2º SIMPÓSIO BRASILEIRO SOBRE REDES DE COMPUTADORES (2º SBRC)

REDES LOCAIS COM INTEGRAÇÃO DE SERVIÇO DE VOZ E DE DADOS

Wagner Luiz Zucchi, EPUSP, FDTE
End: Cidade Universitária "Armando Salles de Oliveira", Setor Amarelo, Bloco N-2, Caixa Postal 11.455, SP

Wilson Vicente Ruggiero, PhD, Scopus Tecnologia S/A.
End: Rua Bela Cintra, 881 - 8º andar, Cerqueira Cesar, CEP: 01415 - São Paulo - SP

Resumo: O trabalho discute os principais problemas para a transmissão de voz em redes locais de computadores. Examina a possibilidade de integração do serviço de voz na arquitetura de redes locais proposta pelo projeto IEEE 802, e apresenta um modelo de um terminal codificador/decodificador de voz que permita a análise dos parâmetros de projeto desta classe de sistema.

REDES LOCAIS COM INTEGRAÇÃO DE SERVIÇOS DE VOZ E DE DADOS

1. INTRODUÇÃO

O rápido desenvolvimento de redes de computadores de pequeno ou médio porte e com dispersão geográfica em torno de alguns poucos quilômetros - que são conhecidas como redes locais - tem chamado a atenção dos pesquisadores para a possibilidade de integração de diversas categorias de serviço numa mesma rede, visando o melhor aproveitamento da capacidade oferecida pela rede e a economia de recursos de comunicação.

Uma categoria de serviços, particularmente importante para tarefas como controle de processos e automação de escritórios são os serviços chamados de "tempo real". Dentro desta categoria destaca-se a transmissão da voz humana como um serviço que tem recebido especial atenção.

O objetivo deste trabalho é fornecer alguns elementos que permitam a construção de um modelo de rede local com serviço integrado de voz e de dados que possa, dado uma certa configuração da rede, avaliar seu desempenho.

A seção 2 caracteriza os serviços e protocolos de tempo real, pondo em relevo os requisitos para que estes protocolos possam operar numa rede de computadores.

A seção 3 mostra como os serviços de tempo real podem ser encaixados na arquitetura de redes locais proposta pelo projeto IEEE 802.

A seção 4 apresenta as técnicas de codificação de voz mais importantes e os problemas típicos dos protocolos de transmissão de voz em redes de computadores.

A seção 5, finalmente, aborda o modelo das estações de voz analisando os elementos básicos que estas devem conter para a realização de sua tarefa, e colocando em evidência os parâmetros que devem ser estudados através do modelo.

2. PROTOCOLOS PARA SERVIÇOS EM TEMPO REAL

A comunicação numa rede de computadores é constituída por mensagens trocadas entre os processos residentes nos diversos nós da rede. Diremos que um processo P exige da rede serviços em tempo real, se existe um intervalo de tempo $\bar{t}$, finito, tal que qualquer mensagem originada pelo processo P no instante t_0 deve ser entregue no seu destino no intervalo $(t_0, t_0 + \bar{t})$, sob pena de perder o seu conteúdo informativo.

É importante observar que o valor de $\bar{t}$ depende exclusivamente da natureza física do processo P, sendo totalmente independente das características da rede que realiza o serviço de comunicação.

Assim, por exemplo, imaginemos um processo que é capaz de receber e codificar as cotações diárias da Bolsa de Valores e transmití-las a um conjunto de investidores, através de uma rede de computadores e respectivos terminais. Imaginemos que este processo receba as cotações no momento em que a Bolsa encerra o seu expediente, no final do dia, e que estas cotações devam ser transmitidas até a manhã seguinte para que os acionistas possam planejar seus negócios. É óbvio que, se as cotações chegarem aos terminais após a abertura da Bolsa na manhã seguinte, elas serão totalmente inúteis aos acionistas. Conclui-se então que este processo exige um serviço de tempo real, pois possui uma restrição de atraso de comunicação que é independente do serviço de comunicação fornecido pela rede.

Chamamos de protocolo de tempo real ao conjunto de regras e procedimentos que é capaz de implementar um serviço em tempo real. O exemplo acima mostrou que estes protocolos nada têm de incomum; de fato, o serviço sugerido no exemplo poderá ser implementado por quase todos os protocolos existentes nas atuais redes de computadores, mesmo por aqueles que foram especificados sem nenhuma pretensão de atenderem serviços de tempo real. Em alguns destes protocolos não existe uma certeza determinística de que o atraso de transmissão seja menor do que $\bar{t}$, porém, estatísticamente falando, ocorre que a probabilidade de

o atraso ser maior do que $\bar{t}$ é desprezível para $\bar{t}$ suficiente - mente grande (o quão grande $\bar{t}$ deve ser, depende das caracterís ticas do atraso na comunicação e da confiabilidade do serviço que se pretende implantar).

Devemos porém, distinguir duas classes de serviço de tempo real: sistemas de tempo real com requisitos críticos (rígidos) e sistemas de tempo real com requisitos não rígidos

Seja $\bar{tc}$ o atraso médio de mensagens numa rede de comunicação. Se $\bar{t} \cong \bar{tc}$, e a não observância do atraso máximo $\bar{t}$ implica na inutilidade do serviço, então esta aplicação utilizando a rede de comunicação com atraso médio $\bar{tc}$ constitui um sistema de tempo real rígido e crítico. Exemplos destes sistemas são comumente encontrados em sistemas especiais de controle de pro cessos.

Por outro lado, se $\bar{tc} << \bar{t}$, podemos dizer que temos um sis tema de tempo real flexível e não crítico (veja o exemplo dado).

Devemos ressaltar ainda, que esta divisão dos sistemas de tempo real depende também da confiabilidade que a aplicação exige do sistema de comunicação. Podemos afirmar, como princí - pio geral que, dado um sistema de comunicação, quanto maior a confiabilidade da transmissão tanto maior o atraso médio. A figura 1 mostra uma relação típica entre a confiabilidade e o atraso médio.

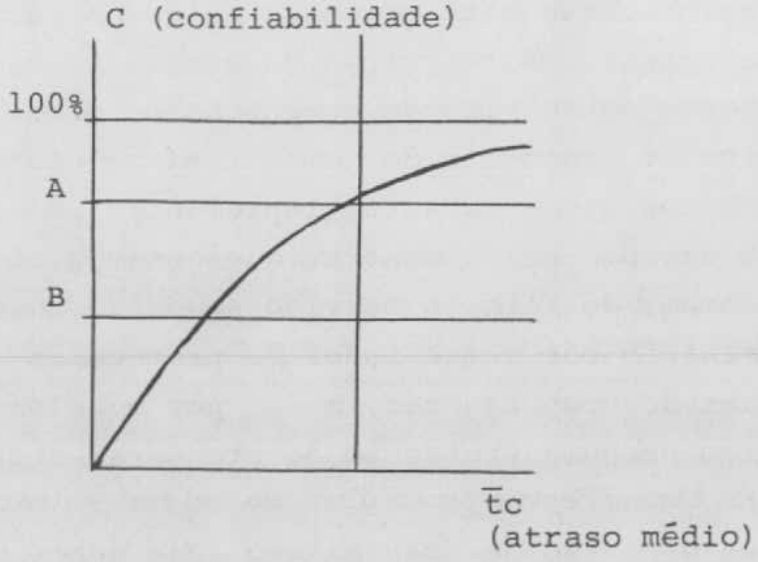

Figura 1

Se a aplicação exigir uma confiabilidade "A" (veja Figura 1), então o sistema será rígido e crítico. Porém, se a confiabilidade exigida for apenas B, o sistema será flexível.

A conclusão destas considerações é que os protocolos já utilizados para os serviços de transmissão de dados desde que convenientemente adaptados, podem ser utilizados para o estabelecimento de serviços de tempo real. Além disso, diversos estudos mostraram que para transmissão da voz humana (que é, talvez, o processo de tempo real mais importante a ser considerado para integração em redes de computadores) os sistemas de comutação de pacotes, ora em uso, podem ser muito eficientes (ref. 1, 2).

Devemos portanto, considerar que existem duas classes de protocolos para transmissão em tempo real. Na primeira classe estão os protocolos que incorporam mecanismos determinísticos para limitação do atraso mesmo às custas de alguma perda de eficiência na transmissão. Por outro lado, temos protocolos que limitam o atraso na comunicação de forma estatística, correndo o risco de alguma perda de informação nas transmissões em tempo real.

Todavia, qualquer que seja o mecanismo utilizado para a limitação do atraso na comunicação, a definição que demos de serviço de tempo real possui dois corolários que o distinguem de um serviço de tempo não real.

a. O controle de fluxo num protocolo de tempo real deve basear-se na taxa e não na quantidade de dados transmitidos.
   Num protocolo de tempo não real o intervalo de permanência de uma mensagem na rede não é limitado e, portanto, devemos limitar a quantidade máxima de dados presentes na rede em qualquer instante, a fim de proteger os recursos desta.
   Nos protocolos de tempo real, sabemos que o máximo tempo que uma mensagem permanece na rede é $\overline{t}$, e portanto, o que devemos limitar é a quantidade de dados que entram na rede durante o intervalo $\overline{t}$, o que significa limitar a taxa de informação. Na verdade, os sistemas de tempo real caracterizam-se tipicamente por uma taxa de produção e de consumo constante.

b. Nos serviços de tempo real as mensagens mais velhas devem ser desprezadas preferencialmente às mais novas.

Nos sistemas de tempo não real, toda vez que um nó da rede precisa escolher uma mensagem para descartar, ele toma a mais nova para preservar a sequencialização das mensagens. Contrariamente, nos sistemas de tempo real, é a mais velha que deve ser descartada, pois tem menor probabilidade de chegar ao seu destino em tempo hábil.

Uma vez definido e caracterizado um protocolo de tempo real, devemos justificar a integração dos serviços de tempo real numa rede de computadores. Podemos dizer, que esta integração é motivada principalmente pelos seguintes fatores:

1. A redução dos custos de processamento em relação aos custos de vias de comunicação verificada nos últimos anos, tornou vantajosa a utilização de uma única via de comunicação para serviços diversos ainda que às custas de um maior processamento em relação à utilização de canais específicos para cada tipo de serviço.

2. A verificação de que em muitos sistemas de comunicação de dados existem normalmente capacidades ociosas que poderiam ser usadas com outras classes de serviço (ref. 4).

3. Os recentes desenvolvimentos na área de interpretação da voz humana e de imagens por computadores, sugere para o futuro próximo um sistema altamente interativo entre homens e máquinas para os quais as redes integradas seriam um veículo de comunicação adequado.

3. SERVIÇOS DE TEMPO REAL E A PROPOSTA DE PADRONIZAÇÃO IEEE-802 PARA REDES LOCAIS

Definidas as características dos serviços de tempo real, dos quais a transmissão de voz interativa é um caso particular, vejamos como este tipo de serviço pode ser integrado numa rede local de computadores.

Em princípio, podemos dizer que se a rede for capaz de atender aos requisitos críticos de atraso e de confiabilidade do serviço de voz então a rede pode suportar esta classe de serviço.

Todavia, como dissemos no capítulo anterior um serviço de tempo real numa rede de computadores possue associado a si um protocolo de tempo real com características próprias que o distingue dos utilizados em comunicação de dados. Uma vez que existe uma proposta de padronização de protocolos para redes locais em vias de aprovação, conhecida como projeto IEEE 802 (ref. 3), devemos discutir se os protocolos oferecidos por esta proposta são compatíveis com serviços de tempo real e, particularmente, com serviço de transmissão de voz.

Examinando a proposta do projeto IEEE 802, verificamos que ela apresenta a estrutura de protocolos mostrada na (Figura 2), sendo que o nível de controle do acesso ao meio pode operar com 3 tipos diferentes de protocolo de acesso: CSMA-CD e Passagem de "token" (permissão) para redes em via comum (" token bus "), e passagem de "token" para redes em anel ("token ring").

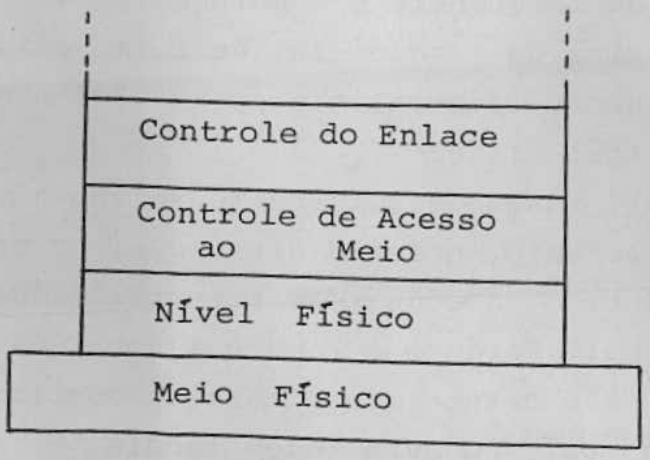

Figura 2 - Arquitetura de protocolos da recomendação IEEE 802

Em primeiro lugar, quanto ao nível físico, observamos que a proposta 802 abrange a maioria das técnicas de transmissão empregadas em redes locais, quer utilizando transmissão em banda base, quer em banda larga. A única limitação imposta pelo nível físico é a taxa de bits transferidos, o que limita a quantidade total de dados que pode ser oferecida à sub-rede.

A utilização de canais banda larga apresenta o atrativo de que diversas classes de serviço podem ser fisicamente separadas evitando-se assim mútuas interferências. Cada serviço possuiria um determinado desempenho que dependeria apenas da porção de espectro que lhe é alocada, mas não das condições de operação dos demais serviços. Todavia, os canais banda larga e as repectivas interfaces são caros e sua confiabilidade e disponibilidade frequentemente ultrapassa as necessidades de transmissão de pequenas e médias empresas.

A configuração em banda base também pode ser utilizada com tráfego de diferentes serviços, porém aparece aí o efeito de mútua interferência. Os mecanismos de controle de fluxo (que são implementados por níveis superiores de protocolos),devem evitar não apenas o congestionamento da sub-rede, mas também que um serviço muito prioritário degrade os demais.

Além de variações na tecnologia de transmissão, podemos pensar também em misturar diferentes tipos de comutação para as diversas classes de serviço. Assim, por exemplo, num canal banda larga poderíamos utilizar comutação de dados por pacotes numa faixa de frequência e comutação de voz por circuitos numa outra faixa de frequência (de forma análoga a um multiplex FDM de grupo). Todavia, a proposta 802 não prevê a utilização de comutação mista.

Em segundo lugar, no que diz respeito ao sub-nível de controle de acesso ao meio, podemos dizer que, em princípio, todos os protocolos recomendados podem ser utilizados para serviços de tempo real. Porém é evidente que protocolos onde o atraso máximo é limitado são naturalmente mais apropriados para aplicações de tempo real. Todavia redes locais integradas

têm sido implementadas com protocolo CSMA-CD ou modificações deste, e bons resultados foram obtidos (ref. 4)quando se garante que a rede opere com baixa carga.

A razão deste fato, a primeira vista surpreendente, é dupla: primeiramente, os protocolos de acesso aleatório (tipo CSMA-CD) possuem baixos atrasos e pequena variância de atraso em condições de baixa carga.

Por outro lado, como dissemos na seção anterior os serviços de tempo real possuem associados a si uma certa confiabilidade crítica menor do que a unidade. Se os pacotes transmitidos com atraso maior do que o crítico puderem ser descartados sem comprometer a confiabilidade da aplicação, então o protocolo de acesso aleatório pode ser utilizado nessa aplicação de tempo real.

Essa conclusão, aliada à relativa facilidade de implementação da interface CSMA-CD, mostra que a utilização deste tipo de protocolo em redes integradas não deve ser descartado 'a priori', pois mesmo este tipo de protocolo pode ser útil em algumas configurações de redes.

Por outro lado, a utilização de protocolos determinísticos para serviços integrados (a proposta 802 inclui o "token bus" e o "token ring") deve levar em conta que o atraso máximo de transmissão não deve superar o atraso crítico do serviço de tempo real e que mesmo nestes tipos de protocolo a variância do atraso pode não ser pequena. Além disto, é importante mencionar que a proposta 802 não permite canais banda larga em protocolos "token ring".

Finalmente, devemos examinar a influência do subnível de controle do enlace sobre a aplicação de tempo real. A proposta 802 divide o serviço oferecido por este subnível em duas classes de serviço.

A primeira classe (chamada classe I) é um serviço de transferência de dados sem o estabelecimento de conexão entre os dois usuários, não há retransmissões, nem confirmações de entrada de mensagens, ele corresponde a um serviço de datagrama. Este tipo de serviço é compatível com protocolos de tempo real. Se um nível de confiabilidade mais alto do que o

oferecido pela sub-rede for desejado, ele deve ser implementado pelos níveis superiores de protocolo.

A segunda classe (chamada classe II) corresponde a um serviço de transferência de dados com estabelecimento de conexão entre os usuários. Ele provê retransmissões e recuperações de situações de erros em nível de enlace. Esta classe é muito parecida com o protocolo HDLC. Devido aos atrasos que podem ser introduzidos pelos mecanismos de recuperação, esta classe de serviço não é recomendável para aplicações de tempo real típicas.

A proposta 802 especifica que todas as estações conectadas à rede devem oferecer aos seus usuários o serviço de classe I, de modo que a implementação de protocolos de tempo real não traz problemas de compatibilidade em nível de enlace.

Outro problema, que diz respeito especialmente à aplicação de transmissão de voz, e que tem repercussões no nível de enlace é a sinalização da chamada telefônica, isto é, a maneira como uma estação de voz avisa a outra estação que uma ligação deve ser estabelecida entre elas para transmissão de voz, e o envio de possíveis respostas (chamado, ocupado, etc.).

Este problema pode ser resolvido de diversas maneiras compatíveis com a proposta 802 sem que nenhuma delas apresente uma vantagem sobre as demais. Apresentamos a seguir algumas das possíveis soluções:

a. Transmitir os sinais de progresso de chamada em pacotes de dados utilizando serviço da classe I. Esta solução torna o serviço de voz totalmente transparente ao nível de enlace, mas esbarra com o problema de que as mensagens de sinalização devem ter alta confiabilidade que não é oferecida na classe I.

b. Utilização de um comando especial de nível de enlace para a sinalização. Ao contrário do sistema anterior, esta mensagem pode ter alta confiabilidade em nível de enlace, porém o serviço não seria mais transparente ao nível de enlace. A proposta 802 dá margem a que mensagens especiais de troca

de identificação sejam utilizadas no serviço de classe I.

c. Utilizar o serviço de classe II para sinalização e o serviço de classe I para troca de informação. Esta solução possui o inconveniente de que todas as estações de voz deverão implementar ambas as classes de serviços. Uma vantagem atraente desta solução é a facilidade de uma futura interconexão da rede local com a rede pública telefônica.

d. Finalmente, poderíamos pensar em implementar a sinalização através de um serviço com reconhecimento (e, portanto, confiável) porém sem conexões. Este tipo de serviço encontra-se atualmente em estudos pelo grupo de trabalho IEEE 802.2.

4. CARACTERÍSTICAS DA TRANSMISSÃO DE VOZ EM REDES DE COMPUTADORES

O esquema básico de qualquer sistema de transmissão de voz (Figura 2) compreende os seguintes elementos:

a. Elemento de Digitalização: é responsável pela amostragem e quantização do sinal de voz analógico. Sabemos que para sinais de voz esta digitalização deve fornecer no mínimo 8000 amostras por segundo para termos uma qualidade aceitável do sinal após recuperação.

b. Elemento de Codificação: deve transformar as amostras fornecidas pelo digitalizador numa sequência de pulsos binários que possam ser transmitidas pela rede digital. Existem diversas técnicas de codificação utilizadas para sinais de voz, das quais a mais conhecida (difundida) é o PCM (Pulse Code Modulation).

Um sistema PCM típico opera com uma taxa de 64.000 bits/s e devido à simplicidade de implementação é o sistema de codificação mais utilizado em redes locais, onde a taxa de transmissão pode chegar tipicamente a 10 Mb/s.

Porém, para muitas redes públicas de longa extensão a taxa fornecida pelos sistemas PCM é demasiado alta, nestas circunstâncias outras técnicas de codificação são utilizadas, pois reduzem a taxa de informação às custas de um maior processamento. Alguns sistemas de codificação atualmente em uso são: "Continuosly Varying Slope Delta Modulation" - CVSDM e "Linear Prediction Code" - LPC (ref. 5). O leitor interessado em técnicas de codificação pode consultar as referências 6, 7, 8.

c. Meio de Transmissão: é o responsável pela transmissão do sinal digital, que sai do codificador até o seu destino final. No caso que estamos estudando, o meio de transmissão é o sistema de comunicação da rede de computadores.

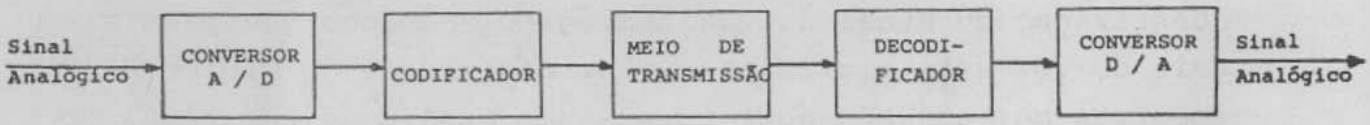

Figura 3 - Sistema de Transmissão de Voz

O serviço de transmissão de voz, por sua natureza de tempo real, e pela interatividade de comunicação que caracteriza a conversa humana, possui exigências próprias que devem ser cuidadosamente examinadas por ocasião do projeto desta classe de sistema:

1. O atraso absoluto entre o instante em que uma sílaba é pronunciada e o instante em que ela é ouvida não deve exceder um máximo atraso permissível (T), que é da ordem de umas poucas centenas de milisegundos. Esta característica, que é comum a todos os sistemas de tempo real, interessa particularmente aos sistemas de voz, pois o atraso máximo permissível já se compara ao atraso de transmissão obtido em redes de grande porte.

O atraso absoluto nos sistemas de transmissão de voz pode ser decomposto em três parcelas:

1a. Atraso de Digitalização e Codificação: corresponde ao tempo necessário para processar o sinal de voz e transformá-lo num sinal digital conveniente para transmissão. Em geral, uma codificação maior traz o benefício de uma menor taxa de transmissão, mas implica num tempo de codificação maior. O valor deste atraso depende, pois, da técnica de codificação escolhida.

2a. Atraso de empacotamento: este atraso só existe nas redes que utilizam tecnologia de comutação de pacotes para transmissão de voz. Uma vez que os codificadores fornecem bits a uma taxa constante e que o tamanho ótimo do pacote para rede de voz é relativamente grande (ref. 9), existe um certo tempo dispendido para montar estes pacotes. Este atraso é igual ao número de amostras coletadas multiplicado pelo período de amostragem.

3a. Atraso de Transmissão: corresponde ao tempo gasto para transportar o sinal de voz de um extremo a outro da rede.

Este atraso é provocado por diferentes causas: atraso de acesso ao meio de comunicação, atraso de filas nos nós da rede, atraso devido aos algorítmos de roteamento e correção de erros, etc. Contrariamente aos anteriores, este atraso não é, necessariamente constante, mas depende da carga da rede, das rotas escolhidas, etc., e em muitos tipos de rede, não é limitado (Ethernet). Ele deve, pois, ser tratado de forma estatística.

A variância do atraso de comunicação é também o fator importante para o serviço de voz. A variabilidade do atraso sofrido pelas diferentes mensagens de voz não podem ser transferidas para o sinal analógico recuperado sob pena de romper-se a continuidade e naturalidade da voz humana.

Existem duas técnicas básicas para superar este problema variância do atraso: a primeira consiste em utilizar-se um otocolo onde o atraso de comunicação é constante, fazendo-se sta forma com que a variância do atraso seja nula. A segunda cnica consiste em corrigir a variância do atraso introduzin- um novo atraso variável (que é chamado de atraso de entre -) de forma a compensar diferentes atrasos sofridos na rede.

A terceira característica importante de um sistema para transmissão de voz em redes integradas diz respeito à política adotada em relação a pacotes inutilizados durante a transmissão, quer por erros de paridade, quer por excessivos atrasos. Diversos estudos mostram que uma certa quantidade de pacotes perdidos não afeta a qualidade da voz transmitida (ref. 4) e, portanto, muitas vezes deve-se preferir descartar simplesmente um pacote do que esperar por uma retransmissão. Como se vê, ao contrário dos sistemas tradicionais de dados, é a continuidade, e não a confiabilidade, a característica mais importante de um sistema de transmissão de voz.

"Um protocolo para comunicação de voz interativa deve, pois

provêr mecanismos de equilíbrio entre o atraso e a confiabilidade de tal forma que o atraso nunca exceda um dado limite (T), mesmo se isto implica em alguma perda de informação" (ref.11).

5. MODELO PARA AS ESTAÇÕES DE VOZ

Um modelo de rede de comunicação com serviço integrado deve incluir, necessariamente, um modelo específico para as estações de voz, pois o tráfego de voz possui características estatísticas que são essencialmente diferentes das características do tráfego de dados.

As funções realizadas por uma estação de voz pode ser divididas em duas partes: as funções de geração de voz, que são responsáveis pela codificação e pelo envio do tráfego vocal ao subsistema de comunicação, e as funções de recepção, que decodificam as mensagens recebidas e os enviam ao destinatário final. A figura 4 mostra a posição de uma estação de voz numa rede local e a figura 5 apresenta a divisão de funções descrita com os componentes básicos de cada parte.

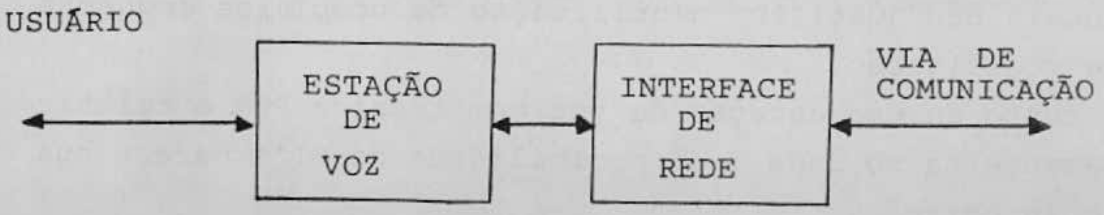

Figura 4 - Posição da estação de voz na rede local

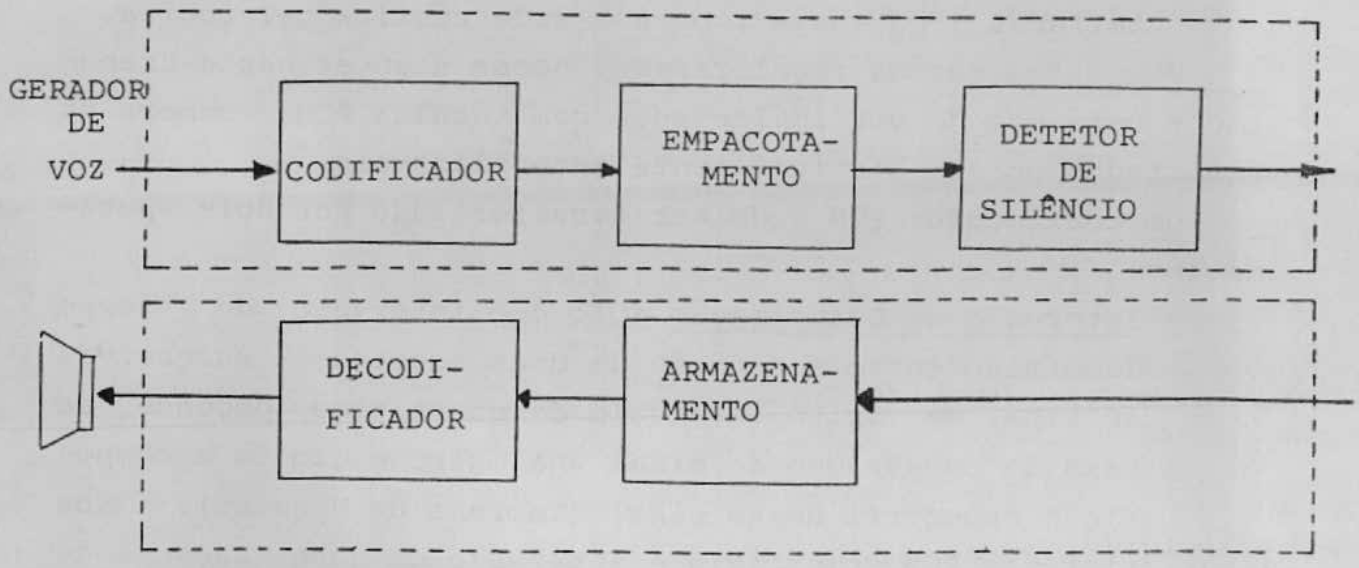

Figura 5 - Componentes básicos de uma estação de voz

Nesta parte do artigo formularemos um modelo para a estação de voz, porque o seu funcionamento influencia o comportamento dinâmico da rede como um todo. Para tanto, descreveremos em detalhes os diversos componentes das estações de voz.

a) Codificador

O codificador é o elemento responsável pela transformação do sinal analógico de voz numa sequência de bits equivalente (conversão análogo digital) e pela compressão desta sequência de bits de modo a resultar numa taxa de bits compatível com o recurso de transmissão disponível.

Diversas técnicas tem sido propostas e utilizadas para a implementação de codificadores em redes de computadores. A seção anterior sugeriu que a técnica preferível para codificadores em redes locais é PCM ("Pulse Code Modulation"), pois:

- A alta banda passante das vias de comunicação em redes locais não justifica a utilização de complexos esquemas de codificação.
- O custo de uma estação de voz com técnica PCM é relativamente baixo dada a disponibilidade de CI's para a sua implementação.
- A técnica PCM permite a construção mais simples de interfaces ("gateways") para a rede pública telefônica.

Por estas razões focalizaremos nossa atenção neste trabalho a estações de voz implantadas com técnica PCM, embora os resultados possam ser facilmente generalizáveis.

Um codificador PCM pode ser caracterizado por dois parâmetros:

- Intervalo de Quantização (Q): é o intervalo de tempo decorrido entre a geração de duas amostras sucessivas do sinal de voz. O intervalo de quantização depende da taxa de amostragem do sinal analógico e limita a composição espectral deste sinal (Teorema de Nyquist). Nos sistemas PCM comerciais o intervalo de quantização é de 125us, o que corresponde a uma frequência de amostragem de 8 kHz.

- Índice de Quantização (I) : é o número de bits gerados pelo codificador PCM a cada intervalo de quantização. O índice de quantização define a precisão com que a amostra do sinal analógico é quantificada e limita a relação sinal/ruído do sistema. Podemos dizer que quanto maior o índice de quantização, tanto melhor será a representação que a sequência digital fornece para o sinal analógico. Nos sistemas PCM comerciais o índice de quantização é de 8 bits.

A partir destes dois parâmetros podemos determinar a taxa de bits (T) fornecida pela estação de voz. De fato, a cada Q segundos, I bits são gerados, logo $T = I/Q$ bits/s.

Nos sistemas PCM comerciais $T = 8/125 \times 10^{-6} = 64000$ bits/s

b) Detetor de Silêncio

Os sistemas PCM não possuem implicitamente técnicas de deteção de silêncio. Se o usuário deixa de falar num determinado instante, o codificador PCM continua a gerar bits na mesma taxa. Se não adotássemos nenhum esquema de deteção de silêncio, a estação de voz forneceria pacotes para a sub-rede numa taxa fixa, não importando se o usuário está falando ou não. Este fato anularia a vantagem resultante da multiplexação estatística decorrente da comutação de pacotes. Sem deteção de silêncio a tecnologia de comutação de pacotes é sempre inferior à tecnologia de comutação de circuitos.

Portanto, algum esquema de deteção de silêncio deve ser provido pela estação de voz para permitir o uso eficiente da via de comunicação. Nos sistemas PCM a introdução destes mecanismos é simples, pois o silêncio é normalmente codificado como uma sequência de zeros.

Devemos observar, porém, que o algorítmo de deteção de silêncio não deve retirar amostras iguais a zero arbitrariamente, pois isto alteraria a fluência da voz, que contém pausas naturais entre as palavras e mesmo entre as sílabas de uma mesma palavra. A regra básica para o projeto de um

detetor de silêncio efeciente é que toda amostra retirada pelo transmissor deve ser resposta pelo receptor.

Esta observação leva à conclusão que o detetor de silêncio deve operar sobre uma fração da voz maior do que uma amostra pois a análise de uma amostra individual não permite decidir se ela deve ou não ser retirada da sequência de bits transmitidos.

A decisão do detetor de silêncio deve pois ser feita a nível de pacotes. É analisando o conteúdo de um ou mais pacotes que o detetor pode decidir se eles devem ou não ser transmitidos.

O algorítmo do detetor de silêncio (para o caso PCM) pode então ser descrito da seguinte forma:

- O detetor espera pela montagem de N pacotes.
- Se nos N pacotes mais do que uma fração f das amostras forem nulas, os N pacotes são descartados.
- Caso contrário o pacote mais antigo é enviado.

Devemos observar neste algorítmo que quanto maior N, tanto mais eficiente será o algorítmo de deteção de silêncio, porém tanto maior será o atraso introduzido pelo detetor. A primeira vista a solução ótima seria tomar N = 1 (condição de mínimo atraso), porém, isto só é verdade se pudermos assegurar que a estação receptora não tomará a ausência de um pacote como uma flutuação estatística do atraso da rede.

Esta última observação sugere que deve haver uma estreita harmonia entre o algorítmo de deteção de silêncio e o algorítmo de compensação da variância do atraso que a estação receptora implementa. Caso contrário, pacotes retirados pelo primeiro podem ser automaticamente compensados pelo segundo, deteriorando a qualidade do serviço de voz. Esta característica será mais detalhada quando tratarmos do algorítmo de armazenamento da estação receptora. Por hora, queremos ressaltar que a adoção de um determinado valor de N está vinculada às características estatísticas do atraso de transmissão.

A escolha de um valor para a fração de amostras nulas que caracteriza um pacote de silêncio (f) está vinculada às características estatísticas da voz humana. A adoção de valores muito pequenos fará com que qualquer ruído seja interpre

tado como voz. A escolha de um valor elevado fará com que trechos rápidos de voz sejam indevidamente cancelados.

A estação transmissora de voz deve ser implementada de forma a permitir a fácil alteração do valor de f, pois o valor ótimo depende do usuário e das condições ambientais de ruído.

O modelo da estação geradora deverá permitir a avaliação da influência do parâmetro f sobre o desempenho do sistema.

c) Empacotamento

A tarefa de empacotamento é talvez, a mais importante realizada pela estação geradora de voz. Como veremos, a qualidade e a estabilidade do sistema de voz é fortemente afetada pela forma como esta tarefa é implementada.

Esta tarefa consiste em reunir amostras codificadas da voz em pacotes de tamanho fixo. Cada pacote conterá uma quantidade fixa de amostras, o que significa que os intervalos entre pacotes serão fixos. Porém esta uniformidade será quebrada pelo detetor de silêncio que transmitirá apenas os pacotes considerados úteis.

A característica fundamental desta tarefa é o tamanho do pacote. Pacotes muito grandes aumentam o atraso total porque o tempo de montagem dos pacotes é diretamente proporcional ao número de amostras nele contidas. Por outro lado, pacotes pequenos tendem a saturar a via de comunicação pelo aumento do número de pacotes oferecidos ao sistema por unidade de tempo. Além disto alguns protocolos de comunicação (por ex. CSMA/CD) tornam-se ineficientes quando os pacotes são muito pequenos.

Desta forma, expressando o atraso total no sistema pela soma do tempo de empacotamento e do tempo de transmissão, devemos esperar que a curva do atraso total seja função convexa do tamanho do pacote, e que, portanto, apresente um ponto de mínimo bem definido. Esta observação intuitiva é confirmada por investigações numéricas do problema (ref. 9).

O tempo de empacotamento depende também da tecnologia de codificação utilizada. Quanto maior for a taxa de quantização do codificador, tanto menor será o atraso de empacotamento. Esta é outra razão pela qual tecnologias de codificação com baixa taxa de quantização nem sempre são convenientes em redes locais. A figura 6 apresenta a relação entre o tamanho do pacote e o atraso de empacotamento para a tecnologia PCM com 64000 bits/s e a tecnologia LPC com 9600 bits/s.

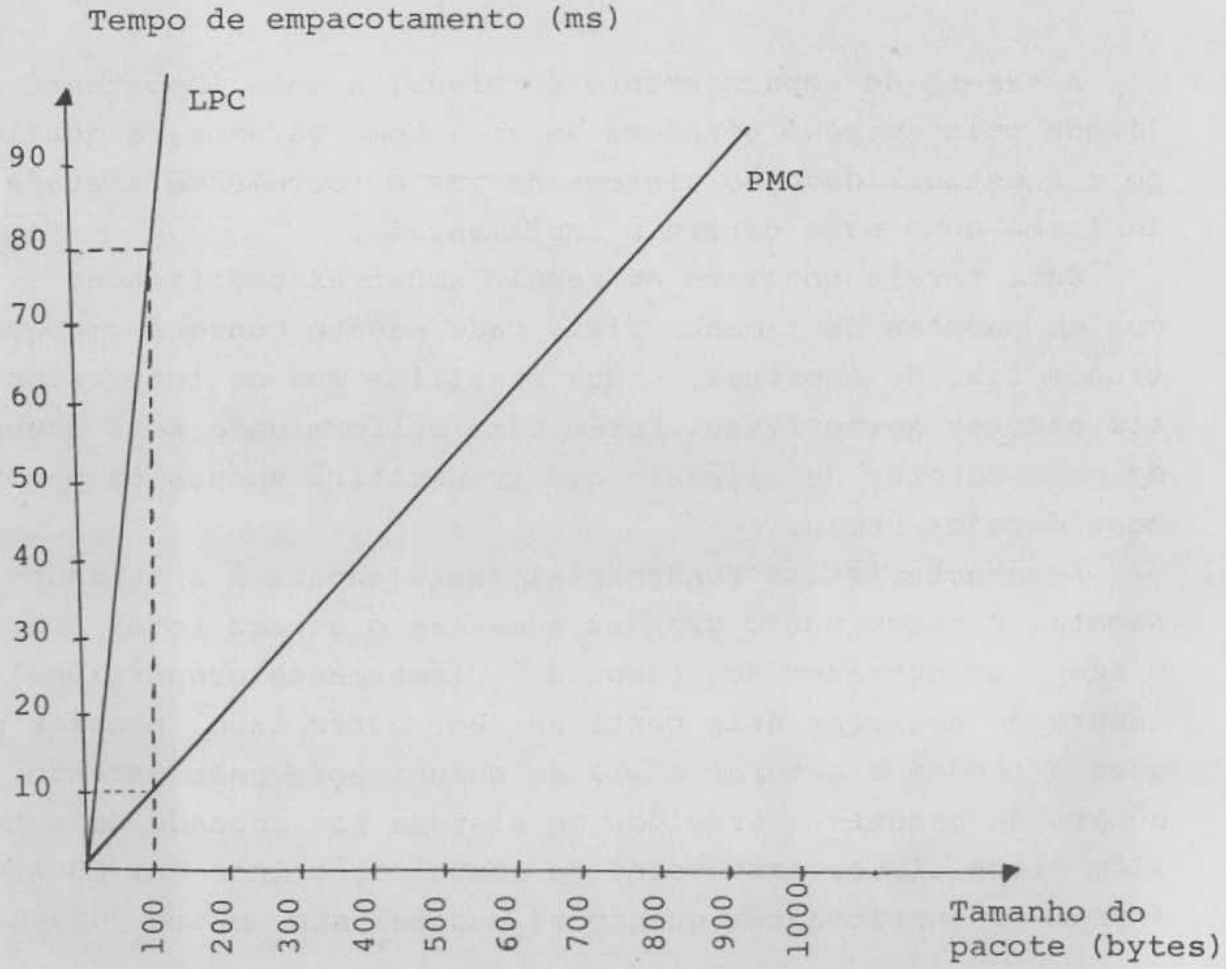

Figura 6 - Tempo de empacotamento x Tamanho do pacote

A figura 7 apresenta um possível comportamento para o atraso total em função do tamanho do pacote de voz (ref. 9).

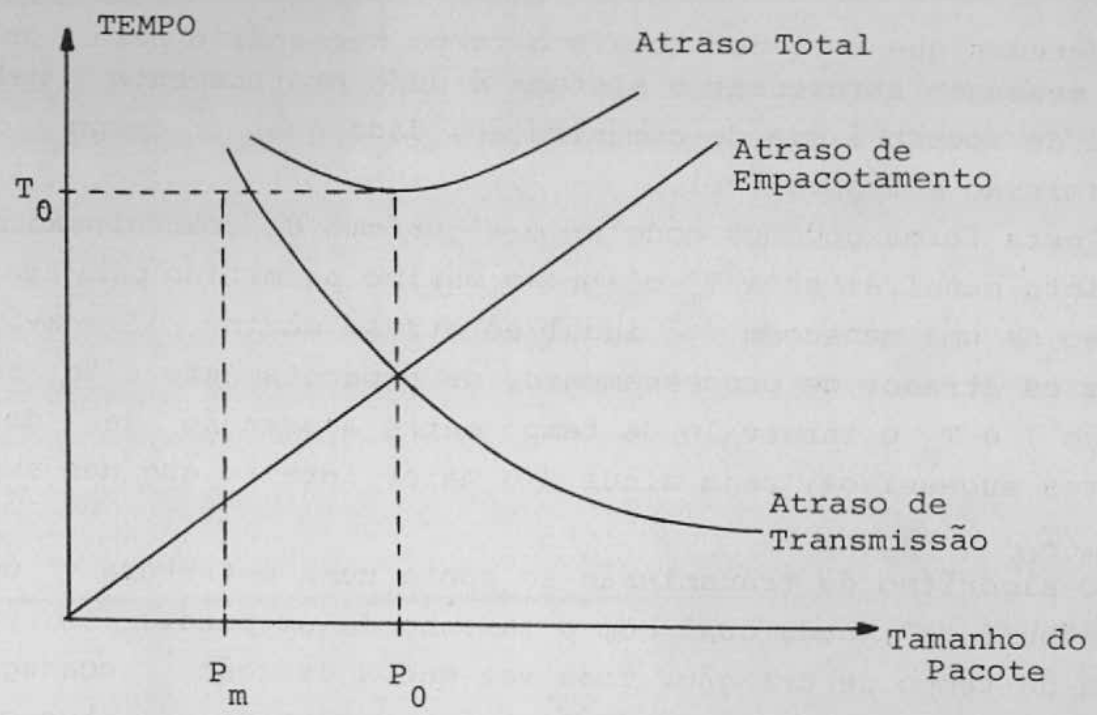

Figura 7 - Composição do atraso total em sistemas de voz

Como vimos na seção 3 deste trabalho, uma das exigências de sistemas de transmissão de voz em tempo real é que o atraso total fim-a-fim (T_t) não ultrapasse um valor em torno de 200ms. Nestas condições, usando a simbologia apresentada na figura 7, podemos dizer que uma condição necessária para o funcionamento do sistema é $T_0 < T_t$. Verificamos também que a estação de voz deve ser programada para operar com pacotes de tamanho P_0.

Estas considerações mostram que um modelo para as estações de voz deve possibilitar o estudo do comportamento do valor de P_0 e de T_0 para diversas condições do subsistema de comunicação.

Outro problema que deve ser levado em conta no algorítmo de empacotamento é o tratamento a ser dado a pacotes que não conseguem ser transmitidos após um certo intervalo de tempo. Como vimos acima todo sistema de tempo real se caracteriza por um tempo máximo de transmissão. Se uma mensagem qualquer demora um tempo maior do que este máximo para atravessar o sistema, o seu conteúdo perde o significado e a mensagem deveria ser descartada.

Sabemos que em redes locais o tempo necessário para uma dada mensagem atravessar o sistema é dado praticamente pelo tempo de acesso à via de comunicação, dado que o tempo de transmissão é desprezível.

Desta forma podemos modelar o algorítmo de transmissão da seguinte maneira: seja T_m o atraso máximo permitido para transmissão de uma mensagem (é igual ao atraso máximo fim-a-fim menos os atrasos de processamento, de empacotamento e de recepção) e T_p o intervalo de tempo entre a geração de dois pacotes sucessivos; seja ainda M o maior inteiro que não supera T_m/T_p.

O algorítmo de transmissão se apoia numa estrutura com $\nu \leq M$ "buffers", cada qual com o tamanho de um pacote e uma marca de tempo de criação. Toda vez que a estação consegue acesso à via, o pacote mais antigo é transmitido e o respectivo "buffer" se torna livre. Toda vez que um pacote é gerado ele é colocado num "buffer" livre, se houver, ou caso contrário no "buffer" mais antigo, cuja marca de tempo é atualizada.

Desta forma uma determinada mensagem será transmitida em νT_p segundos, ou não será mais transmitida. Como $\nu \leq M$, segue que $\nu T_p \leq MT_p \leq T_m$, o que garante a validade da mensagem.

O modelo deve permitir avaliar o comportamento do sistema em função do valor de ν . Para ν grande, o algorítmo de transmissão exigirá mais memória e poderá provocar uma maior variância no atraso das mensagens. Por outro lado valores pequenos de ν redundam numa maior probabilidade de perda de mensagens.

A probabilidade de perda pode ser calculada em função da vazão do subsistema de comunicação. Seja uma rede local com n estações e T_p o intervalo de geração de pacotes num sistema com ν "buffers". Seja também c o número de pacotes que podem ser transmitidos em νT_p segundos. O número máximo de pacotes que podem ser gerados neste intervalo de tempo é $n\nu$. Seja k o número de pacotes efetivamente gerados e vamos calcular o valor médio de k, admitindo que a cada T_p segundos cada estação de voz gera um pacote com probabilidade p.

$$P\ (k=0) = (1 - p)^{\nu n}$$

$$P\ (k=1) = p.(1 - p)^{\nu n-1}$$

$$\vdots$$

$$P\ (k=j) = \binom{\nu n}{j} p^{j}(1 - p)^{\nu n-j} = b\ (j, \nu n, p)$$

O número médio de mensagens geradas é:

$$k = \sum_{k=0}^{\nu n} k.b(k, \nu n, p) = \frac{1}{\nu np}$$ (média da distribuição binomial)

Se $k > c$ então $k-c$ mensagens são descartadas e desta forma o número médio de mensagens descartadas será:

$$\sum_{k=c+1}^{\nu n} (k - c).b(k, \nu n, p)$$

Assim, a fração de mensagens perdidas será:

$$\emptyset = \frac{1}{\nu np} \sum_{k=c+1}^{\nu n} (k - c).b(k, \nu n, p)$$

Em artigo recente, C. Weinstein (ref. 11) afirma que o caso $\nu = 1$ corresponde a um limite superior para a fração de perda e que a utilização de sistemas com $\nu > 1$ provavelmente traria pouca vantagem para o desempenho do sistema, pois a diferença entre k e c pode não ser compensada pelo armazenamento extra.

O modelo para a estação de voz que descrevemos permitirá a verificação da validade destas afirmações.

d) Armazenamento

A função de armazenamento é realizada pela parte receptora da estação de voz e objetiva compensar a variância do atraso de transmissão.

Esta função é realizada introduzindo-se propositalmente um atraso nos pacotes recebidos de forma que o envio de dados

possa ser realizado de forma contínua, mantendo a fluência natural da voz humana.

A figura 8a mostra a recepção de pacotes sem compensação da variância, sendo que as setas indicam os instantes de chegada dos pacotes. Como se vê, nos instantes 1 e 3 haverá uma descontinuidade na recepção de voz. A figura 8b mostra como o mesmo padrão de chegada de voz pode ser corretamente recebido, introduzindo-se um atraso igual a meio intervalo de empacotamento.

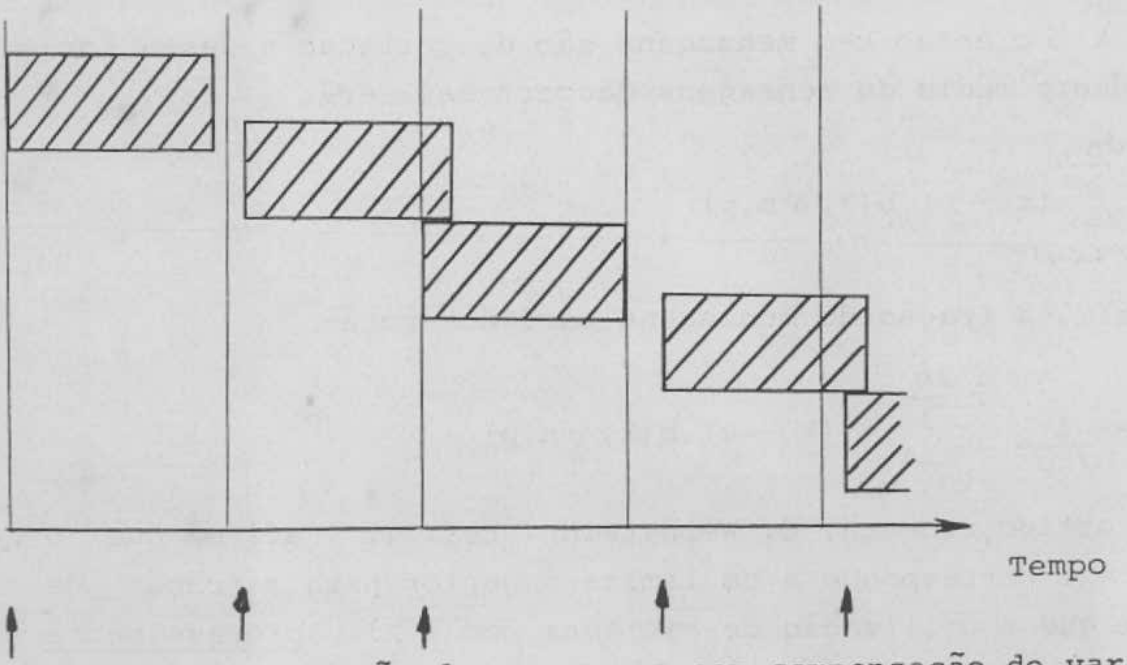

Fig. 8a - Recepção de mensagens sem compensação de variância

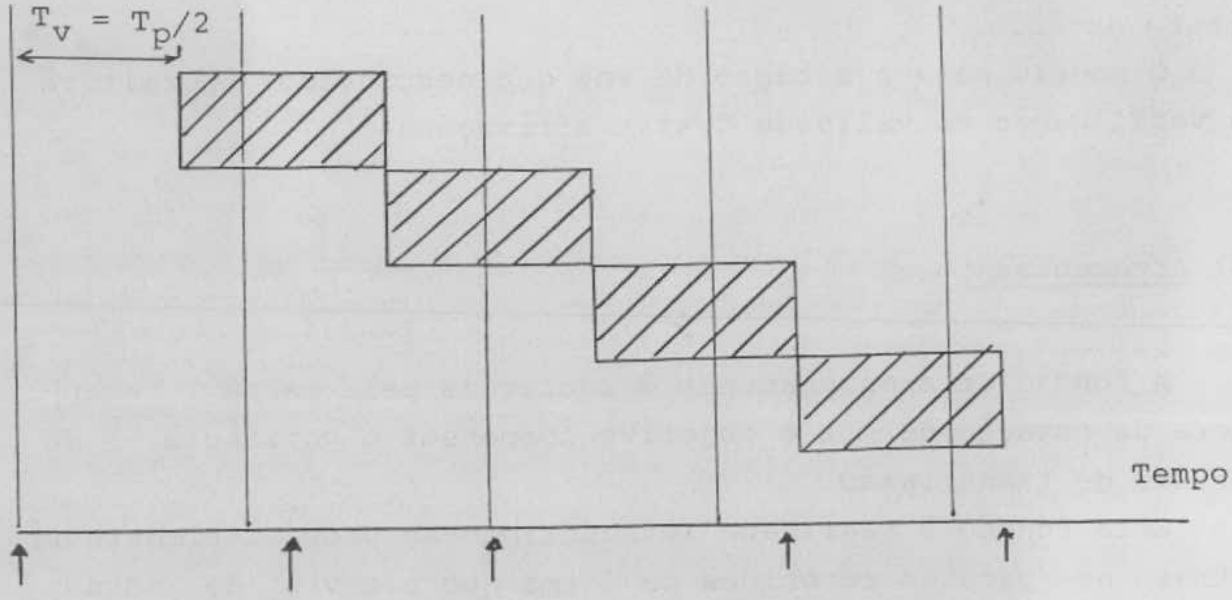

Fig. 8b - Recepção de mensagens com compensação de variância

O algorítmo de armazenamento de mensagens é caracterizado, portanto, pelo atraso introduzido nas mensagens recebidas (T_v). A escolha do valor adequado deste atraso depende:

1. Da distribuição do atraso de transmissão (T_a). Devemos escolher T_v de forma que $P(T_a > \bar{T}_a + T_v)$ seja desprezível.
2. Do algorítmo de deteção de silêncio utilizado. O valor de T_v não deve ser grande o bastante para poder compensar a menor unidade de dados bloqueada por aquele algorítmo. Assim, por exemplo, se o algorítmo de deteção de silêncio opera pacote a pacote, devemos ter $T_v < T_p$, caso contrário, pacotes separados por silêncio poderão ser indevidamente unidos. A simulação do modelo deverá ajudar a determinar o valor de T_v em cada caso.

e) Decodificação

O decodificador realiza a transformação digital-analógica do sinal recebido e o envia do mesmo para o usuário receptor. Os cuidados que devem ser tomados na implementação do decodificador é garantir pequenos atrasos de processamento e coordenação com o codificador.

f) Geradores de Tráfego

O modelo para os geradores de tráfego é apresentado baseado num trabalho de Weinstein (ref.11), o qual faz uso de uma cadeia de Markov simples para descrever o padrão de fala/silêncio de cada gerador de voz (por geradores de voz entendemos os usuários humanos do sistema de transmissão de voz). Ao contrário de outros modelos mais complexos, tais como os apresentados por Brady (ref.12), este modelo se caracteriza por ser facilmente simulável, e ainda apresentar resultados suficientemente acurados (ref. 13).

6. CONCLUSÕES E TRABALHOS FUTUROS

A análise do comportamento de redes de computadores com comutação de pacotes com serviços integrados de voz e de dados pode ser feita com o auxílio do modelo para as estações de voz que descrevemos. As estações de dados podem ser modeladas na forma tradicional como geradores de pacotes de tamanho aleatório a intervalos aleatórios. A subrede de comunicação também deve receber um modelo próprio de acordo com o tipo de arquitetura que se pretenda representar. É interessante que o modelo da sub-rede seja o mais possível independente dos modelos dos geradores de tráfego, de forma que diversas arquiteturas possam ser analisadas e comparadas em condições de carga semelhantes.

No prosseguimento deste trabalho devem ser realizados esforços para se obter uma estimativa do desempenho de redes integradas com base nos parâmetros identificados neste artigo. Presentemente busca-se obtenção tanto de um modelo analítico completo (embora com muitas hipóteses simplificadoras) como de um modelo de simulação bastante realista. Em ambos os casos espera-se encontrar resultados que possibilitem a definição dos métodos e ferramentas para o dimensionamento e a avaliação do desempenho das redes locais integradas.

Nesta linha de trabalhos estamos começando a desenvolver um programa em linguagem SIMULA que implementa o modelo descrito neste artigo. Os parâmetros que este modelo deve ajudar a determinar são:

- O número de pacotes (n) que deve ser examinado pelo detetor de silêncio.
- O tamanho ótimo de pacote (P_o)
- O número de "buffers" utilizados pelo algorítmo de transmissão (ν)
- O atraso introduzido para compensação da variância (T_v).

Além disso deverão ser obtidos curvas de desempenho das arquiteturas mais usuais em redes locais com tráfego misto de voz e dados.

7. AGRADECIMENTO

Agradecemos a FDTE e a EPUSP pelo apoio e incentivo e pelo ambiente propício à pesquisa que temos recebido.

8. BIBLIOGRAFIA

(1) COVIELLO, G. - "Comparative Discussion of Circuit vs. Packet-switched Voice", IEEE Trans - Commun., vol.COM-27, Agosto/1979.

(2) RASNER, R. e SPRINGER, B. - "Circuit and Packet Switching. A Cost and Performance Tradeoff Studdy", Computer Networks, 1976.

(3) IEEE Project 802, "Local Area Networks Standards", Draft D, Novembro/1982.

(4) ROVASIO, P.C. et al. - "Voice Transmission Over an Ethernet Backbone", Proc. IFIP, 1982.

(5) COHEN, D. - "Using Local Area Networks for Carrying Online Voice", Proc. IFIP, 1982.

(6) FLANAGAN, J.L. - "Speech Analysis, Synthesis and Prediction", Springer-Verlag, New York, 1972.

(7) FLANAGAN, J.L. et al. - "Speech Coding", IEEE Trans. on Commun., vol. COM-27, Abril/1979.

(8) JAYANT, N.S. - "Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers", Proc. IEEE, 1974.

(9) MINOLLI, D. - "Optimum Packet Lenght for Packet Voice Communication", IEEE Trans. on Commun., vol. COM-27, Março/1979.

(10) COHEN, D. - "A Protocol for Packet Switching Voice Communications", Computer Network Protocols, Université de Liége, 1978.

(11) WEINSTEIN, C.J. - "Fractional Speech Loss and Talker Activity Model for TASI and Packet-Switched Speech", IEEE Trans. on Commun., vol. COM-26, Agosto/1978.

(12) BRADY, P.T. - "Statistical Analysis of on-off Patterns in 16 Conversations", BSTJ, vol. 46, 1967.

(13) BIALLY, T. et al. - "A Technique for Adaptive Voice Flow Control in Integrated Packet Networks", IEEE Trans. on Commun., vol. COM-28, Março/1980.