一个基于知识的优化入侵检测系统

丁勇 陆晟 龚俭

（东南大学 计算机系，210096 南京）

【摘要】本文介绍了一个脚本描述能力较完备的入侵检测系统的设计和实现，该脚本语言能完整地描述攻击的特征，从而增强了系统的易维护性和可扩充性。本文还提出了对脚本中规则集进行优化的算法，以提高入侵检测系统的运行效率。

【关键词】基于知识的入侵检测；上下文相关分析；优化网络；更新复杂度

中图分类号：TP393

A Knowledge-based Optimized Intrusion Detection System

Ding Yong， Lu Shen， Gong Jian

(Southeast University, Computer Science Dept., 210096 NanJing, P.R.China)

【Abstract】This paper introduces the design and implementation of an intrusion detection system with powerful ability for script description. The script language can describe all features of any intrusive behavior, thus enhances the maintenance and expandability of the intrusion detection system. The paper also puts forward an algorithm for optimizing the set of rules in script, in order to promote the efficiency of running system.

【Key words】knowledge-based intrusion detection; context-related analysis; optimized network; update complexity[1]

1 引言

网络入侵检测系统是保障网络正常运行的重要工具，它可以有效地检测网络用户的使用行为，及时发现有害的入侵活动。

网络入侵检测可以分为基于知识的（knowledge-based）和基于行为的（behavior-based）两类。前者通过对采集的信息按已知的知识进行分类，发现入侵行为，它依赖于所谓的攻击知识库，有较高的处理效率；后者通过观察相对于正常系统行为的偏离来检测入侵行为，这类系统比较完备，但虚警率较高。

本文将提出一个基于知识的网络入侵检测系统的体系结构，该系统的脚本描述功能比较完备。另外，还将给出对脚本中规则集进行优化的思想，以及基于该思想实现的一个算法，并分析了优化能取得的效果。

2 系统的主要特点

本系统是一个面向网络的、基于知识的入侵检测系统，它在单机上运行，适合于低速网络环境下的入侵检测。

该系统使用一种脚本语言来完整地描述各种攻击。该脚本语言是一种说明性语言，它不仅能描述上下文无关的分析，还能描述上下文相关的分析（上下文相关是指对当前安全事件的分析与过去的相关历史有密切联系，如扫描分析）。它相对于过程性的 C 语言抽象级别更高，系统依赖的所有的攻击知识都集中在脚本中，并构成系统的知识库，这使得系统更易于维护和扩充。

脚本语言主要的语法成份包括变量、函数和上下文。变量能灵活地存取报文中的指定字段，它用从报文 IP 层开始的位移和字段长度表示，位移和长度的单位可以为字节和位两种。函数可以描述一些经常用来表述攻击的操作，如字符串匹配等。该脚本语言的最大特点在于能描述上下文相关的分析，例如语句:

```
same {12,4} in 1000 node 1000 do diff {16,4}>=20
```

作者简介：丁勇，硕士研究生，主要研究方向为网络安全。陆晟，博士研究生，主要研究方向为网络安全。龚俭，工学博士，东南大学计算机系教授、博士生导师，主要研究方向包括网络管理、网络安全、网络行为学等。

表示在一段时间内收到的源 IP 地址（用{12,4}表示）相同的报文中，若不同的目的 IP 地址（用{16,4}表示）值的数量大于等于 20 则为 TRUE，否则为 FALSE。其中的“in 1000”表示这种时间间隔，“node 1000”与系统存储分配有关。

为了提高系统的运行效率，本系统还对脚本中的规则集进行了优化，并且取得了显著的效果，这种方法可以在一定程度上缓和 CPU 速度跟不上网络带宽的矛盾。

3 系统的体系结构

该系统的总体结构如图 1 所示，它包含编译模块、优化模块、分析决策模块、数据采集模块、报警模块。

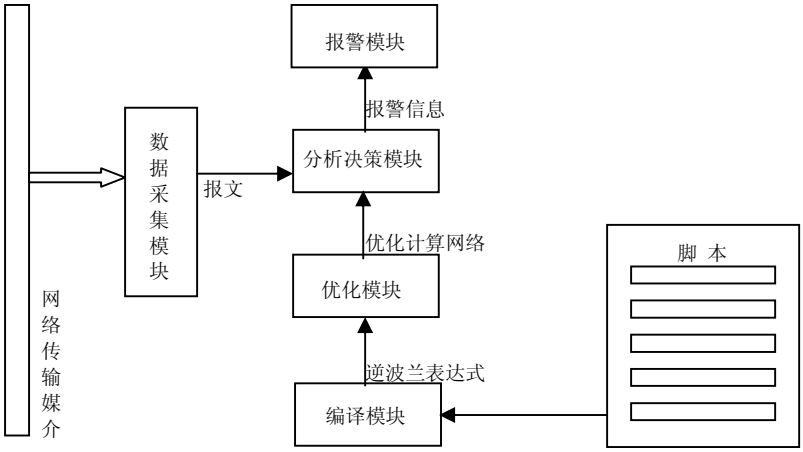

图 1 系统的总体结构

编译模块的功能是将抽象级别较高的脚本语言翻译为抽象级别较低的中间语言，中间语言可以有多种形式，因为本系统的规则都是逻辑表达式，所以采用逆波兰表达式作为中间语言比较适合。

优化模块是该系统最重要也是最复杂的部分，优化效果的好坏对系统运行的效率有很大的影响。该模块的最终目的是将逆波兰表达式集合转换为一个有向无回路的优化网络，该网络中每个节点表示一步操作。对于给定的报文，分析决策模块由入口节点开始计算，并根据当前节点的计算值选择不同的路径进行下一步判断，即对于不同报文的计算将经过网络中不同的路径，当分析决策模块计算到出口节点时，则对于当前报文，所有规则的值均已计算出来。这种网络的特征类似于一个有限状态自动机。

分析决策模块根据优化网络对捕获到的报文进行分析，它应包含计算优化网络中节点的算法，该算法对节点进行解释执行。分析决策模块应当同时具有上下文无关分析和上下文相关分析的能力，当判断发生某种攻击时，将调用报警模块进行报警。

报警的方法可以分为被动方式和主动方式，前者包括记录日志和通知管理员，后者包括切断攻击者的连接，甚至调整防火墙的配置以阻止攻击者的其它动作。报警最常用的方式是记录日志，可以详细记录攻击发生的时间等信息，以便以后分析查阅，但对于实时检测系统而言，还应当在第一时间通知系统管理员，并在主控台实时显示报警。

数据采集模块包括数据采集和预处理两部分，前者将网卡置为混合模式，从数据链路层获取网络报文，后者对获取的报文进行预处理，使其具有符合定义的标准形式。

4 规则集的优化

为了提高系统的运行效率，对于给定报文采用串行计算每条规则的方法显然是不合理的。如果分析决策模块对报文的处理速度跟不上网络流量，那么将造成报文的丢失，从而产生入侵行为的漏判。因此面向规则集的优化势在必行。下面首先分析一下这种优化的可能性和预计的效果，然后是优化算法的基本思想。

4.1 优化的可能性

对本系统的规则集的研究发现，各条规则间存在大量的相同操作，如近一半的规则首先要判断报文是不是 tcp 报文，对于这些相同的操作只做一次当然要比重复做上百次节省大量的时间，因此进行合并相同操作的优化应该能

取得相当不错的效果。

规则是由一些子操作（如判断报文是否 tcp 报文）通过逻辑运算符连接而成，在具体计算每条规则时存在逻辑上的优化。如规则“A and B”的逆波兰表达式形式为“A B and”，要通过逆波兰表达式计算该规则必须先计算出 A 和 B 的值，而对于该规则，若 A 为假则规则为假，B 的值可以不进行计算。

另外规则之间也存在逻辑上的优化。如有两条规则，一条规则要求报文的目的端口必须为 80，另一条规则要求目的端口必须为 1080，则这两条规则不可能同时成立。我们把这两条规则称为不相交。于是我们可以找出这样的规则集合，该集合中的规则两两不相交，这样该集合中最多只能有一条规则成立。

以上三方面的优化结合起来可以显著提高系统的运行效率。

4.2 优化算法的基本思想

由于逆波兰表达式的形式并不适合于规则的快速运算，因此使用了另一种结构取代它，即广义表。假设有一条规则“A and (B or C) and (D or E)”，则它可以转化成如图 2 所示的广义表，其中 child 指针表示 and 关系，而 brother 指针表示 or 关系，图中实心节点为结果节点。

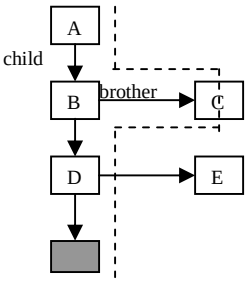

图 2 有回溯的广义表

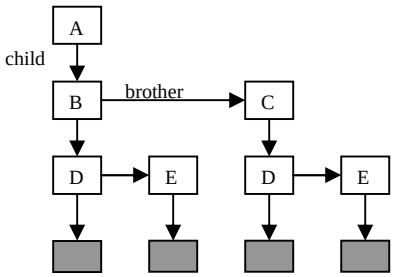

图 3 无回溯的广义表

但根据图 2 的广义表对报文进行分析的过程中可能要遇到回溯，如若 A 为真、B 为假、C 为真、D 为真，则广义表的计算路径在图中用虚线表示，将有一步回溯。

回溯将增加规则的计算时间。为了消除回溯，可以构造图 3 形式的广义表。该表中引入了重复节点，但这是消除回溯所必须的，也就是说，该方法是用空间的牺牲换取了效率的提高。这样，对于给定报文的计算方法可以简单描述为以下两个步骤：

第一步：从入口节点开始判断当前节点，若为真则沿 child 链继续判断，否则沿 brother 链判断；

第二步：若最终达到结果节点，则该规则为真，若无路可走（brother 链为空时）则该规则为假。

广义表的形式只是对规则内部逻辑的优化，而对于相同操作的合并及规则间逻辑的优化，可以在广义表的基础上通过节点的合并和路径的设置来实现。

总之，优化算法最终的目的是生成一个有向网络，不同的报文可能通过网络中不同的路径，且所有路径都是无回溯的，只要计算完该路径，则对于该报文，所有的规则值都已计算出。

5 系统的改进

本系统采用优化规则集的方法可以得到显著的效果，但这种体系结构仍存在缺陷。例如，当规则集进行了部分的改变时，本系统需要重新调用优化算法生成新的优化网络，因此它的更新复杂度（update complexity）比较高，特别是对于频率很高的更新将不适合。解决的方法是当规则集改变时动态修改优化网络，而不是重新生成优化网络，这样可以大大降低系统的更新复杂度，但算法难度比较大。这也是本系统今后进一步改进的方向。

6 结论

一个结构灵活、功能完备的网络入侵检测系统是本研究领域最终所追求的目标，本文的研究工作也是向这个方向的一个尝试。其核心思路是使用了一种比较完备的说明性脚本语言完整地描述各种攻击，使系统依赖的所有攻击知识集中在脚本中，这将有利于系统的易维护性和可扩充性。对脚本中的规则集进行了优化，是本文的另一个重点，这将有利于提高检测效率，该优化方法将适合于所有基于规则的系统。

7 参考文献（略）