使用贝叶斯学习算法分类网络流量

邱 密[1],阳爱民[1,2],刘永定[1],何震凯[1]
QIU Mi[1],YANG Ai-min[1,2],LIU Yong-ding[1],HE Zhen-kai[1]

1.湖南工业大学 计算机与通信学院,湖南 株洲 412008
2.广东外语外贸大学 信息科学技术学院,广州 510006
1.School of Computer and Communication,Hunan University of Technology,Zhuzhou,Hunan 412008,China
2.Information Science and Technology College,Guangdong University of Foreign Studies,Guangzhou 510006,China
E-mail:qiumi229@163.com

QIU Mi,YANG Ai-min,LIU Yong-ding,et al.Application of Bayes learning algorithm to classify network traffic.Computer Engineering and Applications,2010,46(25):78-81.

Abstract: As network applications such as P2P rapidly increase,which makes the efficiency of traditional network traffic classification method that is based on the port and payload reduces.In this paper,it introduces a FCBF feature selection method,which can choose the best feature subsets.It uses Bayes learning algorithm to classify the network.The result of experiment shows a better classification accuracy.

Key words: network traffic;feature selection;naive Bayes classifier

摘 要: 随着网络应用(如P2P)的快速增长,使得传统的基于端口与有效载荷的网络流量分类方法效率大大降低。基于FCBF特征选择方法选择最优特征子集,研究使用贝叶斯学习方法对网络流量进行分类;实验结果显示提出的方法取得了较好的分类准确率。

关键词: 网络流量;特征选择;朴素贝叶斯学习器
DOI:10.3778/j.issn.1002-8331.2010.25.023 **文章编号:** 1002-8331(2010)25-0078-04 **文献标识码:** A **中图分类号:** TP309

1 引言

近年来随着互联网上的网络应用不断增加,除了传统的应用(如E-mail,Web和Ftp),还有许多新的应用(如Streaming,Gaming和P2P)出现。由于一些新的应用不再使用固定的和事先预知的端口号,许多应用没有IANA分配或注册的端口号,仅使用周知的默认端口,而这些通常与IANA分配的端口号存在交迭,这导致基于端口的方法,无法正确识别流量的应用类型(http://www.portsdb.org)。甚至周知的或注册的端口的应用,也会因下列原因使用不同端口号,而无法正确识别和分类。即,非特权用户通常使用1 023以上的端口;用户可能故意隐藏他们的存在或绕过基于端口的过滤器;若干服务器共享单个IP地址(主机);常用的几种P2P应用,也使用不可知的动态端口。基于有效载荷的分析方法通过分析包的有效载荷来进行分类和统计,这种方法甚至可以很好地用来对P2P流量的识别[1-3]。尽管基于有效载荷识别这种技术避免依赖于固定端口号,但它增加了网络识别设备的复杂性和处理的负担,随着P2P应用的增加,特征串的数量也相应增加,使得该方法每检测一个报文所需要匹配的特征串越来越多,从而识别的效率逐渐降低。研究基于网络流量的应用类型的流量分类方法,这种方法以抽取独立于端口、协议和有效载荷的网络流量的信息作为特征,用基于相关性的特征选择方法来优选特征子集,使用贝叶斯学习方法对网络流量进行分类研究。这种分类方式可以避免使用端口、协议和有效载荷信息。

2 网络流的定义及侯选特征产生

2.1 流的定义

研究中流的定义如下:在基于TCP/IP协议的互联网中,按照源IP地址、源端口号,目标IP地址、目标端口号及IP协议五元组(Tuple),将报文(Packets)分成双向TCP或UDP流,同时,规定流与流之间的空闲时间(Idle Timeout)为120 s,即,超过120 s被认为是不同的流。按照上述流的定义,分析跟踪文件中的报文信息,从而形成网络流。用$F=\{F_1,F_2,\cdots,F_i,\cdots,F_n\}$表示流集合,$n$表示样本流的个数,$F_i=\{f_{i1},f_{i2},\cdots,f_{ij},\cdots,f_{im}\}$表示第$i$条样本流,其中$m$表示样本流的属性个数,$f_{ij}$表示第$i$条流第$j$个属性。设$C=\{C_1,C_2,\cdots,C_h,\cdots,C_k\}$表示流所属的类别标签集合,其中$k$表示类别的数量,$C_k$表示第$k$类流,如

基金项目: 中国博士后科学基金(the China Postdoctoral Science Foundation under Grant No.20070410299);广东省高等学校人才引进项目。
作者简介: 邱密(1982-),男,硕士生,主要研究方向:网络流量分类,智能信息处理;阳爱民(1970-),男,博士后,教授,主要研究方向:智能计算,模糊分类,网络流量分类;何震凯(1977-),男,主要研究方向:网络流量分类,智能信息处理。
收稿日期: 2009-02-18 **修回日期:** 2009-04-13

ppstream等。

2.2　候选特征产生

基于贝叶斯学习方法的流量分类研究的一个重要目标是克服基于端口和基于有效载荷方法的缺陷，提高网络流量的分类准确度。因此，在流的特征产生上要考虑特征应独立于协议、通信端口，同时，又希望能很好适合基于贝叶斯学习方法的网络流量分类。网络流量的潜在特征有很多，为了保证流样本标注的正确性，提高分类的正确率，采用手工标注的方法，在局域网内使用多台电脑，每台同时运行单独的相同的应用程序，通过路由器的端口镜像获取数据，获取了4种P2P和4种非P2P应用类型的原始数据，把获取的原始网络报文保存为DMP文件，合成流时依据报文的IP地址直接标注流的应用类型。本文中共使用了网络流的34种特征作为网络流量分类的候选特征集，如表1。

表1　流的候选特征

特征的符号表示	特征描述
fPackets　fBytes bPackets　bBytes	前、后向总的报文的个数，大小
minFpktLen　maxFpktLen meanLenFsum　stdLenFsqsum	前向报文的最小、最大、平均长度、均方差
minBpktLen　maxBpktLen meanLenBsm　stdLenBqsm	后向报文的最小、最大、平均长度、均方差
duration	流的持续时间
fLess100　bLess100 fLess500　bLess500 fLess1000　bLess1000 fBig1000　bBig1000	双向报文长度的范围区间
avePktPerSecond	持续时间
fframNum　bframNum	包/s
fUrgCnt　bUrgCnt　fAckCnt bAckCnt　fPshCnt　bPshCnt fRstCnt　bRstCnt fSynCnt　bSynCnt	前、后向报文中有推送比特位，紧急标志位包的个数

3　基于FCBF的特征选择

3.1　基于FCBF过滤器方法分析

特征选择经常被作为机器学习[4-5]的预处理步骤，它根据一定的评估标准，通过选择原始数据子集来有效降低特征空间。特征选择算法分为两类：过滤器(filter)[6-7]和封装器(wrapper)[8-9]模型。采用了过滤器模式，基本思想是：先用基于类别相关性方法(C-correlation)对网络流的大量候选特征进行初步选择，得到了与类别高度相关的一些特征，然后，对初步选择的网络流用特征相关性方法(F-correlation)进一步选择，称这种方法为Fast Correlation-Based Filter(FCBF)方法。

使用SU(Symmetrical Uncertainty)[10]作为好的特征评价标准，找出在基于相关性的特征(包括类别)基础上来选择好的特征，这里包括两个步骤：(1)怎样决定一个特征是否和类别相关；(2)怎样决定这样一个与类别相关的特征是否与其他特征冗余。

第一个问题通过SU来解决，也可以用其他一些特征权重算法，假定样本流F_i有m个特征，C_h表示某一应用类别，$SU_{j,h}$表示f_{ij}(样本流F_i中第j个特征)与类别C_h的相关性度量SU，然后对于另一个样本流$F_{i'}$与类别相关的特征就由SU的初始值δ决定的，

$$\forall f_{ij} \in F_i', 1 \le i \le n, 1 \le j \le m, SU_{j,h} \ge \delta \quad (1)$$

对于第二个问题，需要决定F_i'中的两个特征之间的相关性是否足够引起冗余，若冗余就把它们中的一个从F_i'中除去。主要的相关性定义如下两种：

(1)一个特征$f_{ij}(f_{ij} \in F_i)$与某一类别C_h的相关是主要的相关性：$SU_{j,h} \ge \delta$并且$\forall f_{ik} \in F_i'(k \ne j)$，不存在这样的$f_{ik}$使得$SU_{k,i} \ge SU_{j,h}$，如果存在这样的$f_{ik}$，就把它称为对$f_{ij}$冗余的变量，使用$Sp_i$来表示对$f_{ij}$冗余的特征集，假定$f_{ij} \in F_i'$并且$Sp_i \ne \varnothing$，把$Sp_i$分为两部分，$Sp_i^+$和$Sp_i^-$：

$$Sp_i^+ = \{f_{ik} | f_{ik} \in Sp_i, SU_{k,h} \ge SU_{j,h}\} \quad (2)$$

$$Sp_i^- = \{f_{ik} | f_{ik} \in Sp_i, SU_{k,h} \le SU_{j,h}\} \quad (3)$$

(2)主要的相关性是指一个特征对于类别是主要的或者说在移除冗余的特征后它变为主要的。

假设是当找到两个特征相互冗余并且要把其中一个移除时，就把那个与类别相关性较小的去掉，从而达到降维的目的。

步骤1　如果$Sp_i^+ = \varnothing$时，把f_{ij}看作是主要的特征，删除所有属于Sp_i^-的特征，快速识别冗余的特征。

步骤2　如果$Sp_i^+ \ne \varnothing$时，在决定f_{ij}之前处理所有属于Sp_i^+的特征，如果它们中没有特征是主要特征，返回步骤1，否则除去f_{ij}并且决定是否要除去属于Sp_i^-的特征，这些特征都是属于F_i'的。

步骤3　拥有最大$SU_{j,c}$值总是被认为是主要特征并且从这里开始移除其他冗余特征。

3.2　FCBF算法

基于前面提出的方案，提出一种FCBF(Fast Correlation-Based Filter)算法，算法思想如下：

给定一个m个特征和某一应用类别C_h的数据集，算法为目标概念找到一个主要的特征集$Sbest$。

算法程序伪代码如下：

输入　$S(F_1, F_2, \cdots, F_i, \cdots, F_n, C_h)$//训练数据集，$\delta$：预先定义的初始值

输出　$Sbest$

```
1  Begin
2  For j=1 to m do begin
3  Calculate SU_{j,h} for f_ij;
4  If(SU_{j,h}>δ)
5  Append f_ij to S'_list;
6  End;
7  Order S'_list in descending SU_{j,h} value;
8  f_ij=getNextElement(S'_list);
9  Do begin
10 f_ik=getNextElement(S'_list,f_ij);
11 if f_ik <> NULL
12 do begin
13 f'_ik=f_ik
14 If (SU_{k,i}>=SU_{j,h})
15 Remove f_ik from S'_list;
16 f_ik=getNextElement(S'_list,f'_ik);
17 Else f_ik=getNextElement(S'_list,f_ik);
18 End until (f_ik==NULL);
19 f_ij=getNextElement(S'_list,f_ij);
```

20 End until（f_{ij}==NULL）

21 *Sbest*=S'_{list}；

22 End；

上述到代码包含两个部分，第一部分是计算每一个特征的SU值代码2~7行，基于预先定义的初始值δ来选择相关特征加入到S'_{list}中，并且根据SU值来对它们进行降序的排序，第二部分进一步处理已排序的S'_{list}来移除冗余的特征代码8~20行，这样就保证了特征集中是主要的相关性特征，根据3.1节的步骤1，特征f_{ij}已被决定为一个主要的特征能够被用来过滤其他特征，迭代从S'_{list}中的第一个元素开始（步骤3），并按下面步骤进行，对于所有剩余的特征（从f_{ij}的下一个元素到S'_{list}中的最后一个元素），如果特征f_{ij}对于特征f_{ik}冗余，则把f_{ik}从S'_{list}中除去（步骤2），在基于f_{ij}的一个过滤特征循环后，算法从当前剩余的特征f_{ij}中的下一个元素作为新的开始来重复过滤过程，算法在确定没有特征从S'_{list}移除之后截止。

提出的算法第一部分的时间复杂度是依据N个特征来计算的，第二部分中，对于每一次迭代，使用主要的特征f_{ij}在第一次的循环中，FCBF能够移除大量的相对于f_{ij}冗余的特征，最好的情况是把剩余的特征全部移除，最坏的结果是剩余的特征全部保留，一般来说，假设在迭代中半数以上的特征能够被移除，FCBF算法不但能过滤掉特征集合中的无关特征，而且能有效地找到特征集合中的冗余特征，得到满意的特征子集，从而提高了分类器的性能。

4 贝叶斯学习方法

4.1 基于朴素贝叶斯学习方法的网络流量分类器

贝叶斯学习方法中实用性很高的一种为朴素贝叶斯学习器，常被称为朴素贝叶斯分类器[11]（naive Bayes classifier）。朴素贝叶斯分类器的学习任务中，样本流F_i可由网络流属性值的合取描述，而目标函数C_h从某有限集合V中取值。学习器被提供一系列关于目标函数的训练样例以及新实例（描述为属性值的元组）$F_i=\{f_{i1},f_{i2},\cdots,f_{ij},\cdots,f_{im}\}$，然后要求预测新实例的目标值（或分类）。贝叶斯方法的新实例分类目标是在给定描述实例的属性值$<f_{i1},f_{i2},\cdots,f_{ij},\cdots,f_{im}>$下，得到最可能的目标值$V_{MAP}$。

假定有k个类$C_1,C_2,\cdots,C_h,\cdots,C_k$，$k$表示流量的应用类型的个数，给定一个未知的数据样本$F_i$，分类法将预测$F_i$属于具有最高后验概率（条件$F_i$下）的类，即朴素贝叶斯分类将未知的数据样本分配给类C_h，当且仅当$P(C_h|F_i)>P(C_g|F_i)$，$1<g<k$，$g\neq h$由此得到朴素贝叶斯分类的公式如下：

V_{NBC}=arg max$(P(C_h|F_i))$其中

$$P(C_h|F_i)=\frac{P(F_i|C_h)P(C_h)}{P(F_i)} \tag{4}$$

由于$P(F_i)$对于所有类为常数，只需计算最大$P(F_i|C_h)P(C_h)$即可。计算$P(C_h)$可以通过公式$P(C_h)=S_h/S$计算，其中，S_h是类C_h中的训练样本数，S是训练样本总数。

但是在实际应用中，对于给定具有许多条件属性的数据集，计算最大后验概率$P(F_i|C_h)$，计算的开销可能非常大。为了降低计算的开销，朴素贝叶斯分类器作了条件独立假设，假定各属性相互条件独立，即在属性间不存在依赖关系，因此，

$$P(F_i|C_h)=\prod_{k=1}^{n}P(f_{ik}|C_h) \tag{5}$$

概率$P(f_{ik}|C_h)$可以由训练样本计算，即

$$P(f_{ik}|C_h)=\frac{S_{ik}}{S_h}$$

其中S_{ik}是在属性f_{ik}上具有值i类C_h的训练样本数，而S_i是C_h中的训练样本数，由此得到NBC算法的分类公式：

$$P(x)=\prod_{i=1}^{n}p(x_i|\Pi_i) \tag{6}$$

为测试未知样本x的分类，对于每个类C_h，计算每个$P(F_i|C_h)P(C_h)$，样本F_i则被指派到$P(F_i|C_h)P(C_h)$最大的类C_h，即

$$V_{MAP}=P(C_h|F_i)P(C_h)>P(C_g|F_i)P(C_g) \tag{7}$$

与其他分类算法相比，NBC算法理论上具有最小的误分类率，但朴素贝叶斯分类器是基于一个简单的假定：即在给定分类特征条件下各属性之间是相互条件独立的。

4.2 基于贝叶斯网的网络流量分类器

贝叶斯网络（Bayes net）[12]是基于概率推理的数学模型，所谓概率推理，就是通过一些变量的信息来获得其他变量的概率信息的过程。一个贝叶斯网络是一个有向无环图（Directed Acyclic Graph，DAG），由代表变量结点及连接这些结点的有向边构成。用符号$B(G,P)$表示一个贝叶斯网络，$B(G,P)$由两部分构成：

（1）一个具有N个结点的有向无环图，图中的结点代表随机变量，结点间的有向边代表了结点间的相互关联关系。结点变量可以是任何问题的抽象用以代表属性、状态、客体、命题或其他的实体，如测试值、观测现象等。结点之间的有向边（弧）反映了变量间的依赖关系，指向结点X的所有结点称为X的父结点。尽管从结点X指向结点Y的弧频繁地被用来表示X引起了Y，但在贝叶斯网络里这不是对弧的唯一解释。例如，Y可能只与X有关联，但是它不是由X引起的。因此，虽然贝叶斯网络可以表示因果关系，但它们并不局限于表示因果关系。除了被称为贝叶斯网络外，它还有另一些术语通常认为有向边表达了一种因果关系，故贝叶斯网络有时叫做因果网（causal network）。重要的是，有向图蕴涵了条件独立性假设，贝叶斯网络规定图中的任一结点X_i条件独立于由X_i的父结点给定的非X_i后代结点构成的任何结点子集，即如果用$A(X_i)$表示非X_i后代结点构成的任何结点子集，用Π_i表示变量X_i的父结点集，Π_i（或Pa_i）表示Π_i的配置情况，Pa_i表示某一具体的配置。对每一个X_i将有一个子集$\Pi_i\subseteq\{X_1,X_2,\cdots,X_{i-1}\}$使得$X_i$与$A(X_i)=\{X_1,X_2,\cdots,X_{i-1}\}\backslash\Pi_i$在给定$\Pi_i$的前提下是条件独立的。那么，对任意的$X$将有$P(X_i|X_1,X_2,\cdots,X_{i-1})=P(X_i|\Pi_i)$，因此，有

$$P(x)=\prod_{i=1}^{n}px_i|\Pi_i) \tag{8}$$

这里变量集$(\Pi_1,\Pi_2,\cdots,\Pi_n)$对应着贝叶斯网络的父结点$(Pa_1,Pa_2,\cdots,Pa_n)$。

（2）一个与每个结点相关的条件概率表（Conditional Probability Table，CPT）。条件概率表可以用$P(X_i|\Pi_i)$来描述，它表达了结点同其父结点的相关关系—条件概率。没有任何父结点的结点概率为其先验概率。因为有了结点及其相互关系、条件概率表，故贝叶斯网络可以表达网络中所有结点（变量）的联合概率分布。

5　实验的结果与分析

5.1　实验数据获取及处理

通过校园网络中心交换机(Cisco 6509)的端口镜像的方式,来采集网络流量数据。采集时,截取报文前面的128 Byte长度,采集的数据形成Libpcap(*.dmp)格式的网络流量踪迹文件(Trace Files)。

在不同时期获取了4种非P2P和4种P2P应用类型的原始数据,这些数据都是存储为dmp格式的。见表2。

表2　原始dmp数据集

应用类型	数据大小/GB	应用类型	数据大小/GB
ftp	1.39	dianlu	2.32
Web	1.65	Xunlei	2.60
pop3	1.08	Bittorrent	3.39
smtp	1.02	ussee	3.64

5.2　特征选择

从5.1节中获取的DMP文件解析为流之后,为了使流样本标注正确,采用基于端口和机器IP的方法,来对流的应用类型进行标注。标注好之后利用第3章中介绍的FCBF特征选择算法进行特征选择,得到最有效的特征子集,共17个,其中特征如下:

fPackets, fBytes, bPackets, bBytes, minFpktLen, maxFpktLen, meanLenFsum, stdLenFsqsum, minBpktLen, maxBpktLen, meanLenBsum, stdLenBsqsum, duration, fUrgCnt, bUrgCnt, fPshCnt, bPshCnt,这些网络流的特征将被用于后面的分类器的构建和测试。

从获取的大量数据流中抽取每种协议中少部分流量样本来进行实验,具体信息如表3。

表3　流应用类型及数量

流应用类型	训练集数量	测试集数量
ppstream	1 000	600
Bittorrent	1 000	600
Xunlei	1 000	600
ussee	1 000	600
ftp	1 000	600
pop3	400	200
smtp	400	200
Web	1 000	600
总数	6 800	4 000

5.3　分类实验结果

实验平台采用的硬件环境:Inter® core 2 duo GHz,内存1 GB软件环境:操作系统Windows XP Professional实验工具为Weka 3.5。

在实验中,使用朴素贝叶斯、贝叶斯网两种贝叶斯学习方法,各分类器的参数分别如下:

基于朴素贝叶斯的网络流量分类参数:使用监督离散(use Supervised Discretization),设置为true,可以转换数字属性为分类属性。

基于贝叶斯网的网络流量分类器参数:这里用的是Weka中自带的参数,阿尔法(alpha)是用来估计概率表(CPT),表示最初期望值,设置*alpha*=0.5,maxNrofparents是表示贝叶斯网中一个结点可以拥有的最大父类数,设置maxNrofparents=1。

实验中,使用10折交叉验证方式进行分类器构建与测试。首先对未进行特征选择和已特征选择的样本集进行分类器的构建与测试,并用平均分类准确率标准进行评估,结果如表4。

表4　34和17个特征时分类器的性能

分类器	平均分类正确率/(%)	
	34个特征	17个特征
naive Bayes	87.980 0	89.100
Bayes net	88.488 9	89.125

同时,还用朴素贝叶斯、贝叶斯网两种贝叶斯学习方法分别对上述8种协议的流样本在特征选择之前和之后分别进行分类测试,最后用查准率来进行评估,结果如图1和图2所示。

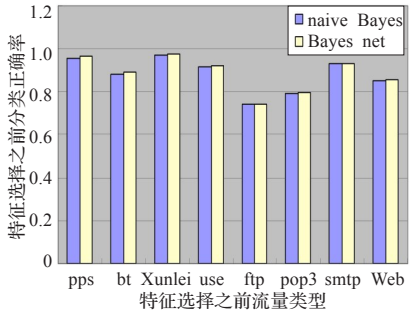

图1　特征选择之前分类正确率

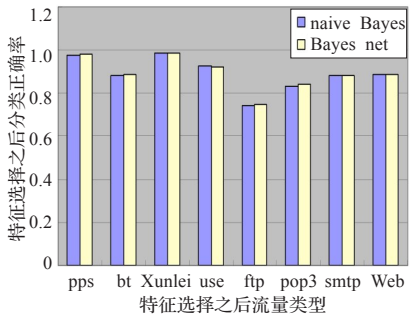

图2　特征选择之后分类正确率

从实验结果可以看出:(1)利用FCBF特征选择方法选择出的特征子集较好地保持分类准确率;(2)对两种贝叶斯学习算法的性能评估,贝叶斯网算法的分类性能是稍好一些的。

6　结论

研究了抽取独立于端口、协议和有效载荷的网络流的统计特征,通过过滤器的的特征选择方法降低了数据维数,利用朴素贝叶斯和贝叶斯网算法对流量进行分类,达到了较好的分类效果,为实际应用奠定了坚实的基础。下一步,将继续研究贝叶斯学习算法,不断改进分类器性能,以适应更加复杂的网络流量分类。

参考文献:

[1] Sen S, Wang J.Analyzing peer-to-peer traffic across large networks[J].IEEE/ACM Transactions on Networking, 2004, 12(2): 219-232.

[2] Gerber A, Houle J, Nguyen H, et al.P2P, the gorilla in the cable[R]. AT & T Labs-Research, 2004

(下转96页)

(2)方案能够抵抗签名接收者的一般性伪造攻击

假设(m,s,e)是签名接收者R收到的有效代理签名,R要想伪造出消息m'的有效签名$(m',s,e_f)(e_f=h(r\|m')(\mathrm{mod}\,q))$,并声称此签名是$B$所为,运用前面提到的签名接收者的一般性伪造攻击方法,需等式(7)成立:

$$g^{s}y_P^{e_f}(\mathrm{mod}\,p)=t^{t}g^{a}y_P^{b}y_R(\mathrm{mod}\,p) \tag{7}$$

等式(7)左边$=g^{a+kt+e^{*}s'+x_R}y_P^{e_f}(\mathrm{mod}\,p)=t^{t}g^{a}g^{(b-e)s'}g^{x_R}g^{s'e_f}(\mathrm{mod}\,p)$,等式(7)转化为:$t^{t}g^{a}g^{(b-e)s'}g^{x_R}g^{s'e_f}(\mathrm{mod}\,p)=t^{t}g^{a}y_P^{b}y_R(\mathrm{mod}\,p)$,即:$g^{(b-e)s'}g^{s'e_f}(\mathrm{mod}\,p)=g^{s'b}(\mathrm{mod}\,p)$。

由于e和e_f不同,所以等式(7)无法成立。因此,签名接收者不能伪造出消息m'的有效签名。

(3)方案具有非关联性

如果B保留了盲消息$\bar{m}$的签名$sig(\bar{m})$及相关数据,当消息m的签名(m,u,s,e)被公开后,通过等式$e^{*}=b-e(\mathrm{mod}\,q)$能计算出$b$。由于在签名提取阶段引入了$R$的私钥$x_R$,所以$B$无法从等式$s=a+s''+x_R(\mathrm{mod}\,q)$解出$a$。因此,新方案具有非关联性。

4.2.2 计算效率分析

这里令T_e、T_i和T_m分别表示模幂、模逆和模乘等计算。在签名阶段,原方案和新方案的计算量分别为:$8T_e+4T_i+7T_m$和$4T_e+5T_m$;在验证签名阶段,原方案和新方案的计算量分别为:$3T_e+T_i+3T_m$和$2T_e+T_m$,在其他两个阶段两方案相同。由以上分析可知,新方案中减少了模幂运算和模乘运算,避免了模逆运算,而影响方案执行效率的主要因素是模幂运算、模逆运算和模乘运算[9]。因此,新方案的计算效率更高。

5 结束语

代理盲签名作为一种新的签名技术具有广泛的应用前景,可用于电子商务中CA证书、电子现金、电子选票的签发等方面,然而目前已提出的许多方案在安全性等方面皆有不足之处。基于文献[5]提出了一个新的基于离散对数的代理盲签名方案,分析可知,新方案能够克服原方案存在的安全问题,并且具有通信量小、算法复杂性低等优点,为代理盲签名在实际中的应用提供了安全性和效率方面的保证。

参考文献:

[1] Mambo M,Usuda K,Okamoto E.Proxy signatures for delegating signing operation[C]//Proc of 3rd ACM Conference on Computer and Communications Security.New York:ACM Press,1996:48-57.

[2] Chaum D.Blind signatures for untraceable payments[C]//Proceedings of Eurocrypt'82.Burg Feuerstein,Germany:Plenum Press,1983:199-203.

[3] Lin W D,Jan J K.A security personal learning tools using a proxy blind signature scheme[C]//Proceedings of International Conference on Chinese Language Computing.Washington:IEEE Computer Society,2000:273-277.

[4] Tan Zhuo-wen,Liu Zuo-jun,Tang Chun-ming.A proxy blind signature scheme based on DLP[J].Journal of Software,2003,14(11):1931-1935.

[5] Tan Zhuo-wen,Liu Zuo-jun,Tang Chun-ming.Digital proxy blind signature schemes based on DLP and ECDLP[J].MM Research Preprints,2002,21(7):212-217.

[6] Zhang F G,Safavi-Naini R,Lin C Y.New proxy signature,proxy blind signature and proxy ring signature schemes from bilinear pairings[EB/OL].(2003-5-29).http://eprint.iacr.org/2003/104.

[7] Li Ji-guo,Wang Shu-hong.New efficient proxy blind signature scheme using verifiable self-certified public key[J].International Journal of Network Security,2007,4(2):193-200.

[8] 秦宝东.对两种基于离散对数代理盲签名的分析[J].计算机工程与应用,2009,45(3):104-105.

[9] Xu Zhong,Dai Guan-zhong,Yang De-ming.An efficient ECDSA-based signature scheme for wireless networks[J].Wuhan University Journal of Natural Sciences,2006,11(6):1707-1710.

(上接81页)

[3] Yuan Huang,Tseng Shian-Shyong,Wu Gang-shan,et al.A two-phase feature selection method using both filter and wrapper[C]//Proc of 1999 IEEE Inter'l Conf on Systems,Man,and Cybernetics,1999,2:132-136.

[4] Mitchell T M.Machine learning[M].[S.l.]:McGraw-Hill Education,1997.

[5] Mitchell T M.Does machine learning really work[J].AI Magazine,1997,18(3):11-20.

[6] Williams N,Zander S,Armitage G.Evaluating machine learning algorithms for automated network application identification,Technical Report 060410B[R].2006.

[7] Kohavi R,John G H.Wrappers for feature subset selection[J].Artificial Intelligence Journal,1997,97(1/2):273-324.

[8] Liu H,Setiono R.A probabilistic approach to feature selection:A filter solution[C]//Proc of Intel Conf on Machine Learning,1996:319-327.

[9] Das S.Filters,wrappers and a boosting based hybrid for feature selection[C]//Proc of the 8th Intel Conf on Machine Learning,2001:74-81.

[10] Yu Lei,Liu Huan.Feature selection for high-dimensional data:A fast correlation-based filter solution[C]//Proceedings of the 20th International Conference on Machine Learning(ICML 2003),2003.

[11] Moore A W,Zuev D.Internet traffic classification using Bayesian analysis techniques[C]//Proc ACM Sigmetrics,2005:50-60.

[12] 邓河,阳爱民,刘永定.一种基于SVM的P2P网络流量分类方法[J].计算机工程与应用,2008,44(14):122-126.