

リングバスの伝送効率の測定と評価

高橋幸男 海老原義彦 池田克夫* 石坂充弘
(筑波大学) *(三菱電機)

1. まえがき

GAMMA-NETのリングバスサブシステムRBSTMには、統計データ収集機構SDG(Statistic Data Gatherer)が付加され、この機構を用いてRBS内のデータ転送に関する伝送効率の測定を行うことができる。GAMMA-NETは複数の異機種計算機システムを高速の光ファイバリングバスで結合したローカルコンピュータネットワークである。そして、SDGは、データリンク制御をハードウェア/ファームウェアで実行するリングバスプロセッサRBP内に実装され、ハードウェアで構成されているために高精度で実時間性を損わない測定が可能である。

データ転送の同期にはランデブー方式が採用されている。この機能を用いることにより、ネットワーク制御層の機能をバイパスしエンドプロセスのバッファ間でデータ転送を可能とするバーストモードが設定されている。この方式により、大容量の高速・高信頼のファイル転送を可能にしている。

本稿では、RBS内のメッセージ遅延とスループットに関してSDGの統計データ収集機能を用いて測定した結果について述べ、その測定結果から伝送性能について評価している。

2. データ転送手順

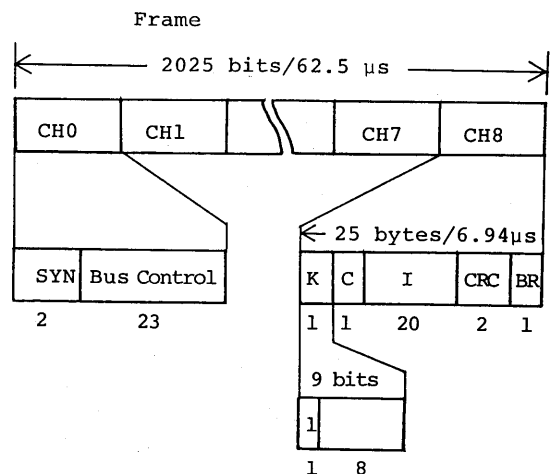
RBSは、優れた伝送特性を持つ光ファイバによる2重のリングバスとハードウェア/ファームウェアでデータリンク制御を実行するRBPから構成されている。表1にRBSの概略仕様を示す。

2.1 チャンネル構成

リングバスの伝送チャンネルは図1に示されるように9個のサブチャンネルに分割され巡回している。チャンネルCH0はフレームの同期制御や後述する統計データ転送用に使用されている。他のCH1からCH8のチャンネルではデータ転送用に使用され、複数のサブシステム間の通信が同時に実行できる。また、各ホスト計算機内のプロセスが高速の

COMMUNICATION LINE	OPTICAL FIBER CABLE 32.4Mbps
NUMBER OF STATIONS	MAX.32
STATION DISTANCE	MAX.1km
CHANNEL	320kBps per channel 8 data channels 1 control channel
NUMBER OF DATA LINKS	MAX.31 per node
DATA TRANSFER MODE	BURST MODE (1 logical link per data link) MULTIPLEX MODE (MAX.4096 logical links per data link)

TABLE 1 SPECIFICATION OF RING BUS SUBSYSTEM



SYN : synchronous word
CRC : cyclic redundancy check
K : channel access key
C : command field
I : information field
BR : busy/response

FIG.1 FRAME STRUCTURE

データ転送を要求する場合には、*RBS*は複数のチャンネルを束ねてデータ転送を実行できる。

2.2 リンク構成

*RBS*の論理的な構成は、図2のような階層構造で表わされる。*RBS*はネットワーク制御層 *NCP* とデータリンク制御層 *DLP* の働きによって通信を行なっている。各通信プロセスは、*NCP* が確立した論理リンクを用いて通信する。*NCP* は図2のようにマルチプレックスとバーストモードの2種のデータ転送モードを同時に管理している。マルチプレックスモードでは、1データリンクを複数の論理リンクに対応させて *TSS*、*RJE* などの短いデータを転送するモードである。そして、2データリンクによる全二重モードで使用される。最大転送量は1データリンク当り2kバイトである。バーストモードでは、1データリンクを1論理リンクに対応させて、ファイル転送のような長いデータを転送するモードである。1データリンクを半二重モードで使用される。そして、プロセス間のバッファ間で直接データ転送するので、大量データを高スループットで転送できる。最大転送量は1データリンク当り32kバイトである。

*DLP*の主な機能は、計算機のチャンネルに接続される *RBP* で実現される。

*DLP*の主な機能を次に示す。

(1) *RBP*間のデータリンクの確立と解放。

(2) ランデブー方式を用いたデータ転送。

データ転送段階では、*RBP*はリングバス上の空きチャンネルを送信が完了するまで捕捉してパケットを送信する。

(3) データ転送のエラー検出と再送制御。

2.3 ランデブー方式

データリンクの両端で送受信要求が発生したことを *RBP* が確認した時、データ転送が開始される。この方式がランデブー方式である。ランデブー方式は送信側と受信側のデータ転送の同期をとる方式である。

CPU-A からデータ送信起動 *WRITE* の発生により *RBP-A* から送信要求コマンド *WR* が *RBP-B* に対して送信される。同様に、*CPU-B* からデータ受信起動 *READ* の発生により *RBP-B* から受信要求コマンド *RD* が *RBP-A* に対して送信される。各 *RBP* でそれぞれ *WR* と *RD* を確認し、*RBP-B* が受信可能状態である場合、*RBP-B* から確認コマンド *STS* が *RBP-A* に対して送信される。この後、データ転送は空きチャンネル確保コマンド *WT* が *RBP-A* から *RBP-B* に送信された後に開始される。*RBP-A* は空きチャンネルを

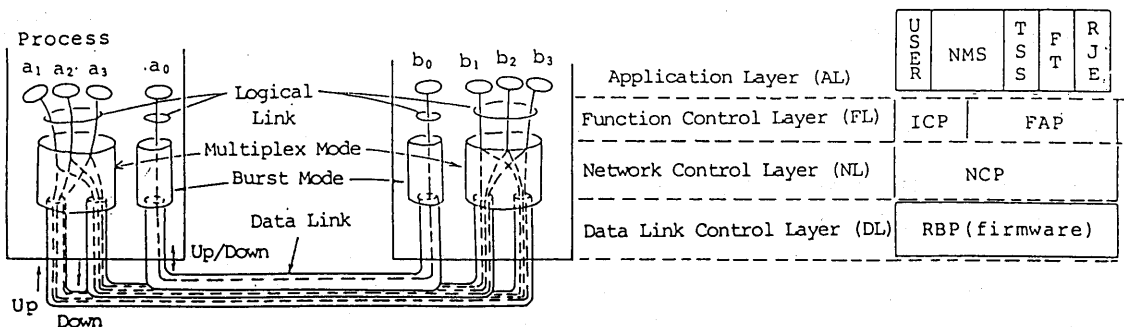
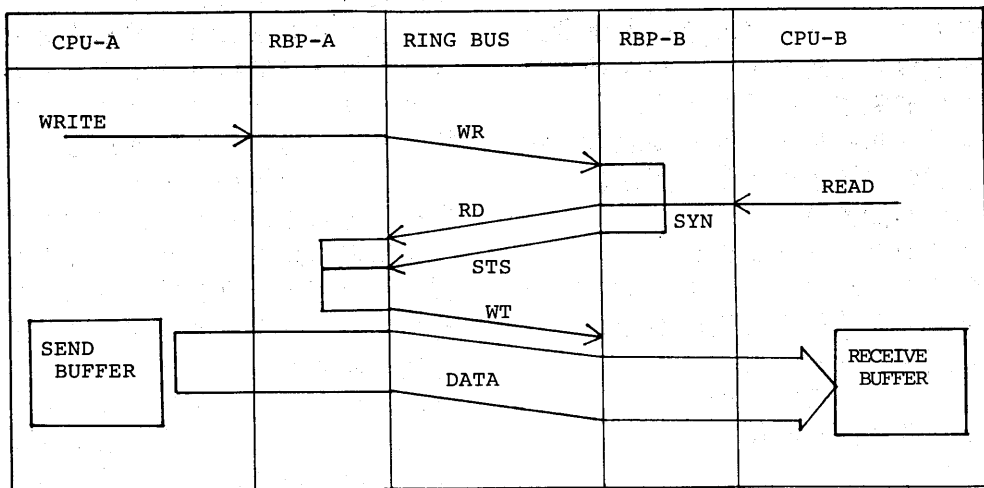


FIG.2 LINKS AND PROTOCOL LAYERS



SYN : synchronizing

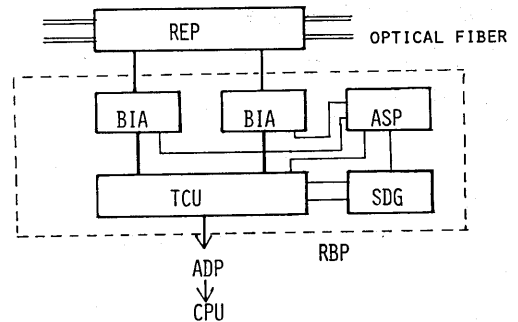
FIG. 3 DATA TRANSFER SEQUENCE ON BURST MODE

探し、そのチャンネルにWTを送信する。WTによって一度チャンネルを確保した後、続くデータは連続して同一チャンネルに送信される。図3は、この通信手順を示している。データ転送はバーストモードの場合である。

3. 統計データ収集判定方法

3.1 統計データ収集機構

GAMMA-NETには、RBS内のデータ転送に関して、動作特性の統計データを収集する機能が実装されている。この機能は統計データ収集機構SDGによって実行される。SDGはリングバス上の各RBP内に実装されている。SDGはハードウェア/ファームウェアで構成されているため、データ転送に関する事象を高精度で実時間性を損わない程度の遅延で記録することができる。図4は、RBPの構成を示している。RBPは、バスインタフェースアダプタBIA、付加サービスプロセッサASP、伝送制御ユニットTCUおよびSDGより構成されている。RBP内のBIAは光リピータを介して光ファイバリングバスに接続



REP : repeater
BIA : bus interface adapter
ASP : attached service processor
TCU : transmission control unit
SDG : statistic data gatherer
ADP : interface adapter
RBP : ring bus processor

FIG. 4 RBP STRUCTURE

されている。データ転送に関した事象、データリンク、相手局RBPの情報がTCUからSDGに伝えられる。ASPは、統計データ収集の開始の指示、統計データ収集の終了の指示、後述する統計1、統計2の指定の指示をSDGに伝える。また、48kHzと16kHzのクロックが用意されていて、このクロック単位時間で記録される。

SDGは、統計1と統計2の2種類の統計データを収集することができる。

(1) 統計1

統計1には、データ転送の各事象の時間を累積した統計データが収集される。次に収集する項目を示す。

- ・送信/受信起動コマンド(WRITE/READ)発生時刻から送信/受信開始までの時刻の時間累積。
- ・送信/受信データ数累積。
- ・送信/受信起動コマンド(WRITE/READ)の正常/異常終了回数の累積。
- ・相手局ビジー回数累積。
- ・CRCエラー回数累積。
- ・リンク確立時間累積。

(2) 統計2

統計2には、1データリンクに対して次の各事象の発生した時刻を連続的に追跡した記録が収集されている。

- ・送信/受信起動コマンド(WRITE/READ)発生時刻。
- ・送信/受信開始時刻。
- ・CRCエラーによる送信/受信再開時刻。
- ・オーバーランによる送信再開時刻。
- ・相手局ビジー時刻。
- ・NG受信時刻(CRCエラー発生)。
- ・送信/受信異常終了時刻。
- ・送信データ数(正常終了の場合)。

SDGには、4kバイトのメモリが付加されており、統計1または統計2のどちらか指定された統計が記録される。

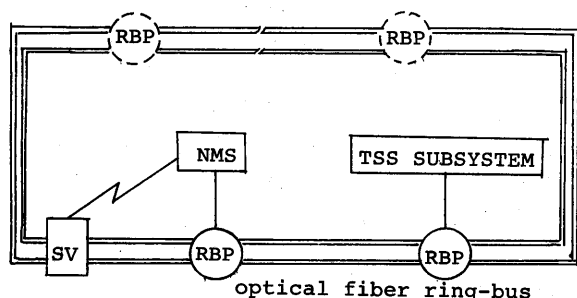


FIG.5 GAMMA-NET CONFIGURATION

3.2 統計データ収集方法

リングバス上には、RBPを介してネットワーク管理サブシステムNMSやリングバススーパーバイザSVが接続されている。(図5) NMSは、主にネットワーク管理やネットワーク案内を行っている。また、SVは主にフレーム同期の維持、不要パケットの除去バスの障害の検出と診断などを行う機能がある。NMSとSVは専用回線で接続されており、両サブシステム間の通信はリングバスを介さずに行える。

統計データ収集の開始および終了コマンドは、NMSの制御卓から入力される。NMSに入力された開始コマンドは、専用回線を介してSVに伝えられる。SVはその開始コマンドをリングバスを巡回しているフレーム内のチャネルCHQを使って各RBPに伝える。各RBP内のSDGは、この開始コマンドに従って統計データの収集を始める。この時、統計の種類およびデータリンク番号を指定することができる。統計データ収集の終了コマンドも同様にリングバス上の各RBP内のSDGに伝えられる。

各SDGは終了コマンドを受信すると、それまでに収集された統計データをリングバスのチャネルCHQを使用してSVに転送する。SVは各SDGから受信した統計データを一定のフォーマットに整えて、NMSに転送する。

NMSには、統計1と統計2のための2種類のファイルが用意されており、SVから転送された統計データは、それぞれのファイルに格納される。

図6は、この手順を示している。

3.3 測定方法

各RBP内に実装されているSDGの機能を用いてRBSのデータ転送に関する動作特性を測定した。

測定は、TSSサブシステム(MELCOM 800Ⅲ, 2CPU, 3MIPS, 4MB)とNMS(

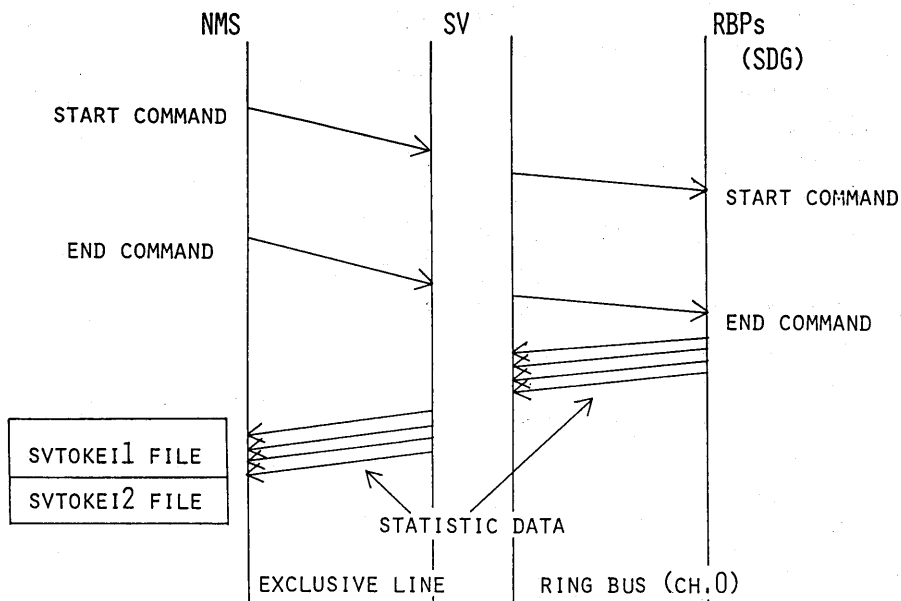


FIG. 6 STATISTIC DATA GATHER SEQUENCE

MELCOM 70/50, 0.6MIPS, 0.8MB)間のデータ転送について行われた。測定は、単方向通信で送信側はTSSサブシステム、受信側はNMSである。

GAMMA-NETの機能制御層FL上にあるプロセス間通信プロトコルICPの機能を用いて測定用のデータ転送プログラムを作成した。これは、メッセージ長を指定して、連続して送信命令SEND/受信命令RECVをくり返し実行する送信用/受信用のプログラムである。このプログラムを用いてRBS内のデータ転送に関する基本的動作特性の測定を行った。

また、測定は統計スガ使用された。

4. 測定結果

図7に示すように、アイドルタイム、ウエイトタイム、トランスミッションタイムを定義する。送信側について考えれば、アイドルタイムは任意のnについて第n-1番目のデータ転送の終了後、第n番目の送信起動コマンドWRITE

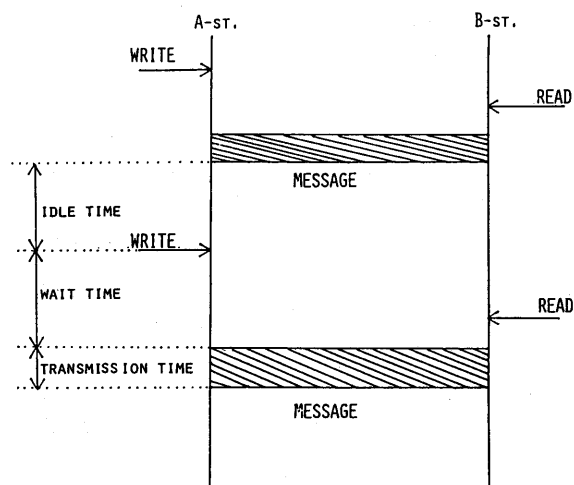


FIG. 7 DATA TRANSFER SEQUENCE

が発生するまでの時間間隔であり、ウエイトタイムは送信起動コマンドWRITEが発生し実際にデータ転送が開始されるまでの時間間隔と定義する。トランスミッションタイムは送信開始から送信終了までの時間間隔とする。受信側

についても同様にそれぞれアイドルタイム、ウエイトタイム、トランスミッションタイムと定義する。収集された統計上によってこれらは算出できる。

4.1 測定結果

図8から図10に測定結果を示す。図8は送信側RBPで測定したデータ転送遅延である。すなわち、送信側RBPにCPUからのデータ送信起動コマンドWRITEの到着時刻からデータ転送の終了時刻までの時間間隔を示している。データ転送遅延は、トランスミッションタイムとウエイトタイムの和である。トランスミッションタイムは転送データ量に比例して増大している。

図9は受信側RBPで測定したデータ転送遅延である。すなわち、受信側RBPにCPUからのデータ受信起動コマンドREADの到着時刻からデータ転送の終了時刻までの時間間隔を示している。同様にトランスミッションタイムは転送データ量に比例して増大している。

ウエイトタイムは、送信側RBPと受信側RBPでの測定で値が異なっている。この理由は、データ転送方向が一方であり、送信側CPUからWRITEが受信側CPUからのREADよりも短時間に発生しているからである。ランデブー方式による送受信の同期が完了するまでデータ転送が実行されないため、各RBPで測定されたウエイトタイムが異なるのである。

図10はスループットを示している。トランスミッションタイムだけを考えれば、チャネル容量にほぼ等しい値を示している。ウエイトタイムまでを含めたスループットは、チャネル容量の約94%に達している。

送信側RBPで測定したアイドルタイムは、ほぼ4.6msecから6.7msecに分布しており、ウエイトタイムは、ほぼ4msecから20msecに分布している。受信

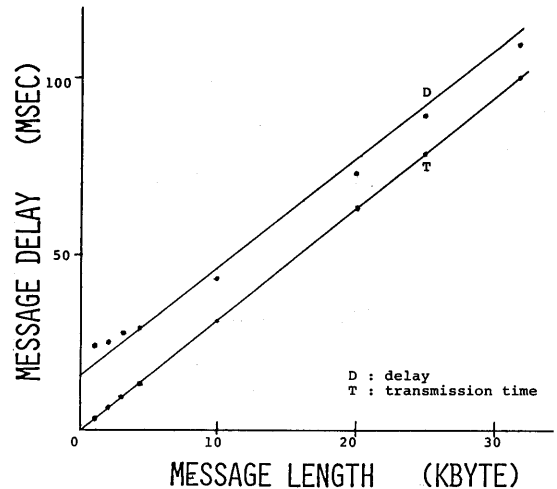


FIG.8 MESSAGE DELAY VERSUS MESSAGE LENGTH (sender)

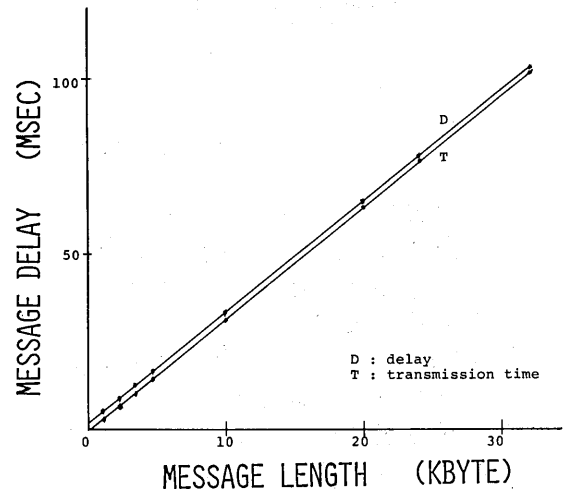


FIG. 9 MESSAGE DELAY VERSUS MESSAGE LENGTH (receiver)

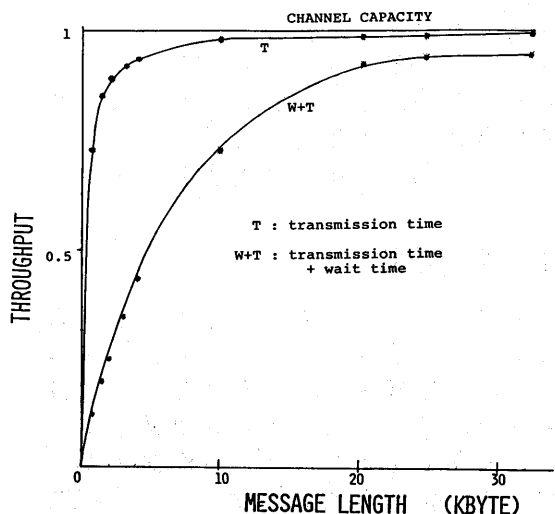


FIG.10 THROUGHPUT VERSUS MESSAGE LENGTH (sender)

側RBPで測定したアイドルタイムは
ほぼ10msecから20msecに分布しており、
ウェイトタイムは0.9msecから1.1msecである。

5. 結論

前章に測定結果を述べた。本測定は
RBPの動作を測定している。GAMMA-
NETのデータリンクの動作はRBP
によって制御され、CPUはチャ
ネルI/Oと同様にRBPを起動してい
る。従来のネットワークのように、通
信用フロントエンドプロセッサのリフ
トウェアで通信処理を行わず²⁾の
でなく、GAMMA-NETでは、通信処理をハ
ードウェア/ファームウェアで実行し
伝送効率を計っている。

図9より、ランデブー方式による送
受信の同期完了後から受信側RBPが
データの受信を始めるまでの時間間隔
(受信側のウェイトタイム)は、0.9msec
から1.1msecの範囲で分布している。
すなわち、前章の図3で示したとおり、
この時間内に受信側RBPがRDおよ
びSTSを送信し、送信側RBPから
のWTを受信できることが示されてい
る。

図10より、送受信の同期完了後デー
タ転送のスループットは1チャンネル分
の容量にほぼ等しいことが測定された。
この理由は、WTを1チャンネルに送出
した後はそのチャンネルを確保しデー
タ転送を実行するため、RBSはTDMA
のような動作をするからである。

以上のことから、RBPは所期の設
計目標どおりに作動していると結論で
きよう。

データリンク制御にランデブー方式
を採用しているために、データ転送は
送受信の同期が完了するまで待ち状態
になりウェイトタイムが長くなる原因
となっている。この理由は、送信およ
び受信側CPUでRBPに送信および

受信の起動を行うリフトウェアオーバ
ヘッドが大きいからである。本測定で
は、アイドルタイムとしてこの時間が
測定されている。また、各CPUに対
する負荷の影響も大きく、測定時の負
荷によってアイドルタイムの平均およ
び分散が大きく変動している。CPU
のサイクルタイムにも関係があり、送
信側CPUは3MIPS、受信側CPUは
0.6MIPSであるため、送信側のアイド
ルタイムは受信側のそれよりも短い。

しかし、ランデブー方式は受信RBP
でバッファブロッキングによる受信
のビジーになる場合を少なくしている。
すなわち、ランデブー方式によって送
受信プロセスが同期するからである。
従来のデータリンク制御プロトコル³⁾
のように、送受信プロセスの同期なしに
送信側RBPがデータ転送を始め、受
信側CPUでは受信状態にないとき、
受信側RBPのバッファがあふれるこ
とがある。これを極力避けるためにラン
デブー方式を採用した。そして、受
信ビジーによる再送が少ないので、チ
ャネルを有効に利用できる。バッファ
ブロッキングの解析は今後の検討課題
である。

なお、測定ではバーストモードが使
用されている。バーストモードはファ
イル転送などの大量データを高スル
ープットで転送するために使用される。

現在、GAMMA-NETにおいてファ
イル転送プロトコルは構築段階にある。
例えば、図8より32Kバイトのデー
タは、約110msecの遅延で転送でき
ることが示されている。また、GAMMA-NET
では、5個のチャンネルを束ねてデー
タ転送を行う機能(マルチチャンネルア
クセス)もあり、本測定の5倍のスル
ープットが期待できる。ファイル転送
の評価はここで深く言及できないが、
十分なレスポンスを期待できると思
われる。

6. あとがき

本論では、最初にRBSの通信機能と測定機能について述べ、次にこの測定機能を用いて測定した結果より、GAMMA-NETの通信特性について論じた。測定結果を用いた通信特性の種々の解析が今後の検討課題となる。

また、これまでのデータリンクの性能評価だけでなく、システム全体を考慮した総合的な性能評価も今後検討すべき課題である。

更に、データ転送において通信ソフトウェアオーバーヘッドによる影響が大きいという測定結果が示されたことより、通信に効率のよいネットワーク向きOSの開発が将来必要であろうと思われる。

7. 謝辞

この研究に関して暖かい支援を頂いた筑波大学学術情報処理センター長中山和彦教授に感謝する。

8. 参考文献

- (1) K. IKEDA, et al, "Computer Network Coupled by 100Mbps Optical Fiber Ring Bus -System Planning and Ring Bus Subsystem Description-" COMPCON 80, 1980
- (2) F.E. Heart, R.E. Kahn, S.M. Ornstein, W.R. Crowther and D.C. Walden, "The interface message processor for ARPA computer network, AFIPS Conf. Proc., Vol. 36, pp. 551 ~ 567 1970
- (3) E.J. Weldon, J., "An Improved Selective-Repeat ARQ Strategy" IEEE Trans. Commun., Vol. COM-30, No.3, March 1982