基于方差分析和支持向量机技术的P2P流量检测

吴 敏 王汝传
(南京邮电大学计算机学院 南京 210003)

摘 要 P2P流量逐渐成为了互联网流量的重要组成部分,在对Internet起巨大推动作用的同时,也带来了因资源过度占用而引起的网络拥塞以及安全隐患等问题,妨碍了正常的网络业务的开展。首先介绍了各种P2P流量识别方法及其优缺点,然后提出一种基于方差分析的P2P流量特征选择方法和基于该方法的支持向量机技术在P2P流量准实时检测中的应用模型。实验结果及分析表明,该方法能较有效地检测P2P流量并具有更好的检测精度。

关键词 对等网络,流量检测,方差分析,支持向量机
中图法分类号 TP393.07 **文献标识码** A

P2P Traffic Identification Using Variance Analysis and Support Vector Machine Algorithm

WU Min WANG Ru-chuan
(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract P2P traffic has taken great portions in the Internet traffic. While having a significant impact on the Internet, it brings serious problems such as network congestion and traffic hindrance caused by the excessive occupation in the bandwidth. The paper firstly introduced methods in identifying P2P traffic and their characters, then put forwards a P2P traffic feature selection method by exploring analysis of variance. Meanwhile a model based on Support Vector Machine (SVM) algorithm was set up to fulfill the quasi-real-time identification of P2P traffic. Experimental results show that the method is efficiency for P2P traffic identification and has a more accurate precision.

Keywords P2P network, Traffic identification, Analysis of variance, Support vector machine

1 引言

随着P2P网络技术在20世纪90年代后期的兴起,P2P流量逐渐成为了互联网流量的重要组成部分。精确地识别P2P流量对于有效地管理网络和合理地利用网络资源都具有重要意义。

目前P2P流量检测技术大致有以下三类:基于端口的检测技术、深层数据包检侧技术和基于流量特征的检测技术[1]。

基于端口的分析方法是在网络流量中探测P2P用户最基本、最直接的方法。但由于现在大多数P2P应用允许用户手动选择随意的端口号来设置默认的端口号或使用随机的端口号,从而使得端口号不可预测,还有一些P2P应用使用默认端口号(例如80端口)来伪装自己的功能端口,因此基于端口号的分析方法的效率变得很差。

深层数据包检测技术,可深入检测其数据包中的有效载荷,即通过应用层数据包的正则表达式的匹配来完成探测工作,以确定特定的P2P应用。当前许多商业P2P应用识别方案都基于该方案,如L7-filter,Cisco'ls PDMLk等。该方法识别准确度可以达到95%,实现简单,维护方便。但该方法是高资源消耗的,在带宽越高的网络,检查时所需要的开销和资源就越多,而且由于必须读取处理所有网络流量,会严重地增加网络设备负担甚至会导致网络的崩溃,因而不适合大型网络。另外该方法对加密P2P流量捕获能力弱,对新的P2P应用必须升级后才能检测且该方法容易和隐私保护法律条款产生冲突。

基于流量特征的检测技术利用P2P在传输层表现出来的流量特征来发现P2P应用。这类方法借用了统计学领域通用的一些概念,分析传输层的信息,不需要任何关于应用层协议的信息,几乎不需要任何额外的软件或者硬件并具有较强的加密和未知P2P流量的捕获能力,因而近年来关于流统计方式测量P2P流量得到了国内外广泛的关注,被认为是最有前途的一种方法。目前主要包括以下几种识别方式:{IP,port}识别、TCP/UDP端口识别、BlockSize识别、基于会话(session)分类的识别、双向识别、流统计状态的识别等等。

实际上,从模式识别的角度而言,P2P流量的识别过程可看作是一个二分类问题:即对流量数据进行分类,分为P2P流和非P2P流。文献[2,3]提出了应用机器学习的方法进行P2P流的检测,例如应用贝叶斯网络、BP神经网络、聚类分

到稿日期:2009-09-10 返修日期:2009-12-10 本文受国家自然科学基金(60973139,60773041),江苏省自然科学基金(BK2008451),省级现代服务业发展专项资金,江苏高校科技创新计划项目(CX09B_153Z,CX08B-086Z),南京邮电大学青蓝工程项目(NY206034,NY208011)和江苏省六大高峰人才项目(2008118)资助。

吴 敏(1976—),女,博士生,讲师,主要研究方向为P2P技术、流量检测、分布式计算、计算机密码学、人工智能;王汝传(1943—),男,教授,博士生导师,CCF会员,主要研究方向为计算机软件理论、移动代理技术、分布式计算和网格计算。

析、支持向量机等方法，但是如何从流特性中获取最有效标识P2P流的特征，建立识别P2P流量的特征库，仍然是该领域内研究的热点。本文提出一个基于方差分析的P2P流量特征选择方案，用于提取对P2P流最具有显著影响作用的特征集合，并将支持向量机技术应用到流量检测的分类问题中来，根据P2P流量的流特性集，事先离线获得大量训练数据，输入到支持向量机中构建最优分类面并以此作为网络P2P流量的测量方法。

本文第2节阐述了多元方差分析方法，提出了一个基于方差分析的P2P流量特征选择方案；第3节描述了支持向量机方法，提出了一个基于支持向量机的P2P流量检测模型；第4节分析了实际实验中应用支持向量机进行P2P流量检测的运行情况，包括实验所用数据、应用方差分析后的特征集合，流量识别的准确度及误检率和漏检率的分析；最后总结全文，并指出下一步研究工作。

2 基于方差分析的P2P流量特征选择方案

2.1 单因素方差分析

方差分析[4]（Analysis Of variance，缩写为ANOVA）是统计学中常用的数据处理方法，其目的是通过数据分析找出对事物有显著性影响的因素。单因素方差分析方法讨论的是在只有一种实验条件下，实验条件对实验结果是否会有显著影响，即将一组数据的总变动量，依可能造成变动的因素分解成不同的部份，并且以假设检定的方法来判断这些因素是否确实能解释数据的变动。目的是通过判断随机变量在只有一种处理因素条件下，经过几种不同处理水平之后得到的处理均值是否相等，从而检验出实验方法对实验结果是否产生显著性的影响[11]。其前提假设是：各样本是相互独立的随机样本；各样本来自正态分布总体且各总体方差相等。其思路是将所有样本的总变动分成两个部分，一部分是组内变动（within groups），代表本组内各样本与该组平均值的离散程度；另一部分是组间变动（between groups），代表各组平均值关于总平均值的离散程度。将这两个变动部分除以它们所对应的自由度，即得到均方差。然后，用组间变动的均方差除以组内变动的均方差，根据统计学原理，组间差异与组内差异的比值呈F分布，从而得到F检验值，根据统计值对应的显著性水平就可以判断不同组间是否有显著性的差异。即：

$$F=\frac{MS_b}{MS_w}=F(df_b,df_w) \tag{1}$$

式中，MS_b 表示组间差异，等于组间平方和 SS_b 除以组间自由度 df_b，MS_w 表示组内差异，用组内方差，等于组内平方和 SS_w 除以组内自由度 df_w，即：

$$MS_b=\frac{SS_b}{df_b} \tag{2}$$

$$SS_b=n_1(\bar{X}_1-\bar{X}_t)^2+n_1(\bar{X}_2-\bar{X}_t)^2+\cdots+n_k(\bar{X}_k-\bar{X}_t)^2 \tag{3}$$

式中，$df_b=N-1$。

$$MS_w=\frac{SS_w}{df_w} \tag{4}$$

$$SS_w=(n_1-1)S_1^2+(n_2-1)S_2^2+\cdots+(n_k-1)S_k^2 \tag{5}$$

式中，$df_w=N-K$，$\bar{X}_k$ 表示各个组内数据的算术平均数，$\bar{X}_t$ 表示所有样本数据的算术平均数，k 表示组数，N 表示样本数。

这样，我们就可以用F检验统计量来表示组间差异和组内差异大小比较的比值，并利用F检验来检验组间差异与组内差异是否相等。如果组间差异与组内差异的比值比较大，大到超过F抽样分布上的显著性水平的临界值，那么总体的差异就可认为是来自于各组总体平均数之间的差异，即各组总体平均数之间确实有本质差异；如果组间差异与组内差异的比值比较小，小于F抽样分布上的显著性水平的临界值，那么总体的差异就可认为是由抽样误差引起的，即各组总体平均数之间没有本质差异。

2.2 应用方差分析进行P2P流量特征选择

方差分析的目的是检验并分析各种流统计特性对P2P流量的作用，使用方差分析方法去检验它们对P2P流和非P2P流是否会产生显著性影响，从而判断出哪些统计量对隐藏信息较为敏感。这一部分属于离线分析。对网络中采集的数据进行预处理，提取249个流相关统计信息，包括包大小、包延迟及各种TCP协议信息等，用方差分析对各个统计量进行检验，选择出对P2P流量更敏感的统计量作为区分P2P流量的特征向量元素。由于流量是随机选择的，其各种统计量可看作互相独立，且其取值范围服从正态分布；因此我们可以用单因素方差分析去检验这些统计量。

例如，给定显著性水平 $\alpha=0.05$，单因素方差分析结果表明任何 $p<0.05$ 的检测结果都是显著的，因此拒绝零假设H0，各处理水平均值相等；转而支持备择假设H1，至少有一对处理水平均值不等。满足备择假设H1的统计量被选作P2P流量的特征值。

在所提取出的全部流统计量中，只有一部分能够较好地反映出是P2P流量和非P2P流量的统计特征变化。将方差分析所选取出的这些统计量集合作为P2P流量的统计特征向量元素，根据特征样本集使用机器学习的方法进行P2P流量检测，必然能提高检测率。例如应用基于核技巧的支持向量机去训练样本数据，找到优化超平面后，再运用SVM检测任何给定流量中是否包含有P2P流量。

3 基于支持向量机的P2P流量检测模型

3.1 支持向量机简介

支持向量机是统计学习理论中最年轻的内容，也是最实用的部分。其核心内容是在1992至1995年间提出的，目前仍处于不断发展阶段[5]。SVM是从线性可分情况下的最优分类面发展而来的，基本思想可用图1的二维情况说明。图中，实心点和空心点分别代表两类样本，H为分类线，H1，H2分别为过各类中离分类线最近的样本且平行于分类线的直线，它们之间的距离叫做分类间隔（margin）。所谓最优分类线就是要求分类线不但能将两类正确分开（训练错误率为0），而且使分类间隔最大。分类线方程为：

$$w\cdot X+b=0 \tag{6}$$

$(x_i,y_i),i=1,2,\cdots,n,y\in\{+1,-1\}$ 满足 $y_i[(w\cdot X_i)+b]-1\geqslant 0,i=1,2,\cdots,n$ (7)

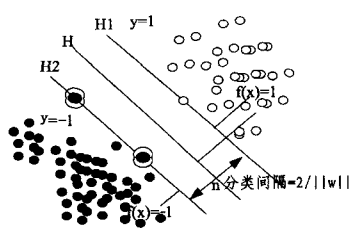

图1 SVM最优分类面

此时，分类间隔为 $2/||w||$，使间隔最大等价于使 $||w||^2$ 最小。满足条件式(1)且使 $||w||^2$ 最小的分类面称为最优分类面，H1，H2 上的训练样本点称为支持向量。使分类间隔最大实际上就是对推广能力的控制，这是 SVM 的核心思想之一。

3.2 一个基于支持向量机的 P2P 流量检测模型

整个模型的系统工作过程分为 3 个模块：基于方差分析的特征预处理模块、基于支持向量机的流量训练阶段和基于 svm 的流量检测阶段，最后根据检测的结果进行流量控制，如图 2 所示。

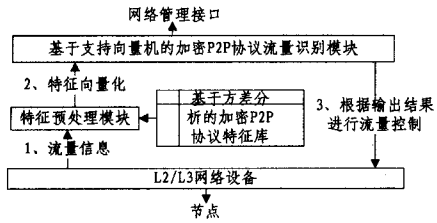

图 2 基于方差分析的 svm 的 P2P 流量识别模型

其中特征预处理模块指根据方差分析方案得到的 P2P 流量特征集合，对已知的正常流量数据、P2P 流量数据根据子项目 1 得到的特征进行数据处理，将之转化为数字向量形式，并归一化使其能量为 1，裁减尺度范围为[−1　1]。

基于支持向量机的训练阶段，将预处理后的统计特征数据作为训练支持向量机的依据，通过训练得到支持向量机的最优分类面。svm 的训练阶段指根据特征预处理后的训练向量训练支持向量机，调整 SVM 参数到最佳状态，得到 Vapnik 意义下的最优分类面 $f(x)=0$。

在识别阶段，将预处理后的流量统计特征数据通过该支持向量机分类器进行检测。主要根据判决函数式对这些数字向量进行分类，并将分类结果提交给流量分析和控制模块，作出对 P2P 流量的控制处理。

其识别流程如图 3 所示。具体步骤如下。

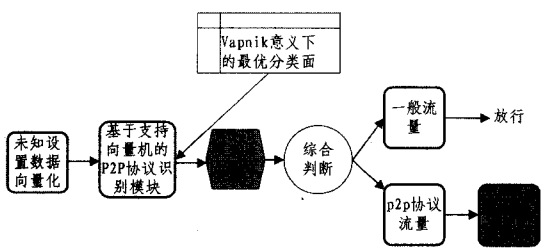

图 3 P2P 协议的网络层流量识别流程

(1) 根据方差分析提出的 P2P 协议的流量特征对实际预测流向量的特征进行预处理。

(2)判断该向量是否已经存在于已经识别的 P2P 记录表中，如已存在，则转 P2P 流量控制模块。

(3)使用基于支持向量机进行未知网络流量性质的预测，如果模式匹配成功，则综合判断为是某 P2P 流量，更新 P2P 记录表，并根据网络实际情况执行 P2P 流量的控制策略。

4 实验结果分析

为了验证本文提出的基于方差分析的支持向量机技术在 P2P 流量识别的有效和准确性，我们设计了一个原型系统，并将之部署在校园网的边缘路由器上。其拓扑图如图 4 所示。

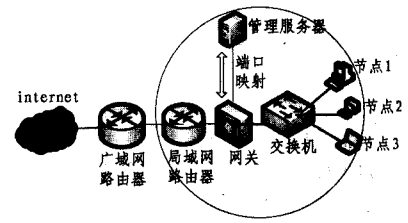

图 4 P2P 流量识别模型部署

实验所用管理服务器主要用于离线训练基于支持向量机的 P2P 检测模型及在线 P2P 流量识别，该服务器型号为 Dell Powereddcdge 2850((Xeon 3.0GHz * 2/256MB/73GB))，其中特征选择、离线模型训练和准实时流量识别在 Weka[6] 中完成。

实验中运行的 P2P 应用包括：BitTorrent、eMule、迅雷、PPlive、Maze 及 Web 应用与 FTP 服务。

具体实验步骤如下。

(1)采集网络流量。

利用 netmate[7] 工具采集网络数据包，按流进行分组(即将具有相同源 IP 地址、源端口号、目的 IP 地址、目的端口号和协议的分组包组成 1 个流)并标注出是否属于 P2P 应用，用“1”表示是，“−1”表示否，考虑到数据量较大，这里采用了文献[8]中提到的半监督聚类方法对流量进行分类。共形成十个原始数据文件。

(2)利用 netmate 从流中提取相关属性信息。

这里只考虑了 TCP 流，不考虑 UDP 和 ICMP 流。其中每个 TCP 流可以提取包括基于 IP 报文首部的协议、流时长、流的字节数、流的包数、前向或者后向分组包长度、包到达时间间隔(最大、最小、平均值、标准差)等共 248 个特征，形成训练数据源，即流量数据向量化及预处理部分。

(3)利用 Weka 软件对训练数据源进行多元方差分析，形成特征集。

为减少计算量和资源(如内存)的消耗，并提高分类速度，运用多元方差分析方法从 10 个数据集的所有流量属性特征中选择了最有代表性的属性特征形成子集。分析发现每个数据集的属性特征集合内容有所不同，表现强度也不尽相同，但是仍然具有较好的稳定性，我们手动选择了 5 个行为特征，这 5 个至少在三分之一的特征子集中出现，且每一个特征子集最起码包含它们中的三分之一。这样做是希望寻找到相对更稳定、更独立于点对点之间的特征[10]。表 2 是单因素方差分析检验结果，第一栏为所选的 5 个特征量。选定显著性水平 0.05，$p<0.05$ 被认为是对隐藏信息有显著影响的特征量。

表 2 单因素方差分析检验结果

属性(流)		组差异	F 值	Sig
包大小标准方差	组间	6.832E9	298.792	.000
	组内	2.287E7		
包交换频率	组间	3.247E7	473.755	.000
	组内	68535.132		
包个数	组间	5.571E9	27.821	.005
	组内	7.124E8		
总字节	组间	1.39E10	25.369	.000
	组内	5.478E8		
平均包大小	组间	1.373E10	26.976	.000
	组内	5.091E8		

(4)利用 Weka 软件进行支持向量机模型训练，形成训练模型。

这是个离线分析的过程。根据步骤(3)从每条流记录中提取5个特征及所属分类,重新形成新的数据源,进行基于支持向量机的P2P流量识别模型。实际训练时采用了10折交叉验证(10-fold cross validation)方法,即将每一个数据集分成10份,轮流将其中9份作为训练数据,1份作为测试数据,进行试验。每次试验都会得出相应的正确率(或差错率)等结果,并以10次结果的正确率(或差错率)的平均值作为对算法精度的估计。

本文对该P2P流量识别模型通过以下参数进行描述:平均建模时间、平均分类时间、识别准确度(Accuracy)、P2P识别精确率(Precision)和反馈率(Recall)。

识别准确率定义如下:

$$Accuracy=\frac{TP+FN}{TP+TN+FT+FN}\times 100\% \quad (8)$$

P2P识别精确率和反馈率定义如下:

$$Presicion=\frac{TP}{TP+FP}\times 100\% \quad (9)$$

$$Recall=\frac{TP}{TP+TN}\times 100\% \quad (10)$$

式中,TP(True Positive)表示成功检测出的P2P流在总体P2P流量中的百分比,FP(False Positive)表示错误地把非P2P流检测为P2P流量的比率,FN(False Negative)表示正确将非P2P流分类成非P2P流的比率,TN(True Negative)表示将P2P流检测为非P2P流量的比率。

表3显示了识别准确度、P2P识别精确度和反馈率。实验最后合并了P2P真阳性率最高,漏检率及误检率相对较低的3个数据集,用来训练支持向量机,得到P2P流量识别模型。

表3 训练集的识别准确度及误检率和漏检率

测试结果	Accuracy(%)	Precision(%)	Recall(%)
entry01	99.823	94.7	95
entry02	96.2228	100	93.2
entry03	98.9709	98	97.5
entry04	98.7391	100	96
entry05	99.7182	98.7	99.4
entry06	99.3603	100	98.5
entry07	99.5003	97.4	98.8
entry08	99.3585	99	99
entry09	99.2845	93.6	97.9
entry10	99.1313	98.4	98.5
平均值	99.01	97.9	97.38

为了体现特征选择方法的必要性和有效性,我们在数据集1—10上对原始属性和其特征子集建立分类模型,并用十倍交叉验证法获得其分类正确率。表4对比了这两种方式的分类准确率、训练时间和分类时间。

表4 特征选择方法与原始特征集的比较

	AVONA方法	原始属性
平均分类准确率(%)	99.01%	80.83
平均模型训练时间(s)	37	1505
平均流量分类时间(s)	1.5	11

从表4可以看到,基于多元方差分析的算法的建模时间和分类时间都远远小于原始数据集所需要的时间,而且分类准确度得到提高。这说明对属性的合适方法约减,能够提高分类精度,节省运行时间。

(5)准实时的P2P网络流量识别。

使用netmate对捕获的网络流量数据数据包提取步骤(3)中的5个特征后进行基于支持向量机的P2P流量识别,这是一个准实时检测过程即在少量流数据包到达后就能尽可能快地给出检测结果,因而存在一定滞后性。考虑到内存开销和计算复杂度问题,实验中没有捕获每一个流的所有数据包,根据文献[8,9],只是捕获了每个TCP流的前5个包,经过特征选择预处理后作为测试集输入支持向量机进行分类(CPU运行时间约:640us/流),后者根据训练模型判别每一条流的类型并打上标签(CPU运行时间约:20us/流)。为了对比验证该方案的性能,在该服务器上同时部署了Analyzer软件[11],后者是完全通过特征关键字达到P2P流识别目的的。我们在试验中比较了该方案和Analyzer软件在在线检测分类的运行时间上的不同,发现该软件所需要的CPU实际运行时间约为1540us/流,几乎是前者的3倍,这主要是因为后者主要依靠payload来检测P2P流量,它必须读取处理所有网络流量,对网络中每一个截获的数据包都必须进行内容解析匹配,因此需要的开销、资源和运行时间较多。

结束语 本文研究了相关P2P流量检测方法和基于方差分析的支持向量机技术,提出了一种基于这种统计方法的支持向量机技术在P2P流量检测中的应用思路,同时提出了一个基于该方案的P2P识别流量模型。该方法是基于P2P流特征的机器学习方法,克服了传统的基于端口和特征字方法的不足,可以用于识别加密P2P流量。实验表明,该方法能够比较有效地判断P2P流量的显著影响因素,适合于准实时流量检测,而且通过调整系统参数可以实现高精度检测。下一步的研究目标首先要进一步地完善本方案,通过仿真实验和原型系统运行,确定流统计的精确时间间隔等,并对比其他机器学习方法如C4.5,聚类分析的训练和分类性能以寻找最合适的识别方法,然后实现该方案和特征字匹配的融合,从而更准确判别P2P流量并进行应用级分类。

参考文献

[1] Basher N, Mahanti A, Mahanti A, et al. A comparative analysis of web and peer-to-peer traffic[C]// Proceeding of the 17th international conference on World Wide Web. Beijing, China, 2008, 4: 21-25

[2] Chen Zhenxiang, Yang Bo, Chen Yuehui, et al. Online hybrid traffic classifier for Peer-to-Peer systems based on network processors[J]. Applied Soft Computing, 2009, 9: 685-694

[3] 王锐. P2P流量检测技术研究[D]. 长沙:国防科学技术大学, 2006

[4] 詹双环,张鸿宾. 基于小波分解和方差分析的图像信息隐藏盲检测[J]. 电子与信息学报, 2007(6): 1460-1463

[5] (美)瓦普尼克. 统计学习理论[M]. 许建华,等译. 北京:水利电力出版社, 2004

[6] Waikato Environment for Knowledge Analysis(Weka). http://www.cs.waikato.ac.nz/ml/weak/

[7] Zander S. NetMate0.9.4 [EB/OL]. [2008201208]. http://www.ip2measurement.org/tools/netmate/

[8] Erman J, Mahanti A, Arlitt M, et al. Offline / realtime traffic classification using semi-supervised learning[J]. 26th International Symposium on Computer Performance, Modeling, Measurements, and Evaluation, 2007, 64(9-12): 1194-1213

(下转第113页)

节的个数由帧头的 Fill_Byte_Num 字段表明。CRC 是对 Header 和数据域进行校验，其生成多项式采用 IEEE 802.3 CRC-32 标准。由于数据帧均在已建立好的专用信道上进行传输，交换网络不进行干预，因此它无需源地址和目标地址信息。

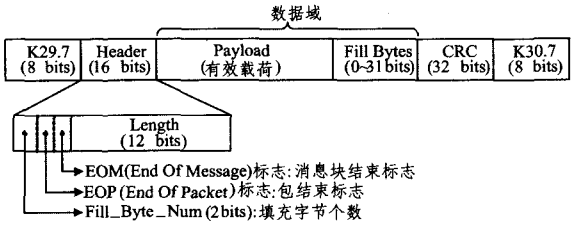

图 2　FCSP-RTF 协议数据帧格式

综上，虽然 FCSP-RTF 帧的控制字段非常少，信息非常简单，但它们能够高效可靠地通信而不会发生混淆。

4　FCSP-RTF 协议通信性能分析

4.1　FCSP-RTF 协议与 FC 协议帧效率比较

已知 FCSP-RTF 和 FC 协议中数据帧的最大长度均为 2k 字节，即 512 个字。图 3 用直观的曲线来说明净荷长度 $i=1,2,\cdots,512$ 时 FCSP-RTF 数据帧效率、FC 协议数据帧最小效率和最大效率的对比情况。

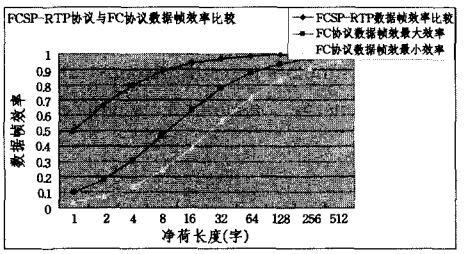

图 3　FCSP-RTF 数据帧与 FC 协议数据帧效率比较

假设数据帧各种载荷长度的概率相同，经计算 FC 协议数据帧的最小平均效率，即带可选头时的平均效率为 85.11%，最大平均效率，即不带可选头时的平均效率为 92.96%；而 FCSP-RTP 数据帧的平均效率则高达 98.86%，比 FC 协议帧的最小平均效率高出 13.75%，而比其最大平均效率还高 5.90%。

4.2　FCSP-RTF 通信效率分析

设 $T_{Setup\text{-}Channel}$，$T_{Data\text{-}Transfer}$ 和 $T_{Remove\text{-}Channel}$ 分别为 FCSP-RTP 建立信道、数据传输和撤消信道时间，$E_{FCSP\text{-}RTF\text{-}Comm}$ 为 FCSP-RTP 的通信效率，则有

$$E_{FCSP\text{-}RTF\text{-}Comm}=\frac{T_{Data\text{-}Transfer}}{T_{Setup\text{-}Channel}+T_{Data\text{-}Transfer}+T_{Remove\text{-}Channel}} \quad (1)$$

式中，$T_{Setup\text{-}Channel}$ 和 $T_{Remove\text{-}Channel}$ 是固定的，它们和通信量无关；$T_{Data\text{-}Transfer}$ 是动态变化的，它与所传输的数据帧数成正比。若建立一次信道之后传输的数据帧越多，即 $T_{Data\text{-}Transfer}$ 越大，则平均每帧所用的信道建立和撤消时间就越小，整个网络的通信效率就越高。当 $T_{Setup\text{-}Channel}+T_{Remove\text{-}Channel}\ll T_{Data\text{-}Transfer}$ 时，$E_{FCSP\text{-}RTF\text{-}Comm}$ 可接近 100%。对于 FCSP-RTP 而言，由于面向的集群系统，如天气预报、基因计算应用，其通信量非常巨大，通常一次数据传输的时间很长，其信道建立时间和撤消时间几乎可以忽略不计，通信总时间绝大部分都用于数据传输，因此其通信效率很高。

而“存储-转发”式网络则不同。由于每帧平均所需的总传输延迟和每帧寻径平均所需的时间与传输的帧数无关，因此效率不会随着通信量的增大而提高，而是保持一个相对稳定的值。

由以上分析可以得知，在集群等通信需频繁使用控制帧的网络中，FCSP-RTP 帧级通信协议具有比 FC 协议更优秀的通信性能。

结束语　本文通过分析 FC 通信协议在高性能集群网络应用中的不足，在“信令寻径式先锋网”基础上提出了信令降频方法，设计了基于此方法的通信协议 FCSP-RTP。FCSP-RTP 协议具有帧格式短、控制简单、通信效率高等特点，能够有效地对短消息数据包通信进行优化，对提高整个集群系统的性能具有重要意义。信令采用降频方法进行编码，简化了网络中交换机的结构，使交换机无需信号降频器件就能够识别高频信令信号并做出正确处理。实验表明，FCSP-RTF 是一种适合集群等通信中需频繁使用控制帧系统的高效率、低开销的通信协议。

参 考 文 献

[1] ANSI. Fiber Channel Physical and Signaling Interface Standard (FC-PH)[S]. Rev 4.3, June 1994: 17-33

[2] ANSI. Fiber Channel Framing and Signaling Standard (FC-FS)[S]. Rev 1.8, March 2003: 19-36

[3] ANSI. Fiber Channel Fabric Standard[S]. Rev 3.3, Oct. 1997: 9-16

[4] Yang Yang, Blum R S. Energy-efficient Routing for Signal Detection under the Neyman-Pearson Criterion in Wireless Sensor Networks[C]// Proceedings of the 6th International Conference on Information Processing in Sensor Networks. 2007(4): 303-312

[5] 张万生，王健. 高速高精度数据采集系统设计与实现[J]. 核电子学与探测技术，2007(9): 809-812

[6] 张曦林，刘海涛，李道本. 高速率的符号干扰传输及其编码的检测译码[J]. 北京邮电大学学报，2007(12): 60-64

[7] 杨书杰，郭宗莲. 基于 FPGA 的高速数据采集系统研制[J]. 电力自动化设备，2008(6): 93-96

[8] 王玮，冯永茂，丁铁夫. 基于 FPGA 的高速数据传输芯片控制器设计[J]. 电子器件，2008(6): 845-848

[9] 倪永军，徐忠. 光纤通道协议技术分析与研究[J]. 信阳师范学院学报，2006(10): 442-445

[10] 雷艳静. 面向 MNWF 的信令寻径式光纤通道先锋交换网研究[D]. 西安：西北工业大学计算机学院，2009: 14-32

(上接第 91 页)

[9] Li Wei, Canini M, Moore A W. Efficient application identification and the temporal and spatial stability of classification schema[Z]. Elsevier North-Holland, Inc. New York, NY, USA, April 2009: 790-809

[10] Canini M, Fay D, Miller D J, et al. Per flow packet sampling for high-speed network monitoring[C]// Proceedings of the first International Conference on Communication Systems and Networks(COMSNETS). January 2009: 1-10

[11] Risso F, Baldini A, Bonomi F. Extending the NetPDL Language to Support Traffic Classification[C]// Global Telecommunications Conference, 2007. GLOBECOM apos;07. Nov. 2007: 22-27