SA算法在基于模型推理入侵检测中的应用

陈 波，于 泠，肖军模

(解放军理工大学通信工程学院 南京 210007)

【摘要】鉴于模型推理的入侵检测方法，需要在庞大的审计记录空间中搜索巨量的攻击脚本子集中的最优值，对于这一NP类完全问题，提出了应用模拟退火算法。并建立了攻击检测的优化问题模型，给出了攻击检测实验中的解空间、目标函数、新解的产生和接受准则，得到了一个合理的冷却进度表，并对实验中的模拟退火算法进行了并行化研究。实验证明，与传统的贪心算法相比，应用模拟退火算法提高了进化速度和全局寻优能力，较好地解决了搜索效率问题。

关 键 词 模拟退火算法； 模型推理； 入侵检测； 网络安全

中图分类号 TP393.08 文献标识码 A

An Application of Simulated Annealing Algorithm in Model-Based Reasoning Intrusion Detection

CHEN Bo，YU Ling，XIAO Jun-mo

(Institute of Communication Engineering , PLA University of Science and Technology Nanjing 210007)

Abstract It is needed to search among all the possible attack subsets and to match the events recorded in the audit trail. To make a decision about the realism of the hypothesis corresponding to a particular subset is difficult in model-based reasoning Intrusion Detection System. We present using Simulated Annealing(SA) algorithm to solve this NP-complete problem. Modeling a optimizing issue of attack detection first, and give the solve space, the target function, the creation of new solution and accept the standard, we got a reasonable cooling schedule. The parallelization of SA algorithm is also presented. The experiments indicate that the SA algorithm can improve the evolution speed and the abilities of seeking the global excellent result, and resolve to the efficiency problem of searching well.

Key words simulated annealing algorithm; model-based reasoning; intrusion detection; network security

1 基于模型推理的入侵检测方法

入侵是指任何试图对资源的完整性、保密性或可用性产生危害的行为。入侵检测是对这些行为的识别。由于入侵模式的多样性，入侵检测策略和模型也具有多种不同的类型[1,2]。基于模型推理(Model-Based Reasoning)的入侵检测方法通过为入侵行为建立特定的模型，结合攻击脚本推理出入侵行为是否出现。入侵检测利用的信息很大一部分是来自系统的日志和审计信息。入侵者经常在系统日志中留下他们的踪迹，因此充分利用系统的日志文件和审计信息是检测入侵的必要手段。日志中包含发生在系统和网络上的不寻常和不期望活动的证据，通过分析日志文件和审计信息，能够发现成功的入侵或入侵企图。

收稿日期：2003 – 06 – 16
基金项目：国家自然科学基金资助项目(69931040)
作者简介：陈 波(1972 –)，男，博士生，讲师，主要从事网络信息安全、人工智能方面的研究.

在Unix操作系统中带有许多审计机制，并且还可以再配置审计工具，因此能够产生大量的审计数据。为从庞大的信息中提取出与安全相关的信息，以产生攻击脚本，首先对历史审计文件进行预处理，产生用户会话矢量。用户会话包括login和logout之间的所有事件。会话矢量$\boldsymbol{A}=<a_1,a_2,\cdots,a_n>$描述单一会话过程中用户进行的各种事件数量。会话开始于login，终止于logout，login和logout次数也作为会话矢量的一部分。可以监视20多种事件，如：登录的时间、登录失败数、写文件数、读文件数等。

然后将产生的用户会话矢量提交给分析模块，分析模块可以采用统计、专家系统等方法建立用于入侵检测的攻击脚本，存入攻击脚本库中。脚本库是一个$m\times n$维的事件矩阵，一个列向量表示一种攻击所对应的会话矢量。

检测时先将这些攻击脚本的子集假设为系统正面临的攻击，然后通过一个预测器程序模块，根据当前行为模式产生需要验证的攻击脚本子集，并将它传给决策器，决策器收到信息后，根据这些假设的攻击行为在审计记录中的可能出现方式，将它们翻译成与特定系统匹配的审计记录格式，在审计记录中寻找相应信息来判定该攻击是否发生。

基于模型推理的入侵检测的实质是在审计记录中搜索可能出现的攻击子集。在实际应用中，发现通常系统的审计记录数据量庞大：一个典型的*C*2级审计机制，在一个多用户系统，不到1 h时便会产生1 GB的数据量；一台4CPU SPARC SUN系统需要用两个CPU和所有磁盘通道来记录其它两个CPU的活动；在网络环境中，数据量将更加膨胀。同时，攻击脚本子集的数量也很巨大，假设攻击脚本中与入侵潜在相关的度量有n个，则这n个度量所构成的子集数便达到2^n个。可以将在审计记录中寻找相应信息来确认或否认这些攻击的问题看作类似于一个非确定型多项式类(Nondterministic Polynomial, NP)完全问题，即在复杂而庞大的搜索空间中寻找最优解或准优解。在求解此类问题时，若不能利用问题的固有知识来缩小搜索空间则会产生搜索的组合爆炸。

因此本文在模型推理方法已有成果的基础上，提出了将模拟退火(Simulated Annealing，SA)算法运用在基于模型推理的入侵检测中。论述了模拟退火算法4要素：解空间、目标函数、新解的产生、接受准则在攻击检测中的选用，给出了实验的主要SA算法代码，并对用于攻击检测的SA算法进行了并行化研究。

2　SA算法

求NP完全问题的最优解有多种方法，如线性规划、贪心算法等，但这些算法往往由于计算时间的限制不具有可行性。而入侵检测一个重要的特征就是要具有实时性，因此上述算法不适合应用于模型推理的入侵检测[3,4]。

模拟退火算法是一种解大规模组合优化问题[5]，特别是NP完全问题的有效近似算法。它源于对固体退火过程的模拟，采用Metropolis接受准则，并用一组称为冷却进度表的参数控制算法进程，使算法在多项式时间里给出一个近似最优解。

模拟退火算法的主要特点是高效、健壮、通用和灵活。尽管该算法本身在理论和应用方法上仍有许多待进一步研究的问题，但实践证明，模拟退火算法对于组合优化中的NP完全问题非常有效。

3　SA应用研究

3.1　问题的描述

本文目标是：根据审计跟踪记录中的事件确定在所有可能攻击子集中哪一个对系统最具威胁性，即对于一个特定的攻击子集，统计所有攻击产生的每类事件的个数，如这一数量少于或等于那种被记录事件的数量，“该攻击子集存在”的假设是成立的。为了用数学形式更精确地描述该问题，设m为整个系统审计事件数；n为整个系统潜在攻击数: $\boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12} & \mathrm{L} & a_{1n} \\ a_{21} & a_{22} & \mathrm{L} & a_{2n} \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ a_{m1} & a_{m2} & \mathrm{L} & a_{mn} \end{pmatrix}$。

这是一个$m\times n$维攻击事件矩阵，包含每个攻击产生的事件集。其中a_{ij}是由攻击j产生的第i种类型的审计事件数，$a_{ij}\geqslant 0$。

$\boldsymbol{W}=(w_1,w_2,\cdots,w_n)$是$n$维权重行向量，$w_j$ ($w_j>0$)是与攻击j相关的权重($1\leqslant j\leqslant n$)；

$\boldsymbol{O}=(o_1,o_2,\cdots,o_m)$是$m$维审计向量，$o_i$为审计记录中第$i$种类型事件的数量($1\leqslant i\leqslant m$)；

$\boldsymbol{H}=(h_1,h_2,\cdots,h_n)^T$是$n$维假设攻击列向量，也就是一个特定的攻击子集。如果攻击j存在则$h_j=1$，否则$h_j=0$。基于模型推理的入侵检测问题可精确描述为：寻找$\boldsymbol{H}$向量，使其最大化$\boldsymbol{W}\times\boldsymbol{H}$，且满足约束条件$\boldsymbol{AH}\leqslant O$，其中$\boldsymbol{AH}\leqslant O$的含义是：$\forall k\in[1,n],\exists a_{k1}h_1+a_{k2}h_2+\cdots a_{kn}h_n\leqslant o_k$。

3.2 SA算法的应用

模拟退火算法应用的一般形式是：从选定的初始解开始，在借助于控制参数t递减时产生的一系列马尔可夫链中，利用一个新解产生装置和接受准则，重复进行“产生新解、计算目标函数差、判断是否接受新解、接受(或舍弃)新解”4个步骤，不断对当前解迭代，从而到达使目标函数最优。因此模拟退火算法要素有：1) 解空间：模型推理的入侵检测问题是一个有约束的优化问题，对此，限定解空间是所有可行解的集合，即：$\boldsymbol{S}=\{(h_1,h_2,\cdots,h_n)^T \mid \boldsymbol{A}\,(h_1,h_2,\cdots,h_n)^T\leqslant O,\ h_j\in\{0,1\}\}$，其中$h_j=1$表示攻击$j$存在。初始解选取(0,0,⋯,0)。2) 目标函数：目标函数是一个需要求最大值的价值函数：$f(h_1,h_2,\cdots,h_n)=\boldsymbol{W}\times\boldsymbol{H}=w_1h_1+w_2h_2+\cdots+w_nh_n$，但必须满足约束条件$\boldsymbol{A}(h_1,h_2,\cdots,h_n)^T\leqslant O,\ h_j\in\{0,1\}$。3) 新解的产生：随机选取攻击$i$。若$i$没被选过，则直接将$i$加入，或去掉已选择的攻击$j$，再将$i$加入；若$i$已被选过，则去掉$i$，并同时随机加入攻击$j$。即：$h_i=1-h_i$，且(或)$h_j=1-h_j$，$i\neq j$。4) Metropolis接受准则：

$$P=\begin{cases}0 & m+\Delta m>O\\ 1 & m+\Delta m\leqslant O\text{且}\Delta f>0\\ \exp(\Delta f/t) & \text{否则}\end{cases}$$

其中Δf是价值函数的差，$1\leqslant k\leqslant n$

$$\Delta f=\begin{cases}w_i & \text{将攻击}i\text{直接加入}\\ w_i-w_j & \text{将攻击}i\text{加入, 同时去掉攻击}j\\ w_j-w_i & \text{将攻击}j\text{加入, 同时去掉攻击}i\end{cases}$$

$$\Delta m=\begin{cases}a_{ki} & \text{将攻击}i\text{直接加入}\\ a_{ki}-a_{kj} & \text{将攻击}i\text{加入, 同时去掉攻击}j\\ a_{kj}-a_{ki} & \text{将攻击}j\text{加入, 同时去掉攻击}i\end{cases}$$

3.3 攻击检测实验

实验中分析模块对模拟攻击产生的日志进行处理后得到攻击矩阵。实验中需要选择一个合理的冷却进度表。冷却进度表是一组控制算法进程的参数，主要包括：控制参数t的初值t_0；控制参数t的衰减函数$t_{k+1}=\alpha t_k$；马尔可夫链的长度L_k以及停止条件。

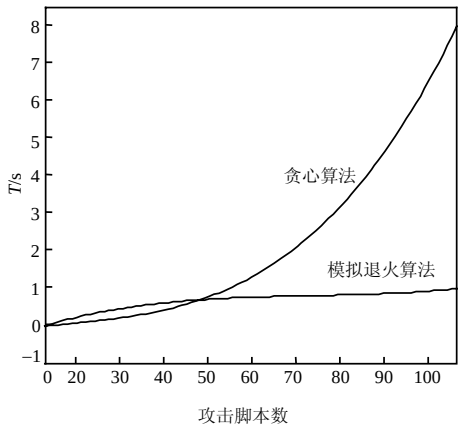

图1 贪心算法和模拟退火算法求解时间

在实验中，取$t_0=0.5$，$\alpha=0.9$，$L_k=10n$(n是问题的规模)，停止条件是在s个相继的马尔可夫链中解无任何变化就终止算法，其中$s=2$。实验证明上述退火算法有较好的收敛性。

本文分别用贪心算法和模拟退火算法对问题进行了求解，结果表明，随着攻击库中攻击脚本数的增加，贪心算法的时间剧增，而模拟退火算法的处理时间近似于线性，如图1所示。用于攻击检测的SA算法伪代码：

```
void annealing()
{initialize(i0,t0,L0);// 初始化解i0，控制参数t0和变换个数L0
 k=0;   i=i0;
 do
 {   for(int l=1;l<=Lk;l++)
     { generate(j from Si);//产生新解
```

```
    // 根据Metropolis接受准则，判断是否接受新解
        if(m+△m> O)    continue;
    else
                if(m+△m≤O&&△f>0)       i=j;
                else    if(exp(△f/t)>random[0,1]    i=j;
    }
  k=k+1;
  calculate_length(Lk);
  calculate_control(tk);
  }while(!stop_condition);
}
```

3.4　算法的并行化

为了进一步提高SA算法的效率，还对其进行了并行化。采用协同试验并行策略：由p台处理机同时产生各自的新解并作出判断，然后对其中满足接受准则者按选取法则选一个予以接受，再在此基础上进行下一轮实验。执行流程如图2所示。

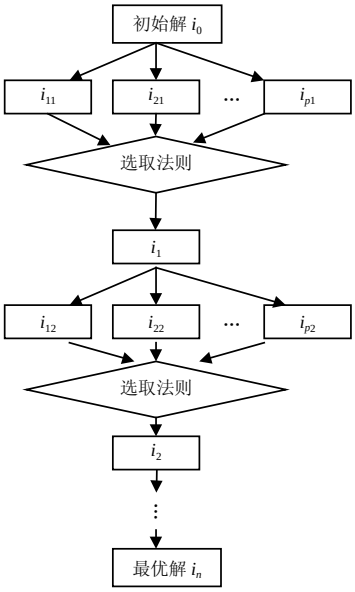

图 2　SA 算法的并行流程图

4　小　结

采用基于模型推理的入侵检测方法具有的一些优点是：

1) 可以根据对攻击的分析来设计不同的攻击脚本；

2) 修改或者增加一个新的攻击脚本较为容易；

3) 需要存储的审计事件只是那些至少在一个攻击脚本中出现过的，这样可以适当减少存储的数据量和需要处理的数据量；

4) 可以为每个脚本赋以权重，使得攻击检测更加有效。

虽然在复杂而庞大的审计记录搜索空间中寻找相应信息来确认或否认攻击的问题被看作类似于一个NP完全问题，但是选用SA算法可以很好的解决这一问题。SA算法在组合优化问题求解、自适应控制、规划设计、机器学习和人工生命等领域的应用中已展现了其特色和魅力，它在解决网络安全问题的应用中同样具有广阔的前景。

参　考　文　献

[1] Sandeep Kumar. Classification and detection of computer intrusions:[Phd Thesis][D]. COAST Laboratory Dept.of Computer Sciences Purdue University. Augest,1995

[2] Sandeep Kumar, Eugene Spafford..An application of pattern matching in intrusion detection[R].Technical Report CSD-TR-94-013, Department of Computer Sciences, Purdue University,1994

[3] Ludovic Me. Gassata a genetic algorithm as an alternative tool for security audit trails analysis[C]. In Proceedings of The Rst International Workshop on The Recent Advances in Intrusion Detection (RAID'98) Beigum, 1998,1123-1127

[4] Jong A K, Spears M W. Using genetic algorithms to solve NP-Complete problems[C]. In Proceedings of the Third International Conference on Genetic Algorithms San Francisco, 1991,124 -132

[5] 康立山, 谢 云, 尤矢勇, 等. 非数值并行算法-模拟退火算法[M].北京:科学出版社,2000:114-131

编　辑　孙晓丹