

# 面向未来的移动宽带音视频传输协议 —现状与挑战



未来移动通信论坛  
FUTURE MOBILE COMMUNICATION FORUM

# 目录

摘要: .....	2
1. 视频制作域和传输域的场景、用例和关键指标.....	4
1.1 场景.....	4
1.2 视频制作域的数据传输.....	5
1.2.1 电视制作域.....	5
1.2.2 互联网直播带货.....	9
1.3 移动传输域的数据传输.....	11
2. 视频制作域传输协议.....	13
2.1 电视级制作使用的协议.....	13
2.1.1 SMPTE ST 2022.....	14
2.1.2 SMPTE ST 2110.....	15
2.1.3 NDI.....	17
2.1.4 无压缩与浅压缩的视频编码技术的简介.....	18
2.2 移动链路回传协议.....	19
2.2.1 RTMP.....	19
2.2.2 SRT.....	20
2.2.3 WebRTC.....	21
2.2.4 ZIXI 协议.....	22
2.2.5 RIST.....	23
2.2.6 GB28181 协议.....	24
2.2.7 ONVIF 协议.....	25
2.2.8 RTMP, WebRTC, ZIXI, RIST 和 SRT 的对比分析.....	26
3. 移动传输域传协议.....	29
3.1 自适应码率传输.....	29
3.2 面向广播组播分发.....	36
3.2.1 基于 FLUTE 和 ROUTE 的广播分发 (Qualcomm) .....	36

3.2.2 基于 BIER 的组播分发.....	38
3.3 单点分发 HTTP/HTTP2/HTTP3.....	41
3.3.1 HTTP/1 协议.....	41
3.3.2 HTTP/2 协议.....	42
3.3.3 HTTP/3 协议.....	44
3.4 VR 内容的分发.....	46
3.4.1 VR 媒体服务总体架构.....	46
3.4.2 VR 视频传输模式.....	47
3.4.3 VR 媒体封装及传输格式.....	48
4. 小结和音视频传输趋势预测.....	50
参考文献.....	53
致谢 .....	55

## 摘要:

随着 5G 网络的部署,手机等移动终端的通信能力得到极大提升。手机、Pad 等移动设备观看视频已经从 WiFi 连接为主转变为移动和 WiFi 连接并重的局面。通过 5G 蜂窝网络观看视频,最大的优点是摆脱了 WiFi 连接的地域限制,使用户能够随时随地观看视频。为了提升用户的体验,手机厂家近年来不断改善手机的 GPU 计算能力和显示技术,为用户呈现 UHD、10 比特 HDR 和高帧率的显示效果。借助 5G 提供的巨大传输能力,提供内容的 OTT 公司顺应这一趋势,为用户提供了大量 4K 超高清和 10 比特 HDR 的视频内容。随着 5G 的广泛部署,大带宽上行能力也逐渐凸显。目前,5G 大上行产品已经可以支持高达 2Gbps 的上行峰值速率,用户平均体验速率也达到了百兆级别。5G 视频背包已经可以提供播出级别的超高清视频,这也开启了基于浅压缩协议的超高清视频云端协同这一新应用场景。可以预期,随着后 5G 网络提供更大的传输带宽和更低的传输时延,基于移动宽带的视频业务将会进一步向前发展,赋能更多的新用例。

视频传输协议构建于 IP 层之上,利用物理层提供的带宽支撑各种视频业务。视频传输协议的设计原则之一就是充分适配物理传输带宽和信道的变化,来满足不同视频业务的需求。与 5G 结合的视频传输协议应用有两个方面的考虑:

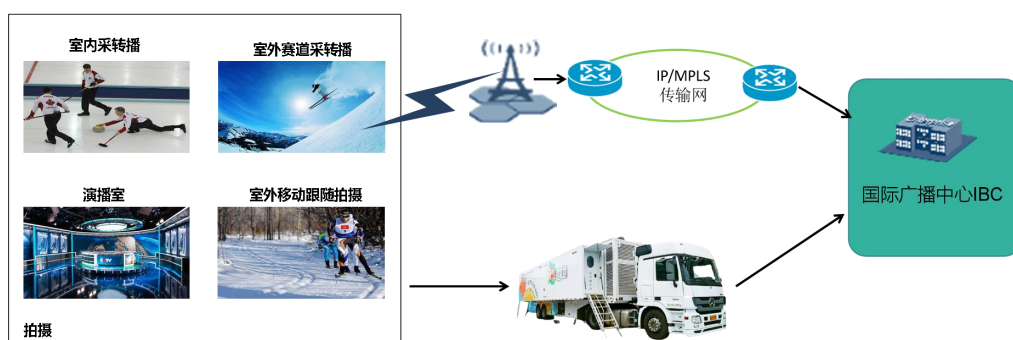
- 首先是被广泛应用在后 4G 时代的传输协议,例如视频背包使用的 RTP/RTMP 协议,可以预期在一段时间内仍然将是主流的传输协议。随着 SRT、WebRTC 等低时延协议逐步成熟,也有可能成为背包传输的主流协议。在服务器分发视频流方面也是类似情况,FLV、HTTP/1 等协议仍然被广泛使用,但 HTTP/3 等协议被越来越多的关注,可以预期新的协议会随着 CDN、OTT/APP 和智能手机几个方面的不断努力,新的协议逐步被广泛应用。
- 其次就是随着 5G 传输速率不断提升,原来用于以太网等固定连接传输的低时延、高速率协议也有可能被用于 5G 链路。例如 JPEG XS、NDI 等制作域协议也有可能被 5G 链路承载,实现播出级的云端制作。

本报告从视频制作域和传输域的需求入手，分别总结了不同场景下传输协议的需求和关键指标，并给予技术分析。最后在此基础上，本报告对面向未来移动通信技术的新型视频制作和分发给予了展望。

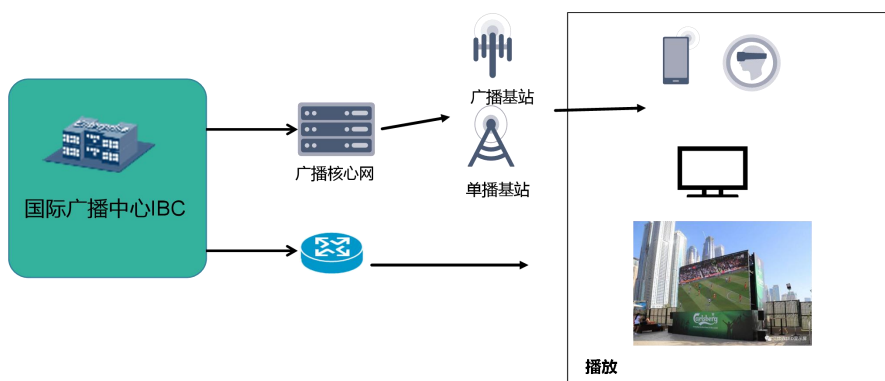
# 1. 视频制作域和传输域的场景、用例和关键指标

## 1.1 场景

根据使用场景区分，可以将屏到屏的视频呈现分为制作域和传输域。制作域主要关注内容生产，通过有线或无线的方式将内容传送到广播中心。传输域关注内容分发，从广播中心通过电视台或者互联网将内容发送给终端用户。



(a) 视频制作域示意图



(b) 视频传输域示意图

图 1 视频制作和传输系统示意图

通常，视频制作域可以分为电视级内容生产、互联网直播和行业视频应用。表 1 列出了视频制作的场景分类。

视频上行业务场景种类繁多，各场景的业务特性、视频编码方式，封装与传输方式，QoE 等均有所不同，与之相对应的技术实现方案及组网架构也有所不同。同时，对网络带宽和时延要求也有所不同。本章的 1.2 章节将按照表 1 给出的分类详述各个分类的典型应用和传输指标。

表 1 视频制作的场景分类

场景分类	典型应用
电视级内容生产	影视拍摄及制作，赛事转播，新闻采访；
移动互联网直播	新媒体直播、电商直播，互联网主播，网红直播，旅游文化休闲直播等；
行业视频应用	视频监控，无人机测绘拍摄，机器视觉的工业应用等；

在传输域，传统的视频分发主要面向电视机、户外大屏等固定接收终端。近年来随着智能手机的快速发展，智能手机、平板电脑的数量已经远远超过了电视机等传统终端。因此，传输域从有线扩展到了移动通信领域。业务形态既包括传统的视频节目推送、视频点播，还包括短视频、交互视频等新的业务类型。

## 1.2 视频制作域的数据传输

### 1.2.1 电视制作域

近年来，电视制作越来越多的利用移动通信网络传输视频内容。通过移动网络聚合网络，快速现场搭建高速、稳定、低延时的制作网络，提供给车载/箱载系统，用于接收和发送各类 IP 化视音频信号和数据信号，面向各类传统媒体和融合媒体的制作发布。

同时，作为快速轻便的传输方案，视频背包近年来得到越来越多的应用。当前的背包主要通过 4G、5G 商用网络，采用 HD/4K 摄像机+编码器+4G、5G 聚合路

由的上行模式为主。同时便携式业务会把编码模块和移动通信聚合模块集成，采用 5G 背包和 5G 手机的形式。目前 5G 背包和 5G 手机以 HD/4K+Sub-6 5G 为主，编码器以 HEVC 为主，预期未来 8K 和 5G 毫米波网络成熟之后，将会演进为 8K+5G 毫米波背包，提供更大的带宽。

目前在专业内容生产制作领域，广电内容制作和信号处理设备之间的连接，主要有三种方式：基带有线传输、IP 有线传输和无线传输（微波、卫星，4G/5G 等）。

基带有线传输，是广电领域传统的传输和连接方式，通过采用 BNC 接口的同轴电缆，传输未压缩的 SDI 信号（串行数字信号）。SDI 接口标准由 SMPTE 标准组织制定，具备不同的速率等级，目前常用于 4K 和 8K 视频传输的通常为 SMPTE ST-2082 定义的 12G-SDI 信号，可以通过 SDI 线缆传输 12Gbps 信号。SDI 信号传输信号随着线缆长度增加而快速衰减，通常 12G-SDI 的传输距离在 110 米以内。

随着视频分辨率的不断提升，基带信号和 SDI 传输方式的传输速率普遍被认为趋近于物理极限，即 12Gbps。当前单线缆最多支持到 4Kp60 标准（12G-SDI），而 8K 内容的制作系统，则需同时采用 4 根 12G-SDI 线缆来传输一路 8K 信号。基带传输能力的限制，增加了系统连接和调整的复杂度，也导致系统风险系数升高。另外，采用 SDI 传输的基带信号基本上都是单向传输，设备间如果存在多种信号的交换需求（视频、音频、控制、同步等），就需要多套线缆提供双向连接。随着目前大型现场节目的复杂度越来越高，对于转播制作系统调整和机动性的要求也越来越高。

随着数字技术的发展，以 IP 网络传输取代线缆传输，采用 IP 网络架构替代传统的基带系统连接，是专业内容制作领域的发展趋势之一。IP 传输一方面可以简化基带结构和连线的复杂度，借助 IP 网络技术实现多种信号的复合与双向传输；同时，IP 结构具备更为优异的带宽性能，比如传输性能高达 100Gbps 的交换机已作为新一代视音频系统路由终端替代传统的基带视音频路由矩阵，可应对单链路 8K 信号的传输需求。在此之上，甚至可以考虑提升视频采样率和比特深度（如 4:4:4、12bit 等），以实现内容制作质量的进一步升级。IP 系统带来的另一改变是实现了高质量远程制作的可能。通过部署专线，将现场的摄像机信



号通过 IP 专线传给远端或异地的演播室转播系统，所有的制作和技术调整，都在远端来完成，这样能够大大提升现场制作的效率。

目前广电领域采用的 IP 系统，遵循 SMPTE 发布的 ST-2110 标准协议，将 SDI 的 AV 信号（视频和音频）转换成 IP，并同时 will 音频、视频和数据三种信号区分形成三路独立的 IP 数据流进行传输，在传输过程中可分别对每种信号进行处理。三种信号的分开与重新合并需要做到高精度的同步，为此，ST-2110 支持采用 PTP 协议进行时钟校准，以确保所有音视频流的同步转换。

除了在演播室、摄影棚的视频制作，一些视频拍摄还需要在室外拍摄并通过无线技术传输回演播中心。这类视频目前通常使用微波和卫星通过无线链路传输视频。

卫星传输制作域信号，一般是负责将演播室或现场转播系统制作完成的基带节目信号，通过地面卫星车内编码器压缩编码后，再通过卫星链路实现上行转发。微波传输，更多用于现场制作过程中，将需要移动拍摄的摄像机信号（如综艺晚会中的无线斯坦尼康摄像机，马拉松比赛中在移动拍摄马车上的无线摄像机）通过摄像机加载的微波发射适配器，借助周围高点的微波中继转发，通常由转播系统来接收使用。

考虑到发射功率、带宽、时延等因素，卫星或微波一般不会用于传输高码率视频。例如，高清 1080 传输码率约为 8-15Mbps，而 4Kp50 约为 30Mbps。考虑到时延和现场环境的要求，微波的传输码率可能会比卫星更低。由于传输码率的局限对画质产生影响，在高要求的广电专业制作域，无线机位目前一般不会作为核心机位来使用，而更多是起到辅助作用。

在 4G LTE 部署后期，已经有公司不断尝试使用 4G 商用网络，通过多卡多链路聚合的方式向制作中心传输 4K 视频信号。由于我国运营商提供了 4G 网络的全国覆盖，这种基于公网的传输方式具有非常大的灵活性，无需提前布线可以随时随地提供超高清视频传输服务。

当前我国 5G 商用网络已经实现了全国大部分城市的覆盖，上行传输速率已接近 80-100Mbps。相比传统卫星和微波，5G 在传输速率及使用成本方面展现出最佳的适用性及更高的性价比。随着 5G 基站在国内的快速布设，5G 信号的覆盖

范围和信号强度将逐步提升，切片等新兴技术的运用也预期将为广电信号传输的高可靠性要求提供精准的服务质量保障。

另外，视频编解码技术的不断升级，特别是针对编码效率及编码时延的进一步优化，为超高清视频信号在制作域借助 5G 进行传输提供了先决条件。在高码率和高画质得到保障的前提下，综合考虑编码延时和 5G 信道延时对方案设计和实际应用而言将非常重要。

云端制作是远端制作的进一步发展，通过 IP 链路将视频信号传回广播中心。图 2 是超高清视频通过移动通信链路传输示意图。摄像机将采集到的数据传输到编码器进行编码，编码后的数据通过路由器和 CPE 作为上行数据发出。信号通过边缘计算节点 MEC、OBS 通信机房进入国际广播中心（IBC）。

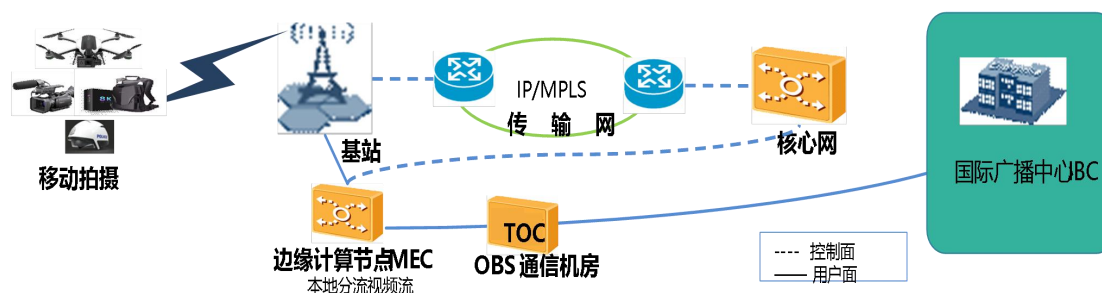


图 2 通过5G链路转播回传超高清视频示意图

超高清视频制作通常采用 SMPTE ST-2110 传输无压缩或浅压缩信号，通过传输大带宽信号获得极低的传输时延。以 4K 为例，摄像机输出的通常是 12Gbps 的 SDI 信号，通过制作域的浅压缩编码器后转换为 IP 码流通常的大小在数吉比特到数百兆比特每秒。根据经验值，用于母盘制作的视频压缩率控制在基带信号的 1/8 - 1/12，即 1 - 1.5Gbps。刨除信号带宽中的部分冗余信号，信号带宽通常需要在 600Mbps 以上。浅压缩信号通常需要预留一帧以保证同步，即在 50p/s 的信号编码需要大约 25-30ms 的时延，同时移动通信系统需要引入 60ms 左右的端到端时延，屏到屏的总时延大约在 100ms 以内。

目前移动视频传输通常采用 5G 背包形式。即将摄像机信号通过 HEVC 等方式进行深度压缩编码，通过 5G 商用网络传输。这种方式对网络适应性好，通常无

需部署专用网络。缺点是受限于商用网络传输能力,通常 4K 信号被压缩至 30Mbps 以内,信号损失大,不适合母盘制作和大屏播出。为了提供适合大屏幕播出的高质量信号,编码后的速率需要保证在 80-100Mbps 左右。

表 2 超高清 (4K) 信号的示例性KPI

场景	视频压缩标准	数据带宽	屏到屏时延
制作域	JPEG XS, SDI over IP	600Mbps - > Gbps	<=100ms
视频背包	HEVC 等	80 - 100 Mbps	<= 2s

### 1.2.2 互联网直播带货

直播带货,是指通过互联网平台,使用直播技术进行商品的线上展示、咨询答疑、导购推销等新型售卖方式。厂家通常借助专业平台开设自有品牌直播间,通过名人主播、商铺店家主播、职业主播等,汇集多种商品进行直播推销。

一方面,“直播带货”互动性好、亲和力强。主播通常具有较高的信誉度,通过积极主动地推荐商品,获得消费者对展示商品产生信赖;另一方面,“直播带货”绕过了经销商等传统中间渠道,能给出更加优惠的价格,实现商品和消费者的无缝对接。“直播带货”通过质量可靠、性价比高的商品推介,引起消费者的关注,并有效降低消费选择过程中的信任成本,促成购物行为。

今年双 11 期间,直播电商依旧是增长点。10 月 20 日-11 月 3 日,今年双 11 第一波预售期间直播电商销售额为 821 亿元。李佳琦以 3615.81 万的销量、106.53 亿元的销售额,稳坐销售第一的位置。截至今年 11 月 11 日,今年双 11 直播销售额累计为 1318 亿元,约为去年的两倍。直播带货需要考虑如下典型的交互需求:

(1) 主持人能接收观众的文字反馈进一步补充商品介绍，增强观众对商品品质的信心，并可借助 AR 技术提供商品试用试穿等功能，让用户身临其境地体验商品使用效果；

(2) 观众通过主持人的商品介绍以及支付链接或二维码进行商品的下单与支付；

(3) 主持人可以进行在线抽奖，或穿插娱乐性小游戏等节目，增加节目趣味性的同时与观众形成有效互动，提高观众粘性。

在一场典型的电视直播带货过程中，一方面需要面向广大观众提供低时延、高质量直播视频的分发；另一方面需要借助双向网络与观众形成交互，形成低时延的交互体验。同时，接入网络交易平台与服务监管机制也将成为此类业务不可缺少的组成环节。

带货的主播有些是地域性的，有些是全国性的。当头部主播开启直播时，往往引发一个地理区域内、或一个时间段内的聚集性在线观看，对直播运营平台的网络传输构成极大的挑战。而基于常规动作，直播运营平台不得不付出高额的经营成本，扩充海量带宽和 IDC 容量，以应对过程中的峰值挑战。

互联网直播的无线传输方式与视频背包类似，通常采用 WLAN 或移动通信将视频传送到 APP 的云服务器。编码器通常采用 HEVC 等主流移动端编码器，分辨率通常为 720p 或 1080p，为了保证传输的稳定性，压缩后的视频通常在 5Mbps 以内。传输时延并没有严格要求，不考虑安全播出需要的额外时延，通常在几秒或者十几秒。为了保障主播的播出质量，通常 APP 服务商会在主播集中的区域部署或租用大量的 CDN 网络资源。对于头部主播，APP 运营方还会租用专线、部署专网保障直播的顺畅。这也是 APP 运营的主要传输成本之一。

5G 网络作为移动传输的高性能纽带，可为头部互联网主播直播、企业级移动直播提供更为可靠、易用性更强、服务质量更高的上行回传保障，而边缘云、切片等技术的同步发展，也将进一步推动 5G 上行业务的发展和商业模式的创新。

### 1.3 移动传输域的数据传输

移动传输面向的接收主要包括电视等固定大屏和智能手机、pad 类的移动小屏幕。随着移动互联网的发展，以智能手机为代表的移动终端成为主流的视频接收端。特别是近年来，中、高端智能手机已经普遍具备 4K 和 HDR 视频处理能力，甚至超过了电视等大屏显示器的处理能力。为了提升用户体验，主流的视频 APP 提供商陆续推出 4K 内容，这也驱动了运营商加速部署 5G 网络以提供相应的流量。

电视的传输方式分为有线和无线两种。在我国，有线电视主要采用 DVB-C 协议，而无线电视传输采用 DTMB 协议。我国还有一定的卫星接收终端，通常采用 DVB-S 和 DVB-S2 协议。智能手机、pad 类终端通常采用 802.11 系列 WLAN 协议和 4G、5G 移动通信协议接收来视频等信号。

中央广播电视总台在 2021 年 1 月颁布《8K 超高清电视节目制播技术要求(暂行)》，其中规定：

- 8K 超高清电视播出信号编码压缩技术要求，采用 AVS3 标准，对 8K 超高清信号(7680×4320/50P/HDR)采用基准 10 位档(profile)、10.0.60 级(level)，视频编码码率不低于 120Mbps。
- 8K 超高清互动点播技术要求，基于 AVS3/H.266/H.265 进行编码，总码率要求大于 80Mbps。

关于 4K 超高清视频，中央广播电视总台在 2018 年发布的《4K 超高清电视技术应用实施指南（2018 版）》规定视频编码采用 AVS2 标准，支持基准 10 位类、8.0.60 级以上的编码方式，1 路视频压缩码率不低于 36Mbps。针对 4K 超高清电视网络传输还规定：

- 卫星分发网络传输可采用 QPSK 调制方式，通过 36MHz 转发器传输 1 路 4K 超高清电视；可采用 8PSK 调制方式，通过 18MHz 传输带宽传输 1 路 4K 超高清电视。
- 有线电视网络传输可采用 DVB-C 64QAM 或 256QAM 调制，在 1 个频点 8MHz 内至少能满足 1 路 4K 超高清电视传输带宽。IPTV 网络传输采用专门的视频网络通道传输 4K 超高清电视节目流，1 路视频的压缩码率应不低于 36Mbps。

表 3 4K、8K超高清视频编码、码率需求

清晰度	业务类型	编码协议	码率
4K	视频编码	AVS2	36Mbps
8K	互动点播	AVS3/H. 266/H. 265	80Mbps
	电视播出信号	AVS3	120Mbps

## 2. 视频制作域传输协议

### 2.1 电视级制作使用的协议

在最近的十几年内，整个广播级专业视频内容制作以及分发行业都在向 IP 网络过度，这种变化不仅仅出现在广播电视内容的交付领域，也在影响专业视频内容的制作领域。

虽然通过 IP 网络向观众和客户分发多媒体视频的工作从上世纪 90 年代中期就开始了，但将 IP 技术彻底的应用在内容生产和制作领域的部署需要的时间和成本却更高。在现场制作设施中，移动数字视频、音频和相关元数据的主导技术一直是串行数字接口（SDI，即使用同轴电缆各种基带视频传输技术，如 SD-SDI、HD-SDI、3G-SDI，12G-SDI 等等）。曾经，IP 技术用来传输大容量视频经常被质疑，而多年以来，内容制作商也已经在 SDI 基础设施的建设上投入了许多资本，用 IP 取代这些基础设施是一项艰巨的任务。

时至今日，得益于 CPU 以及计算能力的提升，如今的 IP 设备可以轻松应对 SDI 的高数据速率的要求。我们可以看到 IP 技术正在被使用于直播内容制作的工作流中、在后期制作中、以及在向多频道视频节目分销商（MVPDs）的内容交付中，机构们使用的摄像机是全 IP 化的，新一代的大容量 IP 切换器和路由器正在取代传统的基于 SDI 连线的设备。

与此同时，随着大众对专业视频内容的多元化和多样性的要求越来越高，专业的内容制作也不能再局限于提供连线的演播室以及专业场馆了，而是需要将专业内容制作的能力部署在更多不同的移动的场地，而随着 5G 网络的流行以及移动通信系统提供带宽的增加，通过移动网络传输无/浅压缩的视频内容（SD/HD/UHD）已经成为了可能。

为了应对以上的应用场景，出现了一批新的基于 IP 的视频传输以及压缩协议，本章节将简要的介绍 SMPTE ST-2022、SMPTE ST-2110、NDI 等 IP 传输协议以及 TICO、JPEG XS、JPEG2000 等相关的视频压缩协议。

## 2.1.1 SMPTE ST 2022

SMPTE ST 2022 标准是由传统 SDI 连线向 IP 基础设施过渡的第一步，该标准的前两部分（2007 年发布）定义了通过 IP（使用 RTP 封装）传输恒定码率 MPEG-2 传输流的方法（可选择使用 FEC 以确保可靠性）。因此，ST 2022-1 和 2022-2 适用于从广播公司到内容分发商的初级分发，例如有线电视、卫星电视和 OTT 的 (v)MVPD 等。通常在这种传输中会使用 JPEG-2000、MPEG-2 或 H. 264/AVC 对视频进行适度的压缩。

然而压缩的视频流对于实时要求极高的前端制作并不适用，在这种情况下，通常需要精确到帧级别的拼接来插入插播广告，以及广播设施中的 Logo 和其他图形或特效的插入。这些操作都需要在无压缩的域中进行。

因此在标准（ST 2022-6）的第 6 部分解决了基于 IP 的 SDI 信号的传输，即将未压缩的音频、视频和辅助数据，如隐藏字幕和定时信息作为捆绑包等组合在一起。ST 2022-6 使用 RTP 将 SDI 数据封装在 1376 字节的数据包中。ST 2022-6 支持在前端的视频内容生产设施中使用 IP 路由器和交换机。

在 ST 2022-5 中则定义了 FEC 的数据容错机制。

此外，在 ST 2022-7 中规定了一种冗余传输保证可靠性的方案，即相同流输入并通过两条不同的路径从发送方发送到接收方。如果其中一条路径的数据包丢失，接收器将使用另一条路径的数据包重建流。接收者必须考虑到延迟抖动和不同的端到端延迟，并为此分配一个缓冲区，缓冲区的大小反映了延迟和弹性之间的权衡。虽然这种技术使带宽的使用增加了一倍，但它排除了对重传、确认等的需要以及与 FEC 方案相关的复杂纠错算法。此外，该方案允许一个自动、自主和无缝的故障切换系统，而不需要监督软件来检测和实施故障切换过程。



下面是关于 ST 2022 相关的协议的总结：

ST 2022-1	实时视频/音频的 IP 网络传输及前向纠错
ST 2022-2	恒定码率 MPEG-TS 流在 IP 网络的单向传输
ST 2022-5	高码率媒体信号的 IP 网络传输及前向纠错
ST 2022-6	高码率媒体信号的 IP 网络传输
ST 2022-7	IP 数据报信号无缝路由切换保护

### 2.1.2 SMPTE ST 2110

上文提到的 ST 2022-6 的特点之一是音频、视频和辅助数据捆绑在一起作为一个流。而在大部分的应用场景中，需要将三者彼此分开（例如，将音频和相关联的隐藏字幕从一种语言转换为另一种语言）。

2015 年，VSF 的 Studio Video over IP (SVIP) 活动小组定义了技术建议 TR-03，该建议将每种媒介作为单独的流或源，使用 RTP 作为基础传输协议。这项建议为 SMPTE ST 2110 标准奠定了基础。

2017 年，SMPTE ST 2110 标准的首批规范发布，以满足这一需求（以及其他），即视频、音频和辅助数据各自在单独的 RTP 流中承载。规范的不同部分定义了压缩视频的传输。该规范也定义了单独部分专门用于系统时序，以确保现在各自的媒体流正确同步。

图 3 展示了 SDI、ST 2022 和 ST 2110 的简要对比。在这个层面上，SDI 和 ST 2022 之间的关键区别是物理连接，而 ST 2022 和 2110 之间的区别是后者的媒体未被捆绑。

图中还列出了 2110 标准的一些具体部分，以及它们具体涉及的内容。例如，ST 2110-22 涉及到浅压缩视频的传输（使用 TICO 或 JPEG XS），这一部分的规范是在 2019 年 8 月发布的。而 JPEG XS 通常提供 10:1 的压缩，但延迟非常低（~1 毫秒，而 JPEG2000 为~60 毫秒）。

在错误恢复方面，ST 2110 标准依赖于 ST 2022-7。注意，ST 2110 和 ST 2022-7 主要用于带宽基本不受限的广播制作域内部（或设施间连接）。

相对于 ST 2022 的一个改进是，ST 2110 引入了对兼容的设备的控制和管理，这里提到的控制和管理是由网络媒体开放规范（NMOS）定义的。NMOS 允许设备自动注册，被其他设备发现并建立连接，大大减少了手动设置和配置的需要，从而提升了自动化和智能化的程度。

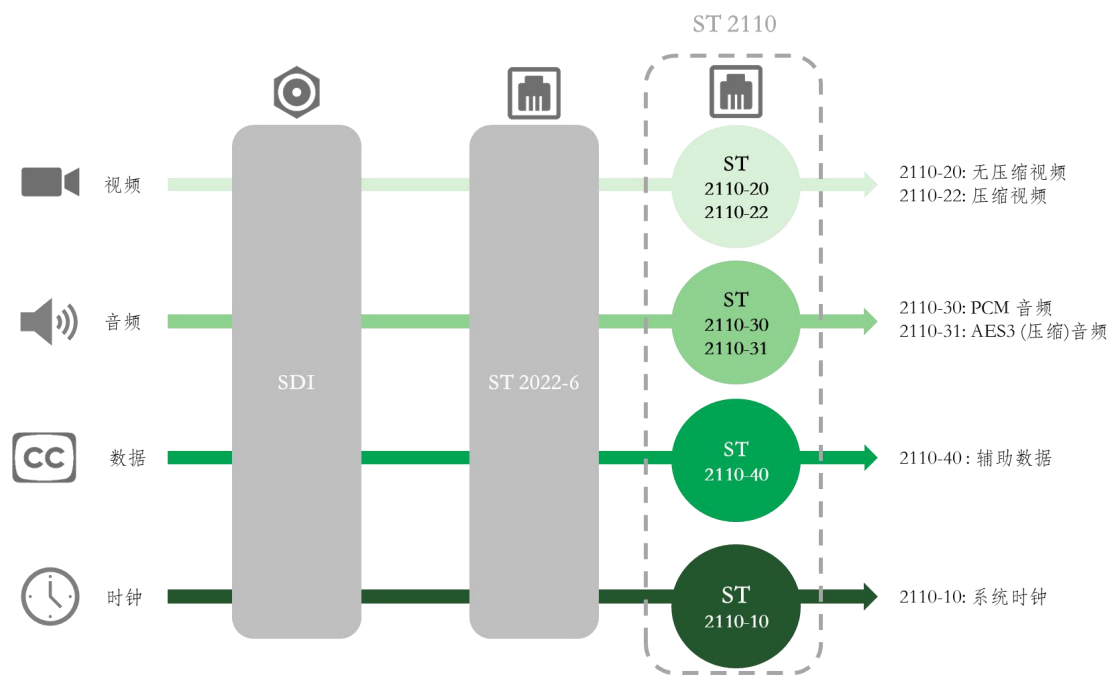


图 3 SDI、ST 2022和ST 2110的简要对比

下面是关于 ST 2110 相关的协议的总结：

ST 2110-10	系统时钟定义
ST 2110-20	无压缩视频的 IP 网络传输
ST 2110-22	压缩视频的 IP 网络传输
ST 2110-30	无压缩音频的 IP 网络传输
ST 2110-31	AES3 音频的 IP 网络传输
ST 2110-40	辅助数据（元数据）的 IP 网络传输
ST 2022-7	IP 数据报信号无缝路由切换保护

### 2.1.3 NDI

NDI 是 Network Device Interface 的简称，是种 IP 网络设备接口协议。NDI 协议技术是美国 NewTek 基于局域网研发的视频编码传输协议，NDI 是一种网络接口协议，是由 NewTek 开创的高质量、低延迟、多通道的 IP 视频传输标准。音视频信号在进行 NDI 编码后，能实时通过 IP 网络对多重广播级质量信号进行传输和接收，同时具有低延迟、精确帧视频、数据流相互识别和通信等特性。

NDI 世界上有两个版本，FULL NDI®和 NDI|HX。FULL NDI®是 I 帧帧内高比特率压缩协议，具备超低延迟（1 帧或更少）和质量几乎无损的优势特点，可取代复杂的 SDI/HDMI 连线，但是需要占用较大的带宽，比如单个 1080p60 的 NDI 流可能需要 140Mbps，单个 4Kp60 的 NDI 流则需要 250Mbps。对于视频传输有高质量要求的，需采用 FULL NDI®进行编码和传输。如果想同时流式传输多个 NDI 源，则受本地网络带宽的限制，如千兆以太网仅能同时传输数个 NDI 源。FULL NDI 可以作为一种带宽占用较小的广播电视级信号传输协议。

NDI|HX 是 NDI 的低带宽版本，是一种压缩版的长 GOP H.264 变体，优点是低带宽要求，允许简单千兆网络承载多个视频流，而不会占用大量网络流量，如单个 1080p60 的 NDI 流仅需 10Mbps 左右。低带宽的优势允许 NDI|HX 采用 WiFi 无线连接，或速度低于千兆位的网络链路，部署十分方便。需要注意的是，与 FULL NDI®相比，NDI|HX 会产生一点延迟，但仍然可以接受，仍优于现有的 RTSP 传输延时。延时大小取决于所采用的设备、帧速率、视频分辨率和网络条件。随着 NDI 技术的火热应用，越来越多的实时流媒体制作软件及设备均已支持 NDI。

另外，NDI 系列标准还推出了 NDI|HX2[10]。作为下一代高效的 NDI 协议，它即可以使用 H.264，又可以使用 H.265（HEVC）。与 NDI|HX 相比的主要区别在于，NDI|HX2 是真正的原生 NDI 流，从源设备一直到目的地。由于有了更高的比特率和 HEVC 的选择，NDI|HX2 可以通过大带宽链路提供类似 Full NDI 的画面质量。NDI|HX2 共享 NDI 所有的一般特性，具备业务发现（mDNS、NDI 访问轮询和 NDI 4 发现服务器）和对低级网络传输的相同控制（TCP、UDP、多播和 mTCP），具备承载任意 NDI 元数据的能力，可以提供更加可靠、低延迟和灵活连接。

## 2.1.4 无压缩与浅压缩的视频编码技术的简介

上文提到的在专业内容制作中比较常见的传输协议，在这些协议中传输的主要内容是视频，由于无压缩的视频的带宽需求相当大，因此引入了基于浅压缩的视频编码技术，在保证低延时的同时降低传输的成本，从而满足在部分特殊场景的需求。下面表 4 总结了上述提及的各种常见的无压缩或浅压缩技术的统计表格供大家参考，用户可以根据自己的应用场景选择适合的技术路线。

表 4 无压缩、浅压缩算法比较

编码格式	延时	压缩比	码率	典型封装格式
Uncompressed	0ms	1:1	12 Gbps	RTP
J2K ULL	~20ms	10:1	1.2 Gbps	MPEG-TS, RTP
TICO/JPEG-XS	~1ms	6:1	2 Gbps	MPEG-TS, RTP
H. 264/AVC (AI)	>20ms	25:1	0.5 Gbps	MXF
H. 265/HEVC (AI)	>20ms	50:1	0.25 Gbps	MXF
H. 266/VVC (AI)	>20ms	95:1	0.13 Gbps	TBD

注：上述提及的码率为 4K/P50 的条件下，延时仅包括视频内容编码所需的典型延时，不包括传输及数据封装等其他环节引入的延时。HEVC 以及 VVC 的 AI (All Intra) 的编码为理论值，并未有大量商用。

## 2.2 移动链路回传协议

随着 5G 的部署，远端采集的信号可以通过 5G 等移动通信网络及时回传到转播中心。比较典型的协议包括：RTP、SRT、RTMP、RIST、ZiXi 等。后 5G 时代，随着带宽的增大预计回传协议会朝向更低编码时延、更低压缩率的方向发展。

随着视频业务的快速发展，众多视频流媒体协议不断涌现，并被用于不同的视频 workflow。根据 Haivision 的调研报告 [5]，大多数受访者使用不止一种流媒体协议。对于视频低延迟回传与分发场景，基于 TCP 的 RTMP 实时消息传递协议仍然是最广泛使用的协议，但 RTMP 的部分缺点在应用中逐渐显现，以至于引起了部分 CDN 提供商的忧虑。伴随着网络与视频应用的发展，以 UDP 为核心的流媒体视频传输逐渐成为新的选择，近年来 SRT 的安全可靠传输、WebRTC 的实时互动性能，在诸多场景下展现出更为优异的适用性，出现了逐步取代 RTMP 的趋势。

目前，在视频上行传输方面，RTMP，SRT 和 Web RTC 协议占比较高。

在将 OTT 视频传输到最终用户方面，HLS 和 MPEG-DASH 是最受欢迎的。由于 CMAF 支持 HLS 和 DASH，如开源低延迟 HLS 项目、LHLS 或苹果自己的版本 LL-HLS，它在与其他低延迟交付协议的竞争中发挥着重要作用，预测使用者数量会逐步上升。

### 2.2.1 RTMP

RTMP (Real Time Messaging Protocol, 实时消息传输协议) 协议最初是由 Macromedia 为通过互联网在 Flash 播放器与一个服务器之间传输流媒体音频、视频和数据而开发的，基于 TCP 协议传输，并由 Adobe 采纳，为其 Flash 技术服务，并成为类似 RTP 等复杂机制的替代。但很长时间以来它都是不开放代码的，因此从 2005 年开始，人们开始逆向此协议并发布了若干开源版本。最终 Adobe 在 2012 年发布了 RTMP 标准，但仍保留了部分技术，如 H. 264 先经过一种加密的握手方式才能进行解码播放。随着视频直播领域的兴起，也成为业内广泛使用的协议。由于其基 TCP 的丢包重传能力和可调缓冲区而享有可靠的声誉，但其也存在着累积延迟和加密方面的问题。

RTMP 协议至今仍然被广泛应用，但是近年来在应用中观察到如下问题急需解决：

1. 协议比较古老，最后更新日期是 2012 年，最高只支持到 H.264，对 HEVC/AV1，VP9 等视频格式没有定义，很多都是各 CDN 厂家自行开发的；

2. RTMP Over TCP 连接时间过长，TCP 需要 RTMP c0/s0 到 c2/s2 三次握手，除此之外，其本身又存在 c0/s0 到 c2/s2 的三次握手，再加上 Connection、Createstream、Play/Publish，总地来说 RTMP 完成一次建连需要进行 9 次会话，因此 RTMP 传输时延通常至少会在 2 秒多；

3. RTMP 的传输完全依赖传输层的算法来进行拥塞的管理，不够灵活；

4. 不能实现自适应带宽编码，依赖于 TCP 协议无法提供实时带宽数据对可变码率的支持。

## 2.2.2 SRT

SRT (Secure Reliable Transport) [6]安全可靠传输协议是新一代低延迟视频传输协议，是一种开源、免费和应用灵活的规范，它的性能与专用的协议一样优秀，同时能够在不同制造商生产的产品之间工作。SRT 最初由 Haivision 和 Wowza 开发，目前 SRT 联盟成员超过 450 个，包括 Bitmovin、Brightcove、Canal Cable、Comcast Technology Services、Cinergy、Deluxe、Ericsson、Huawei、Harmonic 和 NGCodec 等少数知名公司以及数十家小型供应商。

SRT 使用 UDP 协议，旨在利用有损网络来确保可靠性。它通过使用“高性能”发送器和接收器模块来实现这一点，该模块不会通过握手确认来阻塞网络。这允许它扩展并最大化可用带宽。SRT 保证发送的分组节奏（压缩视频信号）与解码器接收的分组节奏相同。SRT 增加了专为高效安全的视频流而设计的其它功能：

- 128/256 AES 加密，通过公共网络在链路级提供安全性；
- 内容不可知，并在单个 SRT 流中汇集多个视频，音频和数据（元数据）流，使其能够轻松支持高度复杂的工作流程；
- 支持发送和接收模式（与仅支持单一模式的 RTMP 和 HTTP 相反）；

- 在发送器/接收器模块内，可以检测网络性能，并使用一种自适应比特率和纠正延迟，丢包和抖动；
- SRT 协议具有良好的丢包重传机制，同时支持 ACK, ACKACK, NACK；
- 基于时间的报文发送，防止流量的突发；
- 丰富的拥塞控制统计信息，RTT, Lost rate, inflight, Send/receive bitrate, 利用这些信息做拥塞控制，或自适应 Bitrate；
- SRT 是开源的，对于应用开发者较为友好；
- 可支持当前和下一代压缩技术，即 H264 (AVC) , H265 (HEVC) , AV1, VP9；

SRT 开发了许多新功能，包括传输大文件、小的对话数据等等。但是 SRT 的“传统优势领域”还是实时的视音频传输，SRT 本质上是一个点对点的传输协议（单播而不是组播）。SRT 的亮点在于能够克服有损网络中的抖动和丢包。SRT 目前还是专注于节目的制作和分发而非交付。

SRT 是一种能够在复杂网络环境下实时、准确地传输数据流的网络传输技术，它在传输层使用 UDP 协议，具备 UDP 速度快、开销低的传输特性，支持点对点传输，无需中间服务器中转，可实现低至 120ms 的低延时传输。

SRT 协议基于 UDP 协议，虽然 UDP 协议是一种不可靠传输协议，在互联网抖动与丢包的网络环境下不稳定，但是凭借 SRT 强大的丢包重传的数据恢复能力，前向纠错技术应用等，可将网络丢包的可能性降到最低，确保了 SRT 传输稳定性。同时 SRT 还可以进行 AES 加密，从而确保数据在传输过程中的信息安全。

此外，针对公司或组织运用防火墙保护私有网络安全的策略，SRT 使用的握手过程支持出站连接，而不需要在防火墙中打开危险的永久外部端口，从而维护了公司的安全策略。

### 2.2.3 WebRTC

WebRTC (Web Real-Time Communication) 即“网页即时通信”是一项在网页浏览器内部进行实时视频和音频通话的技术，于 2011 年 6 月由谷歌开源。

WebRTC 允许开发人员使用 HTML 和 JavaScript API 来创建实时应用，而无需下载安装任何插件。

作为典型的浏览器之间的协议，WebRTC 最大的特点是低延时和无卡顿。WebRTC 可以提供点对多点的安全连接，使得多个音频合视频流可以在其连接上流动。WebRTC 提供了视频会议的核心技术，包括音视频的采集、编解码、网络传输、显示等功能，并支持跨平台（Windows, Linux, Mac, Android）。2020 年的疫情驱使视频通信需求剧增，也驱动 WebRTC 面向更大规模的在线并发会议场景演进。与此同时，去噪、用户隐私、打洞等功能逐步完善，大量新发布的规范也逐步涵盖了 WebRTC 传输、安全、数据通道、拥塞控制等方面。2021 年 1 月 26 日，W3C 和 IETF 同时宣布赋能无数服务的 WebRTC 发布成为正式标准[7][8]。

WebRTC 因传输速度快、延迟低，适合应用于对实时性、互动性要求高，但对画质要求不严苛的应用场景。当前在音视频会议、在线教育、即时通讯、游戏及人脸识别等领域，WebRTC 正被广泛采用。与此同时，WebRTC 的兼容性和浏览器不断升级带来的维护成本是使用者需要考虑的问题。2020 年，谷歌在 WebRTC 及其应用上进行了大举投资，以进行代码优化和加强功能集合，使其在多个平台和设备上的运行更加高效和稳定。

2021 年伊始，大量关于 WebRTC/SIP/RTP/SDP 相关的规范（例如：RFC8830-34, RFC8836, RFC8858）密集发布。这些发布的建议规范涵盖了 WebRTC 传输，安全，数据通道，拥塞控制，SIP/SDP 和 RTP 的支持。

## 2.2.4 ZIXI 协议

ZIXI 的传输流协议是一种内容和网络感知协议，可动态调整以适应不断变化的网络条件，并采用纠错技术实现 IP 上网络上的无差错视频流传输。这种动态机制以最小的物理带宽开销，提供低端到端低时延传输。并将实时消除抖动，恢复和数据包重新排序，平滑视频传输，并将视频重新生成为其原始形式。

ZIXI 提供可预测的低延迟、无数据包丢失的卓越可靠性和广播级视频质量（标清、高清和 UHD），没有延迟、分辨率或断断续续的折中。从一个支持 ZIXI 的设备/服务器到另一个支持 ZIXI 的设备/服务器的数据流保护数据流免受路径上质量下降的影响。它能够在任何距离上传输高质量的视频，同时克服公共互联



网不断变化的网络条件，在公共互联网中，网络错误、数据包丢失、抖动和无序数据包的数量“每秒”都会波动。

ZIXI 的传输流技术通过网络感知、动态去抖动、MPEG 特定的优化、Z-ARQ 误差恢复、Z-FEC-动态内容感知、主动多路径错误恢复、UDP 单播或多播的自适应比特率控制以实现最高质量和可靠性。并通过基于 DTLS 标准的保护提供一流的安全性，采用了包括加密和密码保护在内的多层方法的 256 位 AES 传输加密。

ZIXI [9]发端将数据流封装在 ZIXI 传输协议中，并通过标准的 IP 连接进行点对点或多点传输。ZIXI 可以部署在本地或云中，能够监控路径上任何地方的流。对于管理、处理和更大规模的分发功能，ZIXI 可以支持复杂的实时内容生产制作，并能够支持可靠的和可扩展的集群环境中的转码、录制等。

## 2.2.5 RIST

Video Services Forum (VSF) 于 2017 年初成立了可靠的互联网流传输协议 (Reliable Internet Stream Transport, RIST) 小组，为协议创建了通用规范。视频压缩技术的进步和互联网基础设施的普及，使得流媒体在互联网上广泛传输。但是网络丢包一直是一个困扰人们的问题。市面上已经有许多私有的解决方案用于解决流媒体传输的丢包问题，但是由于是私有协议，各个厂商的设备之间无法实现互操作性。RIST[10]旨在解决在公共网络上的丢包问题，同时解决各厂商设备之间缺乏互操作性的问题。

从 2018 年 10 月推出的 tr06-1 规格，到将在 2021 年推出的 tr06-3 规格，RIST 一直在进行更新调整，紧随着时代变化的脚步，产品随着最新的网络技术，压缩技术等保持更新换代，更加贴合时代的需求。现在 RIST 有两种配置文件：主要配置文件 main profile 和简易配置文件 simple profile，另有高级配置文件 advanced profile 预计在 2021 年发行。

简易配置文件 simple profile 的基础流是基于标准 RTP 协议的，且与非 RTP 设备也可适配，其余特性还包括：基于 ARQ 的数据包恢复；非常好的性能（接近 50%的丢包率）；支持多链接支持；且其他公司可以自由地创新并改进的同时不丢失兼容性。

主要配置文件 main profile 在传输时可以将多个流结合到单个 UDP 接口，

这简化了许多的 IT 调试，且连接可以在任一终端进行初始化，同时对于任何类型的 IP 数据也有可支持选项，另外，在加密时，主配置文件相较于简易配置文件还提供可调试的 AES 加密。这些之外，RIST 还有着一些其他特性：通过删除空包来进行频带最优化，支持高比特率的操作。

目前正在开发的高级配置文件 Advanced profile，它可以自动调试网络参数，并动态地根据网络环境的变化进行修正，可以进行拥塞控制，VPN 隧道，在有多流时的进行时间控制，IGMP 侦听，支持 VBR，NAT 穿越防火墙，以及因特网/卫星混合模式等。

开源的 RIST 接口在 Upipe, libRIST, FFMPEG 中都有支持。RIST 可将数据流分为多个流来进行传输，并支持无缝切换和双向传输。

RIST 的技术建议已经通过类似的标准化过程达成一致，目前已经有 40 余个支持成员，并且在快速增长。鉴于这是迄今为止唯一一个团体开发的开放解决方案，有理由相信 RIST 将是许多广播公司的强制性要求。此外，RIST 还允许通过以仅实时传输协议模式运行与非 RIST 接收器的互操作性，有效地使其与 SMPTE2022-1 和 SMPTE2022-2 接收器兼容。

## 2.2.6 GB28181 协议

GB/T-28181 协议[11]是在国际上通用的 SIP 协议进行私有化定制，流媒体方面就是在国际最流行的编码上进行封装。联网系统的信息传输、交换、控制方面的 SIP 监控域互联结构见图 4，描述了在单个 SIP 监控域内、不同 SIP 监控域间两种情况下，功能实体之间的连接关系。功能实体之间的通道互联协议分为会话通道协议、媒体（视、音频）流通道协议两种类型。

区域内的 SIP 监控域由 SIP 客户端、SIP 设备、中心信令控制服务器、流媒体服务器和信令安全路由网关等功能实体组成。各功能实体以传输网络为基础，实现 SIP 监控域内联网系统的信息传输、交换和控制。在若干个相对独立的 SIP 或非 SIP 监控域以信令安全路由网关和流媒体服务器为核心，通过 IP 传输网络，实现跨区域监控域之间的信息传输、交换、控制。

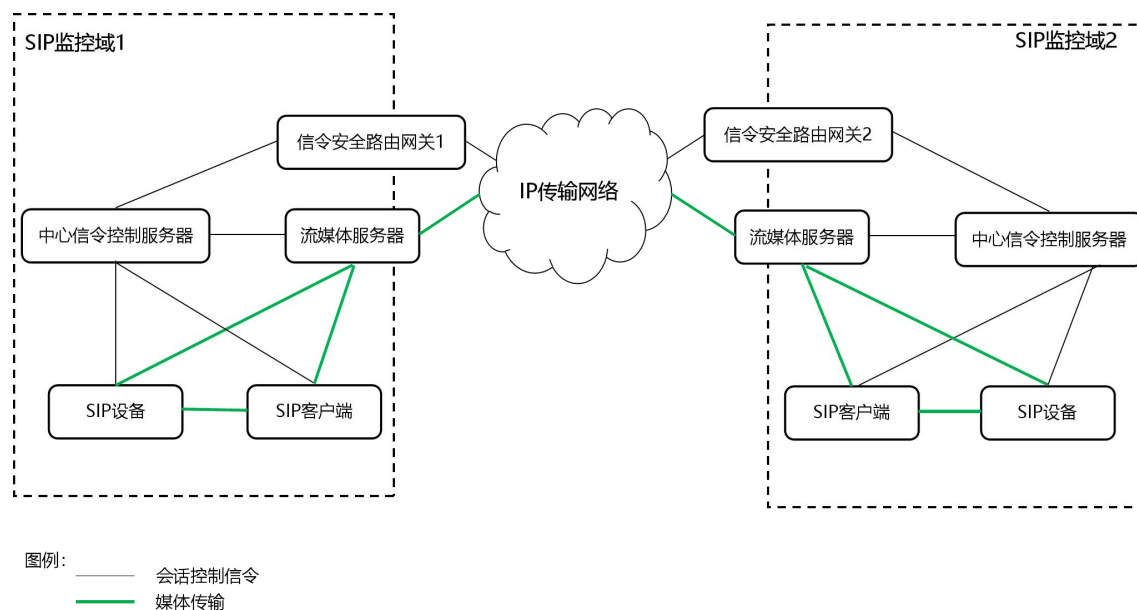


图 4 监控域互联示意图

联网系统在进行视音频传输及控制时应建立两个传输通道：会话通道和媒体流通道。会话通道用于在设备之间建立会话并传输系统控制命令；媒体流通道用于传输视音频数据，经过压缩编码的视音频流采用流媒体协议 RTP/RTCP 传输。媒体流在联网系统 IP 网络上传输时采用 RTP 传输，RTP 的负载应采用如下两种格式之一：基于 PS 封装的视音频数据或视音频基本流数据。

### 2.2.7 ONVIF 协议

ONVIF [12] 是基于网络视频的使用案例，适用于局域网和广域网。规范始于一个核心套接口函数配置和通过定义它们的服务类接口实现控制网络视频设备。网络视频设备包括 NVT、NVD、NVS。该框架涵盖不同网络视频环境下的各个阶段，从网络视频设备部署和配置阶段到实时流处理阶段。这个规范涵盖了设备发现、设备配置、事件、PTZ 控制、视频分析和实时流媒体直播功能，以及搜索，回放和录像录音管理功能。所有的服务使用同一种的 XML schema（XML 文档的结构）。

ONVIF 标准将为网络视频设备之间的信息交换定义通用协议，包括装置搜寻、实时视频、音频、元数据和控制信息等。ONVIF 规范中设备管理和控制部分所定义的接口均以 Web Services 的形式提供。ONVIF 规范涵盖了完全的 XML 及 WSDL 的定义。每一个支持 ONVIF 规范的终端设备均须提供与功能相应的 Web Service。

服务端与客户端的数据交互采用 SOAP 协议。ONVIF 规范的目标是实现一个网络视频框架协议，使不同厂商所生产的网络视频产品（包括摄录前端、录像设备等）完全互通。ONVIF 中的底层流媒体协议主要还是通过 RTP/RTSP 实现。

## 2.2.8 RTMP, WebRTC, ZIXI, RIST 和 SRT 的对比分析

在 4G 时代，RTMP 协议以低延迟著称并被广泛应用。作为基于 TCP 的标准协议，RTMP 兼容性极佳，对用户来说，在现有单向直播架构上，接入成本较低。但随着网络技术的发展，RTMP 协议开始显得相对老旧，不支持 H.265。在缓冲器固定的情况下，当 RTMP 客户端播放端距离服务器越远（RTT 越大），它从服务器获得的流吞吐量就越少。同时，由于 RTMP 基于 TCP 协议，不会丢包，所以当网络状态差时，服务器会将包缓存起来，导致累积的延迟，延迟时间一般在几秒到几十秒。RTMP 无法提供实时带宽数据来动态调整视频编码的 Bitrate，无法实现 NAE 网络自适应带宽编码。因此，逐渐被 WebRTC、SRT 等新进协议取代。

WebRTC 作为 Google 公司的开源技术，降低了音视频通信的接入门槛。WebRTC 使用 UDP 作为传输层协议，它的显著优势在于用户体验好，且仅通过分享链接即可观看视频，非常适合应用于大规模在线并发、有互动性要求的互联网在线直播、培训、会议等场景。许多实时音视频服务专业厂商通过采用 WebRTC 技术，可实现毫秒级的低延迟，远远低于传统 RTMP 的分发延迟。

GB28181 和 ONVIF 都是为了确保安全防范视频监控联网系统信息传输即网络视频在安防市场的应用接口标准的不同，使不同厂商生产的网络视频产品具有互通性的接口标准，其底层流媒体数据传输协议主要还是基于 RTP（实时传输协议），并通过 RTCP（实时传输控制协议）进行数据传输质量控制。在网络传输层，其媒体数据传输可以采用 UDP 或者 TCP，以适配不同的网络组网环境，满足视频监控中各种复杂的应用场景。

ZIXI 协议进入市场较早，目前很多摄像和编解码设备支持 ZIXI 协议。ZIXI 提供卓越的性能（可预测的低延迟）、卓越的可靠性（无数据包丢失）和广播级视频质量（标清、高清和 UHD），没有延迟、分辨率或断断续续的折中。

SRT 协议建立在开源的 UDT 协议上。它强制输入数据加密，可以保护数据安全。它允许在一个连接上混合多个 SRT 流。SRT 试图加快重传速度。SRT 在防火

墙的情况下也可以很好地工作，是目前收到较多关注的传输协议。SRT 在国内应用较多，但目前主要的问题是众多厂家针对 Havision 的 SRT 开源协议做了各自修改，互相之间适配性较差。目前 Haivision 等公司已经推动 SRT 的 IETF 标准制定，预期未来基于统一的 IETF 标准，SRT 的互通性会得到改善。

RIST 需要两个端口，第一个端口用于传输媒体流，并在第二个端口上使用 RTCP 创建了一个控制界面。RTCP 协议是双向的。RIST 的技术路线图分为三种，分别是简单配置、主要配置和先进配置。其中简单配置包括交互 ARQ、重传限制、连接聚合和冗余传输路径。它兼容普通的 RTP 协议，利用 RTCP 协议进行丢包恢复。

RIST 的主要配置包括流加密，多流隧道，高比特率支持等。RIST 的加密采用 DTLS，它和网站采用的 TLS 类似。它还采用预分享的密钥，不需要证明。RIST 采用了 GRE 隧道，非常适合 IPv6 穿透。GRE 隧道消除了额外的端口，并且允许在一个连接中多流复用。GRE 隧道也支持双向流等。RIST 的先进配置包括智能带宽优化，公共通道会话管理，集中呼叫和因特网/卫星混合操作功能。

表 5 ZIXI, SRT和RIST协议的对比

协议	防火墙穿越	ARQ	FEC 前向纠错	加密	链路保护	空包压缩
ZIXI	支持, 两端	支持	支持, 专有的	AES 128/256&DTLS	支持 (SMPTE2022-7)	支持
RIST	支持, 发送方(计划两端)	支持	支持 SMPTE 2022-1	支持(可调试的 AES&DTLS)	支持	不支持(计划)
SRT	支持, 两端	支持	支持	AES 128/256&TLS1.3	不支持(计划 SMPTE2022-7)	支持

ZIXI, RIST 和 SRT 均基于 UDP 协议, 支持高码率的分发能力, 具备 FEC 前向纠错和 AES 加密功能。同时, 在传统的 UDP 协议之上增加了 ARQ 丢包重传。使用 ZIXI, SRT 和 RIST 协议的场景很丰富, 包括摄影机到基站的转播、体育场转播、新闻报道和云转播等等。后期我们将会进行基于 5G 网络的 SRT 和 RIST 传输性能对比测试分析。

### 3. 移动传输域传协议

当前，采用智能手机等移动客户端观看体育赛事、演唱会等实时节目已经成为趋势。基于 IP 的广播协议呈现出了推流效率高、扩展性好、低时延等发展趋势。

#### 3.1 自适应码率传输

在 20 世纪 90 年代，互联网流媒体先驱们在为用户交付视频内容的时候注意到一个严重的问题：不同的用户通过技术连接到互联网，其带宽差异很大：有些人使用拨号，有些人使用 DSL 或 ISDN，还有一些人使用 T1 连接。这个难题使内容发行商陷入两难境地：他们应该用什么视频码率来编码他们的内容？如果他们使用高比特率，他们有可能剥夺大部分观众的权利。如果他们使用一个“最低共同标准”的比特率，那么他们拥有最好连接的最高付费客户将收到低质量的视频。

RealNetworks 公司在 20 世纪 90 年代末推出了一个简单、粗暴的解决方案：以多码率对同一内容进行编码。在会话开始时，让终端用户根据他们的连接速度选择码率。这个简单的解决方案比单一的（而且是次优的）码率编码的方案有了改进，但仍有两个头疼的问题：

1. 它要求最终用户知道他们的连接速度（并考虑到各种开销，如 TCP、IP、以太网和物理层头），这对普通用户来说是不可行的。
2. 它假定带宽在整个流的持续时间内保持固定——鉴于互联网的公共性质和由此产生的高度动态流量和负载，这是一个糟糕的假设。

因此，2007 年，Move Networks 通过多码率编码的概念上增加一些改进，引入了自适应 HTTP 流的概念。他们没有将每个媒体资产编码为一个单一的文件，而是将其编码为一系列可独立解码的小文件或片段。例如将一部两小时的电影编码为 720 个 10 秒的片段。

在多码率编码中，每个变体将被编码为多个片段，并确保每个片段的边界都是对齐的。然后创建一个播放列表，其中包括关于变体数量的信息（内容被编码的不同比特率）以及关于单个文件段的信息（如其位置、持续时间等）。

在会话开始时，播放器设备将下载播放列表，并从一个默认的变体中请求片段（使用 HTTP GET 请求）。当片段被下载时，播放器将不断估计自己和服务器之间的带宽，并在分段边界上的自适应比特率（ABR）阶梯上下移动。播放器的目标是在不超过估计带宽的情况下接收最高质量的视频，同时适应带宽的潜在波动。图 5 用 ABR 阶梯说明了这一概念，ABR 阶梯由三种变体组成（0.5、1.5 Mbps 和 5.0 Mbps）。突出显示的片段是由播放器要求并显示的。

具有讽刺意味的是，这本身并不是一个流传输协议，而是一系列文件下载。文件段下载仍然受益于 CDN 缓存以及通过 TCP 端口 80 的防火墙穿透，并增加了动态带宽适配的好处。该方案的可扩展性得益于这样一个事实，即将动态适配带宽的智能（用于估计带宽和适应带宽）放在客户端，而不是服务器。与 RTP 不同，HTTP 是无会话的，并且不包括 RTCP 消息的交换（尽管底层 TCP 协议确实要求服务器负责错误和流量控制）。



图 5 ABR多码率阶梯示意图

Move Networks 的想法很快被当时的主要流媒体技术提供商采用。微软在 2008 年推出了基于该思路的版本，命名为“Smooth Streaming”。2009 年，苹果和 Adobe 分别推出了 HTTP 实时流（HLS）和 HTTP 动态流（HDS）。虽然这些相互竞争的技术依赖于上述相同的原理，但它们在两个重要方面有所不同。它们使用不同的文件格式来封装音频/视频内容（Smooth Streaming 使用 MP4，HLS 使用 MPEG-2 传输流，HDS 使用 F4V），并且它们对其播放列表文件使用不同的语法。

而这集中不同的技术路线恰恰导致了市场的分裂。不同的设备使用不同的专有技术，无法互操作。内容提供商不得不以多种格式对相同的内容进行编码，而



CDN 则不得不以不同的格式分发和缓存相同的内容—导致存储和带宽的浪费。为了解决这个问题，移动图像专家组（MPEG）在 2011 年制定了 HTTP 动态自适应流（DASH）标准，希望各方都能使用这种单一的方法。DASH 基于碎片化的 MP4（fMP4）容器格式和标准化的清单文件语法，可以实现动态广告插入（DAI）、3D 视频、多摄像机角度、字幕和隐藏字幕以及实时和预先录制内容的混合流式播放等高级功能。

虽然 DASH 现在已经在很大程度上取代了 Smooth Streaming 和 HDS，HLS 仍然被苹果设备广泛使用，这些设备构成了相当大的市场份额。因此 MPEG 推出了多个方案和规范来更好的解决这个问题，其中最成功的就是 CMAF，全称 Common Media Application Format，CMAF 的核心就是促使 HLS 和 DASH 同时使用一种媒体封装格式（加入限制的 ISO BMFF 文件格式），如图 6 所示意。

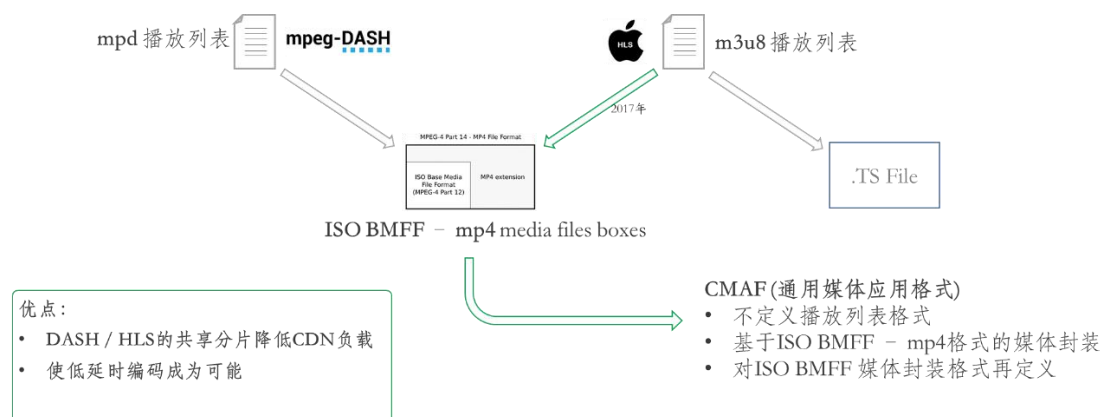


图 6 CMAF为DASH、HLS提供统一的封装

通用媒体应用格式 (CMAF) 的规范旨在为将由 DASH 和 HLS 流式传输的资产提供单一容器格式。此外，通过指定字节范围而不是媒体文件或数据段来请求媒体的概念允许终端设备同时使用 DASH 和 HLS。

与传统的 RTP 相比，DASH 和 HLS（或 CMAF）协议更适合互联网或不可靠网络传输，却有一个让人诟病的缺点：典型的情况下，通过 HTTP 自适应的实时会话会有 30 秒以上的端到端延迟。

为了改善延迟问题，厂商们引入了新的技术，例如分块编码和传输，它们有望将端到端的延迟减少到几秒。其基本思想是编码器将文件段（通常持续时间为 2 到 4 秒）划分为小至 500 毫秒的较小段。然后，每个块都可以单独转发到打包

器(并转发到源服务器,通过 CDN 一直转发到客户端),而不必等待整个数据段编码完成。与数据段本身不同的是,这些数据块不能独立解码,因此码率切换仍然发生在数据段边界上,但通过这个技术,端到端的延迟已经可以缩短到几秒了。

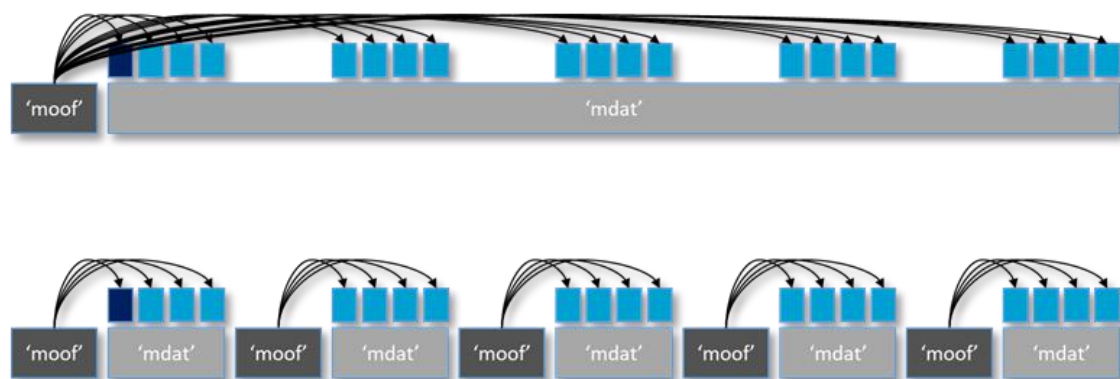


图 7 低时延分片显著降低时延

其中,具有代表性的标准是 LL-HLS (Low-Latency HLS) 和 LL-DASH (Low-Latency DASH)。2019 年 6 月,Apple 在 WWDC 大会上提出了 HLS 的扩展——LL-HLS 标准规范。该扩展与常规 HLS 向后兼容,同时提供了 Apple 认可的方法来降低 HLS 的延迟,主要手段包括: 1) 采用 HLS Partial Segments 机制,允许片段分为更小的部分,类似于 CMAF 定义的 Chunk; 2) 服务端通过 HTTP/2 来实现媒体描述文件与媒体内容的发送传输。2020 年初,Apple 宣布对 LL-HLS 规范草案进行更新。与早期的版本相比,保留了第一个特性,在媒体描述文件(M3U8)中通过#EXT-X-PART 标签来描述 Partial Segment,每个 Partial Segment 可以独立生成、打包和发送,且每个 Partial Segment 可以很短,不必像 Segment 一样必须是 IDR 帧开头,从而达到降低延时的目的。对于第二个特性,此次更新删除了 HTTP/2 的要求,而是引入了一个新标签来描述即将出现的片段。新提出的标签称为#EXT-X-PRELOAD-HINT,借助此标签,发布 LL-HLS 流的服务器可以描述继续播放所需的下一个媒体数据的最可能位置。

2017 年 4 月,LL-DASH 首次被提出,2020 年 3 月,MPEG 发布 LL-DASH Guidelines。LL-DASH 的实现参照了 CMAF 中的 CTE (Chunked Transfer Encoding) 机制,通过使用 HTTP/1.1 Chunked Transport 来传输媒体片段,与 CDN 能很好的兼容,如图所示。此外,LL-DASH 通过“提前通知”的机制,片段

在被完全准备好之前，客户端就可以知道其下载地址，发起下载请求。在传输过程中，通过使用 Chunked Transport，以 Chunk（Chunk 是比 DASH 片段更小的媒体片段，一般为一帧或几帧）为单位进行传输，实现边编码，边传输，从而节省了准备一个完整媒体片段的耗时。在兼容性方面，除了使用 HTTP/1.1 Chunked Transport 来传输媒体片段外，LL-DASH 依然可以解析标准 DASH 的媒体描述文件（MPD），并请求相应的片段进行下载和播放，仅仅是不再具有低延迟特性而已。

尽管 LL-HLS 和 LL-DASH 从设计上可以减低延迟，同时支持多码率自适应，但是这些协议目前仍然在推广当中，实际 CDN 厂商支持不足，缺乏真实大规模使用的成功案例。尤其在国内，目前完全支持 CMAF、LL-DASH 或 LL-HLS 的 CDN 厂家非常有限。

国内流媒体传输使用较多的依然是 HTTP-FLV 的模式，虽然协议比较古老，但基建成熟，架构稳定，终端和操作系统支持较为友好。FLV 协议设计最大的问题是不支持多码率自适应，难以平衡清晰度和流畅度。鉴于 FLV 的产业链成熟优势，业界期望能改进 FLV，使其在保证低延迟的同时，支持多码率自适应。2020 年 6 月，CCSA（中国通信标准化协会）通过了基于流式的直播多码率协议（LAS: Live Adaptive Streaming）的立项（标准计划编号为 2021-0503T-YD）。

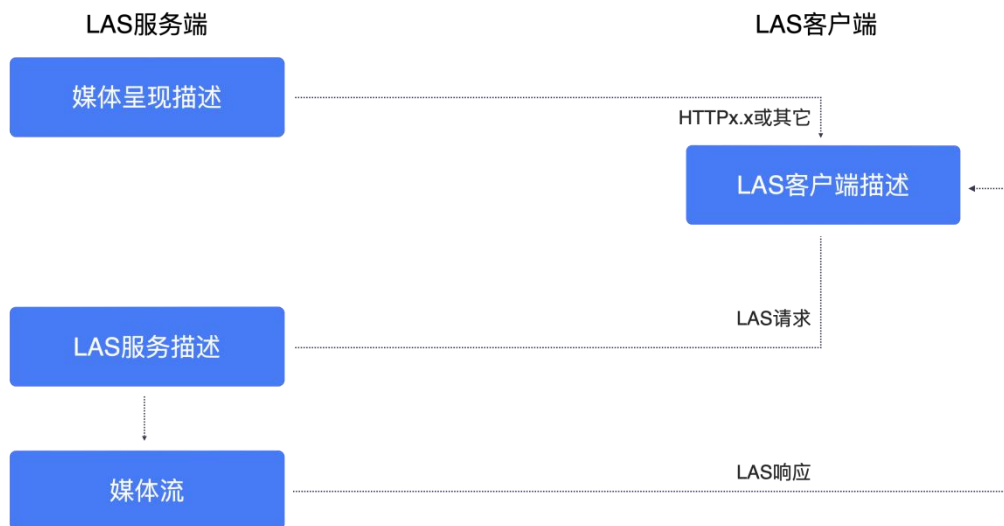


图 8 LAS架构示意图

上图是 LAS 的基本架构图，其基本工作流程为 LAS 客户端首先下载一个媒体

呈现描述文件（类似 DASH 的 MPD 文件或 HLS 的 M3U8 文件），从而得到媒体的相关信息，例如档位信息、地址信息等。紧接着 LAS 客户端向服务端发送 LAS 请求，拉取直播流。LAS 服务端以流式的方式，源源不断的向客户端推送数据。当客户端检测到网络发生变化且需要切换媒体档位时，会再次向 LAS 服务端发送请求，并携带期望吐流的位置信息，服务端会结合请求的地址信息和位置信息，从对应的媒体档位的合适位置开始吐流。值得注意的是，媒体的传输方式是帧级流式传输，从而实现低延迟。同时，在传输过程中，只有在媒体档位发生切换时，才需要发送新的请求，相对基于媒体块的传输方式，可以大大降低请求次数，降低开销。

为了保证不同媒体档位之间的无缝切换，LAS 要求转码时实现关键帧对齐，即不同档位媒体流的 I 帧的 PTS 严格相等。如图 9 所示：

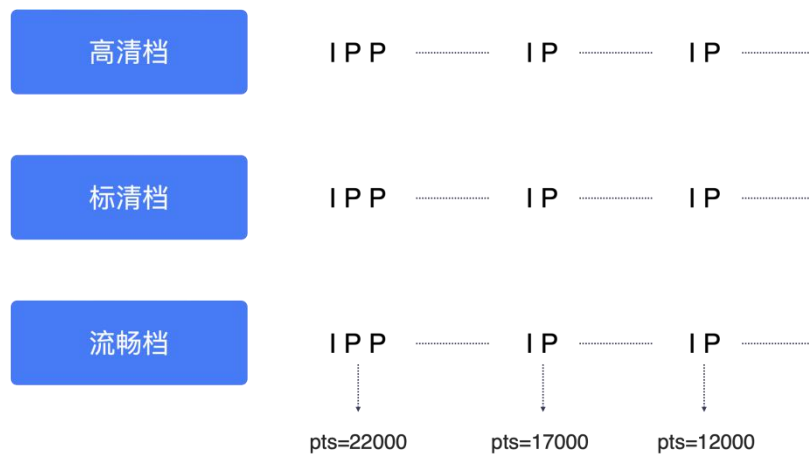


图 9 LAS转码规范示意图

在发生档位切换，需要发送新请求时，通过携带目标档位的 URL 地址信息以及期望开始发送数据的位置信息（I 帧的 PTS），实现不同档位的无缝切换。LAS 支持的媒体切换方式包括 GOP 边界切换和任意点切换。在 GOP 边界切换模式中，只有在一个 GOP 下载结束后，才会触发多码率自适应决策，并允许发送媒体请求，如图 10 所示。这种方式的优点在于实现简单，I 帧对齐天然的保证了无缝切换。其缺点在于自适应决策机会少，当网络发生突变时，不够灵活。

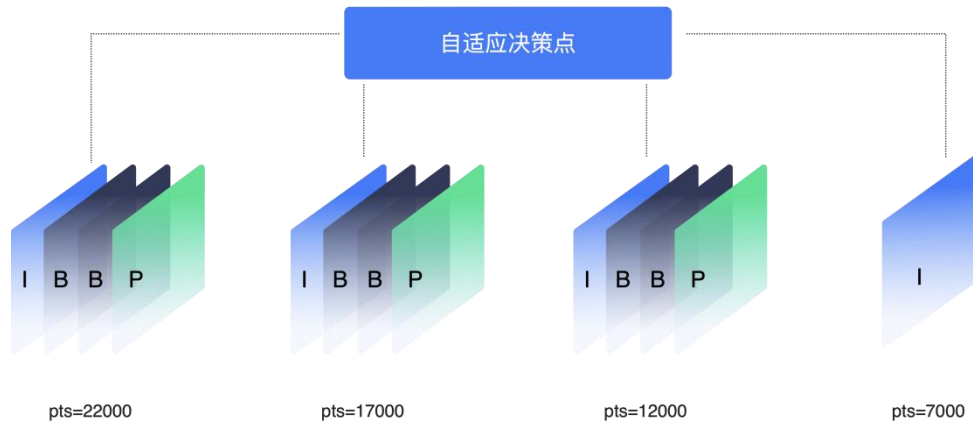


图 10 GOP边界决策示意图

于是，LAS 还支持了任意点切换，如图 11 所示。在任意点决策方案中，任意时刻均可触发自适应算法并发送媒体档位切换请求，易于应对网络的突变，但是实现复杂，需要考虑重复数据和跳帧的平衡问题，详细细节可参考 LAS 标准文档。

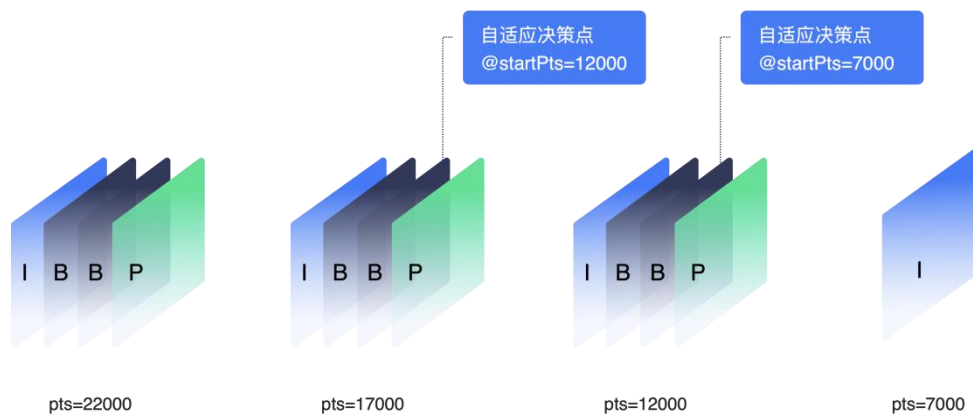


图 11 任意点决策示意图

以目前国内使用最多的 HTTP-FLV 的 GOP 边界决策为例，LAS 客户端通过头信息 (flv\_tag\_header) 来获取当前正在传输的帧的类型。如果为 I 帧，则说明上一个 GOP 下载结束，并记录其对应的 PTS 信息。此时触发自适应策略做码率自适应决策，决定下一个 GOP 的档位。如果需要多码率自适应算法决策需要切换档位，则发送新的请求，并携带上述记录的 PTS 信息即可。

LAS 具有以下几个特性：1) 低延迟自适应：LAS 基于帧级流式传输，相比于传统的 FLV 协议，可以极大的降低延迟。同时，LAS 支持多码率自适应技术，依据实时网络状态动态在多档之间实现无缝切换，平衡清晰度和流畅度；2) 生态

链成熟易部署：LAS 从提出伊始，就得到国内各云厂商的大力支持。目前，阿里、腾讯、百度、金山、华为、网宿、白山等国内主流 CDN 厂商都支持 LAS，且服务行业内头部 APP 企业近两年，服务稳定，部署简单；3) 易扩展：在底层，LAS 支持 HTTP、QUIC 等多种协议，未来将支持 WebRTC，进一步降低延迟，在上层，LAS1.0 版本基于 HTTP-FLV 实现直播拉流，与目前国内使用最广泛的基于 HTTP-FLV 的非多码率架构兼容；4) 开销较小、传输效率高：LAS 没有分片的概念，采用流式传输，直播过程中无需更新 MPD。请求数量少，开销低。目前，LAS 在行业内头部 APP 直播已经全量，每天使用量上亿，稳定性和可靠性得到实际大规模应用的考验。

## 3.2 面向广播组播分发

### 3.2.1 基于 FLUTE 和 ROUTE 的广播分发

FLUTE 是由 IETF (Internet Engineering Task Force) 所制订的通信协议，用于将文件由传送端以多点传送方式，通过 Internet 等 IP 链路传送至多个接收端。FLUTE 构建在 UDP 协议之上，运行时不需要任何由接收端回传至发送端的反馈信息。因此具备优秀的扩展性，对于接收端的数量几乎没有限制，无论是 10 万个或是 100 万个接收端都可以正常工作。FLUTE 被 DVB-H 和 3GPP MBMS 同时采纳为单向传输协议。在两个广播标准中，FLUTE 既可以传输广播的业务信息，也用来 ESG 等传输广播辅助信息，应用非常广泛。

FLUTE 协议的主体部分是 ALC[15]协议，两者的主要差别在于，ALC 是一套单向的“对象”多点传送通信协议，而 FLUTE 则是一套单向的“文件”多点传送通信协议。ALC 所传送的对象本身，并不具任何的属性。FLUTE 通信协议针对 ALC 的最主要扩充，就是将 ALC 传送的对象视为文件，并为每个对象加上档案所需要的属性，例如文件名称、档案长度及档案类型。为此，FLUTE 额外定义了一种叫 FDT (File Description Table, 档案描述表) 的数据结构，里面记录了 ALC 对象的档案属性。

ALC 是一种 IP 多播通信协议，可适用于 IPv4 与 IPv6，传输的基础是 UDP 协议，传输文件采用“尽力而为”方式。传统的 IP 多播协议本身并没有对话管理、拥塞控制、以及提供可靠传输的能力。ALC 通信协议建构于 IP 多播协议之上，并通过 LCT 协议解决了上述三个缺点，具备更好的可用性。LCT 是 ALC 通信协议的主体部分，负责提供前述的 进程管理的功能。拥塞控制是其中一个可选的组成组件，负责 ALC 在 Internet 上的拥塞控制。通常在移动单向广播网并不会发生拥塞问题，所以拥塞控制这个组件有时不会用到。ALC 的可靠性是通过 FEC 提供的前向纠错功能保障的，应用层的 FEC 协议（如 Raptor）不需要来自接收端的回馈信息，通过预先提供冗余信息来弥补 ALC 封包在传送时所发生的遗失或错误。ALC 的 FEC 组件相对独立，目前比较常用的是 IETF 定义的 Raptor[17]和 RaptorQ[18]。

FLUTE 协议本是设计用于传输非实时文件的，随着实时视频流业务的增加，FLUTE 协议的缺点逐渐显露，包括：

- 每次传输只能传输一个文件
- 不能区分传输文件的时延特性，无法针对实时视频等文件优化时延
- 需要接收完整数据后才能播放

在 ATSC 3.0 协议中，为了进一步降低广播传输时延，在 FLUTE 协议的基础上设计了 ROUTE 协议。ROUTE 协议同时针对媒体传输协议 DASH 做了优化。

- ROUTE 协议支持多路传输，多个 LCT 进程可以与一个修复进程绑定，以保证所有业务保持稳定的 QoS
- ROUTE 协议支持乱序接收，以适应互联网传输的时延抖动
- ROUTE 协议支持不同业务封包的可见性，即在传输时可以根据被传输对象（实时媒体或者非实时文件）的时延特性优化传输的参数和设置。实时媒体数据可以在完成整个文件传输之前就开始播放。

ROUTE 协议稍后被纳入 DVB-Adaptive media streaming over IP multicast 规范中[19]。为了更好的将 ROUTE 协议推广到其他支持 IP 广播的协议中，IETF 正在制定独立的 ROUTE 协议[20]。

### 3.2.2 基于 BIER 的组播分发

BIER (Bit Index Explicit Replication) 是一种简化的组播报文转发技术。它是一种基于位索引显式复制的新型组播技术,提供一种无状态的组播转发机制。

该技术不需要显式建立组播分发树,也不需要中间节点保存任何组播流状态。它消除了复杂的组播协议、组播转发表,无需维护任何组播流状态,只需交换组播发送者和接收者的信息, BIER 网络节点仅根据网络拓扑进行转发,就可实现高效的进行组播报文的分发。

IETF RFC8279 将 BIER 组播架构分为 Overlay、BIER、Underlay 三层。

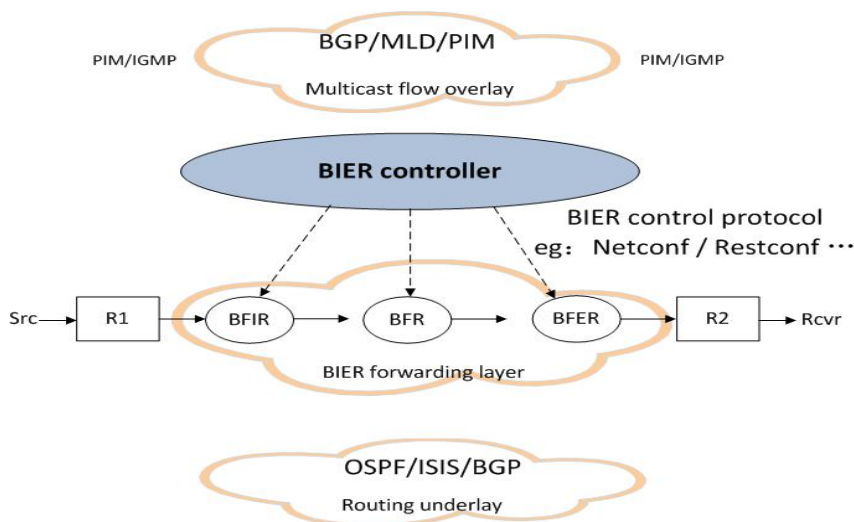


图 12 BIER协议的三层架构示意

**Overlay 层:** 业务层,负责组播业务控制面信息交互, BIER Egress 出口节点和 BIER Ingress 入口节点之间用户组播的加入和离开、组播流进入和离开 BIER 域的封装和解封装转发等。

**BIER 层:** Bier 转发,主要完成 BIER 路由信息发布和洪泛以及本地 BIER 转发表计算更新。

**Underlay 层为:** Bier 链路建设,通过链路状态协议如 ISIS、OSPF 等扩展 TLV 属性携带本节点的 BIER 信息。

BIER 在组播首节点 (BIER Ingress) 确定组播的接收者 (BIER Egress) 信息,中间节点不需要维护任何组播流转发状态信息 (Group、Ingress、Egress),



BFIR 是最靠近组播源的 BIER 路由器。BIER 本地转发表根据 IGP 的 BIER 链路状态库计算生成，BIER 链路状态库则由 IGP（ISIS/OSPF）协议的 BIER 扩展洪泛生成。

大型 BIER 网络可以根据网络拓扑或者地理位置设计多个 SD（Sub Domain）简化管理，如全国性的运营商设立如东部大区 SD 网络、西部大区 SD 网络、南部大区 SD 网络、北部大区 SD 网络等。每个 SD 内的 BSL 和 BFR-id 是独立的，互不影响。

Bier 组播网络中没有维护组播状态，无需组播控制协议运行，中间节点看不到任何组播流，也无需保存任何组播流相关信息，与用户及业务量的增减无关，所以与业务完全解耦，而传统的 PIM 组播技术需要保存组播状态及组播树的维护，对于组播成员的加入、退出都需要进行管理，需要业务参与流程的管控。

表 5 Bier与传统PIM组播的比对

	BIER 组播	传统 PIM 组播
网业耦合度	网络服务层次化，组播业务与组播转发完全解耦，可以任意形式搭配	网络服务扁平化，组播业务与组播传输紧耦合
路径收敛	快：BIER 网络在路由方面，只需要单播路由协议支撑，继承单播的快速重路由（fast reroute, FRR）等路径保护技术。基于路由收敛时，因为不需要组播路由收敛过程，单播路由收敛完成，就代表 BIER 完成收敛，比传统组播收敛快。	慢：所有节点都需要维护组播树状态，网络拓扑变化时都需要先做单播收敛，当单播收敛后，组播协议再为每个组播树重新收敛处理，从而收敛到新的组播树，时间较长，收敛速度慢。
可维护性	中间节点无须维护多播状态及业务相关转发表项，大大减少了	需要支持大量的组播协议簇，运营运维复杂

	业务转发所需的复杂协议，运营运维方便	
可扩展性	BIER 网络不需要组播路由，也不需要维护业务相关的路由表及业务状态，转发表项只和拓扑相关，表项非常小。增删中间转发节点，不需要组播业务相关的配置，相当于即插即用，扩展方便，业务灵活。	所有节点都需要维护组播路由，任何一次增删节点都需要更新组播树，扩展性差。
计费管控	易管控：BIER 在 BFIR 上可以完全控制业务流投递的接收端（BFER），这样基于一个控制点，就可以控制业务的流向和接收者，摆脱了传统组播控制困难、难以收费的困境。	难管控：缺乏灵活的管控手段，难以进行计费管控
移动性支持	灵活快速：当组播源移动时，只需要由控制器切换组播源关联的 BFIR 就能解决问题，避免传统组播树重建过程或者流量绕行的问题。当接收端移动时，只需要在 BFIR 修改位串，就可以正确投递组播业务	源可能在不断在变化，切换速度频繁，每次变更都需要重建组播树，无法满足移动终端的快速切换需要，所以对移动 IP 中的应用支持困难。

### 3.3 单点分发 HTTP1/HTTP2/HTTP3

历史上第一个有记载的 HTTP 版本是 0.9 (The HTTP Protocol As Implemented In W3)，它诞生在 1991 年。这个协议目前已经过时，其最早被设计用于从服务器获取 HTML (HyperText Markup Language) 文档。

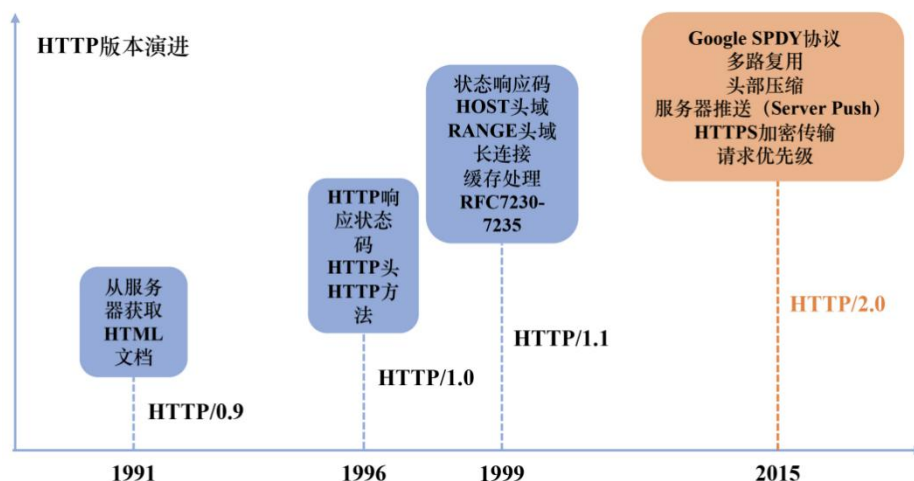


图 13 HTTP 协议发展示意

#### 3.3.1 HTTP/1 协议

1996 年，一个更加完整、更加接近我们目前对 HTTP 认知的版本，HTTP/1.0 (Hypertext Transfer Protocol - HTTP/1.0) 发布了。这个版本中已经包含了很多我们如今都比较熟悉的概念：HTTP 响应状态码、HTTP 头和 HTTP 方法。随着互联网的迅速发展，人们对 HTTP 协议有了更高的要求。1999 年，现在最常见的 HTTP 1.1 (Hypertext Transfer Protocol - HTTP/1.1) 版本诞生了。从此之后，这个 HTTP 协议一直服务至今。并且，在后来的十多年里，这个协议还不断更新和细化，最终在 2014 年形成了 6 个 RFC (RFC7230-7235)。相比于 HTTP/1.0，HTTP/1.1 支持长连接（或持久连接），亦即在一个 TCP 连接上可以传送多个 HTTP 请求和响应，减少了每次建立和关闭连接的消耗和延迟。从而在需要一次请求多个文件/图片等资源的情况下，通过一个 HTTP 连接来完成传输，节约资源加载延迟，提高网络带宽利用率。同时，HTTP/1.1 还增加了

更多的请求头和响应头来改进和扩充 HTTP/1.0 的功能，使得用户站点可以支持更多的功能，比如内容缓存，带宽优化，消息传递，错误提示，内容协商等。

### 3.3.2 HTTP/2 协议

2015 年，HTTP/2.0 (RFC 7540 - Hypertext Transfer Protocol Version 2 (HTTP/2)) 诞生。HTTP/2.0 支持很多新的特性，主要的几个特性如下：

1. 新的二进制格式 (Binary Format)，HTTP/1.x 的解析是基于文本。基于文本协议的格式解析存在天然的缺陷，文本的表现形式有多样性，要做到健壮性的话，考虑的场景必然很多。但是，二进制则不同，该方式只看 0 和 1 的组合。基于这种考虑 HTTP/2.0 的协议解析决定采用二进制格式，实现更加方便，而且健壮。
2. 多路复用 (Multiplexing)，即连接共享，即每一个 HTTP Request 都是用作连接共享机制的。一个 HTTP Request 对应一 ID，这样一个连接上可以有多个 HTTP Request，每个连接的 HTTP Request 可以随机的混杂在一起，接收方可以根据 HTTP Request 的 ID 将 HTTP Request 再归属到各自不同的服务端请求里面。多路复用原理如图所示，在单条 TCP 连接上可以存在多个 HTTP Response，不同的 HTTP Response 对应于不同的 HTTP Request。
3. 头部压缩 (Header Compression)，在介绍 HTTP/1.x 时，我们知道在 HTTP/1.x 的“头”中带有大量信息，包括各种状态响应信息等，而且每次建立 HTTP Request/HTTP Response 都要重复发送 HTTP “头”，HTTP/2.0 使用 Encoder 来减少需要传输的 HTTP “头”大小，通讯双方只需要各自缓存一份 Header Fields 表，这样既避免了重复 HTTP “头”的传输，又减小了需要传输数据的大小。
4. 服务端推送 (Server Push)，和 Google SPDY 一样，HTTP/2.0 也具有 Server Push 功能。所谓 Server Push 指的是 HTTP/2.0 服务器在还没有收到客户端发送的 HTTP Request 时，HTTP/2.0 服务器就可以把各种还未被请求的资源文件推送到客户端。

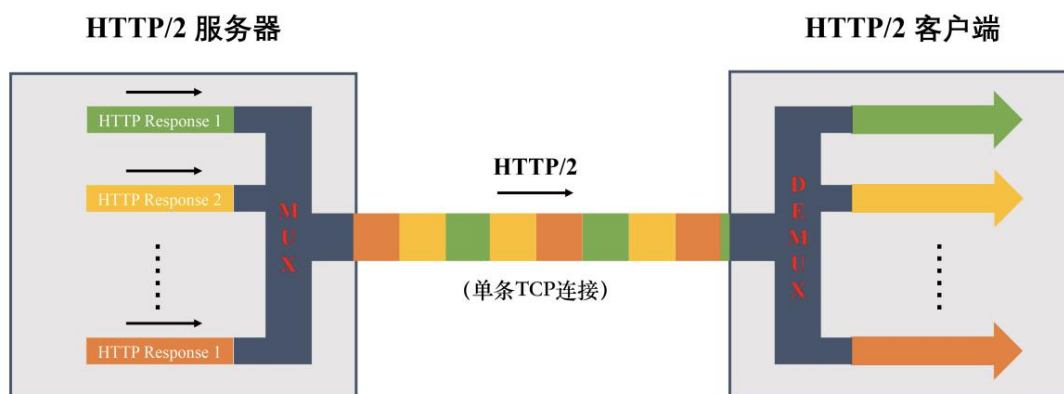


图 14 HTTP/2服务端推动

目前，有大多数网站已经启用 HTTP/2.0，例如 YouTube，淘宝网等网站。此外，Nginx 与 Apache 也已经支持 HTTP/2.0 及其相关特性。

下图表示了 HTTP/1.x、HTTP/2 多路复用与 HTTP/2 Server Push 在网页中请求文件的示意图，在 HTTP/1.x 中，客户端需要请求网站首页 index.html，格式文件 style.css 与图片 home.jpg 等资源时，需要一个一个 HTTP Request 发送，在一个 HTTP Request 发送之后，必须等到对应请求的 HTTP Response 到达客户端之后，才能继续发送下一个 HTTP Request。HTTP/1.1 虽然支持 TCP 长连接，但是其还存在相同的问题，并且一般而言，很多浏览器可支持的长连接数目有限，在资源很多的情况下，同样存在低效的网络带宽利用情况。而 HTTP/2 的多路复用则可以通过 Stream 机制，该机制是一种抽象概念，允许多个 HTTP Request 与多个 HTTP Response 在同一个 TCP 连接上存在，从一定程度上提高了链路带宽的利用率。但是即使如此，客户端与服务端由于每次发起 HTTP Request，等待 HTTP Response 都存在一定的延迟，从而浪费了有限的链路带宽资源。HTTP/2 的 Server Push 则允许服务端主动推送一系列资源文件，而不需要客户端一个一个发起 HTTP Request，从而大大提升了网络请求与响应的性能。可以看出基于 HTTP/2 Server Push 来请求网站首页 index.html，格式文件 style.css 与图片 home.jpg 三个资源文件所需要的时间开销最短，延迟最低，这也意味着用户会得到更好的网页浏览体验。

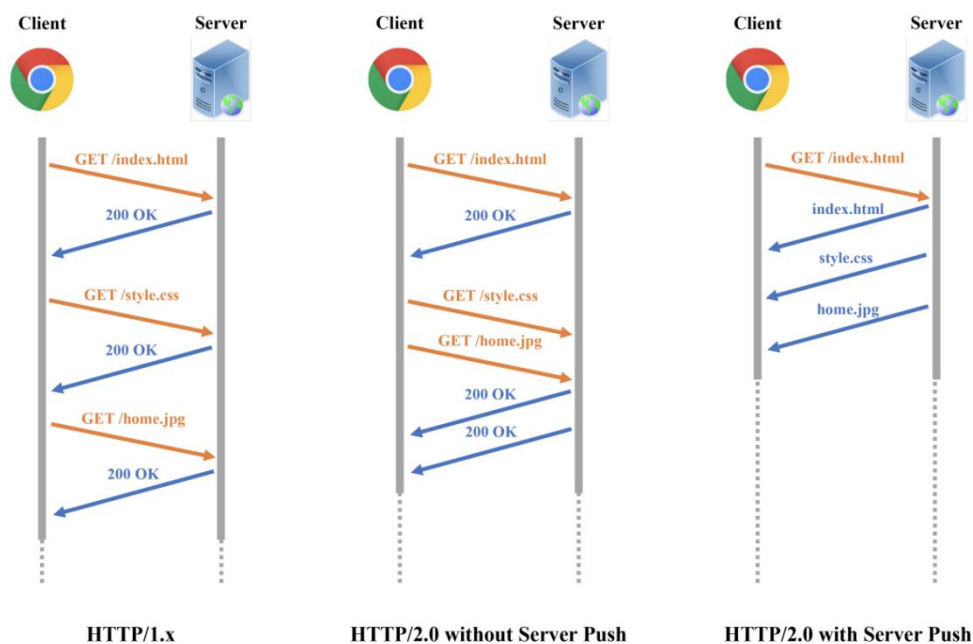


图 15 HTTP/1.x, HTTP/2.0无Server Push和HTTP/2.0结合Server Push流程比较

### 3.3.3 HTTP/3 协议

虽然 HTTP/2 解决了很多之前旧版本的问题，但是，由于其底层还是 TCP 协议，因此，还是存在一些问题。HTTP/2 的缺点主要有以下几点：

1. TCP 与 TCP + TLS 建立连接的高延时问题（一个典型的获取资源的等待延迟可能达到 4 个 RTT）
2. TCP 队头阻塞问题（TCP 多路复用带来的弊端）

Google 在推 SPDY 的时候就已经意识到了这些问题，2012 年便实现部署了一个基于 UDP 协议的“QUIC”协议，让 HTTP 跑在 QUIC 上而不是 TCP 上。2018 年，IETF 相关工作组提出将“HTTP over QUIC”更名为 HTTP/3，这进一步推动了“HTTP over QUIC”逐渐演变为 HTTP 协议的下一个版本。目前，HTTP/3 仍然处于草案状态。HTTP/3 在 HTTP/2 的基础上又实现了质的飞跃，一定程度上解决了“建连高延迟”与“队头阻塞”问题。

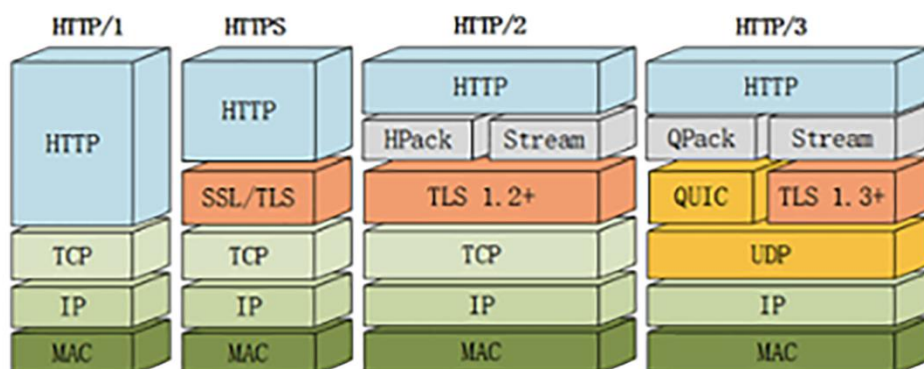


图 16 HTTP/3和前代HTTP协议比较

QUIC 虽然基于 UDP，但是在原本的基础上新增了很多功能：

1. 实现了类似 TCP 的流量控制、传输可靠性的功能
2. 实现了快速握手功能
3. 集成了 TLS 加密功能
4. 多路复用，彻底解决 TCP 中队头阻塞的问题

HTTP/3 有以下优点：

1. 使用 stream 扩展了多路复用特性。在 HTTP3 模式下，一般传输多少个文件就会产生对应数量的 stream。当这些文件中的其中一个发生丢包时，只需要重传丢包文件的对应 stream 即可
2. 通过 UDP 建立，在用户空间保证传输的可靠性，UDP 之上的 QUIC 协议提高了连接建立的速度，降低了延迟
3. 引入 Connection ID，使得 HTTP3 支持连接迁移以及 NAT 的重绑定
4. 含有一个包括验证、加密、数据及负载的 TLS 安全机制
5. 将拥塞控制移出了内核，通过用户空间来实现，不再需要等待内核更新可以实现很方便的进行拥塞控制算法快速迭代
6. 使用 QPACK 压缩方案，优化了对乱序发送的支持，也优化了压缩率

## 3.4 VR 内容的分发

近年来，以 VR、AR 为代表的虚拟和增强现实终端成为行业的热点。在行业中，这两类终端被统称为 XR（eXtended Reality）。有别于传统的智能手机和平板电脑，XR 终端需要集中呈现使用者视角内的图像，而虚化视角边缘的图像。同时，需要及时根据使用者的头部移动方向和角度，调整投射的图像。

在 XR 类业务中，VR 类业务是通过封闭式眼镜提供独立视频，而 AR 业务则是结合周围环境提供交互类视频。本章重点介绍 VR 类业务的呈现和传输协议，AR 类业务呈现方式类似，但需要增加与周围环境的交互，对通信更加依赖，对交互时延也更加敏感。

### 3.4.1 VR 媒体服务总体架构

虚拟现实指通过展现 360 度的视频为终端用户提供沉浸式体验的技术，支持全景视频内容，也称 360 度全景视频或沉浸式视频，支持用户交互性地切换观看视角，终端能够根据用户的观看视角动态渲染图像、视频及其相关联的音频。

如图 17 所示，VR 媒体内容由源站提供，在现实场景中采集到的视频等媒体数据经过 VR 源站拼接服务器进行拼接、投影、旋转等，成为完整视频画面，继而经过编码器编码得到 VR 媒体内容源，利用封装服务器完成系统层封装；用户通过 VR 终端请求 VR 媒体服务，其中传输协议访问模块完成 VR 媒体索引文件的解析，终端根据传感器获取到的视窗元数据（如用户头动信息、观看方向等）请求相应的内容分片，并完成解封装、解码，最终根据视窗元数据完成视频画面的渲染和音频等其他媒体资源的播放。



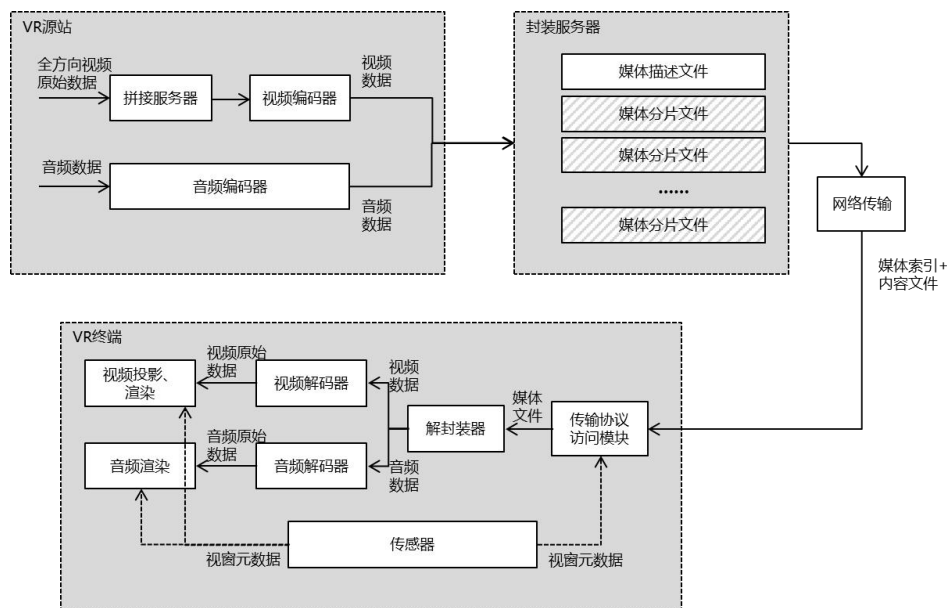


图 17 VR媒体内容传输流程

### 3.4.2 VR 视频传输模式

VR 视频传输模式包括：基于全景传输的虚拟现实基本传输模式，即视窗独立传输模式，以及基于主视场或者辅助视场的多码流切换的虚拟现实视点自适应传输模式，即视窗依赖传输模式。

视窗独立传输模式指将 360° 全方向视频以同等质量、完整地发送给用户。可以保证映射内容完整保留了原始球面的所有内容，保留信息量最大。客户端向服务器请求获取无差别的全景视频文件，当用户观看方向发生变化时，所有的处理都在终端完成。

在沉浸媒体的消费过程中，由于人眼视觉范围有限，用户在某一时刻只能观看局部的内容。按需传输，利用人眼视觉系统的局限性，实现在不降低视觉体验的前提下对数据量进行减少，依赖用户当前的视野对传输内容进行自适应，按需下载视频内容，即视窗依赖传输模式，也称之为 VR FOV 传输。

视窗依赖传输模式可以分为两类：基于区域封装的视窗依赖传输和基于分块编码的视窗依赖传输。

对于基于区域封装的视窗依赖传输模式，全方向原始球面视频内容采用非均匀映射处理。其在对球面内容进行采样时，令球面上的像素点有不同的权重，使

得关键视频内容得到保留，而不重要的区域被下采样，仅保留少部分关键信息。非均匀映射方法用于传输质量不均匀的 360° 全景视频，用户视窗范围内是高分辨率，其他区域是低分辨率，从而减少整体码率。

分块传输技术将 360° 全方向视频按照空间划分为若干个子视频块，客户端可以根据网络状况和用户头部运动有针对性的向服务器端请求视频片段。分块传输仅传一部分内容，或将当前视窗的高质量视频内容以及低质量全景视频内容混合传输，减少了传输数据量，可以自由地选择各个分块的质量。

### 3.4.3 VR 媒体封装及传输格式

VR 视频封装格式可参考 MPEG 制定的 OMAF (Omnidirectional Media Format) 全景视频封装格式，以及表 6 所示的视频基本配置规范及更高配置规范。具体内容参见[22]。

表 6 VR视频基本配置规范

媒体规范	编码格式	视频编码规范	Level	数据盒封装标识	类别
基于HEVC编码格式的视窗独立视频封装	HEVC	Main10	5.1	podv、erpv	hevi
基于HEVC编码格式的视窗依赖视频封装	HEVC	Main10	5.1	podv、erpv与ercm至少一个	hevd
基于AVC编码格式的视窗依赖视频封装	AVC	Progressive High	5.1	podv、erpv与ercm至少一个	avde

虚拟现实音频可参考表 7 基本配置规范及更高配置规范。具体内容参见[22]。

表 7 VR音频基本配置规范

媒体规范	编码格式	音频编码规范	Level	最高采样率	3D 元数据	类别
OMAF 3D 音频基准	MPEG-H	Low Complexity	1, 2 或	48 kHz	编码中已包	oab1



规范	Audio	y	3		含	
OMAF 2D 音频规范	AAC	HE- AACv2	4	48 kHz	无3D元数据	oa2d

VR媒体传输协议要求和智能手机类似，也是以DASH和HLS为主：

DASH: VR 媒体服务可以基于 DASH 传输协议以及针对 VR 媒体服务的信令扩展，具体配置内容参见[22]附录 B。

HLS: VR 媒体服务可以基于 HLS 传输协议以及针对 VR 媒体服务的信令扩展，具体配置内容可参考[23]。

## 4. 小结和音视频传输趋势预测

视频类应用是被业界广泛认可的 5G 发展方向之一。随着移动互联网的发展，以智能手机为代表的移动终端成为主流的视频接收端。特别是近年来，中高端智能手机已经普遍具备 4K 和 HDR 视频处理能力，有 AI 加持的 GPU 甚至超过了电视等大屏显示器的处理能力。为了提升用户体验，主流的视频 APP 提供商陆续推出 4K 内容，这也驱动了运营商加速部署 5G 网络以提供相应的流量。可以预期，随着运营商 5G 网络的部署和升级，用户将会获得更大的体验带宽，届时 APP 服务商也会进一步的提高帧率、HDR 量化比特数等关键指标以提升用户观看体验。

在上行方向，5G 将速率提升了两个数量级，目前上行峰值速率已经接近 2Gbps。这极大的扩展了 5G 上行业务的范畴，背包业务从 20Mbps 以下的超高压缩率、低画质内容扩展到 80Mbps 以上的高画质内容。同时，5G 超大上行链路帮助电视制作域从有线走向无线。利用 600Mbps 的大带宽，5G 可以把 4K 编码器输出的浅压缩信号在 50ms 内传送回节目制作中心，屏幕到屏幕的时延可以控制在 80-100ms 以内。这是电视制作域在从基带信号进入信号全 IP 化之后的又一次革命性变化，信号传输从光纤、网线的有线连接进入全无线连接时代。继光纤取代 SDI 线缆后，演播室和拍摄现场的线路部署被进一步简化，现场部署的复杂度被极大降低，“剪尾巴”后的摄像机可以自由移动，不再被限定到固定拍摄机位。

5G 在上行和下行两个方向提供了超大的传输速率，为了更好的支持制作域的上行传输和传输域的下行分发业务，可能还需在传输协议层面加以适配。本文从制作域和传输域分别总结了当前主流的传输协议，并分别加以分析。

在视频制作域，JPEG-XS 和 NDI 等协议目前主要通过光纤和千兆、万兆以太网等固定链路传输。5G 提供了可以媲美千兆以太网的传输速率，目前已经有一些利用 5G 传输 JPEG-XS 和 NDI 等浅压缩信号的实验，预计在不久的将来就会出现早期的商用方案。视频背包类业务目前主要采用 HEVC 或 J2K 等深压缩方案结合一些基于 ARQ（SRT、ZIXI 以及 RIST）或传统的 RTP、RTMP 的方案进行传输，预计未来会根据场景考虑采用不同的传输方式，例如在专业视频制作使用高可靠的 SRT、ZIXI 或 RIST 协议、在直播或会议类应用中使用交互性好的 WebRTC 协议、在视频监控类应用中使用的 GB28181 协议等。现阶段在远程视频传输协议的选择

上，SRT、ZIXI 以及新推出的 RIST 协议各有所长，如果需要立刻部署的话，使用 SRT 或 ZIXI 会是一个更合适的方案，因为其公网传输的优异性能以及成熟度已经被业界证明；而如果长远规划的话，RIST 会是更面向未来的选择。

在视频传输域，底层传输协议需要智能手机从操作系统层提供良好的支持，HLS 和 DASH 是两大主流自适应码率传输协议，分别在 IOS 和 Android 平台占据主导，具有最佳的生态支持。同时，业界的各 OTT 大厂通过自身 APP 强大的渗透率也在基于自身业务的需要、分别演进适合业务发展的协议。这类演进主要是从 CDN 和智能手机支持的成熟协议出发，加入适合业务类型的技术改进，解决自身业务发展遇到的“痛点”。本文中 LAS 就是根据目前被广泛使用 FLV 协议设计了多码率自适应传输，优化了时延特性。

表 8 5G 场景下的视频传输协议

应用场	传输协议	编码格式	优点	缺点
专业视频制作	MPEG-TS (UDP/RTP)	AVC/HEVC/VVC J2K/JPEG-XS	成熟度高、生态完善、适用范围广、开放	协议陈旧、对网络要求高
	SMPTE 2110	未压缩 /JPEG-XS	业界标准、面向未来、开放	尚未大规模部署、对网络要求极高、部署成本高
	NDI	AVC/HEVC	成熟度高、生态完善、产品化效率高、部署效率高	单一厂商(Newtek)垄断风险、对网络要求高
5G回传或公网传输	RTMP	AVC (HEVC需扩展)	成熟度高、生态完善	延时大、公网传输能力不足、协议陈旧、扩展能力弱
	SRT	ANY	成熟度高、生态完善、延时低、开源、大量成功案例	单一厂商主导技术发展、存在垄断风险；不同厂家实现存在互通问题
	ZIXI	ANY	成熟度高、延时低、纠错能力强	完全私有
	RIST	ANY	业界标准、开放、生态完善、兼容RTP及	实现程度及成熟度低

			SMPTE-2022	
公网分发	DASH	AVC/HEVC/VVC /AV1	业界标准、生态完善、开放、基于HTTP-ABR、MPEG阵营	延时大、低延时方案全产业链支持尚需时间；LL-DASH终端支持问题需解决
	HLS	AVC/HEVC	成熟度高、协议简单、基于HTTP-ABR、Apple阵营	延时大、低延时方案全产业链LL-HLS支持尚需时间；Apple私有协议
	LAS	AVC/HEVC	成本低、延时低、基于HTTP-ABR、生态完善	国外产业链缺乏支持,推广可能会出现问题
直播带货或视频会议	WEBRTC	AVC/HEVC/VP8 /VP9/AV1	业界标准、生态完善、开源、浏览器支持、极延时低、可变码率及可变质量、Google阵营	扩展难度大、CDN支持复杂度高
视频监控	GB28181	AVC/HEVC	N/A	暂无对比

随着国内基于互联网的直播类业务迅速发展，广播和组播这种传统的业务类型得到了重视。本文介绍了移动网络广播分发协议 FLUTE/ROUTE，FLUTE 协议被 3GPP eMBMS、DVB 等移动广播系统广泛采用，ROUTE 是 FLUTE 协议的优化版本，被 ATSC 3.0、DVB 协议采纳。BIER 协议是一种简化的组播报文转发技术，未来在运营商网络中有一定的应用前景。目前直播类节目使用最为广泛的还是基于 IP 的广播技术，在传输时实际使用 HTTP 单播发送到用户。为了优化时延和可靠性，目前 UDP 和 TCP/IP 类的协议互相吸纳了彼此的优点，产生了一些类似 QUIC 的融合类协议。

作为近年来广受关注的终端类型，本文还简述了针对 VR 的媒体服务架构、传输模式、封装和传输格式等信息，以期能够为广受关注的 VR 业务提供参考。但我们也看到目前 XR 类业务还有 AR、MR 等业务仍需研究。

## 参考文献

- [1] 中央广播电视总台，《8K超高清电视节目制播技术要求（暂行）》，2021年1月
- [2] 中央广播电视总台，《4K超高清电视技术应用实施指南（2018版）》，2018
- [3] SMPTE ST 2022, Moving Serial Interfaces (ASI & SDI) to IP, August 2017
- [4] SMPTE ST 2110, Structuring the Future of Broadcasting, Nov. 2017
- [5] Haivision, “BROADCAST IP TRANSFORMATION REPORT 2020-The State of IP and Cloud Adoption in the Broadcast Industry”, 2020
- [6] Haivision, “SRT Protocol technical overview”, October, 2018
- [7] Web Real-Time Communications (WebRTC) transforms the communications landscape; becomes a World Wide Web Consortium (W3C) Recommendation and multiple Internet Engineering Task Force (IETF) standards, <https://www.w3.org/2021/01/pressrelease-webrtc-rec.html.en>
- [8] WebRTC transforms the communications landscape as it becomes a W3C Recommendation and IETF standards, <https://www.ietf.org/blog/webrtc-standardized/>
- [9] ZiXi, Broadcaster User Guide, V1.9,  
[http://pro.jvc.com/pro/attributes/SEG/manual/BR-800\\_User\\_Guide\\_V19.pdf](http://pro.jvc.com/pro/attributes/SEG/manual/BR-800_User_Guide_V19.pdf)
- [10] Rick Ackermans, RIST Activity Group Chair, "RIST; What is the Future";
- [11] GB/T 28181-2016 公共安全视频监控联网系统信息传输、交换、控制技术要求, 2016年7月12日
- [12] IETF, RFC 3551, RTP Profile for Audio and Video Conferences with Minimal Control, July 2003
- [13] IETF, RFC 3926 - File delivery over Unidirectional Transport, Oct. 2004

- [14]3GPP, TS26.346, Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs, v16.4.1, Mar. 2020
- [15]IETF, RFC 3450, ALC protocol instantiation, Dec. 2002
- [16]IETF, RFC 3451, LCT building block, Dec. 2002
- [17]IETF RFC 5053, Raptor FEC scheme, Oct. 2007
- [18]IETF RFC 6330, RaptorQ FEC scheme, Aug. 2011
- [19]ETSI TS 103 769, DVB Adaptive media streaming over IP multicast, Nov. 2020
- [20]IETF ROUTE draft, <https://datatracker.ietf.org/doc/draft-zia-route/>
- [21]IETF RFC 8279, Multicast Using Bit Index Explicit Replication (BIER) , Nov. 2017
- [22]ISO/IEC 23090-2, Omnidirectional Media Format, 2019
- [23]AVS, 《信息技术虚拟现实内容高效编码 第1部分：系统》
- [24] 《BIER组播技术发展白皮书》  
<https://max.book118.com/html/2020/1127/7153042125003023.shtm>
- [25]YD/T 3941-2021 《内容分发网络技术要求VR音视频服务》  
<https://www.antpedia.com/standard/1142100987-1.html>
- [26]ISO/IEC 23090-2 Omnidirectional media format  
<https://mpeg.chiariglione.org/standards/mpeg-i/omnidirectional-media-format>



## 致谢

本报告在撰写过程中收到来自以下高校, 研究机构和公司专家的大力支持和贡献, 在此报告完成之际, 仅表达诚挚的感谢。

4K Garden-4K 花园	于路 (Lu Yu)
ABS-国家广播电视总局广播电视科学研究院	张宇 (Yu Zhang)
Ateme	陈朋奕 (Ben Chen)
Beijing Radio and Television Station-北京电视台	程宏 (Hong Cheng)
B&M Modern Media Inc. -广州波视信息科技股份有限公司	王兆春 (Violet Wang)
China Unicom-中国联合网络通信集团有限公司	黄蓉 (Rong Huang)
Communication University of China-中国传媒大学	林涛 (Tao Lin)
Communication University of Shanxi-山西传媒学院	任石青 (Shiqing Ren)
Haivision	李翠琳 (Cuilin Li)
Hangzhou Arcvideo Technology Co., Ltd.-杭州当虹科技股份有限公司	陈勇 (Yong Chen)
JiangSu Broadcasting Cable Information Network Corp. Ltd. (JSCN) - 江苏有线	李鑫 (Li Xin)
Kwai - 快手	周超 (Chao Zhou)

Qualcomm - 高通	曹 一 卿 (Yiqing Cao) 杜志敏 (Zhimin Du) 李 俨 (Yan Li)
Samsung - 三星	吴越 (Yue Wu)
Syntronic - 新拓尼克 (北京) 科技研发中心有限公司	王 瑞 明 ( Ruiming Wang)
ZTE-中兴	刘 耀 东 ( Yaodong Liu) 陶长标 (Changbiao Tao) 白雅贤 (Yaxian Bai)

