

1 Introdução

1.1. Contexto Atual

Recentemente, o mercado de telecomunicações mundial tem enfrentado muitos dilemas. Como reduzir os custos, aumentar as receitas com novos serviços e manter a base de assinantes frente a crescente concorrência ?

Historicamente, os clientes sempre querem mais por menos, mais flexibilidade, mais rapidez, e às vezes, simplesmente “mais”. Esse comportamento pode ser visto no rápido crescimento no uso dos telefones celulares, que suportam cada vez mais novos recursos. A mentalidade de “*anywhere, anytime*” (à qualquer lugar, à qualquer hora) tem se tornado preponderante em meio ao fenômeno da globalização.

Com o crescente aumento do tráfego de dados na década de 90, e praticamente a estagnação do tráfego de voz, as operadoras começaram a perceber que tinham que integrar suas redes de voz e dados, ou no futuro, teriam problemas com o alto custo da infraestrutura das redes de voz. À medida que o tráfego de dados crescia, a idéia de uma rede única de transporte ganhava espaço.

Foi ficando cada vez mais claro que o pré-requisito para se chegar a idéia da “informação e comunicação à qualquer hora, em qualquer lugar e em qualquer forma” é a convergência das atuais redes – cada uma empregando diferentes tecnologias de transporte e controle – em uma única rede multiserviços baseada em pacotes, oferecendo serviços com diferentes qualidades e custos em uma plataforma aberta.

O conceito da NGN (*Next Generation Networks*) veio a fim de salvar as operadoras de telecomunicações de um futuro nebuloso. A NGN pode ser entendida como uma rede baseada em pacotes onde a comutação por pacotes e os elementos de transporte (roteadores, *switches* e *gateways*) são lógica e fisicamente separados do controle inteligente de serviços/chamadas. Esse controle inteligente é usado para dar suporte a todos os tipos de serviços

baseados em rede de pacotes, incluindo serviços básicos de voz (telefone) à dados, vídeo, multimídia, banda larga e gerenciamento de aplicações, que pode ser pensado como mais um tipo de serviço que a NGN irá suportar.

Como motivadores para o desenvolvimento e implantação da NGN, pode-se destacar os seguintes:

- Menor custo de operação (OpEx - *Operations Expenditures*)
- Menor custo de aquisição (CapEx - *Capital Expenditures*)
- Baixo custo de infra-estrutura
- Oferta de novos serviços
- Serviços com menor *time-to-market*
- Redução do *churn* (migração de usuários para outras operadoras)
- Controle centralizado
- Consolidação dos acessos banda larga

A pergunta pertinente neste momento é: qual a relação do protocolo SIP (*Session Initiation Protocol*) [1] com a NGN ?

O SIP é um dos principais protocolos utilizados na NGN. Seu desenvolvimento está fortemente relacionado com os motivadores listados acima. Ele é um protocolo desenvolvido basicamente para estabelecer, manter e desconectar qualquer tipo de sessão, inclusive chamadas ou conferências multimídia.

1.2. Motivação

A telefonia IP tem sido uma das tecnologias que mais tem se destacado ultimamente no meio científico e começa a fazer parte do planejamento estratégico de várias empresas de telecomunicações, que esperam que, com sua adoção, os custos de manutenção, operação e atualização baixem e a receita com novos serviços aumente.

Neste contexto, o SIP se destacou por ser um protocolo flexível, extensível e de fácil implementação em relação ao H.323 [2], padronizado pelo ITU-T. Até o presente momento, existem mais de 60 RFCs (*Request for Comments*) e 120 I-Ds (*Internet Drafts*) que possuem alguma relação com o SIP, o que demonstra sua atual importância na comunidade científica tecnológica.

No entanto, o desempenho do protocolo SIP para estabelecimento de chamadas telefônicas pouco tem sido estudado [3]. É importante frisar que o desempenho do protocolo SIP não é o único parâmetro a ser levado em consideração para avaliar o desempenho de chamadas telefônicas em redes baseadas no protocolo IP. Quando avaliamos o desempenho de chamadas telefônicas em redes IP, existe a necessidade de fazer uma separação entre desempenho da sinalização, da voz e da rede. Avalia-se o desempenho dos seguintes itens :

- Terminais - Sendo o parâmetro principal, o *mouth-to-ear delay* (M2E) [4], [5] e [6];
- Servidores *proxy/redirect* - Através de parâmetros como chamadas por segundo, BHCA (*Busy Hour Call Attempts*), registros por segundo e carga de processamento [6];
- Rede – Através dos parâmetros retardo fim-a-fim (*end-to-end delay*), perda de pacotes, *jitter* e banda;
- Codificação da voz – É quantificado através do tempo para codificação/decodificação, resistência a perda de pacotes e a rajada consecutiva de perda de pacotes (testando assim o *loss-concealment* do *codec*) [8], [9] e [10].

Com exceção da codificação da voz, os outros itens possuem influência direta em maior (rede) ou menor (terminais) escala no desempenho da sinalização através do protocolo SIP. Existe uma série de modelos para avaliação da qualidade da voz, como MOS [12], PSQM [13], PSQM+ [14] e PESQ [15]. A qualidade da voz é um assunto que já foi muito explorado e existem muitas referências a respeito. Diferentemente, o desempenho da sinalização para chamadas telefônicas via IP, especificamente com o protocolo SIP, é um assunto pouco estudado.

Um requisito básico para a telefonia IP é o de manter ou melhorar a qualidade das chamadas telefônicas através da PSTN (*Public Switched Telephone Network*). Com o desenvolvimento das centrais digitais (CPAs) e da Sinalização Número 7 (SS7) houve um tremendo avanço na qualidade das chamadas telefônicas. Há uma extensa quantidade de recomendações do ITU-T que tratam de requisitos mínimos de qualidade para diversos parâmetros do SS7.

É importante notar que o desempenho da sinalização de chamadas VoIP tem a ver com a arquitetura de rede utilizada [16]. Antes, na PSTN, a inteligência residia na rede e os terminais eram burros. Agora, com a telefonia IP, acontece o inverso. Portanto, os softwares instalados nos terminais, agora, estão intrinsicamente relacionados com o desempenho tanto da sinalização, quanto da voz.

A recomendação E.721 [17] define objetivos para o *post-dial delay* (PDD), *post-pickup delay* (PPD) e *call release delay* (CRD), que serão definidos no capítulo 3, no entanto, o *post-dial delay* das redes telefônicas atuais são bem menores que os definidos na recomendação. A AT&T, por exemplo, afirma que o *post-dial delay* em sua área de concessão é inferior a 2 segundos [18]. Ainda segundo a AT&T, antes do desenvolvimento do CCS (*Common Channel Signaling*), o *post-dial delay* variava de 15 a 20 segundos.

O órgão que regulamenta e fiscaliza as telecomunicações na Malásia [19] definiu seus próprios objetivos de qualidade, entre eles o PDD. Os valores são um pouco mais rigorosos que os definidos na E.721, e não há diferenciação de valores baseado no volume de tráfego. Para chamadas intra operadora fixas, 95% das chamadas devem ter PDD menor que 10 segundos. Para chamadas inter operadoras fixas, 95% das chamadas devem ter PDD menor que 13 segundos.

Já o órgão que regulamenta as telecomunicações na Nigéria [20], fixou os seguintes objetivos, válidos inclusive para o horário de pico (HMM).

Tipo de Chamada	Post-Dial Delay
Local	< 5 segundos
Nacional	< 10 segundos
Internacional	< 15 segundos

Tabela 1 – Objetivos de PDD da Nigéria

No entanto, todos estes objetivos se referem a chamadas na PSTN e portanto, ainda há a necessidade de se estudar e fixar objetivos para chamadas em redes IP [3].

1.3. Trabalhos Anteriores

Em [3], Eysers e Schulzrinne utilizaram curvas de retardo e perda de pacotes UDP na Internet obtidas através do projeto *Surveyor* [21] como base do

estudo para chegar a valores de PDD para o SIP e H.323 entre diferentes localidades. O menor valor e valor de 95% do PDD foram apresentados graficamente durante um período de 90 dias. Foi usado 1 hora de dados em um determinado horário para cada dia. A simulação foi computacional, ou seja, o funcionamento do SIP e do H.323 foram modelados e aplicados às curvas de retardo/perda. Os autores mostram a influência do redirecionamento de uma chamada SIP e das retransmissões por *time out* do SIP (sobre UDP) no PDD. Os valores máximos de 95% do PDD foram inferiores a 2 segundos.

Em [22], os autores se basearam nas recomendações Q.709 [23], Q.725 [24] e Q.706 [25] para chegar aos requisitos de PDD para PSTN. Ainda em [22], os autores calcularam o *call setup delay* para alguns ambientes VoIP (*Voice over IP*) com protocolo Megaco [26] e compararam com os tempos obtidos para PSTN. A conclusão chegada pelos autores, é que os tempos de PDD do ambiente VoIP são comparáveis aos da PSTN. No entanto, a recomendação E.721 [17] não foi citada. De acordo com [3], apesar de [22] se basear na ITU-T para chegar a requisitos de PDD para PSTN, o mesmo não levou em consideração a recomendação E.733 [27], que define o *queueing delay* para mensagens de sinalização.

Já em [16], montou-se vários ambientes experimentais fazendo uso de clientes e servidores *proxy* com o SIP *stack* da NIST [28] baseado em JAIN [29] e usando apenas dois PCs conectados através de LAN. Os resultados mostraram que a carga de processamento nos terminais (com o aumento do número de chamadas simultâneas em um mesmo SIP *client*) e a presença de *proxies* no caminho da sinalização são os fatores que tem maior peso no aumento do PDD. Ainda segundo [16], os fatores que influenciam o desempenho da sinalização de chamadas VoIP em um ambiente LAN é diferente de ambientes como a Internet. Em ambiente LAN, fatores como processamento das mensagens de sinalização, número de sessões simultâneas e número de *proxies* intermediários são os fatores que mais afetam o PDD.

Curcio e Lundan [30] montaram um ambiente experimental e mediram o PDD, PPD e CRD para chamadas VoIP com protocolo de sinalização SIP variando o tipo de chamada (local, nacional, internacional e 3G). Em [31], Curcio e Lundan apresentam valores de PDD obtidos para chamadas VoIP com SIP para diferentes tipos de chamadas em ambiente LAN, WAN e WLAN. Para alguns tipos de chamadas usou-se servidores *proxy* e para chamadas locais, variou-se a perda de pacotes.

Senesi et al. [32] testaram uma solução de telefonia IP (protocolos MGCP e ISUP) com ênfase na avaliação da qualidade da sinalização e na qualidade de voz. O foco do trabalho foi mostrar como a adição de QoS melhora entre outros indicadores, o PDD.

Kist e Harris [33] apresentaram valores de PDD para chamadas realizadas através do protocolo SIP em redes baseadas no IMS (IP Multimedia Subsystem), onde o número de *proxies* geralmente é maior que no caso de telefonia fixa. Os autores fazem suposições em vários casos baseados em valores de outros trabalhos e o trabalho foca principalmente no impacto da resolução de nomes através de DNS [34] no PDD. Já em [35], avalia-se o PDD de uma rede IP baseada no 3GPP [36]. Os resultados foram obtidos através de modelos matemáticos e através de simulações computacionais. Foram obtidos gráficos de PDD x (carga de processamento do servidor), PDD x (tempo de processamento do servidor) e PDD x (*one way delay*). Os resultados foram comparados com a E.721 e ficaram abaixo do recomendado na maior parte dos casos.

Em [37], foram medidos o PDD e PPD para chamadas realizadas com e sem mecanismos de segurança. O ambiente de teste envolveu três servidores de DNS, dois proxies e dois clientes SIP.

Finalmente, em [38], Camarillo et al. simularam computacionalmente e compararam o desempenho do SIP sobre o protocolo UDP [39], TCP [40] e SCTP [41]. Os autores concluem que retransmissões de protocolos de aplicação (como o SIP) via UDP podem ser úteis para pequenos volumes de tráfego de sinalização, mas não apropriado para grandes volumes. Neste caso, recomenda-se o uso do TCP ou SCTP. Além disso, concluem que o SCTP é de alguma forma melhor adaptado ao transporte de sinalização que o TCP, provendo um desempenho um pouco melhor do ponto de vista do usuário. Em geral, quanto piores as condições da rede, melhor o desempenho oferecido pelo SCTP em comparação com o UDP e TCP.

Nota-se que em todos os trabalhos relacionados acima, é mostrada em parte ou não mostrada a influência da variação de parâmetros como perda de pacotes, retardo de propagação, *jitter*, banda e protocolo de transporte nos tempos de *post-dial delay*, *post-pickup delay* e *call release delay*. Portanto, há necessidade de realizar um estudo mais apurado, variando cada um destes parâmetros para se entender qual o impacto de sua variação no desempenho da sinalização através do protocolo SIP.

1.4. Objetivo

O objetivo desta dissertação é avaliar o desempenho do protocolo SIP para estabelecimento de chamadas telefônicas em redes IP.

O desempenho de protocolos de sinalização para estabelecimento de chamadas é classicamente avaliado pelas seguintes medições: *post-dial delay* (PDD), *post-pickup delay* (PPD) e *call release delay* (CRD) [17].

Essas medições serão aplicadas ao protocolo SIP de acordo com as definições feitas em [17] para o SS7.

Serão montados ambientes experimentais nos quais a variação de parâmetros – como retardo, perda de pacotes, *jitter*, largura de banda e protocolo de transporte – será testada, permitindo verificar como esses parâmetros afetam os tempos de PDD, PPD e CRD.

Os resultados serão comparados com [17], que traz os tempos máximos recomendados destas medições para chamadas telefônicas na PSTN, e quando possível, com os resultados de trabalhos anteriores, citados na seção 1.3.

1.5. Principais Contribuições

Como contribuições deste trabalho, pode-se listar as seguintes:

- Realização de medidas em ambiente real de PDD, PPD e CRD mostrando o impacto do jitter e perda de pacotes em diversos tipos de chamadas. Estas medidas foram pouco exploradas em trabalhos anteriores. Os valores podem ser utilizados como referência em trabalhos futuros.
- Identificação de situações onde alguma medida não atenda aos objetivos de qualidade definidos na recomendação E.721. Neste caso, serão sugeridos novos objetivos.
- Comparação do SIP sobre UDP e SIP sobre TCP em um ambiente experimental. Este tipo de teste não foi realizado em nenhum dos trabalhos anteriores citados na seção 1.3. Este tipo de comparação visa a identificação de situações onde um protocolo se adapta melhor que o outro.
- Comparação das medidas com trabalhos anteriores.

- Sugestões de novos testes para avaliação do desempenho do protocolo SIP em Trabalhos Futuros, seção 5.1.

1.6. Estrutura da Dissertação

No capítulo dois, o protocolo SIP será apresentado e detalhado baseado na RFC 3261. Serão apresentadas todas as entidades e mensagens definidas, assim como os principais conceitos. A compreensão do funcionamento do protocolo auxilia, mas não impede, o entendimento dos próximos capítulos. As funcionalidades apresentadas se limitam ao uso do SIP para estabelecimento de chamadas telefônicas. Não serão tratadas de questões de segurança do protocolo ou envio de mensagens instantâneas, por exemplo. Também não serão mostrados os protocolos SDP [42] e RTP/RTCP [43] por não estarem relacionados com o objetivo desta dissertação, que foca o desempenho da sinalização.

No capítulo três, serão descritos os objetivos do experimento e serão definidos os parâmetros de qualidade (PDD, PPD e CRD), assim como as variáveis dos experimentos (tipo de chamada através da variação do retardo de propagação, *jitter*, perda de pacotes e largura de banda). A metodologia adotada será descrita, os *softwares* e *hardware* apresentados e os cenários de experimentação definidos.

No capítulo quatro, as medições experimentais para as diversas situações serão apresentadas e comentadas. Os resultados serão apresentados graficamente através da média e valor de 95% do PDD, PPD e CRD. Quando houver perda de pacotes, os piores valores desses indicadores também serão listados para referência, juntamente com os respectivos motivos que levaram a degradação dos indicadores. Os resultados serão comparados com a recomendação E.721 e com trabalhos anteriores.

No capítulo cinco, serão destacadas as principais conclusões deste trabalho, assim como sugestões de trabalhos futuros.

E finalmente, no capítulo seis, as referências serão listadas.