

White Paper

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and
Network Topologies

Dialogic White Papers

This white paper is brought to you as part of a continuing series on important communications topics by Dialogic.

Dialogic Corporation is a leading provider of world-class technologies based on open standards that enable innovative mobile, video, IP, and TDM solutions for Network Service Providers and Enterprise Communication Networks. Dialogic's customers and partners rely on its leading-edge, flexible components to rapidly deploy value-added solutions around the world.

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

White Paper

Executive Summary

The market place is demanding video content at a rapid rate. For mobile handset users, the challenge of having video on their handsets has been hampered by the handsets' limited bandwidth access and processing power to convert between the many formats available. This white paper discusses how the IP Multimedia Subsystem (IMS) network can be used to convert video content so that it is accessible to any user, and how network-based processing techniques could be applied to a wide variety of applications.

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

White Paper

Table of Contents

Introduction	2
Video Processing Challenges	2
Processing Requirements	3
Processing Architecture.	6
Network Applications.	7
Summary	8
References	9
Acronyms	9

Addressing Video Processing Challenges with the IP Multimedia Subsystem

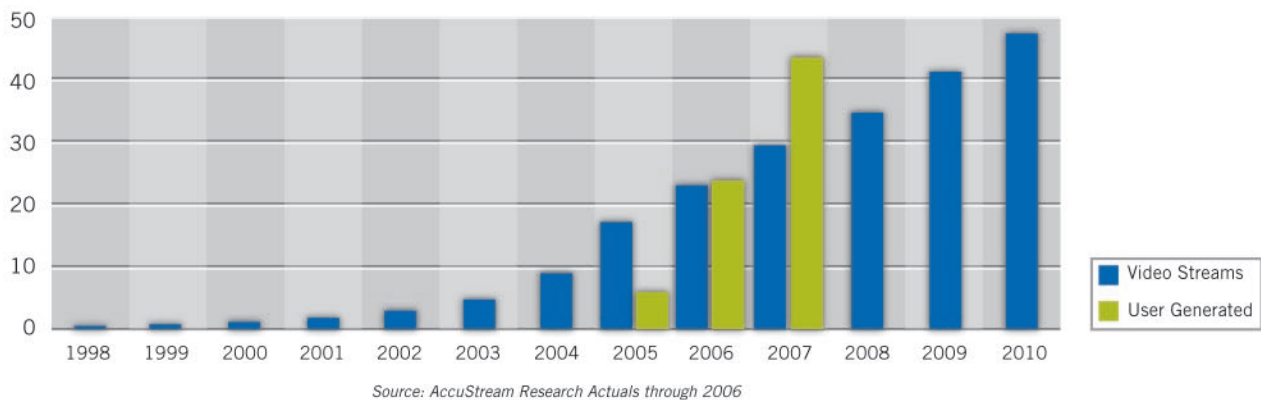
Performance, Architecture, and Network Topologies

White Paper

Introduction

Video content is becoming increasingly prevalent in the marketplace. Figure 1 illustrates the growth in YouTube content access alone. As the content increases, the means of accessing it will most likely also grow. Yet, accessing content could be problematic for many users who employ devices that either do not have the bandwidth to access such content or the processing power to convert between the many formats available.

This white paper discusses the network methodologies, specifically the IP Multimedia Subsystem (IMS) network, that can be used to convert video content so that it is accessible to any user. For this discussion, it is assumed that the user will be using a mobile handset, because it could be argued that the average PC (even ultra-mobile PC) has both the access bandwidth and processing power to compensate for any video format. However, by looking into the future at potential video enhancement applications, the discussion will reveal how the network-based processing techniques could be applied to a wide variety of applications.



"YouTube is driving more capacity today than the entire Internet in 2002..."
Larry Irving, former US Asst. Secretary of Commerce - 2007

Figure 1. Video Content Explosion

Video Processing Challenges

Video applications are typically categorized as "real time" or "streaming." A typical example of a streaming application would be downloading a video clip from a network (for example, YouTube). The most probable real-time application would be video conferencing. Regardless of the application type, moving video in a network requires some sort of compression.

Video compression is used in networks to reduce transport time and storage costs. When discussing mobile or wireless access devices, the transport bandwidth and latency requirements are critical. Mobile access devices rely heavily on video compression in order to view streaming video because of their relatively low download bandwidths. These same devices have an even lower upload bandwidth, making video conferencing in the mobile handset extremely difficult.

At the same time, consumer demand for higher quality video is being driven by High Definition (HD) TV and the extreme graphics quality of video game platforms. In observing the evolution of ultra mobile PCs, video screen size and resolution are among the most important parameters used in measuring the performance of the device.

The challenge is in obtaining the highest quality video in the lowest amount of network bandwidth. Popular video compression techniques such as Flash, MPEG-2, and H.263 have been surpassed by the newer H.264 standard, which can have a two-to-three times advantage in compression.

Addressing Video Processing Challenges with the IP Multimedia Subsystem

White Paper

Performance, Architecture, and Network Topologies

In order to take advantage of this new compression scheme (and others that will most likely follow), or to adapt it to a particular device's screen capabilities, the content is transcoded and transrated. Transcoding is a fundamental network-based video-processing operation in which one video format is decompressed to a raw stream, which is then compressed into another format. Transrating is a key requirement for the connectivity of broadband and mobile networks. These transcoding and transrating operations are optimally done in the network, minimizing the bandwidth required to transmit the content and alleviating the handset of processor-intensive adaptation operations.

Going beyond simple adaptation to the handset, video conferencing requires a video processing operation that combines video streams or images while the video is uncompressed. Applications such as conferencing and image/text overlay use this stream-combination technique. Therefore, compressed video content is collected by a conferencing server and then combined in the uncompressed domain before being re-compressed and forwarded to participants. While techniques exist that can process small conferences at the endpoints, the bandwidth and processing power needed eliminate any mobile device cooperation.

Because an end-user device can only display video in a format it understands, the content provider can either store the same video in many different formats or provide a mechanism by which the content can be adjusted to meet the requirements of the end-user device, whether it is a laptop, ultra-mobile PC, or cell phone. One mechanism has the network extract a default format from the user-account profile and convert the current format so that it is compatible with the end-user device, if necessary. This operation is usually conducted by a type of video gateway. In the IP Multimedia Subsystem (IMS) [Poikselka] network (see Figure 2), which defines the future 3G wireless infrastructure, video processing would most likely be conducted in the Media Resource Function Processor (MRFP).

Conferencing applications have a different set of challenges because communication occurs in real time and involves multiple video streams, which may have incompatible compression formats. Latencies caused by packets moving across the network and by processing must all fit within the acceptable limitations of a real-time conversation – a stringent

requirement considering that packet-movement latencies often consume most (and sometimes all) of the acceptable latency budget. Packet-movement latencies that occur when content is streamed from a storage medium can often be masked if sufficient buffering is provided at the endpoint. Unfortunately, long processing delays caused by format conversion will result in audio-synchronization issues, and long start times will annoy conference participants.

Depending on the application, end users may view a conference in different layouts; for example, view only the loudest talker in one frame or view each individual in an array of frames. Viewing multiple frames requires that the participant video streams be processed and synchronized with the audio content, which itself is being decompressed, combined, compressed, and packetized in a separate stream. But even a simple “loudest talker” conference that switches video feeds (as opposed to combining them) requires analysis and synchronization of the audio stream. Current audio technologies involve a wideband coder (providing more bandwidth for better fidelity) with spatial resolution that drives a high-quality speaker system to provide directional audio that mimics the natural environment in a face-to-face meeting.

Image and text overlay applications have a variety of uses, such as creating special effects, assisting the deaf in interpreting content, and helping ensure Digital Rights Management. Such applications require that video content be converted to its uncompressed (raw) format, and then combined with additional content (such as a text crawl at the bottom of the screen) before being recompressed and packetized. Unlike with conferencing, such content is often streamed from a storage endpoint, and latency is generally not an issue.

Processing Requirements

Video processing operations require that the content be uncompressed and then re-compressed. This process alone consumes a considerable amount of processing power. Any additional processing, such as combining conference streams, synchronizing audio, image overlays, etc., requires even more processing power. The processing power needed for transcoding or combining an image is a function of the number of bits processed, which depends on the overall image quality required. Compression and decompression of an HD stream in H.264 format will involve as much as 200 times more processing power than an H.263 QCIF (Quarter

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

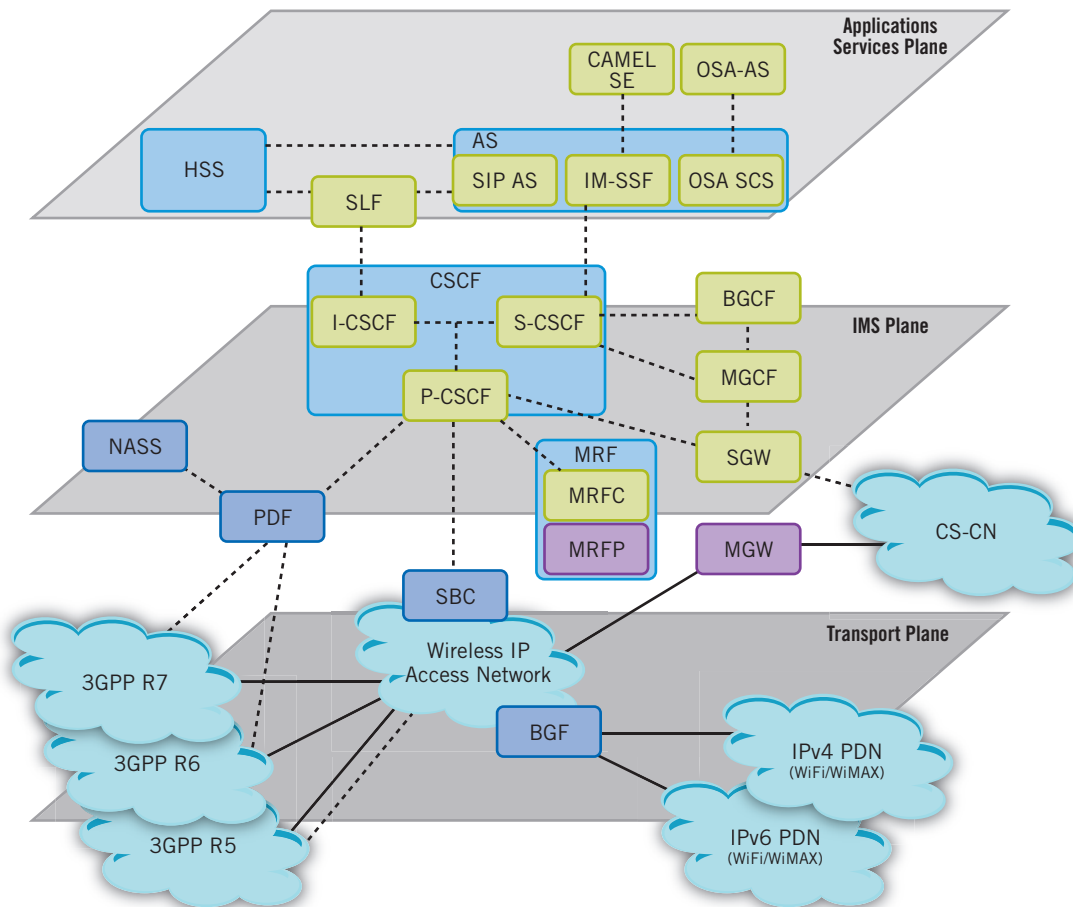


Figure 2. IMS Network

Common Intermediate Format) stream. Also note that an encode-compression scheme for video can be two-to-ten times as complex as the decode-decompression scheme, depending on the coders used. In fact, when transcoding between two formats (such as H.263 and MPEG-4), the asymmetry in processing power can vary an additional five-to-ten times between encode and decode processes.

The encode type chosen will determine video quality, but the network and end user's device need to be able to process the selected video encode type; thus, often what the users get is not what they necessarily want, but what their networks and devices can actually receive and process. Table 1 shows some typical bandwidths and formats.

Encode Type	H.263 QCIF	H.264 CIF	HD 720p	HD 1080p
Resolution	175x144 @ 15 fps	352x288 @ 30 fps	1280x720 @ 60 fps	1920x1080 @ 60 fps
Bit Rate	64 kbps	768 kbps	20 Mbps	50 Mbps

Table 1. Approximate Data Rates for Popular Video Formats [ITU-T]

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

The solution provider who decides to invest in video transcoding can opt to purchase a solution that will scale easily from an initial service introduction of a few hundred simultaneous channels to a mature deployment with tens of thousands of channels. In addition, compression-format standards will most likely change over the lifetime of a service, requiring a flexible transcoding solution, as is happening, for example, with the current launch of the H.265 investigation by the ITU-T Video Coding Experts Group (VCEG).

The design of a video processing gateway or server takes into account the following:

- Scaling over a factor of 100 in channel density
- Scaling over a factor of 200 in processing-density-per-channel
- Field upgradeable to accommodate new algorithms
- Cost effectiveness

Based on these design considerations, four notable design principles can be proposed for video processing architectures:

1. Scalability
2. Versatility
3. Density
4. Usability

Scalability

Any media application must scale in terms of the incremental resolution of service channels. In other words, it is desirable that it be able to provision with a small number of channels at service introduction and be able to add small increments in the event the service grows slowly, or be able to provision large quantities of channels quickly in the event the service becomes widely popular.

In addition, a video processing server must scale in two dimensions: the number of channels and the amount of processing power per channel. It is important to note that the encoding and decoding processing requirements are asymmetrical, and not all applications call for symmetry when encoding and decoding. In most cases, however, it is useful to plan for a worst-case scenario.

In a conferencing solution, the amount of processing power for a single conference is a function of the number of participants. Large video conferences at low resolution can involve two or three processors; moderate conferences at HD resolution can require ten or more processors. Bridging a conference between processors consumes processing resources itself, so it is desirable to consolidate the largest number of participants on one processor as possible. Using multi-casting techniques can also eliminate the need to employ video compression to each leg of a conference; however, this assumes compatible endpoints. If the call is being made from a wide variety of endpoints, additional transrating or transcoding may be required to ensure an acceptable level of quality.

Versatility

New variations on old algorithms, completely new algorithms, and varying demands on algorithmic instances (also known as “algorithm volatility”) require a versatile platform. Such a platform can help a manufacturer to maintain a market-leadership position by quickly introducing new algorithms or new features that differentiate the product line. But versatility stands in stark contrast to the second requirement: heavy-duty processing power. Designs that attain longevity achieve it through versatility, and thus require some level of general-purpose functionality. But because considerable processing power is also needed along with versatility, it is useful for designers to consider a mixture of processor types when choosing processors.

Density

As with any telephony application, the network provider will want the highest density per server possible within the range of an acceptable failure group. Failure of a video call is much more noticeable than failure of an audio call because people are more sensitive to drop-outs in video than the audible clicking sound that may be heard during a fail-over of an audio call. Audio failure groups can be as high as an STM-1, which, in audio terms, can be as large as 2000 phone calls; that is, fine for QCIF @ 15 fps, but would only carry about 170 CIF @ 30 fps channels. It may be assumed that larger backhaul capabilities in the mobile network of the future would allow the network provider to increase this channel density. Going as high as OC192 could provide access to more than 480 720p HD channels at some point in the future.

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

White Paper

Usability

From a standpoint of usability, it is beneficial to look at how applications can be written to the server (MRFP), and how the resources can be provisioned and maintained as well as upgraded in the field.

Most of these usability issues can be addressed with existing XML-based languages that address media, call control, and maintenance needs. Challenges still exist with the media-related interfaces lagging behind the application needs; however, these can be solved with proprietary extensions, if desired.

VXML and MscML suffice for many media applications today, but the work going on in the IETF Media Control Working Group [Boulton] will hopefully consolidate the APIs and expand upon their application set. For the most part, SOAP interfaces will accommodate the maintenance and provisioning needs of the equipment, allowing for updates, servicing, and monitoring.

Processing Architecture

As with any communications application, a considerable amount of control needs to be processed in addition to the pure media manipulation. The application interfaces and OAM&P functions are well served by a general-purpose processing architecture. What is not often appreciated, however, is that the same general-purpose processor used for control purposes can serve as a highly cost-effective and very scalable media processor. In fact, for many applications, especially in lower end video processing and for 1 to 300 video channels, the standard high-volume server model created by the x86 or SPARC general-purpose processor ecosystem is the most cost-effective processing solution available. Where the general-purpose processor is weak is in overall performance density.

A general-purpose processor is normally not competitive with a DSP in terms of processing efficiency. As a result, general-purpose, processor-based architectures cannot produce the same number of channels for a given algorithm as a DSP can in the same space and with the same electrical power. Since the maximum number of channels determines the overall dynamic range of a product offering, for an industry-leading design to achieve the lowest cost, size and power, it must attain the best density possible.

Several disadvantages exist in pursuing DSPs however. DSPs lack the overall versatility of general-purpose CPUs. Some traditional DSP manufacturers have made some strides in giving DSPs more design flexibility, but this is usually restrictive and not cost-effective. What is required is a different approach to algorithmic offload of the general purpose CPU. This offload processor, which is broadly labeled here as “an accelerator”, must retain much of the versatility required to interoperate efficiently with the general-purpose CPU (such as dealing with dynamic requests and algorithmic upgrades), but still provide extremely high performance efficiency.

Using a balance of general-purpose processors and tightly coupled accelerators (for use in very high-density applications) provides a more desirable approach. A wide variety of accelerators, shown in Figure 3, would be appropriate. Notable aspects of this kind of design strategy are the overlying software structure, algorithmic partitioning, and overall communications fabric.

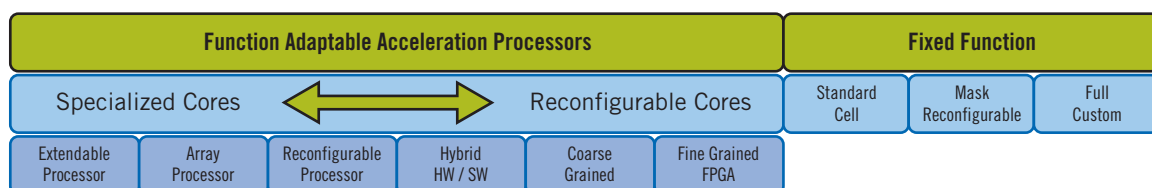


Figure 3. Processing Arrays for Video Acceleration

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

The overlying software structure runs on a general-purpose CPU, and the structure must abstract the type of algorithm used and the acceleration device that is processing the algorithm. Such an approach can allow new algorithms and acceleration technologies to be introduced quickly without impacting the application. The application itself can run on the general-purpose CPU or on a remote server via remote media and call control protocols (SIP, MSML, VXML, H.248, etc.). Local control, management, and data routing are handled on the general-purpose CPU, and this processor can also be used for establishing new algorithms. The goal of algorithmic partitioning is to assure that the older, more stable algorithms or the most processor-intensive operations run on accelerators.

Note that the entire algorithm does not need to run on the accelerator. Generally speaking, only 10 to 30 percent of the tasks in a media compression algorithm consume about 70 percent of the processing power required for the entire algorithm. These media-compression tasks require offload acceleration, but this offloading requires tight coupling between the general-purpose CPU and the accelerator.

Network Applications

After having established the architecture for a video gateway or MRFP in terms of an accelerated server, the next introduction is a cost-effective and scalable architecture that can serve the desired processing performance for any type of video application. The final challenge is to provide a network-scalable solution with the appropriate fail-over capabilities to maintain service availability.

Figure 4 shows a subset approximation of an IMS network in terms of the video streaming functionality. The control signaling via the wireless area network to the P-CSCF establishes the user handset's session with the service. The MRF Controller (MRFC) establishes the call session with the video gateway resources within the MRFP "cluster" — so referred to because it could entail multiple servers stacked in a redundant fashion to provide the necessary channel density for what could be a wide variety of services.

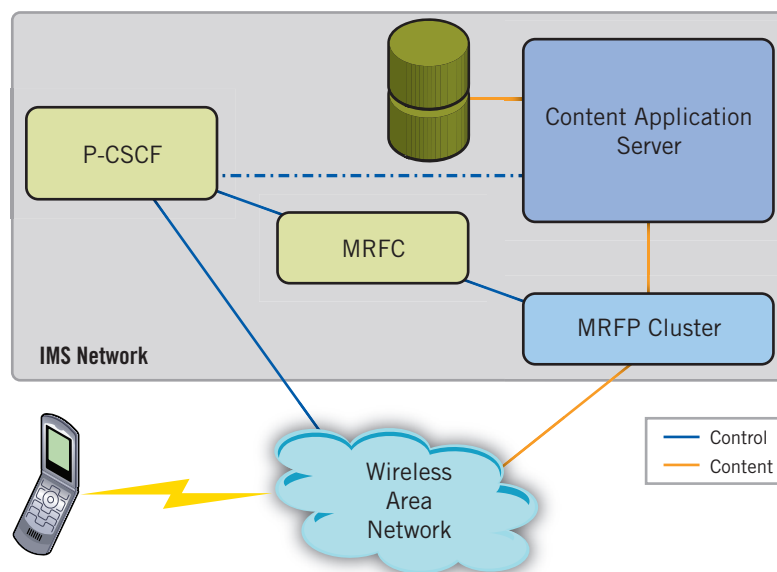


Figure 4. Video Streaming Application in an IMS Network

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

White Paper

These services could consist of the video streaming “play” resources as depicted in the Figure 4 example, or resources that do compression, decompression, conferencing image overlay, etc. These resources could be clustered in a different fashion depending on service requirements and billing dependencies, or they could be shared across applications. An illustration of this service cluster is shown in Figure 5.

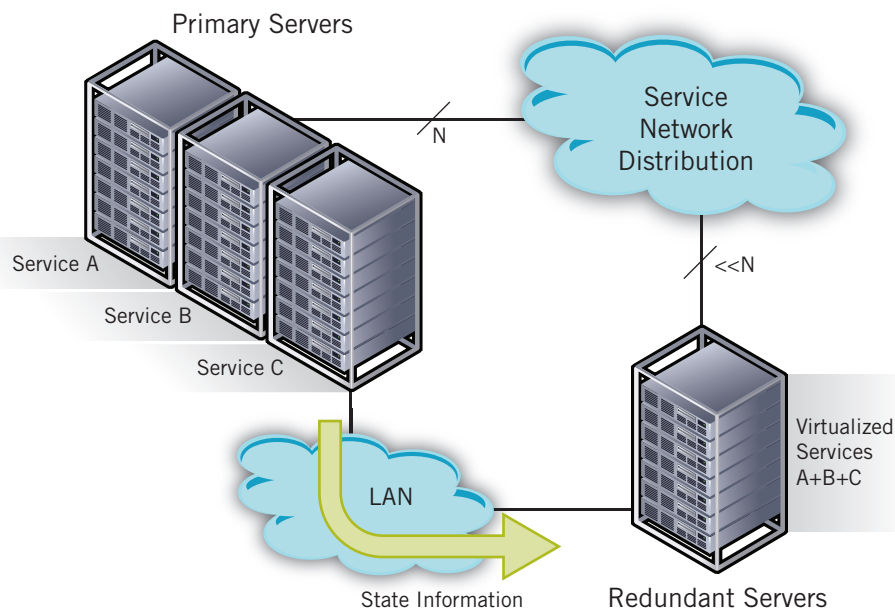


Figure 5. Scalable Media Resource Function Processor (MRFP) Architecture

Notice that the servers themselves can be made redundant in a cost-effective N+1 fashion by allocating additional servers configured as virtualized processes, which could take on the functionality of one of multiple services as needed. This is an additional advantage of using the general-purpose processor architecture. Such processors can be configured with multiple virtual partitions, each poised to take over the functionality of another processor in the local cluster. An accelerator approach is also a preferable option for performance efficiency for high-end applications because the correct acceleration architecture can cooperate in this virtualized environment.

Summary

This white paper briefly evaluated the many types of video services users may encounter in the near future and discussed the challenges of providing some of these services. Also discussed are the architectural and performance requirements of the equipment that would be required to provide video services. Briefly touched upon are the processing architectures that would most suitably meet these requirements. And finally examined are the network topologies that video servers would be deployed in an IMS network.

It was noted that performance scaling, as well as performance density, tempered with an acceptable cost and provisioning model, were critical in any such application. A conclusion could be that the general-purpose CPU model (x86 or SPARC) worked most suitably for most of these requirements and could benefit from a performance boost at the high end with a tightly coupled acceleration processor.

The servers in the network could themselves be made scalable to limit the size of the failure group of any one server while still scaling to the requirements of the service. The service itself can be made fault tolerant in a cost-effective manner by using the virtualization properties of the standard server models enabled by the use of general-purpose CPUs.

Addressing Video Processing Challenges with the IP Multimedia Subsystem

Performance, Architecture, and Network Topologies

White Paper

References

[Boulton] C. Boulton, T. Melanchuk, S. McGlashan; IETF Network Working Group Media Control Channel Framework “draft-ietf-mediactrl-sip-control-framework-01,” February 2008.

[ITU-T] ITU-T Recommendation H.264: Advanced video coding for generic audiovisual services, Approved 03-2005.

[Poikselka] Poikselka, Mayer, Khartabil, Niemi, “The IMS IP Multimedia Concepts and Services,” Second Edition, John Wiley and Sons, 2006.

Acronyms

AS	Application Server
BGCF	Breakout Gateway Control Function
BGF	Border Gateway Function
CAMEL	Customized Applications for Mobile Network Enhanced Logic
CSCF	Call Session Control Function
CS-CN	Circuit Switched Core Network
HSS	Home Subscriber Server
IMS	IP Multimedia Subsystem
IM-SSF	IP Multimedia Service Switching Function
I-CSCF	Interrogating Call Session Control Function
IP-CN	Internet Protocol Core Network
MGW	Media Gate Way
MRF	Media Resource Function
MRFC	Media Resource Function Controller
MRFP	Media Resource Function Processor
NAS	Network Access Server
OSA	Open Services Architecture
PDF	Policy Decision Function
P-CSCF	Proxy Call Session Control Function
SBC	Session Border Controller
S-CSCF	Serving Call Session Control Function

www.dialogic.com

Dialogic Corporation

9800 Cavendish Blvd., 5th floor
Montreal, Quebec
CANADA H4M 2V9

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH PRODUCTS OF DIALOGIC CORPORATION OR ITS SUBSIDIARIES ("DIALOGIC"). NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN A SIGNED AGREEMENT BETWEEN YOU AND DIALOGIC, DIALOGIC ASSUMES NO LIABILITY WHATSOEVER, AND DIALOGIC DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF DIALOGIC PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHT OF A THIRD PARTY.

Dialogic products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Dialogic may make changes to specifications, product descriptions, and plans at any time, without notice.

Dialogic is a registered trademark of Dialogic Corporation. Dialogic's trademarks may be used publicly only with permission from Dialogic. Such permission may only be granted by Dialogic's legal department at 9800 Cavendish Blvd., 5th Floor, Montreal, Quebec, Canada H4M 2V9. Any authorized use of Dialogic's trademarks will be subject to full respect of the trademark guidelines published by Dialogic from time to time and any use of Dialogic's trademarks requires proper acknowledgement.

The names of actual companies and products mentioned herein are the trademarks of their respective owners. Dialogic encourages all users of its products to procure all necessary intellectual property licenses required to implement their concepts or applications, which licenses may vary from country to country.