

インターネット電話の品質評価について

高野 正次

1. はじめに

電話サービスを LAN やインターネットといったデータ系のネットワークの上へ統合し再構築することが注目されている [1, 3]。これを実現するための技術は、VoIP (Voice over IP) 技術と総称されるが、音声系のネットワークをデータ系のネットワークの上に構築し直してコストを下げようとする場合や、既存のデータ系ネットワークの新しいサービスとして、安価に電話サービスを利用しようという場合などを契機として、現在、企業のネットワークへの導入が進んでいる。

本稿では、CATV 網 (ケーブルテレビネットワーク) で加入者に対して新しく電話サービスを提供する場合の End-End 音声品質について検討する。具体的には、通信トラヒック理論に基づいて定量化・評価し比較を行っている。

以下では、通常の加入者電話を指す場合に一般電話と呼び、VoIP 技術を利用してデータ系のネットワークを利用して提供される電話サービスを、IP 電話と呼んで区別する。

2. VoIP

2.1 導入メリット

VoIP 技術が注目されている理由は、以下のよう
明確である。

(1) 通話帯域圧縮による優れた経済性

新しい音声符号化方式を利用すれば、通話当たり
使用する通信帯域を従来よりも大幅に圧縮する
ことができる。その圧縮の程度は、音声符号化
方式毎に異なるが、通常の 64 kbps と比べて
4分の1から8分の1以下にすることができる
(表 1)。

また、他のアプリケーションで利用する通信
帯域との統計多重効果を期待できるようになる
ので、通信回

このしょうじ
NTT サービスインテグレーション基盤研究所
〒180-8585 武蔵野市緑町 3-9-11

表 1 主な音声符号化方式

ITU-T 勧告 (方式)	符号化速度 (bps)
G.711(PCM)	48k,56k,64k
G.726(ADPCM)	16k,24k,32k,40k
G.729/G.729A(CS-ACELP)	8k
G.723.1(ACELP/MP-MLQ)	5.3k/6.3k

線全体の有効利用をさらに高めることになり、
コスト削減メリットが大きい。

(2) アプリケーションとの親和性

音声通話機能を必要とするシステムでは、
VoIP 技術を適用する方がより効率的かつ安
価に高機能のシステムを構築することができ
る場合がある。

具体例を挙げると、通信販売や顧客サポ
ートで使われる顧客対応システムやコール
センターのシステムなどは、現在パソコン
上の Web ブラウザを表示用のアプリケー
ションとして作られることが多いが、音
声機能の部分は VoIP 技術を利用した
方が開発効率が良い。また、多くの拠
点間を結ぶシステムでは、一斉通報機
能が重宝されるが、こちらも IP プロ
トコルのマルチキャストを応用すること
により実現できる。

2.2 課題は音声品質の確保

IP 電話のユーザは音声品質をとて
も敏感に感じている。音声品質を現
状程度に確保した上で、コスト削減
効果を期待している。しかし、これ
が必ずしも容易ではない。その理由
は以下である。

(1) 高速回線と低速回線の混在

LAN が 10 Mbps あるいは 100 Mbps
といった高速回線であるのに対して、
WAN への接続・中継回線の多くが
64 kbps, 128 kbps 程度から数
Mbps の低速回線であることが多い。
そのため、低速回線の部分がネッ
トワークのボトルネックになりやす
い。

(2) データ系の長いパケットの存在

データ系のアプリケーションによ
って使用されるサイズの大きなパ
ケットは、転送されるときの出
送処理時間がかかなり大きくなる。
低速の回線で特に顕著とな

るが、この影響により、同じ通信回線で転送される音声パケットのバッファリング遅延時間とその揺らぎが非常に大きくなる。

(3) IP レイヤでの QoS 技術が未確立

音声パケットとデータ系のパケットが同じ通信回線で転送されることにより、データ系トラヒックの特徴であるバースト性に起因して輻輳が発生しやすい。しかし、このバースト性を制御したり、パケットの損失確率や遅延時間を保証するような QoS 技術が IP レイヤで確立されていない。

本稿では、音声品質の具体的な目標値として、片道の遅延時間 (End-End 遅延時間) を 200 ミリ秒未満と規定した[注 1]。この数値は、電話サービスの End-End 遅延時間の目安として一般的に意識されている。

2.3 遅延と揺らぎの発生

音声パケットが転送される間に、大きな遅延時間やその揺らぎが以下のように発生する。例として、図 1 に 10 ミリ秒間隔に音声情報を符号化し、順々にひとつの音声パケットとして送出させる場合を示す。このときの音声データの大きさは、音声符号化方式に依存するものの、10~80 バイト程度になり、40 バイトの IP ヘッダを加えた音声パケット全体の大きさは 120 バイト以下となる。このように、10 ミリ秒の固定的間隔でひとつの音声情報としてパケット化され、送信側から送出された音声パケットの列は、受信側の端末でも同じ 10 ミリ秒間隔に音声パケットを再生する必要がある (図 1 の VoIP)。

FTP などのデータ系のアプリケーションが使用するパケットは、IP パケットの最大サイズとして規定されている 1500 バイトであることが少なくないが、

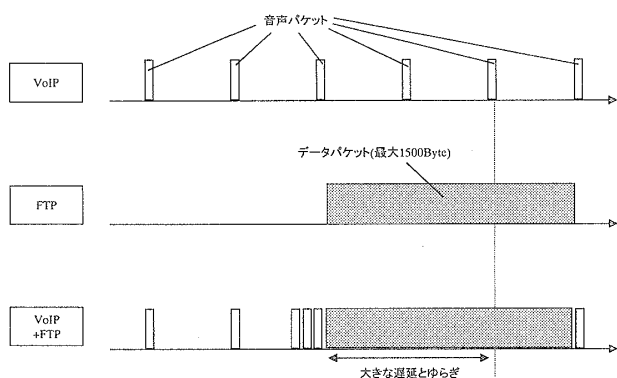


図 1 大きな遅延と揺らぎの発生

[注 1] 音声の途切れとなる音声パケットの損失率などは、本稿が検討するネットワークにおいては問題が顕在化しない。理由は後述する。

通信回線の速度が低速のときは送出処理にかかる時間が長くなる (図 1 の FTP)。例えば、通信回線の速度が 64 kbps の場合では、1500 バイトの送出処理時間は 187.5 ミリ秒である。

10 ミリ秒の一定間隔に整列して連続的に送出されている音声パケットの列に、FTP による大きなパケットが割り込んでしまうと (図 1 の VoIP+FTP)、通信回線が十分に高速でない場合には、元々の音声パケットの列の隙間である 10 ミリ秒のうちに、割り込んだパケットの送出処理が終わらないで、後続の音声パケットの列を大きく乱してしまう。その結果として、後続の音声パケットのバッファリングによる遅延時間とその揺らぎが大きくなる。音声品質の観点から言えば、音声パケットが本来再生されるべきタイミングで着信側端末に到着できないので、その音声品質は悪くなっている。

また、バースト的なデータ系パケットの連続到着によって、トラヒックの輻輳が引き起こされるときにも、同じように音声パケットの列が乱される状況が起こるので音声品質が劣化してしまう。

3. CATV 網での IP 電話

3.1 ネットワーク構成

CATV 網のネットワーク構成を図 2 に示す。CATV 加入者は、加入者端末として Cable Modem (CM) と呼ばれる装置にパソコンや IP 電話端末を接続して IP 電話を利用する。CM はバス回線に收容されていて、数百人~1千人程度の加入者でそのバス回線を共有している。バス毎には、バスルータ (R1)

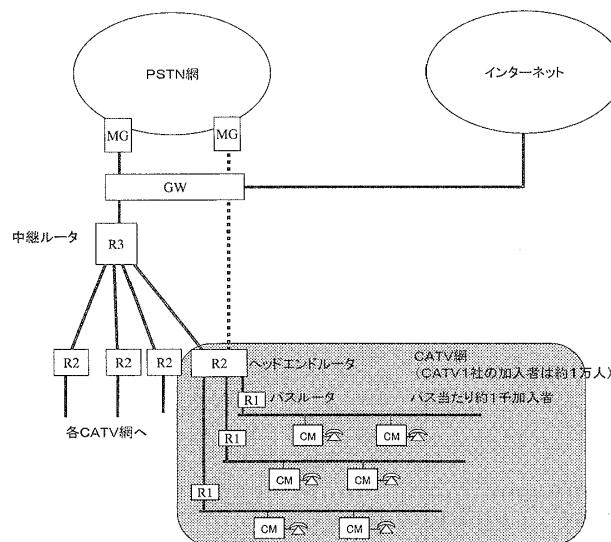


図 2 CATV 網の構成

という装置があり、上位のネットワークとの間の通信経路を制御している。各バスは、CATV会社毎にまとめられて、外部のネットワークとつながっているが、このためのルーティング装置をヘッドエンドルータ(R2)と呼ぶ。CATV1社当たりの加入者数が実勢として平均約1万人であるので、ヘッドエンドルータはおよそ10本程度のバスをまとめている勘定になる。

IP電話による通信を行うためには、ヘッドエンドルータとPSTN網(公衆電話網)をつなぐ(図2の点線)必要があり、この間でのプロトコル変換・メディア変換をMedia Gateway(MG)と呼ばれる装置が行っている。

また、ヘッドエンドルータとMG間の回線の使用効率を上げてコストを圧縮する目的で、複数のCATV会社のヘッドエンドルータからMGへのトラヒックをまとめる装置を中継ルータ(R3)と呼び、この中継ルータとMGを結んで(図2の実線)通話を行う場合もある。

今回の検討では、中継ルータが複数のCATV会社(加入者合計数万人規模)のトラヒックをまとめるネットワークを想定したEnd-End音声品質に対して評価を行う。また、バスルータやヘッドエンドルータの市販製品には、回線の速度やそのインタフェースの数が決まっているので、それに依存して各ルータ配下の加入者数を調整・制限することを行っている。

3.2 End-End 遅延

発信側のCATV網内端末から着信側の一般電話へ通信を行う場合の、End-End遅延要因を図3を用いて説明する。CATV網内のIP電話端末から順番にCM、バスルータ、ヘッドエンドルータ、中継ルータ、GW(後述)、さらに、MGとPSTN網を通して、着

信先の一般電話へつながる。MGは、IPネットワークとPSTN網の間で必要な変換を行っている。つまり、CATV網からPSTN網に向かっては、IP電話端末からのIPプロトコルを終端して、送られて来た音声データを復号化して、PSTN網側の一般電話との間に通常の電話回線を接続して通信を行っている。逆向きに対しては、PSTN網から来た音声データを符号化・パケット化して、そのパケットをCATV網側のIP電話端末に向けてIPプロトコルを使って送出している。

このときの発信側であるCATV網のIP電話端末と着信側であるPSTN網の一般電話端末との間で生じる遅延、すなわち、End-End遅延時間の要因は、以下である。

- (a) 符号化・復号化+パケット化遅延
- (b) 伝送遅延(距離だけに依存する伝搬遅延)
- (c) 遅延揺らぎ吸収バッファ内遅延(4.3節)

End-End遅延時間とは、上記の(a)+(b)+(c)であり、この目標値が200ミリ秒未満となる。

CATV網の加入者は、HTTP、FTP、TELNETなどのインターネットへの接続サービスも利用するが、これらのトラヒックは、Gate Way(GW)で分離されてインターネットへ流れる。このインターネットへのトラヒックは音声パケットの遅延時間を大きくする背景負荷として考慮する必要がある。つまり、本稿の想定では、バスルータを最大約1千人、ヘッドエンドルータを平均約1万人程度、中継ルータを数万人規模の加入者でそれぞれ共用することになるが、各ルータに到着した音声パケットは、他の加入者による音声パケットだけでなく、インターネットへ向かうデータパケットとも通信帯域を競合し、そのためにEnd-End遅

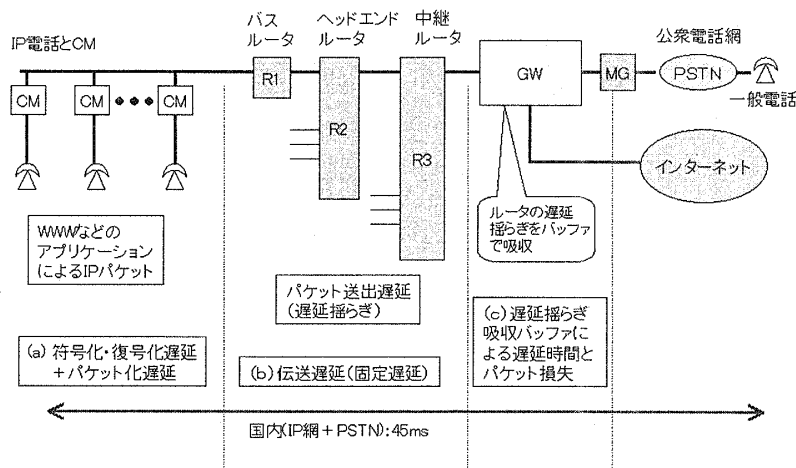


図3 End-End遅延要因

延時間が大きくなる。

4. 遅延を抑える仕組み

4.1 データ・フラグメンテーション機能

送信側で長いデータパケットをより短いパケットに分割することを、データ・フラグメンテーションという。パケットを送出するための処理時間が平均的に短くなるので、バッファリングによる遅延時間も短くなる。図1のように、データパケットが音声パケットの列に割り込んでも、データ・フラグメンテーションによりデータパケットの長さが短くなっていけば、音声パケットの列の送出間隔の乱れが相対的に小さくなり、結果的に音声品質が向上する効果がある。この機能は、通常送信側に備えられる機能であるが、実装が比較的簡単であるため多くの市販製品で利用できる。ただし、分割するパケット毎にIPヘッダが必要になるので、小さく分割しすぎると全体のトラフィックが増大してしまうことになるので、通信回線の速度に応じた適切なパケットサイズにフラグメンテーションすることが重要である。

4.2 音声優先制御機能

音声優先制御機能は、音声パケットと音声以外のデータパケットを識別して、音声以外のデータパケットはバッファ内に滞留させて、送出を待っている音声パケットがないときにだけ、データパケットを送出させる機能である。この機能により、データパケットのトラフィックに関係なく、音声パケットのスループットを確保することができる。音声パケットがバッファリングされる時間が短くなるので、IP電話の音声品質は良くなる。ただし、マルチベンダー環境では音声優先機能を利用することは現在のところ難しい。これは、標準化に時間がかかるためと考えられる。例えば、Cable Modemの仕様には、DOCSIS 1.1という音声優先制御機能を含む仕様があるが、正式に認定を受けた製品は2000年夏の時点で市場にまだ出回っていない模様である。

4.3 遅延揺らぎ吸収バッファ

音声パケットがエッジルータ、ヘッドエンドルータ、中継ルータを通過する間に、バッファリングによる送出処理待ちの時間が発生する。この待ち時間がパケット毎に違っているので、音声パケットの間隔は一定間隔であったものが次第にばらつきが大きくなる。これを再び元の一定間隔に戻す機能が、遅延揺らぎ吸収バッファの機能であり、本稿の網構成ではGW装置に

備えられている。

このバッファは、すべての音声パケットをいったん滞留させて、先頭のパケットの到着から一定時間経過した後で、順番に元の一定間隔で再び送出する。ただし、あらかじめ規定された値以上の大きい遅延時間を持った音声パケットについては廃棄を行う。結局、廃棄されるほど大きい遅延をもった音声パケットの数は非常に少ないので、ほとんどの音声パケットは元の一定の送出間隔に戻すことができる（ごく稀に抜けが生じる）。これは、End-End遅延時間とパケット損失率を犠牲にする代償として、パケット毎の遅延揺らぎを完全に解消する仕組みである。

音声パケットの損失により聞き取れない音や音声の途切れが多少あっても、前後の関係から無意識のうちに補っているもので、会話に支障がないことが普通である。逆に、パケットの損失を減らす目的で遅延揺らぎ吸収バッファでの滞留時間を長くすると、End-End遅延時間がさらに大きくなり、間延びがして会話が不自由になる。

会話に支障をきたし始める音声の途切れの確率は経験的に 10^{-3} 以上とされているので、通信ネットワーク上のパケット損失率と比べると著しく大きい。したがって、遅延揺らぎ吸収バッファ部でのパケット損失率をかなり大きくしても、End-Endの遅延時間をできるだけ小さくする方が、会話を行う上での音声品質を向上させることができる。

本検討では、End-Endのパケット損失率は 10^{-4} と規定し、そのパケット損失率になるように遅延揺らぎ吸収バッファでの滞留時間を求めている。

5. End-End遅延時間の評価

5.1 設計パラメータ

図3に示したIP電話のEnd-Endの音声品質に影響をあたえる設計パラメータの中で、実際に変動させて評価を行った主要な設計パラメータについて以下に列挙する。

- 音声符号化方式
G. 711 (64 kbps), G. 729 (8 kbps) など
- 音声データ長[注2]
20~80 バイト程度
- データ・フラグメンテーション機能
150 バイト, 500 バイト, 1000 バイトなど

[注2] ヘッダがRTPのヘッダを合わせて40バイトなので音声パケット長は60~120バイト程度になる。

- ・音声優先制御機能
ON/OFF
- ・バス当たり加入者数
数百人~1千人
- ・加入者当たり最繁時呼率（音声のみ）
0.05~0.3程度
- ・加入者当たりIPトラヒック（音声除く）
10~100 kbps
- ・中継回線収容率
0.7~0.9
- ・パケット損失率目標値
 10^{-4}
- ・End-End 遅延時間目標値
200 ミリ秒未満

5.2 固定的遅延

End-End 遅延要因の内、その遅延時間分布の分散がごく小さいものは一定であると見なし、固定的遅延と呼ぶことにする。一方、その分散が無視できない遅延要因は、変動的遅延と呼ぶことにして、遅延要因を両者に分類する。そのとき、音声への符号化・復号化遅延とパケット化遅延と伝送遅延を固定的遅延に分類することができ、各ルータでの送出処理待ちのバッファリング遅延時間は、変動的遅延となる。

表2のように、音声符号化方式によって、その符号化・復号化処理の遅延時間がかなり異なる。G.711は符号化速度が64 kbpsであり、G.729は8 kbpsであるが、圧縮効率の優れた符号化方式は、差分符号化などの高度な計算をしているため、その処理時間が大きくなるからである。

また、伝送遅延は、IP網+PSTN網で45ミリ秒とした。実際には、End-Endの距離に依存して変化するが、この値は国内でのEnd-End遅延時間（衛星回線は除く）としてほぼ最大の値である。

5.3 変動的遅延とトラヒックモデル

CATV網の各ルータでバッファリングされて待つ遅延時間の合計時間の分布は、各ルータでのバッファリング遅延時間の分布をそれぞれ算出した上で、それらを畳み込んで求める。それぞれのルータは、前段の

表2 音声符号化方式とその固定的遅延

方式	符号化・復号化遅延	パケット化遅延
G.711	0.75 ミリ秒	5 ミリ秒
G.729	25 ミリ秒	10 ミリ秒

複数のルータからのトラヒックが合流するので、前後のルータ間のトラヒックの相関はないと判断したことがその理由である。また、各ルータへのパケットの到着のモデルは、簡単のため、ポアソン到着とした。

本稿では音声優先制御がOFFの場合を考える。このとき、バスルータ、ヘッドエンドルータ、中継ルータの3段それぞれのトラヒックモデルは、 $M/H_2/1/\infty$ とした[2]。この区間では、音声パケットとインターネットへのデータパケットの両方が混ざっている。それぞれのパケット長の平均とパケット数の割合に従って、サービス過程を超指数分布 H_2 によって表現した。

一方、MGのトラヒックモデルは、 $M/M/1/\infty$ とした。MGを通してPSTN網に行くのは音声パケットだけなので、サービス過程を指数分布 M としたからである。

バッファサイズは、当然、有限になるが、計算の簡単さのために、バッファサイズは無量大 ∞ とした。実際、パケットの損失はほとんど遅延揺らぎ吸収バッファでの廃棄によるもので、各ルータでのバッファ溢れで損失することは非常に稀である。この理由は、遅延揺らぎ吸収バッファでのパケット損失率は 10^{-4} と規定するが、各ルータ装置内でバッファ溢れによりパケット損失となる確率は、これよりもずっと小さくなるように十分なメモリ容量を持っているからである。

したがって、音声優先制御がOFFのとき音声パケットの変動的遅延時間分布は以下の直列待ち行列モデルの遅延時間分布として算出できる。

表3 End-End 遅延時間の内訳

遅延の要因	遅延時間 (ミリ秒)
(a) 符号化・復号化+パケット化遅延	35.00
(b) 伝送遅延(国内IP網+PSTN網)	45.00
(c) 遅延揺らぎ吸収バッファ内遅延	24.54
End-End 遅延 (a)+(b)+(c)	104.54

[注3] 音声符号化方式はG.729で音声パケット60バイト、音声以外のIPトラヒックは加入者当たり10 kbps、データ・フラグメンテーションによりデータパケットは500バイト、音声優先制御機能はOFFとする。ヘッドエンドルータなど上位ルータは、100 Mbps以上の高速回線とした。

[注4] 連続する2つのパケットで、前のパケットが最短の遅延時間で到着し、後のパケットが最長の遅延時間で到着しても規定のパケット損失率を満たすことができるためである。これは、実運用上での一般的なやり方である。

表4 遅延時間分布の99.99%値の一覧

(c) 遅延揺らぎ吸収バッファ内 遅延の内訳	遅延時間分布の 99.99%値 (ミリ秒)
4段のルータ	12.27
バスルータ	12.23
ヘッドエンドルータ	0.18
中継ルータ	1.11
MG	0.15

- $M/H_2/1/\infty$ (バスルータ)
- $M/H_2/1/\infty$ (ヘッドエンドルータ)
- $M/H_2/1/\infty$ (中継ルータ)
- $M/M/1/\infty$ (MG)

5.4 数値例

数値例を紹介する。これは、5 Mbps のバスルータに加入者を 200 人収容した場合である[注3]。4 段階待ち行列モデルによる遅延時間分布を求めて、その裾の確率が 10^{-4} になる値 (分布の 99.99% 値) の 2 倍[注4]を、遅延揺らぎ吸収バッファで滞留させる時間とする。そのとき、End-End 遅延時間の内訳は、表3のようになる。

また、(c)遅延揺らぎ吸収バッファでの遅延時間は、前述のように、ルータ 4 段階の遅延時間分布の 99.99% 値の 2 倍として求めたが、その内訳として、各ルータでの遅延時間分布の 99.99% 値を表4に示す。

この数値例から、5 Mbps のバスルータに加入者 200 人規模では、End-End の遅延時間は 100 ミリ秒程度にしかならないので、目標値の 200 ミリ秒未満と比べた場合には、十分な音声品質が確保されていると判断できる。また、4 段階のルータでのバッファリング遅延のほとんど (約 99%) が 5 Mbps のバスルータで発生していることに着目すると、バスルータがボトルネックであることが分かる。これは、ヘッドエンドルータなど後段のルータが十分高速であるのに比べて、バスルータが極端に低速であるためである。しかし、5 Mbps という回線速度がバスルータの市販品のインタフェース速度であり、さらに高速にすることは容易でないため、結局、バスルータがボトルネックとなることは避けられない。したがって、IP 電話の音声品

質を決定するのは、バスルータの回線速度とそれを共用する加入者の人数などが最も支配的な要因であることが分かる。

6. まとめ

CATV 網で IP 電話を提供する場合の、CATV 網側から発信して PSTN 網側へ着信する通話について、音声品質を決定する End-End 遅延時間に対する評価を行った。これにより、IP 電話の音声品質について定量的に評価することができた。

また、ネットワーク設計の観点から、音声品質を左右する End-End 遅延時間を大きくする要因としては、バスルータでのバッファリング遅延時間が最大のボトルネックになることを示すことができた。バスルータ当たりの加入者数を変動させることなどで、音声品質とそれを提供するためのコストの関係もかなり具体的に評価できる見通しが立った。

一般電話は国内での End-End 遅延時間はほとんど 30 ミリ秒以内とされていることなどから、結論的には、CATV 網での IP 電話は一般電話よりも音声品質の面では悪いと言わざるを得ない。しかし、衛星回線を使用する場合よりも良いので、コスト次第でサービスとしての競争力は無視できないと考えられる。特に、音声優先制御機能を備えた CM などの装置の価格が安くなったときには、End-End 遅延時間が短縮されて音声品質がかなり良くなるので、この段階で、一般電話に対して CATV 網からの IP 電話が競争力を増すことが十分に有り得る。

今後の課題としては、バスのマルチアクセスによる衝突などの現象を適切に反映するように、CSMA/CD 方式などに基づくトラヒックモデルで再評価することを検討したい。

参考文献

- [1] 天野他, “VoFR, VoIP で音声網を再構築”, 日経コミュニケーション, 1999.6.7.-1999.7.19.
- [2] 藤木他, “通信トラヒック理論”, 丸善, 1980.
- [3] 日本フレームリレーフォーラム編, “フレームリレーネットワーク構築ガイド”, 電気通信協会, 1998.