基于游程检测与快速傅里叶变换的加密比特流识别

邢 萌 吴 杨 王 韬 李进东
(军械工程学院信息工程系 石家庄 050003)

摘 要 为获得链路层中的加密与未加密比特流样本,首先提出了基于游程检测方法的链路层加密比特流识别方案,解决了未知网络环境下的加密与未加密比特流样本获取问题。同时,采用快速傅里叶变换分别对加密与未加密比特流样本进行处理,根据最大差异原则确定了快速傅里叶变换结果的特征点位置,并基于正态分布原理确定了特征点的取值,建立了特征模板。最后,以某无线网络链路层加密比特流为识别对象,对提出的方案的有效性进行了验证。结果表明,该方案对链路层加密与未加密比特流的识别率均可达到95%以上。

关键词 加密比特流,游程检测,快速傅里叶变换
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.1.038

Identification of Encrypted Bit Stream Based on Runs Test and Fast Fourier Transform

XING Meng WU Yang WANG Tao LI Jin-dong
(Department of Information Engineering, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract To obtain samples of encrypted data and plaintext in data link layer, an encrypted data identification scheme was provided based on the run test, meanwhile, and the fast Fourier transform was used to process the encrypted data and plaintext. Based on the principle of maximum difference, the characteristic point of the result of the fast Fourier transform was determined. Then the value of the characteristic point and the feature template were determined using the principle of normal distribution. Finally, the identification rate of the proposed scheme was verified, taking a wireless network data as the identification object. The experimental results demonstrate that the rate of the proposed scheme achieves 95% both for the encrypted data and the plaintext.

Keywords Encrypted bit stream, Runs test, Fast Fourier transform

1 引言

网络协议识别的主流方法是基于应用层特征字段的检测技术[1],并且识别对象多为已知网络中的未加密数据。而当前日益严峻的网络安全形势促使研究热点越来越多地转向对未知网络协议的识别。由于链路层协议识别是开展上层协议识别的基础,识别未知网络中链路层数据是否加密,可为进一步更加有针对性地开展协议识别技术研究提供支持。而目前,在对未知网络链路层加密比特流识别的研究方面成果较少,研究方法也主要是基于特征串匹配等算法,即通过将目标比特流与特定模式串相匹配的查找方式,且该方法只适用于已知网络的识别,对未知网络链路层的加密比特流识别并不适用。

Charles等[2]提出了通过加密后保持不变的特征(数据包字节数、持续时间和流方向)来识别加密流量所属的应用协议,但并未给出判别该数据是否加密的方法。Sun等[3]则提出了一种混合多级的加密流量分类识别方法,即在加密网络流量分类识别过程中,以签名匹配方法对SSL/TLS的流量进行了分类识别,而后采用统计分析方法确定加密流量所属的具体应用协议。Talieh等[4]提出了基于支持向量机的P2P流量识别方法,其主要是根据加密数据流的长度、方向等静态统计特性建立对应的特征向量。考虑到机器学习过程的复杂性,Zhang[5]、Du[6]等提出了基于聚类算法的加密流量分类方案,其仍通过统计加密数据流的长度、方向等静态统计特征,实现对加密流量的分类。以上研究主要基于应用层加密流量表现出的静态网络特征而展开,对链路层加密流量识别的有效性还有待验证,且并未从流量自身内容的角度对加密与未加密流量反映出的不同特征进行研究。而文献[7]则以加密流量自身内容为分析对象,利用其随机性特点,提出了基于加权累积和检验的自适应加密流量盲识别算法,其特点是对网络报文逐一实施累积和检验,并将结果进行加权综合,以提高对加密流量的识别率。文献[7]提出的算法在执行过程中,需首先对网络报文进行归类预处理,将属性相同的报文归类到相同的聚类中,其识别率依赖于参与累积和检验的报文具有相似的属性,该条件使其无法有效应用于对单条报文数据的识别,同时也降低了算法的适用性。

本文针对已有研究方案识别率较低以及对单条数据识别的限制等问题,以加密与未加密比特流的随机性差异为依据,

到稿日期:2014-02-26 返修日期:2014-04-20 本文受军内科研资助项目(YJJXM12033)资助。
邢 萌(1990—),女,硕士生,主要研究方向为网络协议识别,E-mail:xingmykx@163.com;吴 杨(1985—),男,博士生,主要研究方向为网络协议识别;王 韬(1964—),男,教授,博士生导师,主要研究方向为网络安全、密码学。

通过随机性检测中游程检测的方法将某无线网络链路层的加密与未加密比特流提取出来，为基于快速傅里叶变换的链路层加密比特流识别提供所需的样本数据。其次，由于经过快速傅里叶变换的加密与未加密比特流在某些特征位有显著差异，提出基于快速傅里叶变换的加密比特流识别方案，通过确定特征位、提取特征值集合以及确定集合的中心值的方法来建立比特序列匹配模板。最后，通过计算待测比特流与匹配模板的差值确定待测序列的性质。

本文第 2 节论述了游程检测及 FFT 的相关理论与方法；第 3 节分析了将游程检测及 FFT 用于链路层加密比特流识别过程存在的问题；第 4 节分析论述了基于游程检测与 FFT 的链路层加密比特流识别方案；第 5 节对实验结果进行了分析比较；最后对本文工作进行了总结。

2 相关理论与方法

2.1 游程检测

信息在网络传输的过程中，既可以通过物理方式来确保其内容的机密性，也可以通过对信息进行加密的方式来确保机密性[8]。已加密数据借助其在统计方面的随机性实现信息的保密功能，并在统计特性上体现出了更大的随机特性。因此，链路层数据的随机分布特性可作为判断其是否加密的主要标志。目前，在众多的随机性检测项目和方法中，应用较为广泛的是美国商务部国家标准技术协会(NIST)于 2001 年 5 月公布的 FIPS140-2 标准中定义的用于密码系统安全性度量的诸多随机性检测方案[9]，以及 2010 年 4 月公布的应用于测试随机数及伪随机数生成器性能的随机性检测标准 SP800-22 revla(Special Publication 800-22 Revision 1a)[10]。其中码元频数检测、游程检测、块内频数检测为使用较多的 3 种随机性检测方法。由于前期准备工作中的实验证实了游程检测对加密比特流识别率较码元频数检测与块内频数检测高，因此实验采用游程检测的随机性检测方法来获得链路层中的加密与未加密比特流样本。

下面主要分析论述 SP800-22 revla 标准中的游程检测方法的基本原理。

游程是序列的一个子串，由连续的“0”或“1”组成，并且其前导和后继元素都与其本身的元素不同。游程检测在码元频数检测后进行，主要对比特序列中的连续特征位进行检测，以检测序列中游程总数是否符合随机性要求。

对一个长度为 n 的比特序列进行游程检测时，定义 $\varepsilon=\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 为该序列中的每个比特，$V_n(obs)$ 为游程总数即 0 游程和 1 游程之和，检测的相关统计分布为 χ^2 分布。游程检测的描述如下：

需要注意的是，游程检测将码元频数检测作为它检测的前提。

(1)计算序列中比特“1”在整个序列中的比例：

$$\pi=\frac{\sum_j \varepsilon_j}{n} \tag{1}$$

理论上，随机序列的 0 游程和 1 游程应是相等的。检测过程中，若 r_i 为游程长度为 i 时的 0 游程或 1 游程的出现频数，则理论游程频数与序列长度间的关系满足：

$$\frac{n}{8}=r_1=2r_2=2^2r_3=\cdots=2^{i-1}r_i=\cdots \tag{2}$$

(2)作为先决条件，首先判断 $|\pi-\frac{1}{2}|\geqslant\tau$ 是否成立。如果成立则不需再执行游程检测，因为如果 $|\pi-\frac{1}{2}|\geqslant\tau$，则该序列不能通过码元频数检测。

(3)计算统计量

$$V_n(obs)=\sum_{k=1}^{n-1}r(k)+1 \tag{3}$$

其中，当 $\varepsilon_k=\varepsilon_{k+1}$ 时，$r(k)=0$，否则 $r(k)=1$。

(4)计算

$$P\text{-}value=erfc\left(\frac{|V_n(obs)-2n\pi(1-\pi)|}{2\sqrt{2n}\pi(1-\pi)}\right) \tag{4}$$

其中，$Run_n=\frac{V_n(obs)-2n\pi(1-\pi)}{2\sqrt{n}\pi(1-\pi)}$ 服从于标准的正态分布函数。

判断条件：若 $P\text{-}value<0.01$，则该序列不是随机序列；否则该序列是随机序列。SP800-22 revla 标准建议每条被检测的序列至少包含 100 比特。

2.2 快速傅里叶变换

傅里叶分析是数学分析的一个分支，它不仅对数学研究起着重要作用，在工程实践中也发挥了重要作用[11]。傅里叶分析提供了信号的频域分析方法，通过变换将时域和频率联系在一起，将时域内隐藏的现象和特征在频域内显现出来。

链路层比特流可以被描述为一维时域信号。一维时域信号通过傅里叶变换后可得到信号的频率和幅值。

函数 $f(t)$ 的一维傅里叶变换由下式定义：

$$\Im\{f(t)\}=F(s)=\int_{-\infty}^{\infty}f(t)e^{-j2\pi st}\,dt \tag{5}$$

其中，$j^2=-1$。在计算采样信号或图像的傅里叶变换时，通常用 DFT(Discrete Fourier Transform，离散傅里叶变换)来实现。一般来说所有的复指数值都存在一张表中，这样计算量实在太大。所以存在一类算法可以将操作降低到 $N\log_2(N)$ 数量级，这就是所谓的快速傅里叶变换算法(Fast Fourier Transform)。

本研究中，关于快速傅里叶变换的实验在 Matlab 环境下进行。将游程检测得到的 i 条待测比特流存储到数组 $m(i,j)$ 中之后，通过调用信号处理工具箱中的快速傅里叶变换函数 $fft(m(i))$ 的方式对待测比特流进行处理，每条比特流经 FFT 变换后的输出结果同样为一条序列，变换结果存入数组 $z(i,q)$ 中。通过将离散的比特流变换为相应的频域上的函数，使得链路层比特流的一些隐藏的特征在频域内显示出来。通过分析计算实验结果得出加密与未加密比特流的显著特征位，将加密与未加密比特流样本特征位的值分别存入集合，从而得到样本数据的特征值集合。

3 相关问题分析

3.1 比特序列长度对游程检测识别率的影响

在实际的检测过程中，SP800-22 revla 标准建议待识别比特序列的最短长度不应小于 100 比特。以某无线网络链路层数据为识别对象，在 Matlab 仿真环境下对不同比特序列长度条件下游程检测的识别能力进行了测试。在比特序列平均长度为 150 比特时，对对未加密比特流的识别率仅为 2.1%，而对对已加密比特流的识别率则为 99%。在比特序列长度为 600 比特时，未加密比特流的识别率为 25.9%，而已加密

比特流的识别率则为 98.8%，如图 1 所示。

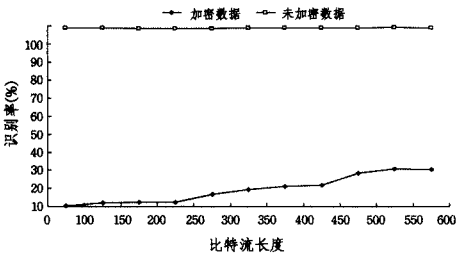


图 1　比特序列长度对识别率的影响

由图中数据可知，当比特流的平均长度为 150 比特时，由于对未加密比特流的识别率过低，需要大量源数据。因此，为减小提取样本数据的工作量，实验选择游程检测识别率较高时的比特序列长度，即 600～800 比特的比特流作为游程检测的实验数据。

3.2　比特序列特征位选择

FFT 实验在 Matlab 仿真环境中通过调用信号处理工具箱中的函数 $fft()$来实现。实验分别对 10000 条加密与未加密比特序列进行快速傅里叶变换。将游程检测得到的 i 条加密与未加密比特流($i=10000$)分别存储到数组 $m(i,j)$中之后，通过调用信号处理工具箱中的快速傅里叶变换函数 $fft(m(i))$的方式对待测比特流进行处理，变换结果存入数组 $z(i,q)$中。然后将各序列变换结果相应位置的数组元素相加取平均值，分别得到加密与未加密序列的平均快速傅里叶变换结果 $A_j=\frac{1}{10000}\cdot\sum_{i=1}^{10000}z(i,j)$，如图 2 所示。

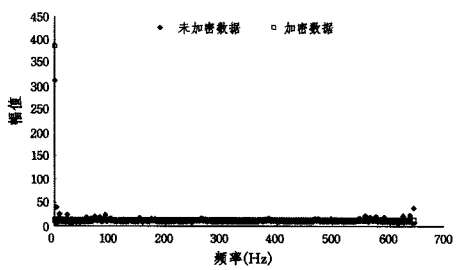


图 2　比特序列的平均快速傅里叶变换结果

从图 2 中大致可以看出，加密比特流与未加密比特流在其快速傅里叶变换结果中有显著差异的一位是第一位。为确定该位是加密与未加密比特流 FFT 结果差异最大的一位，对 FFT 结果进行标准差计算，得到如图 3 所示的结果。由均值与标准差的计算结果可知，第一位的数据差值最大，可达到 $\Delta=\xi_1-\xi_2+\sigma_1-\sigma_2=74.10$。其中，$\xi_1$，$\xi_2$ 分别为加密比特流与未加密比特流第一位的平均 FFT 结果，σ_1，σ_2 分别为加密与未加密比特流第一位的 FFT 标准差计算结果，因此取第一位为特征位。

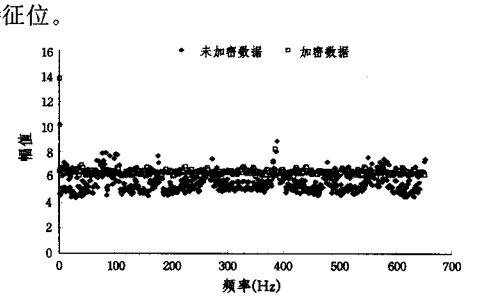


图 3　比特序列的快速傅里叶变换标准差计算结果

4　识别方案

在对加密比特流进行识别的过程中，将采用如图 4 所示的识别方案。

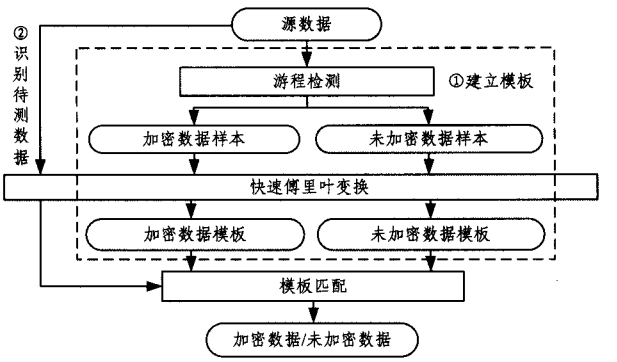


图 4　加密比特流识别方案

首先通过游程检测对源数据进行检测，以获得足够的链路层加密比特流及未加密比特流样本。其次，基于游程检测阶段获得的链路层数据样本，利用 FFT 的信号频域分析方法对数据样本进行变换与分析，根据加密与未加密比特流 FFT 变换结果就一定方法对特征位进行选取并建立对应的特征值集合。再次，采用模式匹配的方法确定特征值集合的中心值，后续的加密比特流识别将以特征值集合的中心值为模板进行识别。最后通过计算待测比特流 FFT 结果的特征值与集合中心值的差值，确定待测比特流的性质，从而实现对链路层加密比特流的识别。

4.1　基于 FFT 的特征值集合提取

随机性检测采用游程检测的方法，检测过程中的 $erfc$ 函数则在 Matlab 仿真环境中实现。得到未加密和加密的比特序列样本后，对未加密和加密比特流进行快速傅里叶变换，并对其特征值进行提取。

4.1.1　比特序列特征位选择

以获得加密与未加密序列的最明显差异为原则，对比特序列的特征位进行选择，可参考 3.2 节。选择过程如下：

1)将游程检测得到的 i 条加密与未加密待测比特流存储到数组 $m(i,j)$中之后，通过调用信号处理工具箱中的快速傅里叶变换函数 $fft(m(i))$的方式对待测比特流进行处理，变换结果存入数组 $z(i,q)$中；

2)将各序列变换结果 $z(i,q)$中相应位置的数组元素相加取平均值，分别得到加密与未加密序列的平均快速傅里叶变换结果 $A_j=\frac{1}{i}\cdot\sum_i z(i,j)$；

3)通过观察分析加密与未加密序列的平均快速傅里叶变换结果并对快速傅里叶变换的结果进行标准差计算，取计算结果差值最明显的位为特征位，详细内容参照 3.2 节。

4.1.2　样本容量的确定

根据随机抽样的基本原理，在建立特征值集合的过程中，需确定合适的样本容量，才能准确全面地反映待识别比特序列的特征信息。样本容量的确定将根据置信水平、允许误差和总体标准差来确定，采用的样本容量计算公式如下所示：

$$n=\frac{Z_{\frac{\alpha}{2}}^2 S^2}{d^2} \tag{6}$$

其中，n 代表建立匹配模板所需的样本容量，a 是置信水平，d

为允许误差。$Z_{\frac{a}{2}}$为概率度。而在$1-a=95\%$的置信水平条件下，取$Z_{\frac{a}{2}}=1.96$；在99%的置信水平条件下，取$Z_{\frac{a}{2}}=2.58$。

S为总体的标准差，公式如下：

$$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\tilde{x})^2} \tag{7}$$

S的渐进无偏估计即样本的标准差则定义为$\tilde{s}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x})^2}$，当样本量足够大时有$\frac{n-1}{n}\to 1$，即$S\approx\tilde{s}$成立。因此，可通过计算样本标准差$\tilde{s}$代替总体标准差$S$，以确定建立特征模板过程中所需的有效样本容量。

4.1.3 特征值集合提取

特征值集合提取，即将待测比特流进行快速傅里叶变换后，根据特征位的位置，将待测序列特征位的值提取到相应的集合中，为下一步基于模板匹配的集合中心值选取做准备。具体步骤如下：

1) 根据4.1.2节中确定的样本容量n，在游程检测的结果中分别选取n条加密与未加密样本比特流并存入相应的数组$mc(i,j)$、$md(i,j)$中；

2) 通过对数组$mc(i,j)$与$md(i,j)$进行快速傅里叶变换$fft(mc(i))$、$fft(md(i))$，得到FFT变换结果数组$zc(i,q)$、$zd(i,q)$；

3) 根据4.1.1节中确定的比特序列特征位将加密比特流与未加密比特流样本的FFT变换结果数组$zc(i,q)$、$zd(i,q)$中的特征位的值提取出来，并分别建立加密与未加密序列的特征值集合。

4.2 基于模板匹配的集合中心值选取

由于上一个步骤中提取出来的样本数据的特征值集合存在特征值的取值较多的特点，因此采取按一定步长进行区间统计的方法分别建立加密与未加密比特流的特征模板。再根据加密与未加密模板相关性的大小确定最佳统计步长，最后按确定的步长进行区间统计，对频数分布最大的区间取均值，将该均值作为相应集合的中心值。

4.2.1 确定特征值统计步长

由观察得，序列的特征值一般为整数，因此为方便计算，在确定统计步长时，采用以1为单位的循环枚举方式进行测试。通过对未加密与加密序列的特征值集合分别按区间进行统计并建立匹配模板的方式，以未加密与加密序列特征模板的相关性大小为依据，选取相关性最小时的步长为特征值统计步长。

a)特征值集合相关性度量

指定$M=\begin{bmatrix} I_1 & I_2 & \cdots & I_h \\ E_1 & E_2 & \cdots & E_h \end{bmatrix}$为特征模板，其中$I_j$表示模板中的第$j$个特征参数所对应的特征值集合统计区间的索引值，而$E_j$表示索引为$I_j$的统计区间的期望值。定义未加密序列特征值集合区间统计模板M与加密序列特征值集合区间统计模板M'之间的相关系数c为：

$$c=\frac{4(T,T')}{(|T|+|T'|)^2} \tag{8}$$

b)步长选定

根据a)中定义的未加密与加密序列特征值集合区间模板之间的相关系数c，从最小区间数2开始(区间数为1时即不进行分区)，采用以1为单位的循环枚举方式进行测试，分别对每轮分区后的集合进行相关系数c计算，将相关系数值最小一轮时的区间数作为选定的区间数，记为δ；步长记为Δ，作为选定的区间长度，其中$\Delta=\frac{z_{max}-z_{min}}{\delta}$。

4.2.2 特征点取值确定

根据4.2.1节中确定的区间数分别对未加密与加密序列的特征值集合进行统计，得到的统计结果分布可以拟合为一个正态分布。由正态分布的原理可知，选取特征值出现频率最高的区间的均值作为加密或未加密序列的特征点取值最具代表性。

1)由4.2.1节可知，以$\Delta=\frac{z_{max}-z_{min}}{\delta}$为步长，分别对未加密与加密序列特征值集合进行区间频率统计；

2)将得到的对应于未加密与加密序列的两个拟合正态分布图中频率最高的两个区间分别作为未加密序列与加密序列的选定区间；

3)分别取该区间的平均值为对应集合的中心值，分别记为α、β，将其作为下一步对待测比特流进行匹配的模板。

4.3 基于差值计算的待测比特流识别

计算待测序列的快速傅里叶变换特征值，将该特征值与4.2节中求得的集合中心值进行差值计算，取差值较小的集合性质为该待测序列的所属性质，从而达到对待测序列的识别，识别步骤如下：

1)首先对待测序列$x(i)$进行快速傅里叶变换，得到相应的FFT变换结果$z(j)$；

2)根据4.1.1节中确定的比特序列特征位选取该变换结果的特征值z；

3)将特征值z与4.2节中确定的未加密与加密序列集合的中心值α、β分别进行差值$|z-\alpha|$、$|z-\beta|$计算；

4)通过比较$|z-\alpha|$与$|z-\beta|$的大小来确定该比特流为加密比特流还是未加密比特流：当$|z-\alpha|>|z-\beta|$时该序列为加密比特流；当$|z-\alpha|<|z-\beta|$时该序列为未加密比特流；当$|z-\alpha|=|z-\beta|$时表明识别该比特流失败。

5 实验结果与分析

实验过程以互联网中链路层比特流为例，采用自主设计开发的网络数据捕获分析工具STREAM_TRACE捕获链路层数据，将其转换为比特流并进行存储。鉴于采集到的比特流均为未加密比特流，在VC++6.0环境下，调用OpenSSL0.98a密码库中的AES等加密函数，对其进行加密处理，每条比特流加密过程均以随机方式产生128位加密密钥，并将已加密的比特流存储到对应的集合中。在构建分析对象集合的过程中，根据已加密与未加密比特流的分布特点，随机从未加密和已加密比特流集合中选择比特流加入到待分析对象集合中。随机性检测采用游程检测的方法，检测过程中的$erfc$函数则在Matlab仿真环境中实现。

5.1 游程检测实验结果

由3.1节知，比特流长度对游程检测识别率有较大影响，因此实验选择600～800比特的比特流作为游程检测的实验数据，在数据数量为500～25000的范围内分别对加密比特流与未加密比特流进行了游程测试，测试结果如图5所示。

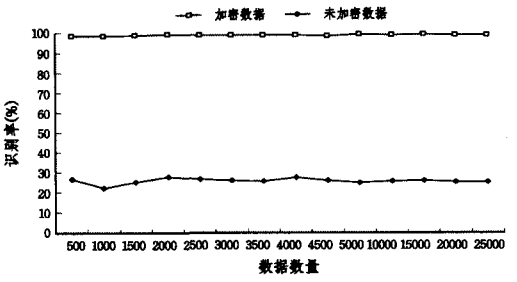


图 5　游程检测结果

其中未加密比特流的识别率在 22.62%到 27.60%之间，加密比特流的识别率在 98.13%到 99.36%之间。由实验结果可知，游程检测对未加密与加密比特流的识别率随比特流数量的变化无较大变化。

5.2　快速傅里叶变换实验结果及分析

5.2.1　样本容量确定与分析

根据式(6)确定样本容量时，取置信水平为 $Z_{\frac{\alpha}{2}}=2.58$，允许误差为 $d=0.3$。在样本量为 10000，未加密比特序列检测率为 25.2%，已加密比特序列检测率为 99.1%的条件下，将各参数代入式(6)中，分别得到未加密比特流所需样本容量为 $n=8036$，加密比特流所需样本容量为 $n=14113$，因此，实验设定样本容量不应小于 14113。

5.2.2　特征值集合提取与中心值确定

a)特征值集合提取

根据确定的样本容量分别选取 15000 条加密比特流与未加密比特流作为实验数据，由 3.2 节可知加密与未加密比特流的特征位为第一位，通过快速傅里叶变换分别将其样本的特征值提取到数组中，建立特征值集合。

b)确定特征值统计步长

实验采用以 1 为单位的循环枚举方式进行测试。由于特征值的范围为 280～430，为方便实验及作图，将实验范围定为 260～460，并将统计步长的取值范围定为 2～50。测试结果如图 6 所示。

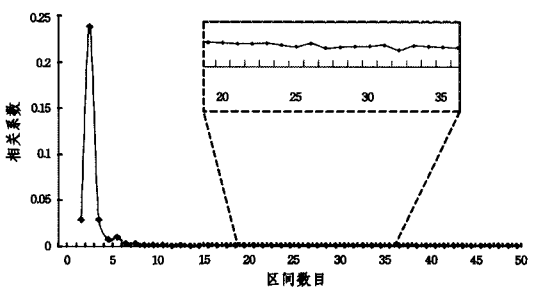
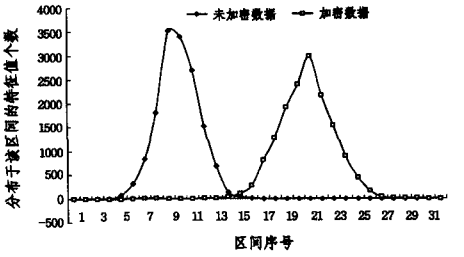


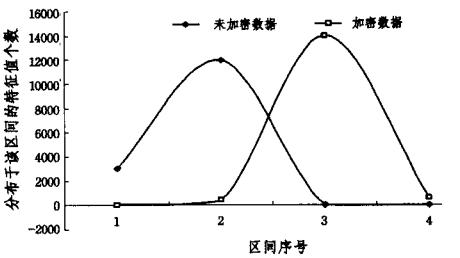
图 6　不同区间数下的未加密与加密比特流集合模板相关系数

由实验结果可知，当区间数为 32 时(由于当区间数大于 5 时相关系数变化微小，因此将区间数为 32 时的局部细节图展示出来，如图 6 所示)未加密序列与加密序列特征模板的相关性最小，相关系数为 $c=0.000552$(由式(8)计算得到)。因此确定特征值统计步长为$\frac{200}{32}$。

由图 6 还可以看出，当区间数为 2、3、4 时未加密序列与加密序列特征模板的相关性较大。当区间数小于 3 时做出的区间特征值个数分布图效果不好，为了达到较好的比较效果，选择区间数为 4 的特征值个数分布图与区间数为 32 的图进行比较。图 7 中(a)与(b)分别为区间数是 32 和 4 时的区间特征值分布情况。由两图比较可知，经过基于模板匹配的相关性度量测试确定的步长，在进行区间统计时可以有效降低未加密比特流与加密比特流特征值集合中心值之间的相关性，从而使得下一步确定集合中心值更加精准，待测比特流的识别率更高。



(a)区间数为 32 时各区间的特征值个数分布



(b)区间数为 4 时各区间的特征值个数分布

图 7　按区间统计的特征值个数分布

c)集合中心值确定与序列识别

根据上一步中确定的统计步长$\frac{200}{32}$，分别对未加密序列与加密序列的特征值进行统计，得到未加密序列与加密序列特征值集合的区间统计频率，如图 7(a)所示。根据正态分布原则分别选取频率最高的区间为未加密序列与加密序列的选定区间，取该区间的平均值为对应集合的中心值，得到未加密比特流特征值集合中心值 $\alpha=313$，加密比特流特征值集合中心值 $\beta=388$，将其作为对待测比特流进行匹配的模板。

对待测比特流的特征值 z 进行 $|z-\alpha|$ 与 $|z-\beta|$ 计算，通过比较 $|z-\alpha|$ 与 $|z-\beta|$ 的大小来确定该比特流为加密还是未加密比特流：当 $|z-\alpha|>|z-\beta|$ 时为加密比特流；当 $|z-\alpha|<|z-\beta|$ 时为未加密比特流；当 $|z-\alpha|=|z-\beta|$ 时表明识别该比特流失败。

5.2.3　基于 FFT 的比特序列识别率

通过实验验证，提出的方案对加密比特序列的识别率为 99.43%，对未加密比特序列的识别率达到 99.88%。在基于游程检测方法的链路层加密比特流的识别方案基础上，保证了对加密链路层比特流识别率并将未加密链路层比特流的识别率提高了 90%以上。同类研究的文献[7]在满足被检测报文序列均具有相似属性及样本数量超过 20 的条件下，其提出的算法对加密流量达到了 90%以上的识别率，在样本数量为 10 时，其识别率达到了 80%以上。而本文提出的方案对 1 条加密比特流的识别率可达到 99%以上，对未加密比特流的识别率可达到 95%以上，且提出方案的识别率不存在报文间的干扰问题，在保证识别率的同时具有较好的适用性。

结束语　对链路层加密比特流进行识别对加强网络安全防护能力及进一步的协议识别技术研究具有重要意义。通过随机性检测中游程检测的方法将未知网络中链路层的加密与未加密比特流作为识别样本，在此基础上提出了基于快速傅里叶变换与模板匹配的链路层加密比特流识别方案，解决了特征位选取、特征值确定等关键问题。结果表明：同已有的加

密数据识别方案相比，本文提出的方案适用性更好、识别率更高。

参 考 文 献

[1] 龙文，马坤，辛阳，等. 适用于协议特征提取的关联规则改进算法[J]. 电子科技大学学报，2010，39(2)：302-305

[2] Charles V W，Fabian M，Gerald M M. On inferring application protocol behaviors in encrypted network traffic[J]. Journal of Machine Learning Research，2006，7(12)：2745-2769

[3] Sun Guang-lu，Xue Yi-bo，Dong Ying-fei，et al. A Novel Hybrid Method for Effectively Classifying Encrypted Traffic[C]// Proceedings of Communications and Systems Security，2010，GLOBECOM 2010. Miami USA，2010 IEEE，2010：1-5

[4] Talieh S T，Mostafa A，Fakhri K，et al. Machine Learning-Based Classification of Encrypted Internet Traffic[C]// 8th International Conference，MLDM 2012. Berlin，Germany，2012：578-592

[5] Zhang Meng，Zhang Hong-li，Zhang Bo. Encrypted Traffic Classification Based on an Improved Clustering Algorithm[C]// International Conference，ISCTCS 2012. Beijing，China，2012：124-131

[6] Du Ye，Zhang Ru-hui. Design of a method for encrypted P2P traffic identification using K-means algorithm [J]. Telecommunication Systems，2013，53(1)：163-168

[7] 赵博，郭虹，刘勤让，等. 基于加权累积和检验的加密流量盲识别算法[J]. 软件学报，2013，24(6)：1334-1345

[8] Menezes A J，Van O P C，Vanstone S A. 应用密码学手册 [M]. 胡磊，王鹏，等译. 北京：电子工业出版社，2005：1-4

[9] NIST FIPS PUB 140-2-2001. Security Requirements for Cryptographic Modules[S]. Washington DC，USA：National Institute of Standards and Technology，2001

[10] NIST SP800-22. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications Revision 1a[S]. Washington DC，USA：National Institute of Standards and Technology，2010

[11] 徐晶，于向军. 基于 FFT 算法的震动信号分析[J]. 工业控制计算机，2005，18(12)：8-9

(上接第 154 页)

在上面的实例中，只需一次协议运行，攻击者就可以成功扮演发起者 C，使 S 误以为客户 C 已经与自己建立连接。

第二种攻击中所提出的安全漏洞，可能会威胁协议的安全性，但我们还未能给出攻击实例。

扮演其它角色的正常主体参与此次协议会话主要取决于其对应的串在丛中的高度与串参数值这两个因素。对于任何角色串，通过我们的强认证测试可以很容易地确定出其它合法主体串的这两个因素值。前者决定扮演协议的其它诚实主体参与此次协议变换的步数，后者决定该诚实主体进行交互消息项的具体值。进而就可以清晰地判断出协议中该角色的认证目标是否可以达到。同时还可通过强认证测试最终确定出其它合法主体串的参数值中不能认证的参数，来指导协议攻击实例的构造。

结束语 本文分析了认证测试类在分析多层嵌套加密协议的局限性，扩展了串空间理论刻画消息项之间及内部结构关系的能力。在此基础上，提出了一种可以分析协议测试组件嵌套加密的通用的认证测试优化方案，并使用串空间理论对其进行了形式化证明，同时通过具体协议实例验证了强认证测试方法在应用范围上的突破及有效性。由于目前认证测试方法在类型缺陷分析方面存在较大局限性，能否进一步改进强认证测试方法，使其能够分析协议类型缺陷攻击，还值得进一步研究[16]。

参 考 文 献

[1] Fabrega F J T，Herzog J C，Guttman J D. Strand space：Why is a security protocol correct[C]// IEEE Computer Society Press Proc. of the IEEE Symp. on Research in Security and Privacy. Oakland，1998：160-171

[2] Meadows C. Open issues in formal methods for cryptographic protocol analysis[C]// Proceedings，IEEE DARPA Information Survivability Conference and Exposition. 2000，1：237-250

[3] Guttman J D，Thayer F J. Authentication tests [C]// Proceedings，2000 IEEE Symposium on Security and Privacy. IEEE Computer Society Press，May 2000：96-109

[4] Guttman J D. Security protocol design via authentication tests [C]// Proceedings Computer Security Foundations Workshop，2002. 15th IEEE. IEEE，2002：92-103

[5] Wang Q，Zhi F H，Ao J H. Design and Security Analysis of Mobile Identity Authentication Protocol[J]. Advanced Materials Research，2012，403：2645-2649

[6] Chen N，Jiang R. Security Analysis and Improvement of User Authentication Framework for Cloud Computing[J]. Journal of Networks，2014，9(1)：198-203

[7] Perrig A，Song D. Looking for diamonds in the desert-extending automatic protocol generation to three-party authentication and key agreement[C]// Proc. of the 13th Computer Security Foundations Workshop. Los Alamitos：IEEE Computer Society Press，2000：64-76

[8] Li Y J，Pang J. Generalized unsolicited tests for authentication protocol analysis[C]// IEEE Computer Society Press Proc. of the 7th Int'l Conf. on Parallel and Distributed Computing，Applications and Technologies. 2006：509-514

[9] Liu J F，Zhou M T. Research and improvement on authentication test's limitation[J]. High Technology Letters，2008，14(3)：266-270

[10] 刘家芬，周明天. 突破认证测试方法的局限性[J]. 软件学报，2009，20(10)：2799-2809

[11] Zhang G，Rong M，Fang Y. One extension of authentication test based on strand space model[C]// Proc. of the 5th International Conference on Wireless Communications，Networking and Mobile Computing. IEEE Press，2009：4506-4509

[12] Zhou Q，Wu X. Extensions to Authentication Test and Its Application[J]. Journal of Zhengzhou University (Engineering Science)，2010，3：014

[13] Wang Q，Zhi F H，Ao J H. Design and Security Analysis of Mobile Identity Authentication Protocol[J]. Advanced Materials Research，2012，403：2645-2649

[14] Muhammad S. Applying authentication tests to discover Man-In-The-Middle attack in security protocols[C]// Eighth International Conference on Digital Information Management (ICDIM)，2013. IEEE，2013：35-40

[15] 余磊，魏仕民. 协议主体密钥在测试组件构造上的性质分析[J]. 计算机工程与应用，2013，49(6)：114-117

[16] Khot R A，Srinathan K，Kumaraguru P. A novel Jigsaw based authentication scheme using tagging[C]// Proc. of the 2011 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011：2605-2614