集的整体覆盖能力和系统对网络入侵的检测能力，具有较好的实用价值。

参 考 文 献

[1] Forrest S, Parelson A, Allen L, et al. Self-nonself Discrimination in A Computer[C]// Proceeding of the 1994 IEEE Symposium on Research in Security and Privacy. Los Alamos, CA, IEEE Computer Society Press, 1994

[2] Forrest S, Hofmeyr S A, Somayaji A. Computer Immunology [J]. Communication of the ACM, 1997, 40(10): 88-96

[3] Haeseleer, Forrest S, Helman P. An Immunological Approach to Change Detection: Algorithms, Analysis and Implication // Proceedings of the 1996. IEEE Symposium on Computer Security and Privacy. IEEE Computer Society Press, Los Alanmitos, CA, 1996: 110-119

[4] Hofmeyr S A. An Immulogical Model of Distributed Detection and Its Application to Computer Security [D]. Ph. D Thesis. University of New Mexico

[5] Kim J, Bentley P. The Human Immune System and Network Intrusion Detection [C]// EUFIT99

[6] Kim J, Bentley P. Evaluation of Negative Selection in An Artificial Immune System for Network Intrusion Detection. GECCO, 2001

[7] Dasgupta D, Attoh-Okine N. Immunity-based Systems: A Survey// Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Orlando, Florida, Oct. 1997: 363-374

（上接第 61 页）

表 1 不同调度算法的性能比较

	UTIL	MRA	THOR	GRD
FC_EDFTBS	0.937	0.003	0.996	637ms
EDFTBS0.15	0.919	0.180	0.816	277ms
EDFTBS0.01	0.887	0.001	0.898	1188ms

5 进一步的研究

文献[12,13]认为，将反馈控制理论应用到计算系统 QoS 保障，关键需要解决 QoS 需求在反馈控制系统中的映射、控制器/执行器的设计、控制环中目标系统（被控对象）的建模等方面的问题。

文献[8,9]和我们的实验显示 CPU 的利用率对任务的时限错过率有显著影响。我们利用改进的 TBS 作为执行器，其规模 U_s 作为控制量，控制 CPU 利用率，从而保证系统的时限错过率满足要求，完成实时 QoS 在控制系统的映射。控制目标建模是确定控制器参数的关键，也是在计算系统 QoS 保障中应用控制方法的难点[12]。目前国内外对特定计算系统精确建模并不多见。文献[13,14]利用离线系统辨识方法对通用 Web 服务器 QoS 建模。文献[15]对排队论进行改进，将请求的时限引入排队模型，提出实时排队论，并利用它对 EDF 调度算法进行建模。在今后的研究中，我们将以上述文献为基础，针对开放式实时系统的特点，建立目标系统精确模型，以进一步提高反馈调度器的性能。

结束语 本文针对负载存在突发变化的开放式硬实时系统（例如适用于工业控制的 EWS），应用反馈控制方法设计负载自适应的 FC_EDFTBS。给出了调度算法的反馈控制结构以及控制器的差分方程。通过反馈控制将硬实时系统的关键性能指标：时限错过率与调度器直接关联，从而明显减弱系统负载动态的变化对实时性能的影响。最后，我们基于 VxWorks+Goahead 嵌入式 Web 服务器系统，开发中间件实现 FC_EDFTBS 算法，通过实验对比传统调度算法的性能。实验证明 FC_EDFTBS 在负载突变时，可以获得满意的时限错过率和较高的资源利用率。

参 考 文 献

[1] Deng Z, Liu J W S. Scheduling real-time application in open environment[J]// Proceedings of the 18th IEEE Real-Time Systems Symposium. Los Alamitos, CA: IEEE Computer Society Press, 1997: 308-319

[2] McCombie B. Embedded Web servers now and in the future[J]. Real-Time Magazine, 1998(1): 82-83

[3] Jane W, Liu S. Real-Time Systems[M]. Beijing: Higher Education Press arrangement with the original publisher, Pearson Education, Inc., 2002: 195-218

[4] Abeni L, Buttazzo G. Intergrating multimedia applications in hard real-time systems // Proc. 19th IEEE Real-time SystemsSymposium. Madrid, Spain

[5] Cervin A, Eker J. Control-scheduling Codesign of Real-time Systems: The Control Server Approach// Proc. Journal of embedded computing, 2004

[6] Crovella M E, Bestavros A. Self-similarity in World Wide Web T-raffic: Evidence and Possible Causes[J]. IEEE/ACM Transaction on Networking, 1997, 5(6): 835-846

[7] Abdelzaher T F, Stankovic J A, Lu Chenyang. Feedback Performance Control in Software Services[J]. IEEE Control Systems Magazine, June 2003

[8] Lu C, Stankvoic J A. Design and Evaluation of a Feedback Control EDF Scheduling Algorithm // IEEE Real-Time Systems Symposium. Phoenix, AZ, Dec. 1999

[9] Lin Suzhen, Manimaran G. A FeedBack-based Adaptive Algorithm for Combined Scheduling with Fault-Tolerance in Real-Time Systems// Proc. Conference on High Performance Computing (HiPC). Bangalore, India, Dec. 2004: 101-110

[10] VxWorks-Programmer's Guide 5.5, Wind River Systems[EB], INC, 2002

[11] Goahead Software Foundation[EB]. http://www.goahead.com

[12] Hellerstein J L, Diao Yixin, Parekh S, et al. Feedback Control of Computing Systems [M]. Wiley-IEEE, John Wiley & Sons, 2004: 31-56, 293-334

[13] Zhang Ronghua, Lu Chenyang, Abdelzaher T F. ControlWare: A Middleware Architecture for Feedback Control of Software Performance // The 22nd International Conference on Distributed Computing Systems(ICDCS'02). 2002

[14] Lu Chenyang, Abdelzaher T F, Stankovic J A. A Feedback Control Approach for Guaranteeing Relative Delays in Web Servers // IEEE Real-time Technology and Applications Symposium. June 2001

[15] Lehoczky J P. Real-Time Queueing Theory[J]// Proceedings of the 17th IEEE Real-Time Systems Symposium. 1996: 186-195