



Pós-Graduação em Ciência da Computação

**“Estudo e Avaliação de Redes
Convergentes de Voz”**

Por

Fabício Barros Cabral

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE, AGOSTO2009



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FABRÍCIO BARROS CABRAL

“ESTUDO E AVALIAÇÃO DE REDES CONVERGENTES DE VOZ”

ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA COMPUTAÇÃO.

ORIENTADOR: DR. DJAMEL FAWZI HADJ SADOK
CO- ORIENTADOR: DR. CARLOS ALBERTO KAMIENSKI

RECIFE, AGOSTO/2009

Cabral, Fabrício Barros

Estudo e avaliação de redes convergentes de voz / Fabrício Barros Cabral. - Recife: O Autor, 2009.

xii, 86 folhas : il., fig., tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. Cln. Ciência da computação, 2009.

Inclui bibliografia.

1. Ciência da computação. 2. Redes de computadores. 3. Qualidade de voz. I. Título.

004

CDD (22. ed.)

MEI2010 – 013

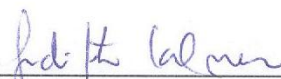
Dissertação de Mestrado apresentada por **Fabício Barros Cabral** Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Estudo e Avaliação de Redes Convergentes de Voz**”, orientado pelo **Prof. Djamel Fawzi Hadj Sadok** e co-orientada pelo **Prof. Carlos Alberto Kamienski** e aprovada pela Banca Examinadora formada pelos professores:



Prof. Nelson Souto Rosa
Centro de Informática / UFPE



Prof. Eduardo James Pereira Souto
Centro de Informática/UFAM



Prof. Judith Kelner
Centro de Informática / UFPE

Visto e permitida a impressão.
Recife, 08 de setembro de 2009.



Prof. Nelson Souto Rosa
Vice-Coordenador da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federal de Pernambuco.

Aos meus pais Rosário e
Lourival.

Agradecimentos

Agradeço primeiramente a Deus, por me conceder a dádiva da vida, e por me colocar entre tanta gente querida as quais amo muito.

Aos meus pais, Rosário e Lourival, que sempre me apoiaram e me apoiarão nos momentos difíceis e alegres e por me mostrarem que as pessoas mais importantes em nossa vida são os nossos pais.

As minhas avós (*in memorian*) Adélia Daniel de Barros e Luzia Batista Cabral, por sempre insistirem em ter um neto “doutor”.

Ao meu orientador Djamel Sadok pela oportunidade de participar do programa de Mestrado da UFPE.

Ao meu co-orientador Carlos Alberto Kamienski, por todas as orientações e conselhos.

A professora Judith Kelner por suas orientações, paciência e disciplina.

Ao meu estimado amigo Arthur Callado, pelas discussões técnicas, se dispor a revisar este trabalho, pela paciência, pelo apoio moral e pela amizade.

Ao GPRT (Grupo de Pesquisas em Redes e Telecomunicações) por disponibilizar toda a infraestrutura e condições para a realização deste trabalho.

Aos meus amigos Carlos Julian e Jalmaratan Macedo, por ouvirem minhas preocupações, amizade e pelas partidas de *Bombberman*.

A todos os colegas de trabalho do GPRT, Ernani Azevedo, Cyrus Dias, Ramide Dantas, Joseane Fidalgo, João Marcelo, Gabriel Fernandes, Ana Cristina, Ricardo Schmidt, Reinaldo Gomes, Glauco Gonçalves, Auristela, Luís Eduardo, Patrícia Endo por toda a ajuda, conversas e brincadeiras.

A equipe do GPRT, Rover de Castro, Richardson Lima, Rodrigo Germano, Nadia Maria, Manuela Melo e Vanderoque da Silva, por todo o suporte e carinho.

Aos amigos do laboratório C1, Érica Souza, Hermano Brandão, Carlos Romero (*in memorian*), Cleber Fonseca, Carla Tanure, Luíz Henrique, Breno Jacinto, Rafael Amorim, Eduardo Miyagawa e Ermerson Carneiro.

E o meu mais sincero obrigado a todos que, de maneira direta ou indiretamente, contribuíram para a realização deste trabalho.

Sumário

Abreviações e Acrônimos	x
Resumo	xi
Abstract	xii
1 Introdução	1
1.1 Apresentação	1
1.2 Motivação.....	3
1.3 Objetivos e Contribuições.....	4
1.4 Escopo deste Trabalho	5
1.5 Estrutura deste Trabalho	5
2 Fundamentação Teórica e Trabalhos Relacionados	6
2.1 A Rede de Telefonia Pública.....	7
2.2 Voz sobre IP.....	8
2.2.1 Sinalização	9
2.2.2 Transmissão da Voz	9
2.3 Serviços Agregados.....	13
2.3.1 Convergência com a PSTN	13
2.3.2 Sessão de Conferência VoIP	14
2.4 Avaliação da Qualidade de Voz.....	16
2.4.1 Qualidade	16
2.4.2 Contexto.....	18
2.4.3 Métodos de Avaliação	20
2.5 Aplicações de Voz sobre IP	23
2.5.1 <i>Skype</i>	23
2.5.2 <i>Yahoo! Messenger</i>	24
2.5.3 <i>GTalk</i>	24
2.5.4 <i>Windows Live Messenger</i>	24
2.6 Trabalhos Relacionados	25
3 Metodologia	27
3.1 Terminologia	28
3.2 Classificação e Descrição dos Cenários de Utilização	29
3.3 Carga de Trabalho	33
3.4 Métricas	35
3.5 Ambiente de Realização dos Experimentos	37
3.6 Experimentos Realizados	39
3.6.1 Topologia para a Mixagem e Distribuição do Áudio.....	41
3.6.2 Experimentos de Calibração	41
3.6.3 Experimentos de Avaliação.....	45
3.7 Medindo Qualidade de Voz no Contexto de Conversação.....	46
3.8 Medindo Jitter	46
3.9 Medindo Perda de Pacotes	47
4 Avaliação de Desempenho	48

4.1	Planejamento dos Experimentos.....	48
4.2	Elaboração da Carga de Trabalho	48
4.3	Tempo de Realização dos Experimentos de Calibração e Avaliação.....	49
4.4	Configuração do Ambiente Controlado.....	49
4.5	Experimentos de Calibração	49
4.5.1	Experimentos de Calibração para o Serviço de Conferência.....	49
4.5.2	Experimentos de Calibração para o Serviço de Convergência.....	52
4.6	Topologia para a Mixagem e Distribuição do Áudio	53
4.7	Experimentos de Avaliação.....	54
4.7.1	Avaliação dos Cenários do Serviço de Conferência.....	54
4.7.2	Avaliação dos Cenários de Correlação entre Métricas.....	63
4.7.3	Avaliação dos Cenários do Serviço de Convergência	74
4.8	Análise dos Experimentos Realizados.....	79
5	Conclusões e Trabalhos Futuros	80
5.1	Considerações Finais.....	80
5.2	Principais Contribuições	81
5.3	Trabalhos Futuros	82
6	Referências	83

Lista de Figuras

Figura 2.1 Composição do sistema telefônico.....	8
Figura 2.2 Distribuição do áudio entre diversas topologias durante uma sessão de conferência.....	16
Figura 2.3 Estados e eventos de uma conversação, observados por um participante.....	20
Figura 3.1 (a) Cenário de classe 1 – Ponto a ponto, homogêneo, sem rede de trânsito. (b) Cenário de classe 2 – Multiponto, homogêneo e sem rede de trânsito.....	30
Figura 3.2 (a) Cenário de classe 3 – Ponto a ponto, heterogêneo e sem rede de trânsito. (b) Cenário de classe 4 – Multiponto, heterogêneo e sem rede de trânsito.....	31
Figura 3.3 (a) Cenário de classe 5 – Ponto a ponto, heterogêneo e com a Internet como rede de trânsito. (b) Cenário de classe 6 – Multiponto, heterogêneo e com a Internet como rede de trânsito.....	32
Figura 3.4 (a) Cenário de classe 7 – Ponto a ponto, heterogêneo e com a PSTN como rede de trânsito. (b) Cenário de classe 8 – Multiponto, heterogêneo e com a PSTN como rede de trânsito.....	33
Figura 3.5 (a) Elaboração do áudio para uma sessão de conferência para três participantes. (b) Elaboração do áudio para uma sessão de conferência para quatro participantes.....	35
Figura 3.6 Ambientes de realização dos experimentos: ambiente controlado e o não-controlado.....	37
Figura 3.7 Posicionamento do cabo de áudio para os participantes emissores de áudio.....	38
Figura 3.8 Caminho percorrido pela chamada durante a avaliação do serviço de convergência.....	44
Figura 4.1 (a) Taxa transmitida e (b) PESQ MOS obtidos através da análise da variação da capacidade (Cenário 1.1).....	50
Figura 4.2 Taxa transmitida durante o experimento de calibração do atraso (Cenário 1.2).....	50
Figura 4.3 Taxa transmitida durante o experimento de calibração da taxa de perda de pacotes (Cenário 1.3).....	51
Figura 4.4 Taxa transmitida durante o experimento de calibração do <i>jitter</i> (Cenário 1.2).....	52
Figura 4.5 (a) Taxa transmitida e (b) PESQ MOS obtidos através do experimento de calibração 8.1, referentes ao serviço de convergência.....	53
Figura 4.6 Qualidade de voz para os Cenários 1.5, 2.1 e 2.2, com a capacidade configurada em 100kbps.....	56
Figura 4.7 Taxa transmitida entre os participantes do serviço de conferência, quando varia a capacidade configurada entre os participantes e o número de participantes (Cenários 1.5, 2.1 e 2.2).....	57
Figura 4.8 Qualidade de voz entre os participantes do serviço de conferência, quando varia a capacidade configurada entre os participantes e o número de participantes (Cenários 1.5, 2.1 e 2.2).....	58
Figura 4.9 Taxa transmitida entre os participantes do serviço de conferência, quando varia o atraso entre os participantes e o número de participantes (Cenários 1.6, 2.3 e 2.4).....	59
Figura 4.10 Qualidade de voz entre os participantes do serviço de conferência, quando varia a capacidade configurada entre os participantes e o número de participantes (Cenários 1.6, 2.3 e 2.4).....	60
Figura 4.11 Taxa transmitida entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes entre os participantes e o número de participantes (Cenários 1.7, 2.5 e 2.6).....	61
Figura 4.12 Qualidade de voz entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes entre os participantes e o número de participantes (Cenários 1.7, 2.5 e 2.6).....	62
Figura 4.13 Taxa transmitida entre os participantes do serviço de conferência, quando varia o <i>DelaySigma (jitter)</i> entre os participantes e o número de participantes (Cenários 1.8, 2.7 e 2.8).....	63

Figura 4.14 Qualidade de voz entre os participantes do serviço de conferência, quando varia <i>DelaySigma (jitter)</i> entre os participantes e o número de participantes (Cenários 1.8, 2.7 e 2.8).	63
Figura 4.15 Taxa transmitida entre os participantes do serviço de conferência, quando varia o atraso de apenas um participante e o número de participantes (Cenários 2.9 e 2.10).	66
Figura 4.16 Qualidade de voz entre os participantes do serviço de conferência, quando se varia o atraso de apenas um participante e o número de participantes (Cenários 2.9 e 2.10).	67
Figura 4.17 Taxa transmitida entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de apenas um participante e o número de participantes (Cenários 2.11 e 2.12).	68
Figura 4.18 Qualidade de voz entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de apenas um participante e o número de participantes (Cenários 2.11 e 2.12).	69
Figura 4.19 Taxa transmitida entre os participantes do serviço de conferência, quando varia o <i>DelaySigma (jitter)</i> de apenas um participante e o número de participantes (Cenários 2.13 e 2.14).	70
Figura 4.20 Qualidade de voz entre os participantes do serviço de conferência, quando varia o <i>DelaySigma (jitter)</i> de apenas um participante e o número de participantes (Cenários 2.13 e 2.14).	71
Figura 4.21 Taxa transmitida entre os participantes do serviço de conferência, quando se varia o atraso de um participante, a taxa de perda de pacotes de outro, e o número de participantes (Cenários 2.15 e 2.16).	72
Figura 4.22 Qualidade de voz entre os participantes do serviço de conferência, quando varia o atraso de um participante, a taxa de perda de pacotes de outro, e o número de participantes (Cenários 2.15 e 2.16).	72
Figura 4.23 Taxa transmitida entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de um participante, o <i>DelaySigma (jitter)</i> de outro, e o número de participantes (Cenários 2.17 e 2.18).	73
Figura 4.24 Qualidade de voz entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de um participante, o <i>DelaySigma (jitter)</i> de outro, e o número de participantes (Cenários 2.17 e 2.18).	74
Figura 4.25 Taxa transmitida entre dois participantes do serviço de convergência (Cenário 8.1).	75
Figura 4.26 Qualidade de voz entre dois participantes do serviço de convergência (Cenário 8.1).	76
Figura 4.27 Taxa transmitida entre três participantes do serviço de convergência (Cenário 8.2).	77
Figura 4.28 Qualidade de voz entre dois participantes do serviço de convergência (Cenário 8.2).	77
Figura 4.29 Qualidade de voz entre três participantes do serviço de convergência (Cenário 8.3).	78

Lista de Tabelas

Tabela 2.1 Relação entre a qualidade de voz, MOS, PESQ MOS e o fator R.	18
Tabela 2.2 Classificação dos métodos objetivos para avaliação da qualidade da voz.	23
Tabela 3.1 Classificação dos cenários de utilização possíveis e suas características.	30
Tabela 3.2 Métricas.	36
Tabela 3.3 Ferramentas utilizadas no ambiente controlado de execução dos experimentos.	39
Tabela 3.4 Fatores utilizados nos experimentos.	40
Tabela 3.5 Cenários relativos aos experimentos de calibração para o serviço de conferência.	43
Tabela 3.6 Cenário relativo ao experimento de calibração para o serviço de convergência.	45
Tabela 4.1 Estimativa das métricas de rede entre dois participantes do serviço de convergência através do experimento de calibração.	53
Tabela 4.2 Cenários relativos aos experimentos de avaliação para o serviço de conferência com os parâmetros de rede variando igualmente entre os participantes.	55
Tabela 4.3 Cenários relativos aos experimentos de avaliação para o serviço de conferência com os parâmetros de rede variando diferentemente entre os participantes.	65
Tabela 4.4 Cenários relativos aos experimentos de avaliação para o serviço de convergência com os parâmetros de rede variando diferentemente entre os participantes.	75
Tabela 4.5 Estimativa das métricas de rede entre dois participantes do serviço de convergência (Cenário 8.1).	76
Tabela 4.6 Estimativa das métricas de rede entre três participantes do serviço de convergência (Cenário 8.2).	78
Tabela 4.7 Estimativa das métricas de rede entre três participantes do serviço de convergência (Cenário 8.3).	78

Abreviações e Acrônimos

codec	Compressor-Decompressor ou Compressor-Descompressor
GIPS	Global IP Sound
IETF	Internet Engineering Task Force
ITU	International Telecommunication Union
MOS	Mean Opinion Score
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
IP	Internet Protocol
VoIP	Voice over Internet Protocol
HTTP	Hypertext Transfer Protocol
WWW	World Wide Web
P2P	Peer-to-Peer
PSTN	Public Switched Telephone Network
QoS	Quality of Service
PESQ	Perceptual Evaluation of Speech Quality
RTP	Real-time Transport Protocol
RTCP	RTP Control Protocol
SIP	Session Initiation Protocol
SDP	Session Description Protocol
SRTP	Secure Real-time Transport Protocol
SVOPC	Sinusoidal Voice Over Packet Coder
SILK	Super Wideband Audio Codec
PABX	Private Automatic Branch Exchange

Resumo

A telefonia convencional é uma tecnologia que permite que pessoas possam, através da fala, se comunicar à distância. Esta tecnologia é basicamente composta pela rede de telefonia pública (PSTN) e por diversos equipamentos, dentre eles o telefone. Com a popularização da Internet, vislumbrou-se a capacidade de utilizar esta rede para efetuar chamadas telefônicas, de maneira semelhante ao sistema de telefonia convencional. A partir de então foi desenvolvida a tecnologia de voz sobre IP (VoIP), que adota um conjunto de técnicas e protocolos permitindo que a voz humana possa se propagar através de uma rede de comutação de dados, como a Internet. Esta tecnologia apresenta diversas vantagens, como a redução dos custos, flexibilidade e o desenvolvimento e agregação de novos serviços. Dentre estes novos serviços, este trabalho foca os serviços de conferência e de convergência entre redes. O serviço de conferência é a capacidade de estabelecer uma chamada entre três ou mais pessoas simultaneamente. O serviço de convergência entre redes é a capacidade de se efetuar uma chamada entre um computador conectado à Internet e um telefone que se encontre na PSTN ou vice-versa. No entanto, por dependerem da Internet, estes serviços estão sujeitos a uma baixa qualidade da voz, devido à variabilidade das condições de rede. Sob este contexto, este trabalho propõe uma metodologia para analisar o comportamento e o desempenho destes serviços, quando submetidas a condições variadas da rede. A partir de um ambiente controlado de rede, foram simulados diversos cenários de utilização destes serviços e configurados diferentes valores para os parâmetros críticos, como o número de participantes, capacidade do enlace, atraso, perda de pacotes e variação do atraso (*jitter*). Este estudo adotou a qualidade do áudio recebido como principal métrica de desempenho cujo cômputo foi efetuado pelo algoritmo PESQ MOS.

Palavras-chave: Voz sobre IP, Avaliação de Desempenho, Conferência, Convergência, Qualidade de Voz, *Skype*.

Abstract

The conventional telephony is a technology that allows people to use their own voices to communicate to each other over long distances. This technology is basically composed by a public switched telephone network (PSTN) and several equipments such as the telephone. As the Internet grew popular, some researchers realized the possibility of using this network to make telephone calls similarly to the telephony system. Thus, the voice over IP (VoIP) technology was developed with a set of techniques and protocols to propagate the human voice through a packet switched network, such as the Internet. This technology presents several advantages such as cost reduction, flexibility and the development and aggregation of new services. Among these new services, this work focuses the conference and the convergence between networks. The conference service is the possibility of performing a telephone call among three or more people simultaneously. The convergence among networks service is the possibility of performing a telephone call between a computer connected to the Internet and a telephone located in the PSTN or vice-versa. However, because of their dependence of the Internet, these services may present a poor voice quality due to the variability of network conditions.

This work proposes a methodology to analyze the behavior and performance of these services when subordinate to different network conditions. Using a controlled network environment, several scenarios of utilization of these services were simulated and different values were configured for critical parameters, such as the number of participants, link capacity, delay, packet loss rate and jitter. This study adopted the received audio quality as the main performance metric, which was computed by the PESQ MOS algorithm.

Keywords: Voice over IP, Performance Evaluation, Conference, Convergence, Voice Quality, Skype.

1 Introdução

“Comece pelo começo, siga até chegar ao fim e então, pare.”

Alice no País das Maravilhas, Lewis Carroll

1.1 Apresentação

A telefonia convencional é uma tecnologia que permite que pessoas possam, através da fala, se comunicar a longas distâncias. Esta tecnologia é basicamente composta por vários equipamentos e a rede de telefonia pública (PSTN – *Public Switched Telephone Network*). Dentre os equipamentos, o mais conhecido é o telefone. Mesmo hoje, mais de um século após o seu surgimento, seu princípio básico de funcionamento ainda é o mesmo. Nele encontram-se dois dispositivos: um microfone e um autofalante. O microfone converte as ondas sonoras, emitidas pelo transmissor, em pulsos elétricos que se propagam através de um par de fios metálicos até o receptor, onde é feito o processo inverso. A PSTN, que corresponde ao par de fios metálicos, é o meio físico que provê a interconexão entre os diversos telefones espalhados ao redor do mundo.

Embora o sistema de telefonia seja tão simples, não significa que não tenha evoluído. Hoje, esta tecnologia é baseada em equipamentos digitais, o que oferece facilidade de programação, redução no tamanho dos equipamentos e o desenvolvimento de novos serviços, tais como: siga-me, chamada em espera, conferência, etc.

Com a chegada da era da informática, surgiu a necessidade de realizar a troca de informações entre computadores localizados geograficamente distantes. Logo, percebeu-se que a PSTN poderia ser usada para interligá-los, formando assim, uma rede para a troca de informações. Esta rede foi a precursora do que hoje é a nossa atual Internet.

Mas não se demorou a perceber que a voz humana nada mais é do que um tipo de informação. Assim, vislumbrou-se a capacidade de utilizar a Internet, uma rede criada originalmente para a troca de dados entre computadores, para propagar a voz humana, de forma equivalente a PSTN. Para atingir este objetivo, foi desenvolvida a tecnologia de voz sobre IP (VoIP), que consiste em um conjunto de técnicas e protocolos que permitem que a voz humana possa se propagar através de uma rede de comutação de dados, como a Internet. A união destas duas redes traria uma

série de vantagens, como a redução dos custos, flexibilidade, o desenvolvimento e agregação de novos serviços [19].

A redução dos custos é obtida através do aproveitamento da infraestrutura de rede já existente e a possibilidade de utilizar gratuitamente outros serviços (conferência, redirecionamento de chamadas, identificador de chamadas, etc.) que normalmente são oferecidos e cobrados pelas empresas de telecomunicações.

A flexibilidade é inerente à tecnologia VoIP, pois esta permite que mais de uma chamada seja realizada na mesma conexão, substituindo o processo da telefonia convencional de instalar uma segunda linha telefônica. Além disso, temos a independência de localização, pois para estabelecer as chamadas necessitamos apenas de uma conexão rápida e estável à Internet.

O desenvolvimento e agregação de novos serviços, como a possibilidade de estabelecer chamadas seguras, através de protocolos padronizados como o *Secure Real-time Transport Protocol* (SRTP) [20] e a integração com outros serviços disponíveis na Internet, incluindo conversação com vídeo, troca de mensagens ou arquivos durante a conversa, conferência de áudio, gerenciamento de endereços e assim por diante.

Mesmo com tantas vantagens e possibilidades, a tecnologia VoIP demorou a proliferar, devido a alguns problemas como a complexidade da utilização e a qualidade do áudio das ligações, quando comparada as ligações efetuadas através da PSTN. Mas esta situação começou a mudar com o surgimento do *Skype* [14]. O *Skype* mudou significativamente a aceitação da tecnologia VoIP, uma vez que ele utiliza uma rede *peer-to-peer* (P2P) para possibilitar a comunicação direta entre os usuários e consequentemente gerar grandes melhorias na qualidade perceptível da ligação. Depois, não tardaram a aparecer novos aplicativos de VoIP concorrentes ao *Skype*, como o *GTalk* [15], *Yahoo! Messenger* [17] e o *Windows Live Messenger* [16].

Com a grande utilização do *Skype*, houve um aumento na utilização nos serviços agregados a este aplicativo. Entre estes, destacamos os serviços de conferência de usuários e o de convergência entre redes.

Os serviços de convergência entre redes, ou simplesmente convergência, se caracterizam por permitir que se efetue uma chamada entre um computador conectado à Internet e um telefone que se encontre na PSTN ou vice-versa. Apesar de haver cobrança neste tipo de serviço, os custos relativos são bem menores quando comparados a uma mesma chamada efetuada exclusivamente pela PSTN, pois a maior parte da chamada é realizada através da Internet e apenas o último trecho do caminho é realizado através da PSTN.

O serviço de conferência entre usuários, ou simplesmente conferência, caracteriza-se por permitir que três ou mais participantes, estando eles conectados diretamente à Internet ou à PSTN (através do serviço de convergência), possam participar simultaneamente de uma mesma chamada.

No *Skype*, os serviços de convergência são chamados de *SkypeIn* (permite efetuar chamadas a partir da PSTN ao *Skype*) e *SkypeOut* (permite efetuar chamadas através do *Skype* à PSTN). O serviço de conferência é um serviço inerente à aplicação, bastando apenas selecionar os participantes que desejam participar da conferência.

Embora existam vários aplicativos VoIP sendo utilizados, restringimos o escopo deste trabalho ao *Skype*, pelo fato deste ser o mais utilizado, analisado pela comunidade científica e de prover os serviços de convergência e de sessões de conferência.

1.2 Motivação

Em 2005, o tráfego internacional de chamadas telefônicas totalizou 272 bilhões de minutos, sendo 97,2% correspondendo ao tráfego da PSTN e VoIP e 2,8% do *Skype*. Em 2008 este tráfego totalizou cerca de 417 bilhões de minutos, sendo 92% (384 bilhões de minutos) correspondente ao tráfego gerado pela PSTN e VoIP e 8% (33 bilhões de minutos) ao *Skype* [37] [21].

Em 2007 o tráfego internacional de voz cresceu 14% e estima-se que em 2008 tenha crescido 12%. No entanto, embora este tráfego esteja aumentando a um ritmo moderado, o tráfego *Skype* apresentou um aumento bem maior. Estimativas apontam que cresceu em 2008, 41%, tornando assim, a *Skype S.A.* a maior companhia de tráfego telefônico internacional do mundo, detendo 8% das chamadas internacionais de voz [21].

Além disso, dados referentes à utilização [34][35] do *Skype* apontam que em 2006 as chamadas *Skype* para *Skype* totalizaram cerca de 28,6 bilhões de minutos e o serviço *SkypeOut* contabilizava cerca de 4,1 bilhões de minutos. Já em 2008, foram 65,5 bilhões de minutos para as chamadas *Skype* para *Skype* (um aumento de 118,5%) e 8,4 bilhões de minutos para o serviço *SkypeOut* (um aumento de 104,9%).

Conforme contextualizado na seção anterior, através dos serviços de convergência e de sessão de conferência, emergiram novos cenários de utilização da tecnologia VoIP por parte dos usuários. Dentre estes cenários, podemos citar:

- Convergência entre dois participantes, através de uma chamada realizada entre um usuário que se encontre na Internet e outro que esteja localizado na PSTN;
- Sessões de conferência entre dois ou mais participantes, localizados na Internet;
- Sessões de conferência entre dois ou mais participantes, localizados na Internet e na PSTN.

No entanto, por dependerem da Internet, estes serviços estão sujeitos à uma baixa qualidade da voz, devido à variabilidade das condições de rede. Portanto, é preciso avaliar e entender como as condições dinâmicas da rede, isto é, a capacidade, atraso, perda e *jitter* (variação do atraso) influenciam na qualidade de voz percebida pelos usuários de uma chamada nestes novos cenários. O entendimento deste comportamento trará uma série de benefícios:

- Os provedores de acesso à Internet podem adequar melhor os seus sistemas e oferecer novos serviços aos seus usuários, tendo como base garantia de qualidade sob estes novos cenários;
- Os usuários podem dimensionar melhor o seu acesso à Internet, de modo a melhorar a qualidade dos serviços oferecidos por esta aplicação;
- Os desenvolvedores de aplicativos VoIP podem adequar ou aperfeiçoar melhor o seu sistema, de modo a obter uma melhor qualidade do áudio e diminuir a utilização dos recursos da rede.

Após uma exaustiva pesquisa, não encontramos uma metodologia ou estudo que aborde estes novos cenários emergentes, bem como uma avaliação de desempenho a cerca destes. Assim, faz-se necessária criar e avaliar uma metodologia sobre como realizar a avaliação de desempenho destes cenários.

1.3 Objetivos e Contribuições

Este trabalho tem como objetivo desenvolver uma metodologia para realizar a avaliação de desempenho de um aplicativo VoIP, incluindo o serviços agregados de convergência e conferência, quando submetidas a condições variadas da rede. Como estudo de caso da metodologia proposta, foi utilizado o aplicativo *Skype*, um aplicativo VoIP P2P gratuito e largamente utilizado sob diversos cenários que podem contemplar dois ou mais usuários simultâneos (conferência) e a integração com a PSTN. Considerando que o *codec* e os seus parâmetros são componentes definidos exclusivamente pela aplicação e não podem ser diretamente manipulados, este trabalho avalia o comportamento desta aplicação quando submetidas a diferentes condições de redes nestes cenários. Por comportamento, entende-se a qualidade de voz e a adaptabilidade. Para atingir tal objetivo, esta dissertação apresenta uma metodologia que compreende um ambiente controlado onde é possível emular o comportamento da Internet para realizar os diversos experimentos. Além disso, a metodologia aponta os meios, ferramentas e técnicas para realizar a avaliação de desempenho. Embora a metodologia aqui apresentada tenha sido aplicada apenas ao *Skype*, ela pode ser perfeitamente aplicada para o estudo e avaliação de outras aplicações VoIP.

Entre outras investigações, esta dissertação avalia como a aplicação se adapta dinamicamente, mudando as características de fluxo de voz nos casos em que a capacidade disponível no caminho entre dois ou mais usuários diminui ou aumenta. Também avalia qual o atraso e a variação do atraso (*jitter*) máximo suportado pela aplicação para que uma conversa seja mantida de forma compreensível, como também investiga quão sensível é a aplicação à perda de pacotes na rede subjacente.

1.4 Escopo deste Trabalho

Como estudo de caso da metodologia proposta, foi utilizado o aplicativo *Skype* versão 3.8.0.139, que utiliza o *codec* SVOPC [4] (*Sinusoidal Voice Over Packet Coder*). Contudo, durante a conclusão deste trabalho, a *Skype Technologies S.A.* havia lançado uma nova versão do *Skype* (4.0.0.226) que agora utiliza um novo *codec*, chamado de SILK [48] (*Super Wideband Audio Codec*), que segundo a própria empresa, oferece uma melhor qualidade de voz e utilização dos recursos da rede com relação às versões anteriores. Assim, não houve tempo hábil de incluir os experimentos relativos à utilização deste novo *codec*, sendo sugerida como trabalho futuro a sua avaliação através da metodologia proposta neste trabalho.

1.5 Estrutura deste Trabalho

Este trabalho está estruturado como se segue: no Capítulo 2 é feita uma fundamentação teórica e o estado da arte das tecnologias e trabalhos relacionados. O Capítulo 3 discute a metodologia adotada, os cenários abordados e como os experimentos foram conduzidos. O Capítulo 4 analisa e discute a avaliação de desempenho, de acordo com os dados coletados através da metodologia proposta. No Capítulo 5 é realizada uma discussão em torno deste trabalho, listamos as principais contribuições e apontamos estudos futuros que podem ser realizados com base neste trabalho.

2 Fundamentação Teórica e Trabalhos Relacionados

“Longo é o caminho ensinado pela teoria, curto e eficaz o do exemplo.”

Lucius Annaeus Seneca

Com a popularização dos computadores, surgiu a necessidade de realizar a troca de informações entre computadores localizados geograficamente distantes. Logo se percebeu que a PSTN poderia ser usada para interligá-los, formando assim uma rede para a troca de informações. Esta rede foi a precursora do que hoje é a nossa atual Internet.

Mas, não se demorou a perceber que a voz humana também nada mais é do que um tipo de informação. Assim, vislumbrou-se a capacidade de utilizar a Internet, uma rede criada originalmente para a troca de dados entre computadores, para propagar a voz humana, de forma similar à PSTN. Para atingir este objetivo, foi desenvolvida uma tecnologia denominada VoIP (voz sobre o protocolo IP), que consiste em um conjunto de técnicas e protocolos que permitem que a voz humana possa se propagar através de uma rede de comutação de dados, como a Internet. A união destas duas redes traria uma série de vantagens, como a redução dos custos, flexibilidade, e o desenvolvimento e agregação de novos serviços [19].

O protocolo IP (*Internet Protocol*) [29] é um protocolo não orientado a conexão, utilizado para o encaminhamento de dados em uma rede de comutação por pacotes. Assim, os dados são divididos em pacotes e individualmente encaminhados através dos nós da rede, por meio de ligações de dados compartilhadas entre eles. O aspecto não orientado a conexão do protocolo IP significa que não é necessário realizar nenhuma configuração ou comunicação inicial antes que uma máquina envie pacotes à outra. A ideia essencial da tecnologia VoIP é utilizar tanto redes privadas quanto públicas, como a Internet, para transportar voz e dessa forma concentrar os serviços de propagação de voz e dados em uma mesma tecnologia de rede. Entretanto, no cenário atual, há empecilhos para inserir com eficácia o serviço de voz sobre IP na Internet.

O serviço de voz sobre IP requer condições mínimas para funcionar satisfatoriamente. Conforme analisado em [49][50][51], um valor alto no atraso de transmissão dos pacotes, na variação do atraso (*jitter*) ou na taxa de perda de pacotes entre a origem e o destino, prejudicam a qualidade de

uma sessão VoIP. Portanto, para a Internet constituir uma alternativa atrativa à rede de telefonia pública tradicional, ela deve prover mecanismos que garantam a qualidade de serviço (QoS) para as aplicações VoIP. Os principais mecanismos de QoS são discutidos em Xiao [75].

As aplicações VoIP não eram relacionadas entre as mais populares da Internet até meados de 2003, quando a *Skype Technologies S.A.* lançou, gratuitamente, uma aplicação VoIP denominada *Skype*. Embora utilize a Internet, uma rede que não oferece garantias de QoS, uma chamada realizada através do *Skype* possui boa qualidade de voz ao utilizar codificadores de áudio proprietários e projetados para funcionar em ambientes com elevadas taxas de atraso, perda de pacotes e variação do atraso (*jitter*). Depois da popularização do *Skype*, vários aplicativos VoIP foram lançados, como o *GTalk* do *Google*, *Yahoo! Messenger* da *Yahoo! Inc* e o *Windows Live Messenger* da *Microsoft*.

Com a popularização dos aplicativos VoIP, houve uma busca do aprimoramento da tecnologia, bem como o desenvolvimentos de serviços agregados, como a capacidade de realizar chamadas de vídeo, sessões de conferência (uma chamada com vários participantes simultaneamente e a convergência com a PSTN (a capacidade de se efetuar uma chamada a um telefone localizado na PSTN e vice-versa). Com este aprimoramento, surgiram novas abordagens e técnicas para avaliar o desempenho destes novos serviços. Entre as formas de avaliação de desempenho, temos a análise da qualidade de voz, utilização dos recursos e adaptabilidade da rede como principais parâmetros.

Este capítulo tem como objetivo contextualizar o leitor entre as técnicas, tecnologias, abordagens e dificuldades relevantes referentes à avaliação de desempenho dos aplicativos VoIP, direcionando-o para a proposta deste trabalho.

2.1 A Rede de Telefonia Pública

A rede de telefonia pública (PSTN – *Public Switched Telephone Network*) foi desenvolvida com o objetivo de transmitir a voz humana a longas distâncias. Para isto, o sistema telefônico utiliza a técnica de comutação de circuitos [22]. Nesta técnica, durante uma ligação, um circuito é reservado do telefone de origem que chama ao telefone de destino, estabelecendo-se um caminho fim a fim entre as partes antes da comunicação. Por utilizar uma linha dedicada, não há o problema de congestionamento durante a ligação, o que existe na Internet.

O sistema telefônico é composto por várias centrais telefônicas: a central local, central tandem, central trânsito e a central privada ou PABX, conforme ilustrado na Figura 2.1. A central local é onde as linhas dos assinantes se encontram conectadas. Ela tem um terminal para cada assinante em um raio típico de 6 km e possui conectores para ligação com outras centrais. Estas centrais dispõem-se organizadas hierarquicamente [22] [24] em vários níveis. Cada telefone está

conectado a uma central local, que recebe o sinal analógico do telefone, o digitaliza e o transmite para uma central do núcleo da PSTN, chamada de central tandem. O sinal da conversa digitalizada segue pelas centrais do núcleo até a central local do outro telefone, onde é convertido novamente para analógico e entregue. Em sistemas de voz sobre IP, os roteadores da Internet desempenham uma função análoga às centrais da PSTN.

Em ambientes corporativos, a estrutura do sistema de telefonia é ligeiramente alterada para fornecer alguns serviços complementares como ramais, transferência de chamadas e conferência. Além disso, empresas necessitam possuir sua própria rede de telefonia para a comunicação interna. Ao contrário de telefones residenciais, onde um telefone é ligado diretamente a uma estação local, em ambientes corporativos, os telefones se ligam a uma central telefônica privada ou PABX (*Private Automatic Branch eXchange*). O PABX promove a redução de custos, através das funcionalidades descritas acima e permitindo que os usuários internos à empresa compartilhem um número limitado de linhas telefônicas externas.

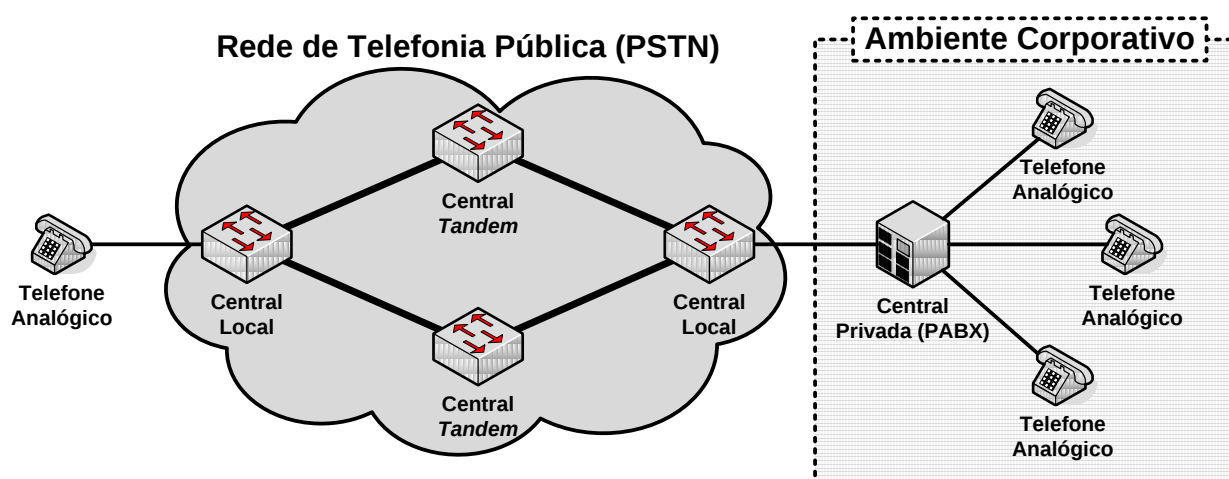


Figura 2.1 Composição do sistema telefônico.

2.2 Voz sobre IP

Embora a PSTN ofereça uma infraestrutura robusta e confiável para a comunicação de voz, existem vários motivos [63] que levam as pessoas a utilizarem uma rede baseada no protocolo IP, ou a Internet, para se comunicarem através da fala. Entre estes motivos, temos que as pessoas buscam maneiras interativas e novas de se comunicar, como a utilização de correio eletrônico, vídeo, mensagens instantâneas ou *blogs*. Também há o fato da expressiva popularidade da Internet, onde aplicações IP estão localizadas não só nos computadores, mas nos dispositivos sem fio e *smartphones*. Por fim, através da utilização da infraestrutura da Internet, há a redução dos custos relativos à realização das chamadas, principalmente as de longas distâncias.

Para estabelecer uma chamada entre duas pessoas, é preciso realizar dois procedimentos: a sinalização e a transmissão da voz. A sinalização corresponde aos processos de estabelecimento e encerramento da sessão entre a pessoa que realiza a chamada e a pessoa que recebe a chamada. Após o estabelecimento da chamada, ocorre o processo de transmissão da voz, que consiste no envio do áudio entre si através de pacotes pela rede.

2.2.1 Sinalização

A etapa de sinalização é responsável por iniciar, gerenciar e finalizar as sessões de voz. Entre os protocolos mais comuns com essa finalidade, temos o H.323 e o SIP.

2.2.1.1 H.323

Em 1998, a ITU (*International Telecommunications Union*), publicou a recomendação H.323 [64], composta de uma metodologia, uma arquitetura e um conjunto de recomendações, com o objetivo padronizar a sinalização de voz e vídeo em redes IP.

Embora o H.323 seja adotado pela indústria, ele possui diversos inconvenientes, como a alta complexidade de sua recomendação, especifica um número maior de serviços que o necessário para a telefonia via protocolo IP, os protocolos utilizam uma especificação complexa, e é necessária uma grande quantidade de mensagens e conexões para iniciar uma sessão de áudio. Assim, visando simplificar este processo, foi desenvolvido um protocolo mais simples, o SIP.

2.2.1.2 SIP

O SIP (*Session Initiation Protocol*) [65] foi desenvolvido com o intuito de sinalizar, de maneira simples, as comunicações multimídia na Internet. O SIP possui várias semelhanças com o protocolo HTTP (*Hypertext Transfer Protocol*) [66], principalmente em suas mensagens, que possuem o mesmo formato e são baseadas no modelo pedido/resposta (*request/response*).

No ambiente VoIP, o SIP utiliza um protocolo auxiliar denominado SDP (*Session Description Protocol*) [67], que troca informações sobre a sessão de áudio. Entre as informações providas pelo SDP estão o horário de início da sessão, os codificadores envolvidos, os canais de comunicação e quem originou a sessão.

2.2.2 Transmissão da Voz

O processo de manipulação de som em um sistema VoIP, desde sua entrada no sistema (um emissor enviando um sinal analógico) até sua saída em um terminal VoIP remoto (um receptor que recebe um sinal analógico), passa por diversas etapas: **amostragem e quantificação, codificação, empacotamento, buffering, desempacotamento e decodificação.**

2.2.2.1 Amostragem e Quantificação

O primeiro componente realiza a amostragem e a quantificação e é responsável por criar uma representação digital da onda sonora. Para isso, extrai um número finito de amostras do sinal analógico. A quantidade de amostras coletadas por segundo é definida como frequência de amostragem do sinal. Um CD de áudio, por exemplo, possui 44.100 amostras de som por segundo, portanto, a sua frequência de amostragem é de 44,1kHz.

Contudo, para se obter uma representação idêntica ao sinal original seriam necessárias infinitas amostras. Em contrapartida, um valor muito baixo de amostragem pode causar degradação do sinal. Assim, é preciso escolher uma taxa de amostragem que maximize a economia do consumo de memória e minimize as perdas do sinal. O Teorema de Nyquist afirma que o número mínimo de amostras necessárias para que um sinal seja reconstruído deve ser igual a duas vezes a frequência máxima encontrada no sinal [76].

A frequência máxima percebida pela audição humana é de até aproximadamente 22kHz e dos principais componentes da fala são 3,8kHz. Portanto, de acordo com o Teorema de Nyquist, para reconstituir o sinal da audição e da fala são necessárias aproximadamente 44.000 amostras por segundo e 8.000 amostras por segundo, respectivamente.

Da mesma forma que um sinal analógico pode ser traduzido em infinitas amostras, uma amostra pode assumir infinitos valores de intensidade. Portanto, o processo de quantificação limita os valores que uma amostra pode assumir. Logo, o processo também introduz erros ou ruídos.

2.2.2.2 Codificação e Decodificação

Um *codec* (contração tanto do termo Codificador-Decodificador ou Compressor-Decompressor) é o elemento de hardware ou software que obtém amostras de som e as converte em bits. Um *codec* também pode realizar compressão com o propósito de economizar capacidade de armazenamento ou de rede. Um *codec* é considerado bom quando mantém a qualidade de áudio e comprime a informação ao máximo.

Os *codecs* podem ser classificados em adaptativos ou não adaptativos. Um *codec* é dito adaptativo quando sua taxa de transmissão pode variar dinamicamente e não adaptativo quando a taxa de transmissão se mantém constante.

Outra classificação utilizada é a que divide os *codecs* em *vocoders* e *waveforms*. Os *vocoders* têm a vantagem de produzirem taxas menores de transmissão, entretanto a voz é semelhante à de um robô e qualidade da voz é geralmente inferior aos *codecs waveforms*. Nos *codecs waveforms*, o sinal codificado preserva a forma de onda sonora original e é o tipo de codificação largamente utilizado. Há diversos

codecs waveforms disponíveis, cada um com suas características particulares. Na sequência, os mais conhecidos são:

- **G.711** [58], padronizado pela ITU em 1988, utiliza a modulação por código de pulsos (PCM). A PCM é uma representação digital de um sinal analógico cuja magnitude é amostrada em intervalos uniformes e então quantizada digitalmente. O PCM é largamente utilizado em sistemas de telefonia pública. Transmite 8.000 amostras por segundo a uma resolução de 8 bits, totalizando assim 64kbps. É o *codec* base para outros padrões de codificação;
- **G.729** [57] e o **G.723.1** [59], padronizados pela ITU em 1996, pertencem à classe de algoritmos que utilizam o modelo CELP (*Code Excited Linear Prediction*), que é o modelo de codificação utilizado pela telefonia móvel. O modelo CELP foi projetado para operar em redes de circuitos chaveados, levando em consideração a perda de bits e não a perda de pacotes. Por isso, este *codec* não se comporta bem em redes com altas taxas de perda de pacotes, mesmo a taxas baixas. Entretanto, é bastante utilizado, principalmente porque foi adotado pela CISCO e posteriormente por diversas outras empresas, como uma alternativa ao G.711;
- **G.722** [56], padronizado pela ITU em 1988, têm a vantagem de ser um *codec* com maior taxa de amostragem (16kHz), o dobro do utilizado pela telefonia (cuja taxa de amostragem é de 8kHz), o que implica em menor perda de informações e consequentemente uma clareza e qualidade de voz superiores comparado aos outros *codecs*. A taxa gerada por este *codec* é de 64kbps. Devido à sua alta taxa de amostragem, não costuma ser usado para integração com a rede telefônica tradicional, que opera a 8kHz;
- **GSM** (*Global System for Mobile*) [55], padronizado pela ETSI em 1989, é utilizado pelo sistema de telefonia celular que leva o mesmo nome, e é conhecido por usar a informação da amostra anterior para prever a amostra atual e com isso gerar uma taxa de transmissão baixa (13kbps) quando comparado a outros *codecs*. É largamente utilizado devido a essa baixa taxa de transmissão e ao fato de ser livre de pagamento de licença.
- **iLBC** (*Internet Low Bit-rate Codec*) [61], foi padronizado pela IETF em 2004, e foi projetado para funcionar na Internet, operando com tolerância a taxas significativas de perda de pacotes. As taxas geradas por seus dois modos de codificação são 13.3kbps e 15kbps. É livre de licenciamento;
- **GIPS iPCM-wb** [61] é um *codec* de alta qualidade de som e baixa complexidade que opera a uma taxa fixa de 80kbps, projetado para suportar altas taxas de perda de pacotes sem degradação do sinal;

- **GIPS iSAC** [61] é um *codec* adaptativo especificamente projetado para ser usado em enlaces de baixa capacidade. Mesmo sob um acesso discado à Internet, possui qualidade de som melhor que o da PSTN. O GIPS iSAC ajusta sua taxa de transmissão dinamicamente de acordo com a velocidade da conexão. Opera na faixa entre 10 e 32kbps, é robusto contra a perda de pacotes e possui taxa de amostragem de 16kHz;
- **GIPS Enhanced G.711** [61] é uma versão aperfeiçoada do G.711 e sua principal característica é a robustez contra perda de pacotes. Opera na taxa fixa de 64kbps;
- **SVOPC** (*Sinusoidal Voice Over Packet Coder*) [4] é um *codec* proprietário desenvolvido pela *Skype S.A.*, que passou a ser utilizado a partir da versão 3.2.0.53 beta, lançada em março de 2007. Possui como principal característica a resistência à perda de pacotes, melhor qualidade do áudio e é menos suscetível a ruídos;
- **SILK** (*Super Wideband Audio Codec*) [48] é o novo codec proprietário desenvolvido pela *Skype S.A.*, lançado a partir da versão 4.0.0.226, visando substituir o SVOPC. Segundo a própria empresa, este *codec* apresenta uma menor taxa de transmissão com relação ao SVOPC, possui baixo custo computacional e utiliza uma taxa de amostragem de 24 kHz, o que lhe confere uma melhor qualidade de voz.

2.2.2.3 Empacotamento, Desempacotamento e Transmissão

As informações de VoIP utilizam como protocolo de transporte o UDP (*User Datagram Protocol*) [28], ao invés do TCP (*Transmission Control Protocol*) [27]. Entretanto, para aplicações VoIP apenas o serviço de entrega de pacotes provido pelo UDP não é suficiente. São necessários outros serviços, como, por exemplo, saber a ordem e tempo de geração de pacotes, além de possuir o conhecimento acerca da qualidade da conexão. Estas atividades competem ao RTP (*Real-time Transport Protocol*) [69] e ao RTCP (*RTP Control Protocol*) [69].

O RTP é um protocolo que, embora não proveja nenhum mecanismo que ofereça QoS, tem como objetivo dispor serviços essenciais para aplicações de tempo-real, como a ordem e tempo de geração dos pacotes, além de obviamente transportar a mídia (e.g. voz, vídeo) destas aplicações.

O RTCP é um protocolo auxiliar ao RTP que tem como objetivo fornecer um retorno às aplicações, informando a qualidade da conexão entre as mesmas, através do envio de relatórios. O RTCP pode ser usado para acompanhar a qualidade da sessão, além de detectar problemas relacionados à rede. Os relatórios contêm informações sobre estatísticas referentes aos dados transmitidos desde o início da sessão até o momento que antecede o envio do relatório, como por exemplo, número de pacotes enviados, número de pacotes perdidos e variações nos atrasos da transmissão.

2.2.2.4 *Buffering*

Na recepção do áudio de um sistema de voz sobre IP, a mídia deve ser reproduzida à medida que o receptor receber os dados e da mesma maneira que foi gerado. Portanto, o processo de envio dos dados pelo emissor está sincronizado com o processo de reprodução.

No entanto, o serviço de entrega dos dados na Internet não fornece um atraso determinístico, ou seja, o atraso é variável devido a diferentes condições de rede enfrentadas por pacotes consecutivos. Com isso, o sincronismo necessário não é garantido. Logo, é necessária uma memória temporária (*buffer*) que minimize ou elimine os problemas que a variação do atraso causa à execução do áudio.

No emissor, ocorre o empacotamento e transmissão, onde pacotes contendo amostras de áudio são periodicamente enviados. Entretanto, os pacotes chegam ao receptor com um atraso não-determinístico. Com o objetivo de suavizar essa variação do atraso, o receptor pode retardar a execução do áudio recebido armazenando os quadros mais recentes em um *buffer*.

2.3 Serviços Agregados

Atualmente, as aplicações de voz sobre IP fornecem vários serviços agregados a sua funcionalidade básica, como o serviço de *chat*, transmissão de vídeo, convergência com a PSTN e sessão de conferência.

2.3.1 Convergência com a PSTN

O serviço de convergência com a PSTN, ou simplesmente convergência, possibilita que se efetue uma chamada entre um computador conectado à Internet e um telefone na PSTN ou vice-versa.

Quando um usuário, através de um computador conectado à Internet deseja realizar uma chamada para outro computador também conectado à Internet, o aplicativo VoIP transforma a voz do emissor, utilizando um codec apropriado a este meio, em um sinal digital que é transmitido através da Internet até o receptor, que por sua vez é decodificado, transformando-se novamente em sinal de áudio.

Entretanto, quando o usuário deseja realizar uma chamada de um computador para um telefone localizado na PSTN é preciso utilizar outros mecanismos e técnicas, devido a diferenças existentes entre as tecnologias empregadas na Internet e na PSTN, pois a Internet é uma rede de comutação de pacotes e a PSTN uma rede de comutação de circuitos, sendo necessária a adequação da sinalização, a utilização de um codec apropriado e o gerenciamento da chamada. Para isto é empregado um equipamento denominado de *gateway* VoIP, responsável por executar estas tarefas.

Assim, quando um usuário que está conectado à Internet deseja realizar uma chamada a um telefone ligado à PSTN, ele utiliza o serviço de convergência do aplicativo VoIP que aciona o *gateway* VoIP mais próximo do telefone a ser chamado. O *gateway* VoIP então realizará a intermediação da chamada entre o computador e o telefone, adequando a sinalizado, recodificando e gerenciando a chamada, para que esta possa trafegar corretamente entre eles.

Quando o usuário deseja realizar o processo inverso, isto é, através de um telefone conectado à PSTN efetuar uma chamada a um computador localizado na Internet, ele novamente utiliza o serviço de convergência, que neste caso, associa um número de telefone a este computador. Este número, por sua vez, está associado ao *gateway* VoIP mais próximo ao telefone.

No *Skype*, os serviços de convergência são chamados de *SkypeIn*, que permite efetuar chamadas a partir da PSTN ao *Skype* e *SkypeOut*, que permite efetuar chamadas através do *Skype* à PSTN. O serviço de conferência é um serviço inerente à aplicação, bastando apenas selecionar os participantes que desejam participar da conferência.

Apesar de haver cobrança neste tipo de serviço, os custos relativos são bem menores quando comparados a uma mesma chamada efetuada exclusivamente pela PSTN, pois a maior parte da chamada é realizada através da Internet e apenas os últimos trechos do caminho são realizados através da PSTN.

2.3.2 Sessão de Conferência VoIP

O serviço de sessão de conferência, ou simplesmente conferência, é um serviço que possibilita que três ou mais participantes, estando eles conectados diretamente à Internet ou à PSTN (através do serviço de convergência), possam participar simultaneamente de uma mesma chamada. Embora este recurso seja facilmente encontrado na maioria dos aplicativos VoIP é importante ressaltar que nem todos os aplicativos oferecem o serviço de conferência, como por exemplo, o aplicativo *GTalk*. Neste serviço, um aspecto importante relativo ao seu funcionamento é a topologia utilizada pelo aplicativo para a distribuição do áudio entre os participantes, ou seja, como o aplicativo interconecta os participantes de modo que o áudio da conferência seja distribuído a todos. Esta característica será detalhada na Seção 2.3.2.1.

2.3.2.1 Topologias Utilizadas para o Serviço de Conferência

Em [52] são abordadas e discutidas várias topologias para sessões de conferência. Existem basicamente três abordagens para a realização de uma conferência: *multicast*, mixagem e *full mesh*.

Na abordagem *multicast*, um ou mais endereços de *multicast* são alocados para a sessão de conferência. Por sua vez, cada participante se junta a um grupo *multicast* e envia os seus dados, isto é mídia, para estes grupos. Mensagens de sinalização não são enviadas para os grupos *multicast*. Como

vantagem, este processo não requer uma coordenação rígida entre os sistemas, os participantes podem se juntar ou deixar a conferência independentemente e a conferência resiste melhor a problemas na rede. Como desvantagens, temos que a transmissão *multicast* pode ter um alto custo computacional para as redes e roteadores, pois cada roteador tem que pelo menos armazenar a identidade do grupo *multicast*, e o acesso a grupos *multicast* geralmente não é autenticado, impossibilitando haver privacidade durante as sessões de conferência. Além disso, o equipamento de cada participante deverá fazer a mixagem dos áudios recebidos para tocar ou então implementar um protocolo de escolha de fluxo de áudio (quando mais de um participante está falando).

A abordagem de mixagem consiste, a partir de um ponto que conecta todos os participantes da sessão de conferência, mixar e encaminhar a todos os participantes o áudio mixado. Existem duas variantes deste modelo: a mixagem via sistema final, onde um participante da conferência fica encarregado de mixar e encaminhar o áudio (Figura 2.2 (a)) e a mixagem via servidor de conferência, onde esta tarefa é delegada à outra entidade da rede, que não é o participante da conferência (Figura 2.2 (c)). Entre as vantagens destas abordagens temos que ambas são simples de ser desenvolvidas e consomem menos recursos de rede se comparadas com outras. Como desvantagens, temos que elas são dependentes do participante ou servidor que irá realizar a mixagem: se ele falhar, a conferência também falhará. Além disso, têm-se o aumento da carga computacional sobre o responsável pela mixagem, pois este se encarregará de mixar e encaminhar o áudio para todos os participantes da conferência.

Na abordagem *full mesh* (Figura 2.2 (b)) os participantes se comunicam diretamente entre si, não havendo algum participante topologicamente especial ou com qualquer direito ou habilidade além dos outros. Cada participante da conferência pode, a qualquer hora, convidar um novo participante da conferência. Se aceitar, este pode estabelecer conexões com os outros participantes. Este modelo apresenta como vantagem o fato de ser independente de um elemento central, responsável por mixar e distribuir o áudio. Como desvantagem, temos o aumento do tráfego de rede, principalmente à medida que se aumentam o número de participantes da conferência. No entanto, este problema pode ser minimizado através da utilização de uma topologia híbrida, conforme ilustrado na Figura 2.2 (d).

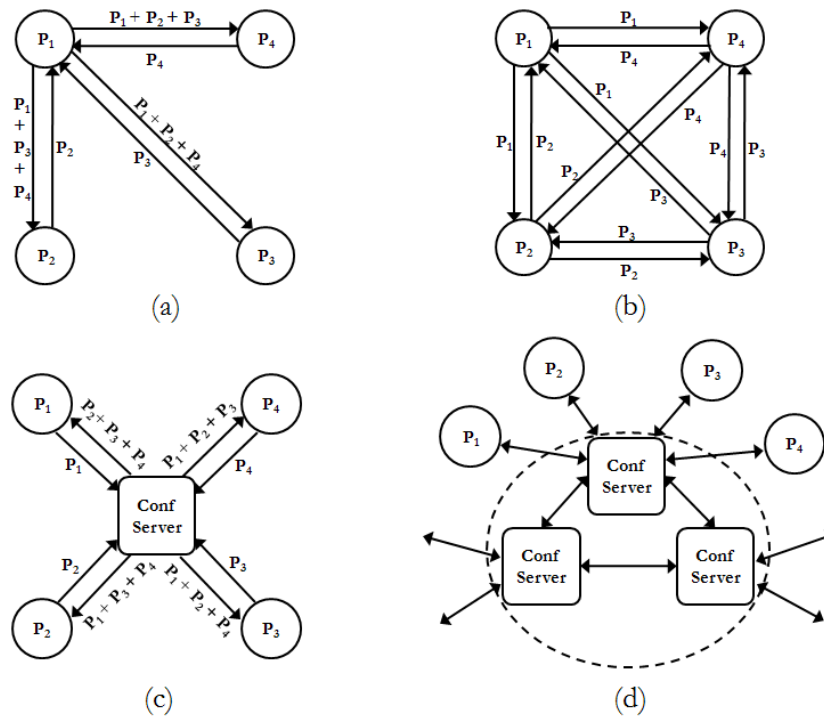


Figura 2.2 Distribuição do áudio entre diversas topologias durante uma sessão de conferência.

2.4 Avaliação da Qualidade de Voz

A avaliação da qualidade de voz tem se tornado um grande desafio para os analistas de telecomunicações devido a evolução das redes de telefonia, principalmente por causa de novas tecnologias, como GSM e VoIP. Estas novas tecnologias têm introduzido novas degradações como perda de pacotes, ruído não estacionário, e distorção da voz, causado pela codificação a uma taxa baixa (*low-bit*) ou longos atrasos devidos ao processamento digital. Assim, faz-se necessário um estudo sobre as questões relevantes sobre a avaliação, como estabelecer o que é a qualidade de voz, o contexto em que ela está inserida e os métodos utilizados.

2.4.1 Qualidade

O termo qualidade, de um modo geral, é subjetivo e depende de quem o julgará. A qualidade de voz é ainda mais complexa de definir, pois cada pessoa tem uma interpretação pessoal de um determinado evento de áudio. Se um evento de áudio corresponde a fala, o conteúdo (a informação semântica) e a forma (o sinal acústico) são analisados. Nos sistemas de telecomunicações de voz, a qualidade se refere à qualidade da forma e do sinal da fala. Contudo, a interpretação da qualidade da voz é influenciada pelo conteúdo do sinal acústico, uma medida que depende de cada pessoa. Jekosch [79] descreve a qualidade de voz como o resultado de um processo de percepção e opinião (avaliação), durante o qual o ouvinte estabelece uma relação entre o que ele entende (o evento de

áudio) e o que ele espera (uma referência interna): a qualidade de voz não é absoluta, mas é atribuída pelo ouvinte.

A qualidade da voz é um fenômeno multidimensional que pode ser avaliado de acordo com diferentes critérios de qualidade. Os dois critérios principais são a intensidade (volume) e a inteligibilidade (o nível e o entendimento/compreensão do sinal de voz, respectivamente), que permite ao ouvinte escutar e entender a mensagem do orador.

2.4.1.1 Medindo a Qualidade de Voz

Dentre as métricas utilizadas para avaliar a qualidade das chamadas telefônicas, o **MOS** (*Mean Opinion Score*) é a mais utilizada. Padronizada pela ITU-T P.800 [11] e P.830 [62], o MOS é uma abordagem de avaliação subjetiva, onde diversas pessoas ouvem um áudio que é submetido ao sistema a ser avaliado, e atribuem uma nota de uma escala, que varia de 1 (ruim) a 5 (excelente), de acordo com a qualidade do áudio ouvido. O MOS é então obtido através da média sobre estas notas. Embora esta técnica forneça um resultado significativo, ela é subjetiva, existindo assim vários problemas inerentes, como a questão da repetibilidade, o custo e o tempo gasto para ser realizado. Assim, com o intuito de minimizar estes problemas, foram desenvolvidas técnicas objetivas, como o Modelo-E e o PESQ.

O **Modelo-E** [12] é um modelo que calcula a qualidade da voz de acordo com fatores que influenciam na qualidade de voz do sistema a ser avaliado. Alguns destes fatores são: os atrasos de transmissão, eco, perdas, distorções produzidas pelos *codecs* e *jitter*. Através destes fatores, o Modelo-E calcula um fator, denominado de fator R, que pode ser convertido para uma pontuação da escala MOS. No entanto, para efetuar o cálculo do fator R são necessárias informações relacionadas ao funcionamento do *codec*, que nem sempre se encontram disponíveis ou são possíveis de obter, como no caso de aplicações VoIP que utilizem *codecs* proprietários (como o SVOPC ou SILK, utilizados pelo *Skype*). Assim, este trabalho optou por não utilizar o Modelo-E como métrica para a qualidade de voz.

O **PESQ** (*Perceptual Evaluation of Speech Quality*) [13] é uma técnica objetiva que realiza uma estimativa do MOS, através da comparação do áudio submetido ao sistema e o áudio obtido, que possivelmente é degradado ao ser codificado e transportado. Pelo fato de ser uma técnica objetiva, o PESQ é um método automatizado para mensurar a qualidade de voz de um sistema. Esta técnica se originou da combinação das técnicas PAMS (*Perceptual Analysis Measurement System*) e do PQSM (*Perceptual Speech Quality Measurement*), com o intuito de complementá-las e melhorar os seus algoritmos para cobrir uma maior quantidade de cenários, como os cenários onde ocorrem *jitter*. A ITU recomenda que utilize como carga de trabalho para avaliar um sistema através do PESQ, vozes

de dois homens e duas mulheres, incluindo pausas. Este trabalho optou por utilizar o PESQ MOS como métrica de qualidade de voz por garantir a repetibilidade dos experimentos, pois é uma técnica objetiva, e oferecer uma avaliação mais exata quando comparada com as técnicas subjetivas.

É importante salientar que embora a pontuação MOS varie de 1 (ruim) a 5 (ótimo), é sabido que a pontuação máxima obtida através do PESQ MOS é de 4,5 (ótimo). Este fato foi comprovado através de testes realizados ao longo deste trabalho através da avaliação de duas cópias de um mesmo arquivo de áudio, não apresentando assim, diferença entre eles. A Tabela 2.1 ilustra a relação entre a qualidade de voz, a pontuação MOS, a pontuação obtida através do PESQ e o intervalo do fator R [11][12][13][62].

Tabela 2.1 Relação entre a qualidade de voz, MOS, PESQ MOS e o fator R.

Qualidade da Voz	MOS	PESQ MOS	Intervalo do Fator R
Excelente	5	4,34 – 4,50	$90 < R < 100$
Bom	4	4,03 – 4,34	$80 < R < 90$
Regular	3	3,60 – 4,03	$70 < R < 80$
Fraco	2	3,10 – 3,60	$60 < R < 70$
Ruim	1	2,58 – 3,10	$50 < R < 60$

2.4.2 Contexto

A percepção da qualidade da voz da pessoa que efetuará a avaliação depende do contexto que ela está colocada [31]. Assim, podemos relacionar três contextos possíveis: o **contexto de audição**, o **contexto de fala** e o **contexto de conversação**.

2.4.2.1 Contexto de Audição

Contexto da audição corresponde à situação em que o participante ouve uma mensagem de voz sem falar, como por exemplo, quando se efetua uma chamada a uma secretária eletrônica. Este contexto pode ser perturbado devido ao *codec*, ruído, perda de informação ou nível do sinal. Estas diferentes falhas diminuem a qualidade da fala, afetando a inteligibilidade, a naturalidade ou intensidade da voz, ou diminuindo a compreensibilidade da mensagem de voz pelo ouvinte.

2.4.2.2 Contexto de Fala

O contexto de fala corresponde à situação em que o participante fala uma mensagem, sem receber uma resposta de retorno, como por exemplo, quando se grava uma mensagem em uma secretária

eletrônica. As falhas neste contexto estão relacionadas principalmente à distorção do sinal *sidetone* (o som da própria voz do orador que é recebido via o autofalante do seu telefone), eco e o ruído. Ao contrário do contexto de audição, onde as degradações são menos óbvias do que as encontradas no contexto da fala, elas podem ser muito perturbadoras para o orador.

2.4.2.3 Contexto de Conversação

O contexto da conversação corresponde à situação em que durante uma conversação, os participantes trocam mensagens alternadamente, revezando-se nos papéis de ouvinte e orador, introduzindo assim, interação entre os participantes. Estes papéis não são mutuamente exclusivos, isto é, em uma conversação, os participantes podem silenciar (silêncio mútuo) ou falar (sobreposição de vozes) ao mesmo tempo. Richards [30] propõe um modelo de quatro estados de conversação entre dois participantes, observado por uma terceira pessoa, conforme ilustrado na Figura 2.3:

- (a) O participante não fala e não escuta;
- (b) O participante está escutando a voz do outro participante, mas não está falando;
- (c) O participante está falando, mas não está escutando;
- (d) O participante está falando enquanto está ouvindo a voz do outro participante.

As transições entre os quatro estados são controlados através de oito eventos:

- i. O participante começa a falar enquanto não está ouvindo a voz do seu companheiro.
- ii. O participante começa a falar enquanto ouve a voz do seu companheiro.
- iii. O participante para de falar enquanto não está ouvindo a voz do seu companheiro.
- iv. O participante para de falar enquanto ouve a voz do seu companheiro.
- v. O participante começa a ouvir a voz do seu companheiro enquanto não fala a si mesmo.
- vi. O participante começa a ouvir a voz do seu companheiro enquanto fala a si mesmo.
- vii. O participante para de ouvir a voz do seu companheiro enquanto não fala a si mesmo.
- viii. O participante para de ouvir a voz do seu companheiro enquanto fala a si mesmo.

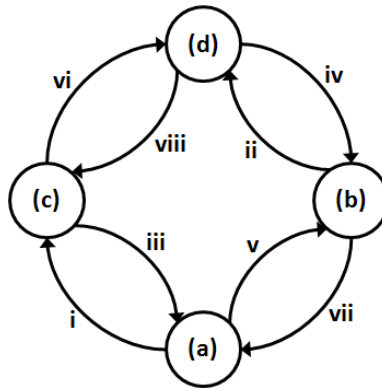


Figura 2.3 Estados e eventos de uma conversação, observados por um participante.

Com essa descrição, alguém que observa esta conversação, como a observada por um dos participantes, é composta de períodos de audição e períodos de fala, alternados de acordo com a interação do participante. Do ponto de vista da qualidade da voz, o contexto de conversação é então afetado por degradações encontradas no contexto de audição, no contexto de fala e aquelas que afetam a interatividade da conversação, como o atraso e a qualidade da voz durante os períodos de sobreposição da voz. O atraso diminui a interatividade da conversação, pois aumenta os períodos de sobreposição das vozes e de silêncio mútuo dos participantes.

2.4.3 Métodos de Avaliação

Existem vários métodos para avaliar a qualidade de voz em sistemas de telecomunicações. Entre eles os métodos subjetivos realizados através de testes utilizando pessoas, e os métodos objetivos, realizados por programas que representam o sistema de escuta e entendimento do som do ser humano.

2.4.3.1 Métodos Subjetivos

Durante um teste subjetivo, participantes são convidados a testar um sistema de telecomunicações sob diferentes condições de degradação, determinadas pelo projetista do teste, e pontuam a qualidade de voz de acordo com uma escala.

No entanto, conforme dito anteriormente, a percepção da qualidade da voz depende do contexto no qual a pessoa que o avaliará está localizada. Assim, diferentes tipos de testes subjetivos estão disponíveis.

2.4.3.1.1 Testes de Audição

Em um teste de audição os participantes são colocados em uma situação de audição e ouvem sinais de áudio correspondente a diferentes condições de degradação. Os sinais de áudio foram pré-gravados por várias pessoas e depois, todos os participantes ouvem os mesmos sinais. As

degradações que podem ocorrer neste contexto são a distorção da voz devido ao *codec* utilizado, ruído, e perda de pacotes.

2.4.3.1.2 Testes de Fala e Audição

Em um teste de fala e audição, os participantes são colocados em uma situação de fala: eles têm que simultaneamente falar em um microfone e ouvir em um altofalante.

2.4.3.1.3 Testes de Conversação

Em um teste de conversação, dois participantes são instalados em salas separadas e têm que conversar através do sistema de telecomunicação a ser testado. As condições deste tipo de teste relacionam-se as degradações encontradas nos contextos de fala e audição, bem como aqueles que afetam especificamente a interação, como o atraso e a sobreposição de vozes. As condições dos testes podem ser as mesmas para dos dois participantes (teste simétrico) ou diferentes (teste assimétrico). Como o objetivo é reproduzir uma conversa telefônica realística, os participantes são submetidos a vários cenários, como uma conversa em uma agência de viagens, um pedido de pizza ou uma informação em uma estrada. Cada participante então pontua a qualidade da conversa de acordo com os métodos definidos nas recomendações ITU-T P.800 e P.831. Neste tipo de teste os participantes são convidados a fazer uma avaliação da qualidade do áudio, da degradação devido ao eco, degradação devido ao ruído e o esforço de interrupção.

Ao contrário dos testes de audição, os testes de conversação que apenas requerem a gravação do áudio em diferentes condições de degradação e a transmissão destas gravações aos participantes, os testes de conversação necessitam a concepção de uma conexão *full-duplex* que degrade a qualidade da voz (em produção). Além disso, a gravação dos sinais da fala tem que ser realizados em tempo real e integrados ao sistema. Logo, estes testes consomem muito tempo e dinheiro e são mais raros na literatura do que os testes de audição.

2.4.3.2 Métodos Objetivos

Os métodos objetivos são uma alternativa aos métodos subjetivos e permitem automatizar a avaliação da qualidade da voz, pois eles apresentam uma grande correlação com os resultados dos testes subjetivos, que representam o julgamento dos usuários. No entanto, a utilização dos métodos objetivos não exclui os métodos subjetivos, tendo em vista que os dados gerados por estes são necessários para construir os modelos objetivos.

De acordo com [10] os métodos objetivos podem ser classificados em três critérios:

- Com o tipo de medição realizada: se são baseadas em medições físicas do sistema (paramétrica) ou através dos sinais de voz (baseados em sinais). Além disso, estes dois tipos de medições ainda podem ser divididos em dois métodos:
 - Intrusivos. É utilizado um sinal de referência que atravessa o sistema a ser testado e é capturado na saída do sistema. Este sinal capturado é o sinal degradado que será comparado com o sinal de referência. Pelo fato de inserir um sinal para efetuar a medição esta técnica afeta o tráfego da rede existente;
 - Não-intrusivos. Estes métodos não utilizam um sinal de referência, requerendo apenas o sinal degradado (sem referência). Pelo fato de não inserir um sinal para efetuar medição, estas técnicas não interferem no tráfego e portanto podem ser utilizados em redes de produção.
- Com tipo de informação necessária: ambos os lados do sistema (fim a fim ou com referência) ou apenas um lado (único-fim ou sem referência);
- Com o contexto do modelo: audição, fala ou conversação.

2.4.3.2.1 Modelos Paramétricos

Os modelos paramétricos usam medições físicas do sistema que está sendo testado para fornecer a pontuação da qualidade da voz. Entre os modelos paramétricos, o Modelo-E é o mais usado e foi desenvolvido como uma ferramenta fim a fim para projetistas de rede, conforme detalhado na Seção 2.4.1.1.

O modelo CCI (*Call Clarity Index*), descrito na recomendação ITU-T P.562 é equivalente ao Modelo-E, mas sem a utilização de referência, onde é avaliada a qualidade da voz sob o contexto da conversação através de medições do sistema a ser testado (por exemplo, nível da voz, nível de ruído e atenuação de eco), obtido através de dispositivos conhecidos como INMD (*Inservice Nonintrusive Measurement Devices*). Este modelo interpreta os dados colhidos através destes dispositivos para predizer a qualidade da voz percebida por cada participante do sistema de comunicação, em um cenário de conversação.

Os modelos paramétricos são facilmente integrados aos elementos de rede e terminais. Contudo, eles não apresentam uma grande correlação com a percepção do usuário, como os modelos baseados em sinais.

2.4.3.2.2 Modelos Baseados em Sinais

Os modelos baseados em sinais, por definição, usam os sinais de referência e o degradado (fim a fim ou com referência) ou apenas o sinal degradado (fim único ou sem referência) para prever a pontuação da qualidade de voz de um sistema.

Modelos com referência passam um sinal de referência através do sistema a ser avaliado, capturam o sinal degradado e comparam estes dois sinais para obter uma pontuação de qualidade, que são bem correlacionados com a pontuação subjetiva. Entre os modelos com referência, os mais utilizados são os que se baseiam na estrutura interna do ouvido humano, chamados de modelos perceptivos. Entre estes modelos, o PESQ foi padronizado através da recomendação ITU-T P.862.

A Tabela 2.2 apresenta um resumo dos principais métodos objetivos para avaliação da qualidade de voz e os contextos nos quais são usados.

Tabela 2.2 Classificação dos métodos objetivos para avaliação da qualidade da voz.

Modelo		Contexto de Audição	Contexto de Fala	Contexto de Conversação
Paramétrico	Fim a fim	G.107 (Modelo-E)	—	G.107 (Modelo-E)
	Um lado	P. 564	—	P.562 (CCI)
Baseado em sinais	Com referência	P.862 (PESQ)	PESQM	—
	Sem referência	P.563	—	—

2.5 Aplicações de Voz sobre IP

Atualmente, existem diversas aplicações VoIP disponíveis e facilmente encontradas na Internet. Dentre elas, destacamos o *Skype*, o *Yahoo! Messenger*, o *GTalk* e o *Windows Live Messenger*.

2.5.1 Skype

Este aplicativo permite que as pessoas através de um computador possam conversar pela Internet, ou seja, sem usar a rede de telefonia convencional (PSTN). Ainda assim, o *Skype* possui *gateways* que se conectam à PSTN oferecendo serviços pagos. O *SkypeOut* permite que um computador conectado à Internet possa estabelecer uma chamada para telefones localizados na PSTN, enquanto o *SkypeIn* oferece o serviço inverso, isto é, permite que telefones conectados à PSTN possam estabelecer chamadas a computadores conectados à Internet.

O *Skype* permite que programadores criem aplicações que funcionem em conjunto com ele através de uma API de código fechado. Como protocolo de transporte, o *Skype* utiliza tanto o UDP como o TCP. Para o tráfego de voz o *Skype* utilizava anteriormente os *codecs* que compõem a suíte da *Global IP Sound*. Atualmente o *Skype* usa o SVOPC e, mais recentemente, o SILK. O desempenho do *Skype* em termos de qualidade do áudio depende das condições da rede pelo fato de que a Internet não oferece garantia de serviço, tal como a rede de telefonia convencional (PSTN).

2.5.2 *Yahoo! Messenger*

O *Yahoo! Messenger* é um aplicativo de voz sobre IP desenvolvido pela *Yahoo!* para concorrer com o *Skype* e com o *GTalk*. Ele possui além do serviço básico de chamada, serviços agregados semelhantes ao *Skype*, como o serviço de sessão de conferência e o serviço de convergência, que neste aplicativo são conhecidos como *Phone In* e *Phone Out*. O *Phone Out* permite que uma chamada seja realizada de um computador conectado à Internet e o *Phone In* permite o contrário, isto é, permite que um telefone conectado à PSTN possa realizar uma chamada a um computador localizado na Internet. Entre os protocolos utilizados, o *Yahoo! Messenger* usa o SIP para sinalização, e o RTP para transmissão de mídia.

2.5.3 *GTalk*

O *GTalk* é uma aplicação VoIP que, diferentemente do *Skype* e do *Yahoo! Messenger*, não oferece os serviços de conferência e de convergência com a PSTN. Atualmente, além dos *codecs* da GIPS, o *GTalk* suporta os *codecs* G.711, G.723 e iLBC.

Um fato interessante a respeito do *GTalk* é que este é desenvolvido utilizando protocolos de rede e componentes de código fonte abertos, como o protocolo XMPP (*Extensible Messaging and Presence Protocol*) [80] e a biblioteca *Libjingle*. A *Libjingle* é um conjunto de componentes escritos em C++ para que outras aplicações possam ser construídas utilizando a infraestrutura VoIP oferecida pelo Google. O protocolo XMPP, além de permitir a comunicação entre os usuários do *GTalk*, provê interoperabilidade com outras redes VoIP, como a rede utilizada pelo *Gizmo Project* [53]. O serviço está hospedado no google.com e pode ser acessado pela porta 5222. O *GTalk* em seu *site* oficial afirma que no futuro suportará SIP. Para transportar mídia é utilizado o protocolo RTP.

2.5.4 *Windows Live Messenger*

O *Windows Live Messenger* é um cliente de mensagens instantâneas que além de permitir o envio de mensagens, permite que se efetuem chamadas de voz e vídeo de um computador para outro computador ou telefone. Portanto, da mesma forma do *Skype* e o *Yahoo! Messenger*, esse oferece os serviços de sessão de conferência e convergência com a PSTN.

O *Windows Live Messenger* utiliza como protocolo o MSNP (*Microsoft Notification Protocol*), que emprega o protocolo TCP para se conectar ao *Windows Live Messenger Service*, um serviço que permite que qualquer um que possua um código de identificação se conectar e comunicar em tempo real com qualquer outra pessoa que já esteja conectada. Atualmente, este protocolo encontra-se na versão 18 (MSNP18).

2.6 Trabalhos Relacionados

Barbosa [1] desenvolveu uma metodologia para realizar a avaliação de desempenho sobre aplicativos de voz sobre IP, utilizando um ambiente controlado para emular a Internet. Através desta metodologia, ele realizou uma avaliação de desempenho entre o *GTalk* e o *Skype*, com relação à adaptabilidade e a qualidade de voz destes dois aplicativos. No entanto, este trabalho não contempla em sua metodologia, bem como na sua avaliação de desempenho, os serviços agregados a estes aplicativos, como a possibilidade de estabelecer uma sessão de conferência ou o serviço de convergência com a PSTN.

Em [7] é argumentado que embora exista uma vasta literatura de como caracterizar, medir e modelar o desempenho de aplicações multimídias quando utilizada por dois usuários, através de uma técnica subjetiva (MOS), não existe uma métrica padronizada para caracterizar o desempenho de uma aplicação de conferência. Motivados por este problema, é proposta um modelo chamado GMOS (*Group Mean Opinion Score*), que se utiliza da medição subjetiva (MOS) e de parâmetros de rede (atraso e perda de pacotes) entre os pares para estimar o desempenho de uma aplicação de conferência. Através de experimentos de conferência do *Skype* é analisada se esta métrica pode ser sistematicamente aplicada para múltiplos experimentos com a mesma aplicação e a mesma população de usuários, com o objetivo de calibrar o modelo. Além disso, foi desenvolvido um método denominado TSMM (*Two-Step Mapping Method*) para prever o GMOS baseado na medição dos parâmetros de rede, semelhantemente ao Modelo-E. No entanto, esse trabalho não utiliza uma técnica objetiva para a avaliação da qualidade do áudio e não aborda os serviços de convergência.

Sat e Wah [77] analisam a qualidade de voz em um contexto de conversação quando submetidos a diversas condições de rede, como atraso, perda de pacotes e variação do atraso (*jitter*) em diversos aplicativos VoIP, como o *Skype*, *GTalk*, *Yahoo! Messenger* e *Windows Live Messenger*. Embora seja analisada a qualidade de voz em um contexto de conversação, não é abordado o serviço de conferência, isto é, múltiplos usuários e nem o serviço de convergência.

Em [10] é proposto um método objetivo para avaliar a qualidade de voz em um contexto de conversação, levando-se em consideração a qualidade da fala e da escuta e o impacto do atraso. A abordagem proposta é aplicada aos resultados de quatro testes subjetivos sob o efeito de eco, atraso,

perda de pacotes e ruído. A abordagem proposta é então, comparada com o Modelo-E. No entanto, esse trabalho não leva em consideração a análise de um aplicativo VoIP popular (como o *Skype*), o número de participantes e o serviço de convergência.

Xie e Yang [78] analisam o desempenho do *Skype* quando submetidos a diversas condições de rede, baseado no tráfego gerado pela rede de super nós. Os parâmetros de rede analisados são o atraso, a perda de pacotes e o *jitter*. Como métricas de qualidade são consideradas o MOS (utilizando o Modelo-E) e a carga do super nó. No entanto, como o tráfego gerado pelo *Skype* é criptografado e não são conhecidos os parâmetros relevantes ao *codec* utilizado, torna a utilização do Modelo-E para estimar a qualidade de voz inadequada. Também não é utilizado um ambiente controlado para a realização dos experimentos e não foi realizada a avaliação dos serviços de conferência e convergência.

3 Metodologia

“Se, a princípio, a ideia não é absurda, então não há esperança para ela.”

Albert Einstein

Este trabalho tem como objetivo desenvolver uma metodologia flexível e automatizada para realizar a avaliação de desempenho de um aplicativo VoIP, incluindo os serviços agregados de convergência e conferência, independentemente do tipo de aplicativo a ser avaliado. Flexível, pois permite que possa ser aplicada para analisar e compreender o comportamento de um sistema VoIP quando utilizada sob diversos contextos (Seção 2.4.2) e cenários (Seção 3.2), aos quais estes sistemas podem ser submetidos; objetiva, diminuindo assim o tempo e os custos relativos a este tipo de avaliação; independente do tipo de aplicativo VoIP, pois não utiliza uma abordagem proprietária ou exclusiva de algum tipo de aplicação, podendo assim, ser aplicado para a maioria dos aplicativos VoIP. Para atingir este objetivo, esta metodologia apresenta uma classificação para os diversos tipos de cenários VoIP, um ambiente controlado para emulação da Internet, destinado à execução dos experimentos, elaboração e utilização de uma carga de trabalho adequada, para ser submetida a este ambiente controlado e um conjunto de critérios (métricas) a ser utilizado para avaliar o desempenho de um sistema VoIP.

A classificação para os diversos tipos de cenários que podem ser utilizados por um aplicativo VoIP ajudará na organização e discussão em torno dos experimentos realizados, pois estes variam de acordo com o número de participantes de uma sessão VoIP, a ocorrência de convergência de redes e rede de trânsito utilizada.

O desenvolvimento de um ambiente controlado para a execução dos experimentos se faz necessária, uma vez que é realizada uma medição do aplicativo sob certas condições de rede. Em um processo de medição, existem muitas variáveis que não podem ser controladas, sendo necessária assim, a minimização das interferências que estas podem causar aos experimentos realizados.

A elaboração da carga de trabalho adequada que possa avaliar corretamente, através das métricas escolhidas, os aplicativos VoIP, evitando assim, uma amostra não tendenciosa, quando aplicada a metodologia proposta. Além disso, a carga de trabalho é utilizada de maneira complementar ao ambiente controlado, ajudando-o na execução dos cenários avaliados nos experimentos.

Como conjunto de critérios (métricas) utiliza-se a qualidade de voz e a adaptabilidade. Por adaptabilidade, este trabalho assume que seja a capacidade e eficiência do mecanismo de adaptação das aplicações com relação às mudanças no comportamento da rede. Uma forma de obter estas métricas é através da análise do conteúdo dos pacotes gerados por estas aplicações, e assim, descobrir o *codec* utilizado, seus parâmetros e as condições da rede, com o objetivo de aplicá-los em um modelo, como o Modelo-E, apresentado na Seção 2.4.1.1, obtendo uma estimativa da qualidade de voz da chamada. Contudo, a utilização do Modelo-E não pode ser utilizada na maioria dos casos, tendo em vista que muitos sistemas VoIP, como o *Skype*, utilizam um protocolo proprietário e criptografa os dados que são trocados entre os clientes deste aplicativo, dificultando, assim, a obtenção de informações necessárias a análise. Outro fato importante é a existência de *codecs* proprietários e recentes, inexistindo assim, estudos ou modelos formados relacionando estes *codecs* proprietários com qualidade de voz. Além disso, nos cenários de utilização onde há a convergência das redes, não é possível obter informações a respeito da PSTN, uma vez que o seu acesso é restrito as companhias telefônicas.

Por estes fatores, este trabalho analisa o comportamento dos sistemas VoIP, através de experimentos controlados, sob o ponto de vista externo, isto é, como uma caixa preta. Para isso, os pontos de entrada e saída do aplicativo são analisados e utilizados para inferir os parâmetros de desempenho. Os pontos selecionados foram a interface de rede de cada participante e a placa de som do receptor, no caso de ser um cenário com apenas dois participantes (ponto a ponto) ou a placa de som do participante *host* da sessão de conferência, no caso de um cenário com múltiplos participantes (multiponto).

3.1 Terminologia

Antes de explicar a metodologia desenvolvida neste trabalho, é interessante fazer uma pequena introdução à terminologia utilizada na área de avaliação de desempenho. De acordo com Jain [25], um **sistema** é um conjunto qualquer de hardware, software e firmware. **Métrica** é o critério utilizado para quantificar ou avaliar o desempenho de um sistema. **Carga de trabalho** é o conjunto de atividades ao qual um sistema é submetido. **Parâmetros** são as características do sistema e da carga de trabalho que afetam o desempenho do sistema. **Fatores** são os parâmetros que variam ao longo da avaliação de desempenho. **Níveis** são valores que os fatores podem assumir. Neste trabalho, definimos que um **cenário** é uma sequência ininterrupta de experimentos, em que para cada experimento, os fatores relacionados à métrica a ser avaliada variam em diferentes níveis durante a sua execução.

3.2 Classificação e Descrição dos Cenários de Utilização

Partindo de um estudo preliminar sobre o uso dos aplicativos VoIP, foram identificados oito tipos de cenários de utilização, classificados de acordo com o **número de participantes**, **convergência entre redes** e a **rede de trânsito utilizada**. Para cada uma dessas características, assumimos, ao longo deste trabalho, que:

- De acordo com o número de participantes, um cenário de utilização será do tipo **ponto a ponto** se houver apenas dois participantes (usuários) em uma mesma chamada, ou **multiponto**, se houver mais de dois. Uma chamada multiponto também pode ser denominada de sessão de conferência ou simplesmente conferência. Nos cenários de utilização multiponto também há presença de um participante especial, denominado de **participante host**, que é o responsável por iniciar a conferência. Dependendo do sistema VoIP utilizado, este participante pode desempenhar algumas funções extras, como a responsabilidade de combinar e distribuir o áudio produzido por todos os participantes da conferência;
- A respeito da **convergência entre redes**, um cenário de utilização apresentará convergência se, para estabelecer uma chamada, é utilizado mais de um tipo de rede, que pode ser a Internet ou a PSTN. A convergência é estabelecida através de equipamentos denominados *gateways*, que encaminham a chamada originada em uma rede, para outra rede de outro tipo, como por exemplo, uma chamada que é originada na Internet e tem como destino um telefone localizado na PSTN, ou vice-versa. Assim, neste trabalho, um cenário de utilização será categorizado do tipo **homogêneo** quando não apresentar convergência ou **heterogêneo**, se houver;
- Com relação à **rede de trânsito utilizada**, isto é, uma rede que utiliza um tipo diferente de tecnologia para transmissão de dados (comutação por pacotes ou comutação de circuitos) da rede que contém a máquina de origem e de destino da chamada, que podem ser a Internet ou a PSTN. Um exemplo é uma chamada originada em um telefone localizado na PSTN, que utiliza a Internet como rede de trânsito, e tem como destino outro telefone localizado também na PSTN.

A Tabela 3.1 apresenta um resumo das classes de cenários de utilização possíveis de serem construídos de acordo com as características apontadas e a classificação proposta neste trabalho.

Tabela 3.1 Classificação dos cenários de utilização possíveis e suas características.

Classe	Número de Participantes	Convergência	Rede de Trânsito
1	Ponto a ponto	Homogêneo	—
2	Multiponto	Homogêneo	—
3	Ponto a ponto	Heterogêneo	—
4	Multiponto	Heterogêneo	—
5	Ponto a ponto	Heterogêneo	Internet
6	Multiponto	Heterogêneo	Internet
7	Ponto a ponto	Heterogêneo	PSTN
8	Multiponto	Heterogêneo	PSTN

A Figura 3.1 ilustra os cenários de utilização de classe 1 e 2, que se caracterizam por serem ponto a ponto ou multiponto, homogêneo, e não utilizarem rede de trânsito. Estes cenários de utilização são os mais utilizados pelos usuários, por apresentar menor custo, pois utilizam apenas a infraestrutura da Internet para estabelecer uma chamada. No entanto, nem todos os aplicativos VoIP oferecem a possibilidade de serem utilizados em um cenário multiponto, isto é, não são capazes de estabelecer uma conferência, como por exemplo o aplicativo *GTalk* [15].

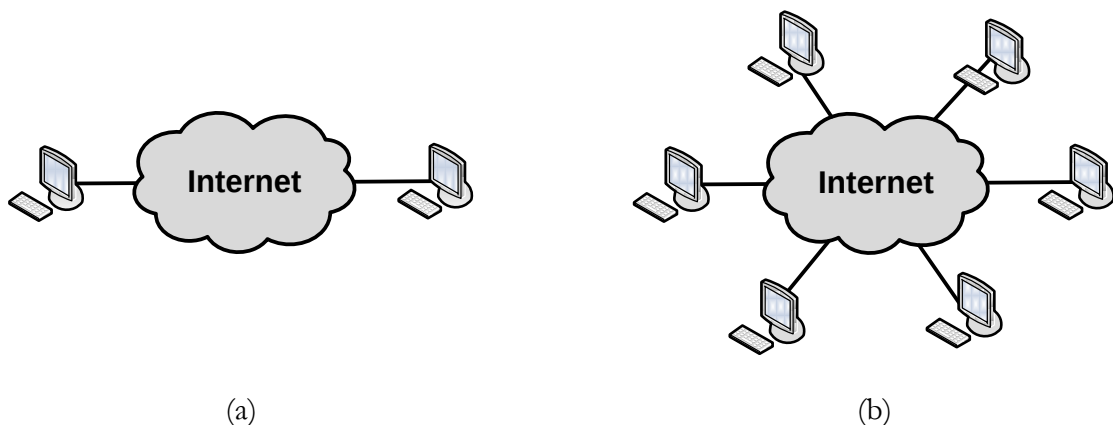


Figura 3.1 (a) Cenário de classe 1 – Ponto a ponto, homogêneo, sem rede de trânsito. (b) Cenário de classe 2 – Multiponto, homogêneo e sem rede de trânsito.

A Figura 3.2 apresenta os cenários de utilização de classe 3 e 4 que se caracterizam por serem ponto a ponto ou multiponto, heterogêneo e não utilizarem rede de trânsito. Como exemplos destes cenários de utilização, temos os serviços *SkypeIn* e *Phone In* (ligação efetuada da PSTN para um usuário conectado à Internet) e *SkypeOut* e *Phone Out* (ligação efetuada da Internet para um telefone conectado à PSTN), oferecidos, respectivamente, pelo *Skype* e *Yahoo! Messenger*.

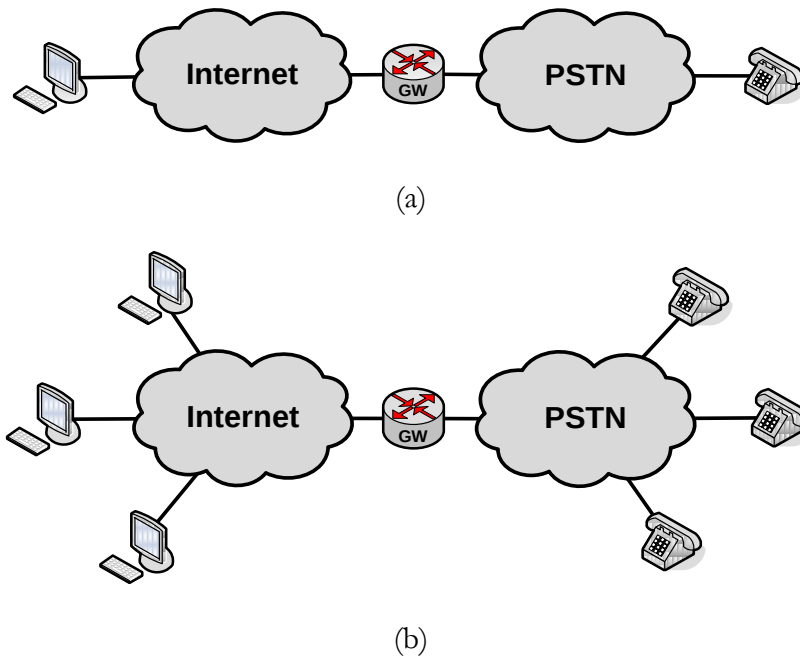
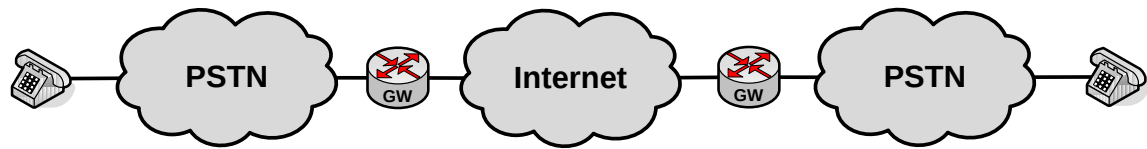
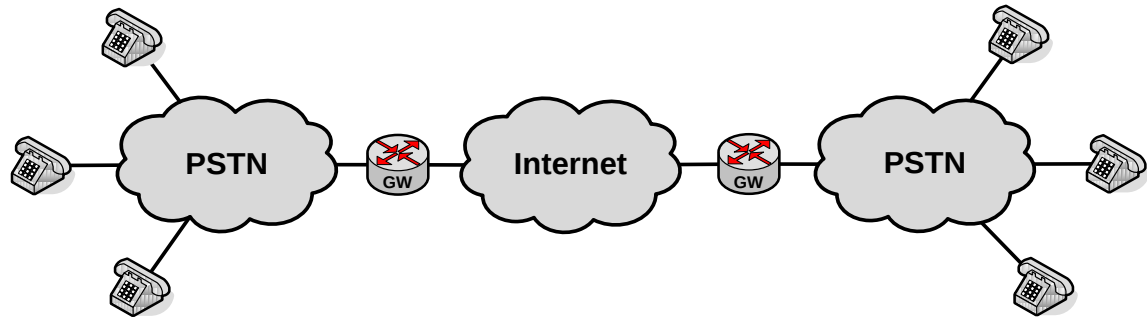


Figura 3.2 (a) Cenário de classe 3 – Ponto a ponto, heterogêneo e sem rede de trânsito. (b) Cenário de classe 4 – Multiponto, heterogêneo e sem rede de trânsito.

A Figura 3.3 representa os cenários de utilização de classe 5 e 6, que se caracterizam por serem ponto a ponto ou multiponto, heterogêneo e utilizarem a Internet como rede de trânsito. O uso deste tipo de cenário de utilização está se tornando cada vez mais comum, por dois motivos: a maior distância da chamada é feita através da Internet, reduzindo o custo da chamada e que o uso deste tipo de serviço dispensa manter uma infraestrutura de Internet adequada à utilização de aplicativos VoIP pelos participantes da chamada, reduzindo ainda mais os custos envolvidos. O funcionamento deste tipo de serviço é bem simples, bastando apenas que a pessoa que queira efetuar a chamada adquira créditos e informe no *site* do prestador deste serviço o(s) número(s) do(s) telefone(s) do(s) participante(s) da chamada. Dentre os prestadores deste tipo de serviço, destacamos o *JaJab* [18].



(a)



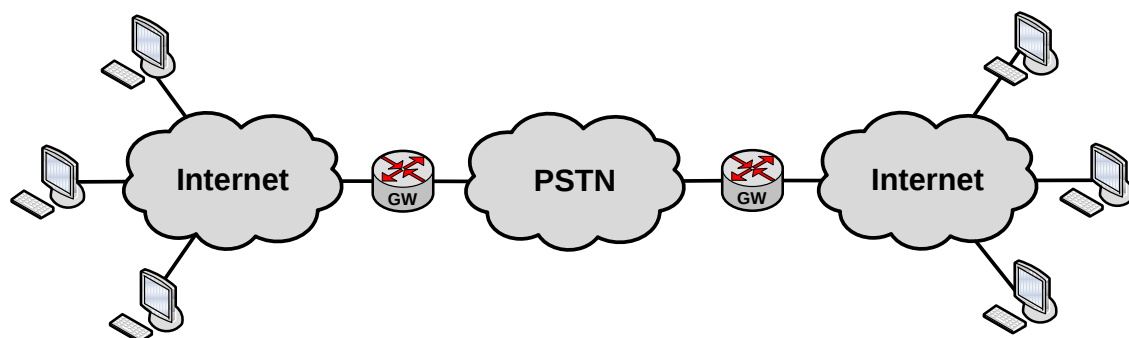
(b)

Figura 3.3 (a) Cenário de classe 5 – Ponto a ponto, heterogêneo e com a Internet como rede de trânsito. (b) Cenário de classe 6 – Multiponto, heterogêneo e com a Internet como rede de trânsito.

A Figura 3.4 ilustra os cenários de utilização de classe 7 e 8, que se caracterizam por serem ponto a ponto ou multiponto, heterogêneo e utilizarem a PSTN como rede de trânsito. Um exemplo deste tipo de cenário é um usuário que, através do serviço *SkypeIn* do *Skype*, deseja efetuar uma chamada para um número do *Phone Out* do *Yahoo! Messenger*. A chamada partiria da PSTN e seria encaminhada para a Internet, que por sua vez seria encaminhada da Internet para a PSTN. Embora estes cenários sejam tecnicamente possíveis, eles provavelmente não são economicamente viáveis, pois a utilização da PSTN como rede de trânsito aumentaria os custos relativos à chamada.



(a)



(b)

Figura 3.4 (a) Cenário de classe 7 – Ponto a ponto, heterogêneo e com a PSTN como rede de trânsito. (b) Cenário de classe 8 – Multiponto, heterogêneo e com a PSTN como rede de trânsito.

Este trabalho propõe a utilização da notação **<classe-do-cenário-de-utilização>. <número-do-cenário-de-utilização>**, onde **<classe-do-cenário-de-utilização>** refere-se à qual classe (de 1 a 8), segundo a classificação proposta, o cenário avaliado pertence, e o **<número-do-cenário-de-utilização>**, que se refere à numeração do cenário com relação à classe que ele pertence. Assim, por exemplo, o cenário 1.2 é o segundo cenário de utilização da classe 1, e portanto, sabe-se que este cenário de utilização é do tipo ponto a ponto, homogêneo e não utiliza rede de trânsito, enquanto que o cenário 8.3 é o terceiro cenário da classe 8, e portanto, este cenário é do tipo multiponto, heterogêneo e possui utiliza a PSTN como rede de trânsito.

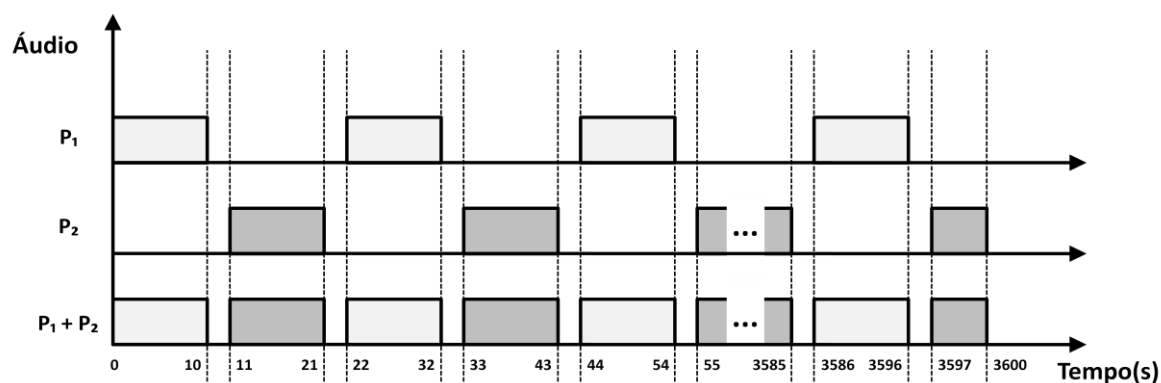
3.3 Carga de Trabalho

Na Seção 3.2 foram descritos vários cenários de utilização para sistemas VoIP, onde se variam o número de participantes, a convergência entre as redes e a rede de trânsito utilizada. Com a finalidade de submeter cada tipo de cenário de utilização às mesmas condições, isto é, a uma mesma carga de trabalho, foi criado um conjunto de arquivos de áudio que visam emular os participantes que compõe estes cenários. Para criar estes arquivos de áudio, foi utilizado como base o mesmo arquivo utilizado por Barbosa [1]. O áudio contido nesse arquivo foi gravado com taxa de amostragem de 16kHz, qualidade mono, e é composto pelas vozes de quatro pessoas diferentes: dois homens e duas mulheres, incluindo pausas, conforme a recomendação da ITU-T P.800, onde cada pessoa fala por 15 minutos, totalizando assim 60 minutos de áudio. Esse arquivo foi utilizado para criar um conjunto de arquivos de áudio para cada classe do cenário de utilização, conforme segue:

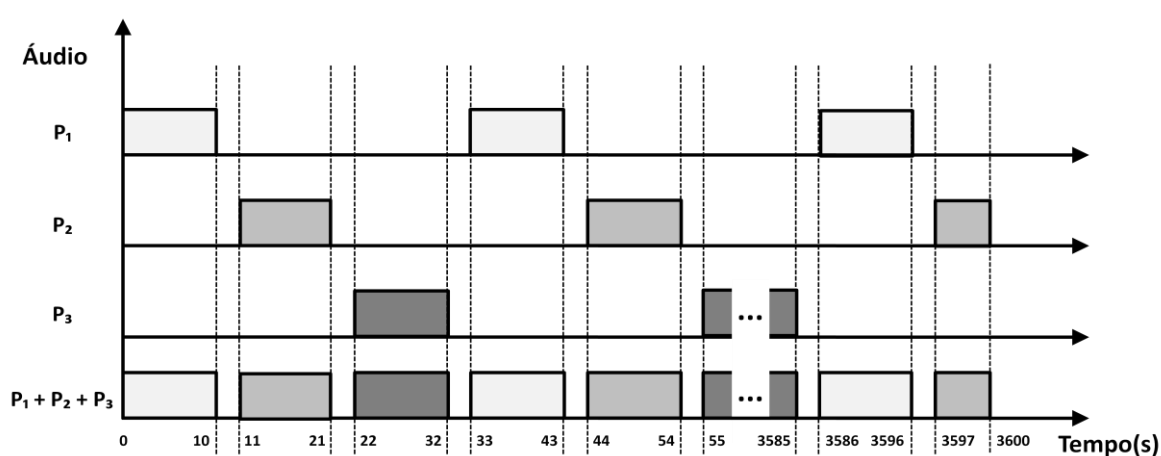
- Para os cenários de utilização de classe 1 foi utilizado o mesmo arquivo base, sem modificações;

- Para os cenários de utilização de classe 2 foi necessário criar uma técnica que emulasse um contexto de conversação (Seção 2.4.2.3), isto é, um cenário onde houvesse vários participantes alternando entre si períodos de fala e de silêncio, durante uma sessão de conferência. Para isto, foi adotada a seguinte abordagem: foi criado um conjunto de arquivos de áudio a partir de uma cópia do arquivo base para cada participante e depois editada, de modo que houvesse um período de 10 segundos de áudio [32], intercalados com períodos de 12 segundos de silêncio. Este período de silêncio será parcialmente utilizado pelo áudio relevante ao próximo participante e assim consecutivamente, até a emissão do áudio relativo ao último participante, retornando novamente ao primeiro. Deve-se atentar que entre o período de emissão do áudio de cada participante haverá um período de 1 segundo de silêncio. Este intervalo foi adotado tendo-se em vista sua relevância, haja vista que neste intervalo várias palavras podem ser pronunciadas, e podem se tornar inteligíveis, caso ocorra uma perda consecutiva de vários pacotes [32]. A Figura 3.5 ilustra a forma de elaboração dos arquivos de áudio relativos aos cenários com 3 e 4 participantes;
- Para os cenários de utilização de classe 3, 5 e 7 foi criado um outro arquivo de áudio base, contendo apenas 7,5 minutos de áudio relativo às vozes de cada pessoa (ao invés dos 15 minutos de áudio) do arquivo original, totalizando assim, 30 minutos de áudio. A diminuição do arquivo de áudio foi necessária devido aos custos relativos a esta classe de cenário;
- Para os cenários de utilização das classes 4, 6 e 8 foi utilizada a mesma técnica de composição de arquivos de áudio dos cenários de classe 2, mas tendo como base, o arquivo de áudio criado para os cenários de classe 3, 5 e 7. O objetivo da construção destes novos arquivos é a emulação de um contexto de conversação de 30 minutos.

É importante notar que em todos os cenários de utilização multipontos não há uma elaboração do áudio relativo ao participante *host* da conferência, tendo em vista que nesta metodologia convencionou-se que este participante seria responsável apenas por iniciar a sessão de conferência e realizar a medição do experimento.



(a)



(b)

Figura 3.5 (a) Elaboração do áudio para uma sessão de conferência para três participantes. (b) Elaboração do áudio para uma sessão de conferência para quatro participantes.

Após a criação dos arquivos de áudio referentes a cada participante, estes são combinados utilizando um programa de edição áudio, resultando em um arquivo que será usado como referência na comparação com o áudio gravado dos experimentos realizados.

3.4 Métricas

Os parâmetros de avaliação de desempenho utilizados são a **qualidade do áudio recebido** e a **taxa transmitida** dos participantes dos cenários avaliados.

Para avaliar a qualidade de áudio foi usado o algoritmo PESQ MOS. No entanto, esta métrica é utilizada para os cenários ponto a ponto, isto é, onde não é utilizado o serviço de conferência. Ao se avaliar o serviço de conferência, que se caracteriza pelo contexto de conversação, é preciso utilizar uma métrica mais adequada, que será determinada pelo tipo de topologia utilizada pelo aplicativo

VoIP para mixagem e distribuição do áudio. Assim, este trabalho propõe a métrica denominada **PESQ MOSconf**, que é baseada no algoritmo PESQ MOS e é detalhada na Seção 3.7.

O cálculo da taxa transmitida é baseado no processamento do tráfego capturado na interface de rede de cada participante do serviço analisado, através da ferramenta *tcpstat* [54].

Embora os experimentos realizados utilizem um ambiente controlado, o fato de se ser efetuando uma medição sobre este não garante um controle sobre todas as variáveis deste ambiente. Por exemplo, ao se configurar um caminho com capacidade inferior à taxa transmitida pela aplicação ocorre um aumento no *jitter* devido ao enfileiramento de pacotes.

Além disso, pode ocorrer a perda de pacotes e a taxa transmitida pode ser diferente da taxa recebida. É conhecido que o *jitter* e a perda de pacotes degradam a qualidade de voz recebida. Portanto, a análise das métricas *jitter* e taxa de perda de pacotes agem indiretamente na avaliação de desempenho, realizando a função de auxiliar o entendimento dos resultados obtidos através do PESQ MOS.

A Tabela 3.2 relaciona o critério de avaliação usado nesta metodologia, o tipo cenário de utilização (ponto a ponto ou multiponto), o dispositivo monitorado, a informação coletada deste dispositivo e o parâmetro derivado de cada informação.

Tabela 3.2 Métricas.

Critério de Avaliação	Tipo de Cenário de Utilização	Dispositivo	Informação	Métrica
Qualidade do áudio	Ponto a ponto	Placa de som do receptor	Onda sonora	PESQ MOS
	Multiponto	Placa de som do participante host	Onda sonora	PESQ MOSconf
Adaptabilidade	Ponto a ponto	Interface de rede do emissor	Fluxo de pacotes	Taxa transmitida
	Multiponto	Interface de rede de cada participante	Fluxo de pacotes	Taxa transmitida

3.5 Ambiente de Realização dos Experimentos

O ambiente para a realização dos experimentos dos serviços agregados de convergência e conferência foi montado de acordo com a Figura 3.6. Esse ambiente foi elaborado de forma a permitir a automatização e a replicação dos experimentos, e é dividido em duas partes: o ambiente controlado e o ambiente não-controlado.

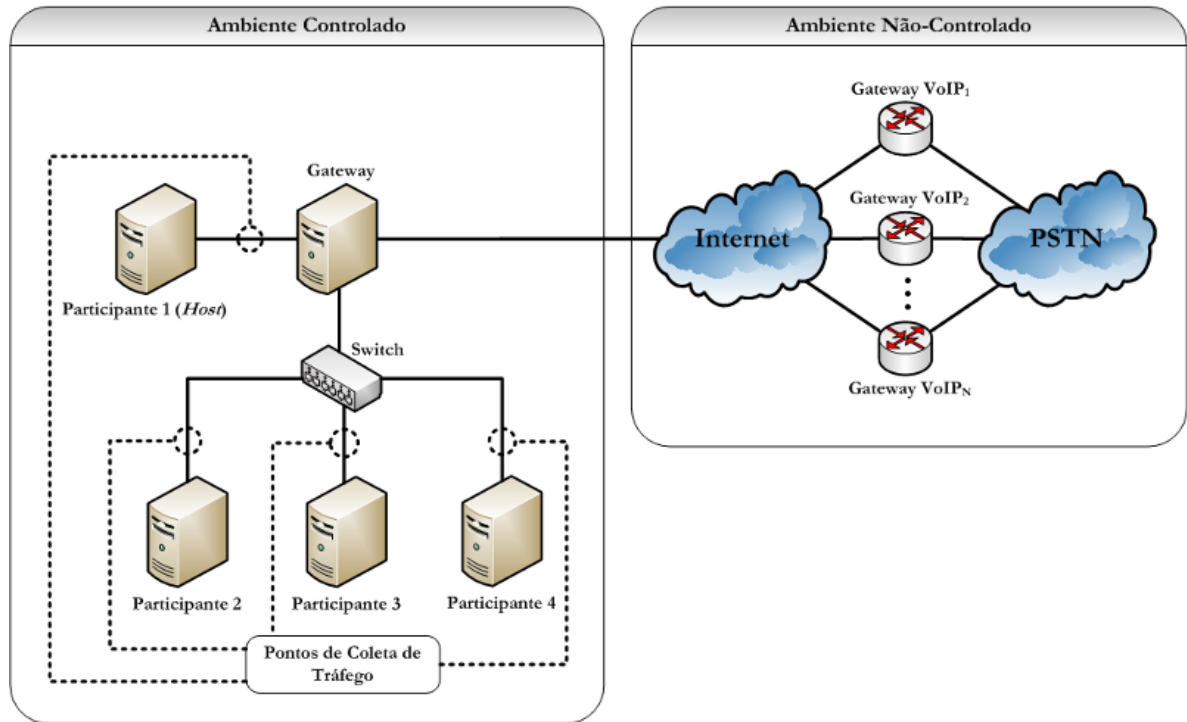


Figura 3.6 Ambientes de realização dos experimentos: ambiente controlado e o não-controlado.

O ambiente controlado é a parte que envolve os serviços de sessão de conferência. Ele é composto de 05 computadores tipo PC. A configuração das máquinas dos participantes é Intel Pentium 4, 1GB de memória RAM, placa de rede 10/100Mbps Intel e *Windows XP SP3*. A máquina *gateway* é composta por um processador AMD Athlon 64 3200+, 1GB de memória RAM, placa de rede 10/100Mbps, Linux 2.6.16 de 64 bits, e um *switch* 10/100Mbps 3Com. As máquinas de todos os participantes deste ambiente possuem tabelas de roteamento ajustadas para que todo o tráfego entre elas seja encaminhado através da máquina *gateway* (emulador Internet), que emula as condições da rede de acordo com parâmetros específicos para cada experimento. Adotamos um emulador de rede ao invés de um simulador ou do uso de medições reais da Internet porque este permite maior controle do ambiente e garante a repetibilidade dos experimentos. A ferramenta de emulação usada foi o *NIST.Net* [44]. Embora o tráfego entre as máquinas dos participantes seja encaminhado através do *gateway*, todas possuem acesso à Internet pública, para que os participantes possam ser autenticados pela aplicação, acessar suas listas de contatos, fazer a chamada e iniciar a sessão VoIP.

O ambiente não-controlado é a parte do ambiente que envolve o serviço de convergência com a PSTN. Essa parte do ambiente não há o controle dos parâmetros de rede e os experimentos envolvendo esta parte serão do tipo caixa preta. Após a execução dos experimentos é realizada uma estimativa do atraso e da perda de pacotes entre a máquina do participante 1 (participante *host*) localizado no ambiente controlado até os *gateways* VoIP utilizados no serviço de convergência. Os endereços IP relativos a estes *gateways* são obtidos através da análise do tráfego capturado no ambiente controlado. O cômputo da estimativa da perda de pacotes e atraso é realizado através do programa *My Traceroute* (*mtr*) [40], que pode ser configurado para utilizar o protocolo UDP ao invés do ICMP (*Internet Control Message Protocol*). A utilização do protocolo UDP ao invés do protocolo ICMP se deve ao fato de que, através de experimentos iniciais, se verificou que o protocolo ICMP apresentava uma diferença considerável quando comparado com o protocolo UDP. Esta diferença pode ser atribuída ao fato de haver roteadores na Internet cujas filas são configuradas com baixa prioridade ao protocolo ICMP.

Em ambos os ambientes, a máquina do participante 1 (*Host*) é encarregada de estabelecer a chamada VoIP e efetuar a gravação do áudio resultante, através do software *Audacity* [41]. A gravação do áudio é utilizada para o cálculo do PESQ MOS, realizado pela implementação de referência do algoritmo PESQ [46]. Todas as máquinas dos participantes possuíam instalado o *Meinberg NTP Client* [45], um cliente NTP (*Network Time Protocol*) para a sincronização dos relógios. Durante todos os experimentos, o tráfego de entrada e saída destas máquinas foram capturados e utilizados no cálculo da vazão e do *jitter*. A captura do tráfego foi realizada usando a ferramenta *tsbark* [43] (uma versão terminal da ferramenta *Wireshark*). A Tabela 3.3 apresenta as ferramentas utilizadas na elaboração deste ambiente controlado.



Figura 3.7 Posicionamento do cabo de áudio para os participantes emissores de áudio.

Para emular os participantes em um contexto de conversação foi utilizado o programa *Winamp* [42], que é um reproduzidor de arquivos de áudio, para realizar a reprodução do áudio utilizado como

carga de trabalho. Cada máquina, com exceção da máquina do participante onde seria feita a captura do áudio resultante da chamada, têm um cabo de áudio conectado a saída do áudio para o microfone do computador (Figura 3.7), de modo que quando o tocador de áudio reproduz os arquivos de áudio, este é conduzido até a entrada do microfone, que por sua vez é levado até o aplicativo VoIP o qual está sendo submetido à avaliação de desempenho. Além disso, deve ser configurado na entrada do áudio do aplicativo VoIP avaliado, a eliminação do volume, de modo que o áudio relativo à chamada, que deve ser entregue ao participante, não seja misturado à carga de trabalho que está sendo reproduzida pelo *Winamp*.

Tabela 3.3 Ferramentas utilizadas no ambiente controlado de execução dos experimentos.

Software	Sistema Operacional	Versão	Funcionalidade
<i>Audacity</i>	<i>Windows / Linux</i>	1.2.6	Gravação do áudio resultante dos experimentos realizados
<i>Winamp</i>	<i>Windows</i>	5.531 Lite	Execução dos arquivos de áudio nos participantes
<i>Meinberg NTP Client</i>	<i>Windows / Linux</i>	4.2.4p@1.520-modena	Sincronização do relógio das máquinas dos participantes e <i>gateway</i>
<i>NIST.Net</i>	<i>Linux</i>	3.0 Alpha	Emulador Internet
<i>PESQ</i>	<i>Linux</i>	2.0	Cômputo da qualidade de voz
<i>IPstat</i>	<i>Linux</i>	1.0	Cômputo da perda de pacotes e <i>jitter</i>
<i>tsbark</i>	<i>Windows</i>	1.0.3	Captura de pacotes
<i>My Traceroute (mtr)</i>	<i>Linux</i>	0.75	Cômputo da perda de pacotes e atraso

3.6 Experimentos Realizados

Com base na classificação dos cenários de utilização proposta neste trabalho, foram elaborados diversos experimentos que visam analisar e estudar o comportamento dos aplicativos VoIP quando submetidos à variadas condições de rede. Esses experimentos estão divididos em duas partes: os

experimentos de calibração e os experimentos de avaliação. Os experimentos de calibração apontam os limiares críticos dos aplicativos VoIP avaliados, enquanto os experimentos de avaliação são usados para a avaliação de desempenho propriamente dita.

Os fatores utilizados nesta metodologia são enumerados na Tabela 3.4. O primeiro fator está relacionado ao número de participantes da conferência. O segundo fator representa a capacidade residual de um enlace, também chamado de banda disponível. O terceiro fator é o atraso unidirecional, conforme definido na Seção 3.8. O quarto fator é definido como sendo a razão entre pacotes enviados perdidos e o total de pacotes enviados, conforme discutido na Seção 3.9. O quinto fator é o parâmetro *DelaySigma* do *NIST.Net* que corresponde ao *jitter*.

Tabela 3.4 Fatores utilizados nos experimentos.

Nº	Fator
1	Número de participantes
2	Capacidade
3	Atraso
4	Perda de pacotes
5	<i>DelaySigma (Jitter)</i>

O fator número de participantes é geralmente configurado no momento de estabelecimento de uma chamada VoIP, enquanto os fatores capacidade, atraso, perda de pacotes e *DelaySigma (jitter)* usados nos experimentos foram configurados no emulador *NIST.Net*, apresentado na seção 3.5. A política de descarte de pacotes disponibilizada pelo emulador é o DRD [26] (*Derivative Random Drop*), uma simplificação da política RED (*Random Early Detection*) [72] que procura descartar os pacotes de maneira aleatória quando a fila atinge um limiar máximo, calculado de acordo com o tamanho médio estimado da fila. O DRD descarta pacotes com uma probabilidade (após a fila alcançar o tamanho de um limiar mínimo) que cresce linearmente com o tamanho da fila até o tamanho da fila alcançar um limiar máximo, a partir do qual os pacotes são descartados com 95% de probabilidade.

Os valores escolhidos para os limiares do DRD foram os mesmos escolhidos por Barbosa [1], 49 e 50. Estes números foram escolhidos para tornar o comportamento semelhante a uma política de gerenciamento de espaço de fila *DropTail* (descarta da cauda os pacotes que chegam após o tamanho da fila atingir o seu tamanho máximo) de tamanho 50 e política de escalonamento FIFO

(*First In, First Out* – o primeiro pacote a entrar na fila é o primeiro a ser encaminhado). Estes parâmetros são largamente utilizados em simulações [47].

3.6.1 Topologia para a Mixagem e Distribuição do Áudio

Antes de realizar os experimentos é necessário determinar qual o tipo de topologia utilizada pelo aplicativo VoIP analisado, pois o tipo de topologia irá determinar qual a melhor forma de realizar o cômputo da qualidade de voz sob o contexto de conversação, necessário para o serviço de conferência, conforme abordado com maiores detalhes na Seção 3.7. Algumas vezes, é possível encontrar o tipo de topologia utilizada pelo aplicativo através da documentação referente a este.

Caso o tipo de topologia utilizado pelo aplicativo não esteja explicitado na documentação, a mesma poderá ser inferida através da análise do tráfego entre os participantes da conferência. Se na análise do tráfego gerado entre os participantes da conferência for detectado que os participantes utilizam o participante que iniciou a chamada (participante *host*) para o envio e recebimento do áudio entre eles, isto indica que o participante *host* é o responsável pela mixagem e distribuição do áudio durante a conferência, caracterizando a utilização da topologia do tipo mixagem via sistema final por parte do aplicativo analisado.

3.6.2 Experimentos de Calibração

Os experimentos de calibração estimam os pontos de adaptação do aplicativo VoIP analisado. Estes pontos determinarão os limiares das condições de rede, utilizados para a composição dos níveis dos experimentos de avaliação.

É importante salientar que o propósito dos experimentos de calibração não é realizar avaliações que gerem conclusões sobre a qualidade das aplicações consideradas neste estudo. Os experimentos de avaliação é que irão efetivamente ser utilizados para analisar um aplicativo VoIP sob os aspectos de qualidade de voz e adaptabilidade. Por isso, o tempo de execução de cada nível dos experimentos de calibração normalmente é pequeno, em torno de alguns minutos, pois o objetivo é apenas determinar quais são os pontos relevantes a serem utilizados na avaliação de desempenho. Este trabalho utilizou o tempo de 3 minutos, que é um tempo considerado adequado e foi o mesmo utilizado por Barbosa [1].

O conhecimento adquirido em [1] foi adaptado para a composição dos níveis dos experimentos de calibração. Também foram utilizadas configurações comuns à Internet, não incluindo configurações de rede inaceitáveis para VoIP, conforme discutido em trabalhos anteriores[1] [9] [32]. Além disso, dentre os experimentos de calibração, deve-se primeiramente executar os de capacidade, pois estes determinam a taxa transmitida máxima, que será utilizada como

informação para a configuração da capacidade do enlace a ser configurado pelos outros experimentos de calibração.

3.6.2.1 Experimentos de Calibração para o Serviço de Conferência

O Cenário 1.1 visa analisar o comportamento das aplicações VoIP quando a rede apresenta enlaces críticos de diferentes capacidades e determinar a taxa transmitida máxima. A capacidade variou de 100kbps até 10kbps, em decrementos de 1kbps, pois a taxa transmitida por estas aplicações está compreendida nessa faixa. Como o interesse deste cenário é apenas analisar o comportamento de acordo com a variação da capacidade, utilizou-se um atraso baixo (10ms), e a perda de pacotes e o *DelaySigma* (fator correspondente ao *jitter*) foram configurados em 0% e 0ms, respectivamente, pois o objetivo deste cenário é verificar o comportamento com relação à capacidade.

O Cenário 1.2 analisa o impacto do atraso. Assim, de acordo com o experimento de calibração relativo à capacidade, a capacidade unidirecional do enlace foi configurada de acordo com a taxa transmitida máxima. O atraso variou de 0ms até 1s, com incrementos de 25ms. A perda de pacotes e o *DelaySigma* foram configurados em 0% e 0ms, respectivamente, pois este cenário tem como objetivo verificar o comportamento com relação ao atraso.

O Cenário 1.3 analisa o impacto da perda de pacotes. A capacidade unidirecional do enlace foi configurada de acordo com a taxa transmitida máxima. O atraso foi de 10ms. A perda de pacotes variou de 1% até 40%, com incrementos de 1%, e o *DelaySigma* foi configurado em 0ms.

O Cenário 1.4 analisa o impacto do *jitter*. A capacidade foi configurada de acordo com a taxa transmitida máxima, obtida dos experimentos de calibração relativos à capacidade. A perda de pacotes foi configurada em 0%. O parâmetro *DelaySigma* variou de 0ms a 80ms, com acréscimos de 2ms. Ao contrário dos cenários de capacidade e perda, onde o atraso era de 10ms, neste cenário configurou-se um valor de 100ms. Este valor foi escolhido devido ao modelo de variação do atraso desenvolvido para o emulador, que permite alcançar valores maiores de *jitter* quanto maior seja o valor do atraso configurado.

A Tabela 3.5 resume os fatores e níveis utilizados nos cenários relativos aos experimentos de calibração para o serviço de conferência.

Tabela 3.5 Cenários relativos aos experimentos de calibração para o serviço de conferência.

Cenário	Fator	Níveis
1.1	Capacidade	100kbps para 10kbps, com decréscimo de 1kbps
1.2	Atraso	0ms a 1s, com acréscimos de 25ms
1.3	Perda de pacotes	0% a 40%, com acréscimo de 1%
1.4	<i>DelaySigma (jitter)</i>	0 a 80ms, com acréscimo de 2ms

3.6.2.2 Experimentos de Calibração para o Serviço de Convergência

Conforme discutido anteriormente, o serviço de convergência permite efetuar chamadas de um computador conectado à Internet a um telefone localizado na PSTN e vice-versa. Para permitir que um computador possa realizar uma chamada para um telefone é preciso que haja um elemento de rede conhecido como *gateway* VoIP, que recodifica a chamada para que essa possa ser transmitida através da PSTN até o terminal desejado. Quando se deseja realizar uma chamada de um telefone para um computador conectado à Internet, o *gateway* VoIP associa um número de telefone a este computador, de maneira que quando se efetuar uma chamada para este terminal, o *gateway* VoIP recodificará essa chamada, para que seja transmitida pela Internet até o computador desejado.

No entanto, não é possível configurar os parâmetros de rede relativos a este serviço, pois o acesso a este ambiente é restrito, o qual se convencionou chamar ao longo deste trabalho de ambiente não-controlado. Assim, é realizada uma avaliação do tipo caixa preta para este serviço, como segue: um participante localizado no ambiente controlado efetuará uma chamada ao serviço de convergência da Internet à PSTN a outro participante que também se encontra no ambiente controlado, e que possua o serviço de convergência da PSTN à Internet, conforme ilustrado na Figura 3.8. Um exemplo desta forma de avaliação é um participante que se encontra no ambiente controlado utilizar o serviço de convergência *SkypeOut* para fazer uma chamada para um número *SkypeIn*, associado a um outro participante que também esteja localizado no ambiente controlado.

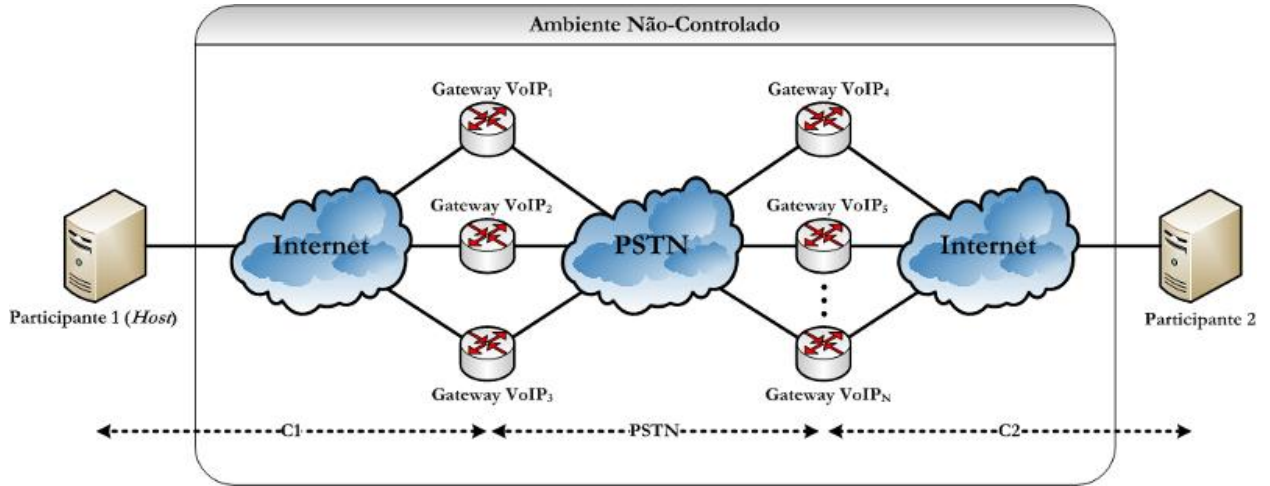


Figura 3.8 Caminho percorrido pela chamada durante a avaliação do serviço de convergência.

Contudo, esta abordagem impossibilita identificar o atraso, taxa de perda de pacotes e *jitter* do caminho percorrido pela chamada, pois uma vez que o pacote de origem é recodificado, não há como identificá-lo no destino. Assim, optou-se por realizar uma medição após o experimento, para obter uma estimativa apenas do atraso e da perda de pacotes, de modo que se tenha:

$$\begin{aligned} \text{atraso}_{\text{conv}} &= \text{atraso}_{C1} + \text{atraso}_{C2} + \text{atraso}_{\text{PSTN}} \\ \text{perda}_{\text{conv}} &= \text{perda}_{C1} + \text{perda}_{C2} + \text{perda}_{\text{PSTN}} \end{aligned}$$

onde:

$\text{atraso}_{\text{conv}}$ é o atraso total, isto é, o atraso relativo ao serviço de convergência,

atraso_{C1} e atraso_{C2} são os atrasos relativos aos caminhos C1 e C2, respectivamente,

$\text{atraso}_{\text{PSTN}}$ é o atraso relativo à PSTN,

$\text{perda}_{\text{conv}}$ é a perda total, isto é, a perda relativa ao serviço de convergência,

perda_{C1} e perda_{C2} são as perdas relativas aos caminhos C1 e C2, respectivamente, e

$\text{perda}_{\text{PSTN}}$ é a perda relativa à PSTN.

Como a PSTN é baseada em comutação por circuitos, o modo de avaliar as métricas de atraso, perda e *jitter* são diferentes se comparadas à comutação por pacotes. O atraso é, basicamente, a velocidade da luz acrescido do atraso ao se passar pelos dispositivos (centrais, etc) e dependem exclusivamente da distância e da quantidade de dispositivos que a chamada terá de percorrer. Não ocorre *jitter*, uma vez que não há enfileiramento de mensagens, causado pela comutação de pacotes. E não existe, a princípio, perda de informações na PSTN. Assim, por utilizarem os mesmos serviços de convergência, os parâmetros $\text{atraso}_{\text{PSTN}}$ e $\text{perda}_{\text{PSTN}}$ serão desprezados, simplificando a forma de cálculo das estimativas para:

$$\begin{aligned}\text{atraso}_{\text{conv}} &= \text{atraso}_{C1} + \text{atraso}_{C2} \\ \text{perda}_{\text{conv}} &= \text{perda}_{C1} + \text{perda}_{C2}\end{aligned}$$

O atraso_{C1} , atraso_{C2} , perda_{C1} , e perda_{C2} serão obtidos através do cômputo da estimativa do atraso e da perda de pacotes, através do programa *My Traceroute* (mtr). Os endereços IP dos *gateways* VoIP serão inferidos através do tráfego capturado dos participantes. A Tabela 3.6 relaciona o cenário relativo ao experimento de calibração e os fatores e níveis utilizados.

Tabela 3.6 Cenário relativo ao experimento de calibração para o serviço de convergência.

Cenário	Fator	Níveis
8.1	Participantes	2

3.6.3 Experimentos de Avaliação

Uma vez realizados os experimentos de calibração, os valores críticos dos cenários avaliados sob estes experimentos são utilizados para a elaboração dos experimentos de avaliação, efetuarão a avaliação de desempenho com as métricas estabelecidas neste trabalho: qualidade de voz e a adaptabilidade. Os objetivos dos experimentos de avaliação são determinar como os parâmetros de rede influenciam na qualidade de voz e na adaptabilidade dos aplicativos avaliados quando utilizados os serviços de conferência e convergência. Também deve ser analisada a correlação entre os fatores escolhidos, caso existam.

Cada nível em um cenário é executado durante uma hora e após isso, muda-se de nível dinamicamente, pois foi detectado que as aplicações mudam seu comportamento quando as condições de rede são estabelecidas antes da chamada ser realizada [1]. Isso é realizado através da modificação de parâmetros do emulador de rede *NIST.Net*. A fim de prover garantias estatísticas, os experimentos onde não há a presença do serviço de convergência duram 60 minutos e são divididos em experimentos menores de 1 minuto. Portanto, para cada nível e em um único experimento, foi possível obter amostras de tamanho 60. Nos experimentos onde há a presença do serviço de convergência duram 30 minutos e são divididos também em experimentos menores de 1 minuto, totalizando assim uma amostra de tamanho 30. Como a entrada de áudio e o nível têm a mesma duração (60 ou 30 minutos), diferentes níveis em um mesmo cenário e em cenários diferentes são submetidos a uma mesma entrada de áudio. Isso permitiu uma reprodução mais fiel dos experimentos e uma comparação mais justa das aplicações. Utilizou-se um nível de confiança assintótico de 95%, pois este fornecerá uma boa aproximação dos resultados de acordo com a quantidade de experimentos realizados.

3.7 Medindo Qualidade de Voz no Contexto de Conversação

Embora não exista ainda um padrão para mensurar de modo objetivo a qualidade de voz sob o contexto de conversação é preciso definir uma métrica que preencha essa lacuna. Assim, este trabalho propõe uma métrica denominada **PESQ MOSconf**, que através do PESQ MOS realiza uma estimativa do MOS em um cenário de conversação. A forma do cômputo da métrica PESQ MOSconf dependerá da topologia utilizada pelo aplicativo VoIP para mixagem e distribuição do áudio da sessão de conferência.

Assim, assumindo que uma sessão VoIP pode ter N participantes e que cada participante i, k é denotado por $P_{i,k} \in \{1, \dots, N\}$, temos:

$$\text{PESQ MOS}_{\text{conf}} = \begin{cases} \text{PESQ MOS}(P_1), & \text{se a topologia for do tipo mixagem via sistema} \\ & \text{final ou mixagem via servidor de conferência;} \\ \frac{\sum_{k=0}^{N-1} \text{PESQ MOS}_i(P_k)}{N-1}, & \text{se a topologia for do tipo } full\ mesh; \\ \frac{\sum_{j=0}^{N_s-1} \text{PESQ MOS}_i(S_j)}{N_s-1}, & \text{se a topologia for um híbrido entre a} \\ & \text{full mesh e mixagem via servidor de} \\ & \text{conferência.} \end{cases}$$

onde:

P_i é o participante *host*, isto é, o participante responsável por mixar e distribuir o áudio de uma sessão VoIP,

$\text{PESQ MOS}_i(P_k)$ é a pontuação PESQ MOS de P_k obtida em P_i ,

N_s é o número de servidores de conferência na topologia híbrida entre a topologia *full mesh* e a de mixagem via servidor de conferência, e

$S_j \in \{1, \dots, N_s\}$ representa cada servidor de conferência utilizada em uma topologia híbrida entre a topologia *full mesh* e a de mixagem via servidor de conferência.

É importante frisar que no caso da topologia ser do tipo mixagem via sistema final ou via servidor de conferência e o número de participantes for igual a 2, a métrica PESQ MOSconf e PESQ MOS serão idênticas. Além disso, a avaliação da qualidade do áudio será efetuada no participante *host*, que por convenção será o responsável por gravar o áudio da conferência e não emitirá áudio.

3.8 Medindo Jitter

A RFC 3393 [38] define o atraso unidirecional, ou simplesmente atraso, como a diferença entre o instante de saída de um pacote p da máquina de origem ao instante de chegada deste pacote p na

máquina de destino. Para obter este atraso, ambas as máquinas precisam ter os seus relógios sincronizados. O termo *jitter* refere-se à variação do atraso, com relação ao atraso médio.

Para o cálculo do atraso é preciso utilizar o relógio das máquinas usadas para capturar o tráfego. Cada relógio possui um erro inerente, que é como o comportamento do relógio difere de um relógio de referência (ideal). Este erro pode ser em fase (um determinado relógio está atrasado t milissegundos em relação a um relógio de referência, em um dado instante) ou em frequência (um relógio atrasa ou se adianta com o passar do tempo, em relação a um relógio de referência). Embora os atrasos da rede sejam imprevisíveis, é possível calcular o *jitter* independentemente da sincronização de relógios. Este método foi desenvolvido por Barbosa [1] e implementado através da ferramenta *IPstat*.

Esta ferramenta funciona da seguinte maneira: através de arquivos de tráfego capturados na máquina de origem e destino do fluxo, no formato da *libpcap* (como o *tcpdump*, *wireshark* ou *tsniff*) este tipo de arquivo contém o momento de captura de cada pacote (*timestamp*), e através do campo *identification* (ID) do cabeçalho do protocolo IP (este campo é usado para fins de fragmentação e pode ser usado como identificador único de um pacote) é identificado o pacote de origem no destino. O atraso é calculado como a diferença entre o *timestamp* da saída do pacote da origem com o *timestamp* de chegada deste pacote no destino. O *jitter*, por sua vez, é calculado como sendo a diferença entre duas medidas de atraso consecutivas. Caso haja a perda do pacote entre a origem e o destino, tanto o atraso quanto o *jitter* serão ditos indefinidos.

No entanto, é conhecido que o emulador *NIST.Net* utilizado na composição do ambiente controlado não permite a configuração direta do *jitter*. Entretanto, este emulador disponibiliza um parâmetro chamado *DelaySigma* que pode ser usado para configuração do *jitter*. De acordo com Barbosa [1], o parâmetro *DelaySigma* apresenta um coeficiente de correlação de 0,999961 com relação ao *jitter* médio, sendo portanto, uma boa aproximação para este.

3.9 Medindo Perda de Pacotes

A RFC 2680 [39] define a perda unidirecional, ou simplesmente perda, como uma falha que ocorre quando um pacote p da máquina de origem que atravessa uma rede não alcança a máquina de destino. Assim, através da ferramenta *IPstat* é possível calcular a perda de pacotes aproveitando a mesma abordagem usada para calcular o atraso e o *jitter*, isto é, utilizar o campo *identification* (ID) do cabeçalho do protocolo IP para identificar se o pacote que saiu da origem não chegou ao destino, pois se não houver um pacote no *trace* de destino com o mesmo ID no *trace* de origem, então é apontada a perda deste pacote.

4 Avaliação de Desempenho

“A verdadeira viagem de descoberta não está em procurar novas paisagens, mas em adquirir novos olhos.”

Marcel Proust, Em Busca do Tempo Perdido.

Este capítulo procura realizar a validação da metodologia proposta neste trabalho através da avaliação de desempenho do aplicativo *Skype* como estudo de caso. Embora existam outros aplicativos que agregam os serviços de conferência e convergência, como o *Yahoo! Messenger*, o *Skype* preterido neste trabalho por sua grande popularidade com relação aos outros aplicativos existentes e, por esta mesma razão, ser um dos mais estudados e analisados [1] [2] [5] [21] [32] [33].

4.1 Planejamento dos Experimentos

Seguindo a metodologia proposta neste trabalho, foram elaborados dois tipos de experimentos: os de calibração e os de avaliação. Os experimentos de calibração irão determinar os limiares críticos dos experimentos de avaliação e podem ser usados para inferir outras informações relevantes à condução desses experimentos, como o tipo de topologia usada pelo sistema VoIP, a taxa transmitida máxima e a qualidade de voz média. O tipo de topologia utilizada pode ser inferido através do estudo da distribuição do tráfego entre os participantes dos serviços de sessão de conferência e convergência, a taxa transmitida máxima será obtida através da análise do tráfego gerado e a qualidade de voz média será conseguida através do algoritmo PESQ MOS.

4.2 Elaboração da Carga de Trabalho

Para a realização dos experimentos de calibração e avaliação dos serviços de conferência e convergência foram construídos arquivos de áudio para dois, três e quatro participantes, como descrito na Seção 3.3. Esses arquivos serão utilizados nas máquinas correspondentes aos participantes no ambiente controlado, com o objetivo de emular uma sessão de conferência. Além disso, os arquivos de áudio dos participantes serão combinados e depois utilizados como arquivos de referência pelo PESQ para estimar a qualidade da voz do experimento realizado.

4.3 Tempo de Realização dos Experimentos de Calibração e Avaliação

Os experimentos relacionados ao serviço de conferência totalizam 26 horas de áudio ou 1560 minutos, onde foram abordados os cenários do tipo ponto a ponto e multiponto, incluindo participantes que estavam utilizando o serviço de convergência.

Os experimentos referentes ao serviço de convergência resultaram em um total de 1,5 horas de áudio, ou 90 minutos, onde foram abordados os cenários ponto a ponto e multiponto e a utilização de sessões de conferência onde existiam participantes utilizando o serviço e convergência e usuários diretamente à Internet.

4.4 Configuração do Ambiente Controlado

Após a elaboração dos arquivos de áudio utilizados como carga de trabalho, estes foram inseridos no ambiente controlado, com o objetivo de simular os contextos de audição e de conversação, conforme selecionados no planejamento dos experimentos. A configuração dos parâmetros de rede foram realizados pelo emulador *NIST.Net*, que simulavam as diversas condições de rede propostas nos cenários apresentados.

4.5 Experimentos de Calibração

Com o objetivo de determinar os limiares de rede críticos que irão guiar a composição dos cenários dos experimentos de avaliação, foram realizados os experimentos de calibração, que abordaram a utilização ponto a ponto (apenas 2 participantes), o serviço de conferência (3 ou mais participantes) e o serviço de convergência com a PSTN.

4.5.1 Experimentos de Calibração para o Serviço de Conferência

Seguindo a metodologia proposta neste trabalho, temos como primeiro passo efetuar o experimento de calibração referente à capacidade, pois através deste poderemos inferir a taxa transmitida máxima e computar o PESQ MOS médio máximo.

O Cenário 1.1 visa analisar o comportamento do *Skype* quando a rede apresenta enlaces críticos de diferentes capacidades e determinar a taxa transmitida máxima. A capacidade foi variada de 100kbps até 10kbps, em decrementos de 1kbps, pois a taxa transmitida pelas aplicações VoIP está compreendida nesta faixa. Como o interesse deste cenário é apenas analisar o comportamento de acordo com a variação da capacidade, utilizou-se um atraso baixo (10ms), e a perda de pacotes e o *DelaySigma* (fator correspondente ao *jitter*) foram configurados em 0% e 0ms, respectivamente.

De acordo com os resultados obtidos neste experimento, conclui-se que a taxa transmitida média do *Skype* entre dois participantes foi de aproximadamente 70kbps, conforme ilustra a Figura

4.1 (a). Com esta informação, configurou-se a capacidade dos outros cenários dos experimentos de calibração em 75kbps, tendo em vista que este era o menor limite superior a não ser atingindo pela taxa transmitida. A pontuação do PESQ MOS variou entre 3,5 e 4, totalizando assim um PESQ MOS médio máximo de 3,89 (qualidade de voz regular, de acordo com a Tabela 2.1), que pode ser verificado através da Figura 4.1 (b). Assim, para a capacidade configurada foi variada de 100 a 30 kbps em decrementos de 10 kbps.

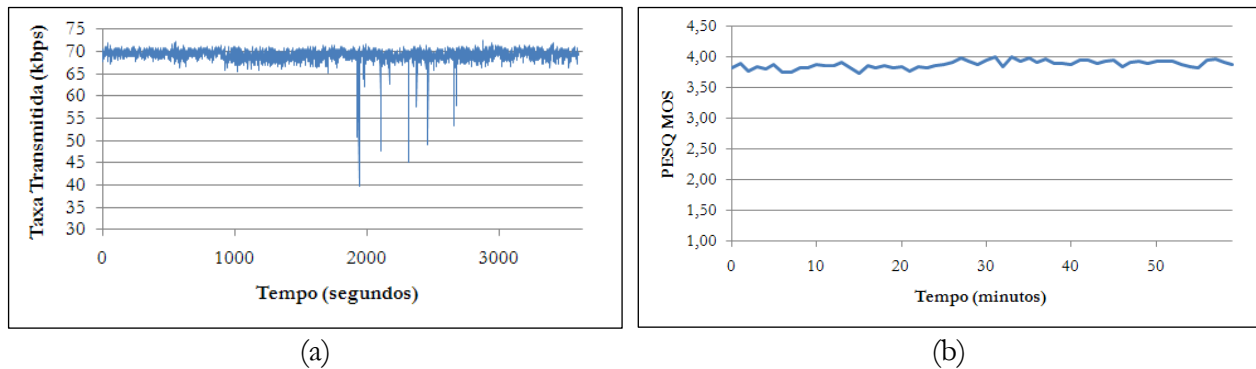


Figura 4.1 (a) Taxa transmitida e (b) PESQ MOS obtidos através da análise da variação da capacidade (Cenário 1.1).

O Cenário 1.2 analisa o impacto do atraso. A capacidade foi configurada em 75kbps e o atraso foi variado de 0ms até 1000ms, com incrementos de 25ms. A perda de pacotes e o *DelaySigma* foram configurados em 0% e 0ms, respectivamente.

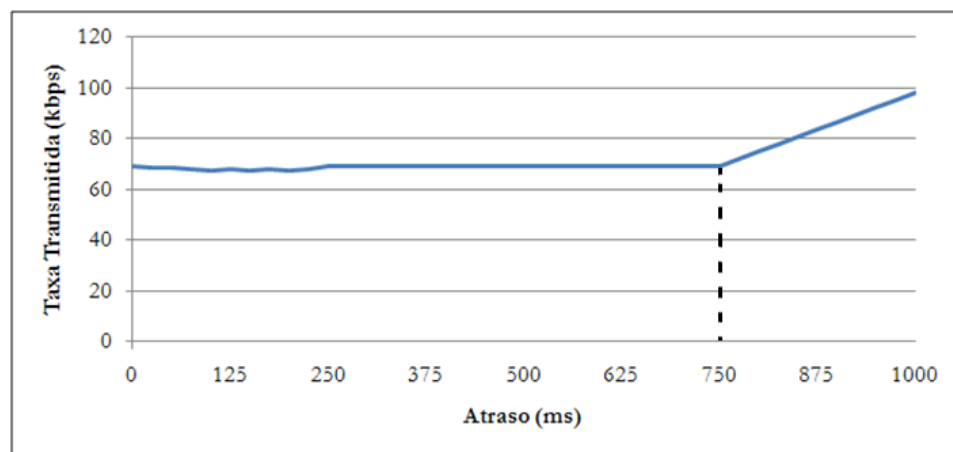


Figura 4.2 Taxa transmitida durante o experimento de calibração do atraso (Cenário 1.2).

Através da Figura 4.2 é possível concluir que a taxa de transmitida apresenta leve oscilação até atingir o atraso de 750ms, onde começa a ter um aumento substancial. Assim, para o atraso foram escolhidos os valores 10, 100, 250, 350, e 500ms, comuns na literatura [1], e a inclusão do atraso de 750 e 1000ms para a composição dos experimentos de avaliação.

O Cenário 1.3 analisa o impacto da perda de pacotes. A capacidade foi configurada em 75kbps, o atraso foi de 10ms, a perda de pacotes foi variada de 1% até 40%, com incrementos de 1%, e o *DelaySigma* foi configurado 0ms.

Por meio da Figura 4.3 conclui-se que a taxa transmitida sofre uma leve baixa quando a taxa de perda de pacotes está entre 1 e 2%. Após este intervalo, há um aumento linear na taxa transmitida à medida que a taxa de perda de pacotes aumenta até o valor de 30%, que a partir deste limite, há uma queda abrupta da taxa transmitida.

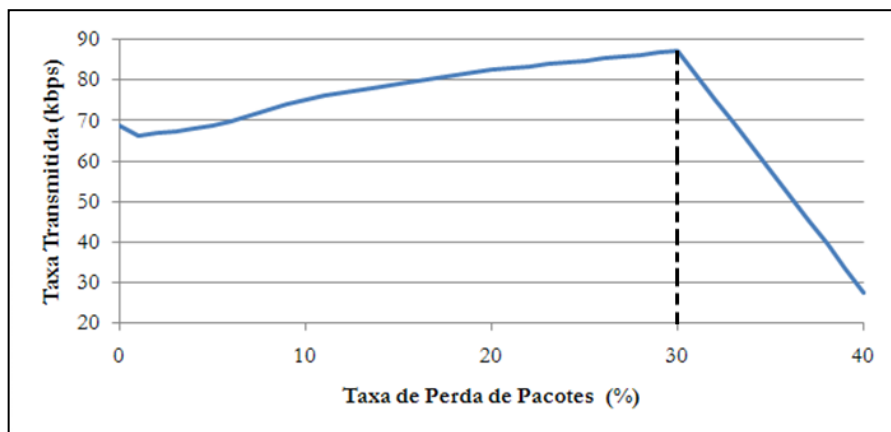


Figura 4.3 Taxa transmitida durante o experimento de calibração da taxa de perda de pacotes (Cenário 1.3).

Observando os resultados obtidos, nota-se que há uma leve queda na taxa transmitida quando a taxa de perda de pacotes atinge 1%. A partir deste valor, a taxa transmitida aumenta à medida que se aumenta a taxa de perda de pacotes até chegar o limiar de 30%, onde começa a ocorrer uma queda abrupta. Assim, os valores selecionados para a taxa de perda de pacotes utilizados na composição dos experimentos de avaliação foram 1%, 5%, 10%, 20%, 30% e 40%.

O Cenário 1.4 analisa o impacto do *jitter*. A capacidade foi configurada em 75kbps. A perda de pacotes foi configurada em 0%. O parâmetro *DelaySigma* foi variado de 0ms a 80ms, com acréscimos de 2ms. Ao contrário dos cenários de capacidade e perda, onde o atraso era de 10ms, neste cenário configurou-se um valor de 100ms. Esse valor foi escolhido devido ao modelo de variação do atraso desenvolvido no emulador, pois os níveis escolhidos para o *DelaySigma* só poderiam ser atingidos se o atraso fosse maior que o maior variação do atraso correspondente.

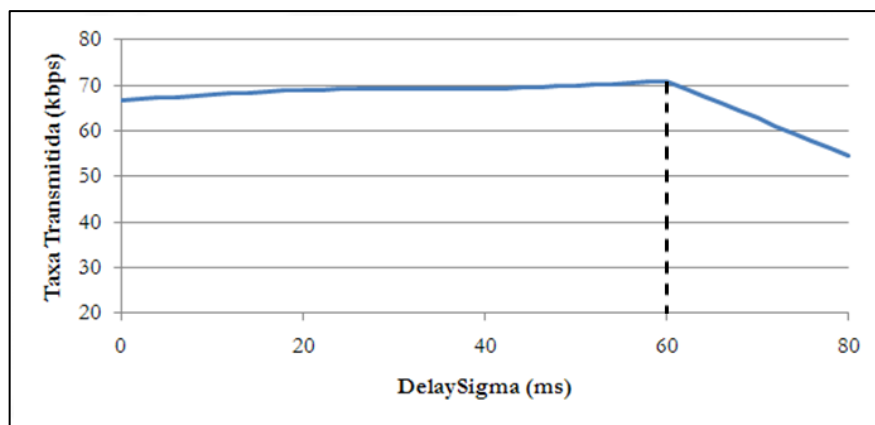


Figura 4.4 Taxa transmitida durante o experimento de calibração do *jitter* (Cenário 1.2).

Através da Figura 4.4 nota-se que há uma leve tendência de crescimento da taxa transmitida quando o *DelaySigma* é configurado entre 0 e 20ms e uma queda significativa quando o mesmo ultrapassa o valor de 60ms. Assim, foram selecionados para o *DelaySigma* os valores 0, 20, 40, 60 e 80ms para a composição dos experimentos de avaliação do *jitter*.

4.5.2 Experimentos de Calibração para o Serviço de Convergência

De acordo com a metodologia proposta neste trabalho, temos como primeiro passo efetuar o experimento de calibração referente à capacidade, pois através deste poderemos inferir a taxa transmitida máxima e como computar o PESQ MOS médio máximo.

Contudo, devido aos custos financeiros envolvidos, não foram realizados todos os experimentos de calibração relativos ao serviço de convergência. Portanto, optou-se por realizar apenas a medição da taxa transmitida máxima e o PESQ MOS médio máximo relativo à utilização deste serviço, ou seja, não houve configuração dos parâmetros de rede através do *NIST.Net* (Cenário 8.1).

Através da análise dos dados resultantes deste experimento, conclui-se que a taxa transmitida média do *Skype* entre dois participantes do serviço de convergência foi de aproximadamente 27kbps, conforme ilustra a Figura 4.5 (a). Com esta informação, configurou-se a capacidade dos outros cenários dos experimentos de calibração do serviço de convergência em 30kbps, tendo em vista que este limite não foi atingido. A pontuação do PESQ MOS variou entre 3 e 3,5, totalizando assim um PESQ MOS médio máximo de 3,71 (qualidade de voz regular, de acordo com a Tabela 2.1), que pode ser verificado através da Figura 4.5 (b).

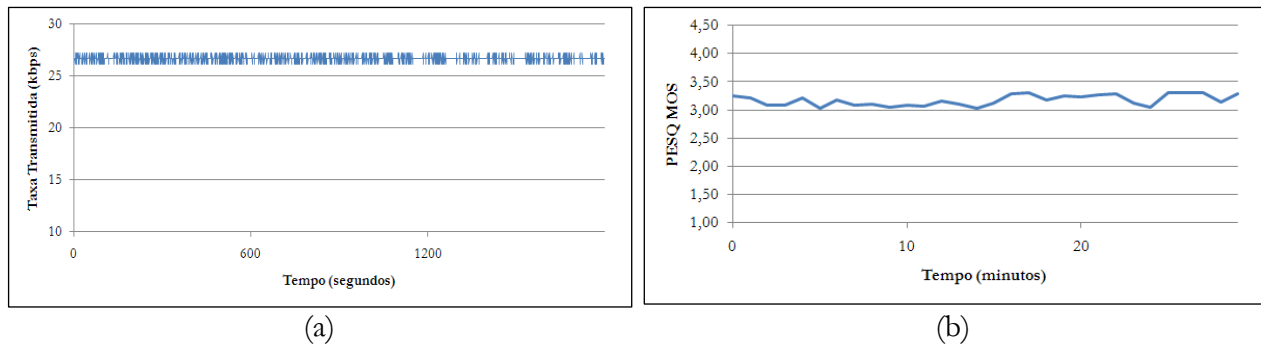


Figura 4.5 (a) Taxa transmitida e (b) PESQ MOS obtidos através do experimento de calibração 8.1, referentes ao serviço de convergência.

Seguindo a metodologia proposta, através da análise do tráfego dos participantes envolvidos neste experimento, extraiu-se os endereços IP referentes aos *gateways* VoIP envolvidos na chamada. O participante emissor do áudio utilizou o *gateway* VoIP 4.53.80.230 e o participante responsável por gravar o áudio se comunicava com o *gateway* VoIP 4.53.80.103. A partir dessa informação, foi realizada a estimativa do atraso e da taxa de perda de pacotes entre cada participante e o seu respectivo *gateway* VoIP, conforme ilustrado na Tabela 4.1.

Tabela 4.1 Estimativa das métricas de rede entre dois participantes do serviço de convergência através do experimento de calibração.

Gateway VoIP	Número de Pacotes	Salto	Taxa de Perda de Pacotes	Atraso
4.53.80.230	1000	10	0,1%	71ms
4.53.80.103	1000	14	0,1%	77ms

4.6 Topologia para a Mixagem e Distribuição do Áudio

Através dos experimentos de calibração foi possível determinar qual o tipo de topologia utilizado pelo *Skype* para a mixagem e distribuição de áudio entre os participantes do sistema de conferência. Utilizando-se dos *traces* gerados pelos experimentos concluiu-se que o *Skype* usa a topologia do tipo mixagem via sistema final, tendo em vista que não há tráfego entre os participantes durante o serviço de conferência e que este mesmo tráfego é concentrado no participante que iniciou a chamada (participante *host*), que por sua vez, realiza a mixagem e a redistribuição entre os outros participantes. Assim, de acordo com a Seção 3.7, a forma de calcular o PESQ MOS_{conf}, que é a métrica para efetuar o cômputo do serviço de conferência é:

$$\text{PESQ MOS}_{\text{conf}} = \text{PESQ MOS}(P_1)$$

onde:

P_i é o participante *host*, isto é, o participante responsável por mixar e distribuir o áudio de uma sessão VoIP.

4.7 Experimentos de Avaliação

Seguindo a metodologia proposta neste trabalho, após a realização dos experimentos de calibração, foram determinados os valores utilizados para a composição dos cenários dos experimentos de avaliação, que realizarão a avaliação de desempenho propriamente dita. Esses cenários visam analisar a influência dos parâmetros de rede na qualidade de voz e na adaptabilidade no serviço de conferência, no serviço de convergência, na utilização de ambos os serviços.

É importante ainda frisar que a finalidade dos experimentos onde há a variação dos parâmetros de rede de apenas um participante é avaliar o impacto da qualidade de voz da conferência, quando apenas as condições de rede deste participante variar. Esses cenários procuram refletir certas situações do mundo real, onde é possível durante uma conferência ter um dos participantes conectado através de uma rede sem fio, ou localizado geograficamente mais distante com relação aos outros participantes da conferência (apresenta um atraso e um *jitter* maior).

Todos os cenários analisados nesta seção tiveram os seus fatores variando simetricamente, isto é, igualmente para todos os participantes, de forma que todos teriam assim as mesmas condições de rede. Os objetivos destes cenários são analisar como o fator escolhido para cada cenário e o número de participantes influenciam nas métricas escolhidas e se há correlação entre eles.

4.7.1 Avaliação dos Cenários do Serviço de Conferência

Os Cenários 1.5, 2.1 e 2.2 analisam o comportamento das aplicações VoIP quando a rede apresenta enlaces críticos de diferentes capacidades. O número de participantes foi variado de 2 a 4 e a capacidade foi variada de 100kbps até 30 kbps, em decrementos de 10kbps, pois a taxa transmitida pelas aplicações VoIP está compreendida nessa faixa. Como o interesse deste cenário é apenas analisar o comportamento de acordo com a variação da capacidade, utilizou-se um atraso baixo (10ms), e a perda de pacotes e o *jitter* foram configurados em 0% e 0ms, respectivamente.

Os Cenários 1.6, 2.3 e 2.4 analisam o impacto do atraso. Assim, a capacidade residual de cada enlace foi fixada em 75kbps e o número de participantes variou de 2 a 4, com acréscimos de 1 participante. A perda de pacotes e o *jitter* foram configurados em 0% e 0ms, respectivamente. Os valores utilizados na variação do atraso foram de 10ms, 100ms, 250ms, 350ms, 500ms, 750ms e 1000ms.

Os Cenários 1.7, 2.5 e 2.6 analisam o comportamento sob diferentes taxas de perda de pacotes. Nestes cenários o número de participantes foi variado de 2 a 4, e a capacidade, o atraso e o

jitter foram configurados com 75kbps, 10ms e 0ms, respectivamente. Os valores utilizados na variação da perda de pacotes foram 0%, 1%, 5%, 10%, 20%, 30% e 40%.

Os Cenários 1.8, 2.7 e 2.8 analisam o impacto do *jitter*. O número de participantes foi variado de 2 a 4, com acréscimo de 1 participante, a capacidade foi fixada em 75kbps e a perda de pacotes foi configurada em 0%. Ao contrário dos cenários de capacidade e perda, onde o atraso era de 10ms, neste cenário configurou-se um valor de 100ms. Este valor foi escolhido devido ao modelo de variação do atraso desenvolvido para o emulador, que permite alcançar valores maiores de *jitter* quanto maior seja o valor do atraso configurado. O fator *DelaySigma* variou de 0ms a 80ms com incrementos de 20ms.

A Tabela 4.2 apresenta um resumo dos cenários, os fatores e os níveis que foram configurados igualmente para todos os participantes.

Tabela 4.2 Cenários relativos aos experimentos de avaliação para o serviço de conferência com os parâmetros de rede variando igualmente entre os participantes.

Cenários	Fatores	Níveis
1.5, 2.1, 2.2	Participantes	2, 3, 4
	Capacidade	100kbps, 90kbps, 80kbps, 70kbps, 60kbps, 50kbps, 40kbps, 30kbps
1.6, 2.3, 2.4	Participantes	2, 3, 4
	Atraso	10ms, 100ms, 250ms, 350ms, 500ms, 750ms, 1000ms
1.7, 2.5, 2.6	Participantes	2, 3, 4
	Perda	0%, 1%, 5%, 10%, 20%, 30%, 40%
1.8, 2.7, 2.8	Participantes	2, 3, 4
	<i>DelaySigma (Jitter)</i>	0ms, 20ms, 40ms, 60ms, 80ms

4.7.1.1 Avaliação do Impacto do Número de Participantes

O gráfico apresentado na Figura 4.6 indica que há uma diminuição na qualidade de voz da conferência à medida que aumentamos o número de participantes. A diferença entre os cenários de

2 e 3 participantes é de 3,45%, enquanto a diferença entre os cenários de 3 e 4 participantes é de 2,07%, o que pode indicar que a degradação da qualidade de voz entre os cenários de conferência, na média, não apresenta uma tendência linear.

Infelizmente, não foi possível identificar com precisão o motivo desta diminuição da qualidade de voz à medida que o número de participantes também aumentava. Uma possibilidade é que esta queda esteja relacionada ao aumento do ruído gerado pelos participantes, uma vez que aumentando a quantidade de participantes, o ruído também aumenta, degradando assim a qualidade final do áudio.

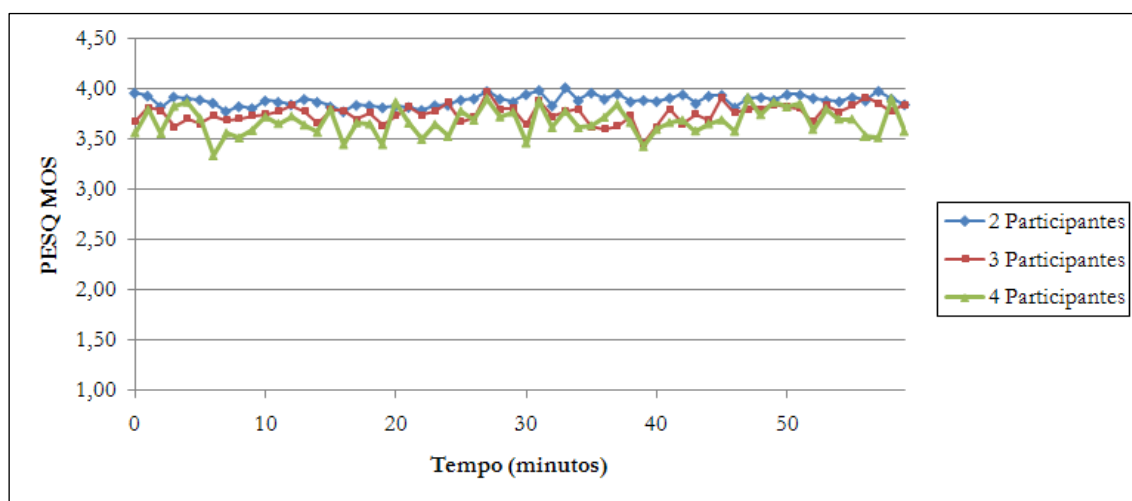


Figura 4.6 Qualidade de voz para os Cenários 1.5, 2.1 e 2.2, com a capacidade configurada em 100kbps.

4.7.1.2 Avaliação do Impacto da Capacidade

A Figura 4.7 apresenta o comportamento da taxa transmitida do *Skype*, com relação à capacidade configurada, quando foi variada a capacidade de 100kbps até 30kbps e o número de participantes de 2 a 4. Através do gráfico, nota-se que no intervalo de 100kbps até 70kbps a taxa transmitida se mantém uniforme e proporcional ao número de participantes. No entanto, quando a capacidade é configurada para um valor menor do que 70kbps, nota-se uma queda acentuada na taxa transmitida em todos os cenários, o que corrobora com o resultado obtido do experimento de calibração, onde se apontou que a taxa transmitida média utilizada pelo *Skype* é de cerca de 70kbps.

Verifica-se que a taxa transmitida sofre um leve aumento em relação ao número de participantes. Quando o número de participantes aumentou de 2 para 3 houve um aumento de cerca de 7% entre as capacidades de 100kbps e 60kbps e 15% entre 60kbps e 30kbps. Quando o número de participantes aumentou de 3 para 4, ocorreu um aumento de cerca de 13% entre 100kbps e 60kbps e 8% entre 60kbps e 30kbps. Infelizmente, não foi possível identificar a causa do aumento

da quantidade do tráfego, tendo em vista que os pacotes trafegados são criptografados, sendo impossível fazer a distinção dos pacotes referentes à mídia ou de controle.

Observa-se que quando a capacidade configurada fica abaixo dos 70kbps, a taxa transmitida pelo *Skype* também diminui, indicando que o mesmo procura adaptar-se às novas condições de rede impostas. É importante frisar que os resultados no cenário de 2 participantes são diferentes dos obtidos por Barbosa [1], onde ele aponta que a taxa transmitida máxima do *Skype* é de 50kbps, enquanto a encontrada neste trabalho foi de 75kbps. Essa diferença é devida a troca do *codec* utilizado pelo *Skype*, que na versão analisada por aquele trabalho utilizava o *codec* iSAC, enquanto a versão aqui analisada usa o *codec* SVOPC.

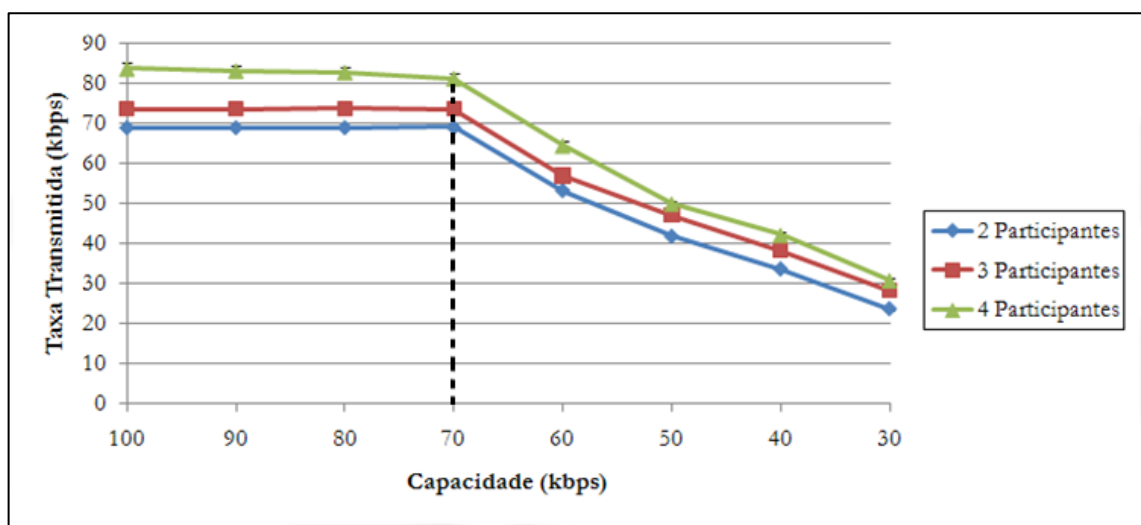


Figura 4.7 Taxa transmitida entre os participantes do serviço de conferência, quando varia a capacidade configurada entre os participantes e o número de participantes (Cenários 1.5, 2.1 e 2.2).

A Figura 4.8 ilustra a avaliação da qualidade de voz ao longo da variação da capacidade e do número de participantes da sessão de conferência. Através dele, nota-se que quando a capacidade configurada diminui além de 70kbps, há uma queda na diminuição da qualidade que continua à medida que a capacidade configurada também diminui. Este fato corrobora com a afirmação que a taxa transmitida média do *Skype* esteja em torno de 70kbps. Além disso, à medida que o número de participantes aumenta, há sempre uma queda gradativa da qualidade de voz, confirmando que à medida que se aumenta o número de participantes, a quantidade de tráfego gerado também aumenta, o que acaba impactando na qualidade de voz.

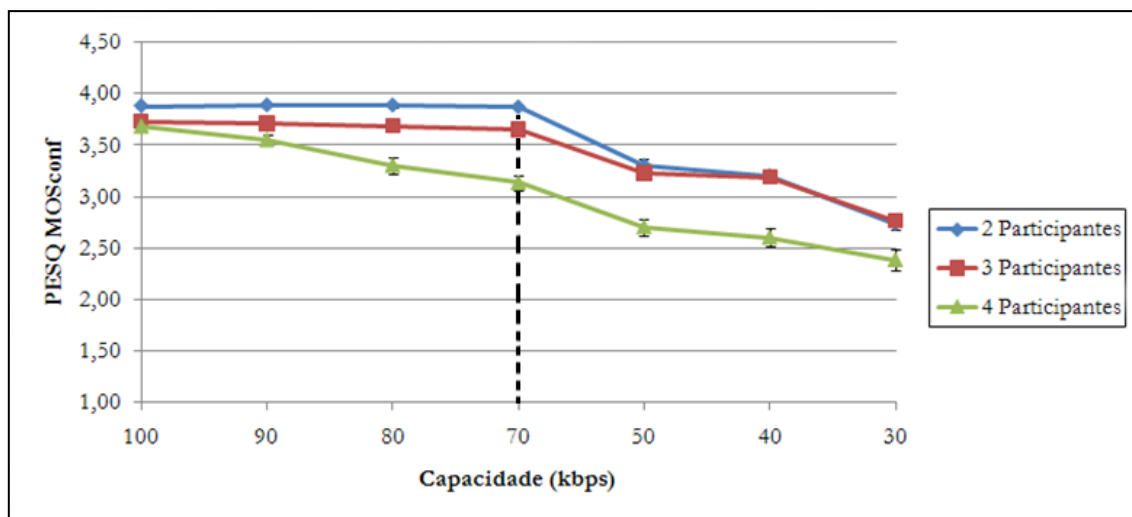


Figura 4.8 Qualidade de voz entre os participantes do serviço de conferência, quando varia a capacidade configurada entre os participantes e o número de participantes (Cenários 1.5, 2.1 e 2.2).

4.7.1.3 Avaliação do Impacto do Atraso

A Figura 4.9 ilustra a evolução da taxa transmitida do *Skype* com relação à variação do atraso entre os participantes de uma sessão de conferência. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelos participantes 1 e 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

Através da figura, conclui-se que o *Skype* apresenta a mesma taxa de transmissão ao longo da variação do atraso, até que o atraso chegue a 750ms, quando ocorre um aumento substancial, em torno de 40%, quando o cenário é composto de apenas 2 participantes e, em torno de 60%, quando o cenário é composto de 3 a 4 participantes. Além disso, ao contrário das versões anteriores do *Skype*, que ao ultrapassar o limiar de 100ms de atraso, diminuía bruscamente a taxa transmitida, esta versão mantém sua taxa até atingir os 750ms, passando a aumentá-la quando transpõe este limiar. A razão deste aumento pode ser que ao ultrapassar o limiar de 750ms de atraso o *Skype* supõe que esteja havendo uma perda de pacotes e assim tente compensá-la, retransmitindo os pacotes, implicando no aumento da taxa transmitida.

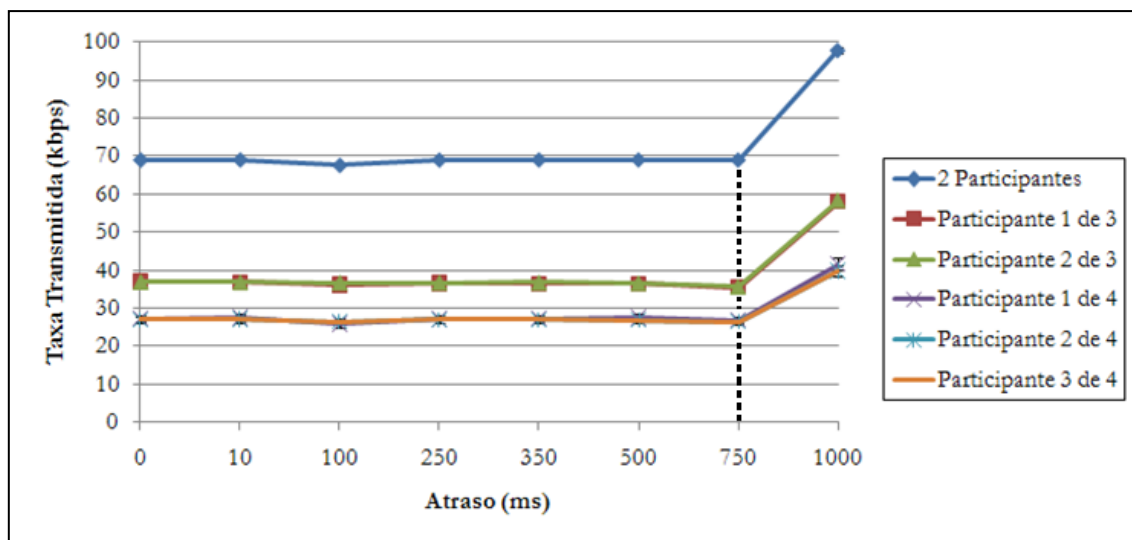


Figura 4.9 Taxa transmitida entre os participantes do serviço de conferência, quando varia o atraso entre os participantes e o número de participantes (Cenários 1.6, 2.3 e 2.4).

A Figura 4.10 apresenta a qualidade de voz da sessão de conferência de acordo com o aumento do atraso e do número de participantes. Quando utilizados 2 ou 3 participantes, verifica-se que o *Skype* mantém a qualidade de voz da sessão de conferência com qualidade satisfatória (pontuação MOS entre 3,60 a 4,03) até atingir um atraso de cerca de 500ms. A partir deste limiar, temos uma queda acentuada da qualidade de voz, atingindo a qualidade pobre (pontuação entre 3,10 a 3,60), sem que haja aumento na taxa transmitida. Uma possível explicação para este fenômeno é que devido ao aumento do atraso entre os participantes esteja havendo sobreposição das vozes no contexto de conversação, o que degradar a qualidade de voz medida através do PESQ. Fato este que pode ser confirmado pela diminuição na qualidade de voz no cenário onde existem 4 participantes a partir do atraso de 350ms, pois com o aumento do número de participantes, há um aumento na probabilidade de sobreposição de vozes. Ainda assim, nota-se uma evolução com relação às versões anteriores do *Skype*, que começavam a apresentar uma queda na qualidade de voz a partir dos 100ms de atraso.

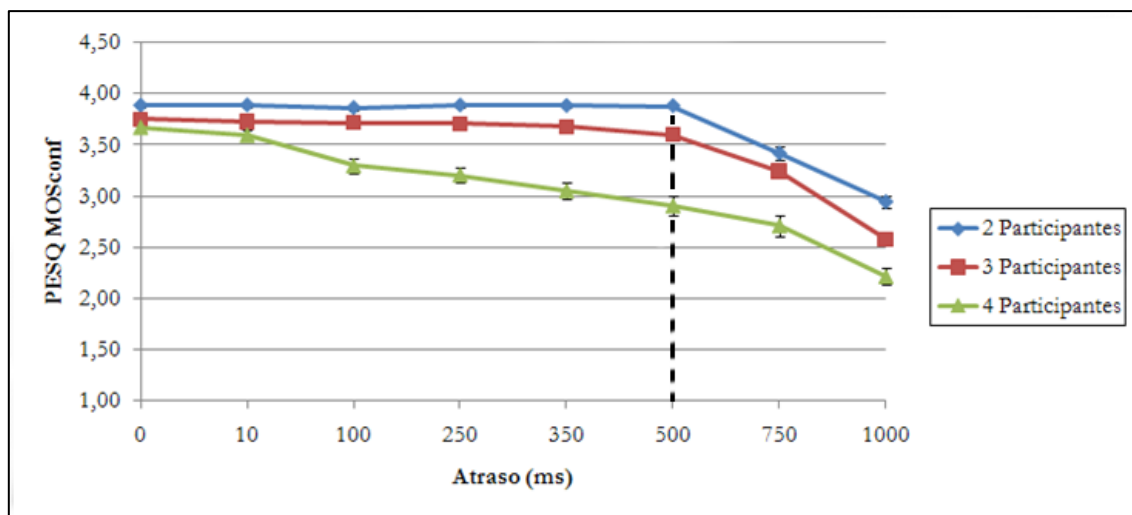


Figura 4.10 Qualidade de voz entre os participantes do serviço de conferência, quando varia a capacidade configurada entre os participantes e o número de participantes (Cenários 1.6, 2.3 e 2.4).

4.7.1.4 Avaliação do Impacto da Perda de Pacotes

A Figura 4.11 ilustra o desempenho da taxa transmitida do *Skype* com relação à variação da perda de pacotes e o aumento no número de participantes de uma sessão de conferência. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelo participante 1 e participante 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

Através do gráfico, é possível perceber um aumento da taxa transmitida quando a perda de pacotes ultrapassa 5%. A partir deste limiar, a taxa transmitida continua a crescer até atingir 30% de perda de pacotes. Além desse limiar, ocorre uma queda muito acentuada. Uma possível explicação para este fato é que ao perceber que a perda de pacotes é muito alta, o *Skype* entende que é irrelevante efetuar a retransmissão dos pacotes, optando assim por suspender a transmissão, aguardando uma possível recuperação do enlace para que ele volte a transmitir.

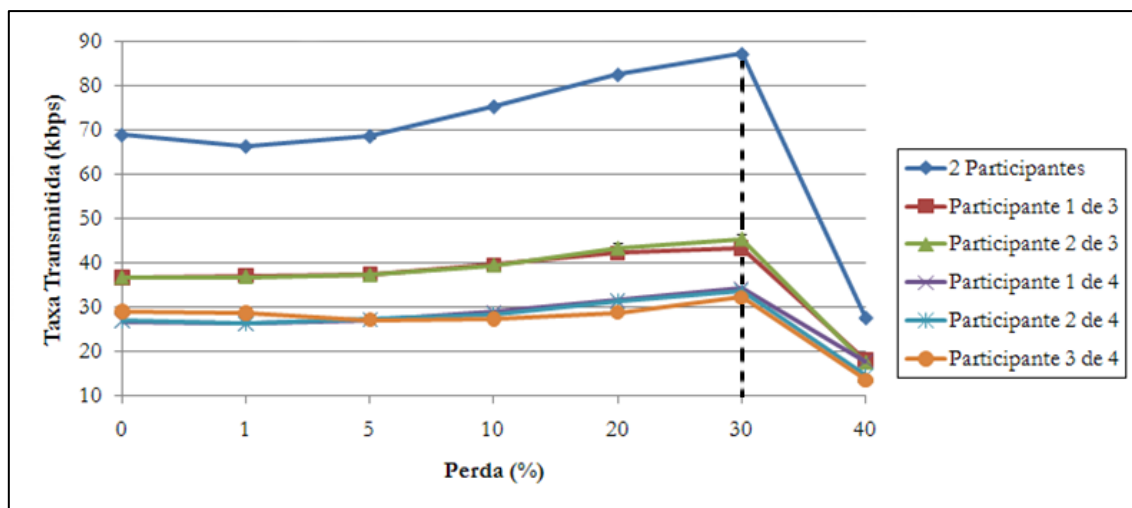


Figura 4.11 Taxa transmitida entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes entre os participantes e o número de participantes (Cenários 1.7, 2.5 e 2.6).

A Figura 4.12 representa a qualidade de voz de uma sessão de conferência de acordo com o aumento da perda de pacotes e do número de participantes. Percebe-se que há uma queda na taxa transmitida de aproximadamente 11% quando a perda de pacotes é de 1%, e de 21% quando a perda de pacotes aumenta para 5%. Após este limiar, a qualidade de voz se estabiliza entre 5% e 20% de perda de pacotes, só ocorrendo nova queda na taxa transmitida ao se aproximar dos 30% de perda de pacotes.

É importante notar que esta versão do *Skype* apresenta uma maior resistência à perda de pacotes com relação às versões anteriores, quando comparados os resultados obtidos neste trabalho com os obtidos por Barbosa [1]. Este fato é comprovado pelo fato de que a qualidade deixa de ser satisfatória somente a partir de 30% de perda, contra o pouco mais de 10% de perda, conforme apontado por aquele estudo. Este fato é devido ao novo *codec* utilizado por esta versão do *Skype*, o SVOPC, que foi projetado para ser mais resistente à perda de pacotes. Acredita-se que esta abordagem esteja relacionada ao fato da *Skype S.A.* planejam aumentar a participação do seu produto no mercado de dispositivos sem fio, principalmente telefones celulares e *smartphones*, conforme coberto pela mídia [73] [74].

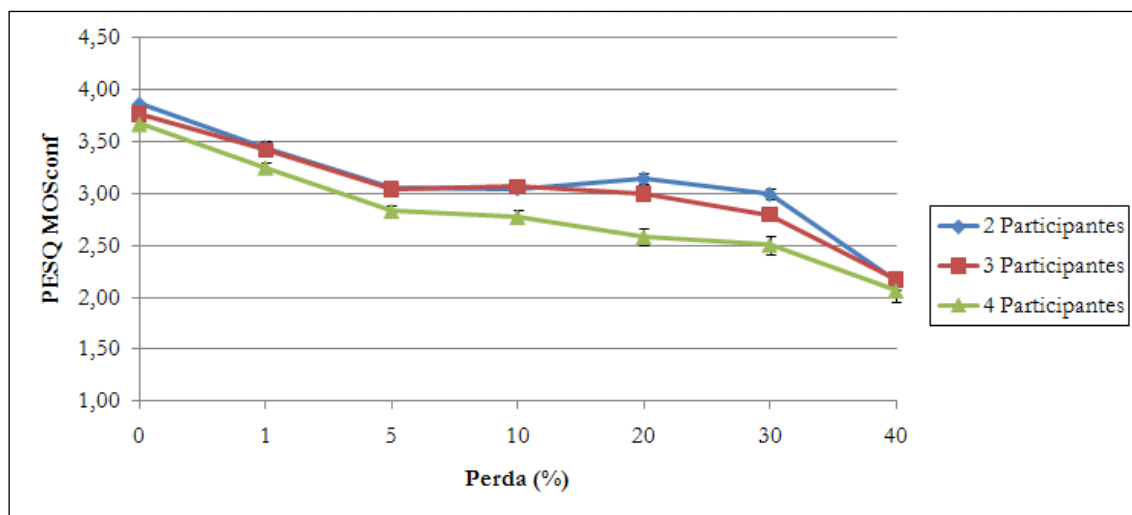


Figura 4.12 Qualidade de voz entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes entre os participantes e o número de participantes (Cenários 1.7, 2.5 e 2.6).

4.7.1.5 Avaliação do Impacto do *Jitter*

A Figura 4.13 ilustra o comportamento da taxa transmitida do *Skype* com relação ao *jitter* (variação do atraso) e o aumento no número de participantes de uma sessão de conferência. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelos participantes 1 e 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

Ocorre um aumento discreto da taxa transmitida quando o *jitter* aumenta de 0ms para 20ms. Entre 20 e 60ms a taxa transmitida permanece estável, até que ultrapassa o patamar de 60ms, onde ocorre uma queda brusca. A explicação para este fenômeno é que semelhante ao que ocorre na avaliação da variação do atraso: ao perceber que o *jitter* é alto, o *codec* diminui a sua transmissão, esperando que as condições do enlace melhorem para continuar a transmissão.

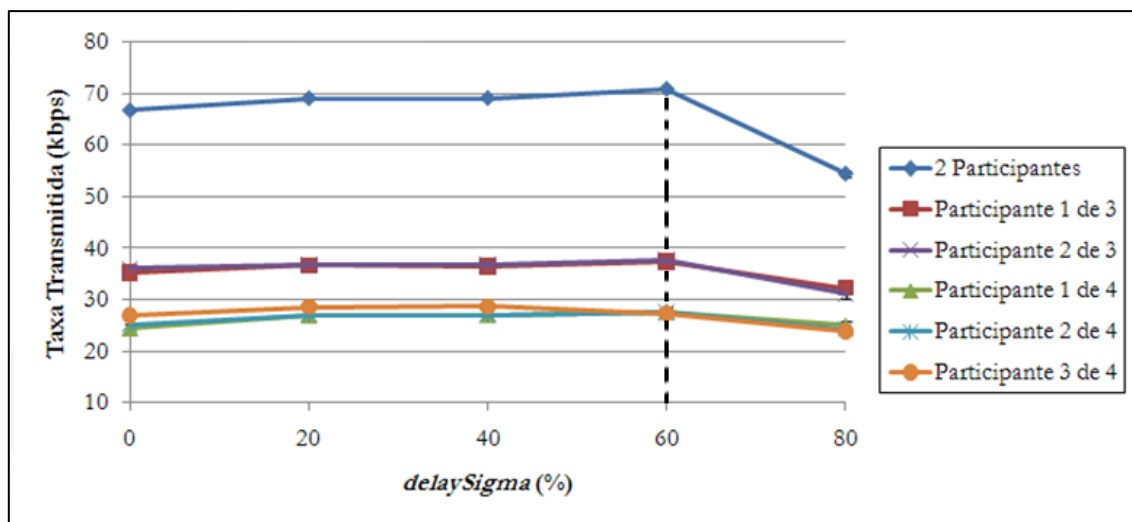


Figura 4.13 Taxa transmitida entre os participantes do serviço de conferência, quando varia o *DelaySigma* (*jitter*) entre os participantes e o número de participantes (Cenários 1.8, 2.7 e 2.8).

A Figura 4.14 representa a qualidade de voz de uma sessão de conferência de acordo com o aumento do *jitter* e do número de participantes. Percebe-se que há uma queda da qualidade da voz à medida que há o aumento do *jitter* e que esta queda também é proporcional ao aumento dos participantes da sessão de conferência. Isto se deve pelo menos motivo do que apresentado no experimento de avaliação do atraso: o aumento do *jitter* e do número de participantes provam a sobreposição das vozes dos participantes, diminuindo assim, a qualidade final da sessão de conferência.

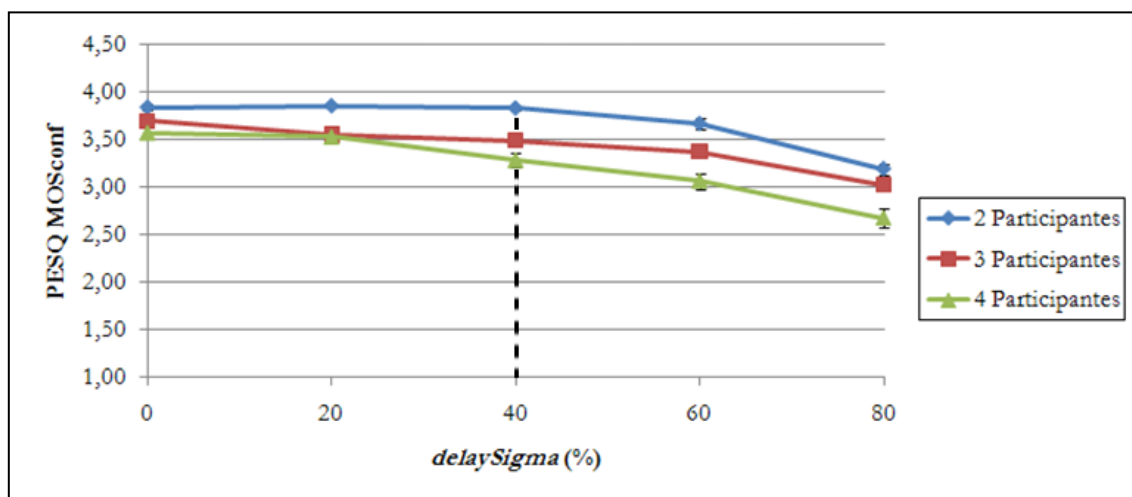


Figura 4.14 Qualidade de voz entre os participantes do serviço de conferência, quando varia *DelaySigma* (*jitter*) entre os participantes e o número de participantes (Cenários 1.8, 2.7 e 2.8).

4.7.2 Avaliação dos Cenários de Correlação entre Métricas

Todos os cenários analisados na Seção 4.7.1 variaram igualmente entre todos participantes da conferência. Nesta seção, a configuração entre os participantes foram alteradas, de maneira que

apresentassem diferentes configurações de rede entre si. Os objetivos destes cenários são analisar como o fator escolhido para cada cenário, e o número de participantes influencia nas métricas escolhidas e se há correlação entre eles.

Os Cenários 2.9 e 2.10 analisam o comportamento das aplicações VoIP quando a rede apresenta enlaces críticos de diferentes capacidades. O número de participantes variou de 3 a 4 e a capacidade de apenas um participante variou de 100kbps até 30 kbps, em decrementos de 10kbps, pois a taxa transmitida por estas aplicações está compreendida nesta faixa. Como o interesse deste cenário é apenas analisar o comportamento de acordo com a variação da capacidade, utilizou-se um atraso baixo (10ms), e a perda e de pacotes e o *jitter* foram configurados em 0% e 0ms, respectivamente.

Os Cenários 2.11 e 2.12 analisam o impacto do atraso. Assim, a capacidade residual de cada enlace foi fixada em 75kbps e o número de participantes variou de 3 a 4. A perda de pacotes e o *jitter* foram configurados em 0% e 0ms, respectivamente. Os valores utilizados na variação do atraso de apenas um participante foram de 10ms, 100ms, 250ms, 350ms, 500ms, 750ms e 1000ms.

Os Cenários 2.13 e 2.14 analisam o comportamento das aplicações sob diferentes taxas de perda de pacotes. Nestes cenários o número de participantes variou de 3 a 4 e a capacidade, o atraso e o *jitter* foram configurados com 75kbps, 10ms e 0ms, respectivamente. Os valores utilizados na variação da perda de pacotes de apenas um participante foram 0%, 1%, 5%, 10%, 20%, 30% e 40%.

Os Cenários 2.15 e 2.16 analisam o impacto do *jitter*. O número de participantes variou de 3 a 4, a capacidade foi fixada em 75kbps e a perda de pacotes foi configurada em 0%. Ao contrário dos cenários de capacidade e perda, onde o atraso era de 10ms, neste cenário configurou-se um valor de 100ms. Este valor foi escolhido devido ao modelo de variação do atraso desenvolvido para o emulador, que permite alcançar valores maiores de *jitter* quanto maior seja o valor do atraso configurado. O fator *DelaySigma* de apenas um participante variou de 0ms a 80ms, com incrementos de 20ms.

Os Cenários 2.17 e 2.18 analisam a qualidade de voz em um ambiente em que um participante apresente atraso e outro a perda de pacotes. O número de participantes variou de 3 a 4 e a capacidade residual foi fixada em 75kbps para todos os participantes. Em um dos participantes foi configurado a variação do atraso em 10, 100, 250, 350, 500, 750 e 1000ms, taxa perda de pacotes de 0% e *DelaySigma (jitter)* de 0ms. No outro participante foi configurado a taxa de perda de pacotes de 0, 1, 5, 10, 20, 30, 40%, atraso de 10ms e *DelaySigma (jitter)* de 0ms. Nos demais participantes, foram configurados o atraso de 10ms, 0% de taxa de perda de pacotes e 0ms de *DelaySigma (jitter)*.

Os Cenários 2.19 e 2.20 analisam a qualidade de voz em um ambiente em que um participante apresente perda de pacotes e *jitter*. O número de participantes variou de 3 a 4, com o incremento de 1 participante e a capacidade residual foi fixada em 75kbps para todos os participantes. Em um dos participantes foi configurado a taxa de perda de pacotes de 0, 1, 5, 10, 20, 30, 40%, atraso de 10ms e *DelaySigma (jitter)* de 0ms. No outro participante foi configurado o *DelaySigma (jitter)* com 0, 20, 40, 60, 80, 80 e 80ms, atraso de 100ms e 0% de taxa de perda de pacotes. Existindo outros participantes, foram configurados o atraso de 10ms, 0% de taxa de perda de pacotes e 0ms de *DelaySigma (jitter)*.

A Tabela 4.3 apresenta um resumo de acordo com os cenários, os fatores e os níveis que foram configurados para os cenários de correlação de métricas.

Tabela 4.3 Cenários relativos aos experimentos de avaliação para o serviço de conferência com os parâmetros de rede variando diferentemente entre os participantes.

Cenários	Fatores	Níveis
2.9, 2.10	Participantes	3, 4
	Atraso	10ms, 100ms, 250ms, 350ms, 500ms, 750ms, 1000ms
2.11, 2.12	Participantes	3, 4
	Perda	0%, 1%, 5%, 10%, 20%, 30%, 40%
2.13, 2.14	Participantes	3, 4
	<i>DelaySigma (Jitter)</i>	0ms, 20ms, 40ms, 60ms, 80ms
2.15, 2.16	Participantes	3, 4
	Atraso	10ms, 100ms, 250ms, 350ms, 500ms, 750ms, 1000ms
	Perda	0%, 1%, 5%, 10%, 20%, 30%, 40%
2.17, 2.18	Participantes	3, 4
	Perda	0%, 1%, 5%, 10%, 20%, 30%, 40%
	<i>DelaySigma (Jitter)</i>	0ms, 20ms, 40ms, 60ms, 80ms, 80ms, 80ms

4.7.2.1 Avaliação da Correlação entre o Número de Participantes e o Atraso

A Figura 4.15 ilustra o comportamento da taxa transmitida do *Skype* com relação à variação do atraso de apenas um participante e o aumento no número de participantes de uma sessão de conferência. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelo participante 1 e participante 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

Através do gráfico, nota-se que a taxa transmitida se mantém proporcionalmente constante de acordo com o número de participantes até a configuração de 750ms de atraso, quando ocorre um aumento da ordem de aproximadamente de 40% para o cenário composto por apenas 2 participantes e aproximadamente de 60% quando os cenários tinham 3 ou 4 participantes. Esta variação mostra-se idêntica ao experimento onde se variam igualmente todos os parâmetros de rede dos participantes da sessão, evidenciando que não existe correlação entre a modificação dos parâmetros de rede de apenas um participante.

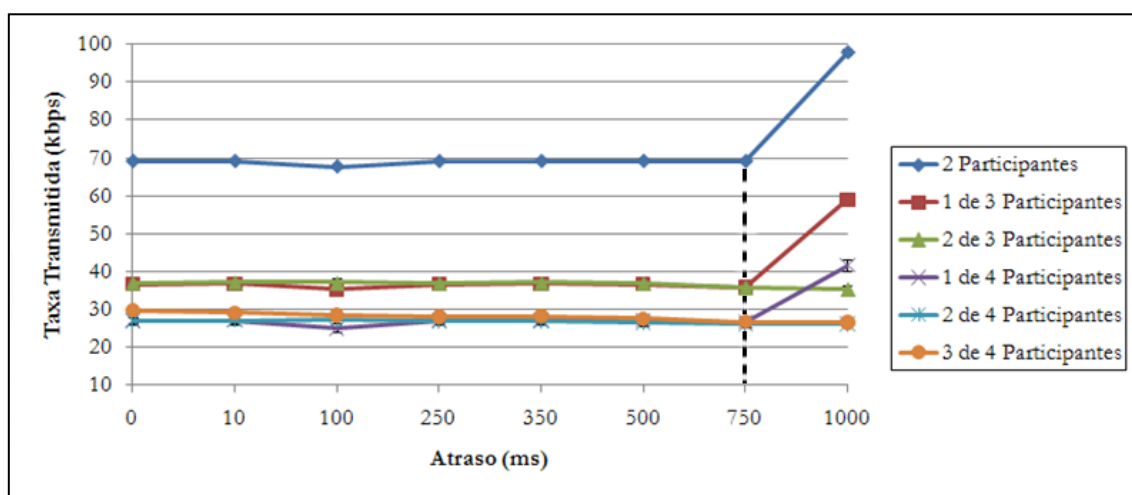


Figura 4.15 Taxa transmitida entre os participantes do serviço de conferência, quando varia o atraso de apenas um participante e o número de participantes (Cenários 2.9 e 2.10).

A Figura 4.16 representa a qualidade de voz da sessão de conferência de acordo com o aumento do atraso de apenas um participante e do número de participantes. Quando utilizados 3 ou 4 participantes, verifica-se que o *Skype* mantém qualidade de voz da sessão de conferência com qualidade satisfatória (pontuação MOS entre 3,60 a 4,03) até atingir um atraso de cerca de 500ms. A partir deste limiar, temos uma queda acentuada da qualidade de voz, atingindo a qualidade pobre (pontuação entre 3,10 a 3,60), sem haver aumento na taxa transmitida. A explicação para este fenômeno é devido ao aumento do atraso entre os participantes esteja havendo sobreposição das vozes no contexto de conversação, o que degrada a qualidade de voz medida através do PESQ. Fato este que pode ser confirmado pela diminuição na qualidade de voz no cenário onde existem 4

participantes a partir do atraso de 350ms, pois com o aumento do número de participantes, há um aumento na probabilidade de sobreposição de vozes do usuário que tem um atraso maior com relação aos demais.

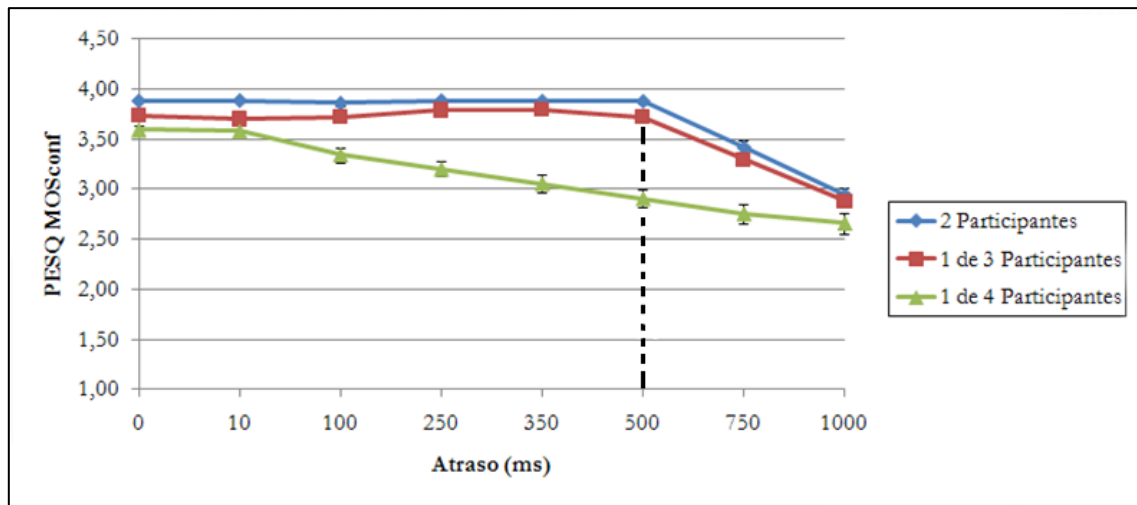


Figura 4.16 Qualidade de voz entre os participantes do serviço de conferência, quando se varia o atraso de apenas um participante e o número de participantes (Cenários 2.9 e 2.10).

4.7.2.2 Avaliação da Correlação entre o Número de Participantes e a Perda de Pacotes

A Figura 4.17 ilustra o desempenho da taxa transmitida do *Skype* com relação à variação da perda de pacotes de apenas um participante e o aumento no número de participantes de uma sessão de conferência. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelo participante 1 e participante 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

Através do gráfico, é possível perceber um aumento da taxa transmitida do participante que têm sua taxa de perda de pacotes variando, quando esta ultrapassa 5%. A partir deste limiar, a taxa transmitida continua a crescer até atingir 30% de perda de pacotes. Além desse limiar, ocorre uma queda muito acentuada. Os participantes que não tiveram o parâmetro de perda de pacotes variando, apresentam uma taxa transmitida constante ao longo de todo o experimento, comprovando que a presença de um participante com perda de pacotes não afeta a taxa de transmissão dos demais.

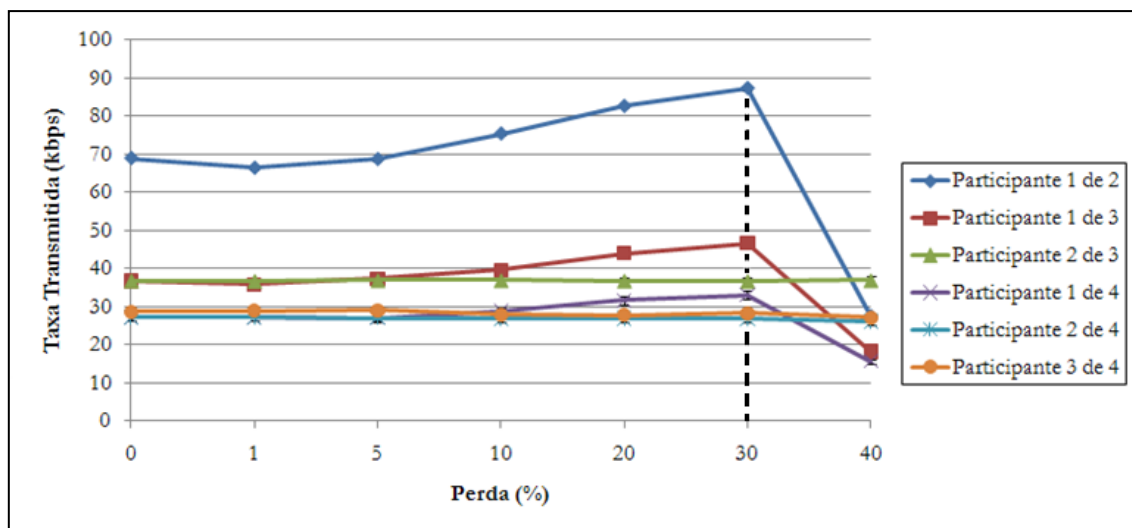


Figura 4.17 Taxa transmitida entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de apenas um participante e o número de participantes (Cenários 2.11 e 2.12).

A Figura 4.18 representa a qualidade de voz de uma sessão de conferência de acordo com o aumento da perda de pacotes de apenas um participante e do número de participantes. Neste gráfico o experimento com 2 participantes apresenta uma qualidade de voz pior do que a de 3 participantes. Isto se deve a métrica PESQ MOSconf utilizada nesta metodologia que avalia a qualidade da conferência como um todo, pois embora haja um participante apresentando perda de pacotes (e consequentemente uma pior qualidade de voz) existem outros participantes que não apresentam este problema, avaliando como melhor a qualidade de voz da conferência.

Apesar de que a métrica PESQ MOSconf leva em consideração a qualidade de voz da conferência como um todo, é importante frisar quando o somatório das degradações afetam a qualidade. Isto pode ser verificado na mesma figura, no experimento relacionado com 4 participantes. Embora seja semelhante ao experimento com 3 participantes, quando a taxa de perda de pacotes de um dos participantes está entre 0 e 30%, a qualidade de voz da conferência é pior do que o experimento com 2 e 3 participantes, pois a métrica avalia que a soma do ruído causado pela perda de pacotes de um dos participantes com o ruído gerado pelo número de participantes (que como visto anteriormente, o número de participantes degradada a qualidade de voz da conferência) degrada mais a qualidade de voz da conferência como um todo do que as condições encontradas nos experimentos com 2 e 3 participantes.

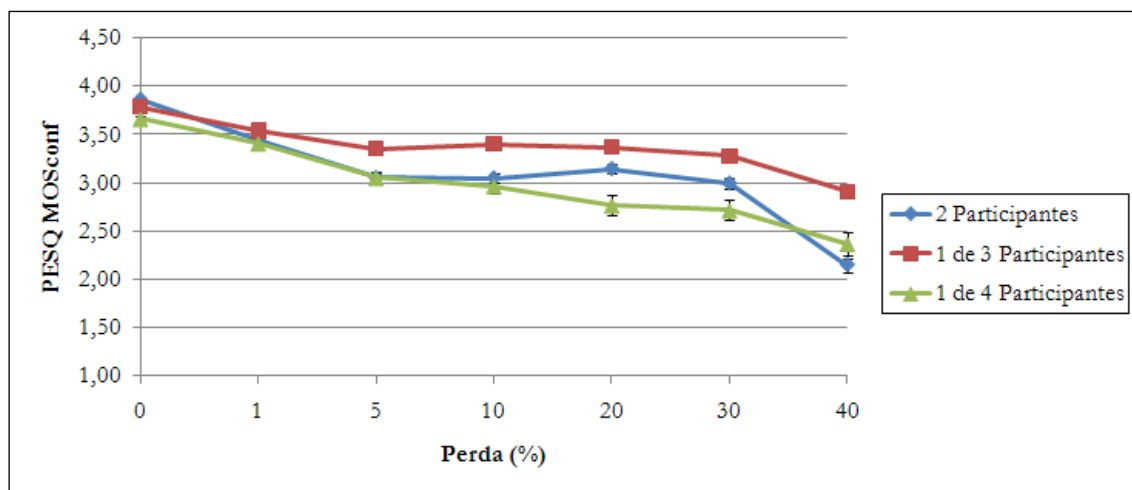


Figura 4.18 Qualidade de voz entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de apenas um participante e o número de participantes (Cenários 2.11 e 2.12).

4.7.2.3 Avaliação da Correlação entre o Número de Participantes e o *Jitter*

A Figura 4.19 ilustra o comportamento da taxa transmitida do *Skype* com relação à variação do *jitter* (variação do atraso) de apenas um participante e o aumento no número de participantes de uma sessão de conferência. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelo participante 1 e participante 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

A taxa transmitida apresentada por cada participante é semelhante ao experimento onde variou o *jitter* igualmente de todos os participantes. No entanto, apenas o participante que teve a seu *jitter* modificado sofreu queda na taxa transmitida quando o *jitter* configurado passou de 60ms para 80ms, o que é esperado, tendo em vista que o *Skype* tenta se adaptar ao *jitter* diminuindo a taxa transmitida. Ainda assim é importante notar que a taxa transmitida dos demais participantes não se altera quando o *jitter* de apenas um participante é alterado.

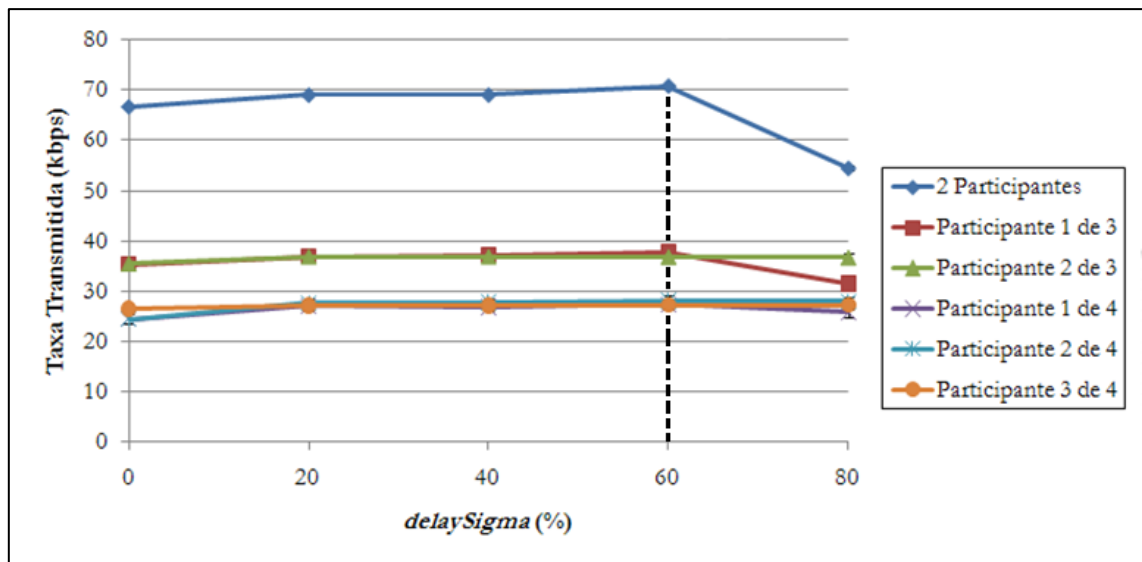


Figura 4.19 Taxa transmitida entre os participantes do serviço de conferência, quando varia o *DelaySigma* (*jitter*) de apenas um participante e o número de participantes (Cenários 2.13 e 2.14).

A Figura 4.20 representa a qualidade de voz de uma sessão de conferência de acordo com o aumento do *jitter* e de apenas um participante e do número de participantes. Quando o *jitter* variou de 0 a 60ms a qualidade da voz do experimento com 2 participantes era melhor do que o com 3 participantes, onde apenas um teria variação no seu *jitter*. Após o limiar de 60ms este quadro se inverteu: o experimento relativos a 2 participantes passou a apresentar uma qualidade de voz inferior ao do experimento com 3 participantes. Isto se deve pelo fato de que no experimento com 2 participantes só há um fator de degradação da qualidade do áudio, que é o *jitter*, enquanto que no experimento com 3 existem dois fatores: o *jitter* e o número de participantes. Quando se ultrapassa o limiar dos 60ms, a conjunção dos fatores de degradação da qualidade de voz do experimento com 3 participantes passa a ser inferior ao dos 2 participantes, ou seja, o *jitter* aplicado a 2 participantes afeta mais a qualidade do áudio do que um cenário com 3 participantes, onde apenas um apresenta *jitter*, quando este alcança o valor de 60ms. Este mesma justificativa também pode ser estendido ao experimento com 3 participantes.

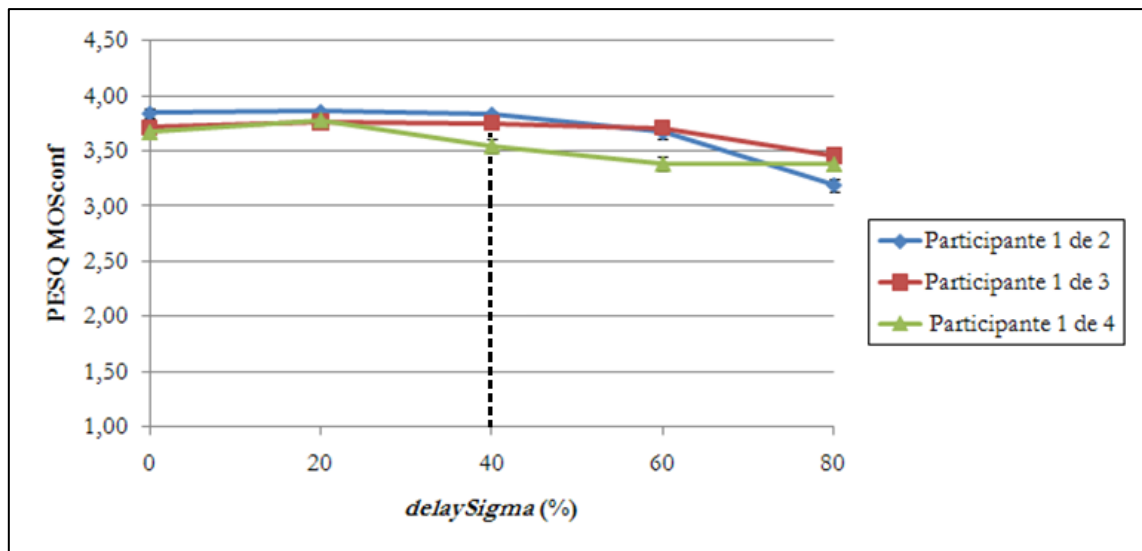


Figura 4.20 Qualidade de voz entre os participantes do serviço de conferência, quando varia o *DelaySigma* (*jitter*) de apenas um participante e o número de participantes (Cenários 2.13 e 2.14).

4.7.2.4 Avaliação da Correlação entre o Atraso e a Perda de Pacotes

A Figura 4.21 apresenta o comportamento da taxa transmitida do *Skype*, onde há um participante apresentando variação do atraso e outro participante, variação de perda de pacotes. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelo participante 1 e participante 2 do cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

No cenário com 3 participantes, observa-se que o participante com perda apresenta uma evolução compatível proporcionalmente ao comportamento do cenário de 2 participantes com perda. O mesmo também pode ser observado para o participante com variação do atraso. Nota-se, também, evolução semelhante para o cenário com 4 participantes, com a diferença de que o terceiro participante deste experimento, o qual não foi configurado nenhum parâmetro de rede, possui a taxa transmitida uniforme.

Neste experimento não foi constatado nenhuma correlação das diferentes configurações dos participantes com relação à taxa transmitida. Elas se comportaram de acordo apenas com os parâmetros de rede configurados individualmente.

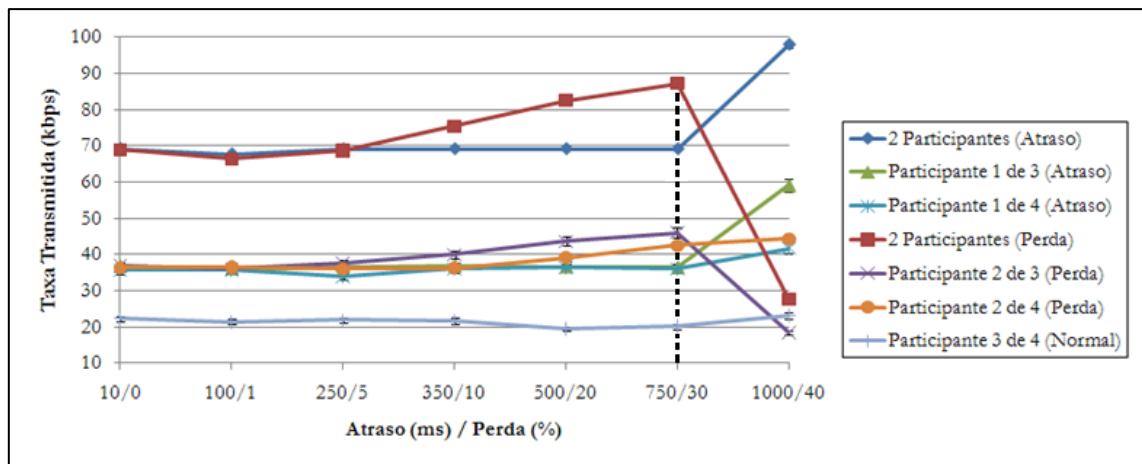


Figura 4.21 Taxa transmitida entre os participantes do serviço de conferência, quando se varia o atraso de um participante, a taxa de perda de pacotes de outro, e o número de participantes (Cenários 2.15 e 2.16).

Embora a taxa transmitida entre os participantes tenha se mostrado semelhante aos experimentos onde variou-se igualmente os parâmetros de rede de todos os participantes, a qualidade de voz da sessão de conferência mostrou-se ligeiramente diferente. No entanto, a curva apresentada assemelha-se a curva apresentada no cenário de perda de pacotes, conforme pode ser observado através da Figura 4.22. A razão deste comportamento, é que o parâmetro de rede perda de pacotes apresenta um peso maior na qualidade de voz da conferência, se comparado ao atraso.

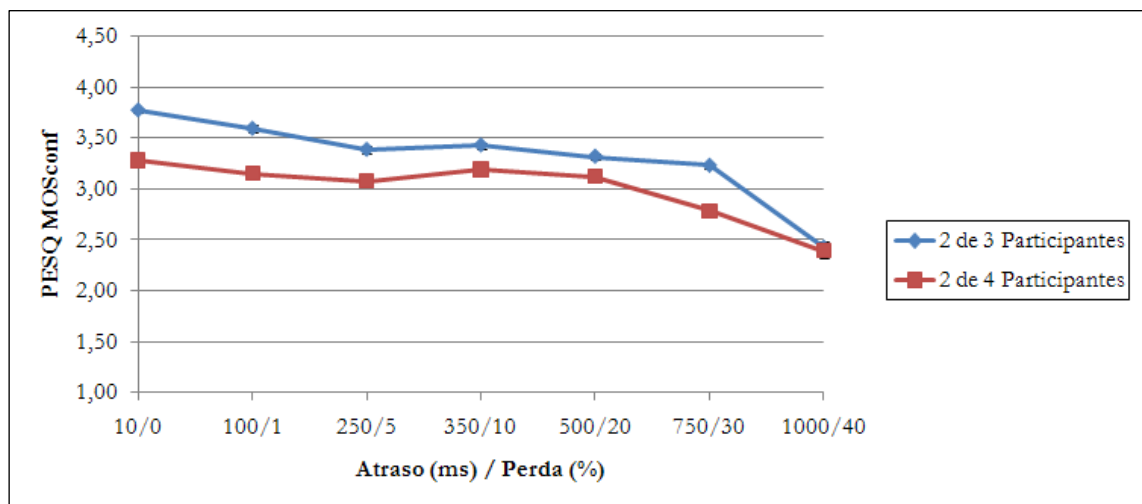


Figura 4.22 Qualidade de voz entre os participantes do serviço de conferência, quando varia o atraso de um participante, a taxa de perda de pacotes de outro, e o número de participantes (Cenários 2.15 e 2.16).

4.7.2.5 Avaliação da Correlação entre a Perda de Pacotes e o *Jitter*

A Figura 4.23 ilustra o comportamento da taxa transmitida do *Skype*, onde há um participante apresentando variação da perda de pacotes e outro participante, variação do *jitter*. A legenda indica a taxa transmitida entre dois participantes, a taxa transmitida pelo participante 1 e participante 2 do

cenário com 3 participantes (Participante 1 de 3 e Participante 2 de 3, respectivamente) e assim por diante.

No cenário com 3 participantes, observa-se que o participante com perda apresenta uma evolução compatível proporcionalmente ao comportamento do cenário de 2 participantes com perda. O mesmo também pode ser observado para o participante com variação do *jitter*. Nota-se, também, evolução semelhante para cenário com 4 participantes, com a diferença de que o terceiro participante deste experimento, o qual não foi configurado nenhum parâmetro de rede, possui a taxa transmitida uniforme. Comparando-se com os experimentos dos Cenários 2.15 e 2.16, onde se apresentou um participante com variação do atraso e outro com variação da perda de pacotes, não foi constatado nenhuma correlação das diferentes configurações dos participantes com relação à taxa transmitida.

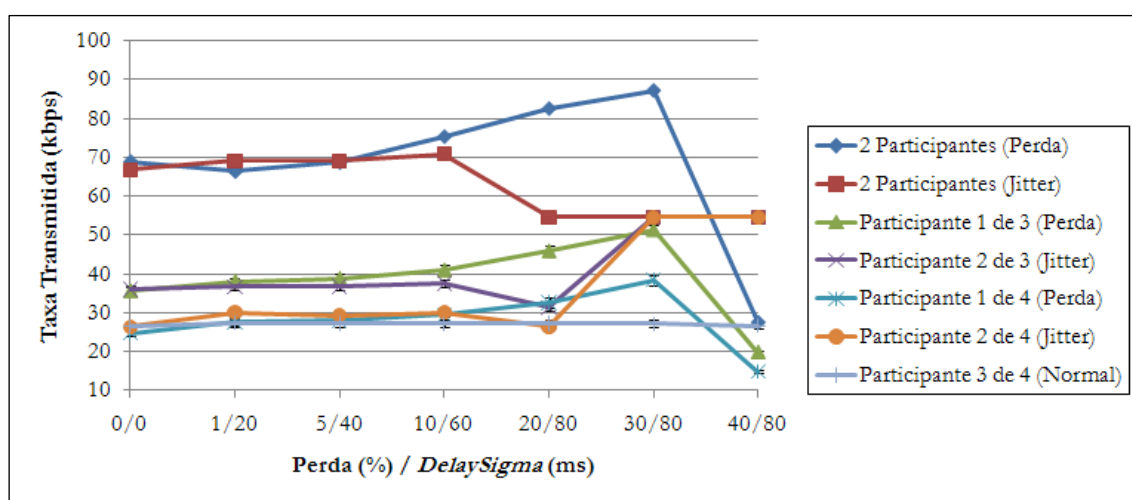


Figura 4.23 Taxa transmitida entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de um participante, o *DelaySigma* (*jitter*) de outro, e o número de participantes (Cenários 2.17 e 2.18).

Embora a taxa transmitida entre os participantes tenha se mostrado semelhante aos experimentos onde se variou igualmente os parâmetros de rede de todos os participantes, a qualidade de voz da sessão de conferência mostrou-se ligeiramente diferente. No entanto, a curva apresentada assemelha-se a curva apresentada no cenário de perda de pacotes, conforme pode ser observado através da Figura 4.24. A razão deste comportamento, é que o parâmetro de rede perda de pacotes apresenta um peso maior na qualidade de voz da conferência, se comparado ao atraso.

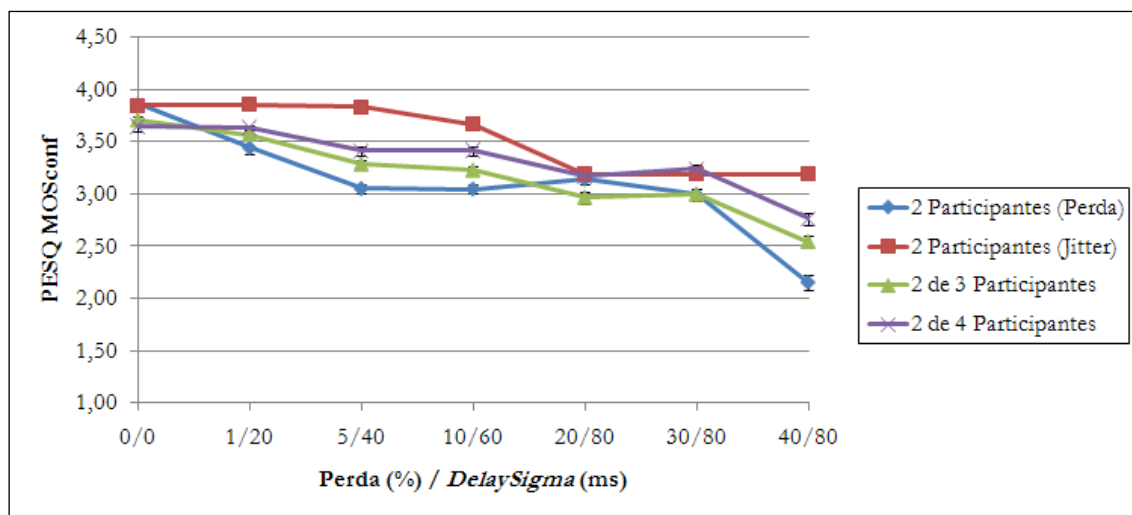


Figura 4.24 Qualidade de voz entre os participantes do serviço de conferência, quando varia a taxa de perda de pacotes de um participante, o *DelaySigma* (*jitter*) de outro, e o número de participantes (Cenários 2.17 e 2.18).

4.7.3 Avaliação dos Cenários do Serviço de Convergência

O cenário 8.1 tem como objetivo analisar o comportamento do serviço de convergência. Assim, conforme a metodologia proposta neste trabalho, foi estabelecida uma chamada utilizando-se o serviço de convergência da Internet para a PSTN (*SkypeOut*) para um número do serviço de convergência da PSTN para a Internet (*SkypeIn*). Esta chamada foi realizada entre dois participantes que se encontravam dentro do ambiente controlado e não foram configurados parâmetros de rede via o emulador *NIST.Net*.

O Cenário 8.2 analisa o comportamento do serviço de convergência quando este é utilizado em conjunto com um serviço de conferência. Assim, foi estabelecida uma conferência entre um participante que utilizava o serviço de convergência e outros dois participantes que utilizavam apenas o serviço de conferência. Não foram configurados os parâmetros de rede via *NIST.Net* do participante que usava o serviço de convergência, enquanto os participantes que usavam apenas o serviço de convergência tiveram seus parâmetros de rede ajustados para capacidade 100kbps, atraso 10ms, taxa de perda de pacotes 0% e *DelaySigma* (*jitter*) 0ms.

O Cenário 8.3 analisa o comportamento do serviço de convergência quando este é utilizado em conjunto com um serviço de conferência e um dos participantes possui condições de rede adversas. Assim, foi estabelecida uma conferência entre 3 participantes: um deles, utilizava o serviço de convergência e outros dois participantes, que se conectavam através da Internet, utilizavam apenas o serviço de conferência. Não foram configurados os parâmetros de rede via *NIST.Net* do participante que usava o serviço de convergência, enquanto que um dos outros dois participantes restantes, teve seus parâmetros de rede ajustados para capacidade 100kbps, atraso 100ms, taxa de

perda de pacotes 1% e *DelaySigma (jitter)* 80ms e o outro para capacidade 100kbps, atraso 10ms, taxa de perda de pacotes 0% e o *DelaySigma (jitter)* 0ms.

A **Tabela 4.4** apresenta um resumo dos cenários, os fatores e os níveis que foram configurados igualmente para todos os participantes.

Tabela 4.4 Cenários relativos aos experimentos de avaliação para o serviço de convergência com os parâmetros de rede variando diferentemente entre os participantes.

Cenários	Fatores	Níveis
8.1	Participantes	2
8.2	Participantes	3
8.3	Participantes	3
	Atraso	100ms
	Perda	1%
	<i>DelaySigma (jitter)</i>	80ms

A Figura 4.25 apresenta a taxa transmitida entre dois participantes que utilizam o serviço de convergência de acordo com a metodologia proposta neste trabalho. É possível observar que a média da taxa transmitida é de aproximadamente 27kbps. Através do aplicativo *Skype*, foi constatado que o *codec* utilizado no serviço de convergência é o GSM.

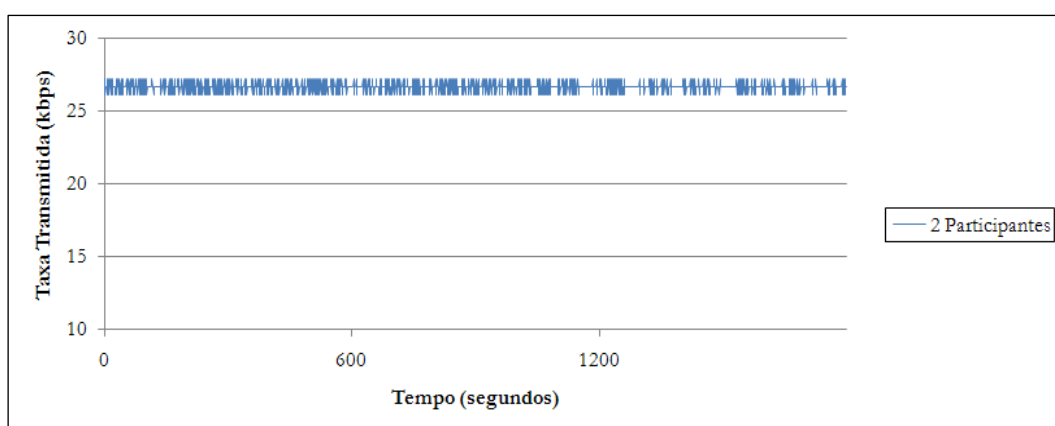


Figura 4.25 Taxa transmitida entre dois participantes do serviço de convergência (Cenário 8.1).

Através da análise do tráfego dos participantes envolvidos neste experimento, extraiu-se os endereços IP referentes aos *gateways* VoIP envolvidos no serviço de convergência. O participante

emissor do áudio utilizou o *gateway* VoIP 4.53.80.230 e o participante que era o responsável por gravar o áudio se comunicava com o *gateway* VoIP 4.53.80.103. A partir desta informação foi realizada a estimativa do atraso e da taxa de perda de pacotes entre cada participante e o seu respectivo *gateway* VoIP, conforme ilustrado na Tabela 4.5.

Tabela 4.5 Estimativa das métricas de rede entre dois participantes do serviço de convergência (Cenário 8.1).

Gateway VoIP	Número de Pacotes	Salto	Taxa de Perda de Pacotes	Atraso
4.53.80.230	1000	10	0,1%	71ms
4.53.80.120	1000	14	0,1%	75ms

A Figura 4.26 ilustra a pontuação PESQ MOS obtida através do serviço de convergência realizado entre dois participantes (Cenário 8.1).

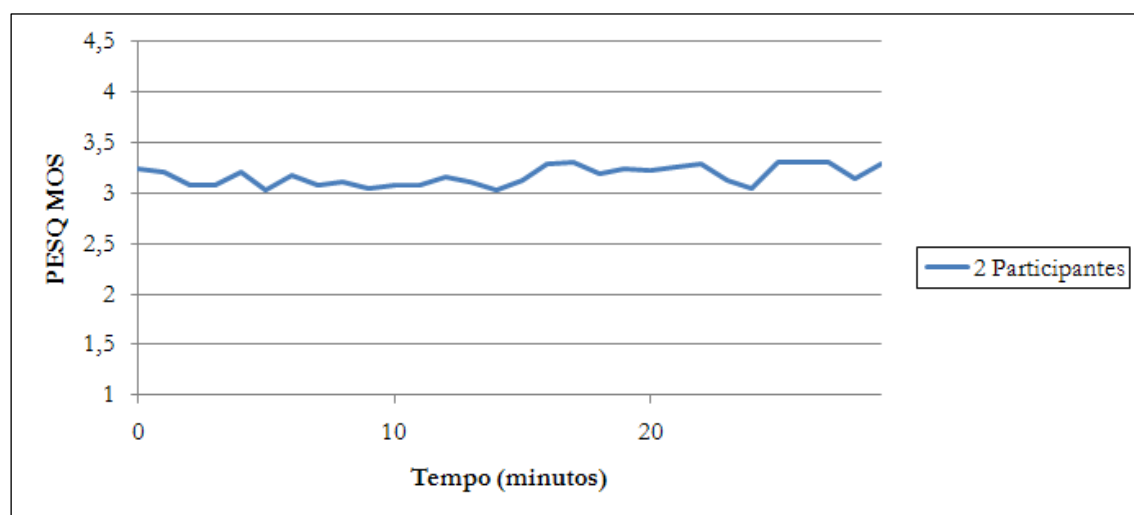


Figura 4.26 Qualidade de voz entre dois participantes do serviço de convergência (Cenário 8.1).

A Figura 4.27 apresenta a taxa transmitida entre 3 participantes (Cenário 8.2), sendo um deles utilizando o serviço de convergência (1 de 3 Participantes) e os outros dois utilizando a Internet (2 de 3 Participantes). Conforme pode ser observado, a taxa do serviço de convergência manteve-se constante, em torno de 30kbps e a taxa transmitida do um dos participantes que utilizavam a Internet, oscilou entre 70kbps e 9kbps. Esta oscilação é devido aos períodos de silêncio da carga de trabalho utilizada para simular o contexto de conversação. Também pode ser notado que a presença de um participante do serviço de convergência em uma conferência não altera a taxa transmitida dos outros participantes.

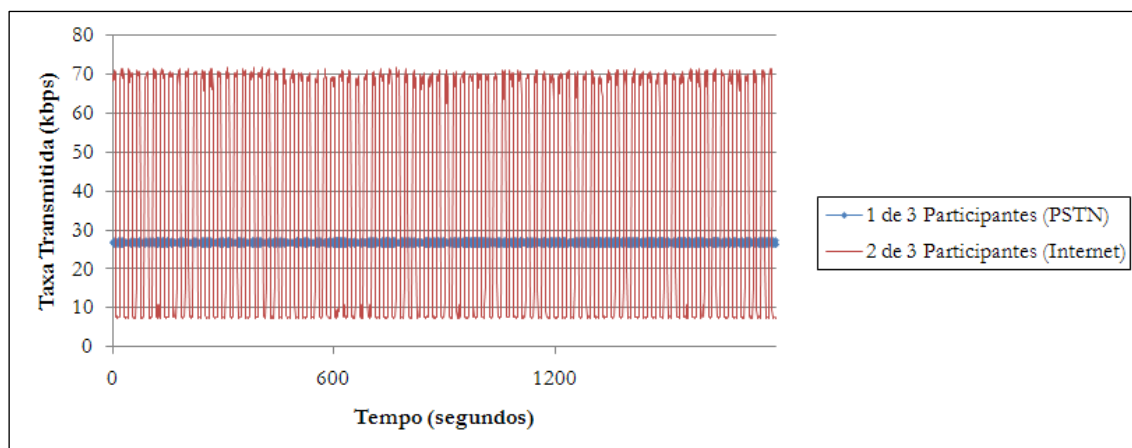


Figura 4.27 Taxa transmitida entre três participantes do serviço de convergência (Cenário 8.2).

A Figura 4.28 ilustra a pontuação da métrica PESQ MOSconf para a avaliação da qualidade de voz de em um cenário de conferência, sendo um dos participantes utilizando o serviço de convergência e dois outros participantes a Internet (Cenário 8.2). O PESQ MOS apresentado aqui é maior do que o cenário com apenas 2 participantes, pois como existem 2 participantes neste cenário que estão utilizando a Internet, estes apresentam uma qualidade de voz melhor do que o participante que utiliza o serviço de convergência. Assim, a qualidade de voz média da conferência tende a ser melhor do que o cenário entre 2 participantes do serviço de convergência.

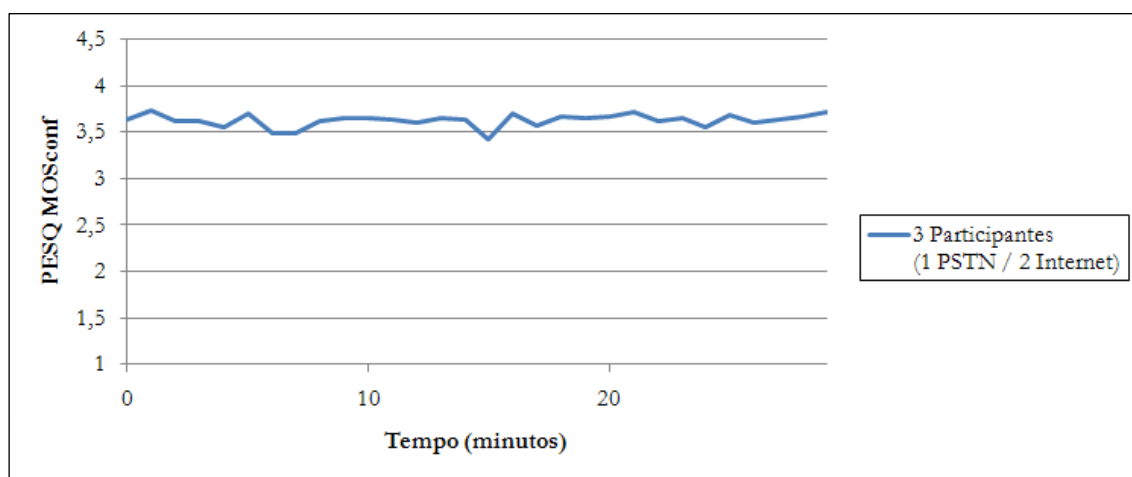


Figura 4.28 Qualidade de voz entre dois participantes do serviço de convergência (Cenário 8.2).

Através da análise do tráfego dos participantes envolvidos neste experimento, extraiu-se os endereços IP referentes aos *gateways* VoIP envolvidos no serviço de convergência. O participante emissor do áudio utilizou o *gateway* VoIP 4.78.191.43 e o participante que era o responsável por gravar o áudio se comunicava com o *gateway* VoIP 4.53.80.99. A partir desta informação foi realizada a estimativa do atraso e da taxa de perda de pacotes entre cada participante e o seu respectivo *gateway* VoIP, conforme ilustrado na Tabela 4.6.

Tabela 4.6 Estimativa das métricas de rede entre três participantes do serviço de convergência (Cenário 8.2)

Gateway VoIP	Número de Pacotes	Saltos	Taxa de Perda de Pacotes	Atraso
4.78.191.43	1000	10	0,1%	74ms
4.53.80.99	1000	14	0,1%	73ms

A Figura 4.29 apresenta a qualidade de voz obtida através do Cenário 8.3, onde existem 3 participantes em um cenário de conferência, sendo 2 deles realizando a chamada através da Internet e um através do serviço de convergência. Um dos participantes conectado através da Internet, apresenta condições adversas de rede, onde foram configurados atraso de 100ms, 1% de taxa de perda de pacotes e *DelaySigma* (*jitter*) de 80ms.

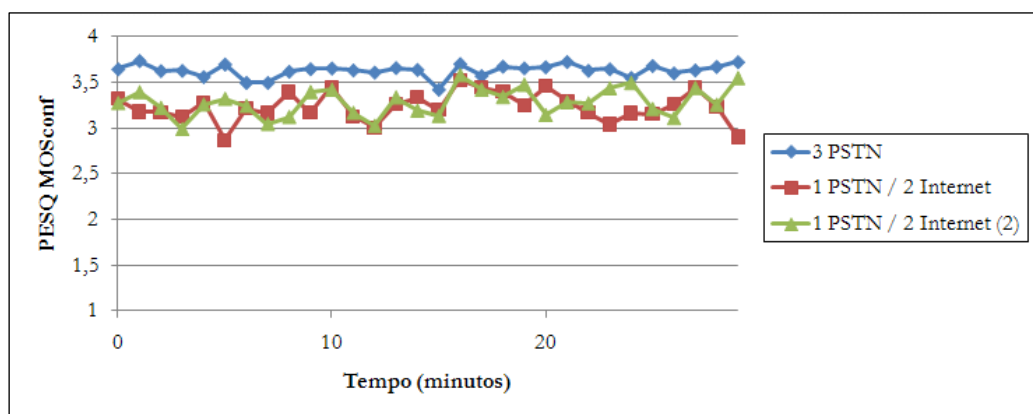


Figura 4.29 Qualidade de voz entre três participantes do serviço de convergência (Cenário 8.3).

Através da análise do tráfego dos participantes envolvidos neste experimento, extraiu-se os endereços IP referentes aos *gateways* VoIP envolvidos no serviço de convergência. O participante emissor do áudio utilizou o *gateway* VoIP 4.78.50.41 e o participante que era o responsável por gravar o áudio se comunicava com o *gateway* VoIP 4.78.82.79. A partir desta informação foi realizada a estimativa do atraso e da taxa de perda de pacotes entre cada participante e o seu respectivo *gateway* VoIP, conforme ilustrado na Tabela 4.7.

Tabela 4.7 Estimativa das métricas de rede entre três participantes do serviço de convergência (Cenário 8.3).

Gateway VoIP	Número de Pacotes	Saltos	Taxa de Perda de Pacotes	Atraso
4.78.50.41	1000	10	0,1%	72ms
4.78.82.79	1000	14	0,1%	76ms

4.8 Análise dos Experimentos Realizados

Através da metodologia proposta neste trabalho, foram realizados diversos experimentos utilizando-se o aplicativo *Skype*. Estes experimentos serviram para traçar um comportamento do aplicativo *Skype* com relação à qualidade de voz e a adaptabilidade, quando utilizando os seus serviços agregados, como a sessão de conferência e a convergência com a PSTN.

De uma maneira geral, verificou-se que a versão do aplicativo *Skype* avaliada apresenta uma maior taxa de transmissão com relação às versões anteriores e uma melhor resistência ao atraso e a perda de pacotes, conseguindo assim, manter uma qualidade de voz melhor em situações adversas de atraso e perda. Este fato se deve a modificação do *codec* utilizado por esta nova versão. Embora não tenha sido divulgada a razão de se ter desenvolvido um *codec* que fosse mais resistente a perdas, especula-se que tenha sido com o propósito de melhorar o desempenho do *Skype* quando utilizado em redes sem fio, pois estas notadamente apresentam atraso e perda de pacotes maiores com relação às redes cabeadas. Este argumento pode ser respaldado graças à cobertura feita pela mídia, onde mostra a utilização do *Skype* em plataformas móveis, especificamente *netbooks* e *smartphones*.

Com relação ao serviço de sessão de conferência, verificou-se que há o aumento da taxa transmitida à medida que se aumenta o número de participantes. Este fato pode ser deduzido pelo aumento da troca de mensagens de controle por parte dos participantes deste tipo de serviço. A respeito da qualidade de voz oferecida por este serviço, apurou-se que o número de participantes de uma conferência é um fator que degrada a qualidade de voz. No entanto, é possível que um cenário com 3 ou mais participantes apresente uma qualidade melhor do que um de 2 participantes. Isto dependerá de quantos participantes têm seus parâmetros de rede variando, o tipo de parâmetro e a qual a sua variação. Outro ponto a destacar é que a variação dos parâmetros de rede dos diversos participantes influencia na qualidade de voz da conferência, conforme pode ser verificado ao longo dos experimentos realizados. A intensidade dessa influência, novamente, dependerá de quais parâmetros estejam sendo variados e as suas medidas.

A respeito do serviço de convergência, constatou-se que apresenta uma taxa transmitida bem menor do que quando utilizada sem convergência. O motivo desse comportamento é que é utilizado um outro *codec* neste serviço, mais adequado a integração à PSTN. A qualidade de voz provida por este serviço, também se mostrou inferior a uma chamada realizada totalmente pela Internet, que novamente também pode ser atribuída a utilização de outro *codec* para esta tarefa.

5 Conclusões e Trabalhos Futuros

“Tudo que tem um começo, tem um fim.”

O Oráculo, Matrix Revolutions

Este capítulo final realiza os últimos apontamentos acerca do tema, menciona as principais contribuições desta dissertação e direciona sugestões de trabalhos a serem feitas.

5.1 Considerações Finais

Este trabalho propõe uma metodologia para avaliar o desempenho de aplicativos de voz sobre IP de acordo com as suas políticas de adaptação dinâmica e a qualidade do áudio das aplicações que visa avaliar os principais serviços agregados encontrados nestes aplicativos: os serviços de sessão de conferência e a convergência com a PSTN. Para atingir esse objetivo, foi desenvolvida uma metodologia que aborda um esquema de classificação das formas de utilização dos serviços, com o intuito de organizar e abordar as principais características dos cenários de utilização dos aplicativos VoIP; o desenvolvimento de um ambiente controlado para a execução dos experimentos, a fim de garantir as mesmas condições de repetibilidade para os experimentos realizados; considerações a respeito da realização da avaliação de desempenho do serviço de convergência, tendo em vista que este se encontra em um ambiente não-controlado, isto é, não é possível a configuração dos parâmetros de rede deste serviço.

À medida que o número de participantes aumenta, também aumenta a complexidade de realizar os experimentos e os custos envolvidos, uma vez que, para cada participante, é utilizada uma máquina. No entanto, a precisão e a automatização inerentes aos métodos objetivos de medir a qualidade de voz compensam todas as dificuldades em se realizar estes experimentos.

A questão de um utilizar/criar uma abordagem para realizar a avaliação de desempenho sob o contexto de conversação através dos diversos serviços agregados constitui um desafio para a avaliação de desempenho dos sistemas VoIP atuais. Eles nos ajudam a guiar quais as melhores decisões e modificações a serem realizadas nos sistemas VoIP com o objetivo de atingir a melhor qualidade de voz.

Esta metodologia pode ajudar os usuários finais nas escolhas de como utilizar, quais serviços Internet adquirir e quais aplicativos utilizar sob os que melhor lhe atendem com relação a sua utilização. A área de avaliação de qualidade de voz carece de uma solução objetiva, isto é, automatizada para avaliar a qualidade de voz sobre o contexto de conversação.

Este tipo de informação é interessante porque permite melhoras no desenvolvimento do aplicativo, como o desenvolvimento de um algoritmo mais evoluído para eleger qual seria o melhor participante para ser o participante host da sessão de conferência, isto é, qual seria o melhor participante para mixar e distribuir o áudio da conferência.

Este trabalho pode ser utilizado para beneficiar as operadoras de acesso a Internet, de modo a oferecer serviços diferenciados para os aplicativos VoIP e seus serviços agregados, como a conferência e a convergência.

Também pode contribuir para o aperfeiçoamento dos *codecs* dos aplicativos VoIP, uma vez que agora eles podem mensurar de uma maneira objetiva e automática a qualidade de voz e a taxa transmitida dos seus aplicativos diante dos mais diversos cenários de utilização.

5.2 Principais Contribuições

Dentre as principais contribuições deste trabalho, apontam-se:

- O desenvolvimento de uma metodologia para avaliar as aplicações VoIP sob os aspectos da qualidade de voz e adaptabilidade na Internet, com ênfase na utilização dos serviços agregados de sessão de conferência e convergência com a PSTN. Esta metodologia contempla: um esquema de classificação dos cenários de utilização destes aplicativos; a emulação do contexto de conversação através da elaboração adequada da carga de trabalho a ser utilizada no ambiente controlado; a configuração de um ambiente controlado para a realização dos experimentos, onde é possível emular o comportamento da Internet com poucos recursos; o procedimento para realizar e coletar os dados oriundos dos experimentos relativos à avaliação dos serviços de conferência e convergência com a PSTN; assinalar as ferramentas para auxiliar a extração de informações relevantes dos dados obtidos dos experimentos;
- A criação de um procedimento para a realização de testes do tipo caixa preta sobre o serviço de convergência, de forma a realizar a avaliação de desempenho do mesmo;
- O esclarecimento de aspectos do comportamento da aplicação de voz sobre IP *Skype*, quando se faz uso dos seus serviços agregados, como a sessão de conferência e a convergência com a PSTN, através da metodologia proposta neste trabalho;

- A elucidação de diferenças relacionadas à adaptabilidade e a qualidade de voz entre diferentes versões do *Skype*, devido à utilização de novos *codecs* por parte deste.

5.3 Trabalhos Futuros

Como possíveis atividades em continuações da pesquisa iniciada nesta dissertação, pode-se citar:

- A construção de um modelo, através dos dados colhidos nos experimentos realizados, de um modelo semelhante ao Modelo-E [12], para o *Skype*, onde, através da especificação do tipo de serviço utilizado, número de participantes e dos parâmetros de rede de cada participante, possa ser inferida a qualidade de voz da conferência;
- A avaliação, através da metodologia proposta neste trabalho, de outros aplicativos de voz sobre IP como o *Yahoo! Messenger* e o *Gizmo Project*, e seus serviços de conferência e convergência;
- Analisar a evolução dos *codecs* utilizados pelo *Skype* ao longo das diversas versões, com relação à sua adaptabilidade e qualidade de voz, e utilização dos seus serviços de convergência e conferência;
- Desenvolver uma simulação, através de ferramentas como o *Network Simulator* (NS), que modelem os serviços agregados de convergência e conferência.

6 Referências

- [1] Barbosa, R., **“Avaliação de Desempenho de Aplicações VoIP P2P”**, dissertação de mestrado, março de 2007, UFPE.
- [2] Chen, K., Huang, C., Huang, P., Lei, C., **“Quantifying Skype User Satisfaction”**, SIGCOMM, setembro de 2006.
- [3] Bacioccola, A., Cicconetti, C., Stea, G., **“User-level Performance Evaluation of VoIP Using ns-2”**, NSTools, outubro de 2007.
- [4] Lindblom, J., **“A Sinusoidal Voice Over Packet Coder Tailored for the Frame-Erasure Channel”**, IEEE Transactions on Speech and Audio Processing, 2005.
- [5] TeleGeography, **“International Carriers’ Traffic Grows Despite Skype Popularity”**, http://www.telegeography.com/cu/article.php?article_id=15656, dezembro de 2006, acessado em setembro de 2008.
- [6] Počta, P., Vaculík, M., **“Impact of the Duration of Speech Sequences on Speech Quality”**, Journal of Telecommunications and Information Technology, abril de 2007.
- [7] Fu, T., Chiu, D., Lui, J., **“Performance Metrics and Configuration Strategies for Group Network Communication”**, 15th IEEE International Workshop on Quality of Services (IWQoS 2007), junho de 2007.
- [8] Callado, A., Fernandes, G., Silva, A., Barbosa, R., Sadok, D., Kelner, J., **“Construção de Redes de Voz sobre IP”**, minicurso, SBRC maio de 2007.
- [9] Barbosa, R., Kamienski, C., Mariz, D., Callado, A., Fernandes, S., Sadok, D., **“Performance Evaluation of P2P VoIP Applications”**, NOSSDAV 2007.
- [10] Guéguin, M., Bouquin-Jeannès, R., Gautier-Turbin, V., Faucon, G., Barriac, V., **“On the Evaluation of the Conversational Speech Quality in Telecommunications”**, EURASIP Journal on Advances in Signal Processing, 2008.
- [11] International Telecommunications Union, **“Methods for Subjective Determination of Transmission Quality”**, Recommendation P.800, 1996.
- [12] International Telecommunications Union, **“The E-Model, A Computational Model for Use in Transmission Planning”**, Recommendation G.107, 2005.
- [13] International Telecommunications Union, **“Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs”**, Recommendation P.862, 2001.
- [14] Website do **Skype**, <http://www.skype.com>, último acesso em fevereiro de 2009.
- [15] Website do **GTalk**, <http://www.google.com/talk/>, último acesso em fevereiro de 2009.
- [16] Website do **Windows Live Messenger**, <http://messenger.live.com>, último acesso em fevereiro de 2009.
- [17] Website do **Yahoo! Messenger**, <http://messenger.yahoo.com>, último acesso em fevereiro de 2009.
- [18] Website do **Jajah**, <http://www.jajah.com>, último acesso em fevereiro de 2009.
- [19] Fingal, F., Gustavsson, P., **“A SIP of IP-Telephony”**, 10 de fevereiro de 1999.
- [20] Baugher, M., McGrew, D., Naslund, M., Carrara, E., Norrman, K., **“The Secure Real-time Transport Protocol (SRTP)”**, RFC 3711, março de 2004.

- [21] TeleGeography, **“Skype’s Share of the International Long-distance Pie on the Increase”**, http://www.telegeography.com/cu/article.php?article_id=27800, último acesso em março de 2009.
- [22] Tanenbaum, A., **“Computer Networks”**, 4th Ed, Prentice-Hall, 2003.
- [23] Collins, D., **“Carrier Grade Voice over IP”**, McGraw-Hill, 2001.
- [24] Walker, J., Hicks, J., **“The Essential Guide to VoIP Implementation and Management”**, NetIQ Corporation, 2002.
- [25] Jain, R., **“The Art of Computer Systems Performance Analysis”**, Wiley, 1991.
- [26] Gaynor, M., **“Proactive Packet Dropping Methods for TCP Gateways”**, <http://www.eecs.harvard.edu/gaynor/final.ps>, Novembro de 1996.
- [27] Postel, J., **“Transmission Control Protocol”**, STD 7, RFC 793, setembro de 1981.
- [28] Postel, J., **“User Datagram Protocol”**, STD 6, RFC 768, agosto de 1980.
- [29] Postel, J., **“Internet Protocol”**, RFC 791, setembro de 1981.
- [30] Richards, D., **“Telecommunication by Speech: The Transmission Performance of Telephone Networks”**, Butterworths, Londres, UK, 1973.
- [31] Gros, L., **“Evaluation Subjective de la Qualité Vocale Fluctuante”**, tese de doutorado, Université d’Aix-Marseille II, Marseilles, França, 2001.
- [32] Sat, B., Wah, B., **“Analyzing Voice Quality in Popular VoIP Applications”**, IEEE Intelligent and Pervasive Multimedia Systems, 2009.
- [33] **Skype Statistics**, http://share.Skype.com/stats_rss.xml, último acesso em junho de 2009.
- [34] Mercier, J., **“Skype Numerology”**, <http://Skypenumerology.blogspot.com/>, último acesso em junho de 2009.
- [35] **eBay Quartely Results**, http://investor.ebay.com/financial_releases.cfm, último acesso em junho de 2009.
- [36] Website do **eBay Inc.**, <http://www.ebay.com/>, último acesso em junho de 2009.
- [37] TeleGeography, **“TeleGeography Update: International Carrier’s Traffic Grows Despite Skype Popularity”**, http://www.telegeography.com/cu/article.php?article_id=15656, último acesso em junho de 2009.
- [38] Demichelis, C., Chimento, P., **“IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)”**, RFC 3393, novembro de 2002.
- [39] Almes, G., et al, **“A One-way Packet Loss Metric for IPPM”**, RFC 2680, setembro de 1999.
- [40] Website do **My Traceroute (mtr)**, <http://www.bitwizard.nl/mtr/>, último acesso em junho de 2009.
- [41] Website do **Audacity**, <http://audacity.sourceforge.net/>, último acesso em junho de 2009.
- [42] Website do **Winamp**, <http://www.winamp.com/>, último acesso em junho de 2009.
- [43] Website do **Wireshark**, <http://www.wireshark.org/>, último acesso em junho de 2009.
- [44] Website do **NIST.Net**, <http://snad.ncsl.nist.gov/nistnet/>, último acesso em junho de 2009.
- [45] Website do **Meinberg NTP Client**, <http://www.meinberg.de/english/sw/ntp.htm>, último acesso em junho de 2009.
- [46] Website da implementação de referência do algoritmo **PESQ**, versão 2.0, <http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en>, último acesso em junho de 2009.
- [47] Kamienski, C., Sadok, D., Cavalcanti, D., Mariz, D., Dias, K., **“Simulando a Internet: Aplicações na Pesquisa e no Ensino”**, anais do XXII Congresso da SBC (JAI), 2002.

- [48] Website do Super Wideband Audio Codec (SILK), <https://developer.skype.com/silk>, último acesso em junho de 2009.
- [49] Cisco Systems, **“Quality of Service for Voice over IP”**, http://www.cisco.com/en/US/docs/ios/solutions_docs/qos_solutions/QoSVoIP/QoSVoIP.html, 2001.
- [50] Miras, D., **“A Survey on Network QoS Needs of Advanced Internet Applications”**, Working document, Internet2 – QoS Working Group, 2002.
- [51] Opticom GmbH, **“State of the Art Voice Quality Testing”**, White Paper, Erlangen, Alemanha, 2000.
- [52] Lennox, J., Schulzrinne, H., **“A Protocol for Reliable Decentralized Conferencing”**, NOSSDAV’03, junho de 2003.
- [53] Website do **Gizmo Project**, <http://gizmo5.com/pc/>, último acesso em junho de 2009.
- [54] Website do **tcpstat**, <http://www.frenchfries.net/paul/tcpstat/>, último acesso em junho de 2009.
- [55] European Telecommunications Standards Institute, **“GSM Full Rate Speech Transcoding”**, GSM Recommendation 6.10, version 3.2.0, janeiro de 1991.
- [56] International Telecommunications Union, **“7 kHz Audio-coding within 64 kbit/s”**, Recommendation G.722, novembro de 1988.
- [57] International Telecommunications Union, **“Coding of Speech at 8 kbit/s using Conjugate-structure Algebraic-code-excited Linear-prediction (CS-ACELP)”**, Recommendation G.729, março de 1996.
- [58] International Telecommunications Union, **“Pulse Code Modulation (PCM) of Voice Frequencies”**, Recommendation G.711, novembro de 1988.
- [59] International Telecommunications Union, **“Speech Coders: Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s”**, Recommendation G.723.1, março de 1996.
- [60] Website do **Global IP Solutions**, <http://www.gipscorp.com/>, último acesso em julho de 2009.
- [61] Website da **Global IP Solutions**, <http://www.gipscorp.com/files/english/datasheets/Codecs.pdf>, último acesso em julho de 2009.
- [62] International Telecommunications Union, **“Subjective Performance Assessment of Telephone-band and Wideband Digital Codecs”**, Recommendation P.830, fevereiro de 1996.
- [63] Collins D., **“Carrier Grade Voice over IP”**, McGraw-Hill, 2001.
- [64] International Telecommunications Union, **“Packet-based Multimedia Communications Systems”**, Recommendation H.323, fevereiro de 1998.
- [65] Rosenberg, J., Schulzrinne, H., et al., **“SIP: Session Initiation Protocol”**, RFC 3261, junho de 2002.
- [66] Fielding, R., Gettys, J. et al., **“Hypertext Transfer Protocol – HTTP/1.1”**, RFC 2616, junho de 1999.
- [67] Handley, M., Jacobson, V., **“SDP: Session Description Protocol”**, RFC 2327, abril de 1998.
- [68] S. Harris, **“The Tao of IETF – A Novice’s Guide to the Internet Engineering Task Force”**, RFC 3160, agosto de 2001.

- [69] Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V., **“RTP: A Transport Protocol for Real-Time Applications”**, RFC 3550, julho de 2003.
- [70] Website do **“Audio/Video Transport (AVT)”**, <http://www.ietf.org/html.charters/avt-charter.html>, último acesso em julho de 2009.
- [71] Website da **International Telecommunications Union**, <http://www.itu.int>, último acesso em junho de 2009.
- [72] Website do **RED (Random Early Detection) Queue Management**, <http://www.icir.org/floyd/red.html>, último acesso em junho de 2009.
- [73] Website do Techwhack, <http://stuff.techwhack.com/6151-nokia-skype>, acessado em junho de 2009.
- [74] Website do **The Wallstreet Journal**, <http://online.wsj.com/article/SB123836849558067525.html>, acessado em junho de 2009.
- [75] X. Xiao, L. Ni, **“Internet QoS: A Big Picture”**, IEEE Network, 13, 2:8-19, março/abril de 1999.
- [76] Callado, A., Fernandes, G., Silva, A., Barbosa, R., Sadok, D., Kelner, J., **“Construção de Redes de Voz sobre IP”**, Simpósio Brasileiro de Redes de Computadores (SBRC), maio de 2007.
- [77] Sat, B., Wah, B., **“Analyzing Voice Quality in Popular VoIP Applications”**, IEEE MultiMedia, janeiro de 2009.
- [78] Xie, H., Yang, Y., **“A Measurement-based Study of the Skype Peer-to-Peer VoIP Performance”**, Sixth International Workshop on Peer-to-Peer Systems (IPTPS 2007), fevereiro de 2007.
- [79] U. Jekosch, **“Sprache hören und beurteilen. Qualitätsbeurteilung von Sprechtechnologien als Forschungs- und Dienstleistungsaufgabe,”** tese de habilitação, Essen University, Essen, Alemanha, 2000.
- [80] Saint-Andre, P., **“Extensible Messaging and Presence Protocol (XMPP): Core”**, RFC 3920, outubro de 2004.