Deploying Video Services in Next-Generation Networks

Josh Adelson

Director of Market Development

cantata
TECHNOLOGY

table of contents

cantata TECHNOLOGY

abstract

Since AT&T's introduction of the Picturephone at the New York World's Fair in 1964, video services over telephone networks have had an aura of futurism. Now, video is really here. Several factors have converged to make this happen, including 3G and fixed broadband service deployment; the proliferation of video-capable mobile handsets; and the emergence of standards for networked multimedia.

For network operators and the equipment providers who serve them, video represents a huge opportunity, potentially a "killer application" for 3G Wireless. But it also poses a risk in terms of selecting the right technology approach. Operators need a technology strategy that will support rapid deployments as well as ongoing flexibility to meet changing application and provisioning requirements.

This paper provides an overview of the market for video services and then describes a standards-based architecture for meeting the market's requirements. It presents the case study of a European mobile operator who has successfully employed this approach to deliver services today. It finally discusses the role of Cantata Technology and its partners in delivering next-generation video solutions.

executive summary

Video promises to be the means for a whole wave of network services through which operators will create market differentiation, new revenue streams, and subscriber loyalty. Deploying successful video applications poses a number of challenges, not least of which is the inability to predict which services and features will be popular with users.

An architectural approach and corresponding standards now exist for the deployment of video services. The architecture is based on next-generation packet voice networks, and notably separates application logic from media processing. These capabilities are typically delivered in the form of independent but interoperable Application Servers and Media Servers. One major benefit of this separation is the ability to select the best components from vendors who specialize in a particular domain. Another is the ability to share resources among different applications and thus gain considerable economies of scale. Key standards that enable this interoperability include SIP, for inter-device connection, and VoiceXML for application control.

This paper explains the main attributes of the next-generation video architecture, and then provides the real-life case study of TMN, a mobile operator who is delivering video messaging and content today.

market overview

growth drivers

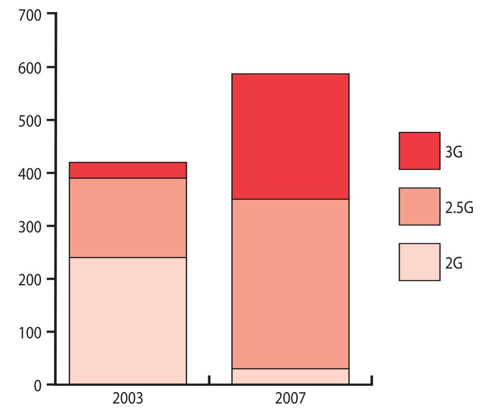
Next-generation video services are poised for tremendous growth. The ARC Group, a market research firm specializing in mobile telecommunications, estimates that by 2008, multimedia service revenues will, for mobile operators alone, exceed $16 billion.[1] Cantata believes that mobile networks present the largest single video service opportunity, but that broadband cable and DSL operators will also have significant presence in these new applications.

Several fundamental trends are fueling this growth. First, operators are continually looking for ways to differentiate their services and increase their annual revenue per user (ARPU). Standard voice service, voice messaging and text messaging, while important, present little opportunity for premium pricing or differentiation. Video is a natural evolution for value-added services. More than a single application or service, video is a foundation for a whole range of different applications and services, as we shall learn later in this paper.

[1] *MMS: Current Status and Future Expectations, The ARC Group, September 2003.*

There are also important technical drivers for video. In the latest generation of mobile handsets, cameras are common. This capability, sold primarily for still photography, means the phone can be potentially a source and playback device for store-and-forward motion video. Camera phones also generate consumer awareness of phones as visual communication devices. These devices have an impact on consumer acceptance as well. When you consider the extent to which younger-generation phone users have come to view their mobile phones as entertainment devices and personality statements, the extension of their behavior to video communication is hardly a leap. The table below illustrates the growth of 2.5G and 3G handsets that is expected to take place over the next three years.[2]

Table 1: 2.5 and 3G Phone Penetration, 2003–2007
(Millions of Handsets)



Technology advances in handsets are being mirrored by equally significant changes in the networks. All of the competing technologies for new networks are capable of supporting video. 2.5G wireless networks provide on the order of 56kbps bandwidth—hardly broadband but sufficient for small-screen video streaming or download of short video clips. 3G wireless networks provide sufficient bandwidth for most video applications. On the landline side, broadband access continues to grow in penetration. While most broadband has been deployed for Internet data access, once deployed it can be used for high-speed telephony including video as well.

Beyond their considerable bandwidth, the architectures of these new networks facilitate the growth of innovative services like video. Most are IP based, supporting standard communication and application protocols like SIP and VoiceXML. This foundation of networking and application standards makes the task of deploying new services far more achievable. It also facilitates interoperability between enterprise and residential IP systems.

a taxonomy for video applications

As mentioned above, "video" is not just one application or service. People sometimes refer to video without any additional definition, which leads to misunderstandings. To give an idea of the variety of applications and services, here are just a few examples:

- **Video Conferencing:** Real-time conversation with sound and moving images between two or more participants.

[2] *Gartner Dataquest, February 2004.*

- **Video Messaging:** The ability to record a brief personal message or video "mail", and send it to another person for their retrieval.
- **Video Greeting Cards:** A variation on video messaging. The video content might be personally recorded, or selected from a library of pre-recorded clips. Greetings may include pre-recorded templates, music and text overlay.
- **Information, Entertainment:** To view streaming content including news, sports, music video clips and movie trailers.
- **Personalized Advertising and Dating Services:** The ability to add video content to "online" advertising. For example, a walk-around view of an apartment, or a personal message for a dating service.
- **Emergency Response:** The ability for public safety personnel or ordinary citizens to provide real-time video capture of a scene, thus assisting in response or legal prosecution.

These applications all have video in common, but they differ considerably in terms of user requirements and back-end infrastructure needed to support them. To better focus on the underlying requirements, we propose a taxonomy that contains three basic categories of video applications:

- **Video Conferencing:** Conferencing is characterized by the real-time, bi-directional nature of the communication. All endpoints must be able to simultaneously create and display video content.
- **Video Messaging:** Messaging is characterized by the need for recording or creation, as well as playback, at the telephony endpoints, but not at the same time. Because of its store-and-forward nature, messaging also requires mechanisms within the network for storage and retrieval. In this context, messaging also includes applications such as greeting cards or dating services.
- **Video Streaming:** Streaming is characterized by its fully one-way nature. Streaming may be pre-recorded, as in the case of film clips, or live, as in the case of television broadcasts or popular "web cams." There is no need for end-user content creation. In fact, creation for streaming is often performed by professionals using sophisticated professional equipment and methods for both pre- and post-processing. While streaming has fewer requirements for end-user content creation, it may have greater requirements for end-user viewing, for example fast-forward and rewind controls.

Throughout the remainder of this paper, we will distinguish as necessary between these three forms of video applications.

Table 2: Video Applications Taxonomy

Category	Examples
Video Conferencing	Conversational 2-Way Video, Multiparty Video Conferencing Video-Enhanced Chat
Video Messaging	Video Mail, Video Greeting Cards, Video Classified Advertising
Video Streaming	Sports Highlights, Weather Reports, Live Television Broadcasts

challenges of video

As video is so valuable, and many of its underpinnings already exist, one might be tempted to ask why it is not more common. In fact implementers of video services still face a number of hurdles. These can be grouped into three major categories: diverse network types, diverse client types, and demand uncertainty.

Diverse network types refers to the different classes of telephone networks: traditional wireline, broadband (cable and DSL), and wireless. Each of these basic network types has its own unique infrastructure and protocols. Broadly speaking, the newer classes of networks support video. However, these newer networks are not yet ubiquitously deployed. This may not be a problem, depending on what one is trying to achieve. For example if a mobile operator wants to offer news highlights to its subscribers, then the service only needs to support a single network. If however the operator wants to offer a video mail service, a constraint of sender and receiver being on the same network will restrict the adoption of the service.

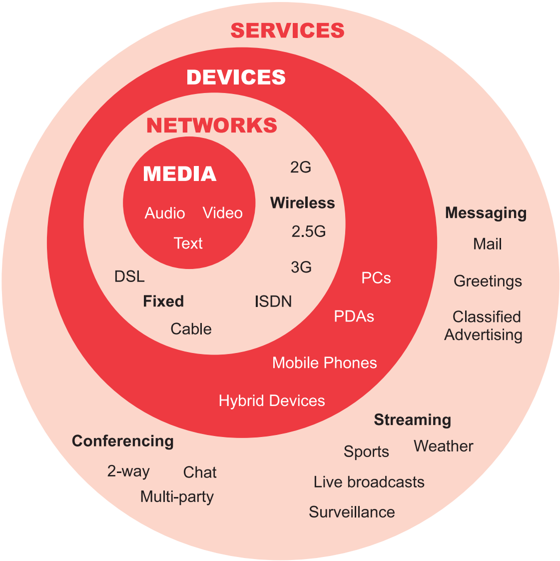
Diverse client types refers to the different types of devices from which users will send or receive video content. While many mobile phones now contain cameras, they are not necessarily capable of capturing motion video. They also may not have sufficient memory or software capabilities to buffer an incoming stream and synchronize the audio and video tracks. Most landline phones have no video capabilities at all, however landline callers may use soft-phone clients running on PCs for video.

Applications differ in the extent to which they rely on commonality among client endpoints. For example, video conferencing is most valuable when the number of possible participants is greatest. In other words, my video-conferencing phone is only useful if others also have video-conferencing phones. This is sometimes known as the "network effect." By contrast, the value to a subscriber of one-way video streaming is not affected by number who can take advantage of it. Messaging most likely falls somewhere in-between. Other things being equal, we expect the penetration to begin with those applications that are least sensitive to the network effect.

As with diverse networks, diverse clients do not necessarily preclude implementing a service, but the service application needs to be aware of the different possible clients and it needs to adapt to the respective capabilities of each.

The third category of hurdles, and perhaps the most fundamental one, is demand uncertainty. By this we mean the inability of an operator to know which of the many potential types of services will be popular, and how quickly they will be commercially adopted. One broadband provider with whom Cantata is familiar has listed over one hundred distinct offerings, spanning audio, video and text, which it is considering as add-ons to its basic telephone service. Video telephony is still novel; it requires new behaviors that are notoriously difficult to predict.

Diagram 1: Large Number of Variables Makes for a Challenging Environment



Narrowing the scope of what is attempted mitigates these challenges: for example by offering a specifically defined service within a single network, to a subset of subscribers who meet specified criteria. However, over-specifying the environment narrows the addressable market and risks placing all the investment in an application that may not even take off. Therefore, even after reasonable boundaries are set, the application developer still faces the question: How can I design an application when so many variables are undefined? Answering this question will be the focus of the next section.

architecture for next-generation video

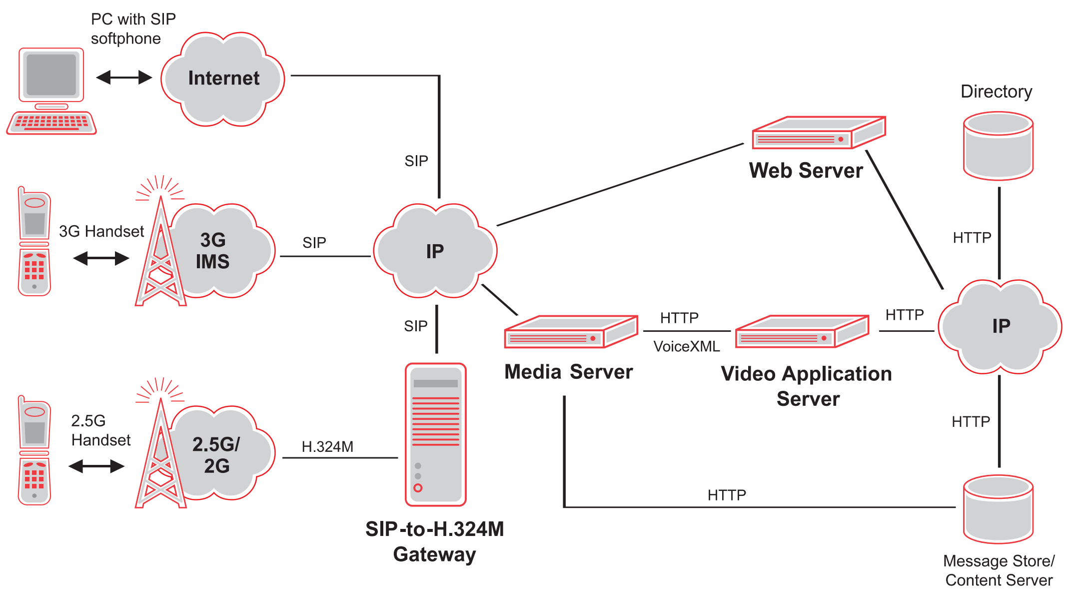
Based on the foregoing discussion, architecture for next-generation video must meet specific requirements. In addition to delivering basic application functionality, the architecture must:

- Fit into next-generation network architectures. This should almost go without saying, but it's important because many delivery systems were designed for legacy networks. Therefore even if they have been subsequently adapted to new protocols, they do not necessarily take advantage of the new generation networks.
- Separate the application from the client device. The diversity and ongoing change in client devices suggests that an application architecture that is tightly coupled to a specific type of end-user device will be difficult to manage. Of course, the application needs some intelligence to adapt to differing capabilities of different devices.
- Be economical at a small scale but allow growth to large. Operators cannot know how new services will be adopted. Therefore they must be able to make initial deployments at relatively low scale but then ramp them up quickly in response to demand.

- Be economical in switching between different applications and services. Again, assuming that operators will not be able to predict which applications will be successful, they cannot afford to dedicate significant asset pools to individual applications. Assets must flexibly support multiple applications.

To meet these requirements, many providers are selecting an IP-based architecture in which SIP is the predominant method for controlling media resources, as illustrated in the diagram below.

Diagram 2: SIP IP-Based Architecture



It should be noted that these basic principles of functional decomposition and use of standard protocols is mirrored in leading industry groups such as the International Packet Communications Consortium (IPCC) and the 3rd Generation Partnership Project (3GPP). Key back-end functional elements of this architecture are:

Application Server: This is the "brains" of the operation. It contains the application logic and business rules, as well as a service creation environment. For example, an action to issue a 384 kbps video stream to a subscriber would be initiated by the Application Server based on its understanding of available content and user privileges among other things, although the actual user information and content reside in other components.

Media Server: This provides the media "muscle", for example playing and recording media streams, transcoding between different formats, or synchronizing audio and video. The Media Server is controlled by the Application Server, through SIP or H.248 and typically a markup language like VoiceXML or MSCML.

Directory: This is the subscriber database, containing user data such as phone numbers and service privileges.

Message Store/Content Server: This is an indexed store of video and audio content. In the case of messaging it would be organized as mailboxes; in the case of streaming it would be organized as clips.

Video Gateway: This extends the video service to non-IP client networks and devices. Typically the gateway translates between SIP, which is used in the IP domain, and H.324M.

In this architecture, the same vendor need not supply the individual elements because the interfaces between them are standards such as SIP, VoiceXML, or Network File System (NFS). This is in stark contrast to traditional so-called "stovepipe" approaches in which elements are consolidated into a single system, from a single vendor.

The benefits of this decomposition are considerable. First, it provides a common infrastructure across multiple applications. By decomposing the different elements of the application, new applications only require changes to the application server. This dramatically reduces the cost of trialing new services. Second, this architecture supports independent scaling of applications and media resources. Since the media server is a networked device responding to SIP commands, it is indifferent to whether the commands are originating from one or more Application Servers. This means that, as demand grows, media servers can be added without the need to duplicate Application Servers or the management overhead implied by such duplication. Likewise, multiple applications can share a common media server resource pool, rather than be dedicated for use with a single application.

Finally, this architecture maximizes vendor choice. Because the different elements are linked via standard protocols, operators can combine the best components for each part of the solution. At the vendor level, this allows each company to concentrate on their particular area of competence, accelerating innovation at each level.

The table below summarizes how this architecture supports the goals that were put forth earlier in this section.

Table 3: Architecture Goals

Architecture Goals	Solution	Benefits
Fit into Next-Generation Architectures	Elements are Networked IP devices, Communicating via SIP and Other Standards	• Use of Existing, Proven Networking Technologies and Protocols • Opportunity to Use Standard Computing Platforms for Many of the Elements
Separate the Application from the Client Device	Application is in the IP domain; Extend to PSTN Via Gateways	• Application Itself not Burdened by Complexities of Multiple Networks • As IP Networks Become More Ubiquitous, Application not Forced to Change
Economical at Small and Large Scale	Independent Application and Media Servers	• At Low/Trial Scale, Single media Server can Support multiple Services. • As Scale Increases, Media Servers can be Added. Shared Resource Pool.
Economical to Switch Between Applications	Independent System Components Joined by Standards.	• New Applications Require Change to Application Server Only.

case study: TMN portugal

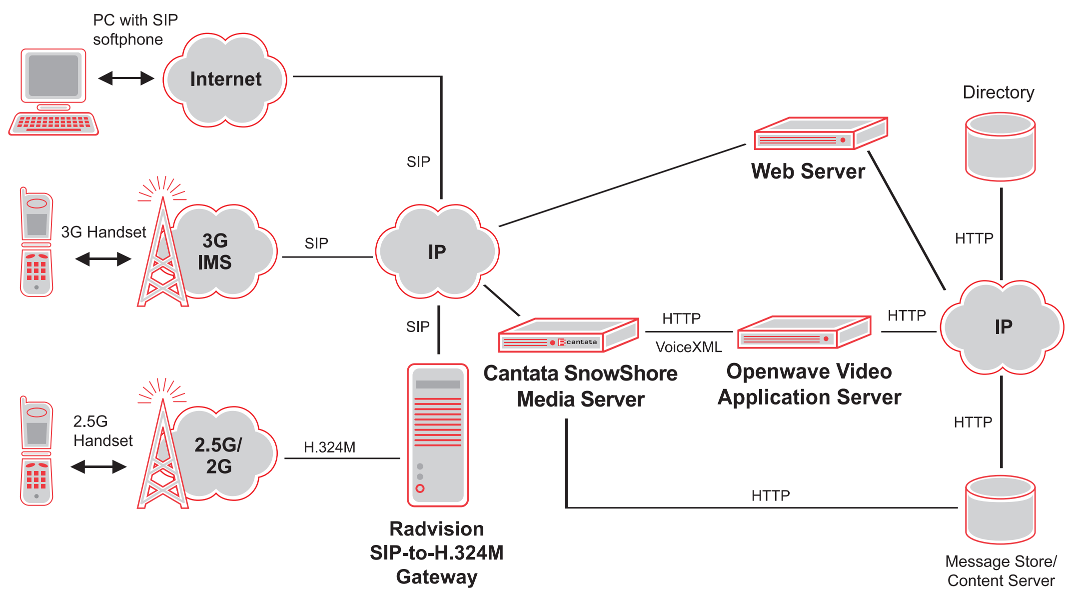
In June 2004, Portugal had the honor of being the host nation for Europe's highest-profile football (soccer) tournament, Euro 2004. TMN, a subsidiary of Portugal Telecom Group, is Portugal's leading mobile telecommunications operator with a 3G/UMTS network and nearly 50% market share. TMN decided to launch its video services in conjunction with the sporting event, to take advantage of the host-nation excitement as well as the popular video content that the games would produce.

TMN's launch strategy, because it was tied to a public event, could not afford any delays in rolling out the new services.

the solution

TMN worked closely with a number of vendors to launch attractive video services quickly and reliably. Openwave was selected as the video application provider for its breadth of video messaging and portal applications, and its flexible service creation environment. TMN was already using Openwave's Mx Messaging Platform for voice-mail, so adding video mail became an evolutionary change, rather than a disruptive one. Cantata Technology was selected as the media server for the messaging service, in part for its leadership in establishing the use of SIP as a means to control media server sessions.[3] Cantata's SnowShore Media Server was the first in the industry to support SIP. It also supports H.263 video compression and adds value beyond the standards with support of features such as video/audio frame synchronization. It has been certified to work with Openwave's product, and the two companies maintain a close working relationship to ensure successful deployments for their common customers. Radvision was selected for its SIP-to-H.324M gateway; this allowed TMN to extend its video services to non-3G users. Finally, Siemens' Information and Communications Group provided project management and integration services.

Diagram 3: TMN's 3G Video Services Architecture



[3] *Cantata's CTO Eric Burger chairs the IETF's speechsc and lemonade working groups. In addition, Cantata has authored or co-authored numerous standards publications in the IETF, IPCC and W3C.*

benefits of the open standards approach

Euro 2004 took place as scheduled, and TMN met its primary goal of launching video services to coincide with the sporting event. Capabilities offered to subscribers include viewing and purchasing video content, as well as sending, receiving, storing and managing voice and video messages. The illustration on the following page shows some of the user interface options that can be offered:

Graphic Courtesy of Openwave Systems, Inc.

Readers can also view TMN's own demonstration of its capabilities at www.tmn.pt/3G/. While full comprehension requires fluency in Portuguese, even non-fluent readers will gain a better understanding of the services and how they are being positioned for the market.

Because the video messaging service was an extension of TMN's voice mail service, TMN was able to share many components, such as the subscriber directory, rather than having to create a new one for the new service and then constantly manage discrepancies between the two.

Clearly, the adherence to open standards by the participating technology providers allowed their individual elements to interoperate successfully with one another. For example, Cantata's and Openwave's use of VoiceXML as the primary application development environment maximizes the available tool-sets and programming skills that can be applied to the project.

conclusion

For operators of next-generation networks, video presents an opportunity to create service differentiation, increase personalization, and build customer loyalty. A standards-based approach to deploying video services now exists, drawing heavily upon recent maturation of voice over IP, SIP and VoiceXML. This approach makes it possible for multiple vendors' equipment to interoperate, and to function within new networks such as fixed broadband and 3G wireless and fixed broadband. Finally, it offers flexibility to experiment with new offerings and adapt to the demands of an evolving marketplace.

Cantata Technology is committed to providing standards-based, advanced media processing capabilities, and partnerships with leading application vendors, in support of reliable and innovative video services.

appendix: cantata technology in next-generation video

The Cantata SnowShore Media Server™ leverages the simplicity and flexibility of SIP and VoiceXML to provide a cost-effective and highly scalable IP media server solution, powering a broad range of voice and video services for next-generation wireline, wireless and broadband networks.

The SnowShore Media Server provides software-based media processing resources that can support a broad range of applications, from basic messaging and multiparty conferencing to network announcements and prepaid services. Because it uses industry-standard server platforms and standards-based IP protocols, the SnowShore Media Server can leverage the ongoing evolution of network architectures and processor technology to deliver a future-proof media processing solution for next-generation IP applications.

Cantata's SnowShore Media Server is supported by a growing list of industry-leading applications, speeding deployment of media-rich, high- value enhanced services for next-generation wireline, wireless and broadband networks.

For more information, please visit www.cantata.com.

cantata
TECHNOLOGY

Corporate Headquarters
410 First Avenue
Needham, MA 02494
USA
Tel: +1 (781) 449-4100
Fax: +1 (781) 449-9009
Email: info@cantata.com

Cantata Technology maintains multiple locations worldwide in North America, Asia and Europe.

www.cantata.com