# 基于多模式匹配的网络视频流识别与分类算法

孙钦东[①②] 郭晓军[①] 黄新波[③]

[①](西安理工大学网络计算与安全技术陕西省重点实验室 西安 710048)

[②](西安交通大学智能网络与网络安全教育部重点实验室 西安 710049)

[③](西安工程大学电子信息学院 西安 710048)

**摘 要**：快速发现网络中的视频流是进行网络视频监督及管理的前提与基础。本文通过分析网络视频流数据包的特征，提出了一种基于多模式匹配思想的网络视频流快速发现与分类算法，该算法利用不同视频流的特征建立匹配机，只需对网络数据包进行一次不完全扫描，就可以判断出数据包中是否含有视频流及类型。实验结果表明，与普通的协议解析方法相比，在满足准确性的前提下，所提算法具有更好的时间性能。

**关键词**：网络视频流；协议识别；多模式匹配

**中图分类号**：TP393 **文献标识码**：A **文章编号**：1009-5896(2009)03-0759-04

# Algorithm of Network Video Stream Recognition and Classification Based on Multi-Pattern Matching

Sun Qin-dong[①②] Guo Xiao-jun[①] Huang Xin-bo[③]

[①](*Key Lab for Network Computing and Security of Shaanxi Province, Xi'an University of Technology, Xi'an* 710048, *China*)

[②](*MOE Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an* 710049, *China*)

[③](*College of Electronics and Information, Xi'an Polytechnic University, Xi'an* 710048, *China*)

**Abstract**: Identifying video stream quickly from the network flow is the foundation and prerequisite of network video surveillance and management. According to the analysis of characteristics of network video streams, a fast recognition and classification algorithm of network video stream based on multiple pattern matching is proposed, which constructs the matching machine with the features of different video streams. The proposed algorithm could determine whether the network packet contains video stream and type only by one incomplete scan. The experimental results show that the proposed algorithm has higher efficiency than traditional protocol analysis method under the requirement of accuracy.

**Key words**: Network video stream; Protocol identification; Multi-pattern matching

## 1 引言

网络视频具有感官性强、信息丰富等特点，目前已成为继文字和图片之后，互联网信息传播的主要形式，目前在我国网络视频的使用率为 71%，用户量已经达到 1.8 亿人，是中国网络应用中的第四大网络应用[1]。如何从海量的网络数据中快速发现视频数据，是进行网络视频服务质量监测、网络流量统计、网络视频用户行为分析及视频内容监管等的前提和基础。

目前针对网络视频内容的监控，大多数研究针对的是静态数据动态查询，还无法对网络视频进行实时监控[2]。其中最大的问题之一是如何在高速网络中实时发现视频流，并将其捕获。目前我国国际出口带宽达 493,729Mbps，在如此高速网络链路中，一方面不同网络应用的数据流数量惊人，另一方面其中视频流的编码类型也极多，要从如此大量的数据中快速发现网络视频流，必须考虑视频流判断依据及方法，即确定采用何种规则及方法来快速地判断网络中是否出现视频流数据包，以及如何实时对其进行捕获，以进行视频内容的处理。

网络视频服务为应用层服务，其数据传输不仅可采用专有应用层协议，如 RTP[3]，RDT[4]和 MMST/MMSU[5]等；还可采用通用应用层协议，如 HTTP 协议[6]等。因此，对网络视频数据流的发现首先是识别应用层协议。然而，由于通用应用层协议携带多种类型信息，如文本、图片及视频等，仅识别出应用层协议是不准确的，还应当对应用层协议内容进一步判断。此外，由于网络视频流存在着多种不同类型，还应当将不同类型的网络视频流进行区分，以方便视频内容监管等后续的处理。

针对应用层协议的识别，文献[7]提出了一种以协议中出现频率最高的字段作为特征串来识别协议的方法，且采用一个特征串来标识一种协议。文献[8]提出了基于签名字串的方法来识别应用层协议，其主要针对的是 P2P 协议的范围，且需要对整个报文通过匹配多个特征串来识别一种 P2P 协议，

2008-03-18 收到，2008-07-18 改回

陕西省自然科学基金(2007F13)和陕西省教育厅产业化中试项目(08JC09)资助课题

时间效率偏低。文献[9]提出了基于先分类后分组的识别应用层协议及流量的方法，但此方法的本质还是基于某些固定端口的，若对于通过随机选择端口而实现的应用层协议，此方法就缺乏准确性和灵活性。虽然网络视频流属应用层服务，但仅靠识别出其所使用的应用层协议来识别网络视频流是不够准确的。所以，上述文献中所提到的方法不能直接应用于网络视频流发现，但具有一定的借鉴作用。由于对网络视频流的识别不仅涉及对其使用的应用层协议的识别，而且还有可能涉及协议内容的识别，而目前多数研究集中于应用层协议的识别，针对网络视频流发现的方法研究较少，尤其是缺少对流量中的视频流进行高效分类识别方面的研究。

本文通过分析网络视频流交互过程的特征，以网络视频流对应的关键特征字串为判断依据，设计了一种基于多关键特征字串匹配的网络视频流识分类算法——VPFS(Video Packets Fast Search)算法，实现了对网络视频流的发现及分类，并通过实验对本文所设计的算法性能进行了分析和验证。

## 2 网络视频流特征分析

在前期研究中，作者通过分析正常情况下网络视频流的交互过程，发现网络视频流客户端和服务器之间的交互过程不仅具有阶段性，而且还具有以下特征[10]：

(1)在交互过程中，各阶段的完成均是利用一些协议来实现的，且不同的网络视频流，其交互过程各阶段使用的协议可能相同，也可能不相同。这些协议可以是专用协议，如 RTSP 、RTP/RTCP 协议；也可以是通用协议，如 HTTP 协议、TCP 协议。

(2)对于交互初期阶段使用的流媒体信令协议(Stream Media Signaling Protocol，SMSP)，如 RTSP[11]，HTTP 协议，数据包具有明显特征，即含有 SMSP 协议数据内容中所包含的字符串称之为 SMSP 的关键特征字串(Crucial Representation String，CRS)。一般情况下，不同类型的网络视频流，其使用的 SMSP 不同，其所对应的 CRS 也不同。因此，不同 CRS 对应着不同类型的网络视频流，一种 CRS 可标识一种类型的网络视频流。

(3) CRS 在数据包中的位置具有稳定性。SMSP 作为应用层协议，必然在格式上遵守其自身的协议规范，而 SMSP 的数据包也是由 SMSP 各字段按照其协议格式规范组成的。因此，SMSP 所含有的 CRS 一般会比较固定地出现在数据包中的某个位置上，其位置是稳定的。

(4)SMSP 数据包中包含有后续视频流连接的关键字段特征参数(Characteristic Parameters，CP)，而且使用相同 SMSP 的不同网络视频流，其所包含的 CP 是不同的。

实验表明，上述特征在各种类型网络视频流交互过程中具有普遍性。含有某种 CRS 数据包的出现不仅可以标志与此 CRS 相对应类型网络视频流的出现，而且还可以唯一标识出此网络视频流的类型。因此，对于网络中的任意数据包，对其应用层数据内容进行分析，若能在其某个位置上发现 CRS，则此数据包为 SMSP 数据包，说明网络中存在与此 CRS 相对应的网络视频流交互过程，从而达到了发现网络视频流的目的；并且可通过判断 CRS 来识别出此网络视频流的类型，并可以根据相应的 CP 参数可对所有的网络视频流数据进行重组与分析。

## 3 基于多模式匹配的网络视频流识别算法

协议识别最为常用方法是简单协议分析方法，本文称之为 SPA(Simple Protocol Analysis)方法[12]。该方法对网络中的数据包进行逐层(IP 层，传输层等)解析，当解析到应用层时，对于采用标准端口的应用层协议(如 HTTP 协议使用 80 端口)，虽然可以通过端口号识别此类协议，但是并不一定说明此协议传输的内容为视频数据。例如，HTTP 协议传输的内容可以是文本，图片或视频数据。因此，在此情况下，SPA 方法需要进一步对协议内容进行检查，才能确定内容是否为视频数据。对于没有采用标准端口的应用层协议，SPA 方法则需要逐个将这些协议的特征关键字串在数据包的应用层数据中进行匹配，才能识别出应用层协议，且在此基础上，SPA 进一步检查协议内容以确定是否含有视频数据。当利用 SPA 方法对网络视频流进行发现时，一个数据包可能需要进行多次处理才能确定最终结果，造成 SPA 方法识别效率不高，难于应付大流量环境。因此，针对网络视频流发现，需要研究一种快速，准确而直接的方法。

根据上一节的分析结果，对任何一种网络视频流是可以提取出相应的识别特征的。在此基础上，结合多模式匹配的思想和文献[13]中的 THT 算法，本文设计了一种基于首字符 ASCII 值的视频流数据包多模式匹配算法 VPFS (Video Packets Fast Search)。该算法仅需扫描数据包一次，就可判断出数据包中是否含有某种 CRS，从而缩短了在数据包中寻找 CRS 的时间，提高了对视频流数据包的搜索效率，完成了对网络视频流的识别，并较好地满足了高速网络环境的应用需求。下面是 VPFS 算法中所使用到的数据结构。

设 $C=\{C_0,C_1,C_2,C_i,\cdots,C_n\}$ 为关键特征字串集合，其中 $C_i(0\le i\le n)$ 为某种关键特征字串。对于任意 $C_i$，为其构造一个关键特征字串节点 CRSNode，其定义如下：

```
typedef struct Node{
int    SuccState;
int    CRSLen;
int    CRSMaxpos;
unsigned char    CRS[CRSSize];
struct Node    *NextCRSNode;
}CRSNode;
```

其中 SuccState 表示与本节点中的关键特征字串匹配成功后的状态，同时也是关键特征字串类型的标识号；CRSLen 表

示 CRS 长度；CRSMaxpos 表示 CRS 在数据包中的最大偏移位置；CRS[CRSSize] 为存储关键特征字串 $C_i$ 的字符数组；CRSSize 为 CRS 数组长度。* NextCRSNode 指向下一个首字符相同的 CRS(为了解决首字符相同的 CRS 所产生的冲突)。

构造一个完全哈希序列 HashSeq，共 256 个元素，类型为指针。$H_j(0 \le j \le 255)$ 表示 HashSeq 中的第 $j$ 个元素，且 $H_j$ 指向 $C$ 中关键特征字串首字符的 ASCII 值等于 $j$ 的 CRSNode；若不存首字符的 ASCII 值等于 $j$，则将 $H_j$ 设置为 NULL。逐次将 CRS 节点添加进 HashSeq，从而构造视频流特征匹配机。

算法匹配过程如下。判断待识别数据包当前的字符 ASCII 值对应的 $H_j$ 是否为 NULL。若 $H_j$ 为 NULL，则继续向后判断数据包的下一个字符；若 $H_j$ 不为 NULL，则将从当前位置开始的字符与 $H_j$ 所指向的 CRSNode 中的 CRS 进行匹配，当匹配成功且满足当前位置小于 CRSMaxpos 时，说明此数据包为 SMCP 数据包，并返回 CRS 类型标识号 SuccState；当匹配失败时，则继续向前扫描下一个字符。可以看出，通过对待识别数据包的一趟扫描，可以对其进行所有特征串的搜索，而且能识别出该数据包的视频流类型。

从上述过程可以看出，VPFS 算法所使用的视频流特征匹配机的构造简单快速。空间使用方面，每个 CRS 结点占用空间为 16+MSIZE，完全哈希序列占用 256×4byte。则整个匹配机所占用空间 $\Omega$ 为

$$W = 256 \times 4 + \mathrm{sum}(16 + \mathrm{MSIZE})$$

其中 sum 为 CRS 数量。

假设有 1000 种类型的视频流，CRSSize 最大值取为 20，则整个匹配机占用的空间仅为 36kB。因此，VPFS 算法的空间性能完全满足实用。

从算法的执行过程可以看出，VPFS 算法具有较高的时间效率，主要体现在如下三个方面：

(1)VPFS 算法只需对数据包扫描一次，即可识别数据包中是否含有 CRS 及其类型，大大减少了扫描数据包的次数，提高了对数据包的处理效率。

(2)通过设置关键特征字串的最大位置偏移量(即 CRSMaxpos)，提高了对 CRS 判断的准确性。在实际的应用中，由于数据包应用层数据中其它位置上也可能出现关键特征字串，但此数据包并不一定就是 SMSP 的数据包。所以，只有 CRS 在数据包中特定位置上出现时，才能说明此数据包是 SMSP 数据包。

(3)本算法不需要扫描完整个数据包的应用层数据就可以达到准确判断数据包中是否含 CRS 的目的，提高了判断效率。在数据包含有某种 CRS 的情况下，本算法至多扫描到 CRSMaxpos，就可以发现此 CRS。而在数据包不含 CRS 的情况下，本算法也至多扫描到 MaxPos 位置，就可以判定数据包中不含 CRS。因此，此算法只需扫描数据包的一部分就可以完成判断任务，缩短了对数据包的处理时间。

另外，算法具有灵活的适用性，当出现新类型的视频流时，只需要将其 CRS 添加进特征集合，对匹配机进行一次重新的构建即可，无需对应用程序及算法进行修改，具有良好的适应性及迁移性。

## 4　实验与结果分析

本节给出实验结果。实验所使用的服务器操作系统为 WINDOWS XP，实验环境为 VC 6.0，处理器及内存分别为 P IV 1.8G /512M，硬盘为 120G。

本文的实验数据来源于局域网中的不同主机访问实际视频网站的过程。本文以每次网络中同时访问两个不同的视频流为例，进行 3 次这样的过程，共访问了 6 个不同的视频网站，同时，利用多台主机同时访问不同的网站作为背景流量。利用抓包软件 Wireshark 分别获取了这 3 次过程中的所有数据包，形成了 3 组实验数据样本，即 Sample1，Sample2 和 Sample3。表 1 给出了此 3 份实验数据的主要参数。其中 S1VS1 和 S1VS2 为访问两个不同在线视频网站上的视频所得到的视频流数据，且包含于样本数据 Sample1 中。由于 S1V1 比 S1VS2 视频播放时间长，且加上抓包工具自身的一些限制，因此，以播放时间较短的 S1VS2 的起始播放时间为抓包开始时间，以其播放结束时间为抓包的终止时间，抓取了这段时间内网络中的所有数据包，得到了实验数据 Sample1。同理，分别获得 Sample2 与 Sample3，它们所包含的视频流分别为 S2VS1，S2VS2 和 S3VS1，S3VS2。6 个视频流所对应的详细信息如表 1，表 2 所示。

**表 1 数据样本的主要参数**

| 实验数据名称 | 文件大小 (Mb) | 数据包总数 | 视频流个数 | 视频流标识及数据包个数 |
|---|---|---|---|---|
| Sample1 | 6.94 | 10168 | 2 | S1VS1：2873<br>S1VS2：1333 |
| Sample2 | 18.6 | 30168 | 2 | S2VS1：6153<br>S2VS2：4829 |
| Sample3 | 39.1 | 52711 | 2 | S3VS1：5060<br>S3VS2：8336 |

**表 2 数据样本的来源**

| 视频流 | 视频源地址 |
|---|---|
| S1VS1 | http://v.youku.com/v_show/id_ce00XMTE5ODEwNjA=.html |
| S1VS2 | http://www.end123.com/play.asp?id=1956&no=1 |
| S2VS1 | http://www.ku6.com/show/wrjVQta-fNKgByx4.html |
| S2VS2 | http://www.9070.com/html/3101.shtml# |
| S3VS1 | http://you.video.sina.com.cn/b/8020111-1290074964.html |
| S3VS2 | http://88ys.com/list/7_3.htm |

经过分析，上述视频流的特征字符串集合为 $C=$ {synacast, RTSP, .flv}，其所对应的 CRSMaxpos 分别为 70、30、100，MaxPos = 100 。S3VS2 对应的视频流关键特征字串为“RTSP”， S1VS1, S1VS2, S2VS1, S2VS2 和 S3VS1 对应的关键特征字串为“.flv”。

实验过程中分别使用普通的协议解析方法 SPA 及本文提出的 VPFS 算法对上述样本数据进行分析处理。实验结果显示，两种算法均能实现视频流数据包的发现与分类，所识别出的数据包数量与表 1 一致。两种视频流发现算法的时间性能比较如表 3 所示。

**表 3 两种视频流发现算法时间性能比较**

| 样本名称 | 花费时间(ms) | |
|---|---|---|
| | SPA | VPFS |
| Sample1 | 37 | 21 |
| Sample2 | 115 | 68 |
| Sample3 | 200 | 112 |

综合上述结果可以看出，两种网络视频流发现算法，均能正确完成视频流数据包的发现但VPFS算法使用的时间较少。表 2 的 3 组实验中，VPFS 算法所使用的时间分别是算法 SPA 的 43.24%、40.87%和 44.00%，其时间花销明显小于 SPA 算法。此外，SPA 算法的时间开销会随着 CRS 长度及个数的增加而增大，即 SPA 算法的时间花销与关键特征字串集合 $C$ 中元素的长度及个数成正比，而 VPFS 算法的时间效率则与 $C$ 中元素的个数无关。

## 5 结束语

针对网络视频流发现与分类问题，本文分析了网络视频流交互过程的特征，以网络视频流对应的关键特征字串为判断依据，设计了一种基于多关键特征字串匹配的网络视频流识分类算法—VPFS 算法。该算法只需对数据包进行一次不完全扫描，就可以判断出数据包中是否含有视频流 CRS，并在保证准确性的前提下，有效提高了匹配效率，缩短了发现网络视频流所需的时间。实验结果证明该算法具有较低的时间复杂度和可接受的空间复杂度。后续研究过程中，将深入研究网络视频流数据包的高效重组等方面。

## 参 考 文 献

[1] 中国互联网络信息中心. 第 22 次中国互联网络发展状况统计报告[R], 北京, 2008.07.
CNNIC. 22nd China Internet Development Statistics Report[R], Beijing, 2008.07

[2] 彭乐, 薛一波, 王春露. 网络视频内容的识别和过滤综述[J]. 计算机工程与设计, 2008, 29(10): 2587–2590.
Peng L, Xue Y B, and Wang Ch L. Survey on recognition and filtering of network video content[J]. *Computer Engineering and Design*, 2008, 29(10): 2587–2590.

[3] 吴永英, 周淼, 陈晓苏, *et al*.. 基于数据包分析的多媒体信息还原方法研究[J]. 华中科技大学学报(自然科学版), 2007, 35(9): 101–103.
Wu Y Y, Zhou M, and Chen X S, *et al*.. Multimedia information extraction by analyzing of network packages[J]. *Journal of Huazhong University of Science and Technology* (*Nature Science Edition*), 2007, 35(9): 101–103.

[4] Real Networks, Real Media Technology, http://www.realnetworks.com/

[5] Microsoft, WMT, http://www.microsoft.com/windows/windowsmedia/default.asp

[6] M H Willebeek-LeMair, K G Kumar, and E C Snible. Bamba: audio and video streaming over the internet. *IBM Journal of Research and Development*, 1998, 42(2).269–280.

[7] 陈亮, 龚俭, 徐选. 基于特征串的应用层协议识别[J]. 计算机工程与应用, 2006, (24): 16–19.
Chen L, Gong J, and Xu X. Identification of application- level protocols using characteristic[J]. *Computer Engineering and Applications*, 2006, (24): 16–19.

[8] Sen S, Spatscheck O, and Wang D M. Accurate. scalable in-network identification of P2P traffic using application signatures[C]. Proceedings of the 13th international conference on World Wide Web, New York, NY, United States, 2004: 512–521.

[9] Kim Myung-Sup, Won Young J, and Hong Won-Ki. Application-level traffic monitoring and an analysis on IP networks[J]. *ETRI Journal*, 2005, 27(1): 1–22.

[10] Sun Q D, Li S L, and Guo X J. Quick Finding of Network Video Stream. 2008. International Conference on Computer Science and Information Technology, 2008. 9: 353–356.

[11] Schulzrinne H, Rao A, and Lanphier R. Real time streaming protocol (RTSP)[S], RFC 2326, April 1998.

[12] 王丽萍, 孙蕾. 基于Ethereal开源代码构建协议解析器的方法研究[J]. 计算机技术与发展, 2007, 17(10): 27–30.
Wang L P and Sun L. Study on establishment of protocol parser using Ethereal public code[J]. *Computer Technology and Development*, 2007, 17(10): 27–30.

[13] 孙钦东, 黄新波, 王倩. 面向中英文混合环境的多模式匹配算法[J]. 软件学报, 2008, 19(3): 674–686.
Sun Q D, Huang X B, and Wang Q. Multiple pattern matching on Chinese/English mixed texts[J]. *Journal of Software*, 2008, 19(3): 674–686.

孙钦东： 男，1975年生，博士，副教授，硕士生导师，研究方向为网络安全、嵌入式系统.

郭晓军： 男，1983年生，硕士生，研究方向为网络安全.

黄新波： 男，1975年生，博士，研究方向为信息处理.