

高速光ファイバープ用データリンクプロトコルの検討

魚住 栄市 星子 隆幸 小柳津 育郎

(日本電信電話公社 横須賀電気通信研究所)

1. はじめに

ループ状ネットワークにおける伝送制御手順の標準仕様として、現在、IEEE 802 トークンリングがある。

複数プロセッサ間を高速光ファイバを用いて接続する大規模な複合システムに、前記 IEEE 802 トークンリングを適用した場合、優先処理のため通信スループットが低下することが予想される。また、トークンおよび応答領域がフレームチェックシーケンスの対象外となっているため信頼度の低下も予想される。

本稿では、多重通信機能および待ち合わせ機能を具備しているプロセッサ間結合装置 (PCU) におけるデータリンクレベルプロトコルについて、前記問題点への対処方法、ならびに、多重通信への適合性、高速光ファイバ通信への適合性、高信頼度化の観点から検討を進める。

2. 検討の前提

以下に本検討の前提、および、その設定理由を示す。

- ①共通予備方式によるシステムの経済化、高信頼度化をねらうため N対Nの対等通信が可能であること。
- ②サブチャネル多重通信ではバスの張られたサブチャネル間で END TO ENDの確認を行うため、データリンクレベルでの応答確認は冗長となる。そこでソフトウェアオーバーヘッドにならない範囲でデータリンクレベルプロトコルの 単純化を図り、また、タスク間での直接転送によるソフトウェアオーバーヘッドの削減、処理時間の短縮を図るための、サブチャネル多重通信に適していること。
- ③機能追加、トラヒックの増大に伴う段階的なシステムの拡大を図るため、システムのライフサイクルを長くできること。

④先端技術の導入による高速性、経済性、拡張性をねらうため、最近の 光ファイバ技術、LSI技術が活用できること。

⑤将来のサービスに充分対応できるためには、イメージ情報やファイル情報等の大量のデータをリアルタイムに参照、更新する必要があるため、高速で効率の良い転送方式が必要となるため、イメージ情報等の高速、大容量の転送が可能であること。

⑥遅延時間が大きいと、プロセッサ間の排他制御時間ネットワークとなりシステムスループットが伸びない。また、大量のデータをPCUに一旦バッファリングしてから転送する場合、イメージ情報やファイル情報の転送遅延時間が大きくなるため、遅延時間の短い方式であること。

⑦データ長を制限すると、ソフトウェアによるフレーム分割処理が必要となりオーバーヘッドとなるため、転送データ長に制限の無い方式であること。

⑧ 経済的であること。

3. ネットワークトポロジー

完全群接続の実現方法として、伝送路 (メディア) を共用して バス接続、又は、ループ接続する方法と、個々にバスを張る スター接続方法があるが、ここでは以下の理由からループ接続について検討を進めることとする。

- ①拡張、及び、障害時において、バイパス、ループバックによるバックアップが可能であり長時間運転、ノーダウン増設が可能である。
- ②複数のプロセッサ対で同時に通信ができる。
- ③ケーブル使用量が少なく、ケーブル費、工事費が削減できる。
- ④総延長を長くでき適用領域が広い。
- ⑤光ファイバを用いた高速通信に向いている。

ループ接続を他接続方式との比較でまとめたものを表1に示す。

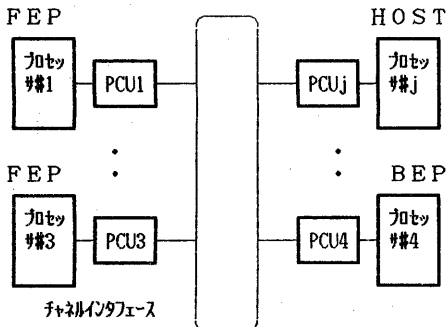
また、ループシステムの構成例を図1に示す。

表1 接続形態の比較(光通信を前提)

項目	ループ	バス	スター
トポロジー			
転送方向ノドの性格	一方向能動	両方向受動	両方向受動
評価項目 光への適合性 信頼度 #2 拡張性 スループット カバー領域 中継再生	◎ 高大 大 大 広い 要	△ 中中 中 中 狭い 不要	○ 低小 中 中 狭い 不要
総合評価	◎	△	○

#1...ノード

#2...二重化を前提



FEP: Front End Processor
BEP: Back End Processor

図1 ループシステム構成例

4. アクセス方式

前記、ループ接続において伝送路を各ノードが共用してアクセスする場合、そのアクセス権の取り方が問題となる。一般にアクセス方式の違いはネットワークトポロジーの相違により発生する。アクセス方式は図2のように分類される。(参考文献1)

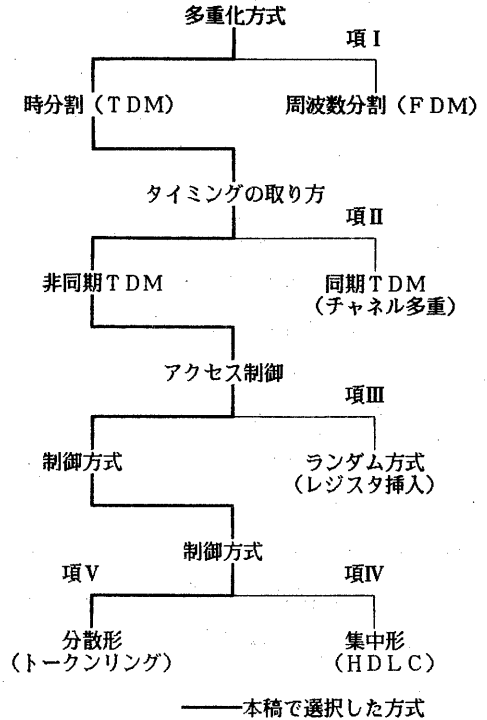


図2 アクセス方式の比較

本稿では、以下の理由により[項V]の分散方式について議論を進めることとする。

項IのFDM方式は、①高速の光リンクに適用しようとすると高価な光分波器および光送受信器を多数必要としコスト高となる。また、②PCU追加の度に波長追加が必要となり、ループ接続のメリットである拡張性も生かせない。

項IIの同期TDMは1チャンネル当たりの転送速度が低くイメージ情報やファイル転送などの高速・大容量転送に向かない。また、遅延時間も増大するため系間の排他制御にも使えない。

項IIIのランダムアクセス制御方式もレジスタ挿入のため遅延時間が増加し項IIと同様の理由から系間排他制御には使用できない。また、ハードウェア量も多い。

項IVの集中制御方式は集中局が信頼度ネックとなること、および、高速伝送路の場合ループのポーリング時間ネックでスループットが伸びない問題がある。

一方、項Vの分散制御方式は、項I~項IVの問題がなく高速、高信頼度で拡張性に優れており高速光ループに適し

た方式と言える。

5. レイヤ分割

N対N対等通信をサポートする方法として、①HDLC手順をN対Nに拡張する方法と、IEEE 802委員会を中心にほぼ固まっているLAN標準化仕様を適用する方法の2案が考えられる。

また、プロトコルは表2に示すように上位レイヤ、データリンクレイヤ、物理レイヤ、媒体に分かれ、さらに、データリンクレイヤを論理リンク制御サブレイヤ(LLC)とメディアアクセス制御サブレイヤ(MAC)に細分される。複数サブチャンネル間リンク制御(SLC)とデータリンク(DLC)をこれらのレイヤにどのように位置付けるかにより以下の4通りの案が考えられる。

【案A】DLCはIEEE802のLAN標準仕様をそのまま適用し、SLCをDLCのすぐ上に位置付ける。

【案B】MACのみIEEEのLAN標準仕様を適用し、LLCはSCLと一体とする。

【案C】MACをPCU独自仕様としLLCは案Bと同様SLCと一体化する。

【案D】DLCはPCU独自仕様とし、SLCはDLCのすぐ上位に位置付ける。

これらの案の関係を表3に示す。

前記4案の内、以下の理由から【案D】について検討を進めることとする。

①IEEEのLAN標準仕様では後述の通り100Mbps程度の高速領域でケーブル遅延時間が性能限界になるため標準化の対象としていない。(A, B)

②IEEEのMACは信頼性、性能、ハードウェア量の点で問題がある。(A, B)

③IEEEのLLCは通信回線をN対Nに拡張したものでありチャンネル結合をN対Nにするには問題がある。(A, B, C)

④案DはSLCを改造することなくDLCの改造が可能であり仕様の簡明さ、拡張性、柔軟性の点で優れている。

表2 各レイヤの機能

レイヤ	機能	
上位	—	
データリンク	論理リンク制御サブレイヤ(LLC)	宛先アドレス、送信元アドレスの付加、および、チェック
	メディアアクセス制御サブレイヤ(MAC)	共通リンクへのデータパケットの送受信を制御する。FCS付加、チェック、応答の付加、チェック
物理	デジタル信号の光電変換、符号化、復号化、デリミタの付加/検出	
媒体	—	

表3 各レイヤの実現案

レイヤ	案 A	案 B	案 C	案 D	
上位	—	—	—	—	
DLC	LLC	IEEE-LLC	SLC'	SLC''	PCU-DLC
	MAC	IEEE-MAC	IEEE-MAC	PCU-MAC	

6. データリンクレベルプロトコル

6.1 フレーム構成

ループ上へ送受信する情報は夫別して送信権(トークン)、情報、応答、ダミーに分けられる。これらの情報をフレームに構成して転送する方法と、IEEE802のようにフレーム外ビットで実現する方法があるが、以下の理由から、送信権、情報、応答はフレーム構成とし、ダミーは特殊コードで実現する方法が良いと考える。

IEEE802トークンリングのようにトークンおよび応答をフレーム外ビットで構成した場合、これらの誤りが検出できず、信頼度上問題となる。

また、フレーム外のビット操作が多くなり高速処理用ハードウェア量が増加して、高価なものとなる。

これに比べ、送信権や応答をフレーム構成とすることによりFCSの対象となり高信頼度化、回路の共通化による単純化が可能になるものと考えられる。

また、ダミーについてはフレーム間を埋めるものであり

IEEE802トークンリングと同様、特殊パターン(タイムフィル)で実現すれば良いと考える。

フレーム内の情報の長さについては情報処理で扱う情報の長さがバイトオリエンテッド(オクテット)なので、これに合わせるのが、ハードウェア量、及び、符号方式の観点からも得策と考える。

以上からフレーム構成の一例を図3に示す。

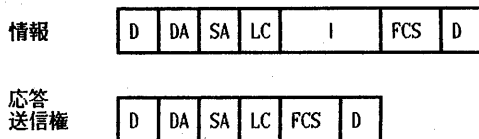


図3 フレーム構成の一例

6.2 送信権管理方式

分散制御方式では制御局が無いため送信権(トークン)はインプリシットな方式がとられる。

IEEE802トークンリングでは情報の緊急度に応じて優先処理を可能としている。優先処理は送信フレームがループを一巡する間に各ノードの送信要求の中から次に一番高い優先レベルを捜し、これに送信権を与えることにより行われる。このため送信権を獲得した局は送信フレームが一巡するまで送信権を解放できない。ループ長が数Km~100Kmで伝送速度が100Mbps程度以上の高速の領域にこのような方式を適用するとケーブル長ネックとなり性能が上がらず優先処理の意味が無くなるという問題が有る。

これらの問題は優先処理機能を削除し、フレーム送信後直ちに送信権を解放(送信権早期解放)することにより解決できる。

この方式によるスループットの向上効果を以下に示す。算出式の詳細については参考文献2に示す。

(1) ループ長の影響

ループ長をパラメータにしてスループットとメッセージ通過時間との関係を図4に示す。これより、ループ長とスループット比(トークン早期解放方式/IEEE802)を求め図5に示す。図5より、ループ長が1Km以下では

両方式に差は無く、10Kmで3倍、100Kmで15倍改善される。これは、トークン早期解放方式ではループ長のスループットへの影響が少ないことを示している。

(2) データ長の影響

データ長(転送単位)をパラメータとしてスループットとメッセージ通過時間との関係を図6に示す。これより、これより、転送単位とスループット比を求め図7に示す。図7より、ループ長1Km、データ長数KB以上では、両方式に差のないことがわかる。ループ長100Kmでは数倍から数十倍改善される。これは、IEEEがループ長ネットワークとなっているためと考える。

(3) ループ接続ノード数の影響

ループ接続ノード数とスループット比の関係を図8に示す。これより、ノード数10以下ではほとんど差が無いが、ノード数60以上で数倍改善される。

(4) ループ伝送速度の影響

ループ伝送速度とスループット比との関係を図9に示す。これより、ループ長がスループット比を決める主要因であることがわかる。

以上の検討より、ループ長が短く、データ長の長い領域ではトークン早期解放方式とIEEE802トークンリング方式に差はなく、ループ長が長くデータ長の短い領域でトークン早期解放方式が有利となる。

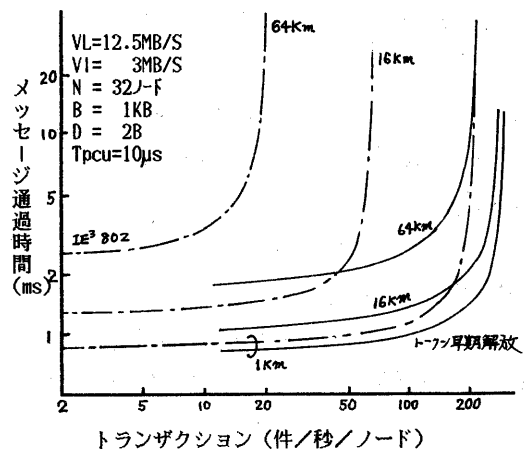


図4 スループットとメッセージ通過時間の関係

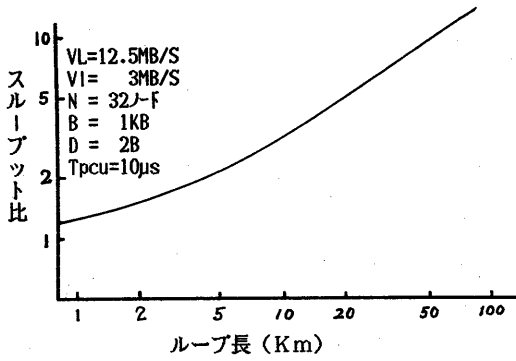


図5 ループ長とスルーパット比の関係

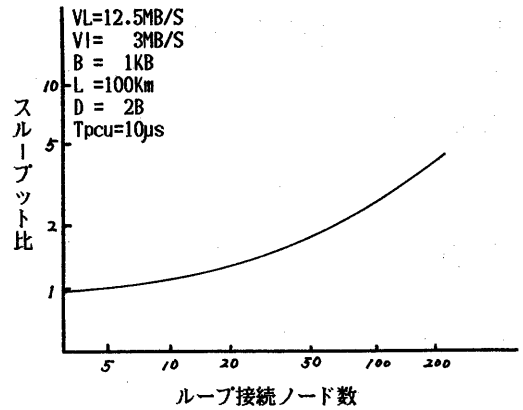


図8 ループ接続ノード数とスルーパット比の関係

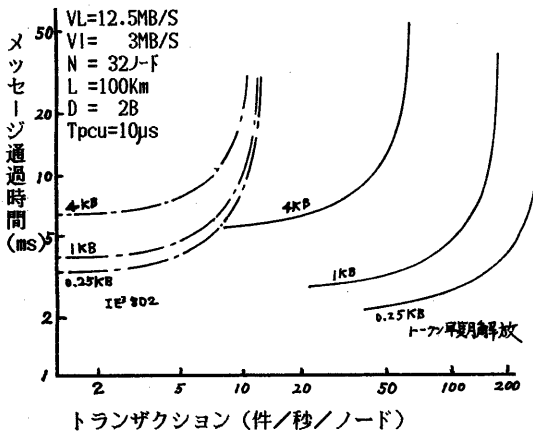


図6 転送単位とメッセージ通過時間の関係

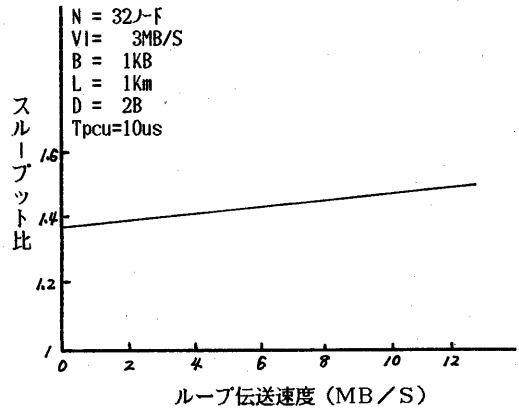


図9 ループ伝送速度とスルーパット比の関係

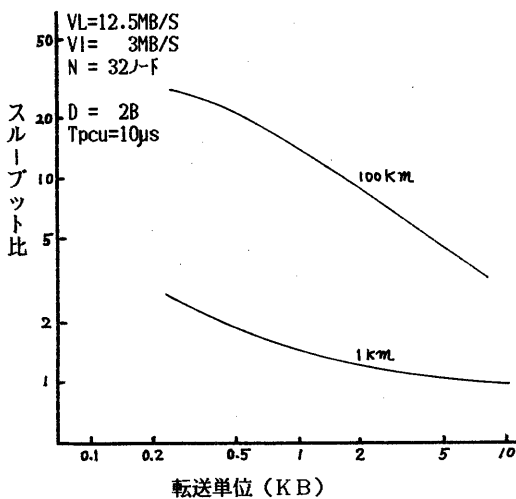


図7 転送単位とスルーパット比の関係

また、送信権を獲得した場合、連続して何フレームまで送信を可能とするかが問題となる。これに対しては、次の理由から1フレーム送信毎に送信権を解放する方式が良いと考える。

①一般に、ループ伝送路に比べて上位プロセッサとのインタフェースの転送速度が遅いため、これがネックとなって送信権を保留するのは、高速伝送路の仕様効率を下げることになり問題である。

②後述の再送をサポートするためには、ビジーまたは無応答に対する応答が返るまで送信フレームを保留する必要がありこのハードウェア量が少なくなる。

非同期TDMで分散制御方式の場合、ループには1つのトークンのみ存在する(シングルトークン)。この場合、

トークンの喪失や二重トークンの問題が発生する。トークンの喪失については、時間監視によるトークンの回復が一般的であり、二重トークンについては自然消滅するものであり、論理矛盾は上位のSLCで検出できる。

6.3 応答方式

送信フレームに対する応答は、正常受信応答、バッファビジー応答、無応答の3通りに大別できる。

ビジー、無応答に対しては高信頼度化、ソフトウェアオーバーヘッドの削減等の観点からDLCLレベルでの再送も必要最小限サポートする必要がある。これは、送信フレームに対する応答が返るまで次の処理を待たせ再送に備えることにより実現できる。

また、前記6.1節フレーム構成で述べたように応答もフレーム構成とした場合、応答契機が問題となる。通常のフレームと同じように送信権を獲得して応答すると、応答が返るまでの時間が特定できず、前記、無応答検出のための時間監視が困難となる。この対処方法としては、送信フレーム送出直後に応答フレームを挿入できるスペースを空けて送信権を解放することにより、応答が返るまでの時間をループ一巡時間に固定でき時間監視が容易となる。

IEEE802では、各フレームに応答インジケータを付加して対処しているが高速伝送路に同期してこのインジケータをON/OFFするのは回路的にも大変である。

6.4 分割転送方式

PCU-プロセッサ間の転送速度はPCU-ループ伝送路間の伝送速度に比べて遅いためPCUで一旦バッファリングしてからループに送出することとなる。このため、プロセッサからの転送データを一フレームでループに送出(一括転送方式)しようとするPCUに最大長のデータバッファを持つ必要があり、また、PCU-プロセッサ間の転送とPCU-ループ間の転送がシリアルとなりメッセージ通過時間が長なること、および、メッセージ長が制限される問題がある。

この対処方法としては、PCUでメッセージを短いフレ

ームに分割し、PCU-プロセッサ間とPCU-ループ間の転送を並列動作させることが考えられ、メッセージの分割長をあまり短くするとPCUの制御オーバーヘッドの増大、ループ使用効率の低下が問題となる。従って、最適なデータ長に区切る必要がある。

分割数をパラメータとしてトランザクションとメッセージ通過時間との関係を図10および図11に示す。(算出式の詳細は付録に示す。)

これより、転送単位4KB、ケーブル長1Kmの場合、トランザクションの少ないところでは、チャンネルネック、多いところではループ伝送路ネックとなることが、また、ケーブル長が10Kmになると、逆に、トランザクションの少ないところでループネックに、多いところでチャンネルネックとなることが分かる。

いずれの場合も、分割数2で約30%~40%、分割数4で約60%~70%メッセージ通過時間が短縮されることが分かる。

また、分割転送を行う場合、データの欠送、冗送の問題が発生する。これについてはシーケンス番号で対処可能である。(シーケンス番号の付与方法として全フレームに付与方法と、データ転送フレームにのみ付与方法がある。前者の場合、サブチャンネル毎にシーケンス番号を管理する必要があり、しかもDLCLとSLCLの切れが悪くなるので後者が優れていると考える。)

シーケンス番号を付加した場合、アウトスタンディング転送するかどうか問題となるが、伝送路速度に比べてプロセッサ間との転送速度が遅いためアウトスタンディングのデータの準備が出来ないこと、および、制御が複雑となり、バッファ量も多くなり得策ではない。

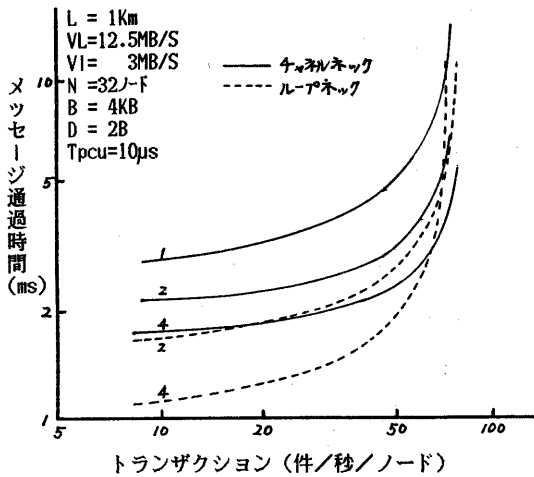


図10 分割数とメッセージ通過時間 (1Km)

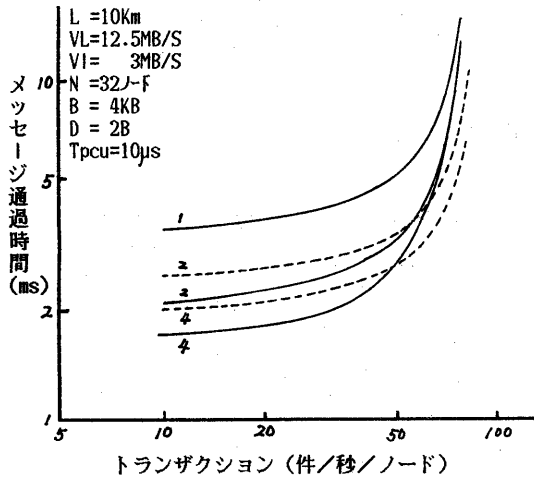


図11 分割数とメッセージ通過時間 (10Km)

6.5 DLCの提案

以上の検討結果より、PCUに適したDLCを提案し、IEEE802データリンクとの比較で表4に示す。

表4 PCUに適したDLCの概要

項目	PCU	IEEE802																				
フレーム形式	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>D</td><td>DA</td><td>SA</td><td>LC</td></tr> <tr><td>I</td><td>FCS</td><td>D</td><td></td></tr> </table>	D	DA	SA	LC	I	FCS	D		<table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>DS</td><td>PFCS</td><td>DA</td><td>SA</td></tr> <tr><td>I</td><td>FCS</td><td>DE</td><td></td></tr> <tr><td colspan="4" style="text-align: center;">PFCE</td></tr> </table>	DS	PFCS	DA	SA	I	FCS	DE		PFCE			
D	DA	SA	LC																			
I	FCS	D																				
DS	PFCS	DA	SA																			
I	FCS	DE																				
PFCE																						
送信権	フレーム形式	D S内に表示 (フリーラン/ビートン)																				
優先処理	無し	PFCS内に表示																				
送信権解放契機	フレーム送信後	送信フレーム受信後																				
応答	フレーム形式 応答領域確保	フレームに 応答インジケータ付加																				
DA	1バイト	6バイト																				
SA	1バイト	6バイト																				
FCS 範囲	DA~FCS	PFCSの央~FCS																				
FCS	2バイト	4バイト																				

8. おわりに

光ファイバーループに接続される任意高速プロセッサ間の通信を高効率に実現するデータリンクプロトコルの方式条件を明らかにした。ここでは高速プロセッサ接続に的を絞って検討を進めてきたが、同一ループにミニコン等の比較的低速のプロセッサをどう効率良く接続して行くかが今後の課題である。

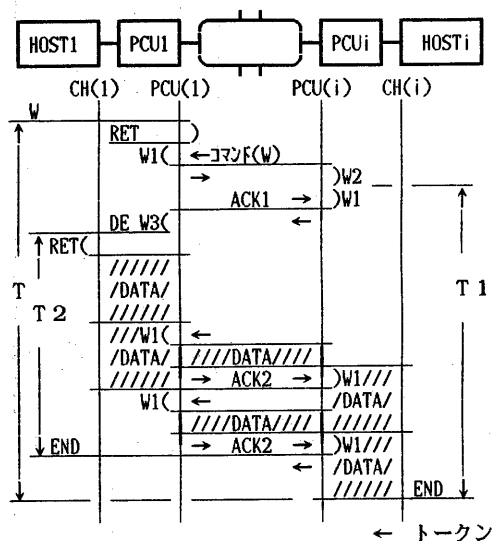
最後に、本検討にあたり有意義なご指導ご助言をいただいた関係各位に深謝します。

参考文献

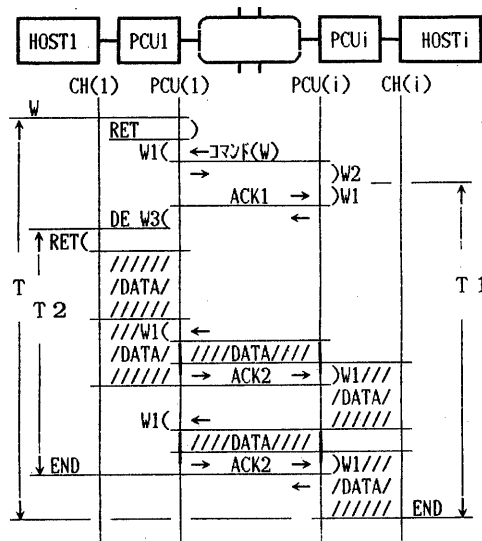
- (1) R.C.DIXON "RING NETWORK TOPOROGY FOR LOGICAL DATA COMMUNICATION" COMPCON 25th 591-605 '82
- (2) 魚住 "トークンリングの送信権制御法に関する一検討" 情報27回全大 3J-1
- (3) 橋田、中村 "多重待ち行列の解析 [I] - 全処理式 - " 研実報、V.19, No.6, 1970

付録 分割転送時の平均メッセージ通過時間

分割転送における分割数と平均メッセージ通過時間の関係の算出モデルを付図1および付図2に、算出式を付表1にします。



付図1 分割転送（チャネルネック）



付図2 分割転送（ループネック）

付表1 平均メッセージ通過時間算出式（P：分割数） 単位 時間：μs データ長：バイト

項番	項目	記号	チャネルネック	ループネック	記事
1	平均メッセージ通過時間	T	$T1+Ts+TL/2+W1+W2+Tpcu$	同左	a:件/秒/ノード N:ノード数 L:ループ長(Km) C:5 μs/Km D:遅延バイト数/ノード(2B) b=a·h
2	ACK 1 送出周期	T1	$(P+1) \cdot Tsp+Td+Ts+TL+2W1+W3+2Tpcu$	$2Tcpt+(Td+Ts+TL+2W1) \cdot P+W3+2Tpcu$	
3	CHとのブロック転送時間	T2	$P \cdot Tsp+Td+Ts+TL+2W1+2Tpcu$	$Tcpt+(Td+Ts+TL+2W1) \cdot P+2Tpcu$	
4	トークン待ち (参考文献3)	W1	$U+(N-1)U/2+(1-N \cdot b)+(2+2P) \cdot N \cdot a+h/2(1-N \cdot b)$	同左	
5	ACK 1 送出待ち	W2	$a \cdot T1 \cdot T1 / (1 - a \cdot T1)$	同左	
6	CH再結合待ち	W3	$a \cdot T2 \cdot T2 / (1 - a \cdot T2)$	同左	
7	平均フレーム送出時間	h	$(Ts+Ts+(Td+Ts) \cdot P) / (2+2P)$	同左	
8	hの2乗平均	h2	$(T_s^2+T_s^2+(T_d+T_s) \cdot P) / (2+2P)$	同左	
9	平均歩行時間	U	TL/N	同左	
10	ループ一周時間	TL	$C \cdot L+D \cdot N/VL$	同左	
11	ショートフレーム長	A	4+36	同左	
12	ショートフレーム送出時間	Ts	A/VL	同左	
13	フレームループ側転送時間	Td	$(B/P+36)/VL$	同左	
14	フレームCH側転送時間	Tsp	$B/P/Vi$	同左	
15	PCU内処理時間	Tpcu	10	同左	