Frédéric Gabin, Markus Kampmann,
Thorsten Lohmar, and Clinton Priddle

3GPP Mobile Multimedia Streaming Standards

Data rates offered by mobile networks are increasing, as are the capabilities of mobile devices. With this, mobile multimedia services are getting wider distribution [1]. Traditional linear mobile TV services, where the viewer has to watch a scheduled TV program at the offered time, were initially dominating but other streaming services like Video on Demand, User-Generated Content (USG) services, or Internet streaming services are becoming more important. This article describes the 3rd Generation Partnership Project (3GPP) standards for mobile unicast multimedia streaming and how they are used in almost 200 mobile networks.

BACKGROUND

3GPP is a collaborative project that defines the 3G mobile communication system and its evolutions like high-speed packet access (HSPA), long-term evolution, and Internet Protocol (IP) Multimedia Subsystem (IMS). To provide a standard for mobile multimedia streaming, 3GPP defined the packet-switched streaming service (PSS) for unicast streaming of speech, audio, video, and subtitles to 2G/3G/4G mobile terminals. It provides an entire end-to-end streaming and download framework for mobile networks spanning from a server file format for storage of streaming sessions and streaming servers for delivery, to streaming clients for reception and a file format for device storage. In PSS, a set of multimedia streaming solutions have been defined as part of the 3GPP service layer: streaming based on the Real-Time Transport Protocol (RTP), progressive download of files formatted according to the 3GPP file format (3GP) and recently also an adaptive Hyper Text Transfer Protocol (HTTP) streaming solution.

The main specification [3] defines the protocols and media codecs for servers and clients. The PSS standard is an umbrella standard covering specifications such as the 3GPP file format (3GP) [5], the timed text format for subtitling [6]; the 3GPP Synchronized Multimedia Integration Language (SMIL) Language Profile [7] for scene descriptions and the IMS-based PSS and Multimedia Broadcast and Multicast Services (MBMS) user services [4].

3GPP PSS first appeared as "simple PSS" in 3GPP Release 4 (March 2001), and its protocols were based on the Real-Time Streaming Protocol (RTSP) [Request for Comments (RFC) 2326][8], the Session Description Protocol (SDP) [RFC 4566] for session setup and control, and the RTP [RFC 3550] for transporting real-time speech, audio and video.

While transport of continuous media was exclusively performed over RTP/UDP/IP, the HTTP protocol was used for transport of session descriptions, SMIL presentation descriptions, and static media files (e.g., pictures). Since then, a number of improvements and enhancements have been introduced. Figure 1 indicates the main PSS features and their time line.

The 3GP file format was introduced in Release 5 and is based on the International Organization for Standardization (ISO) base file format structure. It extends the ISO base file format to support the PSS codecs and other 3GPP specific features such as storage of Digital Rights Management protected content. Capability exchange was introduced based on the User Agent Profile (UAProf) protocol to allow servers to match content to the device rendering capabilities. Progressive download of 3GP audio/video clips over HTTP [RFC 2616] was introduced in Release 6. In the same release, the PSS RTP adaptive streaming mechanism [2] was defined to provide the best content quality and interrupt-free playback even over a variable bit rate wireless bearer.

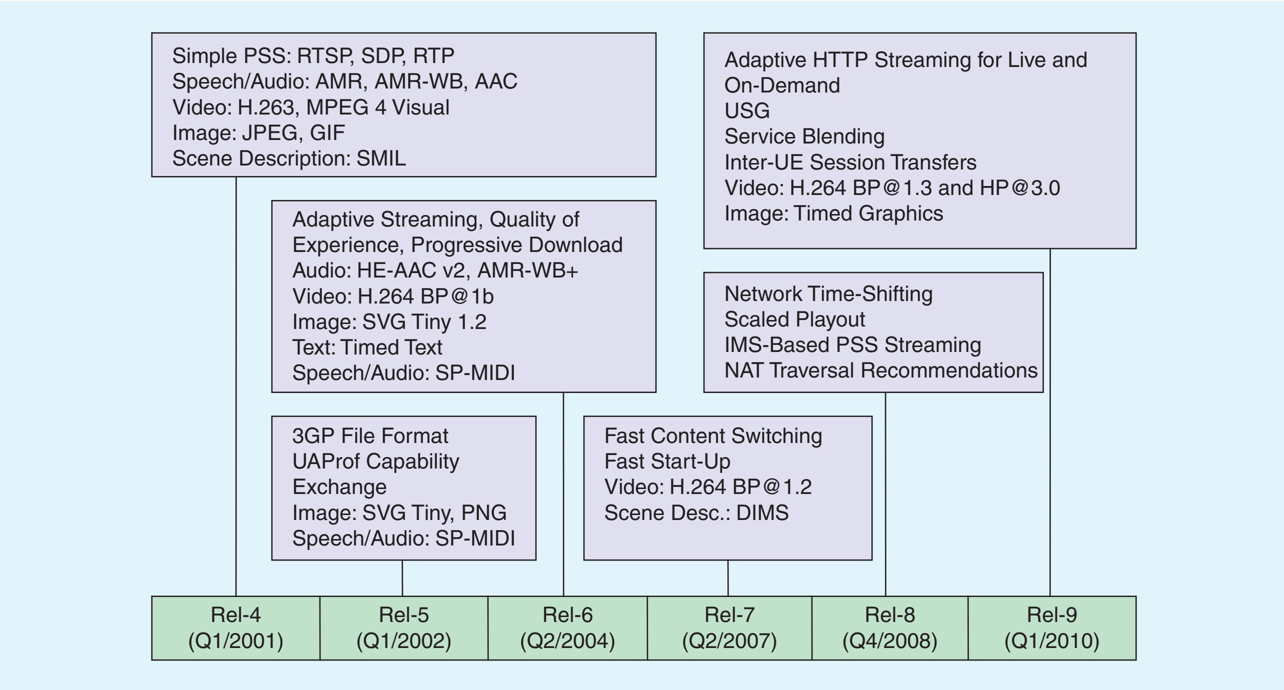
The quality of experience (QoE) reporting mechanisms were introduced in the same release to allow content providers and network operators to retrieve metrics reflecting the level of quality experienced by the user. Fast content switching mechanisms for RTSP-based streaming were introduced in Release 7 to reduce the interruption time when switching from one content stream to the other, thereby greatly improving user experience: now only one client request is needed to switch to new content under optimal conditions.

In Release 8, the network-based time-shifting feature was introduced to enhance the access to live TV sessions, allowing the PSS client to pause live TV and even navigate (rewind, fast forward) within the time-shifting buffer range. The time-shifting buffer is maintained in the server to save network transmission capacity. This is different than in today's broadcast systems, where time shifting is a feature of the set-top-box. Also in Release 8, IMS-based PSS streaming was introduced to define how the 3GPP PSS provides streaming services in the context of the IMS architecture.

State-of-the-art codecs have gradually been added to offer the best compromise in quality versus complexity. In Release 9, the supported codecs for continuous media include H.264/AVC baseline (BP) and high profile (HP) for video and HE-AAC v2 and AMR-WB+ for audio.

Digital Object Identifier 10.1109/MSP.2010.938087

 1053-5888/10/$26.00©2010IEEE



[FIG1] 3GPP releases and multimedia streaming features.

In the last two years, adaptive HTTP streaming proprietary solutions for live and on-demand content have appeared on the market mainly because HTTP/Transmission Control Protocol (TCP)-based transport does not have the same firewall and network address translator (NAT) traversal problems as traditional streaming. NATs are very common in today's Internet to provide a full network behind a single address. The PSS RTSP/RTP solution can also be run over TCP by interleaving RTP into the RTSP session, however, trends are heading towards enabling the use of cheaper HTTP servers (and caches) by moving towards HTTP-based media streaming.

In contrast to RTSP/RTP streaming, the progressive download of 3GP audio/video clips over HTTP [RFC 2616], available in 3GPP since Release 6, does not support client-based adaptive streaming and has very limited support for live service. Therefore, 3GPP decided in January 2009 to open a work item to define an HTTP streaming standard that would allow for dynamic bit rate adaptation and improved live services with 3GP files.

Three important PSS features, specifically, adaptive HTTP streaming, IMS-based PSS streaming, and QoE reporting, are described below in more detail.

ADAPTIVE HTTP STREAMING

Adaptive HTTP streaming was introduced in Release 9 of the PSS standard [3] and was finalized in March 2010. Furthermore, extensions to the 3GPP file format [5] were made related to the timing and synchronization of HTTP streaming content. 3GPP adaptive HTTP streaming provides an entire system for streaming including storage, transport, and media rate adaptation to available link bit rates. As each media segment is identical to all users, the solution is HTTP cache and content delivery network (CDN) friendly.

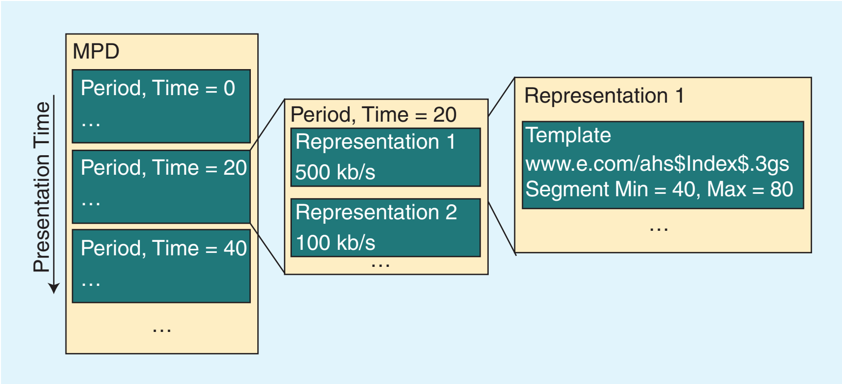
MEDIA PRESENTATION DESCRIPTION

An adaptive HTTP streaming content resource is described through a media presentation description (MPD). An MPD describes the media session as a number of segments, including their timing information and relationship to other segments. Information about the language of the content and the codecs used is also present.

The MPD consists of three major components: periods, representations, and segments. As depicted in Figure 2, period elements are the outermost part of the MPD. Periods are typically larger pieces of media that are played out sequentially. Inside a period, multiple different encodings of the content may occur. Each alternative is called a representation. These alternative representations can have, for example, different bit rates, frame rates, or video resolutions. Finally, each representation describes a series of segments by HTTP URLs. Those URLs are either explicitly described in the representation (similar to a playlist) or described through a template construction. For the template construction, the client inserts a valid segment index and possibly a valid representation index into the template to derive a valid URL. The MPD format is flexible and can support other media container formats such as MPEG-2 Transport Stream. Playlist or ad-insertion functionality can easily be achieved by chaining periods of different content.

MEDIA PRESENTATION UPDATES

Content providers can update the MPD to add other content items such as advertisements to the media presentation. MPDs are updated using standard HTTP polling. The same URL for the



[FIG2] Structure of the MPD for adaptive HTTP streaming.

MPD is used during the entire session. This feature is typically needed for live content but may also be used for very long or very detailed on-demand sessions (i.e., where it is impractical to describe the entire session in a single MPD).

Content in an MPD may be given an availability start and end time in Coordinated Universal Time (UTC) format, which can be used to preschedule content or make live content available for on-demand playback for a period of time after the end of the program broadcast.

CODECS AND FILE FORMATS

The adaptive HTTP streaming protocol is described in the following two parts:

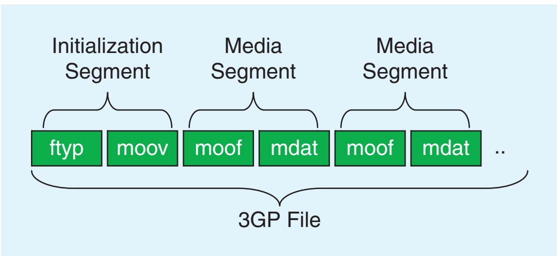
- an MPD for any content file format
- a 3GPP-specific instantiation.

The main part of this 3GPP-specific section is about timing and synchronization in 3GP files. This section is necessary, as although 3GP files (and ISO files in general) have implicit timestamps, in-live streaming content is not necessarily available from the beginning of the clip.

In the context of 3GP files (Figure 3), an initialization segment is simply the configuration data ("ftyp" and "moov" box pairs in ISO file format structure), while the media segment are concatenations of one or more fragments of media pointers and samples ("moof" and "mdat" box pairs). Concatenation of the initialization segment and one or more media segments results in a valid 3GP file.

The media segment files are not stand-alone ISO files and have been given a special Multipurpose Internet Mail Extensions (MIME) type, file extension, and segment type indication. A new track fragment adjustment box has been defined to provide synchronization between tracks when playback starts in the middle of a period. Another new box is the segment index box, which provides an index of movie fragments and other segment index boxes inside a segment. This box may be used to place a segment on the global time line.

Finally, the codecs defined for adaptive HTTP streaming in 3GPP are identical to those defined for other streaming solutions such as PSS streaming.



[FIG3] Structure of a 3GP file for adaptive HTTP streaming.

EXPECTED CLIENT BEHAVIOR FOR MEDIA ADAPTATION

The construction of the MPD with its representations per period allows for an adaptive HTTP streaming solution. In such a case, the content must be available in multiple media quality levels, which are described by alternative representations with different bit rates.

Each time a client downloads a segment, it can choose among the representations of that particular segment. If the segments from different representations are temporally aligned and start with self-decodable pictures (I Frames), switching can be performed directly on segment boundaries. Otherwise, the client needs to search for keyframes and to synchronize the streams of the different representations. The box structure of 3GP files and the segment index structure allow clients to efficiently find keyframes in a segment.

Each representation is characterized by bit rate, language, or codec. On top of this, representations may be given a quality ranking in the MPD. This helps the client decide which representation to choose, as increasing the bit rate does not necessarily increase the quality—especially when different codecs are used.

Adaptation is typically performed through the client by choosing the representation with the highest quality possible, given the constraints of both the terminal (such as codecs supported) and the current network conditions.

STORAGE AND DELIVERY

Several server-side storage modes are possible for adaptive HTTP streaming. With normal HTTP servers, each segment may either be stored in its own file or as part of a larger file and fetched using HTTP byte range requests. In the later case, a byte range is taken from a larger file for each segment request. This larger file can be a valid file containing all segments of a representation and may be used for legacy progressive download clients and adaptive HTTP streaming clients.

Another server-side storage option is to have the HTTP server dynamically prepare the segment for each HTTP request. The server then takes a part of a larger file using a server-side mapping description. This type of content management is, however, transparent to the client and intermediate network and therefore out of the scope of the specification.

As mentioned earlier, segments are requested using standard HTTP. This means that a standard HTTP server can be used, as can standard HTTP caches. Thus, the 3GPP adaptive HTTP streaming solution allows to reuse the same content distribution infrastructure or CDN as for normal Web traffic. The standard is flexible when it comes to file sizes and configurations where all segments have approximately the same size are possible.

Unlike traditional streaming, all segments are intended to be identical to all clients whether it is the first of a series to be downloaded (i.e., when tuning in), if the user has been watching for a period of time, or if the user rewinds and watches a program again. This is a requirement for the system to function correctly with caching.

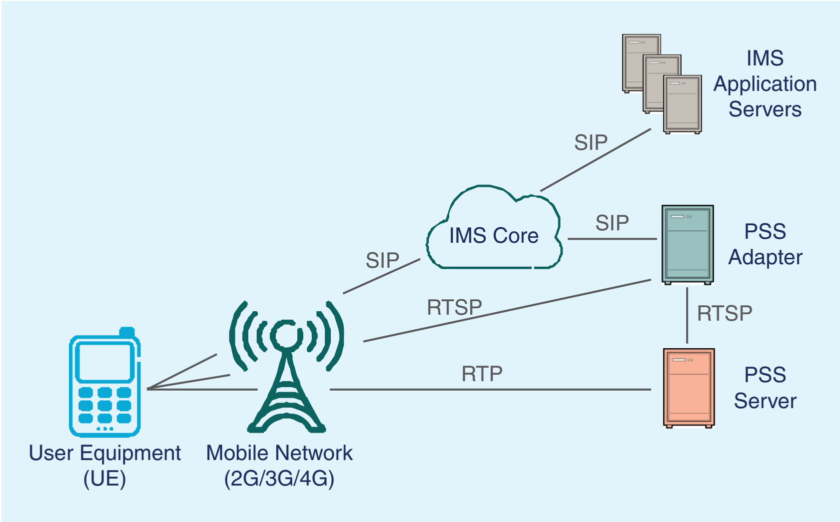
HTTP STREAMING AND OTHER PROTOCOLS

Throughout this section, we have been referring to content delivered over the HTTP Protocol. Both segments and the MPD are described by HTTP URLs, which are typically fetched using HTTP. However, the solution is generic in its description of content and protocols other than HTTP may be used. A concrete example is 3GPP MBMS [9], where files are transmitted using the MBMS file delivery method. File delivery over unidirectional transport (FLUTE) also uses HTTP URLs to provide the content location of the encapsulated files. Thus, segments may be pushed to clients over MBMS broadcast in advance and used by HTTP streaming clients in the same way as if they were in a local HTTP cache.

IMS-BASED PSS STREAMING

IMS [10] is the 3GPP Next Generation Network (NGN) solution. It was introduced in 3GPP Release 5 and combines the world of cellular communication with the Internet. A common control layer is defined that is based on open standards such as the Session Initiation Protocol (SIP) for session establishment, as defined in the Internet Engineering Task Force (IETF). Using IMS, different communication services benefit from the common framework for charging, quality of service (QoS) control and authorization and are thus easy to deploy and integrate. In IPTV standardization, ETSI TISPAN and the Open IPTV Forum define IPTV solutions with IMS as the underlying network framework.

The usage of IMS for PSS user services is defined in [4]. A tight alignment with the procedures for IMS-based IPTV as defined by ETSI TISPAN and Open IPTV Forum was achieved during standardization to offer converged TV services to mobile and fixed receivers.



[FIG4] **IMS-based PSS functional architecture (simplified).**

Figure 4 shows the functional architecture of IMS-based PSS as defined in [4]. The IMS core allows for QoS control, charging, and authorization. IMS application servers enable service provider discovery, user service discovery, and user service description retrieval as described in [4].

Streaming session establishment is enabled through SIP via the IMS core. The PSS adapter provides a protocol translation function for translation between SIP and RTSP controlling the PSS server. The PSS server delivers media streams via RTP to the User Equipment (UE). The media session is then controlled by RTSP commands (e.g., pause or fast forward) exchanged between the UE, PSS adapter and the PSS server. This includes procedures for fast content switching, such as for example fast switching between TV channels using RTSP.

Enhanced features for IMS-based PSS streaming are defined in [4]. These include the blending of PSS streaming with other IMS services, server-based storage and retrieval of bookmarks within VoD sessions, provision and distribution of UGC as well as the transfer of ongoing streaming sessions from one device to another (Inter-UE Session Transfer).

QoE REPORTING

The QoE reporting scheme allows the operator or the service provider to get feedback about the content reception from the streaming client. Typical QoE metrics may describe the initial start-up delay for a streaming service or quality degradations because of packet losses or rebufferings. This makes it possible to understand how the end users are experiencing the quality of the streaming service and that the network provides the correct QoS.

QoE reporting was introduced in Release 6 for RTP/RTSP-based streaming and was adapted in Release 9 to also cover adaptive HTTP streaming as well as progressive download streaming. If supported by the mobile terminal, QoE reporting can be activated by the operator or, for RTP/RTSP-based streaming, by the service provider.

QoE reports can be sent for selected sessions, all sessions, or for a random subset (for example 5%) of the sessions. It is possible to specify which metrics shall be reported and how often these shall be measured. The QoE reporting can be therefore fine-tuned to achieve a good understanding of the service quality, while still not consuming unnecessary uplink radio capacity. The QoE reports sent to the operator can also be selectively forwarded to the relevant service providers, assuming QoE forwarding agreements are made between operators and service providers. However, such forwarding is not covered in the 3GPP standard.

CONCLUSION

3GPP streaming solutions are widely deployed and used in almost 200 mobile networks across the world for media streaming and mobile-TV services. The vast majority of 3G mobile terminals support 3GPP PSS.

3GPP is continuing the development of the PSS standard within Release 10 (which should be finalized Q1/2011) and its further releases. In doing so, PSS will evolve alongside the developments of mobile terminals and networks. For instance, the increased occurrence of mobile terminals with stereo three-dimensional (3-D) displays has been noted by 3GPP and investigations of PSS streaming of stereoscopic 3-D video are ongoing.

Finally, it is important to note that the 3GPP PSS standard and its evolutions extend beyond the 3GPP mobile access. For instance, the 3GPP adaptive HTTP streaming was adopted by the Open IPTV Forum in its Release 2 to offer streaming services over the fixed access and was adopted by MPEG as baseline for its MPEG Dynamic and Adaptive Streaming over HTTP (DASH) specification.

AUTHORS

Frédéric Gabin (frederic.gabin@ericsson.com) is standardization manager at Ericsson France in the area of media and mobile TV.

Markus Kampmann (markus.kampmann@ericsson.com) is senior research engineer at Ericsson Research in Herzogenrath, Germany, in the area of multimedia transport, coding, and signal processing.

Thorsten Lohmar (thorsten.lohmar@ericsson.com) is a master research engineer at Ericsson Research in Herzogenrath, Germany, in the area of multimedia transport optimizations for mobile TV and IPTV systems.

Clinton Priddle (clinton.priddle@ericsson.com) is a senior research engineer at Ericsson Research, Stockholm, Sweden, in the area of video coding and transport.

REFERENCES

[1] I. Bouazizi, K. Järvinen, and M. Hannuksela, "3GPP mobile multimedia services," *IEEE Signal Processing Mag.*, vol. 27, no. 5, pp. 125–130, Sept. 2010.

[2] P. Fröjdh, U. Horn, M. Kampmann, A. Nohlgren, and M. Westerlund, "Adaptive streaming within the 3GPP packet-switched streaming service," *IEEE Network*, vol. 20, no. 2, pp. 34–40, Mar. 2006.

[3] 3GPP TS 26.234, Transparent End-to-End Packet-Switched Streaming Service (PSS); Protocols and codecs.

[4] 3GPP TS 26.237, IP Multimedia Subsystem (IMS)-Based Packet Switch Streaming (PSS) and Multimedia Broadcast/Multicast Service (MBMS) User Service; Protocols.

[5] 3GPP TS 26.244, Transparent End-to-End Packet Switched Streaming service (PSS); 3GPP file format (3GP).

[6] 3GPP TS 26.245, Transparent end-to-end packet switched streaming service (PSS); Timed text format

[7] 3GPP TS 26.246, Transparent End-to-End packet Switched Streaming Service (PSS); 3GPP SMIL Language Profile.

[8] H. Schulzrinne, A. Rao, and R. Lanphier. Real Time Streaming Protocol (RTSP). RFC 2326 [Online]. Available: http://tools.ietf.org/html/rfc2326

[9] F. Hartung, U. Horn, J. Huschke, M. Kampmann, and T. Lohmar, "MBMS—IP multicast/broadcast in 3G networks," *Int. J. Digital Multimedia Broadcast.*, vol. 2009, 2009.

[10] G. Camarillo and M. A. Garcia-Martin, *The 3G IP Multimedia Subsystem (IMS)*, 2nd ed. Hoboken, NJ: Wiley, 2006.

SP

from the GUEST EDITORS continued from page 17

activities such as parameter estimation, filtering, and prediction. In particular, the variational framework—which encompasses methods such as loopy belief propagation, mean-field inference, and expectation propagation—provides a general optimization-based methodology that makes use of the structure of the graph to compute approximations to marginal and conditional probabilities of interest. Rennie et al. discuss the use of these methods in the setting of the factorial HMM, a special case of dynamical graphical model. Their framework has been applied very successfully in the domain of multitalker speech separation and recognition; in particular, they show that the factorial HMM performs better than human listeners on a benchmark speech separation task.

This special issue includes two articles that focus on the intersection of graphical models with another very active area of research: sparse signal recovery. The article by Seeger and Wipf discusses a Bayesian approach to problems of sparse estimation. They make use of the variational framework to develop efficient message-passing algorithms for posterior inference under their Bayesian model. The article by Cevher et al. describes how the sparse reconstruction/compressive sensing formulation can be augmented to include "structured sparsity" priors using tree-based and MRF models that have found use in image modeling and compression. In addition, this article takes a look at how structured sparsity in the measurement matrix can be exploited.

Finally, graphical models have had a major impact in coding theory over the past decade, and this issue includes one article that discusses new directions in this area. In particular, Fresia et al. show how the problem of joint source/channel coding can be tackled within the framework of graphical models. They propose a double low-density parity-check code, in which the first code is used as a source encoder and the second for channel coding and show how it can be decoded using the sum-product algorithm.

Going forward, we see graphical models continuing to play a central, unifying role across engineering and statistics. We anticipate continuing developments in the area of nonparametric graphical models, both in the Bayesian and the frequentist paradigms. In particular, although much research in graphical models has focused on discrete random variables, we expect to see further work on graphical models with continuous variables, where the component conditional probabilities and potential functions will need to make use of the flexible models used in nonparametric regression, nonparametric density estimation, and survival analysis. We also expect to see major growth in applications of graphical models as they become more widely known in applied communities; in particular, we see many opportunities for exciting applications in computational biology, image processing, sensor networks, information retrieval, and social networks.

SP