

Prototipo de un Sistema P2P de Multiconferencia basado en SIP

André Ríos, Alberto José González y Jesús Alcober

Departamento de Ingeniería Telemática, Universitat Politècnica de Catalunya / Fundació i2cat

Av. Canal Olímpic 15, PMT-Edifi C4, Mediacat. Castelldefels

{andre.rios, alberto.jose.gonzalez, jesus.alcober}@i2cat.net

Abstract — En la actualidad, las herramientas colaborativas como la videoconferencia están teniendo un gran desarrollo, motivado principalmente por su capacidad de permitir una comunicación interactiva y facilitar el trabajo conjunto entre usuarios, sin importar que éstos no estén reunidos en el mismo lugar físico. En general, una videoconferencia multipunto o multiconferencia se realiza de forma centralizada utilizando un dispositivo de altas prestaciones llamado MCU (Multipoint Control Unit) y empleando un esquema cliente-servidor. Este esquema es muy usado en ambientes empresariales o de operador. Por otro lado, en Internet, la aparición de aplicaciones Peer-to-Peer (P2P) para distribución de ficheros y contenido multimedia ha permitido un uso distribuido de recursos de red, con lo cual se ha ganado en escalabilidad y versatilidad. Sin embargo, en este ámbito, servicios como una videoconferencia, donde la sensación de interactividad y presencia son puntos claves, el esquema P2P se encuentra en estado embrionario, ya que aún presenta problemas abiertos, relacionados por ejemplo, a la cantidad de participantes en el sistema, al uso de algoritmos ALM que no facilitan la interactividad, a la carencia de mezclado de audio/video, etc. Adicionalmente está la falta de integración con otros servicios de media de valor añadido (IM, VoIP, etc.) que son deseables de tener de forma integrada. Este trabajo presenta un prototipo de un sistema de multiconferencia P2P basado en Session Initiation Protocol (SIP) que intenta resolver algunas de estas cuestiones y sirve como prueba de concepto para desarrollar un esquema flexible para un servicio de multiconferencia, el cual aún no ha sido definido a nivel de estándar P2P-SIP.

I. INTRODUCCIÓN

Las redes P2P han ganado popularidad en los últimos tiempos, gracias al explosivo desarrollo de aplicaciones de compartición de archivos como por ejemplo Emule o Kazaa. Sin embargo, hoy en día, las aplicaciones de media (audio y video) streaming en tiempo real son las que están generando mayor expectación e interés tanto a nivel académico como comercial, debido a que éstas introducen nuevos problemas tecnológicos a resolver, en términos de fiabilidad, escalabilidad, robustez y gestión de las comunicaciones multimedia en un esquema P2P, que permitirán el desarrollo de productos comerciales como la televisión o radio por Internet y más recientemente el despliegue de servicios de multiconferencia P2P. Una solución de media streaming en un esquema P2P debe tener en consideración aspectos tales como la transmisión de grandes volúmenes de información bajo una estricta limitación de tiempo, sin obviar la naturaleza dinámica y heterogénea que hacen que los nodos (“peers”) tengan una conducta impredecible.

Una videoconferencia es un aplicación en tiempo real que permite la comunicación entre dos o más puntos en ambos sentidos de forma simultánea a nivel de audio y de video. Esta aplicación es una herramienta colaborativa que tiene importantes desafíos técnicos en su operación, debido a que es necesario mantener la experiencia de interactividad entre los participantes, en especial cuando éstos son muchos y las condiciones de red son adversas. El esquema básico de una videoconferencia es una comunicación bidireccional punto-a-punto con el estricto requerimiento de mantener acotado el retardo extremo a extremo ($< 150\text{ms}$) [1]. Los servicios de videoconferencia multipunto generalmente son desplegados en dos esquemas: en grupos abiertos y en grupos cerrados. El primero está pensado para distribuir por ejemplo un evento de interés público, como puede ser una clase, una conferencia, un concierto, etc. y el segundo consiste en un grupo reducido y controlado de usuarios que participan de una reunión o discusión o de un seminario. Nosotros en este trabajo estamos interesados en desarrollar un prototipo que soporte ambos esquemas para así tener una plataforma genérica y escalable.

Hay diferentes soluciones para desplegar una videoconferencia P2P. La aproximación más tradicional es IP Multicast a nivel de red, pero actualmente tiene un despliegue limitado debido a sus problemas de complejidad, escalabilidad y seguridad. Una segunda aproximación es mantener una comunicación “full-mesh” entre los participantes de la conferencia, pero claramente esta solución posee problemas de escalabilidad y aplicabilidad especialmente cuando el número de usuarios es grande. Finalmente, la tercera aproximación, la cual es la más utilizada hoy en día, es el Multicast a nivel de aplicación (Application Layer Multicast, ALM) [2]. Su principal ventaja es permitir la reducción del uso de los recursos de red y balancear los enlaces de la red subyacente. Por otro lado, SIP [3] es un protocolo de señalización ampliamente usado que permite iniciar, mantener y terminar una sesión multimedia. Actualmente existen varios trabajos que intentan mezclar técnicas de P2P con SIP siendo dos las aproximaciones más comunes: SIP usando P2P para reemplazar el servicio de localización por un protocolo P2P y P2P sobre SIP la cual es una implementación P2P que utiliza señalización SIP. El grupo P2PSIP WG de la IETF está intentando estandarizar lo que se conoce como P2P-SIP [4], sin embargo, aún no existen borradores IETF (drafts) de una propuesta de esquema para desarrollar un servicio de multiconferencia.

En este trabajo, nosotros estamos interesados en explorar técnicas ALM que permitan mejorar las soluciones existentes y faciliten el ofrecer un sistema de videoconferencia que soporte la mezcla de media (centralizada o distribuida) dependiendo del esquema seleccionado (conferencia pura o multiparty). Adicionalmente, entendemos que un sistema de conferencia debe poseer unas características básicas: facilidad en la puesta en marcha (start-up) y mantenimiento de la conferencia permitiendo agregar o retirar usuarios de forma dinámica. En este sentido, existe un paquete de trabajo llamado SIP based Conference

Control Framework (SIPPING) [5] que ofrece un entorno de trabajo adecuado y que permite la interoperabilidad con otros sistemas SIP. De esta forma, este trabajo presenta un prototipo de un sistema de videoconferencia multipunto basado en ALM y SIP el cual será una primera aproximación para resolver algunos de los problemas asociados a la multiconferencia P2P.

Finalmente, este trabajo está organizado de la siguiente forma. Primero en la próxima sección se describen brevemente algunos trabajos relacionados a media streaming P2P y videoconferencia. En la sección III se analiza como SIP permite desarrollar una conferencia multipunto. La sección IV se presenta el prototipo ALM+SIP y el entorno de test con el fin de validar esta prueba de concepto. Finalmente, en la sección V se presentan las conclusiones generales de este trabajo.

II. TRABAJOS RELACIONADOS

La solución convencional para realizar una videoconferencia multipunto consiste en utilizar una MCU. Esta solución permite una solución centralizada pero carece de escalabilidad debido a la gran cantidad de recursos (ancho de banda y CPU) que se necesitan a nivel central. Esta solución es usada por operadores y empresas, pero su aplicabilidad en Internet es reducida. Otra solución es el uso de IP Multicast mediante aplicaciones MBONE, las cuales hasta la fecha han sólo sido desplegadas localmente en ambientes controlados como universidades y centros de investigación. Por otro lado, el uso de multicast a nivel de aplicación (ALM) ha sido muy difundido, pero siempre enfocado a aplicaciones como P2P-TV o P2P-VoD. Hay sin embargo, algunos trabajos que se han presentado como solución de videoconferencia P2P usando técnicas ALM [6], [7], [8], [9]. En el ámbito comercial existen dos soluciones de videoconferencia P2P basadas en SIP: Damaka y Openwengo, las cuales permiten comunicaciones punto-a-punto usando un concepto de MCU con soporte de pocos usuarios en el sistema. Un aspecto importante a mencionar es que existen una gran cantidad de técnicas ALM, pero en general su enfoque no ha sido el reducir el retardo extremo a extremo entre peers y trabajar en mecanismos de recuperación de fallos, lo cual es importante en una videoconferencia.

Nuestra solución difiere de los trabajos previos en aspectos de operación principalmente. El uso de SIP no sólo nos permite efectuar el "setup" de un videoconferencia, sino también en permitir la ampliación a otros tipos de servicios. Además el prototipo hace que cada peer del sistema, dependiendo de sus capacidades, pueda actuar como mezclador de flujos de video (mixing) reduciendo el tráfico y mejorando la visualización. Adicionalmente se debe mencionar que este trabajo está basado en un trabajo anterior descrito en [10].

III. ENTORNOS DE TRABAJO PARA CONFERENCIA SIP

SIP está basado en diálogos llamados mensajes SIP que permiten iniciar, modificar o terminar una sesión entre User Agents (UAs). El protocolo SIP fue pensando inicialmente sólo para sesiones entre dos puntos y su primera especificación no resolvía las comunicaciones multipunto. Como solución a ello, se han propuesto dos frameworks: SIPPING y XCON. Este trabajo está basado en el primero, ya que éste no define un nuevo protocolo de control, lo cual garantiza la interoperabilidad con otros esquemas SIP y permite de forma simple llevar a cabo una conferencia: invitaciones, abandonos y configuración de los flujos de media durante una sesión. Además permite operar tanto en esquema centralizado como distribuido.

SIPPING ha definido nuevas entidades funcionales cuya descripción se puede ver en la Tabla 1 y Figura 1. Estas son: FOCUS, Conference Notifications Service, Conference Policy Server y Mixer. SIPPING está abierto a cualquier realización física del sistema, lo cual da cierta libertad a la hora emplearlo y permite de esta forma explotar sus funcionalidades en un esquema P2P.

Entidad	Descripción
FOCUS	Es la entidad controladora central que recibe todas las solicitudes de INVITE a la conferencia. Además mantiene las relaciones de señalización SIP con cada participante y toma la responsabilidad de enviar la media a todos los participantes.
Conference Notification Service	Esta entidad mantiene un servicio de Suscripción/Notificación para la modificación de las primitivas de la conferencia.
Conference Policy Server	Cada conferencia es gobernada por ciertas reglas las cuales son llevadas a cabo por esta entidad.
Mixer	En esta entidad es en donde se realiza la transmisión y distribución de la media. Recibe un set de streams, los combina de una manera específica y el resultado lo redistribuye entre los participantes.

Tabla.1 Entidades en SIPPING Framework

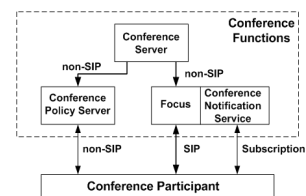


Fig. 1 SIPPING Conference Framework

IV. PROTOTIPO: SISTEMA DE CONFERENCIA P2P

Esta sección explica cómo el sistema propuesto permite crear conferencias para gran audiencia a partir de la aplicación de técnicas P2P para distribuir el flujo de la conferencia entre todos los participantes, empleando SIP como protocolo de señalización y mensajería estándar.

A. Roles de los peers

El sistema distingue entre dos tipos de peers cuando uno se une al sistema: el *líder* y los *participantes*. El líder es el peer encargado de crear la conferencia, es decir, el encargado de crear la sala virtual de conferencia y anunciar al resto de usuarios

el evento. Los peers que reciben el anuncio de la conferencia se convierten en participantes cuando se unen a la sala virtual. Una vez que la conferencia ha comenzado, un participante puede adoptar dos comportamientos: participar activamente en la conferencia o participar pasivamente en la conferencia. Esto se traduce a dos tipos de participantes: el primero se denomina *speaker*, el cual dispone de capacidad para transmitir un flujo de video y audio; mientras que el segundo se denomina *listener*, el cual tan sólo desea estar presente en la conferencia y no toma parte activamente en ésta.

El sistema de conferencia propuesto permite un número limitado de speakers, en concreto se ha definido un límite cuatro speakers simultáneos por conferencia. Cabe destacar que este número no debe ser demasiado elevado, pues en caso contrario la conferencia no sería útil.

B. Arquitectura

En la Figura 2 se muestra la arquitectura basada en SIP implementada por cada peer presente en el sistema. Esta arquitectura permite a cualquier peer en el sistema convertirse en líder, listener o speaker.

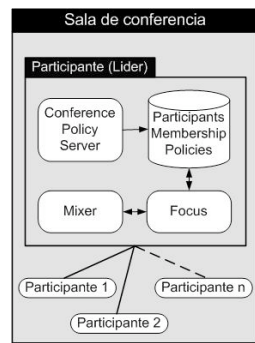


Fig. 2. Arquitectura SIP de un participante

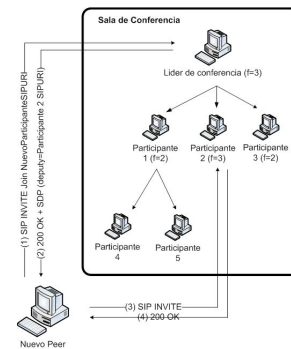


Fig. 3. Unión a una conferencia

C. Construcción de la red de distribución

El flujo multimedia de la conferencia se distribuye a lo largo de toda la audiencia a partir de la construcción de un árbol multicast ALM con origen en el nodo líder. El árbol se construye a partir de la información obtenida de los peers que se unen al sistema por primera vez, los cuales especifican el número de nodos hijo que soportan a partir del cálculo de su fanout (f), obtenido a partir de la división del ancho de banda de conexión del peer entre la tasa del flujo multimedia de la conferencia.

Un nuevo peer que desea unirse a una sala de conferencia envía un mensaje SIP INVITE Join indicando cuál es su fanout en el SDP (Session Description Protocol) adjunto. El líder busca en su estructura cuál es el primer peer *candidato* (deputy) en el sistema con una conexión disponible. Una vez obtenido el peer candidato, el líder le devuelve la SIP URI de éste al nuevo peer empleando el SDP adjunto al mensaje de confirmación 200 OK SIP. A continuación, el nuevo peer envía un mensaje SIP INVITE hacia el peer candidato con el fin de comenzar la conferencia (ver Figura 3).

D. Composición de Speakers

El líder es capaz de mezclar los diversos flujos generados por los speakers de la conferencia. Para iniciar este proceso, los participantes envían mensajes SIP REFER con la finalidad de añadir su flujo a la conferencia. Si el máximo número de speakers simultáneos es alcanzado, el participante candidato a ser speaker pasa a la cola de espera hasta que su turno llega. Una vez que el speaker puede participar en la conferencia éste envía su flujo de video al líder, el cuál mezcla (ver Figura 4) los flujos generados por los speakers (S_1 , S_2 , S_3 y S_4) y crea un único flujo (S) que será distribuido (ver Figura 5) a toda la audiencia.

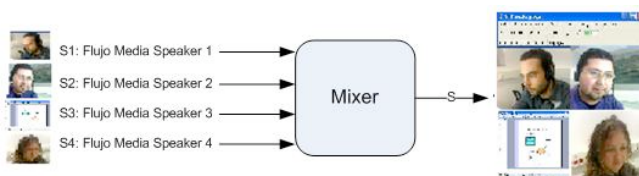


Fig. 4. Mezcla de 4 speakers (S_1 , S_2 , S_3 y S_4) en un único flujo S

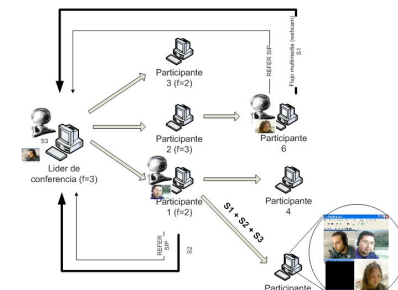


Fig. 5. Distribución del flujo compuesto

E. Implementación y Banco de Pruebas

Se ha desarrollado un prototipo con el fin de evaluar la distribución de un único flujo de media compuesto. La función de mezcla se ha realizado empleando VideoLan Client (VLC) 0.8.5, el cual permite crear un mosaico con diferentes flujos de audio/video. Básicamente, se ha construido un mosaico mezclando cuatro flujos de video obtenidos de unas cámaras

conectadas a unos PCs y se ha distribuido el flujo resultante a través de la red de participantes empleando un único flujo MPEG-TS. El escenario sobre el que se han realizado las pruebas consta de seis máquinas conectadas a un switch Fast Ethernet, en el que tres actúan como speakers y el resto tan sólo son listeners. La distribución del flujo a nivel lógico se puede representar como un árbol binario en el que cada nodo es padre de dos hijos. En la Tabla 2. se refleja la carga de CPU y consumo de memoria RAM que sufre el nodo líder en el proceso de mixing. Se ha probado con diferentes flujos de distintas calidades procedentes de fichero. Los resultados reflejan que el coste computacional que tiene trabajar con flujos de alta tasa (25Mbps, High Definition) es muy alto para un PC de sobremesa (Pentium 4 a 3,2GHz, 1GB RAM). Además, se puede observar que el coste computacional crece cuando hay realizamos una transcodificación.

# videos	Tasa de video	Tasa de salida	Res. Salida	CPU	RAM
4	500 Kbps	2 Mbps	200x200	10 %	12 KB
			400x400	52 %	30 KB
		1 Mbps	200x200	42 %	27 KB
			400x400	55 %	29 KB
4	1,5 Mbps	6 Mbps	200x200	67 %	75 KB
			400x400	77 %	80 KB
		3 Mbps	200x200	68 %	77 KB
			400x400	78 %	80 KB
4	25 Mbps	100 Mbps	200x200	100 %	70 KB
			400x400	99 %	74 KB
		50 Mbps	200x200	100 %	70 KB
			400x400	100 %	70 KB

Tabla. 2. Medición de la carga en el nodo Mixer (líder)

El nodo líder configura la interfaz de transmisión de red para recibir los flujos de los speakers y también para enviar el flujo compuesto resultante a todos sus hijos. La capa de transmisión del prototipo desarrollado fue implementada empleando diferentes lenguajes de programación (Java y C), diferentes implementaciones de sockets (java.io blocking socket, java.nio non-blocking socket, Packet Reflector C based) y diferentes sistemas operativos (Windows XP y Ubuntu Linux). Tras observar los resultados obtenidos, se comprobó que el mejor rendimiento lo proporcionaba la implementación basada en Packet Reflector corriendo bajo el sistema operativo Linux. Se obtuvieron tasas de hasta 25Mbps al enviar cuatro flujos compuestos. Los resultados de las distintas implementaciones se pueden ver representados en la Figura 6. En resumen, la prueba de concepto fue satisfactoria, pero falta depurar aspectos de implementación que hagan que el sistema tenga mayor rendimiento.

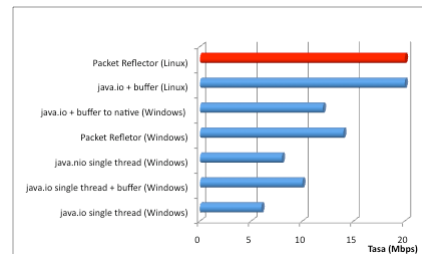


Fig. 6. Tasas máximas conseguidas según implementación

V. CONCLUSIONES

En este trabajo presentamos un prototipo para crear una multiconferencia por medio del uso de técnicas ALM para la distribución de media, permitiendo hacer un uso eficiente de los recursos de red. El sistema además utiliza SIP como protocolo de señalización, lo cual proporciona flexibilidad a la solución, ya que permite el desarrollo de nuevos servicios de valor añadido. Por otro lado, SIP en un esquema de multiconferencia P2P es un desarrollo que ha sido poco explorado y que creemos, dado los resultados de la prueba de concepto efectuada, puede ser una alternativa real en aspectos de gestión y simplicidad. El prototipo desarrollado posee un esquema de mixing de flujos de video en cada peer, lo cual permite la reducción del tráfico generado y facilita la visualización a nivel de terminal, con lo cual es posible ofrecer tanto un servicio de conferencia pura como de multiparty en un ambiente heterogéneo. Entendemos que es preciso continuar este trabajo con el fin de optimizar las técnicas ALM para así brindar más escalabilidad y mejorar los mecanismos de recuperación de fallos y gestión de salas. Finalmente, como trabajo futuro, pretendemos implementar mensajería SIP (SUBSCRIBE) para poder de manera más eficiente detectar fallos y efectuar pruebas de medición de overhead de mensajería.

REFERENCIAS

- [1] E. Alexandros, C. Reha, and S. Ofer. "Multipoint Videoconferencing with Scalable Video Coding". Journal of Zhejiang University SCIENCE A. 2006 7(5).
- [2] WP. Ken, X. Jin, and H. Gary Chan. "Challenges and Approaches in Large-Scale P2P Media Streaming". IEEE Multimedia, Volume 14, Issue 2. Apr 2007.
- [3] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol" IETF RFC 3261, June 2002.
- [4] P2PSIP WG IETF, <http://www.p2psip.org/>
- [5] J. Rosenberg. "A Framework for Conferencing with the Session Initiation Protocol (SIP)" IETF RFC 4353. Feb 2006.
- [6] X. Wu, K. Kishore, and V. Krishnaswamy. "Enhancing Application-Layer Multicast for P2P Conferencing". IEEE Consumer Communications & Networking Conference (CCNC). Las Vegas, NV. Jan 2007.
- [7] M. Dowlatshahi and F. Safei. "Multipoint Interactive Communication for Peer to Peer Environments. IEEE International Conference on Communications (ICC). Istanbul. Jun 2006.
- [8] Ch. Luo, J. Li, and S. Li. "DigiMetro: An Application-Layer Multicast System for Multi-party Video Conferencing". Globecom. Dallas, Texas. Nov 2004.
- [9] H. Horiuchi, N. Wakamiya, and M. Murata. "A Network Construction Method for a Scalable P2P Video Conferencing System". IASTED European Conference Internet and Multimedia Systems and Applications. Chamonix, France. Mar 2007.
- [10] JA. Ríos, A.J. González, A. Oller, J. López, and J. Alcober. "P2P Multipoint Conference System using SIP". HET-NETS 2008, Karlskrona, Sweden. Feb 2008.