An MPEG Standard for Rich Media Services

Jean-Claude Dufourd and Olivier Avaro
Streamezzo

Cyril Concolato
ENST

Lightweight Application Scene Representation (Laser) is the Moving Picture Expert Group's solution for delivering rich media services to mobile, resource-constrained devices. Laser provides easy content creation, optimized rich media data delivery, and enhanced rendering on all devices.

A *rich media service* is a dynamic, interactive collection of multimedia data such as audio, video, graphics, and text. Services range from movies enriched with vector graphic overlays and interactivity (possibly enhanced with closed captions) to complex multistep services with fluid interaction and different media types at each step. Demand for these services is rapidly increasing, spurred by the development of the next-generation mobile infrastructure and the generalization of TV content to new environments. However, despite long-lasting deployments and significant investment from various industries, mobile and, more generally, embedded interactive services (mobile Internet and interactive mobile TV, for example) have failed to reach the masses. In addition to conjectural (such as economic) and structural (such as a lack of compelling business models) problems, current technologies (see the related sidebar on p. 63) have failed to provide an effective user experience.

Using rich media services is more challenging in embedded devices than on PCs, where various interfaces are available and homogeneously implemented (such as a mouse or keyboard), and ergonomic concepts have been tested and validated. On the move, when it's not always easy to interact and time is limited, users expect to be one click away from the information they need. In addition, end users are accustomed to quality Web interfaces, so to be successful in the embedded domain, service interfaces must leverage the online experience. Finally, because users pay for these services, they also expect a decent level of quality, efficiency, and readability.

The Moving Picture Experts Group (see the "Related Standards Organizations" sidebar on p. 64 for brief descriptions of MPEG and other groups involved in this effort) has specified MPEG-4 part 20 (formally known as ISO/IEC 14496-20) as the new rich media standard dedicated to the mobile, embedded, and consumer electronics industries. MPEG-4 part 20 defines two binary formats: Lightweight Application Scene Representation (Laser) and Simple Aggregation Format (SAF). Laser enables a fresh and active user experience on constrained networks and devices based on enriched content, including audio, video, text, and graphics. It addresses the requirements of the end-to-end rich media publication chain: ease of content creation, optimized rich media data delivery, and enhanced rendering on all devices. As such, it fulfills the need for an efficient open standard.

Key features and use cases

Four key features distinguish Laser from existing technologies:

- In Laser, graphic animations, audio, video, and text are packaged and streamed together. Unlike existing mobile technologies that are mostly aggregations of various components, Laser's design is based on a single, well-defined, and deterministic component that integrates all of the media (the same design that made Macromedia Flash successful on the Web). This integration ensures the richness and quality of the end-user experience.

- Laser provides full-screen interactivity with all streams. It uses vector graphic technology so users can easily fit content to the screen size. Laser therefore provides optimal content display despite high variations in screen resolution. In addition, it can use virtually all pixels as elements of the user interface, letting users design rich and user-friendly interfaces.

- Laser efficiently delivers real-time content over constrained networks. More specifically, it delivers media content in packaged pieces, letting the device display a piece as it's received (as opposed to download-and-play mechanisms). Laser generalizes the streaming con-

1070-986X/05/$20.00 © 2005 IEEE

ONE YEAR LATER: VISION COMING INTO FOCUS

Some day in the far off future, when women and men from Earth are living and working on Mars, they may well look back on Jan. 14, 2004, as the first step of their amazing journey.

(a) **(b)** **(c)**

Figure 1. Laser-based rich media services: (a) rich media portal, (b) interactive mobile TV, and (c) interactive screen saver.

cept—already in place for audio and video data—to scene description and rich media. As such, content providers can design services to keep some information of interest on the screen at all times.

- Finally, Laser delivers rich media service at rates from 10 kilobits per second (Kbps) using vector graphic compression and dynamic scene updates. This drastically reduces end users' waiting time compared to standard Web-like approaches in which the system resends the complete page even if only small changes have been made. This functionality is useful not only in low bit-rate networks such as General Packet Radio Service, but also in higher bit-rate networks in which rich media services can be sent at low rates, preserving bandwidth to improve audio and video quality.

Three use cases demonstrate the new standard's benefits.

Rich media portal

In this application (see Figure 1a), a Laser engine enhances an existing Wireless Access Protocol (WAP) or Extensible Hypertext Markup Language (XHTML) service with rich media much like Flash enhances Web sites. When a user accesses the WAP portal, a hyperlink provides access to the site's rich media part, giving the user a complete, deterministic, and consistent rich media experience. An intuitive interface with a pop-up menu makes navigating the site simpler and the screen feel larger. Figure 1a shows rich text with small—but readable—Arabic fonts. At any time, a hyperlink can bring the user back to the original portal.

Interactive mobile TV

Interactive mobile TV (see Figure 1b) aggregates multiple rich media use cases, from interactive mosaic, electronic program guide, and voting to personalized newscast. All of these applications require a system that can provide deterministic rendering and behavior of rich media content (including audiovisual, text, graphics, and images, along with streamed TV and radio channels) in the user interface. The system must allow fluid navigation through content in a single application or service, as well as local or remote synchronized interaction for voting and personalization (for example, related menus or submenus, advertising, and content in the user profile or service subscription).

Interactive screen saver

The interactive screen saver (illustrated in Figure 1c) is an instance of a larger class of applications that receive content updates in the background (such as fixed or mobile convergent services). The screen saver uses mostly static data—that is, text, graphics, and images—arranged with transitions similar to those in a slide show, with an element of randomness in the presentation. The server adds new elements and removes expired elements from the application stored on the device. Developers should design this application with care to avoid overuse of the device's power resource.

Technical aspects

Part 20 of the MPEG-4 standard consists of two specifications: Laser, which specifies the coded representation of multimedia presentations for rich media services; and SAF, which defines tools for fulfilling the requirements of rich media service design at the interface

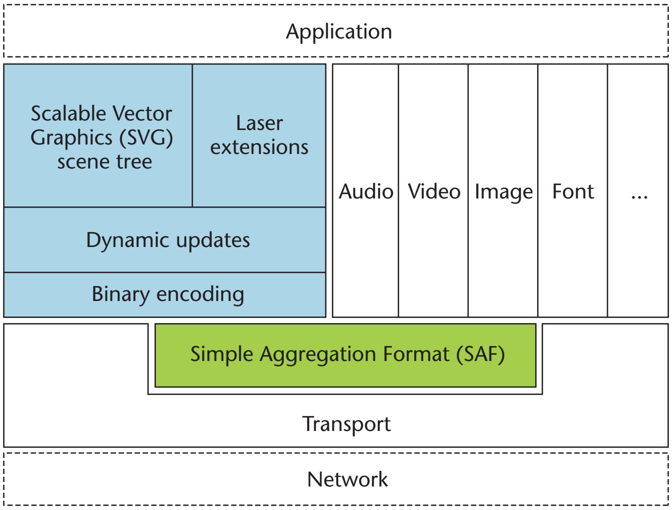


Figure 2. Laser engine architecture.

between scene representation and transport mechanisms.

Laser

In Laser, a multimedia presentation is a collection of a scene description and media (zero, one, or more). A media is an individual content item of the following type: image (still picture), video (moving picture), audio, and by extension, font data. A scene description consists of text, graphics, animation, interactivity, and spatial and temporal layout.

A Laser scene description specifies four aspects of a presentation:

- how the scene elements (media or graphics) are organized spatially (for example, the visual elements' spatial layout);
- how the scene elements are organized temporally (for example, if and how they're synchronized, or when they start or end);
- how users interact with the elements in the scene (for example, when a user clicks on an image); and
- how changes occur in a scene.

A Laser scene description changes through animations. The scene's states during the animation can be deterministic (that is, known when the animation starts) or not (for example, when a server sends modifications to the scene). A *Laser stream* is the scene description's sequence and its timed modifications. The notion of Laser access units is key to streaming Laser content while guaranteeing tight synchronization between the scene and the media assets composing the rich media presentation.

MPEG-4 part 20 defines a Laser engine as the viewer for Laser presentations. Such an engine has rich media composition capabilities on top of the capabilities common to classic multimedia players. These composition capabilities are, as a result of the technology selection process, based on Scalable Vector Graphics (SVG) Tiny 1.1.[1] The composition capabilities rely on the use of an SVG scene tree and are enhanced with key features for mobile services, such as binary encoding, dynamic updates, state-of-the-art font representation, and the stable features of the upcoming SVG Tiny 1.2 specification (described in the "Current Technologies" sidebar), including audio and video support. Figure 2 illustrates the Laser engine architecture.

SVG scene tree. Laser uses an SVG scene tree at its core. It imports composition primitives from the World Wide Web Consortium's (W3C) specifications (all of SVG Tiny 1.1, some of SVG 1.1, and Synchronized Multimedia Integration Language, or SMIL, version 2) and uses the SVG rendering model to present the scene tree. Laser specifies hyperlinking capabilities, audio and video media embedding, vector graphics representations, animation, and interactivity features.

Scene tree extensions. After selecting SVG as Laser's core technology, MPEG identified several areas needing extensions to allow the development of efficient services:

- Because SVG Tiny doesn't have clipping, MPEG added simple axis-aligned rectangular clipping to let developers create such common user interface widgets as ticker tapes and simple transitions.
- SVG lacks a restricted, nonresampling rotation for video and images and a full-screen mode, which MPEG added.
- SMIL and SVG only allow the signaling of one synchronization master, whereas MPEG allows the specification of multiple synchronization references.
- MPEG defines new events: long key press and pause and resume for video, audio, and other timed elements.

Current Technologies

Several technologies are competing to achieve the vision of rich media service to resource-constrained devices.

Flash and Flash Lite

Flash (http://www.macromedia.com/flash) is the current de facto standard for distributing rich media content on the Internet, but it doesn't efficiently address other industry requirements. For example, Flash isn't an open standard, which is critical for massive industry support, especially on mobiles. When using Flash, content creators, services operators, and device manufacturers are tied to Macromedia for the creation, distribution, and playback of rich media content. Because it's difficult to evolve a media infrastructure once it's deployed, organizations tend to avoid proprietary solutions and promote open standards.

Because Flash was designed for PCs, it's unsuitable for the mobile environment, as Flash Lite (the mobile version of Flash) demonstrated. To both remain compatible with its existing PC format and fit constrained device requirements, Macromedia had to compromise on technology. The first version of Flash Lite is a downgraded version of the PC version. Moreover, the problem of having a single vendor and a proprietary format remains.

Therefore, a rich media solution for mobiles must be open and allow easy conversion from the many existing types of Flash and other proprietary content into the new standard.

Two standardization groups have tried to specify standards that would satisfy this requirement: the Moving Picture Experts Group (MPEG) and the World Wide Web Consortium (W3C).

MPEG-4 Binary Format for Scenes

MPEG-4 Binary Format for Scenes (BIFS)[1] is MPEG's first attempt in the composition coding field. It features innovative tools that let users create multimedia content mixing 2D and 3D graphics, introduces the notion of incrementally updating the scene, enables streaming of long-running scenes, and ensures a tight synchronization between a scene's audiovisual elements. We attempted to profile MPEG-4 BIFS to create a small enough subset for use on mobile phones, but to no avail. The inherent content and binary encoding structure makes it inappropriate for mobiles. Therefore, instead of compromising on the technology's performance, MPEG decided to create a rich media standard for constrained devices.

SMIL and SVG

The W3C also attempted to define languages for creating rich media content as an alternative to Flash, including the Synchronized Multimedia Integration Language (SMIL)[2] and Scalable Vector Graphics (SVG)[3,4] standards. Both languages are getting traction in the mobile industry, where both the Third Generation Partnership Program (3GPP) and Open Mobile Alliance (OMA) consortia have adopted their mobile profiles. However, SMIL and SVG are XML languages and rely on the HTML model for content consumption: download-and-play or progressive download and rendering. The streaming of SMIL and SVG content is unspecified, making these models inappropriate for fast, dynamic, interactive, and interoperable content. A new standard for rich media content for mobiles must extend SVG consumption scenarios to streaming and broadcast.

Media File Format

Creating rich media content services for mobile systems isn't only about composition coding. To be successful, a mobile service must provide a reactive and fluid user experience, achieved through efficient delivery mechanisms. However, efficiency is difficult to achieve when distributing rich media content made of individual audio, video, and image content. Separately delivering all of these media streams to a mobile device incurs high latency unless the system uses an efficient aggregation mechanism. In this case, high-latency networks attach a specific penalty to multimedia content consumed in download-and-play mode because the content's waiting time is the sum of each media's waiting time requested separately. Receiving all streams in a single package reduces waiting time to a single request delay.

Such aggregation mechanisms must be efficient and simple to implement. The International Organization for Standardization's (ISO) Base Media File Format[5] is a natural candidate for this task because the mobile industry has already adopted it. However, this file format was designed for storing large amounts of media data, easy editing, and streaming operations. It's inefficient for storing small amounts of timed media data. Thus, we need a simple-to-implement yet efficient aggregation format for mobile to complement the ISO Base File Format.

References

1. *Information Technology—Coding of Audio-Visual Objects—Part 11: Scene Description and Application Engine*, ISO/IEC 14496-11, Int'l Organization for Standardization/Int'l Electrotechnical Commission, 2005.
2. *Synchronized Multimedia Integration Language (SMIL 2.0)*, 2nd ed., World Wide Web Consortium recommendation, 7 Jan. 2005; http://www.w3.org/TR/2005/REC-SMIL2-20050107/.
3. *Mobile SVG Profiles: SVG Tiny and SVG Basic*, World Wide Web Consortium recommendation, 14 Jan. 2003; http://www.w3.org/TR/2003/REC-SVGMobile-20030114/.
4. "Scalable Vector Graphics (SVG) Tiny 1.2 Specification," World Wide Web Consortium working draft, 13 Apr. 2005; http://www.w3.org/TR/2005/WD-SVGMobile12-20050413/.
5. *Information Technology—Coding of Audio-Visual Objects—Part 12: ISO Base Media File Format*, ISO/IEC 14496-12, Int'l Organization for Standardization/Int'l Electrotechnical Commission, 2004.

Related Standards Organizations

The following are the standards organizations that matter in the mobile multimedia services domain.

- *Internet Engineering Task Force* (IETF, http://www.ietf.org). A standards organization responsible for the Internet Protocol (IP), Real-Time Protocol (RTP) for streaming, Hypertext Transfer Protocol (HTTP) for download, and many others. IETF standards are called requests for comments (RFCs).

- *International Organization for Standardization/International Electrotechnical Commission* (ISO/IEC, http://www.iso.org and http://www.iec.ch). A joint standardization group.

- *International Telecommunications Union* (ITU, http://www.itu.int). A standards organization.

- *Moving Picture Experts Group* (MPEG). A group within ISO/IEC responsible for many standards, including Binary Format for Scenes (BIFS), a binary standard for 2D/3D vector graphics; MPEG-J, a standard Java interface to BIFS scenes and Lightweight Application Scene Representation (Laser) for 2D vector graphics; and Simple Aggregation Format (SAF) for lightweight streaming. Both Laser and SAF are part 20 of MPEG-4.

- *Open Mobile Alliance* (OMA, http://www.openmobilealliance.org). A standards organization for mobile applications, formerly the WAP forum, responsible for the Wireless Application Protocol (WAP) standard.

- *Third Generation Partnership Project* (3GPP, http://www.3gpp.org). A standards organization for 3G mobile telecommunications.

- *World Wide Web Consortium* (W3C, http://www.w3.org). A standards organization responsible for Scalable Vector Graphics (SVG) with its SVG Tiny (SVGT) mobile profile; Synchronized Multimedia Integration Language (SMIL), an XML-based standard for interactive audiovisual presentations; Cascading Style Sheets (CSS), a standard for document styling; Document Object Model (DOM), a standard software interface to access XML documents; and Extensible Hypertext Markup Language (XHTML), a reformulation of HTML in XML.

- Finally, SVG Tiny doesn't have simple text underlining, so MPEG added a set of font styles imported from digital television captions.

Dynamic updates. SVG currently supports the following content consumption use cases:

- In classical download-and-play mode, the user waits until the download ends to start viewing the content.

- A progressive rendering mode lets the user view the content while it's downloading. However, the downloaded content only adds new content to the existing content, making it difficult to manage long-running documents.

- Using scripting and a Document Object Model network software interface and an ad hoc protocol, the server communicates scene modifications to the client.

However, SVG can't satisfy the following use cases, currently permitted by Flash or MPEG-4 Binary Format for Scenes (BIFS), in an efficient and interoperable way:

- representation of streamable cartoons,

- partitioning scenes into small packets that fit in size-limited delivery mechanisms (such as cell broadcast),

- dynamic creation of answers to user requests and their integration into the current scene, and

- dynamic push of content into an existing scene.

To enable these use cases, Laser complements SVG with a dynamic updating mechanism that uses Laser commands. Using these commands, a server can, for example, insert or delete graphical elements or modify an object's visual properties. Developers can also use these commands to enable Web cookie-like mechanisms.

Binary encoding. As the W3C specifies, content providers create, store, and transmit SVG content in XML form. Although XML is well suited for Web browsing with powerful PCs and high-bandwidth Internet connections, it incurs severe penalties in performance, code size, and memory requirements for small, predetermined vocabularies. Although the debate still rages over the generality of these statements, MPEG chose to use binary streams in the Laser specification and thus avoid much of the complexity of XML parsing.

Laser's binary format allows SVG content encoding. It uses a compact representation for the SVG elements' structure and specific coding algorithms to encode the SVG elements' attribute values. Because mobile platforms usually lack

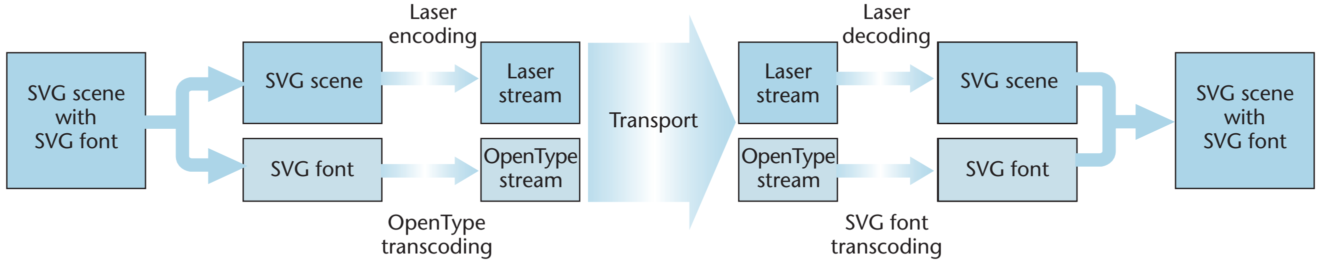


Figure 3. Transmission of font data with Laser content. (SVG = Scalable Vector Graphics.)

hardware float processing, compressing these attribute values must be simpler than on a PC. During the standardization process, MPEG rejected complex computations that would slightly improve the compression ratio while doubling the decoding time. The binary encoding of Laser is therefore straightforward, and its quality resides in the balance of complexity and efficiency. MPEG took special care in the encoding of values for some attribute types, such as float coordinates, vector graphics paths, and transformation matrices.

The Laser binary syntax is extensible, letting developers mix private extensions with normal Laser elements and attributes that can be ignored by decoders that don't know how to process them.

A generic MPEG-7 Systems (BiM) decoder can decode the Laser binary syntax, which is compatible with a predefined BiM configuration. BiM compatibility lets developers encode other XML syntaxes, such as XHTML, with Laser. However, the Laser standard doesn't mandate the use of BiM. The Laser binary syntax is specified such that a specific BiM-agnostic decoder can process it.

Audiovisual support. A key feature of rich media is the support of audio and visual information. SVG 1.1 doesn't support audio or video but SMIL 2 does, and such support will likely be included in SVG Tiny 1.2. The Laser specification includes SMIL 2 audio and video elements as well, including the additional MediaClipping module for VCR-like media control. Binary identifiers refer to the audio and video streams, which travel alongside the Laser stream.

Use of font information. Both SVG and Laser let content creators embed font information in the content to ensure that it's presented as designed and that the terminal doesn't substitute any fonts at viewing runtime. However, MPEG deemed the SVG fonts solution too limited for the following reasons:

- It doesn't leverage device support of current font-rendering engines.
- It doesn't support OpenType fonts. These fonts are more readable at small screen sizes and better equipped for internationalization than SVG fonts.
- It doesn't let the SVG engine share fonts with other applications on the device, such as an XHTML browser.

Laser also carries font information alongside scene information as a media stream. The exact format is optional. One option is to use MPEG-4 part 18,[2] which defines a font data stream and can carry OpenType fonts, possibly compressed.

As a functional subset of OpenType fonts, SVG fonts can be transcoded into OpenType font data without losing SVG font data, and then carried with Laser and possibly reconstructed for playback in an SVG player, as Figure 3 shows.

Services as incremental scenes. Many rich media services rely on Laser's incremental scenes, made possible by the specification's *append* mode. The append mode lets content providers create Laser streams that aren't independent scenes, but add-ons to existing scenes.

The following are typical use cases of incremental scenes:

- *Streaming style*. The scene is designed as a sequence of frames, with a continuous stream of updates changing the current frame into the next frame. Bandwidth usage varies, but never drops to 0. Incremental scenes of this type are usually best transported over streaming protocols like the Real-Time Transport Protocol (RTP).[3]
- *Interactive style*. The scene is interactive with the server processing user requests. The response to a user request is a change to the

existing scene, not a new scene. Such scenarios also require continuous scene updates, but the transmission statistics differ from the streaming style. Bandwidth usage is heavy for a short time after a user request and then drops to 0 until the next user request. Given that mobile usage varies greatly, the next user request could come a few seconds or a few hours later.

From the server's viewpoint, interactive transmissions are a series of separate connections, as opposed to the continuous connection of the streaming style. Application developers typically implement interactive style using separate HTTP connections, because each data burst results from a user request. From a Laser viewer viewpoint, however, the same scene or service is modified. Hence, the server must be capable of signaling an append mode that says, "This stream doesn't contain a totally new scene, but an improvement to the scene the viewer is currently processing."

Append mode also lets the server create multiple responses to possible user requests in advance. If we model the service as a state machine, each transition of the state machine represents a change to the current scene and we can implement it as an append component. This requires careful authoring and scope management, particularly to avoid ID clashes between elements added by different append components. Still, this functionality lets servers cache most of their responses to users, and thus could dramatically improve the service's performance.

Simple Aggregation Format

SAF functionalities include

- simple aggregation of any type of media stream, resulting in a SAF stream with a low overhead multiplexing schema for low-bandwidth networks, and
- possibility of caching SAF streams, as described later.

Multiplexing media streams produces a SAF stream that any delivery mechanism—download-and-play, progressive download, streaming, broadcasting, and so on—can carry.

Aggregation mechanism. The SAF specification defines the binary representation of a compound data stream composed of various elementary streams such as Laser scenes, video, audio, image, font, and metadata streams. Multiplexing the data from these streams into one SAF stream gives us simple, efficient, and synchronous delivery. A SAF stream consists of SAF access units carrying

- configuration information for the media or Laser decoder;
- configuration information for elementary streams not carried in the SAF stream, such as streams that start interactively or that travel over another protocol;
- media or Laser access units;
- an end-of-stream signal, indicating that no more data will arrive in an elementary stream;
- an indication that no more data will arrive in the SAF session; and
- cache units, as explained in the next section.

We can use Laser and SAF independently, but MPEG currently only specifies the use of a SAF stream carrying a Laser stream. In this case, the first SAF access unit carries the Laser engine's configuration information.

Caching capabilities. One way to reduce server response time in high-latency networks is to anticipate what the user will need next and send it with the previous request. The CacheUnit is SAF's mechanism for doing this.

The CacheUnit is a package of information attached to a URL. When it receives this package, the Laser viewer stores it in the cache, associated with the provided URL. If a user requests that URL, the viewer uses the cached version. CacheUnits have expiration dates. This feature, together with the Laser append mode, significantly improves rich media service fluidity.

Relationship with relevant parts of MPEG-4

The Laser standard relates mainly to two existing MPEG-4 specifications:

- *Part 1: Systems*,[4,5] including the synchronization layer and Object Descriptor Framework (ODF); and

- *Part 11: BIFS and MPEG-J* (a standard Java interface to BIFS scenes).[6]

Currently, no specification defines Laser's use in a complete MPEG-4 systems environment. But Laser can replace BIFS in mobile applications, applications not requiring integration of 2D and 3D content, and MPEG-J. Theoretically, we could build an MPEG-4 systems application using a Laser stream instead of a BIFS, but with one major difference: A Laser stream doesn't use the ODF.

SAF maintains compatibility with the synchronization layer but doesn't mandate its use. SAF packet syntax follows synchronization layer packet syntax, with a particular configuration. The SAF specification serves as a multiplexing scheme at the synchronization layer level.

SAF and Laser are independent, but aware, of the rest of the MPEG-4 specification. Therefore, SAF can carry and use MPEG-4 video or audio streams as defined in MPEG-4 part 2, 3, or 10.

Storing Laser content

We can store Laser content in files compatible with the ISO Base Media file format (see the "Current Technologies" sidebar). Because Laser streams are timed streams consisting of access units, storing them in ISO files is straightforward and similar to storing audio or video streams. We store each Laser access unit as a sample. All of the samples form a Laser track identified by a four-character code. We store the decoder's configuration as an entry in the sample description box, and Laser streams comprising only one access unit as a primary item of the file using the meta box structure (as in the Third Generation Partnership Project, or 3GPP, SMIL presentation specification).

Although a SAF stream is timed, it's an aggregation of several elementary streams (such as Laser, audio, and video), and it would therefore be inappropriate to store it as-is in an ISO file. However, SAF can serve as a delivery protocol, letting content creators define hint tracks to generate one SAF stream per ISO file, aggregating the file's Laser, audio, and video tracks.

Streaming

MPEG and the Internet Engineering Task Force have jointly specified, in MPEG-4 part 8[7] and RFC 3640,[8] the transport over IP of any kind of MPEG-4 elementary stream, as long as it's packaged using the synchronization layer syntax. Hence, MPEG-4 part 20 doesn't specify any new elements in this domain. As elementary streams, Laser streams can be wrapped in synchronization layer packets and transported in RTP packets, as RFC 3640 specifies. Currently, no specification exists for carrying SAF packets over RTP.

Conclusion

MPEG will likely extend the Laser specification to support new features on mobile devices. Another direction for future work relates to the link between the Laser specification and the W3C Compound Document Format Group's work on the interactions among W3C standards such as SVG, SMIL, and XHTML. It's therefore a natural extension for MPEG to study how this work will affect Laser.

Finally, the Laser standard's development might impact other MPEG activities such as MPEG multimedia middleware (M3W) activity. M3W aims to improve the portability of applications and services by defining a series of application programming interfaces (APIs) that can evolve as middleware technology itself evolves. Dedicated APIs for accessing a Laser engine could be an interesting extension of M3W. **MM**

Acknowledgments

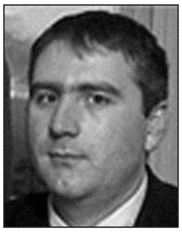
The European Commission in the course of the Information Society Technologic's (IST) Dynamic and Distributed Adaptation of Scalable Multimedia Content in a Context-Aware Environment (DANAE) Project (http://danae.rd.francetelecom.com) partially funded Cyril Concolato's work.

References

1. *Scalable Vector Graphics (SVG) 1.1 Specification,* World Wide Web Consortium recommendation, 14 Jan. 2003; http://www.w3.org/TR/2003/REC-SVG11-20030114/.
2. *Information Technology—Coding of Audio-Visual Objects—Part 18: Font Compression and Streaming,* ISO/IEC 14496-18, Int'l Organization for Standardization/Int'l Electrotechnical Commission, 2004.
3. H. Schulzrinne, A. Rao, and R. Lanphier, *Real Time Streaming Protocol,* Internet Eng. Task Force, RFC 2326, Apr. 1998; http://www.ietf.org/rfc/rfc2326.txt.
4. *Information Technology—Coding of Audio-Visual Objects—Part 1: Systems,* ISO/IEC 14496-1, Int'l Organization for Standardization, 2004.
5. *Information Technology—Multimedia Content Description Interface—Part 1: Systems,* ISO/IEC 15938-1, Int'l Organization for Standardization/Int'l Electrotechnical Commission, 2002.

6. *Information Technology—Coding of Audio-Visual Objects—Part 11: Scene Description and Application Engine*, ISO/IEC 14496-11, Int'l Organization for Standardization/Int'l Electrotechnical Commission, 2005.
7. *Information Technology—Coding of Audio-Visual Objects—Part 8: Carriage of ISO/IEC 14496 Contents over IP Networks*, ISO/IEC 14496-8, Int'l Organization for Standardization/Int'l Electrotechnical Commission, 2003.
8. J. van der Meer et al., *RTP Payload Format for Transport of MPEG-4 Elementary Streams*, Internet Eng. Task Force, RFC 3640, Nov. 2003; http://www.ietf.org/rfc/rfc3640.txt.

Jean-Claude Dufourd is a cofounder and chief scientist of Streamezzo, chair of the MPEG committee's Integration subgroup, and a member of the W3C's Scalable Vector Graphics (SVG) and Compound Document Formats (CDF) working groups. His research interests include mobile multimedia and MPEG-4 systems. Dufourd has a PhD in computer science from Ecole Normale Supérieure de Paris.



Olivier Avaro is a cofounder and vice-president of Streamezzo in charge of business development. Before this, he was CTO of the Hyper-languages and Multimedia Dialogs division of France Telecom R&D. He is chair of the Systems subgroup of the MPEG committee. His interests include mobile multimedia, and MPEG-4 systems. Avaro is a graduate of Ecole National Supérieure des Télécommunications de Bretagne.

Cyril Concolato is a research assistant within the Communication and Electronics Department of Ecole Nationale Supérieure des Télécommunications in Paris, currently pursuing a PhD on Scene Description Representations at ENST. He is member of the SVG working groups of W3C and of the MPEG standardization group. His research interests include Scene Description Languages, their compression, and efficient rendering.

Readers may contact the authors at Streamezzo, 83 Boulevard du Montparnasse, 75006 Paris, France; jean-claude.dufourd@streamezzo.com.