

MCU e SFU integrados num sistema de videoconferência em software na nuvem

Valter Roesler
Instituto de Informática, UFRGS –
Universidade Federal do RGS
Porto Alegre, RS, Brasil
roesler@inf.ufrgs.br

Mário Gasparoni Júnior
Instituto de Informática, UFRGS –
Universidade Federal do RGS
Porto Alegre, RS, Brasil
mariogasparoni@gmail.com

Felipe Cecagno
Mconf Tecnologia Ltda
Porto Alegre, RS, Brasil
felipe@mconf.com

Rafael Valle
RNP – Rede Nacional de Ensino e
Pesquisa
Rio de Janeiro, RJ, Brasil
rafael.valle@rnp.br

Ronaldo Husemann
Engenharia Elétrica, UFRGS -
Universidade Federal do RGS
Porto Alegre, RS, Brasil
ronaldo.husemann@ufrgs.br

ABSTRACT

This paper presents the final results of the development of an MCU module that can be integrated with the Mconf web conferencing service. The development took over three years: two to develop the MCU module in software and one to refine and integrate into the Mconf code. This article presents the main characteristics developed to obtain a viable product. Among them, we can mention the integration of architecture, audio and video web / MCU, document sharing, BFCP protocol, moderation interface, recording and registrar SIP. The paper also presents evidences of this successful integration, as well as load testing on the developed system.

KEYWORDS

Mconf, Cloud videoconference, web conferencing.

1 Introdução

Existem dois grandes modelos de comunicação de videoconferência multiponto (mais de dois participantes): MCU (*Multipoint Control Unit*) e SFU (*Switching Forwarding Unit*). Os próximos parágrafos resumem sua diferença.

O modelo MCU utiliza um elemento central que recebe todas as mídias, COMPÕE todas numa mesma imagem, e envia essa imagem para todos participantes. Dessa forma, cada participante envia UM vídeo (da sua câmera) e recebe UM vídeo (composição do vídeo de todos participantes). O modelo SFU utiliza um elemento central que recebe todas as mídias e DISTRIBUI essas mídias para cada participante. Dessa forma, cada participante envia UM vídeo (da sua câmera) e recebe MÚLTIPLOS vídeos (um de cada participante da conferência).

Ambos modelos possuem vantagens e desvantagens. No modo MCU existe uma diminuição de tráfego de rede, entretanto, o elemento central tem maior sobrecarga de processamento e gera um atraso. O modo SFU gera mais tráfego, porém diminui a carga no elemento central. O computador do participante final é mais exigido no modo SFU, pois o mesmo deve decodificar cada um dos vídeos recebidos. Para muitos vídeos simultâneos e máquinas de baixo poder computacional (dispositivos móveis) isso tem se mostrado muitas vezes um problema.

O objetivo desse artigo é apresentar um desenvolvimento para uma evolução do sistema de webconferência Mconf, de modo que o mesmo tenha capacidade de MCU e também de SFU, agregando a inteligência para utilizar da melhor forma um ou outro modelo. Essa evolução também permite a interoperação de forma transparente com *endpoints* de hardware, pois seria feita a adequação para utilizar o protocolo SIP (*Session Initiation Protocol*).

2 Análise da realidade

Um conceito muito buscado por empresas é a *Comunicação Unificada*, ou *Unified Communication*¹, que tem o objetivo de unificar e simplificar o acesso a serviços de comunicação. Ao propor uma solução de comunicação unificada no lugar de webconferência, entrega-se maior valor com a solução, o que permite ao produto atender um público antes não atingido, como também proporcionar uma maior satisfação aos usuários². Como se trata de uma área de alto valor de mercado³, sendo

In: VII Workshop "O Futuro da Videocolaboração" (WCT-Video 2020), Online. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2020.
©2020 SBC – Sociedade Brasileira de Computação.
ISSN: 2596-1683

¹ <https://www.cisco.com/c/en/us/products/unified-communications/index.html>

² <https://www.bluejeans.com/press-releases/bluejeans-receives-2018-unified-communications-product-year-award>

³ <https://edition.cnn.com/2019/04/18/tech/zoom-ipo-video-conference/index.html>

transversal, ou seja, impactando educação, saúde, segurança, entre outros, existem empresas de grande porte trabalhando nisso. Alguns são os seguintes: Zoom, BlueJeans, Vidyo Connect e Microsoft Skype for business. Para analisar os diferentes sistemas, contratou-se uma sala ou se utilizou a versão gratuita dos mesmos.

O produto da Zoom (<https://zoom.us/>) oferece um aplicativo que necessita ser instalado e oferece boa qualidade, integrando *endpoints* com o aplicativo da empresa.

O produto da Bluejeans (<https://www.bluejeans.com/>) também oferece boa qualidade e integra com *endpoints* de hardware, suportando SIP e H.323 TCP. Não suporta SIP / H.323 UDP. Ele já possui interface HTML5, melhorando a experiência do usuário, que não precisa instalar aplicativo.

O Vidyo Connect (<https://www.vidyo.com/>) oferece uma solução instalável ou via HTML5. Foi testada a versão HTML5, com ótima qualidade. Houve dificuldade para integrar com SIP / H.323, apesar do sistema dizer que suporta. Talvez precise de um módulo adicional.

Outras soluções de mercado seguem na linha dos produtos analisados.

As próximas seções apresentam os desenvolvimentos efetuados no sistema Mconf para que o mesmo funcione tanto na modalidade MCU como na modalidade SFU, de forma transparente. Os esforços foram de integrar vários aspectos do sistema, como sua arquitetura, áudio, vídeo, documentos, gravações, entre outros.

3 Arquitetura Integrada

Na arquitetura única implementada através do desenvolvimento de uma API própria, há um sistema central que recebe vídeos (não importando se a origem é de um *endpoint* ou webconferência) e toma uma decisão de como é a melhor forma de enviar esse vídeo para o destinatário (se de forma composta como se fosse um MCU ou de forma individualizada tipo SFU, ou mesmo um misto de ambos para otimizar banda). O fato é que o sistema central, que é o núcleo do GT-Videocolaboração, efetua uma tomada de decisão de forma mais inteligente e otimizada do que existe hoje com os sistemas trabalhando separadamente.

A Figura 1 apresenta a arquitetura unificada final. No lado esquerdo (participantes) pode-se ver a forma que o sistema se adapta a sistemas heterogêneos, ou seja, através de gateways que convertem um determinado protocolo (SIP, WebRTC, etc) no padrão definido pelo grupo (API unificada). Esse modelo permite vários tipos de expansão. Por exemplo, se for necessário integrar com RTSP, basta criar um gateway RTSP. O mesmo vale para protocolos novos que ainda não foram criados.

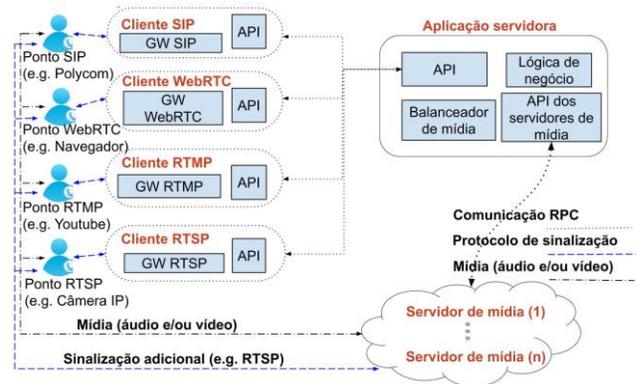


Figura 1: Arquitetura da solução desenvolvida

O lado direito superior (*Aplicação servidora*) mostra o núcleo principal do sistema, onde as decisões são tomadas. Ali existem os blocos de *API* (interface com a API unificada), *Lógica de negócio* (inteligência do sistema), *Balanceador* (permite escalabilidade quando uma máquina virtual atinge seu limite), e *Interface com os servidores de mídia* (que lidam com o áudio e vídeo de uma determinada sala de conferência).

4 Áudio, vídeo e documentos de conferência web/MCU integrados

O sistema permite, de forma transparente, o ingresso de participantes via *endpoint* de hardware (protocolo SIP) e também via navegador web (no computador, tablet ou dispositivo móvel). Essa integração pode ser vista na Figura 2 (para *endpoints*) e na Figura 3 (para web). Observa-se que não tem como saber qual vídeo é de *endpoint*, pois aparece da mesma forma que os outros. Vários *endpoints* podem se conectar simultaneamente (nesse caso o sistema faz o papel de MCU), e vários usuários via web também podem se conectar simultaneamente, como já é usual no Mconf.



Figura 2: visão da integração no *endpoint* (QDX 6000).

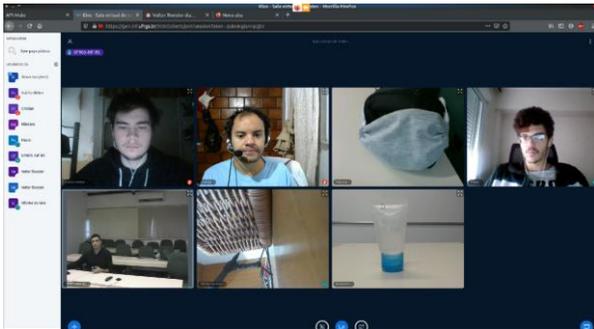


Figura 3: visão da integração no navegador web.

A Figura 4 mostra o compartilhamento de tela visto pelo usuário do endpoint Polycom QDX 6000. Esse compartilhamento está sendo transmitido via protocolo BFCP (*Binary Floor Control Protocol*), e é visualizado de forma transparente pelos participantes no navegador Web.



Figura 4: visão da integração no endpoint (protocolo BFCP).

5 Interface de moderação

A Figura 5 apresenta a interface de moderação para controle dos endpoints. Essa interface é acessada pelo moderador da sala no navegador Web, bastando clicar um botão chamado "Configurações MCU". A interface permite o seguinte:

- Mudança de layout:** no menu é possível escolher entre três tipos de layout: 1) padrão: tamanho de vídeo igual para todos; 2) Destacado: um usuário maior em relação aos outros; 3) Tela cheia: mostra somente o usuário definido no Floor.
- Floor:** define qual usuário será o "floor". Esse usuário é o escolhido para ficar em destaque (no layout 2) ou em tela cheia (no layout 3).
- Legendas:** habilita ou desabilita legendas para os usuários de endpoint.

Configurações MCU para Endpoints

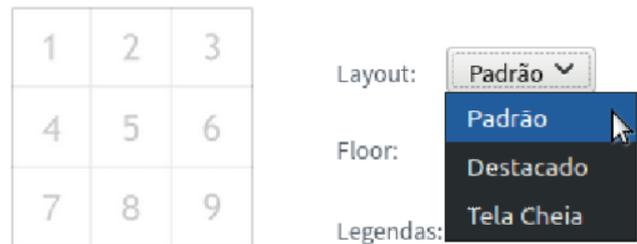


Figura 5: interface de moderação.

6 Sistema de gravação

O sistema de gravação do Mconf agora integra também o ambiente SIP, gravando o seguinte:

- Todos fluxos de áudio e vídeo, independente se o participante está utilizando navegador em desktop, dispositivo móvel, hardware ou software SIP.
- Todos os compartilhamentos, independente se foram originados na Web (documento ou tela) ou se via SIP (protocolo BFCP).
- Os itens que já eram gravados anteriormente pelo Mconf Web, como Chat e Bloco de notas.

7 Registrar SIP

O "Registrar SIP" é responsável por disponibilizar um número único para que endpoints façam chamadas SIP e ingressem em uma sala de webconferência, mesmo sem saber qual a URL dessa sala.

Para que isso seja possível, os servidores desta infraestrutura distribuída devem ser autorizados a se comunicar com o "Registrar SIP" (por usuário e senha ou autorização por IP). Quando uma sala de webconferência é aberta no sistema, o servidor "avisa" o "Registrar SIP" que a sala foi aberta, e que ele é o servidor responsável por aquela sala. Assim, quando o "Registrar SIP" receber uma chamada de endpoint para o número definido para aquela sala de webconferência, ele sabe para qual servidor direcionar a chamada. Quando a sala for fechada, o servidor novamente "avisa" o "Registrar SIP" que aquela sala foi fechada.

O "Registrar SIP" funciona como um proxy, ou seja, ele atua de maneira ativa na conexão. Quando uma chamada SIP conecta em uma sala, a comunicação da sinalização flui através da seguinte configuração:

Endpoint SIP <=> Registrar SIP <=> servidor de videoconferência
Caso a sala chamada pelo endpoint SIP não esteja aberta, o "Registrar SIP" tem a capacidade de responder 404 NOT FOUND. Além de centralizar a comunicação SIP para chamadas de entrada, o "Registrar SIP" funciona como ponto único também para chamadas de saída. Ou seja, caso se deseje integrar com uma central telefônica ou com um serviço de VoIP como o

Fone@RNP, é necessário liberar apenas o endereço do “Registrar SIP”, e não de todos os servidores de videoconferência que fazem parte da infraestrutura.

Um último recurso possível com o “Registrar SIP” é a criação de contas de usuário, que poderiam realizar e receber chamadas de salas abertas na infraestrutura ou usuários externos (embora no contexto da RNP, as contas de usuário idealmente residiriam apenas no Fone@RNP).

A solução para “Registrar SIP” utilizada é o Kamailio (<https://www.kamailio.org/w/>), utilizada principalmente pela sua robustez e flexibilidade de configuração e operação..

8 Testes de carga

Trabalhando com uma VM de 8 núcleos da Digital Ocean, obteve-se os seguintes resultados:

a) **Um usuário Web e aumenta-se usuários SIP:** esse caso representa o sistema funcionando como MCU (*Multipoint Control Unit*), e suportou o máximo de 30 participantes com vídeo HD 720p. Nessas condições a CPU da VM atingiu 70% e o teste foi interrompido para ter margem de segurança.

b) **Um usuário SIP e aumenta-se usuários Web:** esse caso representa uma reunião via conferência Web onde um dos participantes é um *endpoint*. O sistema suportou 140 participantes em navegador Web e a CPU atingiu 50%. O teste foi interrompido para ter margem de segurança.

9 Visão de Futuro na área

O sistema hoje escala horizontalmente, ou seja, se aumenta a capacidade total do sistema aumentando o número de servidores na infraestrutura. Cada servidor é tipicamente uma máquina virtual, e nele é executado o sistema completo.

Assim, o Datacenter deve se adaptar ao PICO de uso, mantendo dezenas de VMs desnecessárias na madrugada (por exemplo), gerando custo para a empresa. Supondo uma demanda de pico de ~100VMs buscando suportar uma carga máxima de 10.000 usuários simultâneos, tem-se um custo fixo alto, que pode ser muito reduzido com a elasticidade de VMs.

No futuro, espera-se uma evolução para que a infraestrutura na nuvem seja alocada somente durante a utilização, ou seja, se adaptando a poucas salas de videoconferência simultâneas até milhares de videoconferências simultâneas. Assim, à medida que o número de conferências cresce, o número de VMs também cresce, atendendo à demanda. Quando o número de conferências diminui, o número de VMs também diminui. Isso faz com que a infraestrutura do Datacenter só seja utilizada quando necessária, otimizando seu uso, reduzindo custos e facilitando a manutenção.

Em uma arquitetura de microsserviços, como é o Mconf, é possível que componentes sejam escalados de maneira assimétrica, ou seja, que componentes sejam replicados conforme a necessidade. Para o melhor aproveitamento dos

microsserviços, é natural que seja utilizado algum mecanismo de orquestração de containers, como por exemplo o Kubernetes.

Além de proporcionar maior dinamicidade na operação da solução, o Kubernetes poderia ser utilizado para orquestrar o aumento e a redução do cluster de maneira elástica – ao atingir determinado gatilho, novos servidores são incluídos no cluster, e após a utilização, os mesmos são descartados, trazendo assim uma otimização significativa nos recursos.

10 Visão de Futuro em videoconferência

A Pandemia do Coronavírus gerou uma grande mudança no mundo atual. Pessoas morreram, empresas faliram, a economia global está em recessão, entre outros acontecimentos ruins. Por outro lado, com o isolamento social, milhares de reuniões e eventos que antes aconteciam de forma presencial passaram a acontecer por videoconferência, provocando um enorme crescimento nessa área.

Um fato inesperado para muitos foi o crescimento no número de pessoas participando de eventos. Eventos presenciais que aconteciam com 100 pessoas se transformaram em eventos virtuais para 1000 pessoas, fora o fato das palestras serem gravadas e terem milhares de acessos posteriormente.

Num futuro retorno ao presencial, o aprendizado sofrido pela população mundial da efetividade das videoconferências não voltará ao que era. Os eventos do futuro serão naturalmente virtuais ou híbridos, misturando um misto de presencial e remoto, permitindo a facilidade de conversar nos corredores e intervalos com as pessoas, mas também permitindo o acesso para quem está distante e não tem as condições (seja de tempo ou dinheiro) para viajar e estar presencialmente no evento.

AGRADECIMENTOS

Este trabalho foi executado na UFRGS e Mconf Tecnologia, sendo financiado pela RNP e pela Mconf Tecnologia Ltda.

REFERÊNCIAS

- [1] ROESLER, Valter; COELHO, L. ; LONGONI, G. ; MARINS, A. ; DARONCO, L. ; CIUFFO, L. ; DUARTE, R. . Multipresence: towards videoconference and collaboration in multi-use environments. In: TNC16 Networking Conference, 2016, Praga. 32th TNC Networking Conference, 2016.
- [2] ROESLER, V.; GASPARONI JUNIOR, M. ; CECAGNO, F. ; VALLE, R. ; HUSEMANN, R. . Sistema de Videocolaboração interoperável entre webconferência e endpoints de videoconferência. In: Webmedia: Simpósio Brasileiro de Sistemas Multimídia e Web, 2019, Rio de Janeiro. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (Webmedia), 2019.