

# 基于可伸缩编码的高清 IPTV 系统<sup>①</sup>

刘小龙, 王 雷, 况亚萍, 黄承真

(中国科学技术大学 自动化系网络传播系统与控制安徽省重点实验室, 合肥 230027)

**摘 要:** 随着移动互联网技术与智能手持设备的发展, IPTV 面临多屏融合的需求, 可伸缩编码技术可以很好地解决网速和设备分辨率的差异问题. 本文设计和实现了一套基于 H264/SVC 标准的高清实时 IPTV 系统, 通过对视频在时间、空间、质量上进行可伸缩编码, 以及 TS 解封装、AVC 解码和 TS 流传输 H264/SVC 等模块的实现, 通过实时编码, 使拥有不同带宽和设备分辨率的用户能并发实时地观看节目. 实验证明了系统的有效性和稳定性.

**关键词:** 多屏融合; 可伸缩编码; TS; AVC; 实时编码

## HD IPTV System Based on Scalable Coding

LIU Xiao-Long, WANG Lei, KUANG Ya-Ping, HUANG Cheng-Zhen

(Department of Automation, USTC Key laboratory of network communication system and control, Hefei 230027, China)

**Abstract:** With the rapid development of mobile Internet technology and smart handheld devices in recent years, IPTV facing the demand of multi-screen integration. Scalable video coding technology can be a good solution to the differences in network speed and resolution of device. In this paper, we design and implement an HD real-time system of IPTV based on H264/SVC standard. The finalized H264/SVC standards are able to achieve scalable coding in the temporal, spatial, or fidelity domain. This paper presents a detailed introduction of TS encapsulation, AVC decoding and the encapsulation of SVC. By encoding the video in real-time, users with different bandwidth and device resolution concurrent viewing program in real-time. In the end of the paper, the experiments proved the effectiveness and stability of the system.

**Key words:** multi-screen integration; scalable video coding; TS; AVC; real-time encoding

## 1 引言

三网融合推动了广电系统视频业务的发展, 原来仅面向数字电视终端的 IPTV 系统开始建设面向更多类型终端的基础设施. 近年来, 随着不同类型终端用户对高清视频业务需求的不断增长, 三屏融合<sup>[1]</sup>已成为一个热点话题. 所谓三屏融合, 就是使手机、PC 机、电视实现内容的共享和无缝对接<sup>[2]</sup>, 这正好解决了不同设备对视频的需求, 使 IPTV 手机用户、PC 机用户、电视用户能无缝的观看高清电视节目, 由于各类设备的网络传输速率和分辨率的差异, 所以不同类型终端的用户对视频码率和分辨率的需求各不相同, 如果 IPTV 系统针对每一类设备进行不同的编码, 那么服务

器端的压力就会明显增大, 而且存储服务器的开销也过大, 特别是对于 IPTV 这种高清实时播放系统, 需要对高清视频进行实时的编码, 那么服务器的压力就会更大, 因此, 急需一种新的编码方式来满足不同的用户对于高清视频的需求.

2003 年, ITU-T/VCEG 和 ISO/MPEG 合作组建的联合专家组 (JVT) 制定了到目前为止最新的视频编码标准-H264/AVC<sup>[3]</sup>. 而后为了应对视频通信中各种内容、各种不同的网络条件、和不同能力终端之间失配问题, JVT 于 2007 年 11 月将可分级视频编码扩展(Scale Video coding Extension)纳入到了 H.264 标准中, 通常该扩展被称作 H264/SVC, 到目前为止, 该扩展可支持 3 种可

① 项目基金: 国家高技术研究发展计划(2011AA01A107); 中央高校基本科研业务费专项基金(WK2100100012)

收稿时间: 2013-03-21; 收到修改稿时间: 2013-04-28

分级性: 时域可分级、空间可分级和质量可分级, 由于其在压缩效率、实现复杂度、兼容性等方面的优势, 使 H264/SVC 成为视频编码领域的热点. H264/SVC 的可分级特性恰好符合 IPTV 中的三屏融合的需求, 对同一个视频源, H264/SVC 编码器编码出不同层次的码流来适应不同用户对高清视频的需求, 整个过程中, 只需要进行一次编码, 减轻了服务器的负担, 但是现有对 H264/SVC 的研究中, 并没有涉及到对高清视频实时编码和传输的研究, 因此, 需要设计并实现一个基于可伸缩编码的高清实时编码系统, 来满足 IPTV 的需求.

本文描述 H264/SVC 实时高清 IPTV 系统的设计与实现, 结构如下所示: 第 2 节介绍系统设计, 第 3 节介绍高清节目源解析过程, 第 4 节介绍 SVC 的封装实现, 第 5 节对系统进行测试, 并对实验结果进行分析, 第 6 节进行总结.

## 2 系统设计

为了让不同终端(手机、PC 机、电视)的用户能流畅的观赏高清电视节目, 系统对高清视频进行 H264/SVC 实时编码来满足用户对高清视频质量的需求, 为了方便 H264/SVC 在 IP 上进行传输, 相关的传输协议也做了适当的修改, 其中有两个传输协议用得最广, 一个是实时传输协议(RTP: Real-Time Transport Protocol)<sup>[4]</sup>, 另一个是 MPEG-2 传输流(TS: Transport Protocol)<sup>[5]</sup>, 这两个协议都很好的实现了对 SVC 的封装<sup>[6-9]</sup>, 由于 MPEG-2 传输流在数字视频广播中使用非常广泛, 且性能出色, 所以现在很多 IPTV 还是使用它进行传输, 由于本系统主要面向电视节目, 因此本系统选择用 MPEG-2 传输流对 SVC 进行封装传输. 在视频进行可伸缩编码以前, 视频已被编码为 AVC 格式, 并被封装在 MPEG-2 传输流当中.

在基于 H264 可伸缩编码的实时高清编码系统中, 主要有 4 大模块: 高清节目源解析模块, SVC 实时编码模块, SVC 封装模块, SVC 重写模块, 其系统框架如图 1 所示:

- 高清节目源解析模块的主要功能是接收信源发送来的码流, 然后解析为 YUV 码流.
- SVC 实时编码模块的主要功能是对 YUV 码流进行实时的编码, 以达到直播的效果.
- SVC 封装模块的主要功能是对 SVC 编码进行视频流化处理, 使之能进行实时播放.

- SVC 重写模块的主要功能是根据用户的网速以及设备等因素抽取不同码率的视频.

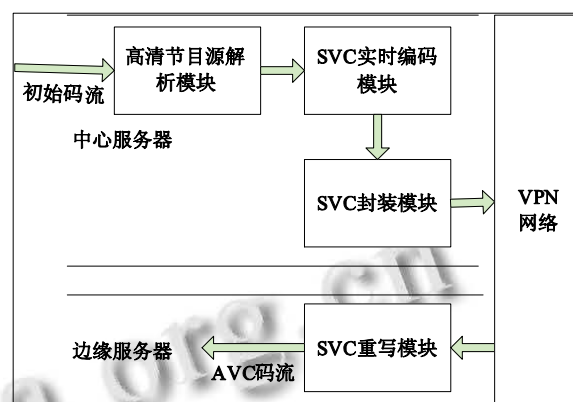


图 1 系统架构

基于 H264 可伸缩编码实时高清编码系统的主要流程是:

- ① 通过视频服务器组播初始码流, 初始码流为 TS 封装的 AVC 高清码流, 高清节目源解析模块接收到了初始码流后, 对 AVC 高清码流进行解码, 解码后的格式为 YUV, 然后通过共享内存的方式传递给 SVC 实时编码模块.
- ② SVC 实时编码模块在接收到 YUV 流后, 对其进行实时的 SVC 编码, 编码完成后的 SVC 码流也通过共享内存的方式传递给 SVC 封装模块.
- ③ 当 SVC 封装模块对接收到 SVC 码流后, 直接把码流打包为传输的 TS 流, 然后 TS 流通过 VPN 网络传输给边缘服务器上的 SVC 重写模块.
- ④ SVC 重写模块根据用户的设备以及网速抽取不同质量的码流, 使用户能更好的观看视频.

以上就是基于可伸缩编码实时高清编码系统的整个流程, 其中 SVC 编码和重写模块主要是算法实现, 不在本文的讨论范围, 下文重点介绍高清节目源解析模块和 SVC 封装模块的实现原理.

## 3 高清节目源解析

这个模块的主要任务是将初始码流解码为 YUV 码流, 主要有两个部分: 首先把从视频服务器接收到的 TS 封装的 AVC 码流进行解封装, 解封装后为 AVC 码流; 然后再把 AVC 码流解码为 YUV 格式.

### 3.1 TS 封装

对于解 TS 封装, 首先要了解 TS 封装 AVC 码流的

格式，而解封装就相当于封装的一个逆过程，因此，下面我们将会讲解 TS 封装 AVC 码流的过程。

AVC 在概念上分为两层：视频编码层(VCL: Video Coding Layer)和网络提取层(NAL: Network Abstraction Layer)，视频编码层主要负责视频的高效压缩和表示，而网络提取层则主要负责以网络要求的方式进行打包和发送，网络提取层包括头部信息、段结构信息和实际载荷信息，即视频编码层数据，整个网络提取层的实现称为一个 NAL 单元，NAL 单元的一个可变长字节字符串。

为了把 AVC 的 NAL 单元封装在 MPEG-2 TS 中，MPEG-2 TS 定义了一个打包基本码流-PES(Packetized Elementary Streams)，简单来说，就是把 AVC 的 NAL 单元加上一个 PES 头，PES 的头信息包括帧率、包大小信息、初始流类型以及视频帧相关的时间信息(解码时间和 PCR 等)。帧率表示携带的视频每秒有多少帧图像，包大小包含 PES 包的长度信息，而根据规定，PES 打包 AVC 时的初始流类型为 0X1B。TS 是以包为单位的，一个包为固定的 188 个字节，其中头为 4 个字节，其余 184 字节为负载区，格式如图 2 所示，图 2 中括号中的数字为每个字段占有的位，这些字段中，PID 最为重要，是每个 TS 包的标识。由于在 AVC 码流中，每个 NAL 单元的大小是不固定的，因此每个 PES 包的大小也是不固定的，而每个 TS 包的大小是固定的，因此必须对不满足要求的 PES 包进行分割，然后分别打包到不同的 TS 包中，在 TS 流中，没有限定视频流的数量，因此，可以在一个 TS 流中传输多个视频流，但是在本系统应用中，只传输一个视频流。在传输的 TS 包中，大部分包的负载数据都是视频和音频数据，但是也有一些包含描述信息的包，MPEG-2 中规定了 4 种：节目关联表(PAT: Program Association Table)、节目映射表(PMT: Program Map Table)、条件访问表(CAT: Conditional Access Table)和网络信息表(NIT: Network Information Table)。在本系统中，主要涉及到两个表：PAT 和 PMT。

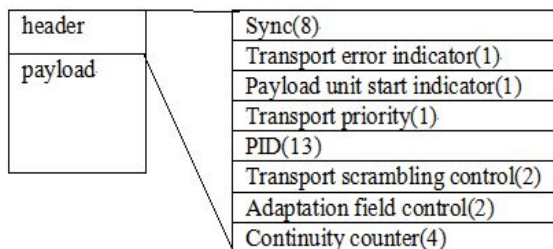


图 2 TS 包结构

节目关联表：在节目关联表中，记录了节目号和该节目的节目映射表的 PID，携带节目关联表包的 PID 为固定的值 0，节目关联表具体格式如图 3 所示：

图 3 中最主要的就是记录了节目号和节目映射表 PID 的 N loop 字段， program map PID 字段是表示该节目的节目映射表 PID，用以确认节目映射表对应的 TS 包，reserved 字段表示预留字段，section length 字段表示从它之后到 CRC 的长度，这个字段的主要用途就是可以推算出整个表中承载了几个节目，N loop 字段表示在一个节目关联表中，可以同时存储多个节目的信息，即在同一个 TS 流中，可以传输多个节目。

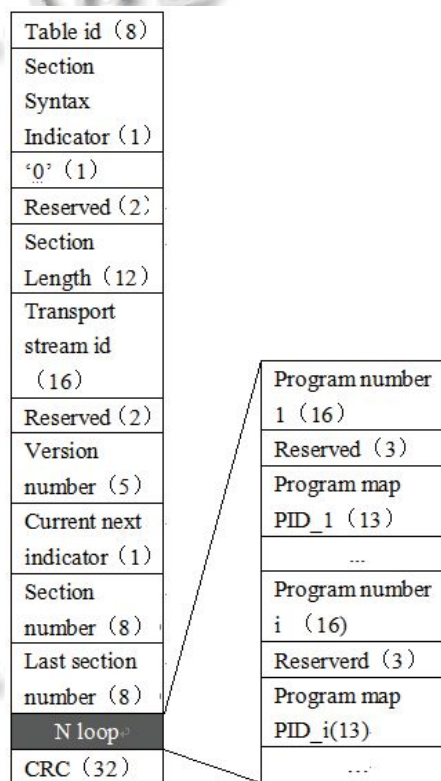


图 3 节目关联表结构

节目映射表：节目映射表主要存储该节目里包含的音频和视频流的 PID，它的结构与节目关联表在结构上相差不多，在 N loop 中，program number 和 program map pid 分别换成了 stream type 和 elementary pid，stream type 表示该流的类型，包括音频和视频，而 elementary pid 则指示了该流所在 TS 包的 PID，同样的，根据节目映射表的结构，一个节目可以有多个视频流合成，携带节目映射表的包的 PID 不固定，但是会在节目关联表中指出。

在了解这两种表之后就会发现，如果要解码 TS 流，需要实现以下三个步骤：

① 找到携带节目关联表的包，该包的 PID 为固定的 0 值，然后解析出节目关联表，在表里面就会有节目信息和携带节目映射表对应 TS 包的 PID。

② 通过携带节目映射表对应 TS 包的 PID 找到 TS 包，然后解析出节目映射表，就能确认该节目中包含的流以及码流所在的包的 PID。

③ 通过码流所在包的 PID 就能解析出需要找的码流，然后对 TS 包进行重组，整个解析过程完成。

通过以上 3 个步骤，就能从 TS 封装的 AVC 高清初始码流中解析出 AVC 高清码流，然后进行后续操作。

### 3.2 解码 AVC

在解 TS 封装后，得到了 AVC 编码的码流，而接下来为了满足用户对视频的需求，需要对视频进行可伸缩编码，即 SVC 编码，因此，首先要把 AVC 格式转码为 SVC 格式，但是到现今为止，还没有一个高效的方法能使 AVC 编码直接转码为 SVC 编码，所以就采取了先对 AVC 解码为 YUV 格式，然后再把 YUV 编码为 SVC。

为了实时解码 AVC 编码视频，我们采用了已经非常成熟的开源软件 FFMPEG 来进行实时解码，关键代码如表 1 所示：

表 1 解码 AVC

```

procedure: decode
While(bytesRemaining>0)//判断是否有剩余数据
{
    bytesDecoded=avcodec_decode_video2( pCodecCtx, pFrame,
    &frameFinished, &packet);
    if(bytesDecoded < 0)
    {
        解码错误，进行处理
        break;
    }
    bytesRemaining=bytesRemaining - bytesDecoded;
    对得到的 pFrame 进行处理，得到 YUV
}
    
```

在解码过程中，首先通过一个循环数组来存储接收到的数据，然后通过表 1 中的代码调用 FFMPEG 开源代码进行 AVC 循环解码，直到没有数据为止。解码过后的数据为 YUV 格式，得到的数据通过共享内存的方式传输给 SVC 编码模块。

## 4 SVC封装

### 4.1 SVC 码流结构

SVC 码流结构是以 AVC 码流结构为基础，但是为了支持其可分层的特性，对 AVC 的 NAL 单元的头结构进行了适当的扩展，图 4 是 SVC 的 NAL 单元扩展头结构，在图 4 中，深色部分就是扩展的 NAL 单元头结构，这个时候的 NAL 单元的类型值必须为 14 或者 20。

在 SVC 扩展头部中，有 3 个字段标识了 3 个不同的伸缩编码方向的等级，字段 D 表示的是空间可分级的等级，字段 Q 表示质量可分级的等级，字段 T 表示时间可分级的等级。对于基本的 AVC 层，如果单元类型为 14，那么这是一个特殊的前缀 NAL 单元，也会有扩展 SVC 头结构，扩展头用来确认该基本层属于哪一个时间等级。在实际应用当中，为了得到特定质量的视频，只需要读取 NAL 单元中 D、Q、T 字段的值，丢弃大于所要值的单元，就能得到想要的视频质量，实现可伸缩编码。

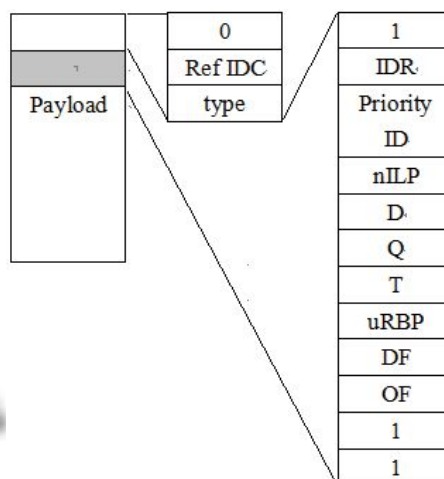


图 4 NAL 单元扩展头结构

在 NAL 单元中，可以负载不同类型的内容，根据负载内容的不同可以分为 VCL 单元和非 VCL 单元，VCL 单元负载着视频图像的编码数据，VCL 数据负载着各种附加信息，如参数集、提高性能的附加信息(SEI: Supplement Enhancement Information)等，参数集包括序列参数集(SPS: Sequence Parameter Sets)和图像参数集(PPS: Picture Parameter Sets)，一般不会发生变化，SEI 不是解码视频序列的必要信息，但是为视频的现和管理提供了附加信息，包含了当前视频中包含图像尺寸、可分级层次、帧率等信息。NAL 单元的负载



类型可以通过 NAL 单元头结构中的 NAL 单元的类型来判断, 如果 NAL 单元中包含 SVC 的视频编码层数据, 那么这时 NAL 单元的类型值为 20, 然后就会有图中所示的 3 个字节的扩展头部。

#### 4.2 TS 封装 SVC

TS 封装 SVC 码流与 TS 封装 AVC 码流的不同之处在于 SVC 有增强层, 怎么处理 SVC 的增强层成为关键, 而根据规定, TS 在封装 SVC 时, 会把 SVC 的不同层放在不同的 PES 包中, 以此来区分码流的不同层次。在具体实现时, AVC 基本层被打包在一个 PES 流中, 然后对于 SVC 增强层, 则根据 SVC 扩展包头中字段 D 的值不同打包到不同的 PES 流中, 即如果字段 D 有 3 个不同取值, 那么就会有 4 个 PES 流(包含一个基本层的 PES 流), 由于在对 SVC 码流进行抽取时, 采取丢弃高层次的 PES 流, 那么这中封装方式就保证了低层次包含的访问单元也包含在高层次的码流中。对于 AVC 基本层, PES 头部的流类型字段值为 0X1B; 而对于 SVC 增强层, PES 头部的流类型字段值为 0X1F。

由于不同的层被放在不同的 PES 流中进行传输, 那么在接收端就必须定义一个方法来对码流进行重组, 得到正确的码流序列, 而根据规定, 重组是基于 PES 头部中的解码时间标签(DTS: Decoding Timestamps)来实施的。由于 SVC 当中可能存在时间可伸缩编码, 因此, 不同时间层次的 DTS 不一定严格匹配, 图 5 就显示了这一特点, 在图 5 中, PES1 负载的是 AVC 基本层, 是时间可分级最低的一层, PES2 负载的是第一个 SVC 增强层, 帧率是 PES1 的两倍, PES3 的帧率同样是 PES2 的两倍, 但是为了解码的连续性, DTS 是等间距的, 因此, 时间可分级低的层的解码时间比时间可分级高的层的解码时间长。在解码过程中, 如果遇到同意访问单元中的不同层次中的 DTS 不一致, 那么这时候就会用到 PES 头部中的参考时间标签(TREF: A Timestamp Reference)字段, 层次高的 PES 中的 TREF 值调整层次低的 PES 中的 DTS 值, 在图 5 中, 如果 PES2 中的 DTS 于 PES3 中 DTS 对于同一访问单元不一样, 那么 PES3 中的 TREF 值调整 PES2 中 DTS 值, 即图 5 中 DTS 调整为 DTS'。

为了方便解码, MPEG-2 系统定义了一个关于时序和缓存的模型-系统目标解码(STD: System Target Decoder)模型, 这个模型主要定义了缓冲区大小、缓存之间的传输等, 图 6 显示了解含有 SVC 的 TS 封装的

缓存模型, 与文献[7]定义的解码模型相对应, 根据没有 TS 流的不同被分配到不同的传输缓冲区(TB)中, 然后把 TS 包的头部去掉, 传给复用缓冲区(MB), 当把 PES 包头去掉后, 传入码流层缓冲区(DRB), 在这里, 会等待这个层里面的所有数据都已经接收到, 然后就会检查该 PES 包中的 DTS 查看是否能够解码, 如果时间满足要求, 那么就会进行解码。

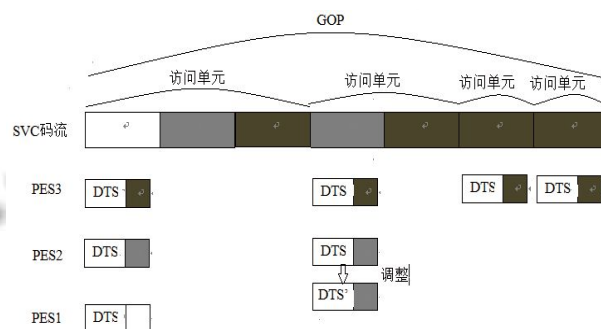


图 5 基于 DTS 重组及 DTS 调整

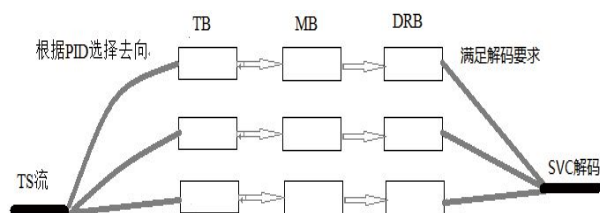


图 6 缓存模型

## 5 实验结果及分析

### 5.1 实验环境

本文在实际系统上进行了实验, 实验设备包括一台 32 核 CPU 的服务器(单核频率为 2.3GHz, 内存为 64G), 其上运行源码解析模块、SVC 实时编码模块和 SVC 封装模块, 一个转码板(1.0GHz CPU, 1G 内存, 32G 存储), 其上运行 SVC 重写模块, 网络环境是连接 32 核服务器和转码板的 VPN 网络。

实验目的是验证系统的可靠性以及编码过后的码流能否适应不同类型终端的需求, 为此, 信号源选择为广电系统标准的 7.5Mbps 高清源码流, 分辨率为 1920\*1080i, 码流的帧率为 50fps。对源码流进行 3 层空间编码, 4 层时间可伸缩编码, 然后抽取 3 种不同分层码流, 根据典型的终端设备能力(广电网络上的高清电视机, 电信网络上的 PC 机, 移动互联网上的智能手机), 将抽取不同空域层码流, 码率分别为

400kbps、2.25Mbps、7.4Mbps,这三种码率的视频流被分别传输给手机终端、PC机终端、电视终端.实验中,对这3个码流视频分别进行传输播放,图7显示了3个不同码流对应的播放结果,图7中右边的图像是7.4Mbps码率对应的视频,中间的图像是2.25Mbps码率对应的视频,左边的图像是400kbps码率对应的视频,这三幅图从右到左的分辨率明显减少,但是码率也相应的减少了很多,而且刚好满足手机用户、PC机用户、电视用户对视频的需求,使手机、PC、电视都能正常的观看电视节目.



图7 三个不同码流对应的图像

实验结果表明,整个系统能稳定正常的运行,且在大幅度降低总码率的同时,通过一次编码,即可在边缘抽取出不同码率的码流,然后传输到不同种类的终端,并且可以正常播放视频.

## 6 总结

三网融合推动了广电系统视频业务的发展,使IPTV系统开始需要面向多个不同类型的终端,为了让手机终端、PC机、电视能无缝的观看高清电视节目,本文设计并实现了基于H264/SVC可伸缩编码技术系统.本文首先设计了该系统的框架,然后详细介绍了系统主要模块的实现,其中源码解析模块为SVC实时编码模块准备需要编码的数据,关键的部分就是对源码流

进行TS解封封装;SVC封装模块主要是对SVC进行TS封装,便于码流在IP网络上传输.实验结果证明,该系统能面向多终端进行实时高清视频的统一编码传输,提供高质量且稳定的视频输出.以上系统已完成原型实现,下一步的工作是基于现有广电网络进行工程化实现.

## 参考文献

- 1 蒋力,施唯佳.三屏互动下的IPTV融合业务探讨.电信科学,2009(3):17-21.
- 2 郭静.三屏融合时代的手机视频内容建构.新闻实践,2010(002):59-60.
- 3 ITU-T and ISO/IEC JTC 1. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC). Advanced Video Coding for Generic Audio/Visual Services: Version 9. 2009.
- 4 Frederick R, Jacobson V. RTP: A transport protocol for real-time applications. IETF RFC3550,2003.
- 5 ISO/IEC 13818-1: 2007, Information technology-generic coding of moving pictures and associated audio information (MPEG-2) -part 1: systems. 2007.
- 6 Wenger S, Wang YK, Schierl T. RTP payload format for SVC video. Internet Engineering Task Force (IETF), 2008.
- 7 ISO/IEC13818-1: 2007/Amd 3:2009.Transport of scalable video over Rec.ITU-T H.222.0 | ISO/IEC 13818-1. 2007.
- 8 Schierl T, Gruneberg K, Wiegand T. Scalable video coding over RTP and MPEG-2 transport stream in broadcast and IPTV channels. IEEE Wireless Communications, 2009, 16(5): 64-71.
- 9 MacAulay A, Felts B, Fisher Y. WHITEPAPER - IP Streaming of MPEG-4: Native RTP vs MPEG-2 Transport Stream.http://www.envivio.com/technology/white\_papers.php, 2005.

(上接第26页)

- 8 王慧强,杜晔.入侵检测技术研究.计算机应用研究,2003,20(10):90-95.
- 9 鲁剑锋,李华文.访问控制策略的分类方法研究.武汉理工大学学报.2011,33(6):878-882.
- 10 Sandhu RS, Coyne EJ. Role-Based Access Control Models, IEEE Computer, Feb. 1996: 13-20.
- 11 齐忠厚.Kerberos 协议原理及应用.计算机工程与科学, 2000,22(5):11-13.
- 12 胡美燕,刘然慧.DES 算法安全性的分析与研究.内蒙古

大学学报(自然科学版),2005,36(6):693-697.

- 13 唐正军.黑客入侵防护系统源代码分析(第一版),北京:机械工业出版社,2002:41-52.
- 14 李孟珂,余祥宣.基于角色的访问控制技术及其应用,计算机应用研究,2000,10(1):87-89.
- 15 游新娥,胡小红.Kerberos 身份认证协议分析与改进,计算机系统应用,2012,21(4):216-219.
- 16 乔颖,须德,戴国忠.一种基于角色访问控制(RBAC)的新模型及其实现机制.计算机研究与发展,2000,37(1):37-44.