恶意网络攻击防范中的规则匹配方法研究

李小兵[1],杨浩[2]

（巴中职业技术学院，四川巴中 636000）

摘要：本文基于网络入侵检测系统的基本需求，分析了当前常用的规则匹配算法原理，在此基础上提出了基于改进的BMH算法的恶意网络攻击的规则匹配检测方法，并且验证了该规则匹配算法在性能和资源占比上相较于传统的规则匹配方法的优越性。此后，在Snort软件的环境下对攻击规则进行检测，将网络入侵的信息和历史数据展开规则匹配，这样就很好的将对字符的逐次计算转化为了对数据规则的匹配，极大的降低了网络检测所需要的时间，同时提升了在网络流量增大过程中检测系统的稳定性。

关键词：恶意网络攻击防范；规则匹配；BMH算法；性能测试

网络通信手段的飞速变化使得互联网在各行各业中都得到了更加深入的应用，随着互联网逐渐步入各行各业的核心部位，其不良影响也逐渐出现。因为当前互联网TCP/IP结构的特性，各种网络攻击手段可以轻易的攻入一些网络系统并造成破坏，继而对网络使用者产生不良的影响。入侵检测技术作为一种预防性的网络安防方法，可以有效地在网络系统遭受内外部网络攻击的过程中予以保护。故而研发行之有效的互联网入侵检测方法，尤其是对入侵检测系统匹配规则进行探究对于提升当前网络环境的安全性有着重要的意义。

1 理论基础

1.1 入侵检测系统

网络入侵检测系统 (Intrusion Detection System，简称IDS)是对任何意图对计算机网络状态进行恶意篡改的行为展开监测的应用系统，其系统形式有纯软件模式也有软硬件结合的模式。入侵检测系统对本地计算机的各个网络接口展开监测，对任何意图对计算机网络状态进行恶意篡改的行为展开应对，比如隔断文件系统访问权限、存留记录、切断通信接口、提醒系统使用人员等凡是来确保计算机系统的安全。其体系结构如图1所示。

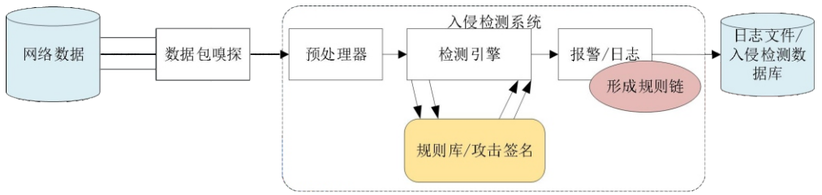

图1 入侵检测系统常见结构

作者简介：1.李小兵，（1979-），男，陕西宝鸡人，本科，四川巴中职业技术学院讲师，计算机与应用专业；2.杨浩，（1988-），男，四川巴中人，硕士，四川巴中职业技术学院讲师，计算机技术专业。

1.2常见网络检测方法

（1）量化分析

量化分析是网络入侵检测系统判断恶意网络访问行为的最常见的方法，主要手段是通过对信息字段来判断其行为的属性。在这种方法之中往往会有大量的算法组合，从基本的聚合分类到高级的神经网络算法都有，其算法处理的结果往往用作系统判断的参照数据之中。

（2）统计分析

对于大部分的异常数据监测系统来说，统计学都是常用的技术手段。在网络入侵检测系统之中，通常使用统计方法来构建系统在正常运行状况下的基本状态，并且对诸如资源占用时间、CPU使用状态、IO调用情况等参数计算出一个标准阈值，并且基于此作为参照值来判定异常状况的出现与否。

在互联网网络检测算法发展之初，最为常用的算法为BM匹配算法，即基于配置好的两个数组进行平移的一种后缀匹配模式，将模式串的比较从右到左，模式串的移动也是从左到右的匹配过程。后来逐渐演进到使用AC算法，即多模式匹配算法。然而无论是BM或是AC规则匹配，均难以在匹配过程中进行字串的跳跃，故而导致匹配过程中效率低下，且系统资源占用率高。

2 恶意网路攻击的规则匹配方法

2.1 BMH匹配算法的理念

对于大部分的异常数据监测系统来说，为了探知系统在正常运行状况下的基本状态，都会使用系统硬件接口来采集正常的工作状态数据以及通信接口中的可疑数据，同时把这些数据整合成字符串送入CPU进行分析。网络入侵监测系统的计算过程往往会调用到CPU的资源，比如BM　算法就是依托在CPU的资源对网络流量数据进行匹配的。在一般的入侵监测方法之中，BM算法往往被用于字符串序列的匹配计算。通常来讲就是按照一个固定模式A来与字符序列X进行校准，进而展开字段匹配。并且在算法结束后将结果记录入数据库，再重新对下一个字符序列进行匹配。使用者可以根据配置匹配模式以及具体的字段偏移量阈值来调整匹配结果。为了方便计算，在算法处理过程中通常将目标字段的信息导入到一个二元二次方程之中进行计算，并根据计算结果来判定匹配状态。匹配过程如图1所示。

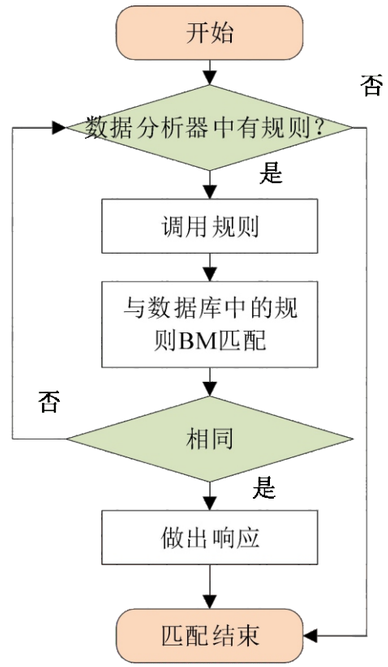

图1 BM匹配规则的流程图

虽然BM算法输出流程较为简单，但是算法本身的时空损耗度较大。故而在入侵检测系统中要规避使用这些函数以降低字符串的对比量，从而提升系统检测效率，为此，也出现了新的BM型算法——BMH（Boyer-MooreHorspool）规则匹配算法。在BMH算法的匹配方式之中，第一步会将将文本和模式两个字符串左侧对齐，然后从模式串最右侧与文本串逐次向左进行匹配，假如文本串字符T_i和模式串末字符P_{m-1}一样，那么就会基于此向左继续对比T_{i-1}和P_{m-2}，直到遍历完所有文本串为止。当文本串中出现和模式串不匹配的字串的情况下，就使用坏字符规则来决定匹配位置平移多少个单位距离，然后持续遍历，直到规则匹配成功或对整个文本串的遍历完成为止。虽然做出了一定的改进，但是BMH只是在传统的BM算法上提升了预处理过程的效率，其最坏情况下的时间复杂度依然为$O(m*n)$。

2.2 改进的BMH规则匹配方法

基于BMH算法本身的缺陷，本文围绕着文本串与模式串右移过程提出改进的BMH规则匹配算法。模式匹配算法在匹配失败时，都会向右移动一段距离再次进行匹配，同时该距离值要尽可能的大而且不能跨过文本串中与模式串匹配的字符。故而本文在考虑右移距离的设定时，主要考虑匹配窗口右侧紧接着的文本串字符是否在模式串中出现，以及当文本串字符在模式串出现时，是否与模式串首字符相同。这种思路可以在规则匹配过程中忽

略不属于模式串的字符，减少与模式串的比较，减少匹配次数，提升匹配效率。其基本思想如下所示：

（1）比较时首先将模式串与文本串左端对齐，然后按照从右至左顺序匹配，如果$P_{m-1}=P_{m-2}$，则继续判断左侧字符是否匹配，依次进行，直到字符$P_0=P_{m-i}$匹配成功或者出现不匹配的字符为止；

（2）发生不匹配情况时，查找当前匹配窗口对应文本串T的后s位字符，直到某个字符$T_{i+m+s-1}$（$0\leqslant i\leqslant n-m-1$，$1\leqslant s\leqslant n-m$）属于模式串P，则转入第三步；

（3）比较文本串的字符$T_{i+m+s-1}$与模式串的首字符$P0$，若二者相同，则模式串右移$s+m+1$个距离，继续比较；反之，则将模式串向右平移$s+m$个单位，然后继续进行匹配；

（4）按照上述步骤匹配，循环进行，直到完成对文本串的搜索。当在文本串字符X处出现不一样的情况时，根据本文拟定的改进BMH算法思路，可以得到跳跃函数如下所示：

$$Skip[x]=\begin{cases} s+m-1, T_{i+m+s-1}=P_0 \&\& P, \neq T_{i+m+t-1} (0<i\leq n-m-1, 1<t\leq s)) \\ s+m, T_{i+m+s-1}\neq P_0 \&\& P\neq T_{i+m+t-1} (0<i\leq n-m-1, 1<t\leq s)) \end{cases}$$

下面以文本串T为“psmpronbmpxohroproper”，模式串P为“proper”作为案例进行本文拟定的改进BMH算法流程的说明。首先第一轮匹配与一般的匹配算法相同，如图2所示。

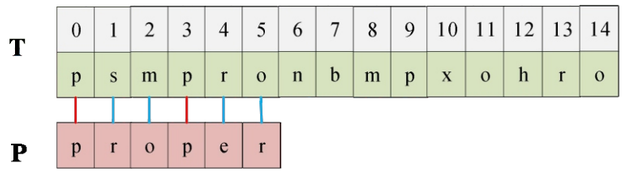

图2 第一轮匹配

可以看出，模式串末字符已与文本串发生不一致，匹配失败，$T_6\neq P_0$，字符“n”、“b”、“m”不属于模式串，因此模式串右移个距离，进行第二轮匹配，如图3所示。

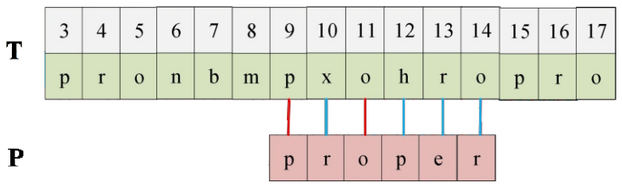

图3 第二轮匹配

在第二轮匹配过程中，模式串自右向左与文本串匹配，第一次比较$T_{14}\neq P_5$时，发生不匹配。当前窗口后一文本串字符“p”属于模式串，且$T_{15}\neq P$，因此模式串右移6(1+6-1)个距离，进入下一轮比较，如图3所示。

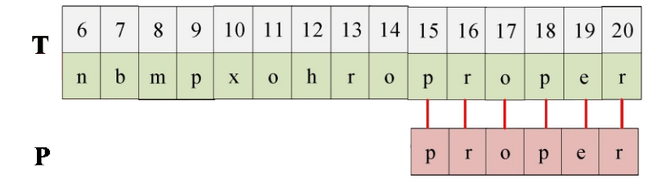

图3 第三轮匹配

通过第三轮查找，匹配成功。按照本文拟定的改进BMH算法流程，共进行了2次右移，8次字符匹配。改进算法相对于BMH算法来说，增大了跳跃距离，减少了字符比较次数，效率得到提高。该算法在最好情况下复杂度为$O(n/m+1)$，虽然最坏情况下时间复杂度与BMH算法相同，但整体上算法性能优于BMH算法。

2.3 改进的BMH规则匹配的性能测试

为了验证性能，本文对改进后的BMH算法进行性能测试，选取对比算法为BM算法和BMH算法。测试硬件环境为：CPU：Intel Core i9-9700K，内存16GB，操作系统为Windows10旗舰版。代码编译环境Python3.7。

测试对象为两个不同的恶意网络攻击流量包数据集“machine”和“area”，均来自于wireshark网络流量范例数据库。测试结果如图4所示。

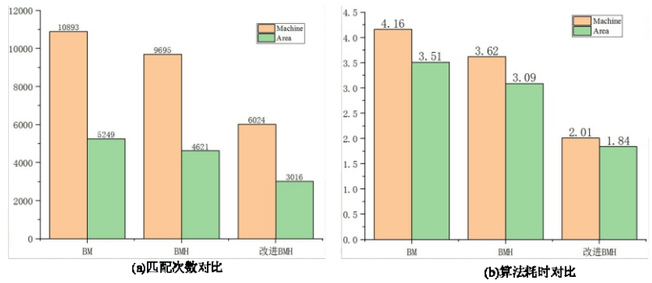

图4 算法对比结果

从图中两组数据的对比我们可以看出，改进算法在匹配次数上均有下降，但两者下降幅度有所不同。通过测试一的数据对比可知，与BM算法相比，本文提出的改进的BMH算法匹配次数下降了 44.7%；与BMHS算法相比，改进的算法匹配次数分别下降了37.9%。通过数据对比我们可以看出：本文提出的改进BMH算法的效率都比BM算法和BMH算法的效率高。

3 基于改进BMH规则匹配算法的恶意网络攻击防范

3.1 恶意网络攻击的规则匹配流程

按照标准的网络入侵检测流程一般分为两步，首先由入侵检测设备进行规则解析流程，检测引擎从规则文件中读取用户编写的多条规则，然后逐条规则进行解析并用对应的语法表示出来，接着在内存中建立规则树，该规则语法树是由链表连接组成的，规则头和若干规则选项构成了规则，其中源和目的IP地址及端口、指向下一规则头的指针和指向归属于它的规则选项的指针存放在规则头中。

第二步为规则匹配流程，入侵检测设备的检测引擎对捕获到的每条网络数据报文同规则树中的每条规则进行匹配，首先匹配该规则树中的规则头，在规则头完全匹配的情况下，再对规则头所指向的规则体进行匹配。如果发现存在某条规则完全匹配该报文，那么就认为入侵检测设备检测到一个入侵行为，然后再根据配置文件或命令进行输出操作，如生成日志或报警，否则就忽略该条报文。无论报文是否匹配上述的规则链，处理完毕后都将循环处理下一条被捕获的数据报文。完整检测流程如图5所示。

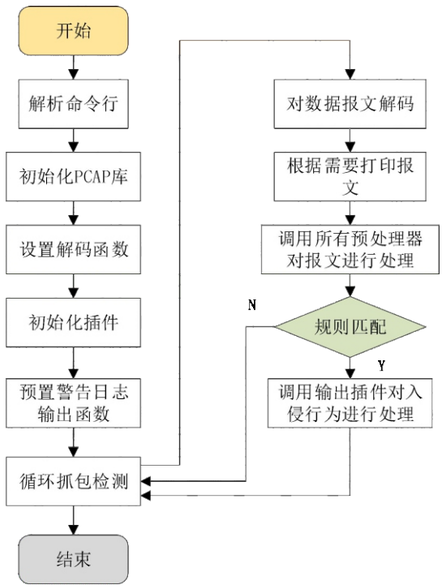

图5 网络入侵检测流程

3.2 性能测试及结果分析

本节将通过实验将本文提出的改进的BMH算法Snort系统中，与BM算法和BMH算法做比较，来检验改进算法对于Snort系统效率的提高。其中的Snort系统是一个Snort是一个以开放源代码形式发行的一个功能强大、跨

平台、轻量级的网络入侵检测系统，可以用于各种网络流量验证以及入侵检测算法和规则匹配方式的实验。本文选取Defcon5中的数据文件33156057.gz解压后的网络数据集来测试改进的BMH算法在恶意网络攻击防范中的性能表现。测试硬件环境为：CPU：Intel Core i9-9700K，内存16GB，操作系统为Windows10旗舰版。

为了使实验结果具有普遍性，实验采用不同数目的规则集来测试，规则数分别为9832、7526、4894、2622以及869，采用不同算法的系统运行消耗时间如图6所示。

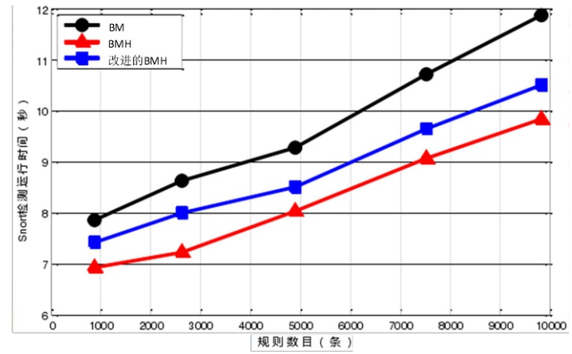

图6 不同算法的系统运行消耗时间

从图中可以发现，当规则数目减少时，无论使用哪种算法的系统检测所消耗的时间都在减少；其次，当规则数目不变时，使用本文提出的改进BMH算法检测的效率比BM和传统BMH算法都要高。

同时，为了测试匹配方法的资源占用性能，本文使用内存占用率作为主要参数来对比改进BMH算法、BM和传统BMH算法的资源占用性能。在进行上述实验的同时，记录的三种不同算法运行时的内存占用率对比图如图7所示。

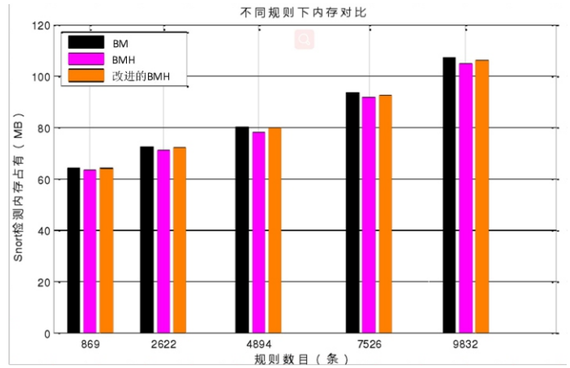

图7 不同规则匹配算法运行时snort的占用内存对比图

从图中可以看出，同样的算法版本，随着规则数的增多，系统占用内存都会有不同的增加；当规则数一定时，使用本文提出的改进BMH算法检测的snort软件内存占用率比BM算法要低，略高于传统BMH算法，但是从整体算法效果的角度来考虑，1-2MB的内存占用不会对整体性能产生大的影响。故而可以认为改进BMH规则匹配算法可以作为恶意网络攻击防范的模式匹配算法的一种选择。

4 总结

因为当前互联网TCP/IP结构的特性，愈加多元化的网络攻击手段可以轻易的攻入一些网络系统并造成破坏，继而对网络使用者的信息、财产安全造成影响。入侵检测技术作为一种预防性的网络安防方法，可以有效地在网络系统遭受内外部网络攻击的过程中予以保护。本文基于网络入侵检测系统的基本需求，分析了当前常用的规则匹配算法原理，在此基础上提出了基于改进的BMH算法的恶意网络攻击的规则匹配检测方法，并且验证了该规则匹配算法在性能和资源占比上相较于传统的规则匹配方法的优越性。此后，在Snort软件的环境下对攻击规则进行检测，将网络入侵的信息和历史数据展开规则匹配，这样就很好的将对字符的逐次计算转化为了对数据规则的匹配，极大的降低了网络检测所需要的时间，同时提升了在网络流量增大过程中检测系统的稳定性。

参考文献：

[1]古平,欧阳源遊.基于混合采样的非平衡数据集分类研究[J].计算机应用研究,2015,32(02): 379-381+418.

[2]曾铖、韩桂华，基于网络的入侵检测系统分析与设计[J]，成都信息工程学院学报，2016.12(21)：86-89.

[3]王昕阳等,浅析串模式匹配算法 KMP 及应用[J], 智能计算机与应用,2017.2(11): 57-60.

[4]K. Fukushima，“Neocognitron:A self-organizing neural-network model for a Mechanism of Pattern recognition unaffected by shift in position，Biol. Cybern.，vol.36，pp.193-202，1980.

[5]C.Neubauer，Shape, position and size invariant visual pattern recognition based on principles of neocognitron and perception in Artificial Neural Networks，1.AlexanderandJ.Taylor，Eds. Amsterdam the Netherlands: North—Holland，vol.2，2012，11: 833-837.

[6]Van Ooyen and B. Nienhuis，Pattern Recognition in the Neocognitron Is Improved—by Neuronal Adaption，Biological Cyberneties70,pp.47-53(1993).

[7]Fukushima K. A hierarchical neural network capable of visual pattern recognition . Neural Networks，1989:2:413-420

[8]Y.Bengio.YLeCun，and D.Henderson，“Globally Trained Handwritten Word Recognizer using Spatial Representation, Convolutional Neural Networks and Hidden Markov Models in Advances in Neural Information Processing Systems，Jack D.Cowan，Gerald Tesauro，and Joshua AlsPector，Eds.1994，vol.6，PP.937-944，Morgan Kaufmann Publishers，Inc.

[9]崔璐,张鹏,车进.基于深度神经网络的遥感图像分类算法综述[J].计算机科学,2018,45(S1):50-53.

[10]谢红,刘人杰,陈纯锴.基于误用检测与异常行为检测的整合模型[J].重庆邮电大学学报(自然科学版),2015, 24(01):73-77.

[11]蒋玉娇,王晓丹,王文军,毕凯.一种基于PCA和ReliefF的特征选择方法[J].计算机工程与应用,2016,46(26):170-172.

第2卷第1期 教育与研究 Volume.2.No.1
2020年01月 Education and Research January .2020