一种基于损失函数的SVM算法在P2P流量检测中的应用

吴 敏 王汝传
(南京邮电大学计算机科学学院 南京 210003)

摘 要 P2P流量逐渐成为互联网流量的重要组成部分,但在对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,妨碍了网络业务的正常开展。首先介绍了各种P2P流量识别方法及其优缺点,然后提出一种基于损失函数机制的支持向量机算法,用于实时P2P流量检测,并构建了一个基于本算法的检测控制模型。实验结果显示,该算法更符合P2P流量的实际检测要求,具有更好的检测精度。

关键词 对等网络,流量检测,损失函数,支持向量机
中图法分类号 TP393.08 **文献标识码** A

Risk Function-based SVM Algorithm and its Application in P2P-traffic Detection

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

Abstract P2P traffic has taken great portions in the network 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 Support Vector Machine (SVM) algorithm based on risk function which can be used to identify P2P traffic. Finally, a model was set up to fulfill the identification of P2P traffic. Experimental results show that the method is more suitable for practical requirements of traffic identification and has a more accurate precision.

Keywords P2P, Traffic identification, Risk function, 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3类:基于端口的检测技术、深层数据包检测技术和基于流量特征的检测技术。

基于端口的分析方法[1]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很差。

深层数据包检测技术[2],通过深入检测其数据包中的有效载荷来进行检测,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应用识别方案如L7-filter, Cisco'ls PDMLk等都基于该方案。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流量的捕获能力弱,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[3]是利用P2P在传输层表现出的流量特征来发现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流量的方法得到了国内外的广泛关注[4-6],被认为是最有前途的一种方法。目前主要有以下几种识别方式:{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。

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[7]采用该思想提出了一种新的基于支

到稿日期:2009-03-16 返修日期:2009-06-14 本文受江苏省自然科学基金(BK2008451),国家高科技863项目(2007AA01Z404,2007AA01Z478),南京市高科技项目(2007软资127),江苏省博士创新项目(CX08B-086Z)和南京邮电大学青蓝工程项目(NY206034,NY208011)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,E-mail:wumin@njupt.edu.cn;
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。

持向量机的 P2P 流量识别办法，将支持向量机技术应用到流量检测的分类问题中。该方法通过 P2P 流量的{IP，port}特性，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。然而，该方法显然没有考虑两类误判在造成损失上的差异[9,11]，只考虑了最优分类面上的样本成为各类别的概率是相同的，没有考虑其损失是不同的，即将非 P2P 流量错误地识别为 P2P 的流量，从而错误地实施对该流量的控制，带来的损失远大于将 P2P 流量识别为非 P2P 流量带来的损失。因此本文在文献[7]的基础上，定义了一种代价敏感的分类算法，即基于损失函数的诊断可信度函数，并在特征空间中对传统支持向量机技术形成的最优分类面进行重新设计，使其符合损失等的要求。

本文第 2 节描述这种基于损失函数的 SVM 机制；第 3 节提出一个基于该方法的 P2P 流量检测模型；第 4 节通过实验结果证实该方法的性能效率及识别精度，包括误检率和漏检率的分析；最后总结全文，指出下一步的研究工作。

2 基于损失函数的支持向量机技术

2.1 支持向量机简介

支持向量机(SVM)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 年至 1995 年间提出的，目前仍处于不断发展阶段[10]。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：

$$w \cdot X + b = 0 \tag{1}$$

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

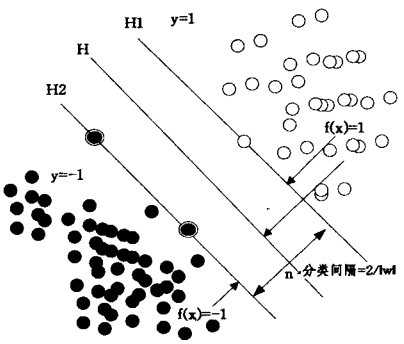

图 1 SVM 最优分类面

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。根据统计学习理论，求最优分类面的问题可转化为优化问题[10]，即在式(1)的约束下求函数(2)的最小值。这是一个二次规划问题：

$$\phi(w) = \frac{1}{2}||w||^2 = \frac{1}{2}(w \cdot w) \tag{2}$$

可转化为在约束条件 $\sum_{i=1}^{n} a_i y_i = 0, a_i \geqslant 0, i=1,2,\cdots,n$ 之下求式(3)的最大值：

$$Q(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j (x_i \cdot x_j) \tag{3}$$

其中，$a_i > 0$ 为 Lagrange 系数。若 a_i^* 为最优解，求解上述问题后得到的最优分类函数是：

$$\mathrm{sgn}((w \cdot x)+b) \tag{4}$$

其中，sgn()为符号函数，b^* 是分类的阈值，可以由任意一个支持向量用式(4)求得，或通过两类中任意一对支持向量取中值求得。对于给定的未知样本 x，只需计算 $\mathrm{sgn}((w \cdot x)+b)$，即可判定 x 所属的分类。对于线性不可分情况，用内积 $K(x_i \cdot x_j)$代替最优分类面中的点积，就相当于把原特征空间变换到了某一新的特征空间，此时优化函数变为：

$$w(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j}^{n} a_i a_j y_i y_j K(x_i \cdot x_j) \tag{5}$$

相应的判别函数也应变为：

$$f(x) = \mathrm{sgn}\{\sum_{i=1}^{n} a_i * y_i K(x_i \cdot x) + b^*\} \tag{6}$$

2.2 引入损失函数的 SVM 机制

考虑图 1，结合上面的讨论可以看出，由于支持向量的存在，对于检测样本 X，如果存在 $f(X) \leqslant -1$ 或 $f(X) \geqslant 1$，则可对其给出明确的诊断结果。

而在区间 $-1 < f(X) < 1$，特别是在 $f(X)=0$ 的邻域内，诊断的结果存在一定的不确定性，将该区域称为可疑诊断区。为了定量描述可疑诊断区诊断结果的不确定性程度，现引入诊断可信度函数。定义诊断可信度 $T(X,y)$是衡量将当前运行状态 X 诊断为 y 的可信程度。显然针对两类诊断问题{正常 1；故障 -1}，对任一诊断样本都有：

$$T(x,1) + T(x,-1) = 1 \tag{7}$$

考虑到 SV 对诊断结论的支持及分界面的含义，可定义一种诊断可信度函数，如图 2 所示。具体的表达式为：

$$T(x,1) = \frac{1}{2}[1+f(x)] \tag{8}$$

$$T(x,-1) = \frac{1}{2}[1-f(x)] \tag{9}$$

显然，在分类面上 $T(x,1)=T(x,-1)$，因此，$f(x)=0$。

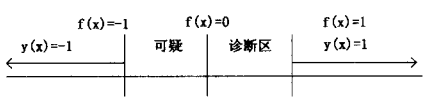

图 2 $f(x)$与诊断决策关系示意图

传统的由 Vapnik 提出的 SVM 理论没有考虑两类误判在造成损失上的差异[10]，即只考虑了最优分类面上的样本成为各类别的概率是相同的，没有考虑其损失是不同的。而在实际问题中，如故障诊断、流量识别等，有时需要考虑一个比误判率更为广泛的概念——风险。例如，对于 P2P 流量检测，将正常流量错误地识别为 P2P 的流量，从而错误地实施对该流量的控制，带来的损失远大于将 P2P 流量识别为正常流量带来的损失。因此，决策规则中必须考虑到不同的分类错误引起的损失，可以用决策表的形式表示出来。表 1 为关于 P2P 流量识别两类问题的决策表。

表 1 P2P 流量决策表

决策	实际状态	
	P2P 流量	正常流量
P2P 流量	L_{11}	L_{12}
正常流量	L_{21}	L_{22}

其中，$L_{11}=L_{22}=0$，即正常分类时不存在选择。而对于流量识别、故障诊断、癌细胞识别等问题，一般认为 $L_{21}\geqslant L_{12}$，即把非 P2P 流量错误地识别为 P2P 的流量，从而错误地实施对该流量的控制，带来的损失远大于将 P2P 流量识别为非 P2P 流量带来的损失。但是，(L_{21},L_{12}) 的选择并非是随意的，实际中损失函数的确定往往要根据分析错误决策造成的损失的严重程度，综合经济性、可靠性及专家经验等因素共同确定。

引入损失函数后，记 $T'(x,1)$ 为考虑损失样本 X 属于第一类的可信度函数，$T'(x,-1)$ 表示考虑损失样本 X 属于第二类的可信度函数。此时

$$T'(x,1)=\frac{L_{21}T(x,1)}{L_{12}T(x,-1)+L_{21}T(x,1)} \tag{10}$$

$$T'(x,-1)=\frac{L_{12}T(x,-1)}{L_{12}T(x,-1)+L_{21}T(x,1)} \tag{11}$$

显然，在分类面上 $T'(x,1)=T'(x,-1)$，因此

$$f(x)=\frac{L_{12}-L_{21}}{L_{12}+L_{21}} \tag{12}$$

式(12)为修正后的最优分类面。

3 基于改进 SVM 机制的 P2P 流量检测模型

3.1 检测模型

基于损失函数的 SVM 的 P2P 流量识别模型如图 3 所示。

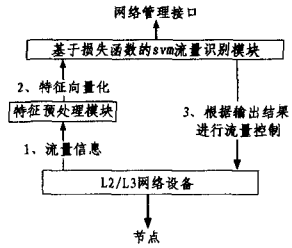

图 3 基于损失函数的 SVM 的 P2P 流量识别模型

整个模型的系统工作过程分为两个阶段：训练阶段和检测阶段。在训练阶段，对已知的正常流量数据、P2P 流量数据进行特征处理，将之转化为数字向量形式，作为训练支持向量机的依据。选择合适的 L_{21} 和 L_{12} 作为损失函数，得到基于损失函数的支持向量机的最优分类面。在检测阶段，将 L_2 传输层或 L_3 网络层采集的数据包提交给流量分析和控制模块，该模块先将未知流量数据处理成数字向量的形式，然后通过该改进的支持向量机分类器，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。

3.2 特征预处理模块

支持向量机的分类器只能对维数相同的数字向量进行分类，但由于捕获的数据包长度不同，而且很有可能不是数字类型，故必须将原始流量数据转换成支持向量机能够识别的数字向量，所以需要数据预处理器来对大量的数据包进行特征的选取。根据文献[3]，一个识别 P2P 流的具体方法为：检测所有源{srcIP，srcport}和目的{dstIP，dstport}，如果特定连接 IP 数与特定连接端口数相同，就认定为 P2P 流；反之，认定为非 P2P 流。考虑到两类流量源端口变化差异的不明显性，本文改进了文献[7]提出的三维特征向量选取办法，如下：

$$v(flow)=\langle f(src,dst),g(src,dpt)\rangle$$

其中，

$$f(src,dst)=dif(src,dst)/same(src,dst)$$

$$g(src,dpt)=dif(src,dpt)/same(src,dpt)$$

$dif(src,dst)$ 计算与源 IP 有连接的不同目标 IP 的数目，$same(src,dst)$ 计算该目标 IP 与源 IP src 建立的连接数，$dif(src,dpt)$ 计算与源 IP 有连接的不同目标 Port 的数目，$same(src,dpt)$ 计算该目标 port 与源 IP src 建立的连接数。对于 P2P 流量，由于 dif 值很大，$same$ 值偏小，因而其 f 值与 g 值较大，而非 P2P 流量则较小，这样就可以以此作为区分 P2P 流量的依据。

3.3 基于损失函数的支持向量机 P2P 流量识别

3.3.1 SVM 的训练阶段

(1)选择合适的 (x_i,y_i)，$i=1,2,\cdots,n$，$y\in\{+1,-1\}$；

(2)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 $f(x)=0$；

(3)由式(12) 建立适合故障诊断需要的分类面 $f(x)=\frac{L_{12}-L_{21}}{L_{12}+L_{21}}$ 作为实际决策函数。

3.3.2 SVM 的识别阶段

基于损失函数的 SVM 流量识别流程如图 4 所示。具体步骤如下。

图 4 基于损失函数的 SVM 流量识别流程

(1)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识别上的 P2P 记录表中。如已存在，则转 P2P 流量控制模块。

(3)使用基于损失函数的支持向量机进行流量性质的预测。如果是 P2P 流量，则更新 P2P 记录表。

4 实验结果分析

为了验证本文提出的基于损失函数的支持向量机技术在 P2P 流量识别上的有效性和准确性，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 5 所示。

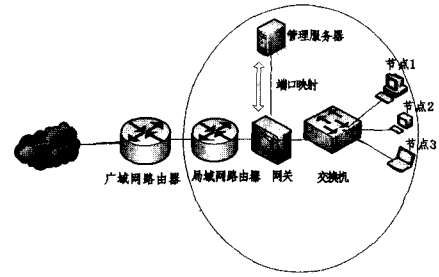

图 5 P2P 流量识别模型部署

实验所用管理服务器主要用于离线训练基于损失函数的 P2P 检测模型及在线 P2P 流量识别。该服务器型号：Dell Poweredcdge 2850(Xeon 3.0GHz * 2/256MB/73GB)，其中离线模型训练和在线流量识别在 Weka[12] 中完成。

实验中取得的训练样本集截表如图 6 所示，其中前 14 条记录基本上为 BT 流量，后 9 条记录为 Web 流量，可以发现 P2P 流量的 $v(flow)=\langle f(src,dst),g(src,dpt)\rangle$ 较后者大，说

明实验中以该向量作为区分 P2P 流量和正常流量是可行的。

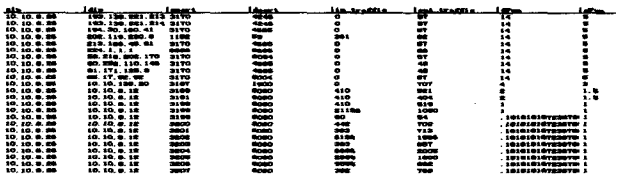

图 6　训练数据表分析

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。实验所用服务器性能：Dell Poweredcdge 2850（Xeon 3. 0GHz * 2/256MB/73GB），训练模型和测试数据集在 Weka 中完成。

具体实验步骤如下：

(1)利用 netmate[13] 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[14]中提到的半监督聚类方法对流量进行分类。采样流量数据的采集是每隔 1min 采集一次数据，采集的数据流量分成 3 个数据集，共形成 2 个原始数据文件，分别命名为 enry01. final. arff，enry2. final. arff，enry3. final. arff。

(2)从数据源文件中提取流相关信息，根据文献[7]提出的三维特征向量进行数据的向量化，形成训练和分类数据源，即流量数据向量化及预处理部分。

(3)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。实验中采用的支持向量机核函数为 RBF 核函数，采用了 10 折交叉验证(10-fold cross validation)方法，即将每个数据集分成 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。表 2 显示了当(L_{21}，L_{12})等于(1，2)时的识别精度及误检率和漏检率。其中 P2P 真阳性率指成功检测出的 P2P 流在总体 P2P 流量中的百分比，误检率 FP 表示错误地把非 P2P 流量检测为 P2P 流量的比率，漏检率 FN 表示错误地把 P2P 流量检测为非 P2P 流量的比率。实验最后采取了 P2P 真阳性率最高、漏检率及误检率相对较低的数据集 1 来训练支持向量机，得到 P2P 流量识别模型。

表 2　当(L_{21}，L_{12})等于(1，2)时训练集的识别准确度及误检率和漏检率

测试结果	分类准确度(%)	P2P 真阳性率(%)	P2P 漏报率 FN(%)	P2P 误报率 FP(%)
entry01	99. 823	94. 7	1. 3	6
entry02	96. 2228	1	2. 1	5. 2
entry03	98. 9709	98	0. 8	6. 7
平均值	98. 3388	97. 56	1. 4	5. 9

为了更好地分析方案识别准确性，本文对比了一般 SVM 分类后得到的最优分类面，即(L_{21}，L_{12})＝(1，1)(如图 7 所示)及当(L_{21}，L_{12})等于(1，2)时的最优分类面(如图 8 所示)。其中 P2P 流量用实心点表示，正常流量用空心点表示。

从以上 2 种不同(L_{21}，L_{12})的分类结果可以看出，随着 L_{21}/L_{12} 的增大，也就是说，随着将正常误判为 P2P 的损失函数值的增大，分类面向 P2P 方向偏移，减少了将非 P2P 的正常流量识别为 P2P 流量的概率。从实际运行的结果分析来看，当(L_{21}，L_{12})＝(1，1)时的平均分类精度为 95. 667%，但实际误报率却达到 9. 2%，这是因为在其训练模型时没有考虑到错分向量对系统性能的影响。而(L_{21}，L_{12})＝(1，2)时误报率却有较明显的下降，这说明加大对非 P2P 流量错误地识别为 P2P 的流量的惩罚力度的必要性，因为损失函数的引进可以使得支持向量机的经验误差更小(即错分支持向量减少)。但是，(L_{21}，L_{12}) 的选择并非是随意的，算法必须依据合适的损失函数。实际中，损失函数的确定往往要根据所研究的具体问题及分析错误决策造成损失的严重程度，综合经济性、可靠性及专家经验等因素共同确定，在整个训练过程中是固定不变的。

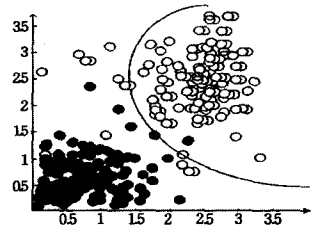

图 7　(L_{21}，L_{12})＝(1，1)

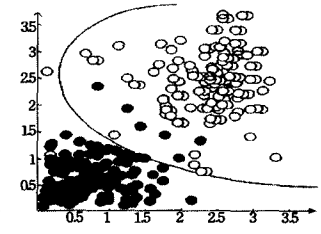

图 8　(L_{21}，L_{12})＝(1，2)

(4)对待分类的数据源进行基于支持向量机的 P2P 流量识别，这是一个准实时的检测过程，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[15]只是捕获了每个 TCP 流的前 5 个包，经过预处理后作为测试集输入支持向量机进行分类，后者根据训练模型判别每一条流的类型并打上标签。为了对比验证方案的性能，在该服务器上同时部署了 Analyzer 软件[16]，后者是完全通过特征关键字达到 P2P 流识别目的的。

图 9 显示了当(L_{21}，L_{12})等于(1，2)时的部分 P2P 软件检测准确率(图 9(a))及检测误检率(图 9(b))。其中检测准确率指成功检测出的 P2P 流在总体 P2P 流量中的百分比，误检率 FP 表示错误地把非 P2P 流量检测为 P2P 流量的比率。从图 9(a)可看出，针对 BitTorrent 的 P2P 文件共享应用，检测准确率较高，其次是 eMule 和 PPLive 的。并且随着时间的增长，检测准确率逐渐上升。这是因为随着采集数据的增多，支持向量机的分析性能更为有效，准确度随之上升；在图 9(b)中，3 个 P2P 应用的误检率都相对比较低，其中 PPLive 的误检率比较高，这是因为 PPLive 是一种混合式 P2P 应用，有一部分流量是 C/S 模式，影响了其流量的数字特征，造成了误检率的上升。

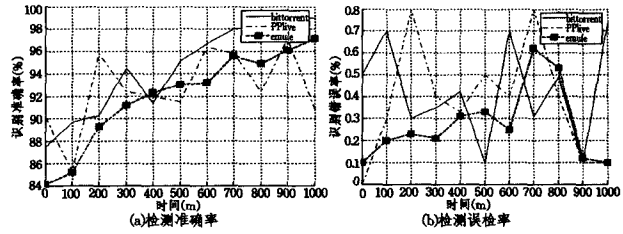

图 9　部分 P2P 软件检测结果

结束语　本文研究了相关 P2P 流量检测方法和有限样本下机器学习的支持向量机技术，提出了一种基于损失函数的支持向量机技术在 P2P 流量检测中的应用思路，同时提出

了一个基于此方案的 P2P 识别流量模型。本方法基于 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，并实际考虑了由于流量误识别带来的风险。实验表明，本方法适合实时流量检测，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，如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、(L_{21}，L_{12}) 的选择等；然后实现方案和特征字匹配的融合，从而更准确地判别 P2P 流量并进行应用级分类。

参 考 文 献

[1] Sen S, Wang J. Analyzing Peer to Peer Traffic Across Large Networks[J]. ACM/IEEE Transactions on Networking, 2004, 12(2): 137-150

[2] Sen S, Spatscheck O, Wang D M. Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures[C]// Proc. of 13th International Conference on WWW. New York, NY, 2004: 512-521

[3] Karagiannis T, Broido A, Faloutsos M, et al. Transport Layer Identification of P2P Traffic[C]// Proc. of ACM SIGCOMM IMC. Taormina, Sicily, Italy, 2004: 121-134

[4] Moore A, Zuev D. Internet traffic classification using bayes ananalysis[C]// Proceedings of International Conference on Measurement and Modeling of Computer Systems. 2005: 50-60

[5] Zuev D, Moore A. Traffic classification using a statistical approach[J]. Lecture Notes in Computer Science, 2005, 3431: 321-324

[6] Constantinou F, Mavrommatis P. Identifying Known and Unknown Peer-to-Peer Traffic[C]// Proceedings of Fifth IEEE International Symposium on Network Computing and Applications. 2006: 93-102

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

[8] Liu Hui, Feng Wenfeng. A Peer-To-Peer Traffic Identification Method Using Machine Learning[C]// International Conference on Networking, Architecture, and Storage. 2007: 155-160

[9] 翟永杰，韩璞，王东风，等. 基于损失函数的 SVM 算法及其在轻微故障诊断中的应用[J]. 中国电机工程学报，2003，23(9)

[10] (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

[11] 肖健华. 智能模式识别方法[M]. 广州：华南理工大学出版社，2006

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

[13] Zander S. NetMate0. 9. 4 [EB/OL]. [2008201208]. http://www.ip2measurement.org/tools/netmate/

[14] 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

[15] Li Wei, Moore A W. A Machine Learning Approach for Efficient Traffic Classification[C]// Proceedings of the IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS). Istanbul, Turkey, 2007

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

(上接第 75 页)

$DA(S).credit \geqslant price(O)$

$id(S) = "SA1"$

$preUpdate(DA(S).credit): credit' = credit - price(O)$

其中，DA(S)表示多域条件下主体 S 的动态属性集。

在 DB 系统中，当主体 SA 使用来自 DV 的动态多域属性信誉卡时，DB 也为 SA 动态地创建某属性 T(如电影票)：

$AttSetPreUpdate\ (DA(S)): DA' = DA \cup \{ticket\}$

$AttValPreUpdate\ (DA(S).ticket): ticket' = 20$

$AttTypePreUpdate\ (DA(S).type): type' = "Dynamic\text{-}Multidomain"$

这样 SA 就可以跨系统使用该属性 T，进而满足了多域交互时刻支持动态属性的系统需求。

结束语 本文对多安全域条件下的动态使用控制问题进行了研究。当前的访问控制模型都是预先定义它们的属性及组件，通过场景示例，论证了这种静态定义不能满足移动的和动态多域的交互需求。为了支持多域的应用环境，我们明确主客体的属性、授权、义务和条件动态特性，在下一代使用控制模型 UCON 的基础之上提出了一种使用控制扩展模型 DAB-UCON，以表示动态授权策略。基于属性定义时间和作用范围，将属性、授权、义务和条件进行分类，扩展后的动态组件有助于动态策略的构建。隐私特性是一个重要的系统安全需求，主体还应具备将属性转移到其它主体的能力。要支持这些特性，还有待于进一步研究。

参 考 文 献

[1] Park J, Sandhu R. The ucon abc usage control model[J]. ACM Transactions on Information and System Security, 2004, 7(1): 128-174

[2] Freudenthal E, Pesin T, Port L, et al. drbac: Distributed role-based access control for dynamic coalition environments[C]// Proceedings of the Twenty-second IEEE International Conference on Distributed Computing Systems (ICDCS). 2002: 411-420

[3] Covington M, Sastry M. A contextual attribute - based access control model[C]// Second International Workshop on Context-Aware Mobile Systems, LNCS 2006. November 2006

[4] Damiani E, Vimercati S, Samarati P. New paradigms for access control in open environments[C]// Proc. of the 5th IEEE International Symposium on Signal Processing and Information. December 2005

[5] Lepro R. Cardea: Dynamic access control in distributed systems [M]. NASA Advanced Supercomputing (NAS) Division, 2003

[6] Hayton R J, Bacon J M, Moody K. Access control in an open distributed environment[C]// 1998 IEEE Symposium on Security and Privacy. May 1998: 3-14

[7] Lin Chuang, Feng Fu-jun. Access control in new network environment[J]. Journal of Software, 2007, 18(4): 955-966

[8] Sandhu R, Park J. Usage control: A vision for next generation access control[C]// Proc. of the 2nd Int'l Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security. LNCS 2776. Berlin: Springer-Verlag, 2003: 17-31