改进的空间协议识别算法

郑天明[1,3]，王韬[1]，郭世泽[2]，李华[1]，赵新杰[1]

(1. 军械工程学院 计算机工程系，河北 石家庄 050003；2. 北方电子设备研究所，北京 100083；
3. 中国人民解放军第六九零九工厂，江苏 苏州 215300)

摘 要：提出了一种适合空间协议识别的改进 BM 算法。首先给出了一种基于比特距离的空间数据预处理算法，增大字符集数量，并通过引入小数跳进机制，提高 BM 算法协议分组头匹配效率；然后应用正则表达式进行协议识别，利用层次关系法提高多层空间协议识别效率；最后对提出的算法进行了复杂度分析和实验验证。结果表明：对于识别模式串长度为 m 的单层协议，算法时间复杂度可降低到 BM 算法的$(1+m/4)/m$，对多层协议识别效率可提高 2.5 倍；同时，与 BM 算法相比，提出的算法可有效解决模式串长度不足与存在大量不确定数据的问题，在数据量较大情况下具有更高的识别效率，且所形成的分组可有效抑制正则表达式 DFA 匹配引擎状态膨胀。

关键词：空间协议；协议识别；BM 算法；比特距离；小数跳进；正则表达式；层次关联

中图分类号：TP393.08 **文献标识码：**A **文章编号：**1000-436X(2012)05-0183-08

Improved space protocol identification algorithm

ZHENG Tian-ming[1,3], WANG Tao[1], GUO Shi-ze[2], LI Hua[1], ZHAO Xin-jie[1]

(1. Department of Computer Engineering, Ordnance Engineering College, Shijiazhuang 050003, China;
2. The Institute of North Electronic Equipment, Beijing 100083, China; 3. 6906 Military Enterprises, Suzhou 215300, China)

Abstract: An improved boyer-moore (BM) algorithm for space protocol identification was proposed. First, a space data preprocessing technique based on bit distance was used to increase the size of data set. A decimal jumping technique was introduced to enhance the matching efficiency for the col head part of the BM algorithm. Then, the regular expression method was applied to proceed the protocol identification and the hierarchy associated analysis technique was proposed to improve the efficiency of multi-level space protocol identification. Finally, the complexity of the proposed algorithm was analyzed and verified with concrete experiments. The results show that with the proposed algorithm, as to pattern string length m, the time complexity of the single layer protocol identification can be reduced to $(1+m/4)/m$ of the BM algorithm. The efficiency of the protocol identification for multi-level layers can be improved about 2.5 times. Meanwhile, comparing with BM algorithm, the proposed algorithm can solve the problem of pattern string shortage and large wildcards in the space data. The identification iciency in case of huge data packages can be improved and the new formed data block can restrain the state expansion for the DFA matching engine in regular expression.

Key words: space protocol; protocol identification; BM algorithm; bit distance; decimal jumping; regular expression; hierarchy association

1 引言

当前，各种卫星网络已被广泛应用到民用、军事领域的各个角落，对卫星网络安全性进行分析十分重要，其中协议识别是卫星网络安全性评估的研究热点[1]。空间协议是卫星网络的骨架和

收稿日期：2011-04-25；**修回日期：**2011-12-13

神经，是维系网络正常通信的纽带，各层协议在空间信息结构中占有核心地位，通过对卫星网络协议进行识别与分析，可确定各层协议类型，获取网络拓扑结构、路由算法、IP 地址分布等信息，探测协议漏洞或获取高层信息，在此基础上进行更高层次的安全评估。

BM 算法(boyer-moore algorithm)[2]是最为经典的协议识别算法，算法主要根据常用协议模式串规则，通过对待识别数据特征进行分析，从中查找模式串并确定其所在位置，在此基础上进行协议识别。BM 算法主要适用于模式串较长、字符集较大的协议识别。然而，在卫星网络中，由于空间数据具有特征位长度短、数据字符集小特点（如典型的 SCPS-NP[3] 协议特征仅为 3bit），传统的 BM 算法很难从中获取协议特征，存在执行效率和识别准确度较低问题。

为此，本文提出了一种改进的 BM 算法。提出的算法给出了一种计算空间数据中相同比特距离的数据预处理方法，降低字符串长度；并引入小数跳进机制，提高了 BM 算法分组头匹配效率；在此基础上应用正则表达式方法进行规则判断，然后给出了一种基于层次关系法的多层协议识别方法，提高了多层网络协议的识别效率；最后以应用广泛的 SCPS(space communication protocol specification)协议识别为例，应用提出的算法进行了实验。

2 空间协议特征分析

空间数据在传输过程中具有传输时延长、信道误码率高、非对称、适合同地面终端进行协议转换等特点[4]。为适应这种特点，CCSDS（consultative committee for space data systems）委员会制订了空间传输协议栈 SCPS。SCPS 协议栈结构如图 1 所示，其自下而上包括：物理层、数据链路层、网络层、运输层和应用层，以及介于网络层与传输层之间的安全保护机制等多层协议，其中每层又包括若干个可供组合的协议[5~9]。

经分析，空间数据结构特征可归纳如下。

1) 协议结构层次化，对空间数据进行分析需要首先解调出比特流数据，并逐层开展协议分析；

2) 协议特征位较短，数据中模式串不易被识别；协议长度可变，典型模式串识别方法不具优势；

3) 链路层协议兼容各种上层协议，空间数据各层具有紧密相关性，分析网络层协议可获取传输层协议类型，而分析链路层协议却无法降低网络层协议分析难度；

4) 数据量大，同一时段内截获的数据类型多样，针对单一协议的识别效果不佳；业务需求大、种类多，卫星通信技术以及协议更新速度快。

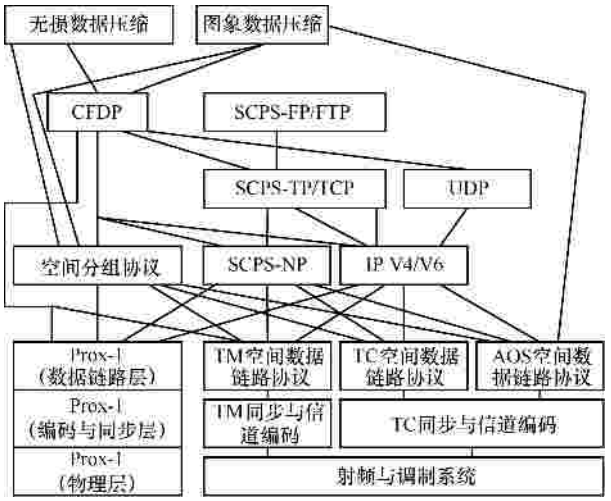

图 1 SCPS 协议栈结构

综上，空间协议识别应根据协议栈结构进行层次分析，分析可选取协议特征位作为识别基准点，参照协议结构拆分数据进行分析，并在单层协议识别基础上对多层协议开展分析。

3 基于 BM 算法的空间协议识别分析

将截获数据存储到一个字符数组，将协议特征转化为特征模式串，可将协议识别归结为模式串匹配[10,11]问题，通过对协议特征和数据进行逐位匹配，可分析出协议类型。BM 算法[2]是最具有代表性的模式串匹配算法，识别速度较快，执行时间复杂度为 $O(n+sm)$[12]，s 为模式串在数据中出现频率相关变量，m、n 分别为模式串、目标串长度。如果模式串频繁出现，BM 算法在查找阶段时间复杂度可以表示为 $O(mn)$，最好时间复杂度是 $O(n/m)$。

图 2 为应用 BM 算法对基于比特、英文字母字符集的匹配对比。易见，字符集{0，1}所生成的数据，每位出现 0、1 的概率为 50%。应用 BM 算法进行模式串匹配时，在模式串跳进过程中遇到干扰项概率较高、跳进距离较短、误判率较高、任意 4bit 出现模式串概率为 6.25%。相比较而言，由字符集合{a~z}所生成的模式串，每项出现概率仅为 3.8%，跳进过程中遇到干扰项概率较小，跳进距离也相对较大，误判率较低，任意 4bit 出现模式串的概率约为 0.000 207 9%。

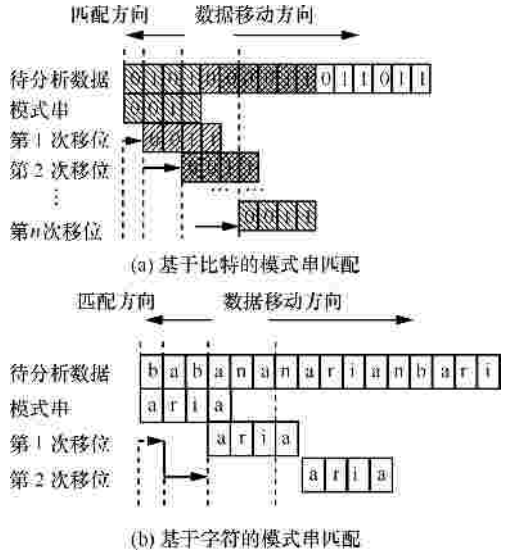

图 2　应用 BM 算法对不同字符集的模式串匹配对比

为提高模式串匹配效率，现有研究大都着眼于改进 BM 算法的模式串跳进方式，研究者相继提出了 QS[13]、RSPS[14]、Wu Manber 和 SBOM[15](set backward Oracle matching)等算法。以 SBOM 算法为例，算法的预处理过程是根据所有模式串构造匹配使用的 Oracle 结构，以识别匹配窗口内最长字串，其算法复杂度为 $O\left(n\left[\frac{pm+(1-p)^{\log_a mr}}{m-\log_a mr}\right]\right)$，$p$ 为命中密度，a 为字符集大小。

通过对上述几种算法进行对比分析不难发现：改进的 BM 算法在命中密度低（模式串较长、字符集较大[16]）情况下识别效率极高[17]，应用到汉字搜索等领域，可有效提高匹配效率。但随命中密度的增大，上述算法的识别效率将急剧降低（在处理 0、1 数据时，字符集较小所以 p 极大，SBOM 的复杂度约为 $O(mn)$，接近于 BM 算法）。

总的来说，应用 BM 算法及其改进算法进行空间协议识别存在以下不足：模式串长度较短，字符集元素有限（只有 0、1 组成），模式串跳进和识别效率低下；即使增大模式串长度，将不可避免增多模式串中通配符数量，降低 BM 算法识别效率。

4　改进的空间协议识别算法

4.1　主要思想

为解决应用 BM 算法进行空间协议识别问题，我们提出了一种改进算法。通过对空间数据进行预处理，增大模式串字符集；然后改进模式串跳进规则，提高数据识别效率，形成数据分组；通过分组标识使用特定 DFA 匹配引擎，抑制 DFA 状态数膨胀，实现单层协议识别；在此基础上结合多层协议间关联关系，缩小多层协议识别范围，并调整匹配顺序，提高多层协议识别效率和准确度。应用改进算法进行空间数据分析过程如图 3 所示。

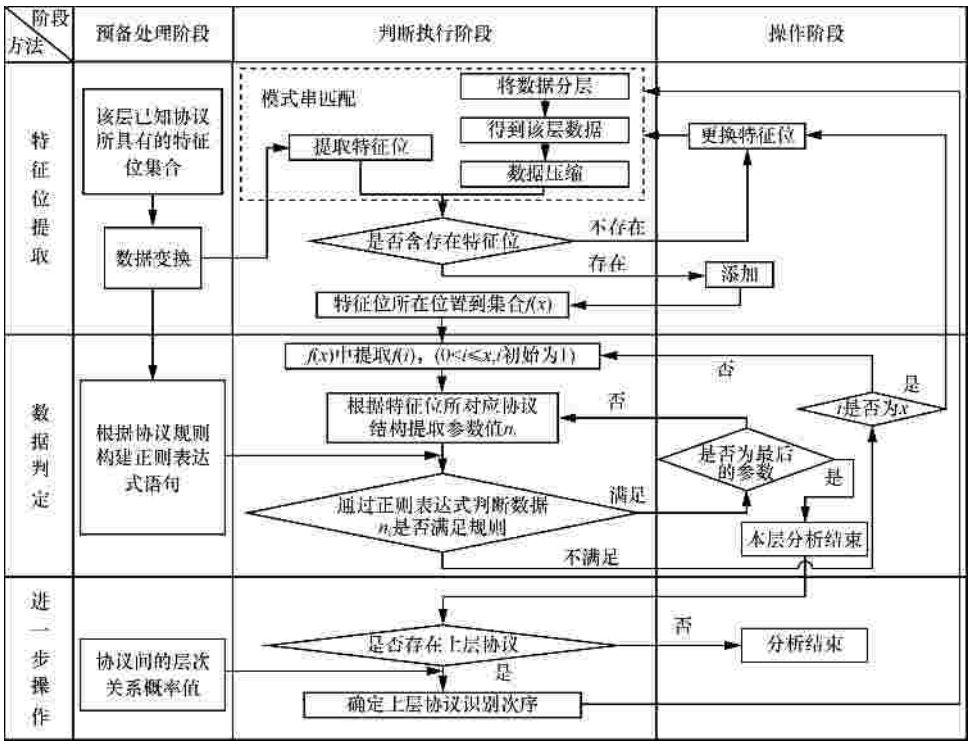

图 3　基于 CRB 算法的数据分析过程

4.2 基于比特距离的空间数据预处理

在协议识别前，常需先将数据转换为比特流，基于比特流进行协议分析。由于信道误码等原因可造成部分比特数据误码，主要表现为 2 个方面：首先，数据内容倒置，在这种情况下，如果直接将比特数据按照十六进制方式进行存储，将 1bit 误码存储后造成 1byte 误码，即误码放大，因此，基于软件系统的协议分析，比特流中的每 1bit 通常按照字符（字节）进行存储，文中称为字节流；其次，数据首字符难于确定，通常的分析方式无法判断数据起始位置，将造成大量错误分析，将 BM 算法同正则表达式方法结合起来可有效解决该问题。

传统 BM 算法基于字符进行比对跳进，虽然能够有效解决起始位置难于确定问题，但由于空间比特流转换成的“0”和“1”字符较多，且由“0”和“1”组成的字符集数目较小，导致协议分析效率低下。为此，文中给出了一种计算相邻比特距离的空间数据预处理算法。具体算法如下：选择字符“1”为特征位，并记录连续 2 个字符“1”之间的字符“0”的个数 N，并将 N 以字节形式保存。该预处理算法可增大模式串字符集数量、提高字符跳进距离、降低数据分析时间复杂度。

图 4 为应用 BM 算法对原始空间数据和预处理后数据进行识别的过程。对于 16byte 存储的“0101000011011011”字节流，经预处理后可转换为 8byte“11401010”。

根据第 3 节，模式串频繁出现时，查找阶段时间复杂度 $O(mn)$同目标串、模式串长度成正比。文中数据预处理算法可减小模式串长度 m 和目标串长度 n，降低模式串匹配复杂度。数据查找阶段复杂度可降低到 $O(mn/4)$，预处理时间复杂度为 $O(n)$（在长度为 n 数据中查找“1”所耗时间），整体时间复杂度为 $O(n+mn/4)$，显然 $O(n+mn/4)<O(mn)$，提出算法可加速数据分析效率。

由于 BM 算法效率同模式串字符集数量成反比，在空间数据识别过程中，传统模式串字符集仅为“0”、“1”2 种，单个跳进距离最多为 1，跳进次数等同于数据串长度。应用本节预处理算法，模式串字符集可扩展至“0 到 255”256 种。当数据串字符为 0 时，跳进距离为 1，等同于 BM 算法；而当数据串字符 A 大于 0 时，单次跳进距离为 A+1。由图 4 可知，应用 BM 算法模式串匹配的平均跳进距离为 1，跳进次数为 16，而应用本节预处理算法，平均跳进距离为 2，跳进次数为 8。易见提出算法增大了 BM 算法的跳进距离，减小了跳进次数，进而提高了识别效率。

4.3 基于小数跳进的模式串查找

在 4.2 节对空间数据预处理的基础上，本节给出一种基于小数跳进机制的数据查找算法，可有效改进 BM 算法后缀转移和不良字符转移机制。

小数跳进机制基本思想如下：依据变换后数据内在意义，即每比特代表变换前数据 1 的前后间隔比特数，将目标串自起始位置起往右一个模式串长度的字符与模式串最小值进行比较。如果模式串长度的字符小于模式串最小值，则应从目标串的下一比特开始新一轮匹配，相当于把模式串在目标串中向右滑过，即一次跳过“滑过距离”个字符，从而实现数据匹配的快速跳进。

算法 1 给出了基于小数跳进机制的数据查找算法，同 BM 算法一样，文中小数跳进机制也采用从右向左逆向检索匹配窗口。

算法 1 基于小数跳进机制的数据查找算法

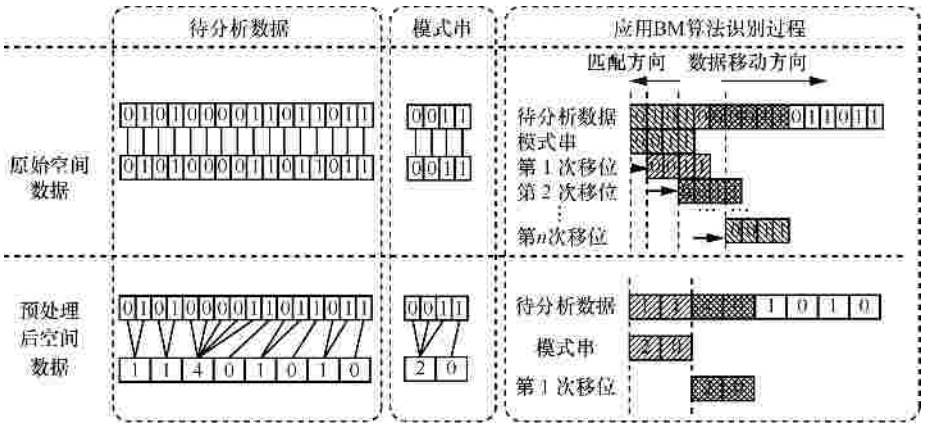

图 4 算法对比

输入：预处理后数据串 T，模式串 P

输出：模式串匹配位置集合 S

$L(T)$？　变换后数据串 T 长度

$L(P)$？　模式串 P 长度

min(P)？　模式串中最小值

max(P)？模式串中最大值

L？　模式串匹配起始位置

1) 如果字符 $T_{L+L(P)}$小于 $min(P)$则可根据小数跳进机制

2)　模式串 P 左移 $L(P)$位

3) 反之

4)　进行循环比较，如果 $L(P)$大于 2

5)　在 T 中查找下一个值等于 $P_{L(P)-1}$ 的字符位置 $T_{k+L(P)-1}$

6)　模式串右移 k 位

7)　如果 $L(P)$小于等于 2

8)　则判断 $P_{L(P)}$是否等于 0，如果等于 0

9)　如果 $T_{L+L(P)+1}$ 大于等于 P_1

10)　右移 $L+L(P)+1$ 位

11)　如果 $T_{L+L(P)+1}$ 小于 P_1

12)　搜索下一个 $T_{L+L(P)+w}>=P_1$的位置 w

13)　如果 w 存在

14)　模式串 P 向右移动 w 位

15)　如果 w 不存在

16)　匹配失败，返回

17)　相对于第 8 步，如果 $P_{L(P)}$不等于 0

18)　在 T 中查找下一个值等于 $P_{L(P)}$的字符位置 $T_{O+L(P)}$

19)　模式串 P 右移 O 位

20)　查询下一个 i，使得 T_i 等于 $P_{L(P)-1}$

21)　直接将模式串 P，移动 i-$L(P)+1$ 位

小数跳进机制正确性证明如下。

命题 1 已知数据 T 长度 $L(T)$和模式串 P 长度 $L(P)$，在数据匹配过程中，模式串最小值 min(P)，若 $T_x<\min(P)$，数据 T 中下一个大于 min(P)位置为 T_s，则 T 从 $x$bit 开始的子串 T_{x-1+i}（$0<i<s-x$）与模式串 P 不匹配，即$\neg(T_{x-1+i+k}=P_k)$。

证明　假设命题不成立，则 $T_{x-1+i+k}=P_k$，$0<=k<L(P)$，即 $T_{x+i+k-2}$ 还原数据所得值最后 1bit 应为“1”，$T_{x-1+i+k}$ 最后 1bit 也为“1”，即 2 个“1”之间“0”的个数应为 $T_{x-1+i+k}$ 个。

因为 $T_{x-1+i+k}=P_k$，所以根据数据变换规则可知 P_{k-1} 之后下一个“1”所在位置应距离 P_{k-1} 为 $T_{x-1+i+k}$ 位。同样对模式串 P 进行逆变换后，可知 P_{k-1} 后下一个“1”所在位置应该距 P_{k-1} 为 $P_k$bit，与 $T_x<min(P)$ 已知条件相矛盾，故假设不成立，命题结论成立。

4.4　基于正则表达式的单层空间协议识别

由第 2 节可知，空间数据串中查找到的标识比特并不一定为模式串特征比特。通过较短特征比特匹配方式查找数据可形成集合 S_N，而真正特征比特集合为 S_T，S_N 中元素数量应大于等于 S_T 中元素数量，且 $S_T\in S_N$ 式恒成立，所以需要通过规范限制方式减少 S_N 数量，即在 S_N 中查找真正特征位集合 S_T，文中基于正则表达式[18] DFA 方法实现。

DFA 方法却受限于状态数膨胀[19,20]问题，解决方法一般通过将规则集分成 N 组来实现[21]，但这样会将处理速度降为原来的 $1/N$，增加存储器所需带宽，限制匹配速度提高。而应用文中改进算法，在对数据集进行整合的前提下，依据当前所处理子集，将相互之间导致状态膨胀的正则表达式分在不同分组，在不影响匹配速度前提下可有效抑制 DFA 状态数膨胀。通过对空间协议结构进行分析，将空间数据类型归纳为 4 种，并用正则表达式进行规则判定，具体类型及分析方式如表 1 所示。

表 1　数据类型及分析方式

类型	代表字段	分析方式
确定标识比特	协议标识比特等	要求数据码流完全匹配规则的每个分量可能作为分类前提或为分类提供依据，常需确定标识比特所在位置和标识特征属性
数据区间	数据长度等	一个多比特集合，取值是一个范围区间，对其判断应该采用范围控制方式
离散值	上层协议等	根据协议可能结果，列出可能的数据项
任意值	控制比特等	此类数据常对分析带来较大困难，分析过程中应尽量跳过，并减少其出现概率，不对确定数据、数据区间、离散值产生影响

4.5　基于层次关联关系的协议选择

单层协议识别完成后，可分析出单层使用协议、数据长度、源目的地址等参数。如果要进一步分析下层协议，一般做法是通过分析本层数据长度，跳转到下层数据起始位置，逐层进行协议识别。设当前识别的底层协议包含 m 个协议，上层包含 n 个协议，在等概率的协议分析条件下，应用一般的协议分析模型，一共有 mn 个协议组合，需进行 mn 次协议分析判定。该方法忽略了本层协议往往规定、限制上层协议这一规则，导致识别效率低下。

事实上，协议上下层协议之间存在一定的关联

关系，论文通过分析这种关联关系，并计算上层协议的出现概率，将上层协议分析次序进行了合理调整，减小了不必要的判定，进而降低了协议识别复杂度。如图 1 所示，如果在网络层使用“空间分组协议”，那么在传输层就不会使用 SCPS-TP、TCP 或 UDP 协议，进行传输层协议识别时就无需选取 SCPS-TP 等协议特征位，可在一定程度上提高识别效率。

基于该思想，给出了一种基于层次关系的协议选择机制，根据已识别协议所兼容上层协议，缩小上层协议识别范围，提高识别效率。算法在实现自底向上分层识别后，统计并记录该层协议同上层协议关系，在大量数据样本基础上，对上层协议出现概率进行归纳，后续识别可根据归纳概率确定多层协议识别顺序，提高识别效率。基于层次关系的协议选择机制算法如算法 2 所示。

算法 2 基于层次关联的协议选择机制

输入：变换后数据串 T,某层协议 $C(X_i)$

输出：整条数据所使用的协议

$C(X_{i+1})$、$C(X_{i+2})$？ 上层所使用的协议

$T(X_{i+1})$、$T(X_{i+2})$？ 上层所有可选协议

$P(C(X_{i+1})|C(X_i))$、$P(C(X_{i+2})|C(X_i))$? 协议出现概率

r_1、r_2？ 权重值

L？ 模式串匹配起始位置

1) 如果现有的协议特征库中有 $U(X_{i+1})$协议符合规则 $C(X_i).NextProtocol$

2) 则把该协议视为已知协议

3) 本算法将提供基于临近层协议权重比较的方式，将数据进行分类

4) 之后重新设置 $P(C(X_{i-1})|C(X_i))$、$P(C(X_{i+1})|C(X_i))$、$P(C(X_{i+2})|C(X_i))$的出现权重值 r_1、r_2、r_3。

5) 给出上层协议类型 $C(X_{i+1}).Pattern$

6) 回到第 1)步，进行下一轮比较

7) 如果现有的协议特征库中没有协议符合规则

8) 本算法将对临近层协议的置信区间进行比较，将数据进行分类

9) 出现概率分别为 $P(C(X_{i-1})|C(X_i))$、$P(C(X_{i+1})|C(X_i))$、$P(C(X_{i+2})|C(X_i))$，而权重值分别为 r_1、r_2、r_3。如果对于未知协议 $C_1(X_i)$，同该未知协议集合中临近协议相对比

10) 如果待分析的数据段？$r_tP(C(X_k)|\ C(X_i))$ {其中，$k\in(i-1、i+1、i+2)$, $t\in(1、2、3)$}较大

11) 更新临近层协议的置信区间

12) 返回该未知协议 $C(Xk).Protocol$ 的所属类别

13) 将该类别进行归纳

14) 更新协议类别库

15) 返回到第 7)步，循环进行之后的识别

16) 反之

17) 判定该数据使用一种新协议

5 实验结果与分析

在 Windows 2000 环境下，基于 Visual C++6.0 语言，设计实现了空间数据分析系统，可生成符合 SCPS 空间协议标准的数据流，并存入数据库。此外，针对 SCPS 协议族，对几种当前常用的网络层、传输层、应用层协议特征进行了归纳，建立了相应的协议特征规则的正则表达式，并应用文中改进算法进行了大量的数据分析实验。

5.1 单层空间协议识别实验

实验中，首先读取待分析数据，应用 4.2 节算法进行数据预处理，然后应用 4.3 节小数跳进机制算法实现协议模式串的快速查找，实现单层协议识别。

图 5 为应用 4.3 节方法同传统 BM 算法识别空间数据的比较，可以看出，提出算法识别效率可提高至 BM 算法的 2 倍。

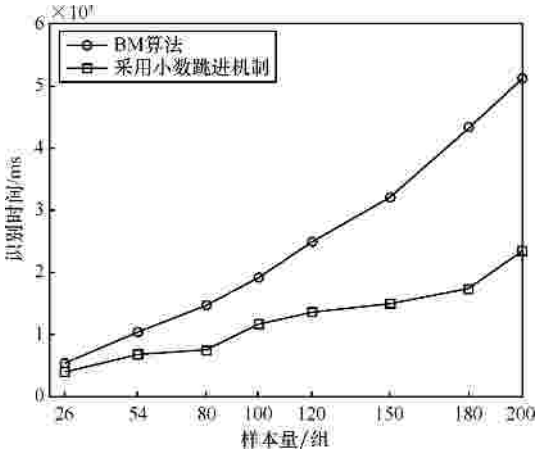

图 5 空间数据单层协议识别比较

由上可知，提出算法在规则复杂度较高，数据量较大的情况下，通过变换空间数据的方法有效减小了 n、m，增大了跳进距离，解析单条数据复杂度为 $O(4y(mn/(4+a\ )+n)$ （y 表示经筛选后某一层可能的协议个数，$y<P$，a 与 m 中连续“1”出现的频率有关）。

$$\left.\begin{aligned} &O\left(4y\left(mn\right)/\left(4+a\right)+n\right)<O\left(y\left(mn\right)+n\right)\\ &O\left(y\left(mn\right)+n\right)<O\left(P\left(mn\right)+n\right)\\ &O\left(P\left(mn\right)+n\right)\le O\left(3Pn+n\right)\end{aligned}\right\} \quad (1)$$

s.t.　$a \ge 0$，$y < P$，$m \le 3$

由式(1)可以得出：

$$O(4yO(mn/(4+a\ )+n)$$
$$\le O(3Pn+n) < O(12Pn) < O(4P(nm)) \quad (2)$$

式(2)成立。易见，对于单层协议识别，文中提出算法时间复杂度可降低到 BM **算法的**$(1+m/4)/m$**。**

5.2　多层空间协议识别实验

在单层空间协议识别基础上，又应用 4.5 **节算法对多层协议识别进行了实验。**

以 SCPS **协议族为例，传输层协议包括** SCPS-TP、TCP **和** UDP 3 **种典型协议，应用层包括** CFDP、SCPS-FP **和** FTP 3 **种典型协议。图** 6 **给出了卫星网络协议识别中，传输层和应用层协议的架构及出现概率示例。按照单层的协议分析方法，识别由“**SCPS-TP、CFDP**”组成的一组数据，在最差的情况下，需要执行** 3×3 **次协议分析。而应用文中算法，通过分析协议结构和出现概率，首先计算出多层协议分析组合的优先级。**

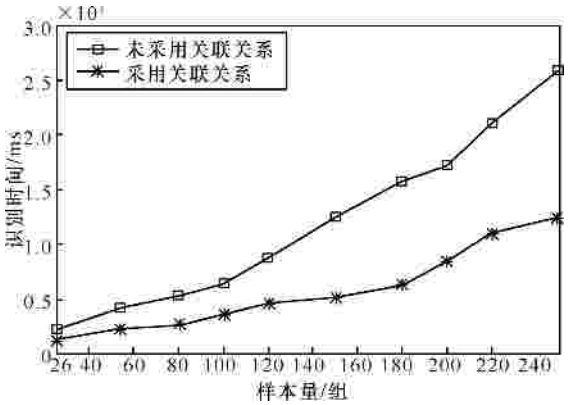

图 6　**空间数据多层协议识别比较**

从表 2 **可知，应用本文方法，最优仅需** 1 **次、最差仅需** 7 **次协议分析，复杂度要低于传统的** 9 **次分析。首先应用文中方法按照表** 2 **的优先级顺序进行了多次实验，并在随机的协议组合顺序下进行了识别实验，二者的结果比较如图** 6 **所示。**

可以看出提出算法根据相邻层协议出现概率进行识别，可将多层空间协议识别效率提高 2.5 **倍。**

表 2　　**协议组合优先级**

序号	传输层协议	应用层协议	出现概率	分析优先级
1	SCPS-TP	CFDP	0.508 4	1
2	SCPS-TP	SCPS-FP	0.270 6	2
3	SCPS-TP	FTP	0.041	5
4	TCP	CFDP	0.027 6	6
5	TCP	SCPS-FP	0.085 2	3
6	TCP	FTP	0.007 2	7
7	UDP	CFDP	0.06	4
8	UDP	SCPS-FP	0	—
9	UDP	FTP	0	—

对于多层协议，设传输层协议为 L_{T}**、应用层协议为** L_{A}**，概率系数为** b **，提出算法的时间复杂度**

$$o\left(b\,L_{\mathrm{T}}L_{\mathrm{A}}\right)\le o\left(L_{\mathrm{T}}L_{\mathrm{A}}\right) \quad (3)$$

6　结束语

空间协议分析是一个倍受关注的研究领域，本文在对 SCPS **协议族进行研究基础上，提出了一种改进的空间协议算法。提出算法在空间数据预处理、数据查找跳进、多层协议识别方法进行了改进，在一定程度上解决了特征位难于区分和查找问题，提高了协议识别效率。**

参考文献：

[1] **李雄伟. 网络对抗系统及其关键技术研究**[D]. **北京: 北京邮电大学**,2003.
LI X W. Research on Network Operation System and Key Technologies[D]. Beijing: Beijing University of Posts and Telecommunications, 2003.

[2] BOYER R S, MOORE J S. A fast string searching algorithm[J]. Communications ACM, 1977, 20(10):762-772.

[3] CCSDS 311.0-M-1 Reference Architecture for Space Data Systems[S]. Washington, DC, USA, 2008.

[4] CCSDS 130.0-G-2 Overview of Space Communications Protocols, Report Concerning Space Data System Standards[S]. Washington, DC, USA, 2007.

[5] CCSDS 211.0-B-4 Proximity-1 Space Link Protocol—Data Link Layer, Recommendation for Space Data System Standards[S]. Washington, DC, USA, 2006.

[6] CCSDS 713.0-B-1. Space Communications Protocol Specification (SCPS)-Network Protocol (SCPS-NP)[S]. Washington, DC, USA, 1999.

[7] CCSDS 714. O-B-2. Space Communications Protocol Specification

(SCPS)-Transport Protocol (SCPS-TP)[S]. Washington, DC, USA, 2006.

[8] CCSDS 717.0-B-1 Space Communications Protocol Specification (SCPS)-File Protocol (SCPS-FP)[S]. Newport Beach, California, USA, 1999.

[9] **许枫, 尤政**. CCSDS **空间通信协议及其与互联网通信协议的比较**[J] . **中国航天**, 2007, (5): 26 - 29.
XU F, YOU Z. CCSDS space communications protocol and its comparison with Internet protocols[J] . Aerospace China, 2007,(5): 26-29.

[10] HOOKE J .Evolutionary Paths to Internationally-standardized Space Internetworking[R] . AIAA , 2008. 2 - 9.

[11] **谭建龙**. **串匹配算法及其在网络内容分析中的应用**[D]. **北京:中国科学院研究生院**, 2003.
TAN J L. String Matching Algorithm and Application of Network Content Analysis[D]. Beijing: Graduate School of Chinese Academy of Sciences, 2003.

[12] RICHARD O D, PETER E H, DAVID G S. Pattern Classification[M]. America: Library of Congress, 2000.

[13] SUNDAY D M. A very fast substring search algorithm[J]. Communications ACM , 1990, 33(8): 132-142.

[14] KAGHAZIAN L, MCLEOD D, SADRI R. Scalable complex pattern search in sequential data[A]. Proceedings of the 17th ACM Conference on Information and Knowledge Management[C]. Napa valley California, 2008. 1467-1468.

[15] **范洪博, 姚念民**. **一种高速精确单模式串匹配算法**[J]. **计算机研究与发展**, 2009, 46(8): 1341-1348.
FAN H B, YAO X M. A fast and exact single pattern matching algorithm[J]. Journal of Computer Research and Development, 2009, 46(8): 1341-1348.

[16] CROCHEMORE M, ALLAUZEN C, RAFFINOT M. Factor oracle:a new structure for pattern matching[A]. Proceedings of the 26th Conference on Current Trends in Theory and Practice of Informatics[C]. London, UK, 1999.291-306.

[17] **刘萍, 刘燕兵, 郭莉等**. **串匹配算法中模式串与文本之间关系的研究**[J]. **软件学报**, 2010, 21(7):1503-1514.
LIU P, LIU Y B, GUO L, *et al*. Research on relationship between patterns and text in string matching algorithms[J]. Journal of Software, 2010, 21(7):1503-1514.

[18] **范慧萍**. **基于正则表达式的协议识别研究与实现**[D]. **长沙:国防科学技术大学研究生院**, 2007.
FAN H P. Research and Realization of High Speed Protocol Identification Based on Regular Expression[D]. Changsha: Graduate School of National University of Defense Technology, 2007.

[19] **陈曙晖, 苏金树, 范慧萍等. 一种基于深度报文检测的** FSM **状态表压缩技术**[J]. **计算机研究与发展**, 2008, 45(8): 1299-1306.
CHEN S H, SU J S, FAN H P, *et al*. An FSM state table compressing method based on deep packet inspection[J]. Journal of Computer Research and Development, 2008, 45(8): 1299-1306.

[20] **杨毅夫, 刘燕兵, 刘萍等**. **正则表达式的** DFA **压缩算法**[J]. **通信学报**, 2009, 30(10A): 36-41.
YANG Y F, LIU Y B, LIU P, *et al*. Effective algorithm of compressing regular expressions' DFA[J]. Journal on Communications, 2009, 30(10A): 36-41.

[21] FANG Y, CHEN Z F, DIAO Y L, *et al*, Fast and memory-efficient regular expression matching for deep packet inspection[A]. ANCS'06: Proceedings of the 2006 ACM/IEEE Symposium on Architecture for Networking and Communications Systems[C]. New York, NY, USA, 2006.93-102.

作者简介：

郑天明（1985-），男，辽宁锦州人，军械工程学院硕士生，主要研究方向为卫星网络安全。

王韬（1964-），男，河北石家庄人，博士，军械工程学院教授、博士生导师，主要研究方向为信息安全和密码学。

郭世泽（1969-），男，河北石家庄人，博士，北方电子设备研究所研究员，主要研究方向为信息安全和密码学。

李华（1981-），男，河北石家庄人，军械工程学院博士生，主要研究方向为卫星网络安全。

赵新杰（1986-），男，河南开封人，军械工程学院博士生，主要研究方向为网络安全和密码分析学。