White Paper | January 2014

radisys.

Real-Time Multimedia Transcoding

IP Media Conditioning for Mobile, WebRTC, and IMS Architectures

Overview

Internet Protocol (IP) networking technology is the clear winner in delivering converged networks between fixed and wireless domains. But even with a converged IP network, different media codec standards and encapsulation formats are creating challenges to achieve interworking between a diversity of endpoint technologies.

This whitepaper discusses real-time multimedia transcoding, voice quality enhancement, and media adaptation—what we collectively called IP Media Conditioning—in IP communication networks with respect to service provider requirements and the impact on network architecture.

CONTENTS

Introduction

The evolution of telecommunications has resulted in a proliferation of network technologies and media encoding approaches and standards. Internet Protocol (IP) networking technology is the clear winner in delivering converged networks between fixed and wireless domains. But even with a converged IP network, different media codec standards and encapsulation formats are creating challenges between a diversity of endpoint technologies.

Codecs are the encoding/decoding technologies that define the algorithms used to convert data streams. For example, many VoIP solutions use the ubiquitous G.711 codec to encode analog voice to digital voice for 64 Kbps circuit-switched networks or Real Time Protocol (RTP) packet streams for IP networks. However, alternative codecs provide similar audio fidelity with lower bandwidth, while newer IP broadband networks support high-definition (HD) audio codecs providing superior audio quality.

Similar innovations are occurring with codecs used in video communication services. While H.264 is the prominent video encoding standard today, new video encoding standards and approaches are improving video resolution and resiliency with less bandwidth.

Codec diversity is increasing the complexity of the network equipment that must process IP media streams. Hence, the requirement for real-time multimedia transcoding, the process of converting from one encoding format to another, will continue to increase in modern IP communication networks. For video endpoint interworking, requirements also include transrating the video streams between different picture sizes and frame rates. A well-designed multimedia transcoding architecture maximizes interoperability between disparate devices and technologies, while allowing operators to standardize on a smaller subset of codecs. Additionally, the same equipment performing transcoding can perform other media conditioning functions, including voice quality enhancements and bearer channel adaptation.

This whitepaper discusses the delivery of real-time multimedia transcoding in IP communication networks with respect to service provider requirements and the impact on network architecture.

Transcoding and Media Conditioning in IP Communication Networks

The explosion in the number and variety of audio and video codecs used in modern telecommunication networks is making it challenging for network architecture design. Although some codec standards help carriers by requiring less bandwidth, the need to support numerous codecs increases network management complexity and requires greater media processing investment.

Audio Transcoding

The G.711 codec (circa 1972) is still widely used to convert voice signals to a digital stream, but other codecs have emerged to fulfill different needs, particularly low bandwidth requirements or high-definition (HD) audio codecs. Hence, the scope of audio transcoding in a carrier network might include narrowband codecs like G.711, G.726, G.729, AMR, EVRC and iLBC, high-definition (HD) audio codecs like G.722 and AMR-WB, and the newest HD audio codecs including SPEEX, SILK, or OPUS.

Transcoding allows Communication Service Providers (CSP) to reduce the variety of codecs in their core networks and centralized service platforms. Imagine a CSP supporting an Interactive Voice Response (IVR) application for users in New York. Figure 1 shows an example of 4 IVR subscribers accessing the IVR service using 4 different devices—each device using a different codec. The CSP could deploy an IVR service platform with multi-codec flexibility, however this would add extra media processing costs to the services infrastructure.

Instead, a more pragmatic and economical approach might be to implement the IVR service using the commonly-used G.711 codec. The implication of this approach is that somewhere in the call path between the service platforms and endpoint, a transcoding function would be required to convert G.711 media streams to G.729, AMR-NB, and AMR-WB, as shown in Figure 1.

This simple example already highlights a number of key points and observations:

- Transcoding allows devices, like the IVR platform and the mobile phone in the example, to communicate with each other even when they support different codecs.
- One of the key benefits of transcoding is that once applied near the network edge, the remaining network elements in the core network now only need to support a single codec—G.711 in this example—thus lowering the computing requirements and, consequently, equipment cost in the core service platforms.
- In the case of the PSTN phone subscriber, they are already using G.711 as well. So no transcoding is required between the IVR platform and the PSTN phone. Instead, a media gateway function would be required to convert a G.711 RTP packet media stream from the IVR platform, to a 64 kbps G.711 circuit connected to the PSTN phone. Media gateways are a common location for transcoding, hence the relationship between transcoding and media gateways is a key theme that will be explored throughout this whitepaper.
- Finally, while a VoLTE subscriber has the capability and benefit of using an HD quality AMR-WB codec, the IVR services would be heard and experienced using a conventional G.711 narrowband audio experience. Hence, with the increasing market penetration of HD audio devices, including VoLTE or even WebRTC services and endpoints, service platforms in the core will also need to evolve to support HD quality as well.

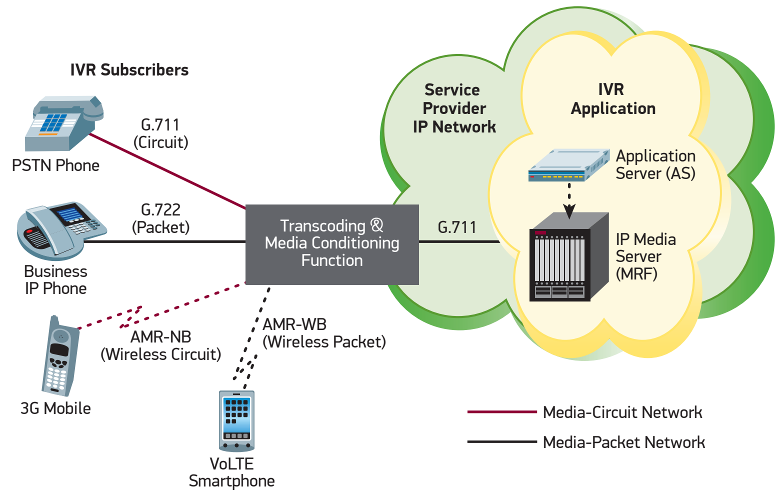

Figure 1. *Transcoding for a Hosted IVR Application*

Audio Transcoding in VoLTE

3GPP standards have specified the use of AMR-Wideband (AMR-WB) codecs for VoLTE audio communications. While the audio quality between two VoLTE devices will be excellent, and far superior to our baseline expectations around 3G mobile communications, the reality is that the migration from 3G to VoLTE will take a few years. Hence, chances are high that early adopters of VoLTE devices will still do a majority of their voice calls connecting with 3G devices (using AMR-NB codec), or to the PSTN (via G.711 circuits). Hence, AMR-WB to AMR-NB (or G.711) transcoding capacity is an important network requirement for mobile operators rolling out VoLTE services.

Video Transcoding and Transrating

In the video world, the diversity of various video stream formats introduces additional dimensions of disparity, including video picture size, frame rate, sampling speed and bandwidth. New codecs are emerging to deal with the accelerating consumer demand for video across mobile networks.

H.264 Adaptable Video Codec (AVC) is the prominent video encoding standard used in IP communications today. H.264 Scalable Video Codec (SVC) delivers more resilience in lossy networks, while WebRTC standardization efforts are advocating the use of VP8 video codec. Soon, emerging H.265 and VP9 codecs will grow as a percentage of endpoints. Video transcoding will be increasingly required in a modern IP network to achieve interworking between various video codecs.

Transrating is also often required and applied to a video stream. Transrating re-encodes a video stream to a different bit-rate or frame rate without changing video content, which ultimately saves a significant amount of bandwidth. For example, transrating could be used to convert high-resolution video clips to a lower resolution (and lower bandwidth) video stream suitable for a handheld device with a small screen or limited availability of bandwidth, particularly in wireless networks. Therefore, video transrating and re-scaling is an important media processing function depended upon to deliver what is often described in the industry as a 4-screen strategy for television, desktop, tablets, and mobile devices.

Additional Media Conditioning

In addition to audio transcoding, and video transcode/transrate, carrier networks often require other media stream processing capabilities, referred to as media conditioning in this paper. These requirements might include:

- Voice Quality Enhancement (VQE)
- Bearer Channel Adaptation

Voice Quality Enhancement

Voice quality is still a major concern for many network operators that offer telephony services based on VoIP technology. Despite providing tremendous economic benefits, VoIP also presents new voice quality challenges such as dropped packets, variable packet delay, and packet reordering and corruption. Cost-effective solutions, generally referred to as Voice Quality Enhancement (VQE), are available to effectively address these issues, thereby increasing end-user satisfaction.

Video Transcoding for WebRTC-to-IMS Interoperability

A current topic in IETF standardization efforts surrounds the selection of a video codec standard for WebRTC services. The web community, including the originators of WebRTC, are advocating the VP8 codec, which has excellent video encoding qualities but (so far) limited adoption beyond early WebRTC services. Meanwhile, the majority of the video endpoints deployed today are based on H.264 AVC codecs, so telecom standards experts believe that H.264 should be the standard, or at least a 2nd choice alongside VP8. Picking only one video codec for the WebRTC standard would have an immediate side effect of increasing the need for video transcoding infrastructure in the network for any services requiring interworking between H.264 (IMS) and VP8 (WebRTC). And even in a scenario where both video codecs eventually get standardized in WebRTC standards, video transcoding or other media conditioning processing would still be required.

VQE solutions typically incorporate features designed to overcome background noise, packet loss, acoustic echo and variable network delays, the four most common sources of audio quality problems in a VoIP network. A comprehensive VQE solution also measures VoIP quality metrics, which are used in ongoing voice quality measurement associated with service level agreements. Many VQE features require sophisticated real-time digital signal processing, particularly noise reduction and acoustic echo cancellation.

Bearer Channel Adaptation

In addition to the audio and video media stream processing, media conditioning might include broader bearer channel adaptation features such as normalizing between IPv4 and IPv6 addressing schemes, dynamic bit rate adaptation, or implementing access and bandwidth controls. Media conditioning might also include adapting media streams for security, such as interconnecting Real-Time Protocol (RTP) media streams from IMS endpoints, to Secure RTP (SRTP) endpoints used with WebRTC. All these requirements are applicable to the scope of media conditioning in a modern next-generation IP network.

Implications for Media Processing Platforms

Audio transcoding, Video transcoding and transrating, and media conditioning functions require powerful platforms and architectures that deliver both fast digital signal processing and real-time performance. These functions are often best performed on hardware platforms leveraging digital signal processor (DSP) technologies, although the performance of software media processing running on COTS servers, and virtualized media processing for cloud deployments, are increasingly suitable for scalable IP media conditioning in a modern communications network. Sufficient real-time performance is needed to minimize the delay introduced by applying a media conditioning function against the media stream. Finally, carrier class communications systems require fully redundant services and fast service restoration times.

Architectural Considerations

Transcoding and media conditioning requirements are broad based, arising from a wide range of factors and service requirements such as codec diversity, network bandwidth consumption and media quality enhancement. The challenge is to implement trancoding, transrating and other media conditioning functions at the lowest cost, while optimizing quality. Every transcoding operation requires processing against a media stream, which introduces a slight delay that could negatively impact voice or video quality. Hence, it is important for network designers to address transcoding and media stream processing requirements based on informed architectural decision making.

Network Locations for Media Conditioning

Three alternative network locations have emerged as the key contenders for performing media conditioning in a packet communications network. These network locations and functional definitions are consistent with specifications put forward by leading standards organizations such as the 3rd Generation Partnership Project (3GPP) specifications for IP Multimedia Subsystem (IMS), Telecoms & Internet converged Services & Protocols for Advanced Networks (TISPAN) and the MultiService Forum (MSF). The three network locations include the:

- *Multimedia Resource Function (MRF),* also known as an IP media server, provides media related functions such as media manipulation (e.g., voice and video stream mixing), playing media clips (tones, announcements), or recording and later playback (messaging applications), under the control of the application servers in a network.
- *Media Gateway (MGW)* provides media conversion functions between Time Division Multiplex (TDM) circuits and Real Time Protocol (RTP) media packet streams in the IP core network.
- *Border Gateway Function (BGF)* provides security with IP network connectivity at the border between autonomous IP networks.

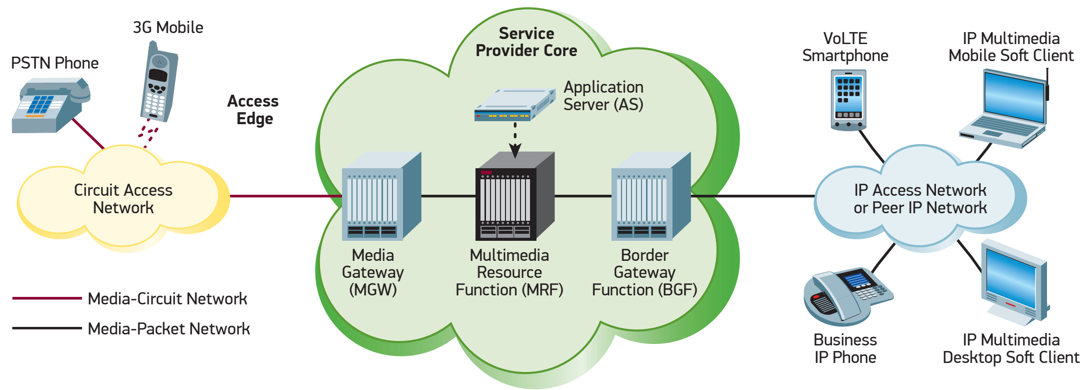

Figure 2. *Network Locations for Transcoding and Media Conditioning: Core and Edge*

Figure 2 shows the location of these network elements in a simplified next-generation network. The MRF is normally associated with service processing in the service provider core, while the MGW and BGF elements are associated with the access edge. The following section describes the particular merits and implementation characteristics related to the network location (i.e., service provider core or access edge) where media conditioning is performed.

Media Conditioning in the Service Provider Core

Located in the service provider core, the MRF already performs automatic transcoding today depending on the service processing requirements. Hence, the MRF is also a desirable network location for many media conditioning functions in the network.

The MRF provides media processing services for IP communication applications, including media play/record, phone digit collection and audio/video mixing (see sidebar). However, in this primary role, the media server may also need to perform automatic transcoding or voice quality enhancement. For example, an IP media server supporting heterogeneous conference call mixing uses transcoding to normalize the call legs to a common codec before it mixes the audio and then re-encodes each conference leg to the end-points codec requirements.

Media Resource Function (MRF)

Carriers deliver numerous telecommunication services and capabilities such as unified communications, multimedia conferencing, voicemail and ringback tones. While all these services have differing application logic and signaling requirements, the underlying media processing requirements share many common characteristics, such as playing an audio or video clip, collecting digits from a phone, bridging multiple signals into a conference mix, or transcoding media streams using different encoding standards. The MRF provides a common, shared media processing resource for supporting a wide range of audio, video, fax and speech applications in VoIP and IMS networks. There are many similarities when processing voice and video media streams.

In addition to supporting service-level media processing, the MRF also performs various media conditioning functions, such a automatic transcoding or voice quality enhancement. Because an MRF is designed to deliver sophisticated IP media stream processing capabilities and capacities, the MRF is also a fitting technology for multimedia transcoding and media conditioning at network and access edge locations, beyond its more common deployment location in the IMS services core.

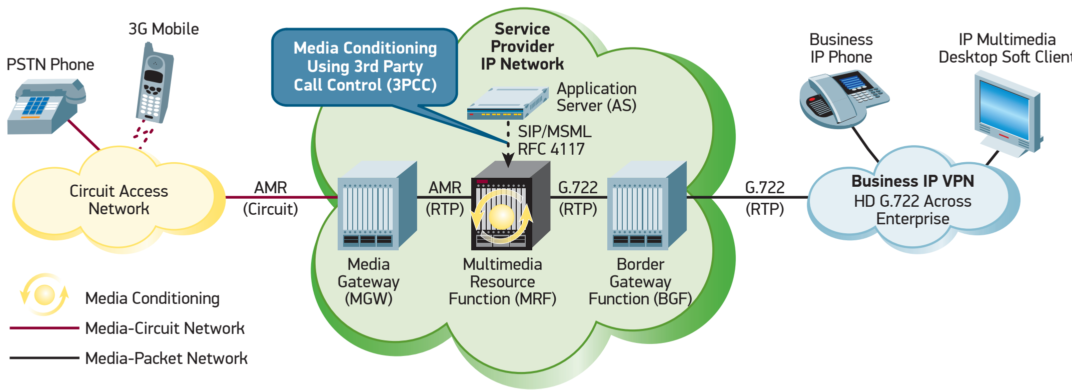

Figure 3. *Media Conditioning in the Core using 3rd Party Call Control (3PCC)*

When media conditioning is controlled by an application server, it supports a topology called 3rd Party Call Control (3PCC), as shown in Figure 3. The topology makes a distinction between signaling and application logic processing as separate and distinct functions from real-time media processing performed by the MRF against Real-Time Protocol (RTP) audio and video packet streams. The application server "controls" the MRF, typically using Session Initiation Protocol (SIP) commands. Media conditioning controlled by 3rd party application logic facilitates granular control to apply specific media conditioning treatments within the IP media server against each RTP media stream.

Transcoding control can be achieved using a full-featured IP media server interface, such as SIP with Media Server Markup Language (MSML– RFC 5707), used for rich IVR, conferencing or voice quality enhancement (VQE) service creation and feature control. Alternatively, a SIP-based interface, such as RFC 4117 could also be used, which can support transcoding for simpler two leg or two party RTP connections.

Supporting SIP and RTP

The two widely used protocols for VVoIP signaling control and media plane processing are the Session Initiation Protocol (SIP) and the Real-Time Transport Protocol (RTP), respectively. At a high-level, signal plane processing vs media processing could be separately addressed using two separate network elements (decomposed model) or by one network element (integrated model).

The decomposed model is illustrated in Figure 3, where the IP media server handles RTP flows, and the application server manages the SIP flows. The integrated model example in Figure 4 shows media gateways or border gateway elements handling both RTP and SIP flows.

The decomposed model allows service providers to scale SIP and RTP traffic processing independently and avoid burdening an edge element, like a border gateway, with two workloads. Alternatively, the integrated model does not require an extra network element (i.e., signaling control), which can save equipment cost and simplifies deployment, presuming the edge element has sufficient media processing power.

There are several benefits of performing media conditioning tasks, such as transcoding, on the MRF in the service provider core, including:

- Media processing is consolidated into a centralized function, which enables functional synergies, simplifies management and achieves processing economies of scale.
- The close proximity and integration of service-level media processing and stream-level media conditioning facilitates new service deployment and expands capabilities.
- Edge devices, like media gateways and border gateway functions, don't have to support multiple codecs, which saves cost on network edge equipment with a high number of network instances.
- Voice quality is often better than performing transcoding on multiple edge elements because there are fewer transcoding operations in the end-to-end call path.

It is worth noting that many basic point-to-point calls typically do not require media processing applied by a centralized MRF. Furthermore, if all calls that require media conditioning are routed into the service provider core, backhaul bandwidth requirements may be unnecessarily large and possibly introduce network bottlenecks. While centralized media conditioning offers benefits for calls already requiring MRF processing, most network architects should also consider incorporating media conditioning in access edge elements as well.

Media Conditioning at the Access Edge

Situated on the edge of IP networks, media gateways and border gateway functions, pictured in Figure 4, are also logical options for transcoding or media conditioning. Their location is ideal for achieving codec "normalization" (e.g., converting all access codec varieties into G.729) since transcoding is performed before media streams enter the core network.

Media Gateways (MGW) are deployed at the border of circuit-based access networks, like the public switched telephone network (PSTN), 2G, or 3G cellular networks, and the IP core. The primary function of a media gateway is to convert media circuits on one side into RTP over IP media packet streams on the other side. In performing this function, the media gateway may also perform transcoding or other basic media conditioning enhancements, like acoustic echo cancellation. However, media gateways have limitations that impact their ability to effectively perform transcoding for the entire network.

- Primarily supporting G.711, media gateways often lack the computing headroom to take on more complex codecs, a critical requirement for transcoders.
- More importantly, media gateways are not a viable network location with the growing number of video transcoding requirements, or IP-to-IP media conditioning operations required in a modern IP network.

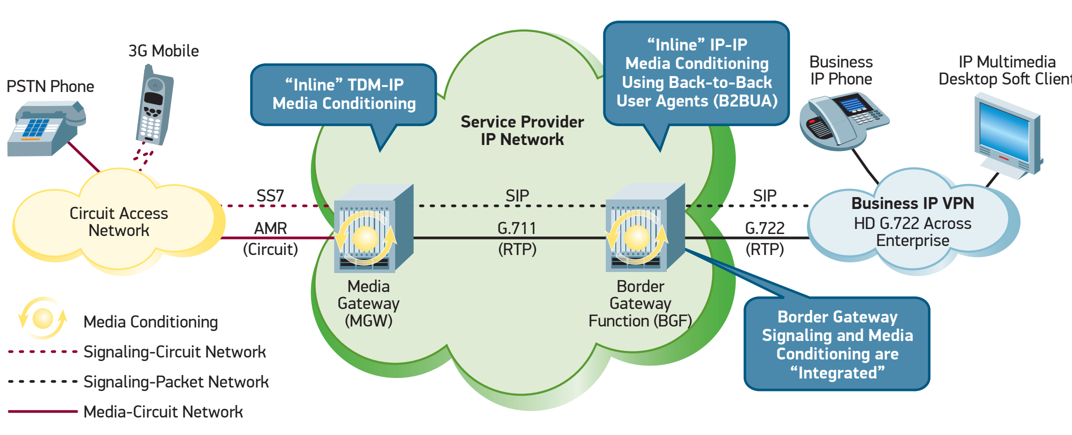

Figure 4. *Inline Media Conditioning at the Edge*

The **Border Gateway Function (BGF)** is located at the border of two IP networks in order to provide security with IP network connectivity. For example, a BGF can be deployed between service provider's IP network and an enterprise IP VPN.

Media conditioning at the access edge is often achieved using an "inline" approach. Media gateways provide inline TDM-IP media conditioning, whereas a BGF performs inline IP-to-IP media conditioning typically using a Back-to-Back User Agent (B2BUA) topology, as shown in Figure 4. Unlike the 3PCC topology, a B2BUA approach does not require a 3rd party controlling element. Instead, media conditioning running on a network edge element is typically configured once, and then the BGF processes both the signaling and media streams passing through.

Figure 4 also shows an increasingly common situation where a large enterprise might standardize their IP-based phonesets or softphones using a high-definition G.722 audio codec. While internal calls would achieve great audio quality, the enterprise still needs to make external calls to customers and suppliers using a more common G.711 baseband codec. Hence, the BGF peering point between the service provider's network and the Business IP VPN is an ideal location for performing G.722-to-G.711 transcoding.

A Border Gateway Function that processes both signaling and media streams is sometimes called an integrated Border Gateway Function, or more commonly, a **Session Border Controller (SBCs)**.

While there's industry consensus that SBC resources are best applied to border gateway functions, as confirmed by IMS architecture, the same IMS standards do not specify transcoding and media conditioning as an SBC function. Although an SBC occupies an optimal network location for media conditioning, including media conditioning in an integrated SBC product presents some design and economic drawbacks:

- SBC hardware platforms are primarily designed for signaling security applications and network connectivity, and they are typically not well-suited for media processing.
- SBCs often lack the computing resources to cost-effectively scale multimedia transcoding and media conditioning functionality, a burden that can lower their capacity and increase price per port.

Alternatively, the Border Gateway Function could be decomposed, which separates signaling gateway and signaling security functions, performed in a Border Control Function (BCF), from media encryption and media conditioning in the decomposed BGF. Figure 5 shows how this implementation starts to resemble the 3PCC approach applied to the access edge, for both Access Border Gateway Functions (A-BGC) to enterprise IP VPNs, and Interconnect Border Gateway Functions (I-BGF) with 3rd party carrier peer IP networks.

A key benefit of decomposed border gateway architecture is the independent scaling of the border gateway function, on a platform optimized for a "firehose" of high volume IP media stream processing,

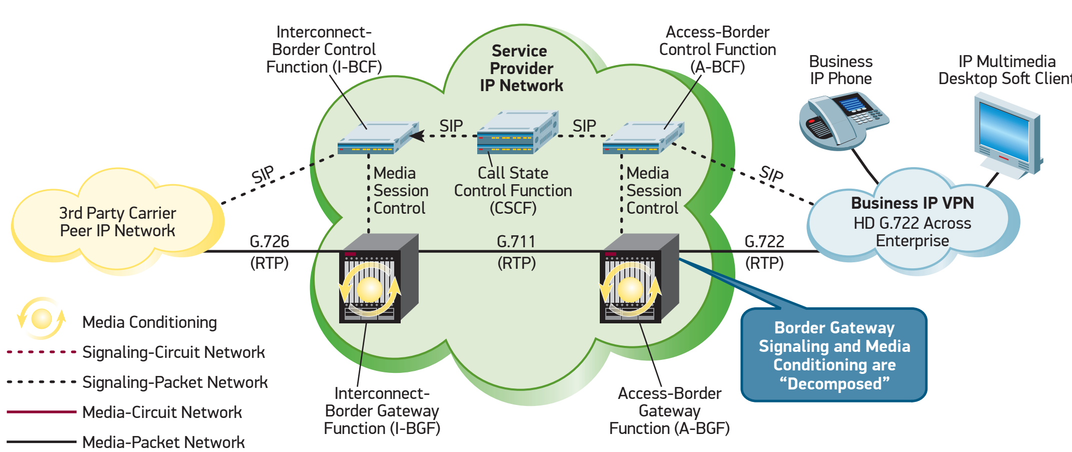

Figure 5. *Media Conditioning at Edge using 3rd Party Call Control (3PCC) in Decomposed Architecture*

and border signaling elements, on a platform optimized for bursty signal processing. This allows the network architect increased flexibility in designing a cost-efficient access edge architecture, while providing the signaling network more granular control of media conditioning.

In addition, by decoupling the border signaling requirements from the BGF, the remaining media plane processing requirements in the BGF increasingly align with the product specifications of an MRF. Therefore, MRF products, which are already optimized for fast, efficient IP packet processing, will see an increasing role for applying media conditioning in access edge applications as well.

In summary, Table 1 examines various criterions when deploying media conditioning at the service provider network edge or in the core network.

Media Conditioning Placement

Criteria	Preferred Transcoding Location	Comments
Reduce edge element complexity	Core	Edge elements are simpler and more cost efficient, lowering CapEx and OpEx
Increase multi-service media processing capabilities	Core	Centralization of media conditioning resources with multi-service media processing creates application synergies and economies of scale
Decrease transcoding delay	Core	Fewer transcoding steps minimizes delay and improves voice quality
Reduce network backhaul bandwidth	Edge	Many calls never need media processing services of an MRF, therefore forcing all calls into the core for media conditioning needlessly increases backhaul bandwidth
Enable codec normalization	Edge	Transcoding is performed before streams enter the core, providing efficient codec reduction
Eliminate application servers for media conditioning call setup	Edge	Controlling elements are not required for inline edge transcoding applications
Increase edge element scalability	Edge	3PCC is optional for edge-based deployments

Table 1. *Comparison of Edge and Core Deployment Scenarios*

Radisys Solutions for Transcoding and Media Conditioning

Radisys is a global leader in carrier-class IP media processing with its Multimedia Resource Function (MRF) product family. Radisys MRFs already support extensive IP media processing features for a broad range of IP service applications including multimedia conferencing, ringback tones, unified communications, IP contact centers and Interactive Voice and Video Response (IVVR applications). In addition, Radisys MRFs support a comprehensive media conditioning solution with both automatic multimedia transcoding and voice quality enhancements as integrated components of existing IP media processing features.

The carrier-class MPX-12000 Broadband MRF was specifically designed to deliver industry-leading performance and capacities for audio and video transcoding, transrating, and media conditioning requirements between IMS, WebRTC, and Mobile Video endpoints.

For smaller deployments, Radisys also offers the Software MRF, optimized for real-time performance on variety of COTS server platforms, or virtualized media processing solutions for cloud network deployments.

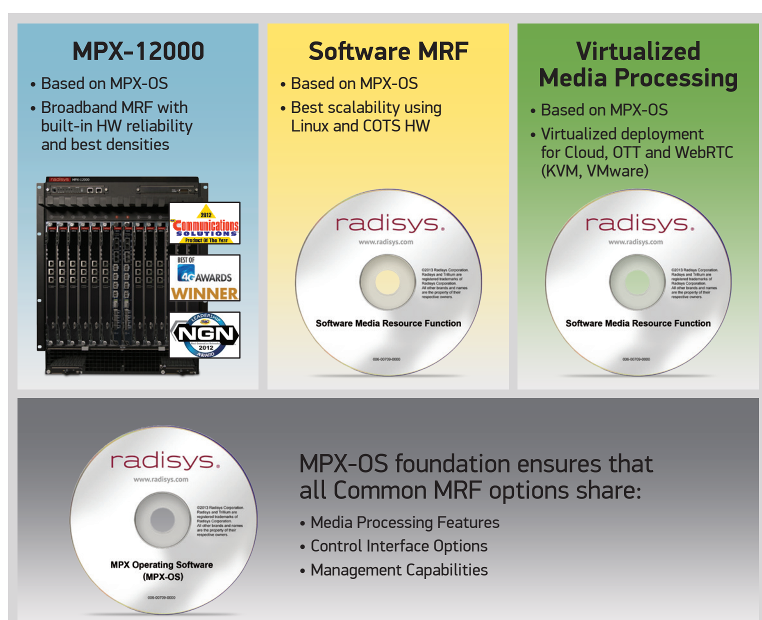

Figure 6. *Radisys MRF Product Family*

Underlying all three media processing platform options—hardware, software, or virtualized media processing—is Radisys MPX Operating Software. MPX-OS provides a common real-time multimedia processing foundation for IMS MRF, OTT, cloud, and WebRTC communication services, transcoding, and media conditioning requirements.

The MPX-OS foundation ensures that all Radisys MRF products support similar integration capabilities. For 3rd Party Call Control (3PCC) architectures, MPX-OS supports a feature-rich SIP with Media Server Markup Language (MSML–RFC 5707) interface. Alternatively, MPX-OS also supports RFC 4117 for two-party call transcoding applications. Alternatively, MPX-OS can also support "inline" transcoding and media conditioning through support of a Back-to-Back User Agent (B2BUA). MPX-OS commonality allows an operator to start small with a software media processing solution, yet scale to virtualized cloud or MPX-12000 platforms with minimized changes in the media conditioning control or network architecture.

Exceptional Media Processing

For over thirty years, transcoding has been a critical technology for bridging TDM- and IP-based networks. Moving forward, it is playing a vital role in delivering new services, while optimizing network bandwidth in modern IP telecommunication networks. At the same time, the scope of media stream processing is reaching well beyond transcoding, and today it includes other media conditioning functions such as voice quality enhancement, video transrating and other bearer channel adaptation features.

Helping carriers deliver cost-effective and higher quality services, Radisys offers leading-edge media processing hardware and platforms for multimedia transcoding and media conditioning applications. With over twenty five years of experience in the telecom industry, Radisys is at the forefront of IP media processing and network architecture design.

Corporate Headquarters
5435 NE Dawson Creek Drive
Hillsboro, OR 97124 USA
503-615-1100 | Fax 503-615-1121
Toll-Free: 800-950-0044
www.radisys.com | info@radisys.com