

複数チャネルを用いた最適データ転送方式の提案

平田 俊明 近藤 毅 宮崎 聡
(株)日立製作所 システム開発研究所

2つの通信ノード間で複数のリンクを用いてデータ転送を行なう方式には、複数のリンクに順次データ単位を割り当てるマルチリンク方式と1つのデータ単位を分割し、これを各々のリンクに割り当て、受信側でこれを組み立てるパラレルリンク方式がある。このうち、パラレルリンク方式では、データ転送の並列性と逐次的なソフトウェア処理のトレードオフにより転送時間を最小にするデータ単位の分割数、すなわち、リンク上の転送データ長が決まる。本報告では、パラレルリンク方式において、データ転送時間を最小化するデータ転送方式を提案し、これを典型的なモデル上で定量的に評価する。

A Study of optimal data transfer method using plural channels

Toshiaki HIRATA Takeshi KONDOU Satoshi MIYAZAKI
Systems Development Laboratory, Hitachi, Ltd.

There are multi-link method and parallel-link method which transfer data between two nodes using plural links. The former is the method that sending node assigns data to plural links in order. The latter is the method that sending node divides data into several transfer units and assigns each unit to plural links and assemble them at receiving node. In parallel-link method, data division number that minimizes data transfer time exists because of trade-off between data transfer time on link and data processing time on CPU. This paper proposes optimal data transfer method with respect to parallel-link method and evaluate the method using typical data transfer model.

1.はじめに

2つの通信ノード間のデータ転送性能を向上させる方法の1つに、通信ノード間のデータ転送を複数のリンクを用いて行なう方法がある。複数のリンクを用いたデータ転送方式には、複数のリンクに順次データ単位を割り当てる方式^{1)、3)}や、1つのデータ単位を分割し、これを各々のリンクに割り当て、受信側でこれを組み立てる方式³⁾がある。前者の方式をマルチリンク方式、後者の方式をパラレルリンク方式とよぶ。マルチリンク方式は、データを連続的に転送する場合、すなわち、各種のアプリケーションでリンクを多重化して使用するようなスループット性能を重視する場合に有効な方法である。マルチリンク方式は、OSI (Open Systems Interconnection)第2層に位置付けられるマルチリンク手順として規定されており¹⁾、待ち行列の構成やデータの分配アルゴリズムについて論理解析やシミュレーションを用いた評価が多数行なわれている^{4)、5)、6)、7)、8)}。一方、パラレルリンク方式は、個々のデータの転送時間、すなわち、応答時間性能に注目した場合有利であり、Telnet等の応答時間を重視するアプリケーションが専用にリンクを使用する場合に有効である。また、パラレルリンク方式は、マルチリンク方式に比べて以下の点で有利である³⁾。

(1) パラレルリンク方式では、データ単位を送信側と受信側で正しい順序で受け渡せる。

(2) 1つのリンク上の転送データ長に制限がある場合でもパラレルリンク方式では、受信処理が短時間に行える。

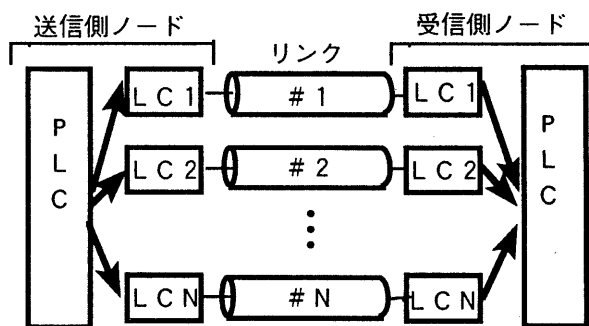
上記(1)は自明である。上記(2)は次の理由による。マルチリンク方式では、あるリンクに対して1つのデータ単位を分割して順次送信し、受信側で組み立てる必要がある。このとき、送信データ長が同一の場合、受信側では、最初の分割データを受け取ってから、ある一定の間データの組立を行なうことができない。そして、しばらくした後、到着したデータを複数のリンク上で一斉に組み立てなければならず、通信ノードのCPUが一時的に混雑する。それに対して、パラレルリ

ンク方式では、同一サイズのデータを送信した場合でも、一定時間間隔でデータを受信することができるため、一時的に通信ノードのCPUが混雑することを防げる。

本報告では、複数のリンクを用いたデータ転送方式として、パラレルリンク方式を対象とする。通信ノード間のデータ転送時間は、通信ノード上のソフトウェア処理時間とリンク上の転送時間の和になる。ソフトウェア処理は、逐次処理であり、リンク上の転送処理は、複数リンクによる並列処理が可能である。データ転送を複数のリンクに分散して行なう場合、データ転送の並列性とソフトウェア処理(データの分割数を多くするほどオーバーヘッドが大きくなる)のトレードオフにより転送時間を最小とするデータ単位の分割数、すなわち、リンク上の転送データ長が決まる。従来のパラレルリンク方式では、1つのデータ単位をリンクの数に分割し、これを各々のリンクに割り当てる方式であり³⁾、データ転送時間の最小化をはかるための考慮がなされていない。本報告では、複数のリンクを用いたパラレルリンク方式において、データ転送時間を最小化するデータ転送方式を提案し、これを典型的な通信ネットワークのモデル上で定量的に評価する。提案方式は、使用できるリンク数を有限とし、1つのリンク上の転送データ長の上限を考慮したものとする。また、リンクのデータ転送速度は、全て同一の場合と、速度が異なる場合の両方について、データ転送方式を提案する。以下、2章では本報告で対象とするパラレルリンク方式を示し、3章で最適データ転送方式を提案する。4章で提案方式の定量的な評価を行う。

2.パラレルリンク方式

パラレルリンク方式では、図1に示すように2つの通信ノード間をN本のリンクで結び、これらのリンクによりデータを転送する。LC(Link Controller)は、各リンク上のデータ転送制御を行なう。PLC(Parallel Link Controller)は、図2に示すように上位レイヤから送信要求のあった1つ



LC : Link Controller
PLC : Parallel Link Controller

図1 システム構成

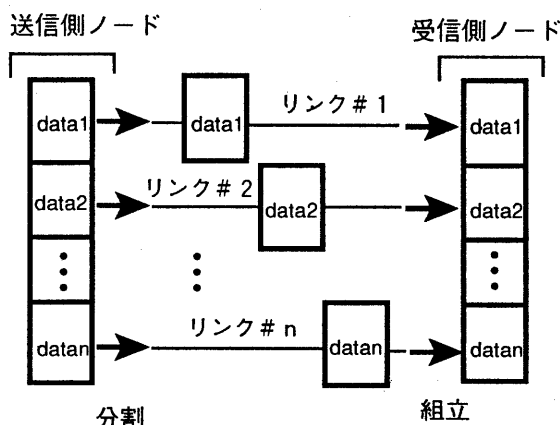


図2 パラレルリンク方式

のデータ単位を分割し、これを配下のLC（リンク）に分配し、受信側でこれを組み立て1つのデータ単位として上位レイヤに渡す。PLCでは、上位からのデータ単位を要求順に転送し、送受信ノード間でデータ単位の順序性が保証されるようにする。すなわち、1つのデータ単位の転送でN本のリンクを占有することになる。各リンク上で転送可能なデータ長には上限があるものとするが、PLCは上位レイヤに対して、転送可能なデータ長の制限を設けない。すなわち、PLCは上位レイヤに対して、最大データ長の制約のない1本の仮想的な高速リンクを提供する。

パラレルリンク方式では、受信側のデータ組立は、データ分割の際、順序番号の入ったヘッダと

最終データか否かを示すビット（X.25のMbit相当）を挿入することにより実現できる。この方式では、受信側では順番通りにデータを接続していき、最終データを受信した時点で、上位レイヤにデータ単位を渡せばよい。このほか、データ組立の方法として、送信側と受信側で次にデータを送受信すべきリンクをあらかじめ認識しておく、つまり、あらかじめ、リンクに順序番号をふる方法がある。この方式は、通信を開始するにあたって、各リンクのコネクション確立用のデータ送受信でリンクの番号を取り決めることによって実現できる。この方式では、送受信データ中に最終データか否かを示すビットは必要であるが、順序番号の入ったヘッダは不要である。

3.最適データ転送方式

3.1 従来方式の問題点

通信ノード間のデータ転送時間は、通信ノード上のソフトウェア処理時間とリンク上の転送時間の和になる。ここで、ソフトウェア処理時間とは、PLCのデータ分割、組立処理、データのリンクへの割当や、LCのプロトコル処理、バッファ確保/解放、リンクの起動処理等をいう。

ソフトウェア処理は、CPUは1つであるため、逐次処理であり、リンク上の転送処理は、複数リンクによる並列処理が可能である。

パラレルリンク方式では、データ転送の並列性から、分割数を増やすほどリンク上の転送時間は短くなるが、データ送受信に必要なソフトウェア処理のオーバーヘッドは多くなる。このため、データ転送の並列性とソフトウェア処理のトレードオフにより転送時間を最小とするデータ単位の分割数、すなわち、リンク上の転送データ長が決まる。

従来のパラレルリンク方式は、1つのデータ単位を単純にリンクの数に分割し、これを各々のリンクに割り当てる方式であり、データ転送時間の最小化を図るための考慮がされていない。

3.2 提案方式（その1）

N本のリンクのデータ転送速度がすべて同一の場合について、最適データ転送方式を示す。パレレルリンク方式では、3.1で示したように転送時間を最小にするデータ単位の分割数およびリンク上の転送データ長が存在する。L(KByte)を上位からの転送データ長、V(bps)を各リンクの転送速度、 T_p (msec)をソフトウェア処理時間、 T_h (msec)をリンク対応のハードウェア処理時間（データ長には非依存）、nをデータ分割数、 T_c (msec)を長さLのデータの転送時間、 l_{max} (KByte)をリンク上で転送可能な最大データ長とする。このとき、 T_c を最小にするデータ分割数nは問題1の解として与えられる。

問題1

$$\min_{n \in I} T_c = q (L / (V n) + T_h) + n T_p \quad (1.a)$$

$$\text{subj.to } 0 < L / n \leq l_{max} \quad (1.b)$$

ここで、qは n/N を切り上げた整数値、Iは自然数の集合である。 T_c は、ソフトウェア処理時間($n T_p$)、リンク対応のハードウェア処理時間($q T_h$)およびリンク上の転送処理時間($q (L / (V n))$)の和で表される。ハードウェア処理時間およびリンク上の転送処理時間は、 $n > N$ の場合、リンクをラップラウンドで使用するため、ラップラウンドの回数qを乗じてある。

問題1から具体的なデータ転送アルゴリズムを導く。まず、その前に、問題1において使用可能なリンク数Nおよび最大データ長 l_{max} を考慮しない問題2を考える。

問題2

$$\min_n T_c = L / (V n) + T_h + n T_p \quad (2)$$

問題2の解は、

$$dT_c / dn = -L / (V n^2) + T_p = 0 \quad (3)$$

を満たす。式(3)をnについて解くと、

$$n = \sqrt{L / (T_p V)} \quad (4)$$

となる。nは自然数であるので、式(4)で求め

た値を T_c が小さくなるほうに切り上げまたは切り捨てて。これを n' とする。これより最適なデータ分割長 L_o (KByte)は、

$$L_o = L / n' \quad (5)$$

となる。

次に、上記 L_o を用いて具体的な転送アルゴリズムを以下に示す。

(1) $L_o < l_{max}$ の場合

(a) $L \leq L_o \times N$

この場合、問題1は問題2に置き換えられる。従って、データ転送アルゴリズムは、図3に示すように、上位からの転送データを L_o に分割し、先頭データから順次、空きリンクに割り当てる方式となる。

(b) $L_o \times N < L \leq l_{max} \times N$

この場合、問題1は次の問題3に置き換えられる。

問題3

$$\min_n T_c = q (L / (V n) + T_h) + n T_p \quad (6)$$

ここで、 $q = n / N$ とすれば、

$$T_c = L / NV + n T_h / N + n T_p \quad (7)$$

となる。条件より、最適なデータ分割数はN以上であり、式(7)はnの正勾配の一次関数であるため、 $n = N$ が問題3の解である。従って、データ転送アルゴリズムは、図4に示すように、上位からの転送データをN分割し、先頭データから順次、空きリンクに割り当てる方式となる。

(c) 転送データ長 $> l_{max} \times N$

この場合、問題1の解は、制約条件式(1.b)の境界(右境界)に存在し、問題3で述べたように、 T_c はnの正勾配の一次関数であるため、 $n = L / l_{max}$ となる。従って、データ転送アルゴリズムは、図5に示すように、上位からの転送データを l_{max} に分割し、先頭データから順次、空きリンクに割り当て、データ分割数がNを越える分については転送が終了したリンクに順次割り当てる方式となる。

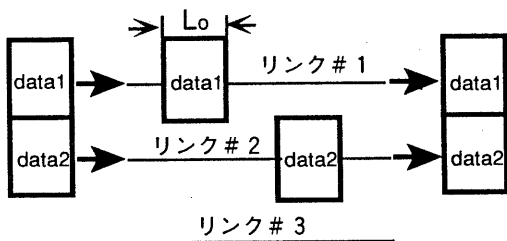


図3 $L < L_o \times N$ で $N=3$ の場合

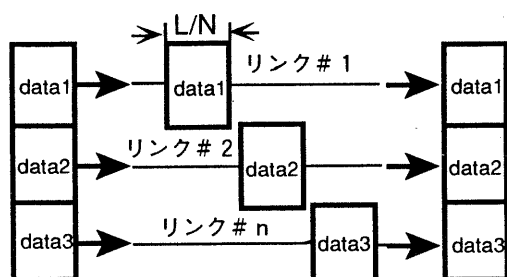


図4 $L_o \times N < L \leq l_{\max} \times N$ で $N=3$ の場合

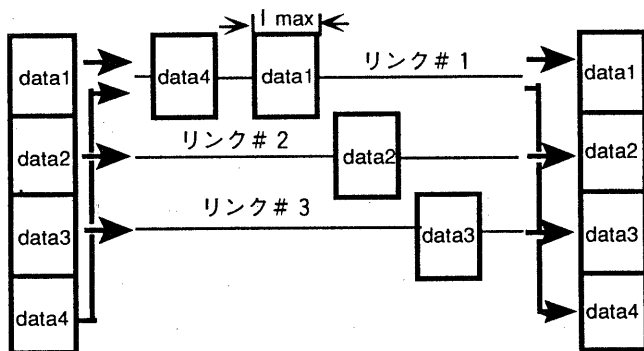


図5 $L > l_{\max} \times N$ で $N=3$ の場合

(2) $L_o \geq l_{\max}$ の場合

この場合も、問題1の解は制約条件式(1.b)の境界(右境界)に存在し、 T_c は n の正勾配の一次関数であるため、 $n = L / l_{\max}$ が解となる。従って、データ転送アルゴリズムは、上位からの転送データを $l_{\max}$ に分割し、先頭データから順次、空きリンクに割り当て、データ分割数が N を越える場合は、その分については転送が終了したリンクに順次割り当てる方式となる。

3.3 提案方式(その2)

N 本のリンクのデータ転送速度が異なる場合について、最適データ転送方式を示す。

$v = (v_1, \dots, v_N)$ (bps)を各リンクの転送速度とする。 $v_1, \dots, v_N$ は転送速度の順とする。 $l = (l_1, \dots, l_n)$ (KByte)を各リンクの転送データ長とする。 $l_1, \dots, l_n$ は、速度 v_1 のリンクから順に対応させ、 N を越える分は速度 v_1 のリンクから再度対応させる。このとき、長さ L のデータの転送時間 T_c を最小とする l, n は、問題4の解として与えられる。

問題4

$$\min_{n \in \mathbb{N}} q(f(l) + T_h) + nT_p \quad (8.a)$$

subj.to

$$f(l) = \min_i \max_{i \in \{1, \dots, n\}} (l_i / v_i) \quad (8.b)$$

subj.to

$$0 < l_i \leq l_{\max} \quad (i=1, \dots, n) \quad (8.c)$$

$$\sum_{i=1}^n l_i = L \quad (8.d)$$

$f(l)$ はリンク上の転送処理時間である。これは、制約条件(8.c)、(8.d)のもとで転送時間が最大のリンクの転送時間を最小とする l に対する目的関数値である。このことは、転送時間が最大のリンクの転送時間が全体のリンクの転送時間を決めることを意味する。

問題4は、 n を1から順に与え、それぞれの n に対して、式(8.b)

～(8.d)を解き、 T_c が最小となる n を求めればよい。式(8.b)～(8.d)は、問題5と等価である。

問題5

$$f(l) = \min_{l, \epsilon} \quad (9.a)$$

subj.to

$$l_i / v_i \leq \epsilon \quad (9.b)$$

$$0 < l_i \leq l_{\max} \quad (i=1, \dots, n) \quad (9.c)$$

$$\sum_{i=1}^n l_i = L \quad (9.d)$$

問題5は線形計画問題であり、容易に解くことができる。

以上より、データ転送アルゴリズムをまとめると以下ようになる。

(1) 問題4を解き、 T_c を最小とする l 、 n を求める。

(2) 上位からの転送データを先頭から長さ $l_1, \dots, l_n$ に分割し、転送速度 $v_1, \dots, v_N$ のリンクに順次割り当てる。データ分割数が N を越える分については対応するリンクのデータ転送が終了しだい、転送データを割り当てる。

本データ転送方式を実現する場合、データ転送ごとに問題4を解くことは処理オーバヘッドの点から好ましくない。そこで、 L に対する最適な n 、 l をあらかじめ計算し、その結果をテーブル化しておき、データ転送時にこのテーブルを検索する方法が考えられる。テーブル化すべき L の範囲は使用する上位プログラムから予測し、もし、テーブルに記載の範囲外のデータ長のデータを転送するときは、以下のような簡略化したアルゴリズムを用いればよい。

(a) 転送データ長 $\leq l_{\max} \times N$

上位からの転送データを n 分割し、先頭データから順次、空きリンクに割り当てる。

(b) 転送データ長 $> l_{\max} \times N$

上位からの転送データを $l_{\max}$ に分割し、先頭データから順次、空きリンクに割り当てる。データ分割数が N を越える分については転送が終了したリンクに順次割り当てる。

4. 転送時間の評価

本報告で提案したデータ転送方式の有効性を1つのデータ単位の転送時間に注目して定量的に評価する。

4.1 リンクの転送速度が同一の場合

(1) 前提条件

通信制御処理装置間を複数の通信回線で接続する通信ネットワークのモデルを考える。リンク数

$N=5$ 、リンクの転送速度 $V=1$ Mbps、リンク上で転送可能な最大データ長 $l_{\max}=4$ KByteとする。また、CPUの処理能力を1 MIPS、1つの分割データの転送に必要なソフトウェア処理を4 Kstep（通信制御処理装置では標準的な値）とすると、ソフトウェア処理時間 $T_p=4$ msecとなる。リンク対応のハードウェア処理時間は、ソフトウェア処理時間の約 $1/10$ とし、

$T_h=0.5$ msecとする。

(2) 計算結果

(1)の前提条件に対し、3.2で示したデータ転送アルゴリズムを適用した結果を表1および図6に示す。これらには、比較のため、従来のパラレルリンク方式（データ単位を単純にリンク数に分割する方式）でのデータ転送時間とシングルリンクでのデータ転送時間を併記する。

パラレルリンク方式のシングルリンクに対する効果は明らかである。提案方式は、従来方式に対して $L \leq L_0 \times N$ の範囲で有利である。

$L > L_0 \times N$ の範囲では、転送データは、 N 分割または $l_{\max}$ で分割されるため、提案方式と従来方式は同等になる。パラレルリンク方式が有効な応答時間を重視するアプリケーションでは、転送データ長は比較的短く、提案方式が有効である。

4.2 リンクのデータ転送速度が異なる場合

(1) 前提条件

リンクの転送速度を $v_1=1.5$ Mbps,
 $v_2=1.25$ Mbps, $v_3=1$ Mbps,
 $v_4=0.75$ Mbps, $v_5=0.5$ Mbpsとする。他の条件は、4.1と同様とする。

(2) 計算結果

(1)の前提条件に対して、3.3で示したデータ転送アルゴリズムを適用した結果を表2および図7に示す。これらには、比較のため、従来のパラレルリンク方式（データ単位を単純にリンク数に分割する方式）でのデータ転送時間とシングルリンクでのデータ転送時間（速度 v_1 のリンクを用いた場合）を併記する。提案方式は、従来方式、シングルリンクに対して、全ての場合で有利である。

表1 データ転送時間の評価（リンクの転送速度均等）

転送データ長	提案方式		従来方式の 転送時間(msec)	シングルリンクで の転送時間(msec)
	分割数	転送時間(msec)		
1 K B	1	1 2.3	2 2.1	1 2.3
2 K B	2	1 6.3	2 3.6	2 0.1
4 K B	3	2 2.9	2 6.7	3 5.7
8 K B	4	3 2.1	3 3.3	7 1.5
1 6 K B	5	4 5.5	4 5.5	1 4 3.0
3 2 K B	8	9 5.5	9 5.5	2 8 6.0

表2 データ転送時間の評価（リンクの転送速度不均等）

転送データ長	提案方式		従来方式の 転送時間(msec)	シングルリンクで の転送時間(msec)
	分割数	転送時間(msec)		
1 K B	1	9.7 1	2 3.6	9.7 1
2 K B	2	1 4.2	2 6.7	1 4.9
4 K B	2	2 0.0	3 3.0	2 5.3
8 K B	3	2 9.2	4 5.5	4 6.2
1 6 K B	4	4 4.3	7 0.5	8 7.8

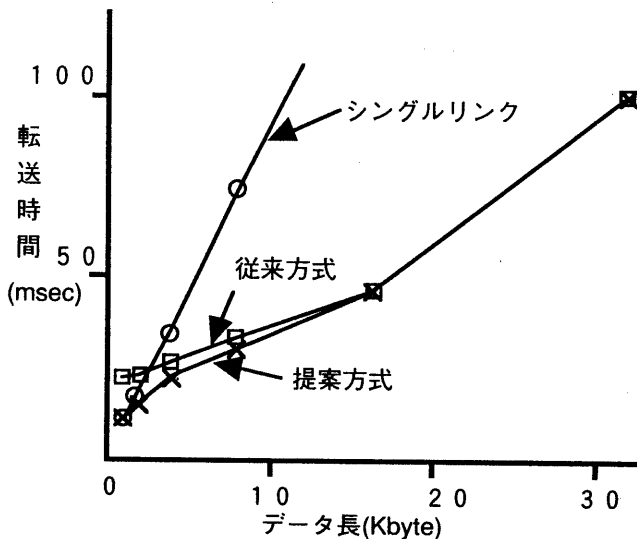


図6 リンクの転送速度均等の場合のデータ転送時間

提案した。また、提案方式を典型的な通信ネットワークのモデル上で定量的な評価を行なった。この結果、提案方式は、従来方式に対して、リンクの転送速度が同一の場合で最大50%、リンクの転送速度が異なる場合で60%～35%の転送時間の改善が図れることを示した。

なお、リンクのデータ転送速度が異なる場合、問題4からリンクのデータ転送速度が同一の場合で示したようなケース分けされたデータ転送アルゴリズムが求められれば、3.3で示したテーブルを持つ必要がない。これについては、

今後の検討課題とする。また、提案方式が有効である範囲に関する考察も今後の課題とする。

本報告で示したデータ転送方式は、通信ネットワークにおけるデータ転送ばかりでなく、ディスクの入出力におけるデータ転送等、幅広い範囲に適用できる。

5. おわりに

本報告では、複数のリンクを用いたデータ転送方式として、パラレルリンク方式を対象とし、データ転送時間を最小化するデータ転送方式を提案した。データ転送方式は、リンクのデータ転送速度が同一の場合と異なる場合のそれぞれについて

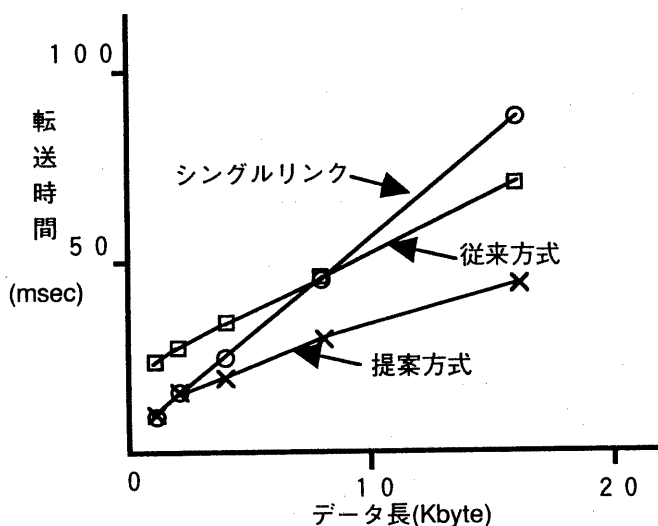


図7 リンクの転送速度不均等の場合のデータ転送時間

[謝辞]

日頃ご指導いただき、本研究の機会を与えていただいた、当研究所所長堂免信義氏ならびに部長寺田松昭氏に深謝致します。また、本検討に当たり、有益なご討論、ご助言をいただいた関係各位に感謝致します。

[参考文献]

- (1) 高橋修：3.3 データリンク層、情報処理 Vol.26、No.4、pp.315-320、1985
- (2) 西園敏弘：異速度回線からなるマルチリンクパケット伝送方式の評価、信学論、Vol.J6a-B、No.12、pp.1647-1655、1986
- (3) 篠崎郁生他：ネットワーク間接続のソフトウェアの開発、大学内ネットワークの相互接続の諸問題シンポジウム、pp.51-57、1990.2
- (4) 西園敏弘他：パケット交換におけるマルチリンク制御、情報処理学会、分散処理システム研究会資料、DP16-4 (昭57-11)
- (5) 吉田他：M/M/2待ち行列系の順序待ち特性、信学論 (B)、J66-B、No.7、pp.926-927、(昭58-7)

(6) 西園敏弘他：マルチリンクパケット伝送方式の解析、信学論 (B)、J67-B、No.1、pp.47-53、(昭59-01)

(7) 西園敏弘他：ランダム分配によるマルチリンクパケット伝送方式の解析、信学論 (B)、J67-B、No.10、pp.1148-1154、(昭59-10)

(8) 西園敏弘他：マルチリンクパケット伝送方式の評価、信学論 (B)、J68-B、No.10、p.1112-1118、(昭60-10)