マルチメディア通信と分散処理 60－3
（1993．5．14）

バースト的な高速通信システムにおける負荷分散方法

日野村 聡　　濱田 晃　　藤田 克孝
NTT交換システム研究所

データ通信アプリケーションの高速化に伴い、データ通信ネットワーク内トラヒックのバースト性が増大しつつあり、その最も顕著な例として、センタ-エンド通信形態における一斉データ送受信が上げられる。センタ端末を収容している交換機に、バースト的なトラヒックが集中すると、通信品質の低下や交換機利用効率の低下等の様々な問題を引き起こすが、本稿で対象とするマルチプロセッサ形交換機においては、通信パスを割り付ける通信処理装置の選択方法により、上記問題をかなり改善できる。本稿では、特に一斉データ送受信に着目し、適切な通信処理装置の選択原理やその基本的なアルゴリズムについて、フレームリレーを例に上げて提案する。

Load Sharing in a Multi-processor Switching System that Handles High-speed and Burst Data Traffic

Satoshi HINOMURA　Akira HAMADA　Katsuyuki FUJITA
NTT Communication Switching Laboratories

As telecommunication networks transmit data more rapidly, the resultant increase in the amount of burst traffic reduces the efficiency with which resources are used and aggravates delay and data loss problems. An especially severe problem is the 'simultaneous burst traffic' that occurs when data on many virtual paths from the same terminal must be transferred at the same time. This paper therefore suggests a load sharing method that overcomes these problems by selecting suitable devices for virtual paths in a multi-processor switching system.

1. はじめに

近年、コンピュータ通信の普及に伴い、特に企業を中心として、パケットやフレームリレー等のデータ通信ネットワークに対する高速化への要求及びトラヒックの需要が増大しつつある。データ通信は、アプリケーションの性質上バースト的に行なわれ、そのバースト性は、ネットワークの高速化に伴い増大する方向にある。端末がバースト的な通信を行なうと、交換機には、通信品質及び交換機リソース利用効率の低下等、様々な問題が起き、この最も顕著な例として、一斉データ送受信がある。

マルチプロセッサ構成の交換機においては、一斉データ送受信を行なう論理パス(データリンク等)を適切な通信処理装置に割り付け、一斉データ送受信のバースト性を通信処理装置間で負荷分散することにより、かなりの通信品質及び交換機リソース利用効率の向上を図ることが出来る。本稿では、以上に示した負荷分散方法について、最近データ通信に用いられつつあるフレームリレーを例に上げて、報告する。

2. 一斉データ送受信

2-1. 一斉データ送受信の定義

今日、企業間ネットワークではセンタ-エンド通信形態をとっている場合が多い。それは、1台の大型ホストコンピュータ(センタ端末)と多くの小型端末(エンド端末)との間に多くの通信パスを設定している通信形態で、一般的に通信パスの数が多い(図2-1参照)。アプリケーションは同時刻に全てのデータリンク上でデータ送受信を行なうように自動化されていることが多く、例えば、POSシステムでは、一定時刻にになると、各店舗の小型端末から本社のホストコンピュータに、一斉に売上等のデータを送出する。このように、多くの通信パスで、同時刻に集団でデータ通信を行なうことを一斉データ送受信と言い、センタ端末が送信する場合を一斉データ送信、その逆の場合を一斉データ受信と言う。

以上のことをまとめると、一斉データ送受信は以下の特徴を持つ通信のことを言う。(図2-2参照)

複数通信パス…複数通信パスが存在する
バースト的通信…各通信パスは一定時刻になると、集団で(連続して)データの送受信を行ない、その他の時刻にはほとんど通信を行なわない。
相関関係がある…全ての通信パス上で、ほとんど同時にバースト的通信が行なわれる。

フレームリレーでは、この通信パスがデータリンクに相当し、Q.922プロトコル上、最大で2の23乗、即ちで800万本以上の多重が可能である。フレームリレーは主にLAN間通信に用いられるが、ホストコンピュータやデータベースを持つLANと他のLANとの間におけるデータリンク上で、センタ-エンド通信形態と同様の一斉データ送受信を行なう可能性は充分考えられる。

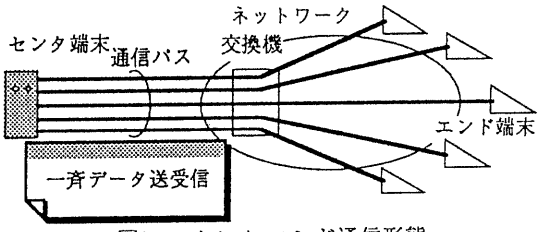


図2-1 センタ-エンド通信形態

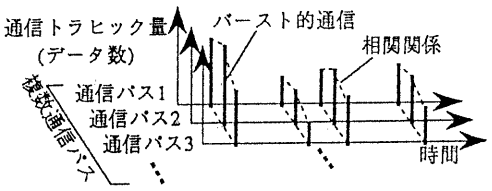


図2-2 一斉データ送受信

2-2. 一斉データ送受信が交換機に与える影響

1端末からのデータが交換機にバースト的に到着した時、交換機リソース(プロセッサ処理能力、バッファ)が不足し、データ廃棄(又は輻輳)やデータ転送遅延の増加等、通信品質の低下を引き起こす。また、端末は、特定の時間帯以外には、ほとんど通信を行なわない為、平均すると交換機にデータが到着していない時間の割合が大きくなり、交換機リソースの使用率が低下する。

このように、バースト性の強いトラヒックを扱うデータ処理交換機は、通信品質及び交換機リソース利用効率の低下と言う2つの問題をかかえている。プライベート網等の比較的小さなネットワークにおける交換機は、交換機リソースが小さい場合が多く、バースト的通信に対する問題はより深刻になる。

一斉データ送受信時、センタ端末を収容するデータ処理交換機には、多くのデータリンクのバースト的通信が集中し、上記の問題はより一層深刻なものとなる。

3. マルチプロセッサ交換機(図3-1参照)

本稿の負荷分散はマルチプロセッサのデータ処理交換機であることを前提としている。本稿で述べるマルチプロセッサ交換機は、中央制御装置及び同一機能の

複数通信処理装置、を具備し、その間はバス等で接続されている。通信処理装置は、データを処理するプロセッサとデータを格納するバッファメモリから成り、データ転送処理の他に、課金やトラヒック観測等の機能を持つ。この様なアーキテクチャを採ることにより、処理能力の小さな通信処理装置を用いても、交換機としては、多くのデータリンクを収容することが可能になる。更に、端末と通信処理装置の間には、パススイッチング機能を設け、任意のデータリンクを任意の通信処理装置に割り付けることを可能とする。そして、通信処理装置間の負荷分散は、中央制御装置で加入者のデータリンクを割り付ける通信処理装置を選択することにより行なう。

データリンクにはリンク設定契機により、以下の2種類がある。

VC…通信開始時、端末からの発呼要求によりデータリンクを割り付ける通信処理装置を選択

PVC…サービスオーダ投入時等に、データリンクを割り付ける通信処理装置を選択(固定的にデータリンクを通信処理装置に割付)

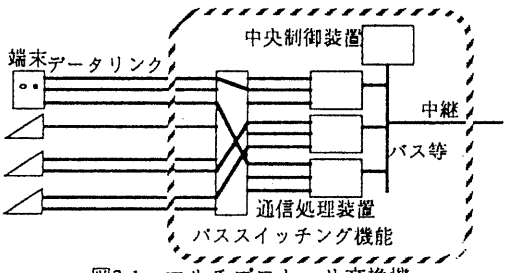


図3-1　マルチプロセッサ交換機

4. 負荷分散論理

4-1. トラヒック特性(図4-1参照)

一般的に、バースト性の強いトラヒックは、以下に示す3つのパラメータで表現される。

平均到着データ数[個/sec]…長時間でトラヒック監視した時の単位時間当たりに到着する平均データ数

集団到着データ数[個]…集団到着する平均データ数

集団データ到着率[/sec]…単位時間当たりに集団データが到着する確率

[平均到着データ数
=集団到着データ数*集団データ到着率]

本稿では、各データリンクはこの様なトラヒック特性を持ち、通信処理装置に到着するトラヒックは、各データリンクが通信するトラヒックの和になることを前提として、負荷分散を行なっている。

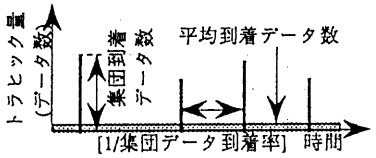


図4-1　トラヒック特性

4-2. 負荷分散の基本的な考え方

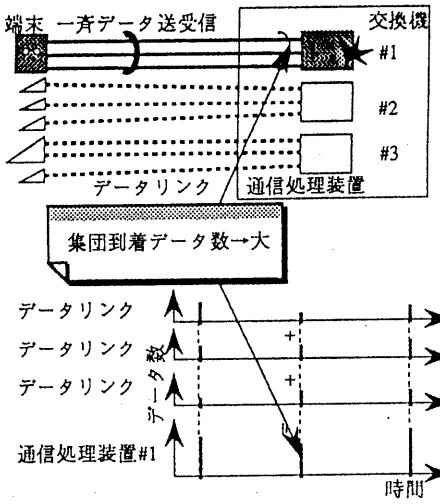
通信処理装置への平均到着データ数が同じでも、集団到着データ数が大きくなると(集団データ到着率は小さくなる)、これらのデータ処理負荷が通信処理装置の処理能力を越える確率が高くなる為に、データ廃棄率やプロセッサ処理待ち時間が大きくなり、通信品質が低下する。この様に、通信品質は、集団到着データ数にも大きく依存し、通信処理装置の処理能力が小さいほど、集団到着データ数の影響度が大きくなる。

しかし、現実には、各通信処理装置毎の平均到着データ数が同じになることのみを考慮にいれて、各通信処理装置にデータリンクを割り付ける場合が多い。この場合、各通信処理装置への集団到着データ数による平均到着データ数からの変化を、マクロ的にトラヒック変動率と見なし、設備算出を行なっているだけであり、集団到着データに対しては消極的な対策しか採られていない。

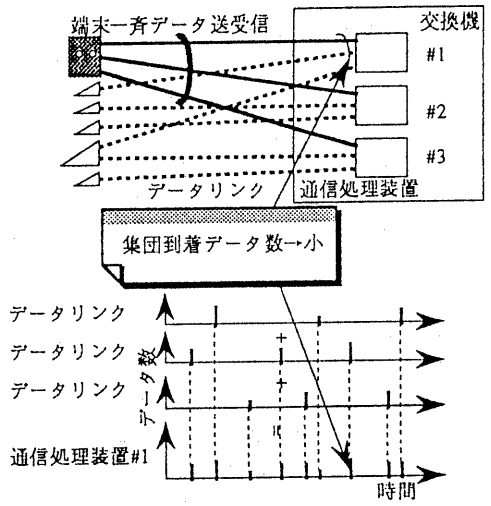
それに対して、本稿で述べる通信処理装置間負荷分散方式では、一斉データ送受信による集団到着データが最も交換機に高負荷をかけることに着目し、その集団到着データ数を積極的に低減する方式を取り入れている。その方法として、相関関係のある複数データリンクを出来るだけ多くの通信処理装置に分散して割り付けることで、一斉データ送受信による集団到着データを複数の通信処理装置に分散する。1通信処理装置には相関関係のないデータリンク同士が割り付けられ、割り付けられている各データリンクからの集団到着データが同時に到着する可能性は低くなる。その為、1通信処理装置に到着する全体としてのトラヒックにおける集団到着データ数が小さくなる。(図4-2参照)

4-3. 負荷分散用パラメータ

本稿で提案する負荷分散方式は、各データリンクに、4-1項で示したトラヒック特性を表すパラメータ(負荷分散用パラメータ)を付与し、そのパラメータを基に行なうことを特徴とする。



(a) 一斉データ送受信するデータリンクを1通信処理装置に割り付けた場合



(b) 一斉データ送受信するデータリンクを複数処理プロセッサに分散して割り付けた場合

図4-2 相関関係のあるデータリンクの割付方式と通信処理装置が受信するトラヒック特性

フレームリレーのように可変長のデータを転送するネットワークにおいては、負荷分散用パラメータを決定するときに、4-1項に示したトラヒック特性だけでは無くデータ長も重要な要素となる。通信処理装置のデータ処理時間はデータ長にほとんど無関係だが、そのデータが使用するバッファの大きさはデータ長に比例する為である。即ち、データ転送遅延時間は通信処理装置の処理待ちデータ数に依存し、データ廃棄率は処理待ちデータ量(データ数*平均データ長)に依存する。よって、データ転送遅延時間の短縮に重点をおく場合と、データ廃棄率の低減に重点をおく場合で、負荷分散用パラメータ値が違ってくる。

ここでは、主にコンピュータネットワークを考慮にいれ、データ廃棄率の低減に重点をおいて負荷分散用パラメータ(BS/CIR/RG)を決定する。各パラメータの意味を以下に示す。

(1) BS(バーストサイズ)

4-1項中における集団到着データ数に相当する。標準化が進んでおり、その値をそのまま用いてもほとんど問題は無い。ただ、データ廃棄率の低減に重点を置くと、1データリンクの集団到着データがバッファを使用する大きさをBSと定義することが最も適切である。すると、一斉データ送受信により使用されるバッファ量を、各通信処理装置に均等に分散させることが出来る。この条件において、回線速度や通信処理装置処理速度等を考慮に入れると、BSの算出式は以下の様になる。

BS[bit]=(集団到着データ量)
-(集団データ到着中の処理データ量)
=8*DL*Max{Int{ND-DS*T},1}
=8*DL*Max{Int{ND(1-8*DL*DS/TS)},1}

T[sec] 集団データ到着時間
ND[個] 集団で到着したデータ数
DL[byte] 平均データ長
TS[bit/sec] 回線速度
DS[個/sec] 通信処理装置処理速度
Int{} 整数化(繰り上げ)関数
Max{a,b} 最大値選択関数

(2)CIR(Committed Infomation Rate)

4-1項中における平均到着データ数に相当する。標準化が進んでおり、その値をそのまま用いてもほとんど問題は無い。ただ、データ廃棄率の低減に重点を置くと、1データリンク上の通信が平均的に使用しているバッファ量をCIRとして定義することが最も適切である。この条件においては、CIRの算出式は以下の様になる。

CIR[bit/sec]=8*ND*DL/T
T[sec] 平均到着データ数が見込める長時間
ND[個] 長時間(T)に到着したデータ数

DL[byte]　平均データ長

(3)RG(Relation Group)

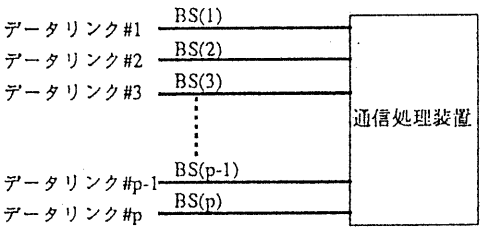
本負荷分散方式で取り入れた新しい概念であり、同時刻に、集団で(連続して)データ通信を行なうデータリンクの関係(相関関係)を表す。同時刻に集団でデータ通信を行なうデータリンクをグループ化し、RGを付与する。そして、RG内で、BSが各通信処理装置に均等に割り振られるように、データリンクを各通信処理装置に割り付けることにより、本負荷分散を行なう。

RGの概念が無いと、1通信処理装置に割り付けられているデータリンク全てに相関関係がある可能性を残す。よって、通信処理装置における最大集団到着データ数は、その通信処理装置に割り付けられている各データリンクのBSの総和の値(図4-3(a)、及び下式参照)となる可能性があり、その最大集団到着データ数を下げることは出来ない。

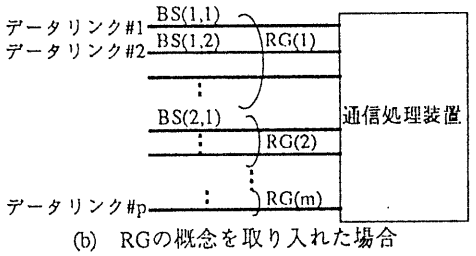
最大集団到着データ数 = ΣBS(i)

それに対して、RGの概念を取り入れると、相関関係があるデータリンクが明らかになる。よって、最大集団到着データ数は、各RG内データリンクのBSの総和の中での最大値(図4-3(b)、及び下式参照)となる。

最大集団到着データ数
= Max { ΣBS(1,i), ΣBS(2,i), ⋯, ΣBS(m,i) }



(a)　RGの概念が無い場合



(b)　RGの概念を取り入れた場合

図4-3　1通信処理装置に到着するトラヒックの変動率

ここで、各「ΣBS(n,i)」を小さくすると、それだけ最大集団到着データ数も小さく出来る。本負荷分散方式は、上式における各「ΣBS(n,i)」を最小化する様にデータリンクを各通信処理装置に割り付けるアルゴリズムになっている。

4-4.　BS/RGによる負荷分散アルゴリズム

リンク管理データで各データリンクの負荷分散用パラメータ(BS/CIR/RG)を管理する(図4-4参照)。更に、各データリンクがどの通信処理装置に割り付けられているのかという情報とそのデータリンクのBSの情報を持つ。そして、各通信処理装置において、RGに閉じて、割り付けられているデータリンクのBSの総和を算出し管理する。この機能はRG単位に持っているBSテーブルにて実現する(図4-5参照)。

端末より、データリンク設定要求が上がった時、まず、そのRGを識別する。そして、BSテーブルにて、そのRG内で、BSの総和が最小になっている通信処理装置を選択し、データリンクを割り付ける。具体的な負荷分散アルゴリズム例を図4-6に示す。

以上の論理を用いて、加入者のデータリンクを適切な通信処理装置に割り当てることにより、一斉データ送受信の集団到着データ数が通信処理装置間に均等に分散される。例えば通信処理装置がn個ある交換機で、本負荷分散方式を用いると、相関関係のあるデータリンクを1通信処理装置に割り付ける場合に比べて、平均通信量を変えずに、理想的には、1通信処理装置への集団到着データ数を1/nに下げることが出来る。その為、ある程度小さい処理能力のプロセッサを用いても、通信品質及び交換機リソース利用効率の向上が可能となる。

4-5.　負荷分散を行なった時の効果

本負荷分散方式は、各通信処理装置への集団到着データ数を均一化する事(平均到着データ数は一定)により、通信品質及び交換機リソース利用効率の向上を図っている。ここでは、収容する端末全てが一斉データ送受信を行なうと仮定し、理想的に集団到着データ数を下げた場合の効果について記す。

1通信処理装置を図4-7に示す通りに待ち行列モデルに疑似する。そして、図4-1で表されるトラヒックを入力し(M[x]/D/1モデル)、その集団到着データ数を変化させた時の呼損率や処理待ち時間の値をシミュレーションで求めることにより、負荷分散効果の確認を行なった。

一例として、一斉データ送受信による集団到着デー

タを1通信処理装置で処理する場合、それに到着するトラヒック条件は以下の通りとした。

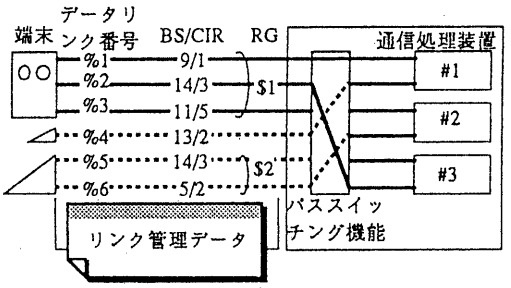


図4-4　負荷分散用パラメータ

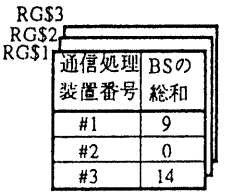


図4-5　BSテーブル

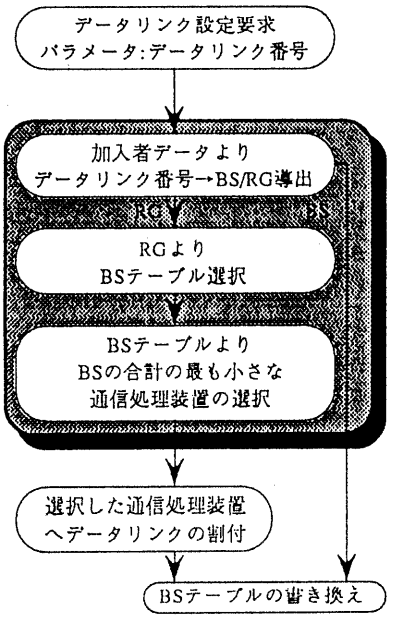


図4-6　BSによる負荷分散アルゴリズム例

集団到着データ数[個]…x　(0～2xの場合があり、その確率は一定である。平均の集団到着データ数はxである。)

集団データ到着率…1/150(プロセッサの1データ処理時間で正規化)

データ長……1データ当たり1バッファを使用(一定)

そして、n台の通信処理装置間で負荷分散を行なうと、1通信処理装置に入力するトラヒック特性は、理想的には、4-4項に示す通り、以下の通り変化する。

集団到着データ数は1/n、集団データ到着率はn倍
(平均到着データ数は一定)

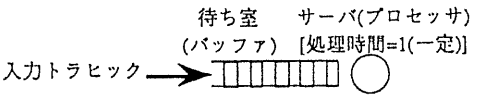


図4-7　シミュレーションモデル

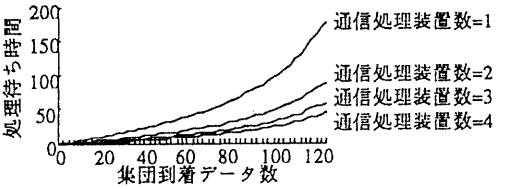


図4-8(a)　集団到着データ数-処理待ち時間率　特性
(バッファ量=∞)

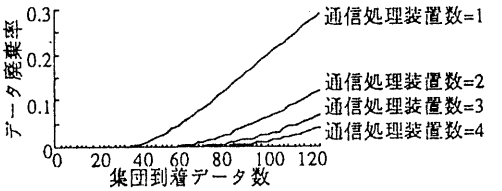


図4-8(b)　集団到着データ数-データ廃棄率　特性
(バッファ量=100)

この条件でシミュレーションをした結果の一部を図4-8に示す。図4-8(a)においては、バッファ量を∞にとったものなので、集団到着データ数は主に処理待ち時間に影響を及ぼし、図4-8(b)はバッファ数に制限を設けたので、集団到着データ数は主にデータ廃棄率に影響を及ぼす。

通信処理装置数を大きくする程、呼損率や処理待ち時間が減少していることが分かる。ただ、通信処理装置数が2台の場合でも、大幅に減少しており、本負荷分散は充分効果があることが分かる。

本結果より、通信処理装置の性能や端末からの入力トラヒック特性が同じでも、1台の通信処理装置で一斉データ送受信処理を行なう場合に比べて、本負荷分散方式を用いると、通信品質の向上が行なえることが確認出来た。実際には、これほど理想的な負荷分散は

行なえないものの、本シミュレーションに近い効果を期待することは出来る。

5. 実際の負荷分散アルゴリズム

5-1. 各通信処理装置に割り付ける上限値の設定

1通信処理装置に割り付けられているデータリンクのBSの総和が、その装置の処理能力を上回ると、その装置は輻輳を起こしやすくなる。よって、本負荷分散方式を実際のシステムに導入する場合には、その総和の上限を決め、上限を越えた場合には、その通信処理装置への割り付け回避を行なわなければならない。

5-2. 輻輳状態中の通信処理装置への割付回避

バースト性の強いトラヒックにおいては、短時間的には、割り付けられているデータリンクのBSの総和が最も小さい通信処理装置が輻輳する場合が考えられる。輻輳中の通信処理装置に新たにデータリンクを割り付けると、輻輳を助長する結果となるので、本負荷分散論理に通信処理装置の輻輳状態をチェックする機能が必要になる。割付予定の通信処理装置が輻輳していたら、他の通信処理装置へ割付回避を行なう。

5-3. CIRによる負荷分散との関係

本負荷分散は、BSのみを考慮に入れて、行なっている。しかし、実際には、各通信処理装置の平均使用率に大きな違いが起こらない様に、CIRによる負荷分散も同時に行なわなくてはならない。

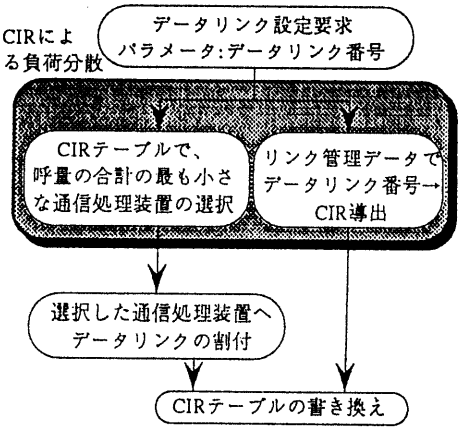
CIRによる負荷分散は次の通りである。まず、各データリンクがどの通信処理装置に割り付けられているのかという情報、及び割り付けられているデータリンクのCIRより、各通信処理装置に割り当てられているデータリンクのCIRの総和をCIRテーブルにて管理する。その時、各データリンクの相関関係は意識する必要が無い為に、CIRテーブルは管理単位を収容データリンク全体とする(図5-1(a)参照)。そして、データリンク設定要求が上がった時には、CIRテーブルを参照することにより、CIRの総和の最も小さな通信処理装置にデータリンクを割り付ける(図5-1(b)参照)。

具体的なBSとCIRの負荷分散の組み合わせ例を以下に示す。一般的に、交換機には一斉データ送受信を行なうデータリンクと行なわないデータリンクが混在して収容されている。この場合は、一斉データ送受信を行なうデータリンクについては、BSによる負荷分散を行ない、一斉データ送受信を行なわないデータリンクについては、CIRによる負荷分散を行なう(図5-2参照)。この時、BSによる負荷分散で割り付けられたデータリンクのCIRもCIRテーブルに反映されることとする。このことにより、BSによる負荷分散で生じた各通信処理装置間の平均使用率の違いを、CIRによる負荷分散で平均化することが出来る。

全データリンクで1テーブル

通信処理装置番号	CIRの総和
#1	1
#2	4
#3	3

(a) CIRテーブル



(b) CIRによる負荷分散アルゴリズム例

図5-1 CIRによる負荷分散例

6. パラメータの最適化

本方式で、正確な負荷分散を実現する為には、負荷分散用パラメータがトラヒック特性を正確に反映している必要がある。4-3項で示している通り、フレームリレー網においては、プロトコルでCIR及びBSが規定されている為、その値をそのまま、負荷分散用パラメータにマッピングすることが出来る。しかし実際には、加入者がアプリケーションより推定したものを元に、CIR及びBSを決める為、初期段階においては、そのパラメータの信頼性は低い。

ただ、コンピュータネットワークにおいて、アプリケーションや通信端末数等、一つのデータリンクの通信環境が変化することはまれであり、データリンクのトラヒック特性もほとんど変化しない。その為、データリンク単位のトラヒック観測と、その観測結果によって負荷分散用パラメータを最適化する学習アルゴリズムを組み合わせると、本負荷分散を正確に実現する為の極めて有効な手段となり得る。そして、上記方法で求めた負荷分散用パラメータで、より正確な通信処

理装置間負荷分散を行なうことが理想である。

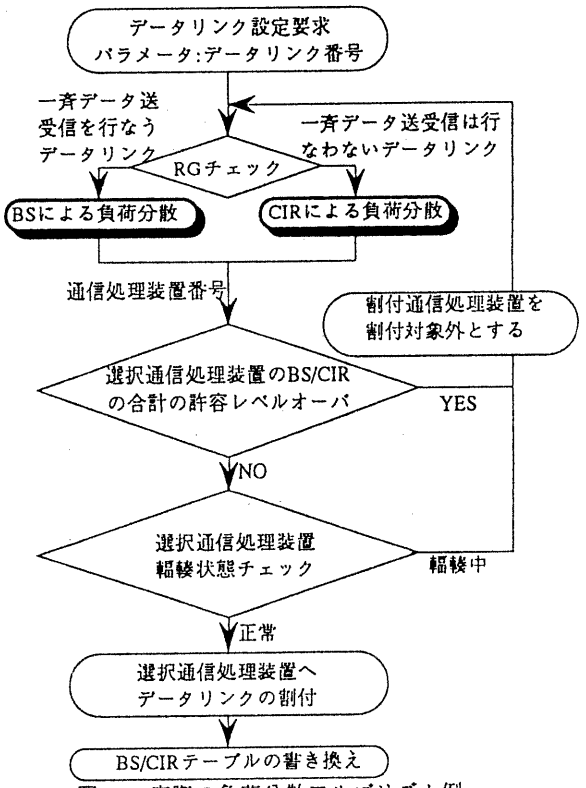


図5-2　実際の負荷分散アルゴリズム例

7. その他の効果

本負荷分散の目的は、1通信処理装置の集団到着データ数を低減することにより、通信品質及び交換機リソース利用効率を向上させることである。しかし、他にも、以下に示す効果が考えられる。

①各通信処理装置に到着するトラヒック特性を均一化し、各通信処理装置に収容されているデータリンクの通信品質を均一化する効果がある。

②交換機のオペレータが、通信処理装置を意識すること無しに、自動的に通信パスを通信処理装置に割り付けることが出来、また、あるデータリンクのトラヒック特性が変化した時も、そのデータリンクの負荷分散用パラメータ値を変えるだけで、適切な通信処理装置間負荷分散を行なうことが出来る。その為、オペレータの保守稼働が小さくなる。

8. 今後の課題

一斉データ送受信を行なうデータリンクを通信処理装置間で分散して割り付けることで、1通信処理装置への集団到着データ数を下げ、通信品質及び交換機リソース利用効率の向上を図る方法について、更に、それらを実現する為の負荷分散アルゴリズムについて報告した。特に小規模なネットワークで、一斉データ送受信を行なう端末を収容する交換機に本方式を採用すると、通信品質が飛躍的に向上すると思われる。

本稿では、フレームリレーを例にとって報告したが、本負荷分散方式は、X.25パケットやATM等、他のデータ処理交換機にも適用出来る。X.25パケットの一斉発着呼はもちろん、今後ATM通信が普及した場合も、TV会議等、一斉データ送受信を行なう通信形態が存在すると考えられる為、本負荷分散方式は有効である。

本稿では、大まかなアルゴリズムの提案に留まったが、今後は特に以下の項目を課題として、引き続き検討していく予定である。

- 1通信処理装置に割り付けられるデータリンクのBS/CIRの総和の上限と通信品質の関係を明確化
- 上記課題は交換機の設備設計との関係が深い為、設備設計へフィードバックする方法
- 小規模から大規模まで、また、一斉データ送受信を行なうデータリンクの数や集団到着データ数をパラメータとして振らせて、トラヒックモデルを作り、シミュレーションを行なうことで、本負荷分散方式による効果の定量的評価
- 上記シミュレーション結果より、トラヒックモデルと、適切なBSによる負荷分散とCIRによる負荷分散の組み合わせの関係
- 負荷分散アルゴリズムとそれに必要なテーブル構成の詳細化
- 本負荷分散方式を実現する為に必要な処理ステップ数とテーブルメモリ量の評価
- 本負荷分散方式のX.25パケット/ATMへの適用方法

参考文献

(1) 有満他:データ交換処理における呼接続方式の検討,情報処理学会,92-DPS-58,1992/11/19
(2) 岩瀬他　分散処理CHMにおける負荷配備方式の検討,93年春期連学全大B-538,
(3) CCITT,RECOMMENDATION X.25
(4) CCITT,RECOMMENDATIONS Q.931,Q.933
(5) CCITT,RECOMMENDATION Q.922
(6) CCITT,RECOMMENDATION I.370