一种快速检测入侵的方法

韩仲祥[1,2]，史浩山[1]，董淑福[2]，余　征[2]
(1.西北工业大学 电子信息学院，陕西 西安　710072；2.空军工程大学 电讯工程学院，陕西 西安 710077)

摘　要：研究了网络入侵检测的最优检测问题。针对攻击事件类型和规则的匹配问题，改进了限定条件下规则集的选取，通过建立凸松弛函数，确定问题的最佳解，最后通过模拟实验，验证了用最佳解匹配攻击数据包，效果明显优于BM方法。

关键词：入侵检测；规则集；匹配；凸松弛函数
中图分类号：TN915　　**文献标识码**：A　　**文章编号**：1009-3516(2008)01-0054-04

基于网络的实时入侵检测系统(Real-Time IDS：RT-IDS)主要面对的是：如何快速地处理网络中的海量数据，从中检测出带有入侵行为的数据包。国内外学者在这方面作了不少的研究[1-3]。本文在文献[4]研究的基础上，采用限定条件下的最佳规则来检测网络数据包，实验表明，此方法检测速度快于SNORT的BM方法。

1　事件驱动型模型结构

建立基于网络的实时入侵检测系统(RT-IDS)，基本组成部分如图1所示。系统中每一个信息包都被复制到RT-IDS。RT-IDS中的事件引擎，传输原始的数据文件包，以供事件分析；计算引擎要为每一事件作出声明，声明既包括非入侵也包括入侵，并对入侵进行分类。依靠复杂的事件驱动，1个RT-IDS能够区分包驱动或者事件驱动。在包驱动型的IDS中，如SNORT[5]，原始数据包自动送到计算引擎部分，检测规则会为每1个数据包作出相应的检测[6-10]。作者在文献[4]中就事件类型、攻击、检测规则和系统重组作了详细的讨论，仍然参照这些定义。

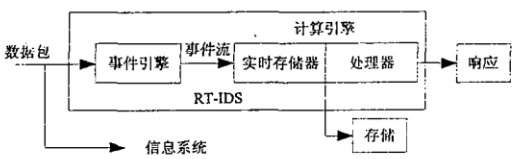

图1　实时入侵检测系统的组成
Fig.1　RT-IDS structure

2　限制条件下的规则集的选取

2.1　第1个限制条件：接收事件的系统时间

事件在到达系统之前排成序列(如图2所示)。序列事件的系统服务取决于事件的类型，即：事件类型i的规则对应于规则集R_i。规则的应用顺序如图3所示。在某时刻事件类型i的一个事件到达系统，此时假定已经有m_i个事件类型i($i=1,2,\cdots,N$)的事件到达系统，则系统处理这些事件的系统时间的期望值$T_{S_i}=(\sum_{i=1}^{N} m_i T_i)+T_i$。系统的服务时间$T=\sum_{i=1}^{N}\pi_i T_{S_i}=\sum_{i=1}^{N} m_i\pi_i T_i+\sum_{i=1}^{N}\pi_i T_i$，式中：$T_i=\sum_{j=0}^{N_i} p_{ij}T_{ij}$，$T_{ij}$为事件类型

收稿日期：2006-12-29
基金项目：教育部博士点基金资助项目(20050699037)；陕西省自然科学基金资助项目(2006F16)
作者简介：韩仲祥(1971-)，男，山东莒南人，讲师，博士生，主要从事计算机网络安全研究．E-mail：zhongxianghan@sina.com
史浩山(1946-)，男，陕西西安人，教授，博士生导师，主要从事数据通信与计算机网络研究．

i 中的一个事件与规则集 R_{ij} 进行匹配的服务时间。这里 $T_{i0} = T_{iN_i}$,如图 3 所示,一个类型 i 的正常事件仍然占用相同的规则匹配时间,$T_{ij} = \sum_{l=1}^{j} t_{il}$,因为规则 R_{i1}, R_{i2}, R_{iN_i} 按顺序进行检测,结合以上式子及 $\sum_{j=0}^{N_i} p_{ij} = 1$,即有:$T = \sum_{i=1}^{N}\sum_{j=1}^{N_i} v_{ij}t_{ij}$,式中:$v_{ij} = m_i q_{ij}$,$q_{ij} = 1 - \sum_{l=1}^{j-1} p_{il}$。因此第 1 个限制条件就是 Knapsack[2] 限制:

$$\sum_{i=1}^{N}\sum_{j=1}^{N_i} v_{ij}t_{ij} \leqslant D_{\max} \tag{1}$$

p_{ij} 和 t_{ij} 已知,并假定 m_i 的估计是可行的,实际中 m_i 的选取在一个门限内。$D_{\max}$ 的选取有 2 方面的因素:响应速度和顺序的稳定性。用 D_{emt} 来表示有效最大响应时间,选取 $D_{\max} = \min\{D_{\text{emt}}, T_{\text{iar}}\}$,$T_{\text{iar}}$ 表示事件平均到达时间。这种选择可以确保系统服务时间期望值不大于 T_{iar}。

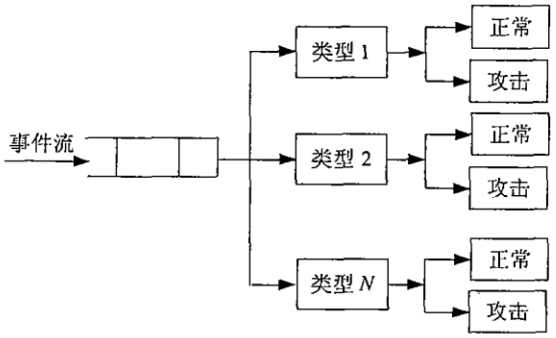

图 2　计算引擎系统

Fig. 2　Computing engine system

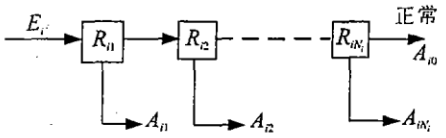

图 3　系统处理事件类型 i 的一个事件

Fig. 3　The system to process a event of event type i

2.2　第 2 个限制条件:攻击的规则匹配

假定 $x_{ijk} \in \{0,1\}$,并作如下定义:$\begin{cases} x_{ijk} = 1, \text{若规则 } R_{ijk} \text{ 存在于 } R'_i \text{ 中,且起作用;} \\ x_{ijk} = 0, \text{若规则 } R_{ijk} \text{ 存在于 } R'_i \text{ 中,但不起作用。} \end{cases}$

x_{ijk} 值的数量表征了规则集,利用这些数量就可以来讨论优化问题。很明显:$t_{ij} = \sum_{k=1}^{n_{ij}} t_{ijk}x_{ijk}$。对式(1)中的 Knapsack 限制进行扩展,有

$$\sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} a_{ijk}x_{ijk} \leqslant D_{\max} \tag{2}$$

式中:$a_{ijk} = v_{ijk}t_{ijk} = m_i q_{ij} t_{ijk}$,则可以知道每一规则的系数 a_{ijk} 由 3 部分决定,即:m_i、q_{ij} 和 t_{ijk}。

继续应用前面的假定:每一个攻击至多有一个规则对应,限制条件可以表示为

$$\sum_{k=1}^{n_{ij}} x_{ijk} \leqslant 1, i = 1,2,\cdots,N;\quad j = 1,2,\cdots,N_i \tag{3}$$

如果在规则集中确保一个攻击对应一个规则,这时,x_{ijk} 限制条件为:$\sum_{k=1}^{n_{ij}} x_{ijk} = 1$,对所有的 $i = 1,2,\cdots,N$;$j = 1,2,\cdots,N_i$。

为了实现最佳化问题,应用贝叶斯定理来表示规则 R_{ijk} 的期望值:

$$V_{ijk} = C_{ij}^{\beta}\pi_i p_{ij}(1 - \beta_{ijk}) - C_{ij}^{\alpha}\pi_i(1 - p_{ij})\alpha_{ijk} \tag{4}$$

式中:$C_{ij}^{\beta}\pi_i p_{ij}(1 - \beta_{ijk})$ 为漏报部分;$-C_{ij}^{\alpha}\pi_i(1 - p_{ij})\alpha_{ijk}$ 为误报部分。最终得到关于 x_{ijk} 为变量的线性方程,即

$$V(R') = \sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} c_{ijk}x_{ijk} \tag{5}$$

式中:$c_{ijk} = V_{ijk} = C_{ij}^{\beta}\pi_i p_{ij}(1 - \beta_{ijk}) - C_{ij}^{\alpha}\pi_i(1 - p_{ij})\alpha_{ijk}$。

3　轻量级的优化方法

结合式(1)、式(3)和式(5)可以得到问题的表达式:$\max_{x_{ijk}} V(R') = \sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} c_{ijk}x_{ijk}$,式中:$x_{ijk} \in \{0,1\}$ (

$\sum_{k=1}^{n_{ij}} x_{ijk} \leqslant 1$ 或 $\sum_{k=1}^{n_{ij}} x_{ijk} = 1$) 及 $\sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} a_{ijk}x_{ijk} \leqslant D_{\max}$,当数据 a_{ijk} 和 c_{ijk} 准确已知,关键的问题是寻找 1 个规则集,使它接近线性方程组,即符合 Knapsack 约束,这就是优化问题。

当数据 a_{ijk} 和 c_{ijk} 不能确定时,此情况则是最优化问题的鲁棒性,为了叙述方便,假定 a_{ijk} 和 c_{ijk} 已知且符合以下情况:$\bar{A} = \{(a_{ijk}): a'_{ijk} \leqslant a_{ijk} \leqslant \bar{a}_{ijk}$,对所有的 $i,j,k\}$;$D = \{(c_{ijk}): c'_{ijk} \leqslant c_{ijk} \leqslant \bar{c}_{ijk}$,对所有的 $i,j,k\}$。式中上下限都知道,这种情况对 $\bar{A}$ 中任意一个 a_{ijk} 都有一个规则集确保满足 Knapsack 约束,即

$$\sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} a_{ijk}x_{ijk} \leqslant D_{\max} \quad , \quad \text{对所有的 } a_{ijk} \in A \tag{6}$$

用这种方法定义以下集合:

$$\left\{\begin{array}{l} x_{ijk}: 0 \leqslant x_{ijk} \leqslant 1, \text{对所有的 } i,j,k \\ X_r^c = \sum_{k=1}^{n_{ij}} x_{ijk} \leqslant 1, \text{对所有的 } i,j \\ \sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} a_{ijk}x_{ijk} \leqslant D_{\max}, \text{对所有的 } a_{ijk} \in A \end{array}\right\} \quad (7a) \quad X_r = \{(x_{ijk} \in X_r^c: x_{ijk} \in \{0,1\}), \text{对所有的 } i,j,k\} \quad (7b)$$

需要修改函数以确定未知的因素,考虑 1 个合适的规则集(x_{ijk}),则 $\min_{(c_{ijk}) \in C} \sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} c_{ijk}x_{ijk}$ 就是与规则集(x_{ijk}) 的实现有关的最坏值,所以最大值就是最坏情况的数值。

最佳问题(ROPT) 的解:$\max_{(x_{ijk}) \in X_r} \min_{(c_{ijk}) \in C} \sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} c_{ijk}x_{ijk}$。

组合优化领域建议利用凸松弛函数和随机算法。其基本思想是首先建立一个合适的凸松弛函数,求解函数值,分析解的可能性,抽取其中合适的一个解作为次最佳化解。这里,根据统计检测原理,改进以上处理过程:① 建立凸松弛函数;② 分析解的可能性,为 ROPT 建立凸松弛函数,先建立如下集合:

$$\hat{X}_r^c = \left\{\begin{array}{l} (r,(x_{ijk})): r \in X_r^c, \quad x_{ijk} \in X_r^c \\ r - \sum_{i=1}^{N}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} \leqslant 0, \text{所有的}(c_{ijk}) \in C \end{array}\right\} \quad (8a) \quad \hat{X}_r = \{((r,(x_{ijk})) \in \hat{X}_r^c: x_{ijk} \in \{0,1\}), \text{对所有的 } i,j,k\} \quad (8b)$$

下面的命题为 ROPT 建立一个凸松弛问题:

命题:利用上面的知识,有最佳问题解:

$$\max_{(r,(x_{ijk})) \in \hat{X}_r} r \tag{9}$$

$\hat{X}_r^c$ 是凸集合。

最佳解即为合适的规则集,利用这一个规则集与 RT - IDS 接收到的事件比较,并计算其攻击概率 π_i,以判定攻击情况。

4 实验及分析

实验在局域网上的 3 台计算机之间实现,分别作为攻击终端机、客户机、装有 SNORT2.1 的局域网入侵检测终端机,并装载有图 1 所示的检测引擎,操作系统为 Red-Hat7.2,实验采用 SNORT2.1 的规则集,并与 SNORT 的规则匹配方法——BM 方法相比较,假设攻击的可能性大于 0.5,为有攻击存在,小于 0.5 则不存在攻击,$D_{emt} = 50$ s,p_{ij} 初始值设为 0.01,N 为 2 316(SNORT2.1 事件类型数目),为实验方便,用 scanner3.0 对客户机的 80 端口进行扫描以代替 TCP 的一种攻击,结合式(7)、(8)和(9)在 Matlab6.5 环境下编程,模拟实现。

模拟实验表明,在同等条件下,利用本文方法来检测入侵比 BM 方法要快,如图 4 所示。

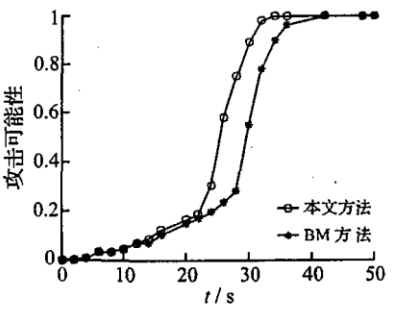

图 4 本文方法与 BM 方法比较

Fig. 4 Comparing of fast method and BM method

参考文献:

[1] Nwanze N, Summerville D H, Skormin V A. Real - Time Identification of Anomalous Packet Payloads for Network Intrusion De-

tection:2005 Proceedings from the Sixth Annual IEEE[C]. [S. l.]:System,Man and Cybernetics(SMC) Information Assurance Workshop IEEE,2005:448 - 449.

[2] Cabrera J B D,Lee W,Praseanth R K. Optimization and Control Problems in Real Time Intrusion Detection:Proceeding of the 41st IEEE Conference on Decision and Control[C]. Las Vegas:IEEE Press,2002:1408 - 1413.

[3] Boyer S. A Fast String Searching Algorithm [J]. Communications of the Association for Computing Machinery, 1997,20(10): 762 - 772.

[4] 韩仲祥. 实时入侵检测的优化问题研究[J]. 计算机工程与应用,2004,29(10):15 - 18.
HAN Zhongxiang. Study on Optimization in Real - Time Intrusion Detection System[J]. Computer Engineering and Applications, 2004,29(10):15 - 18. (in Chinese)

[5] 陈铁柱. Snort 规则集的优化[J]. 海军航空工程学院学报,2005,20(6):664 - 666.
CHEN Tiezhu. Optimization on the Rules Sets of Snort[J]. Journal of Naval Aeronautical Engineering Institute,2005,20(6):664 - 666. (in Chinese)

[6] 韩仲祥. 一种分布式入侵检测系统的实现研究[J]. 空军工程大学学报:自然科学版,2004,5(5):85 - 88.
HAN Zhongxiang. The of a Discributed Intrusion Detection System[J]. Journal of Air Force Engineering University:Natural Science Edition,2004,5(5):85 - 88. (in Chinese)

[7] 韩仲祥. 基于 MIB II 的 IDS 实现研究[J]. 空军工程大学学报:自然科学版,2006,7(4):55 - 59.
HAN Zhongxiang. Implementation of IDS Based on MIB Ⅱ[J]. Journal of Air Force Engineering University:Natural Science Edition,2006,7(4):55 - 59. (in Chinese)

[8] Paxson V B. A system for detecting network intruders in real - time[J]. Computer Network,1999,31:2435 - 2463.

[9] Cabrera B D,Mehra R K. Control and Estimation Methods in Information Assurance - A Tutorial in Intrusion Detection Systems [J]. In Proceedings of the 41st IEEE Conference on Decision and Control,2002,41:1402 - 1407.

[10] Roberto P. Alarm Clustering for Intrusion Detection Systems in Computer Networks[J]. Engineering Applications of Artifical Intellingence,2006,19:429 - 438.

(编辑:田新华,徐楠楠)

A Fast Method for Detecting Intrusion

HAN Zhong - xiang[1,2], SHI Hao - shan[1], DONG Shu - fu[2], YU Zheng[2]

(1. Electric Information College, Northwestern Polytechnical University, Xi'an 710072, China; 2. The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, China)

Abstract: Most papers on detection intrusion problems made by past researchers usually employ rule portfolio to match attack data packages and appear unable to keep the system in optimized detection status when the detection probability is unknown or varying. In this paper, the method of rule portfolio election is improved and the optimal solution via establishing convex relaxation function is gained, then attack package is matched with it. Finally simulation is done with computer, which shows that the effect is better than that of the BM method.

Key words: intrusion detection; rule portfolio; matching; convex relaxation function