

# GAMMA-NET の プロトコル構成

海老原義彦, 鶴岡知昭, 池田克夫, 中山和彦 (筑波大学)

西川正文, 新沢誠 (三菱電機)

## 1. はじめに

現在の計算機システムはますます高速化, 大型化が進み, 年々経費対性能比が飛躍的に向上している。しかし一つのシステムに多様な仕様を盛り込もうとするため, いたずらにオーバヘッドの増大を引き起こしている面もある。この問題に対処する一つの方法には, 多様な利用形態を専門的に分担するシステムの分散化が考えられる<sup>(1),(2)</sup>。また RAS の向上の面からも単一の大型システムを運転するよりも実際的であると考えられる。本稿の目的は最新の LSI 技術や光ファイバ伝送技術と積極的に導入して超高速伝送路による複合システムを構築し, 異機種間を効率良く,かつ柔軟性のある結合を達成するためのプロトコルを確立することにある。従来の計算機ネットワーク<sup>(3),(4),(5)</sup> (ARPA, CYCLADES や ALOHA など) のプロトコルに関する論文は数多くあるが, これらのプロトコルは遠距離間を結ぶ比較的低速回線に適用されたものである。また代表的なローカルネットワーク (NBS net, Ether net<sup>(6)</sup> や Milre net など) も 1~10 Mbps 程度の転送速度を対象としているのが, 大多数である。

本稿では筑波大学学術情報処理センタに設置されている GAMMA-NET (General purpose And Multi-Media Annular NETWORK) システムに実装されたプロトコル体系について論ずる。特にプロセス間通信における両プロセスの読み/書きの I/O 状態を一致させる制御方式 (ランデブー方式) を導入することにより, 過負荷トラフィック時のチャネル効率を高めるためのデータフロー制御を提案している。さらに提案方式の具体的なアルゴリズムを示し, その妥当性を論述する。

次にネットワークユーザがネットワーク資源

を効率良く利用できるような支援する網管理サブシステム (NMS: Network Management sub-System) について述べる。RAS の向上から考察した NMS のネットワーク全体での位置付け, 資源情報の収集・管理・提案方式や NMS の有効性を論ずる。

以上の2つのプロトコルを中心にその特徴を述べ, 最後に稼働中のシステムをとおいて得られた諸問題をまとめる。

## 2. GAMMA-NET のシステム構成

本システムの伝送路には光ファイバケーブル (32 MBPS) が使用されており, 伝送路の経済性から環状形が採用されている<sup>(7)</sup>。このリングバスとは端末制御 (TIP) 用 [MELCOM 70], CAI 教育用 [MELCOM 700 III], 会話型 [MELCOM 700 III], 事務用 [FACOM M-160 F] などのように, 利用形態により専門的に調整したサブシステムがチャネルインターフェースを介して接続している (図1)。

その他バッチ用 [FACOM M-200], ファイル, バルク型サブシステムなどが計画中である。伝送路は時分割多重チャネルのデュアルバスである。1つの伝送路は, バス監視制御などのための制御用チャネル (以下制御用スロットと呼ぶ) と8つのデータ用チャネル (以下データスロットと呼ぶ) から成る。前者は図1のリングバス監視装置 (RBSV) が使用する。1スロット当り 3 Mbps の転送能力を持つ。

## 3. GAMMA-NET のプロトコル構成

異機種間の通信を達成するため, 本システムには4つの階層からなる通信規約が設定されている。サブシステム間のデータ通信を規約するデータリンクプロトコル (DLP), プロセス間の論理リンクを制御するネットワーク制御プロトコル (NCP), 機能制御プロトコル (FCP) とアプリケーションのためのアプリケーション (ALP) である。ALP には会話処理 (TSS), 遠隔端処理 (RJE), 網管理 (MSP) とファイル転送

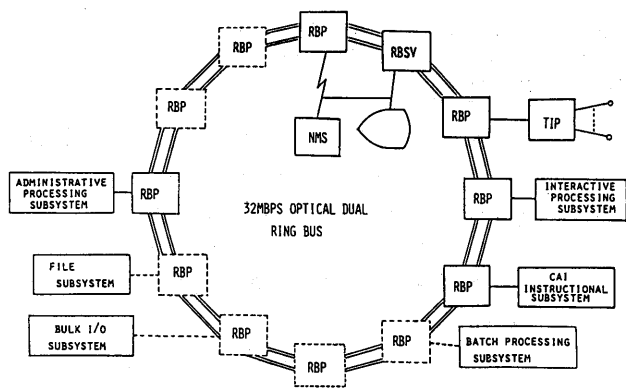
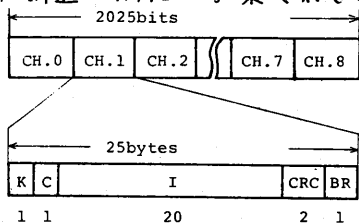


Fig. 1 GAMMA-NET STRUCTURE

機能(FTP)が存在する(図3)。特にMSP機能は前述のNMSに実装されている。



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

Fig. 2 Frame structure

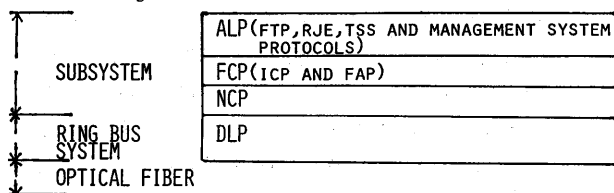


Fig. 3 GAMMA-NET PROTOCOL STRUCTURE

### 3.1. データリンク プロトコル

基本的にはDLPはサブシステム間に設定された論理的パス(データリンク)を介してパケットの送受信を行うための規約である。DLPの主な機能を次に示す。

- (1) データリンクの設定と切断機能
- (2) データの送受信 および データのフロー制御機能
- (3) パケットのエラー検出 および 再送制御機能

これらの機能は計算機チャンネルに接続されるリングバスプロセッサ(RBP)で実現され、ソフトウェアからはデータリンクを通常の入出力チャンネルと同様に入出力処理を行うことができる。即ち計算機サブチャンネルとデータリンクは1対1の関係にある。またDLPの大部分の機能をソフトウェアにより実現し、高速なDLP処理を計り、このレベルの通信制御オーバーヘッドを極力少ないものとしている。ARPAのIMPは最大4台の計算機と接続することができ

るが、本システムのRBPは1つの計算機だけである。以下に(1)と(2)の基本的動作手順を示す。説明の前提としてRBPの物理アドレスRAをもつサブシステムSAとRBをもつサブシステムSBが存在し、今SAからSBへのメッセージ転送要求があったものとする。

#### (a) データリンクの設定動作

(i) SAは相手SBのRBPアドレスR<sub>B</sub>を知っているものとする。SAは自分のRBP(RBP-A)にR<sub>B</sub>を通知し、データリンクの設定要求を行う。RBP-Aは未使用のリンク番号(i)を選択し、iをSAに通知すると同時にリンク設定コマンドCLを相手RBP(RBP-B)に知らせる。このことによりRBP-BにR<sub>A</sub>とiを知らせたことになる。

(ii) CLを受信したRBP-Bは未使用リンク番号(j)を選び、jをSBに通知すると同時に、設定確認コマンドCAを返送する。これにより、RBP-AにR<sub>B</sub>とjを知らせる。この時点でiとjの対応したデータリンク(i, j)を設定したことになる。

(iii) 他方CAを受信したRBP-Aは情報R<sub>B</sub>から設定中の相手対応番号はjであることを知る。

#### (b) データの送受信動作

(iv) データ送受信が可能となる条件はRBP-Aからの送信要求コマンドWRとRBP-Bからの受信要求コマンドRDとの相互確認がなされることである。その後RBP-Bからの確認コマンドSTSが送信される。

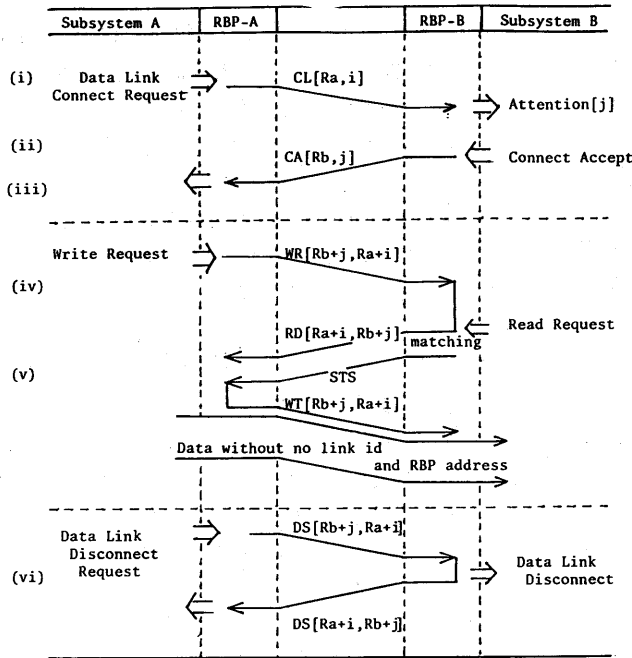


Fig. 4 Data Link Control Flow

(V) データ転送の開始に当り、空きスロット確保コマンド WT が RBP-A へ送り送信される。WT 自身の送信は空きスロットを利用するが、両方の RBP は WT が送信されたスロット番号を記憶し、続くデータは連続して同スロットで転送される。1 回のデータ転送量は最大 32 K バイトに現在、設定してある。データ転送はメッセージ交換タイプである。

(C) データリンクの切断手順

(VI) データリンクの切断はサブシステム間 A と B で切断コマンド DS を交信し合い相互に確認された後行われる。

一般に、プロセス間通信においては送信プロセスが書き状態にあり、かつ受信側の受信バッファが空いており、受信可能な状態にある時にメッセージが転送される。ARPA の RFNМ制御メッセージや ALL 制御メッセージなどが上記のデータフローを実現するための主な役割りを演じている。しかし受信プロセスがメッセージを受信できる条件としては、上記の条件に加えて、受信プロセスが読み込み状態にあることが必要である。本システムのプロセス間通信はシステムプロセスばかりでなくユーザ・プロセス(ユーザ・プログラム)間通信も含まれている。一般論としてユーザ・プロセスはスワップイン/アウトの対象となるため、常に読み込み状態にある訳ではない。

このため本システムではバッファ管理ばかりでなく、同時にプロセスの巧み状態を考慮したデータフロー制御を採用している。本システムの1スロットで転送できる単位をパケットと呼ぶ。図4に示したように制御用コマンドはパケット交換方式であり、データ転送はメッセージ交換方式で行われる。メッセージ交換の長所はパケット毎の発信着信地の宛先を制御情報にとられるパケットヘッダのオーバーヘッドを極力少なくし、実効転送効率を高める点にある。かついったんデータ転送が開始すると RBP はデータ転送に専従するため、DLP 実装が簡易化することができる。反面ファイル転送のような長時間データ転送に対しては、使用中スロットの長時間確保のため、他の RBP の空きスロット確保を困難

	データリンクと論理リンクの対応	送信確認メッセージ	全2重/半2重モード	適応分野	フロー制御	優先制御
バーストモード	1 対 1	拡張機能以外なし	1 データリンクを半2重モードで使用	ファイル転送などの長いデータ	原則としてなし	なし
マルチプレックスモード	1 対 n 最大 n=4096	同上	2 データリンクを全2重モードで使用	TSS, RJEなどの短いデータ	有り	高/低あり メッセージ単位の優先処理

Table 1 Message Transmission Modes

にする。また データ転送中にある RBP は他の RBP からの パケットをビジー状態 (図2の BR が示す情報) として受け取れない。このための パケット再送オーバーバインドが発生する場合もある。

1 回の書き込み (ライト) 要求で転送できる最大バイト長はファイル転送を対象としたバースト転送では 32K バイト、会話処理のマルチプレックス転送では 2K バイトに現在は設定している。

### 3.2. ネットワーク制御プロトコル (NCP)

NCP はデータリンク上の論理的な通信路を規定し、プロセス間通信制御を行うものである。NCP は上位プロトコルの指示で決まる 2 つのデータ転送モードを用意しており、データの性格に適合したデータ転送を行う。表1 に本システムのもつバーストとマルチプレックス・モードの 2 つの転送モードの特徴を示す。

### 3.3. 機能制御プロトコル (FCP)

FCP はデータメッセージおよび割込みメッセージの送受信機能を規定する。また FCP はプロセス間通信プロトコル (ICP) と ICP の上位に位置し、すべてのデータのアクセスを仮想論理ファイルのアクセスとみなして制御するファイルアクセスプロトコル (FAP) から成る。但し仮想論理ファイルの実ファイル (ディスク、CR、LP 装置など) への対応処理は上位のアプリケーションプロトコルで行われる。以下の表2 に FAP の主な機能を示す。

<ul style="list-style-type: none"> <li>・システム構成情報 (OS タイプ、ファイル機能、データ転送長) の交信</li> <li>・ファイル情報 (データタイプ、ファイル編成と属性、レコード形式) の交信</li> <li>・ファイルのアクセス制御</li> <li>・ファイル内のレコードの読み書き</li> </ul>
---

Table 2. Main Functions of FAP

て、応用レベルの特有な機能制御を規約したものである。この規約を支援するプログラムをプロセスと定義する。本システムには 4 つの応用プロセスがある。

- (1) TSS プロセス ; 端末の標準化を計るため、仮想端末プロトコルを定義し、それを支援するプロセスである。
- (2) RTE プロセス ; ジョブの送受信、ジョブの実行および問い合わせのための標準化を計る RTE プロトコルを定義し、それを支援するプロセスである。
- (3) ファイル転送プロセス ; ファイルの削除、ファイル名変更とファイル・レコードの参照、格納、追加および変更の標準化を計りそれらの機能を支援するプロセスである。
- (4) ネットワーク管理プロセス ; サブシステムのもつ資源案内、負荷分散処理、ネットワーク統計データの収集・管理を行うためのプロセスである。

以上の応用プロセスの他にユーザプロセス間通信のために、プロセス間ライブラリとファイルアクセス・ライブラリの 2 種類のライブラリが提供されている。現在までに、TSS プロセスと NMS プロセスのサービスが稼動中である。ALP レベルから見た実効転送速度は、リングバスの転送速度の  $\frac{1}{3} \sim \frac{1}{5}$  程度である。

以下特にネットワーク管理システムの基本的構成方式を中心に述べる。

### 3.4. アプリケーションプロトコル (ALP)

#### 3.4.1 ALP の基本概要

ALP は FAP または ICP 機能を使

### 4. NMS の構成方式

まず最初にネットワーク・ユーザがネットワーク資源を利用するためには、利用すべき資源がどの計算機にあり、利用

でまうる状態の否かの情報を入手する必要がある。そこで情報を問い合わせるネットワーク・ユーザ(以後ユーザの代りにTIPとも呼ぶ)と情報を提供する側(以後ホストと呼ぶ)との関係から資源情報の管理方式を述べる。前提として、各ホストはそれ自身の資源を独自に管理しており、それらに関する情報(これをローカル情報と呼ぶ)をそのホストが定める形式に従って提供できるものとする。

以下ネットワーク資源管理の方法を幾つか示し、その特徴を論述する。

#### 4.1 資源情報管理方式

第1の情報管理方式は図5に示すように各ホストが自己の資源の提供のみ関与し、ユーザは要求する資源を探し出すまで、TIPとホストの接続/切断を繰り返してゆく利用形態である。システム作成上、最も単純な方法であり、ホストのシステム・ダウンがネットワーク全体に与える影響は最も少ない。ネットワーク・ユーザはシステム・ダウンのホストがもつローカル情報を失うが、そのローカル情報が致命的情報でない限り、ネットワークシステムの利用を続行することができる。しかも個々のホストが定める情報形式のまま、ユーザ端末に出力することができる利点がある。

しかしホストの数が増加するに従い、ネットワークユーザの資源情報の検索操作を非常に困難なものとし、ネットワークシステム利用の容易性を損うことになる。

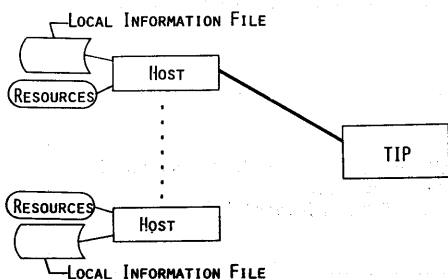


FIG.5 DISTRIBUTED MANAGEMENT

第2の情報管理方式は図6に示すように、各ホストのもつローカル情報を集中的にNMSが管理する利用形態である。TIPは必要な資源をNMSに問い合わせ、資源の存在するホストにアクセスする方法である。第1の方式に比べて、ネットワークユーザは個々のホストの物理アドレスを意識することが比較的軽く済み、ネットワーク全体を一つのシステムとみなせうる。第2の方式はネットワーク資源の有効的利用からも、またネットワークユーザのネットワーク・アクセスの容易性からも最も良い情報管理形態であると思われる。さらにNMSの管理する情報はマッシュリダブルな形式で統一する必要がある。このことはTIPばかりではなく、他のホストからも資源情報の検索も可能にする利点をもつ。反面、すべての資源情報の形式を統一することの労力は多大なものである。かつ情報の内容が頻繁に変更するものはNMSとホスト間の転送オーバーヘッドが増大する恐れがある。さらにサブシステムの1つであるNMSの障害・システムダウンがネットワークシステムに与える影響は甚大であり、RASの向上の面から、必ずしも最適な形態とはいえない。

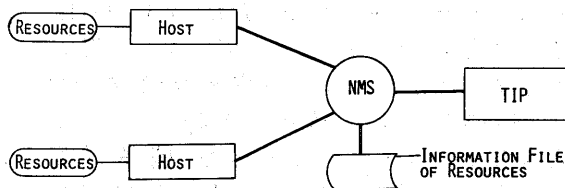


FIG.6 CENTRAL MANAGEMENT

第3の情報管理方式は、第1と第2の折衷案とも言えるものである。即ち提供する情報をネットワーク全体に共通なものとしてホスト固有な情報に分割することにより、ネットワーク上でNMSの占める情報管理の役割を前者に限定し、ネットワークユーザに利用させる形態である。図7に示すようにローカル情報は各ホストの定める形式をもつ。ローカル情報のうち、ネットワーク共通

な情報は ネットワークが規定する標準形式に変換され、ネットワーク共通資源情報として NMS が管理する。NMS は 共通資源情報の収集・管理機能の他に、これらの情報を検索する機能を持ち、TIP の問い合わせに対して、適切な共通情報を提供する機能も有する。

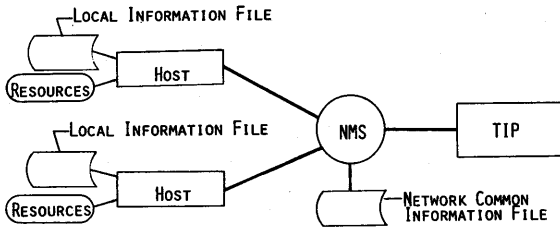


Fig.7 MANAGEMENT OF GAMMA-NET

このことにより、NMS のシステムダウンが発生してもネットワークシステムは自動的に NMS を切り離し、少なくとも第 1 の利用形態にネットワークシステムを縮退化して運営を続行することができる。GAMMA-NET のもつ資源情報管理の形態は第 3 の管理方式を採用している。NMS が管理の対象としていふネットワーク共通資源情報はネットワーク全体に関する情報である。かつ時間がある程度経過しても、その情報の提供がネットワークユーザによって、有効である。またホストが特定の資源を頻繁に使用するもの、機種種の内部形式に依存するオブジェクトファイルなどや機種特有のハードウェア資源に関する情報提供は各ホストの管理にまかせま方が良いと思わ

れる。具体的に、本システムの NMS が管理するネットワーク共通資源情報を表 3 にまとめる。

次に NMS の管理形態を遂行するための情報収集・提供の方法を以下に述べる。

#### 4.2. 共通資源情報の収集方法

共通資源情報は情報形式、情報のもつ内容がネットワーク全体で一意に定義されている。情報収集は原則的には一定時間間隔で周期的に、NMS とホスト間で行われる。ネットワーク共通資源情報収集の方法は、NMS が自動的に発する情報収集コマンドによりホストから情報を収集する形態(図 8 の a)とホスト側から情報収集コマンドを発して、情報を NMS に転送し、NMS が記録する形態(図 8 の b)の 2 種類がある。

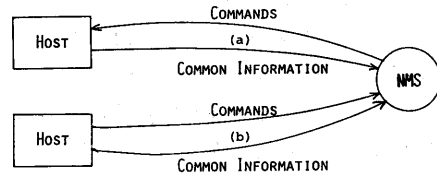


Fig.8 INFORMATION COLLECTING FLOW

#### 4.3. 資源情報の提供方法

NMS が管理する情報を TIP またはホストに提供する方法を述べる。

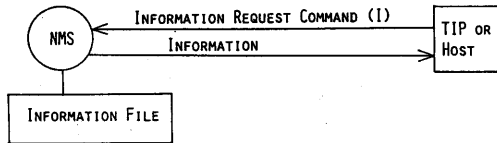
##### (1) 共通資源情報の提供

TIP またはホストは情報提供要求コマンド

1. リングバス構成情報 (バイパス, 折返し状況, etc.)
2. サブシステムの構成 (機種, OSタイプ, 特殊周辺機器, etc.)
3. サブシステムのもつ利用可能なソフトウェア (言語, アプリケーション, データベース, etc.)
4. センタニュース情報 (ネットワーク運営, 管理, 日程スケジュール, etc.)
5. サブシステムの負荷状況 (サブシステム毎のジョブ数, TSSログオン件数, etc.)
6. GAMMA-NETシステム仕様のドキュメンテーション
7. ネットワークコマンドの機能と利用手引き

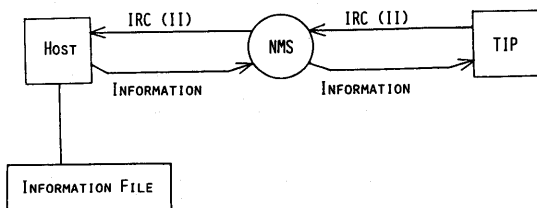
Table 3 Network Common Information

(I) NMSに送信すると、NMSの所有する記録ファイルから必要な情報を読み出し提供する。情報の形式、内容がネットワークで一意に定まっているので、TIPはかりでなく、この規約に従う他のホストも情報を解釈することが出来る。



## (2) ローカル資源情報の提供

NMSはTIPから情報提供要求コマンド(II)を受信すると、NMSはホストとの接続を始め、受信コマンドを送信する。最終的に情報提供要求コマンド(II)を受け付けたホストはそれ自身のローカル資源情報をNMS経由でTIPに提供する。即ちNMSは情報提供要求コマンド(II)を解釈し、TIPに代って目的のホストにアクセスし、必要な情報を入力する。従って情報の形式および内容はホスト固有の形態を有しているのでTIP以外のサブシステムは情報を解釈するのは困難である。下図に情報提供の基本的制御フローを示す。



IRC : INFORMATION REQUEST COMMAND

ローカル資源情報の提供例を具体的に述べる。ネットワークユーザが以前に

作成したファイルがどのホストに蓄積されているかを知る場合がこの例に当る。ネットワークユーザはネットワークコマンドの一つであるカタログリストコマンドを入力すれば、このユーザがもつ各ホストのファイルカタログを端末に出力することが出来る。即ちカタログリストコマンドを受信したNMSは次々とホストを起動し、ネットワークユーザのファイルカタログをTIPに転送する。

## 4.4. 負荷分散機能

NMSはTIPから送られたホスト選択要求コマンドに対して最適負荷分散となるホストを選択して、そのホスト名をTIPに返送する。バッチとRJEに対しては実行待ち時間とCPU速度(MIPS)の値が最小のもの、TSS利用に対してはCPU速度(MIPS)とCPU負荷率が大きいホストを選択する方法を取っている。しかしホストの選択方法は種々の提案があり、今後のシステムの測定・評価により選択方法の最適化を計ってゆくつもりである。

## 5. まとめ

高速光ファイバ伝送路による密結合した計算機ネットワークシステムを構築するためのプロトコル体系を記述し、各層に実装されたプロトコルの持つ特徴を中心に論じている。以下本システムの主眼点をまとめる。

- (1) 本システムは実験システムではなく、実用化を目的とした異機種計算機システムである点が特徴である。
- (2) リンクバスの高速性を十分に生かすためGAMMA-NET専用の身軽なプロトコルを考慮している。さらにDLPレベルのフォーマット化を行い、高速処理を計る。GAMMA-NET専用のため、標準化を指向した他のネットワークプロトコルとの互換性はない。
- (3) 従来のフロントエンドプロセッサによる網構成はプロセッサの性能からデータ処理の隘路となりかねないが、本システムは実効転送能力の向上のためチャネルイン

ターフェースによるリングバス直結方式を採用している。NCP以上のプロトコルは、ホスト内蔵の標準アクセスソケットに組み込まれる。このための改造に延べ約2人年程度が必要であった。

(4) プロセス間のメッセージ送受信に対して送受信バッチ制御ばかりでなく送受信プロセスの送受信タイミングも考慮したデータフロー制御を導入し、過負荷トラフィック時の突効チャネル効率の向上を計る。

(5) NMSを支援するネットワーク管理プロセスをALPレベルの1つのプロセスと位置付け、システムのRAS向上を計る受動的ネットワーク管理方式を提案する。さらにNMSに負荷分散機能を加え、ネットワーク全体の負荷の最適化を計る。

今後の問題点として幾つかの点を指摘しておく。次の段階で100Mbpsの光サブリングによる超高速システムの計画があるが、本システムのプロトコル設計には、最初から超高速通信に对应できるような考慮した上で、プロトコル体系そのものの大幅な変更はないものと思われる。

セッションファイル、インボックスやキーワードのファイル共有化を実現するファイルサブシステムおよびLP、CRやMTなどのサブ装置の共有化を実現するバルクサブシステムのためのプロトコル標準化が残された問題である。

次にプログラムやデータの互換性について述べる。本システムにおいては、計算機の内部表現に關係しない高級言語で記述されたプログラムの携帯性(ポータビリティ)は十分にある。ただし数値データや論理データの内部表現、数値計算の精度に影響するワード長または計算機特有の機能に対してはプログラムの互換性が極めて乏しい。将来内部表現やワード長の問題に関しては、データ変換機

作をプロトコルに内包する必要がある。

最後にNMSのもつ統計データ収集機能を活用して、システムの測定および性能評価を行ってゆくつもりである。これらの測定結果を踏えて、本システムに採用されたプロトコルおよびその実装方式の妥当性を検討してゆく。

おわりに

本システムの開発は筑波大学と(株)三菱電機との協同研究によるものである。関係者各位に深く謝意を表する次第である。特に電子情報工學系坂口環助教授、小川晴彦、中村奉夫講師から有益な助言をいただいたこと深く感謝致します。

#### 参考文献

- (1) M. V. Wilkes and D. J. Wheeler, The Cambridge digital communication ring, Proc. Local Area Communications Network Symposium, May, 1979.
- (2) K. Ikeda, et al, Computer Network Coupled by 100 Mbps Optical Fiber Ring Bus - System Planning and Ring Bus Subsystem Description -, COMPCON 80, 159-165, 1980.
- (3) H. Frank, et al, Computer Communication Network Design - Experience with Theory and Practice, AFIPS Conference Proc. 40, 255-270, SJCC, 1972.
- (4) L. Pouzin, CIGALE, The packet Switching Machine of the CYCLADES Computer Network, IFIP Congress, 155-159, 1974.
- (5) N. Abramson, The ALDHA system - Another Alternative for Computer Communications, AFIPS Conference Proc. 37, 281-285, FJCC, 1970.
- (6) R. Metcalfe and D. Boggs, Ethernet: Distributed Packet Switching for Local Communications Networks, CACM, Vol. 19, 395, 1976.
- (7) P. Zafirovulo, Performance Evaluation of Reliability Improvement Techniques for Single-Loop, Communication System, IEEE Trans. on Comm. COM-22, 742-751, 1974.
- (8) D. Walden, A System for Network, Comm. of the ACM 15, 221-230, 1972.