4
Transmissão de Voz em Pacotes nas Redes Celulares

Nos últimos anos, aplicações baseadas em voz sobre IP (VoIP) têm sido cada vez mais difundidas. O VoIP tradicional é uma aplicação de tempo real em modo *full duplex* com requisitos de atraso e perda de pacotes que procuram manter a qualidade da interação entre os usuários. Nos últimos anos alguns trabalhos procuraram demonstrar que a transmissão de voz em modo pacote pode aumentar a eficiência das redes celulares permitindo um maior número de chamadas simultâneas por canal [22].

O *Push-to-Talk over Celullar* (PoC) também é um serviço de transmissão de voz sobre IP. Ou seja, a voz capturada é digitalizada e separada em quadros para formar pacotes que são transmitidos através da interface aérea celular e, uma vez na parte terrestre da rede, são roteados por um protocolo padrão (que pode ser o próprio IP). A diferença crucial é o fato de que no PoC a transmissão acontece em modo *half duplex*. Ou seja, durante uma sessão, apenas um usuário de cada vez pode falar (*talker*) enquanto todos os outros escutam (*listeners*). A sinalização para estabelecimento, controle e finalização das sessões é baseada no SIP e, durante a sessão, a sinalização para gerenciamento do acesso ao meio é feita através de mensagens RTCP (protocolo de controle do RTP).

Serviços de rádio do tipo PTT (*Push-to-Talk*) têm sido explorados há anos. As soluções eram tipicamente baseadas em transmissão analógica e tinham cobertura limitada. Contudo, não existia um padrão. Em 2003 um consórcio de empresas difiniu especificações para o PoC que foram passadas para o OMA a fim de que fosse estabelecido um novo padrão.

Devido às suas características singulares, o PoC, possui requisitos de desempenho bem menos rigorosos que os de uma sessão VoIP tradicional. Mesmo assim, ele não deve ser oferecido para o usuário final sem que sejam observadas certas recomendações feitas com o objetivo de massificar o serviço, como o fato de que ele deve ser acessível para estações mais simples capazes de utilizar apenas 01 *slot*. Este tipo de recomendação tem impacto direto na implementação do PoC [19].

4.1. Requisitos de Qualidade em Aplicações VoIP

A transmissão de voz em redes de pacotes enfrenta problemas como perdas devido a congestionamentos, atrasos e armazenamento (pacotes que chegam fora dos limites de tempo para reprodução), além de atrasos devido à manipulação dos pacotes (geração, processamento dos cabeçalhos, filas em roteadores, etc). No caso de uma rede celular devemos adicionar ainda as perdas e atrasos devidos à interferência na parte aérea do percurso fim-a-fim e imperfeições do canal. Ressalte-se que, neste caso, “percurso fim-a-fim” significa todo o caminho percorrido entre o som emitido pelas cordas vocais do emissor e o som recebido pelo ouvido do receptor, ou seja, há ainda que se considerar o ruído introduzido no processo de codificação e decodificação da voz.

Em especial, devemos diferenciar o **atraso absoluto** da **variação do atraso** (conhecido como *jitter*) na chegada dos pacotes de um mesmo fluxo. Os efeitos do *jitter* são diminuídos através do armazenamento de uma certa quantidade de pacotes antes do início da reprodução no lado receptor (*bufferização*). A IETF padronizou um esquema de FEC que adiciona gradativamente informação redundante de cada quadro de voz aos pacotes seguintes (RFC 2198). Quanto maior for o tamanho deste *buffer*, melhor será a robustez da aplicação aos efeitos do *jitter*. Porém, o tempo de espera em *buffer* é um valor constante somado ao atraso absoluto de cada pacote. O atraso absoluto não deve ultrapassar um valor limite que geralmente varia entre 150 a até 400 ms (dependendo da aplicação) para o VoIP tradicional (*full duplex*). Vários algoritmos foram propostos a fim de garantir o melhor compromisso entre a solução destes dois problemas conflitantes. Em [18] podem ser encontradas várias referências de trabalhos sobre este tema.

Outro problema é a existência de lacunas geradas no fluxo de reprodução pela **perda de pacotes**. Nestes casos, algoritmos de recuperação de erro utilizados por alguns codificadores de voz (como o AMR) podem evitar, até certo ponto, os efeitos negativos no lado do receptor. Por outro lado, não é recomendada a recuperação de pacotes utilizando esquemas de retransmissão como o do TCP tradicional. Isso porque um pacote retransmitido dificilmente chegaria a tempo de ocupar sua posição dentro de um fluxo já em reprodução. Por este motivo, técnicas de correção de erros baseadas em FEC são mais utilizadas em aplicações de voz sobre IP, embora ainda assim deva haver uma preocupação com o tempo de processamento. Quanto mais poderoso for o FEC, em geral, maiores serão atraso e a banda utilizada, o que nos leva novamente à necessidade de uma solução de compromisso.

A **taxa de transmissão** pode variar de acordo com o tipo de codificador utilizado. Um exemplo disso é o AMR que foi adotado a fim de implementar esta filosofia na versão 98 do GSM (*Rel'98*). Neste padrão de codificação é possível variar a taxa de transmissão da voz entre **4.74** e **12.2 kbps** dependendo das condições do meio. Em caso de congestionamento, diminuir a taxa de codificação pode ser a saída mais rápida para evitar a perda excessiva de pacotes pelo simples fato de que a fonte em questão passa a contribuir menos para a degradação da rede. Mas essa medida, ao mesmo tempo, significa diminuir a qualidade da voz no receptor.

Por outro lado, sob boas condições, os quadros de voz podem ser codificados com taxas mais elevadas aumentando a qualidade na recepção. Isso, do mesmo modo, irá contribuir para aumentar o tráfego na rede e, em conseqüência, os níveis de interferência no caso de serviços oferecidos em redes celulares.

O interesse pelas aplicações de voz sobre IP deu origem a uma série de trabalhos com o objetivo de buscar soluções para as questões acima. A maioria tendendo para a implementação de soluções na camada de aplicação e optando-se, na camada de transporte, pelo UDP (a despeito do fato de que também existem trabalhos que propõem modificações no TCP). Ou seja, uma aplicação VoIP deve se adaptar dinamicamente ao estado corrente da rede (atrasos, taxa de perdas, banda disponível e nível de interferência) modificando **o tamanho do *buffer* de reprodução**, **o tipo de codificação** e a **quantidade de redundância**.

A questão que se segue é: Como a **qualidade de uma chamada de voz em modo pacote** pode ser mensurada? Um caminho óbvio seria a análise das opiniões subjetivas de um grupo suficientemente grande de pessoas para uma determinada quantidade de chamadas reais sob condições controladas. Esta idéia é explorada pelo **MOS** (*Mean Opinion Score*) muito utilizado para atribuir fatores de qualidade que variam entre 1 (péssima) e 5 (ótima) à voz ouvida pelo receptor após a sua passagem por um sistema.

Uma outra abordagem se baseia no monitoramento das condições da rede (taxa de perdas, atraso médio, etc) e no mapeamento dos valores observados em níveis de degradação do serviço de acordo com resultados pré-existentes. Um exemplo desse paradigma é o *E-model* proposto pelo ITU e que procura resolver o problema de uma forma objetiva. A principal contribuição do *E-model* é o fato de que a partir dele os efeitos de degradação causados por fontes independentes podem ser mapeados em fatores de uma equação cujo resultado (o fator R) define o nível de qualidade da chamada.

Outro modelo de performance, o PESQ (*Perceptual Evaluation of Speech Quality*),

compara o sinal original com a sua versão degradada no destino. O PESQ, por ser computacionalmente intensivo, geralmente não é utilizado para medir a qualidade de uma aplicação em tempo real. As referências sobre o assunto revelam que os valores medidos utilizando o PESQ e os resultados tabelados pelo ITU para o padrão MOS possuem uma correlação de aproximadamente 93% [9, 17].

4.2. Padrões para Medir Qualidade (MOS e o *E-Model*)

O *E-Model* é baseado no fato de que cada fator de degradação da qualidade de uma chamada pode ser calculado separadamente e somado. O resultado é denominado fator R e é dado pela eq. (1) cujo valor varia entre 0 (péssima qualidade) e 100 (ótima qualidade):

$$R = R_0 - I_S - I_D - I_E - A \quad \textbf{(1)}$$

Tanto R_0 quanto I_S dependem apenas da aplicação e, uma vez definidos, não variam de acordo com as condições de transmissão no meio. O termo R_0 representa a qualidade na presença de degradação causada por fontes de ruído (representadas pelos outros termos). O termo I_S representa as distorções causadas no processo de quantização, digitalização, etc, isto é, todas as distorções que acontecem durante a geração do sinal digitalizado.

A é baseado no fato de que a qualidade de uma chamada é julgada de forma diferente pelo usuário em função da conveniência de certas aplicações. Ou seja, a maioria dos usuários está disposta a aceitar um certo nível de redução na qualidade do serviço em troca de maior mobilidade, menor custo, etc. Valores típicos de A estão entre 0 e 20 dependendo do tipo de serviço.

As condições variáveis do meio de transmissão são mapeadas através do parâmetro I_D, que reflete a degradação causada por atrasos, e do parâmetro I_E, que reflete a degradação causada pela perda de pacotes.

A eq. (2) mapeia o fator R em MOS conforme estabelecido em [11]:

$$\begin{aligned} &Para\ R < 0,\ MOS = 1; \\ &Para\ R > 100,\ MOS = 4.5; \\ &Para\ 0 < R < 100,\ MOS = 1 + 0.035R + 7x10^{-6}\ R\,(R-60)\,(100-R) \end{aligned} \quad \textbf{(2)}$$

Além disso, conforme foi feito em [4], a equação do fator R pode ser reduzida à eq. (3) na qual R depende apenas dos parâmetros Id e Ie. O primeiro parâmetro é calculado como função do atraso absoluto (Ta), e o segundo, como função do codificador de voz utilizado (*codec*) e da taxa de perdas na rede (*loss*).

$$R = 93.4 - Id(Ta) - Ie(codec, loss) \quad \textbf{(3)}$$

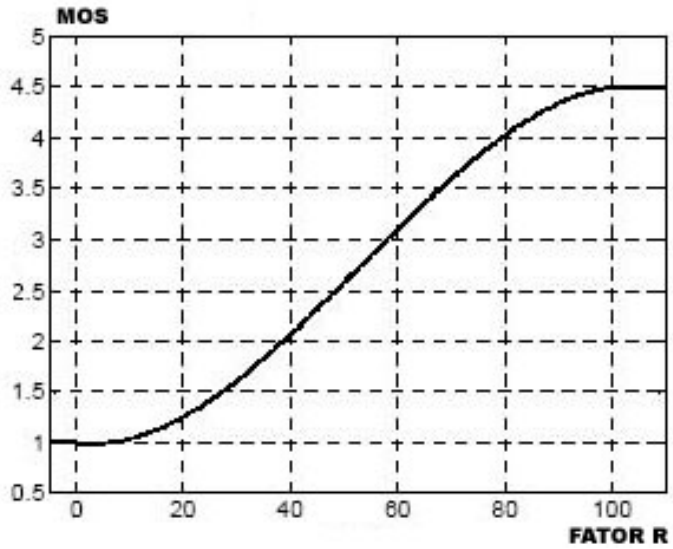

Figura 6: MOS versus *E-model* [11].

O trabalho realizado por HOENE em [9] faz um mapeamento dos valores do MOS através de medidas baseadas no PESQ de acordo com as recomendações do ITU utilizando um esquema como o mostrado na figura 7:

Figura 7: Esquema para a obtenção de valores MOS a partir do PESQ [9].

Os resultados obtidos mostram a variação do MOS de acordo com a taxa de perda de pacotes nos casos em que o AMR é implementado sem supressão de silêncio e, em outro caso, no modo DTX, ou seja, com supressão de silêncio (figuras 8 e 9, respectivamente). Em ambos os casos apenas um quadro de 20 ms de voz é encapsulado por pacote.

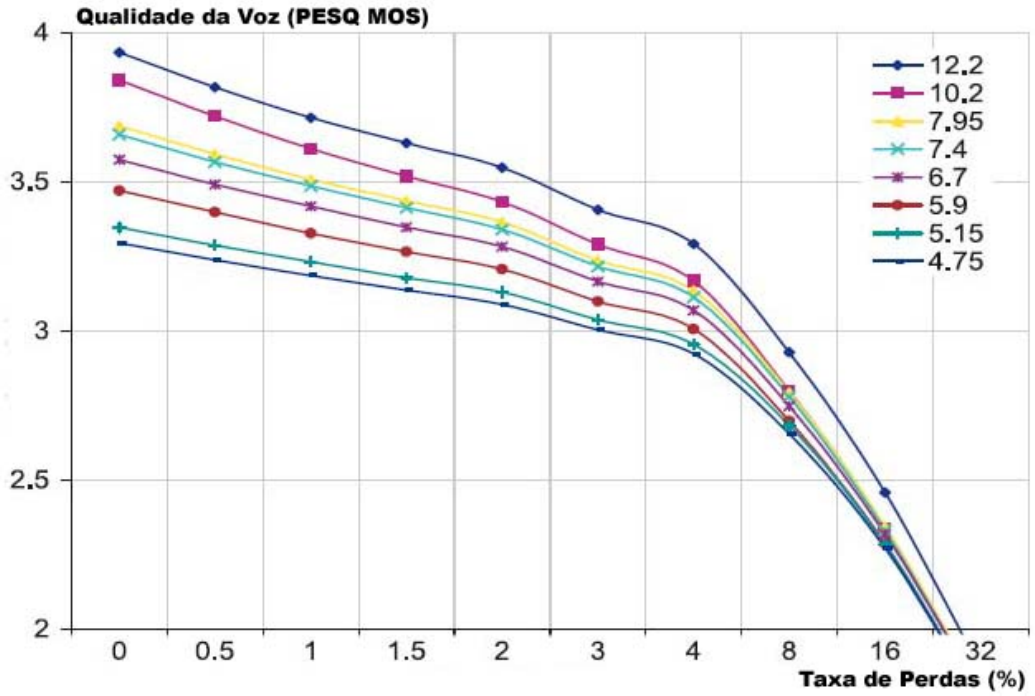

Figura 8: Variação do MOS para o AMR sem supressão de silêncio [9].

No gráfico da figura 9 a variação do MOS para as curvas em que não há supressão de silêncio (figura 8) é comparada com a variação do MOS onde a supressão de silêncio foi implementada. Neste caso é mostrado apenas o intervalo onde as perdas ficam entre 0% e 4%.

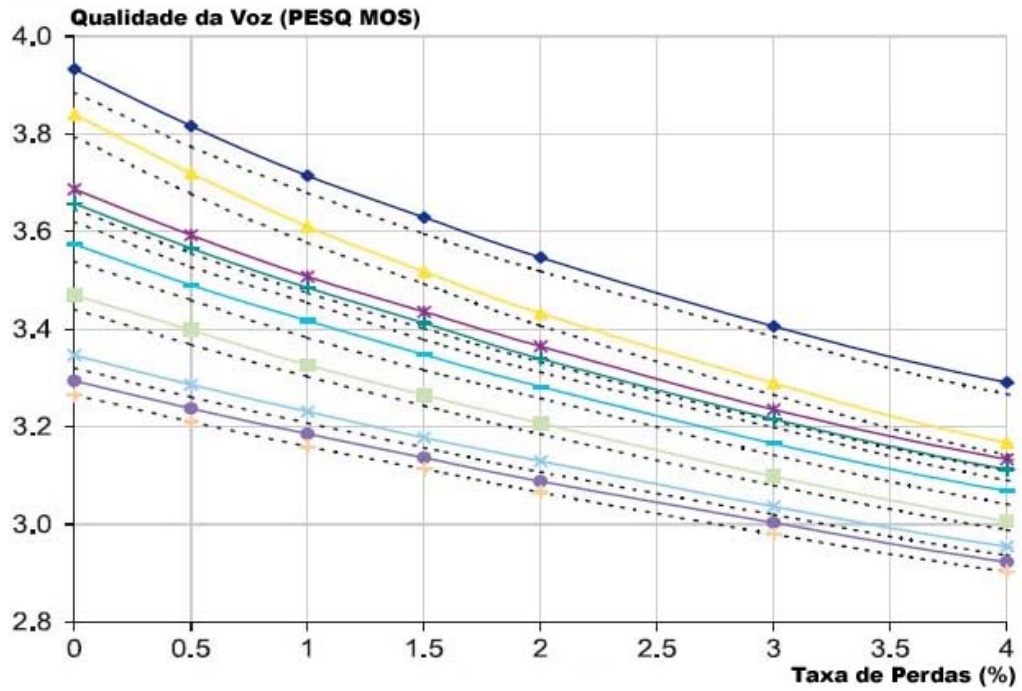

Figura 9: Variação do MOS para o AMR com (pontilhada) e sem supressão de silêncio [9].

Em HOENE [9] os resultados de simulação permitem afirmar que o nível de qualidade conforme medido acima não se altera de forma significativa quando um número maior de amostras por pacote é enviado (desde que o limite onde começam a haver perdas também por atraso não seja atingido).

4.3. Codificação com Taxas Adaptativas (AMR)

Inicialmente o GSM foi utilizado com os codificadores GSM-FR e GSM-HR. Com o intuito de oferecer qualidade equivalente às redes PSTN, foi desenvolvido o GSM-EFR e, em 1998, o AMR foi adotado na fase 2+ do GSM (*Rel'98*). Atualmente o AMR-WB (*Adaptative Multi-Rate Wideband*) é recomendado pelo 3GPP como padrão a ser utilizado em aplicações e serviços de terceira geração definidos pelo IMT-2000. Até a adoção do AMR, a maioria dos padrões de codificação de voz utilizava taxas fixas.

O AMR fornece 08 taxas de codificação conforme mostrado na tabela 2. Ele também implementa um detector de atividade da voz (VAD) para as estações que operam em modo DTX. Este detector de atividade decide se cada quadro é composto por voz ou por silêncio com base na energia do sinal amostrado. Os trechos de silêncio no discurso do emissor são codificados a uma taxa denominada SID (*Silence Descriptor*) que reproduz as características do silêncio produzindo o chamado "ruído de conforto". Além

disso, o AMR possui mecanismo de cancelamento de quadros perdidos que diminui os efeitos da perda de pacotes na rede.

Tabela 2: Taxas de codificação do AMR.

Modo	Taxa (kbps)	Classes			Total
		A	B	C	
AMR475	4.75	42	53	0	95
AMR515	5.15	49	54	0	103
AMR590	5.90	55	63	0	118
AMR670	6.70	58	76	0	134
AMR7.40	7.40	61	87	0	148
AMR7.95	7.95	75	84	0	159
AMR10.2	10.2	65	99	40	204
AMR12.2	12.2	81	103	60	244
SID	1.80	39	0	0	39

Após a geração da fala, o codificador AMR faz a amostragem do sinal a uma taxa de 8 KHz para gerar quadros de 20 ms (correspondendo a 160 amostras). Cada quadro de 20 ms de voz produz, 95, 103, 118, 134, 148, 159, 204 ou 244 bits de informação dependendo da taxa de codificação utilizada. Após a codificação da voz, os bits são separados em três categorias (A, B e C) conforme a sua importância. Durante a codificação do canal, tais bits são protegidos de acordo com a importância que lhe foi atribuída (codificação de canal mais poderosa ou menos poderosa).

Ou seja, utilizando AMR, cada quadro de voz enviado dentro de um pacote pode ter um tamanho que varia entre 95 e 244 bits. Apesar disso, o tamanho dos cabeçalhos permanece constante em 40 bytes sendo 12, 20 e 08 para o RTP, o IP e o UDP, respectivamente. Assim, a escolha da taxa de codificação AMR tem influência em vários fatores que afetam a qualidade do serviço, tais como atraso fim-a-fim e o *overhead* de protocolo. Alguns trabalhos procuraram demonstrar o ganho de desempenho obtido com taxas AMR variáveis de acordo com as condições do meio [21]. Porém, vale ressaltar que o trabalho de SEO [21] considera a variação entre todas as taxas possíveis, algo que não é previsto pelo padrão PoC [8], pois o chaveamento entre os modos durante uma rajada de voz contribui para a degradação no sinal [9]. No capítulo 05 desta dissertação são apresentados resultados referentes ao desempenho do PoC quando as taxas do AMR variam entre os valores 4.75, 7.40 e 12.2 kbps.

4.4. Sinalização de Sessões (SIP)

O SIP é um protocolo de sinalização implementado a partir da camada de aplicação padronizado pela IETF para controlar a **criação**, **alteração** e o **encerramento** de sessões entre dois ou mais participantes (RFC 3261). Neste contexto uma sessão pode ser uma vídeo-conferência, uma chamada de voz sobre IP, etc. Ele utiliza outro protocolo, denominado SDP (RFC 2327), para descrever uma sessão no momento de seu estabelecimento ou para modificar uma sessão ativa. Através dele, os terminais podem negociar (ou renegociar) parâmetros da sessão como, por exemplo, os tipos de mídia possíveis, codificadores pemitidos e até mesmo as velocidades de transmissão.

O SIP foi criado inicialmente para ser utilizado na Internet e é um protocolo cujas mensagens, baseadas em texto ISO, consomem uma quantidade de banda considerável. Porém, em sistemas celulares as restrições de banda são bem maiores. Nestes meios a taxa média de erros por bit (BER) assume valores típicos em torno de 10^{-3} (normalmente oscilando entre 10^{-2} e 10^{-4}).

Especialmente para resolver este problema, a IETF desenvolveu o padrão SigComp para compressão de protocolos de sinalização (RFC 3320) e, com base no SigComp, definiu logo em seguida um padrão de compressão das mensagens SIP (RFC 3486). A principal finalidade da redução do tamanho das mensagens SIP é atender às exigências estabelecidas para certos serviços multimídia em redes celulares principalmente com relação ao tempo máximo para o estabelecimento de uma sessão (*call setup*).

Na (RFC 3322), *SigComp Requirements & Assumptions*, através de um cálculo baseado numa rede WCDMA, foi mostrado que o tempo necessário para trocar todas as mensagens SIP (sem compressão) durante o estabelecimento de uma sessão pode ser maior que **04 segundos**. Isso sem considerar a necessidade de sinalização do sistema para a ativação de um canal de tráfego (autenticação do usuário, reserva de recursos pela BSS, etc). Nas redes celulares a sinalização SIP poderá concorrer com os dados do usuário em muitas aplicações. Isso porque esta abordagem é mais apropriada do que utilizar canais adicionais somente para este propósito. O próprio *Push-to-Talk over Cellular* se encaixa neste contexto. A recomendação para o tempo de estabelecimento de uma sessão PoC, no pior caso, não pode ficar acima de 04 segundos. No caso de aplicações oferecidas em redes EGPRS (cujas taxas são menores que as do WCDMA), tal problema de sinalização para estabelecimento de uma sessão se torna ainda mais problemático. Em [2] foi mostrado que, obedecendo ao critéro de que o terminal em uma

sessão PoC utilize somente um *slot* , o nível de compressão das mensagens SIP deve ser de 70%, no mínimo. No caso das mensagens RTCP, que são relativamente menores, a compressão não é necessária e nem exigida.

Além da compressão das mensagens SIP, a compressão de cabeçalhos (RTP, UDP, IP, etc) também é possível. Para isso a IETF definiu padrões de compressão na (RFC 3095) e na (RFC 2507). Vale ressaltar que a utilização de compressão deve ser usada de forma criteriosa uma vez que, por si só, ela tende a aumentar o atraso no processamento da informação. Além disso, todos os elementos intermediários (componentes da BSS, servidores, roteadores, etc) por onde o tráfego irá passar devem possuir meios para o tratamento da compressão principalmente nos casos em que ela ocorre na camada de rede.