

蓄積交換網におけるルート選択方式と評価シミュレーション

西垣 通, 池田 智明, 池ヶ谷 浩
(日立製作所システム開発研究所)

1. はじめに

データ通信のための蓄積交換網において、メッセージはルート選択方式により定められるルートに沿って転送される。Periodic Updating 方式³⁾、Shortest Queue plus Bias 方式等に代表される適応型ルート選択方式は、トラフィックの変動に対応して動的にルートを与える方式であり、高性能が期待できることから注目されている¹⁾。適応型ルート選択方式の大きな欠点は、メッセージが複数のネットワーク・ノード間で循環するループ現象が発生し、この結果、性能および転送の確実性が損われる可能性があることである。Last M Nodes Visited 方式²⁾は、これを防止しようとするものであるが、一定数 ($M+2$) 以上のネットワーク・ノード間で循環するループ現象に対しては効果がない。

本報告では、ループ現象を完全に防止した、適応ループ・フリー方式とよぶ新しい適応型ルート選択方式を提示する。適応ループ・フリー方式においては、各ネットワーク・ノードを、宛先ノードに到るまでに経由するノード数の最小値により、相異なる集合に分類する。最小にすべき目標函数としては、メッセージが、現在存在する集合から一段階宛先ノードに近い集合に到るまでの時間とする。メッセージが同一の集合内に長時間滞留したり、宛先ノードからより遠い集合に転送されることを禁ずることにより、ループ現象の発生が防止される。

2. 蓄積交換網におけるルート選択

現在ノード A に存在し、宛先がノード Z であるメッセージに着目する。各ノードは、以下に定義される直和集合のいずれかの元となる。(なお、ここで、すべての回線は同一容量と仮定する。)

$$\Omega(Z, n) \triangleq \{ \text{ノード} \mid \text{出発したメッセージは、最小 } n \text{ ホップで } Z \text{ に到着} \} \quad (1)$$

($n = 0, 1, 2, \dots$)

いかなるノードも、ネットワーク内の宛先ノード数に等しい数の、相異なる集合に属する。

ノード A が $\Omega(Z, n)$ に属すると仮定するとき、 $\Omega(Z, n)$ の部分集合 $\omega(Z, n, A)$ を次のように定義する。

$$\omega(Z, n, A) \triangleq \{ \text{ノード} \mid \Omega(Z, n) \text{ の元で } A \text{ と直接結ばれている} \} \quad (2)$$

ここで、次式がなりたつ。

$$\omega(Z, k, A) = \{ \emptyset \} \quad k \neq n-1, n, n+1 \quad (3)$$

メッセージが、 Ω に属するいずれかのノードに最初に到着した時刻を $\tau(n)$ とするとき、 $\tau(n-1)$ と $\tau(n)$ との時間間隔を $d(n)$ と表わす。メッセージが、ノード A からノード Z に転送される時間 D は、次式で与えられる。

$$D = d(n) + d(n-1) + \dots + d(m) + \dots + d(1) \quad (4)$$

$\Omega(Z, n)$ に属するノードから、宛先ノード区に到る最適ルートは、 D を最小にするものである。しかし、 $d(m)$ ($m=1, 2, \dots, n$) は、未来の状態に依存する未知変数であり、最適解を常に与えるようなアルゴリズムは存在しない。Periodic Updating 方式では、 $D \rightarrow \min$ を目標とするが、予測の誤りのため、達成されるとは限らない。 D に比べて $d(n)$ の予測の誤りは比較的小さいため、本報告で提示する適応ループ・フリー方式では、 D のかわりに $d(n)$ を、最小にすべき目標函数とする。局所的最適化という点では、Shortest Queue plus Bias 方式と類似している。

$d(n) \rightarrow \min$ が $D \rightarrow \min$ と一致するのは、 $\Omega(Z, m)$ 内のノードが互いに連結されており、しかもそれらのノード間での転送時間が無視できる場合である。逆に、 $\Omega(Z, m)$ 内のノードが互いに連結されていないような場合、 $d(n) \rightarrow \min$ と $D \rightarrow \min$ とは必ずしも一致しない。

ループ現象は、メッセージが $\Omega(Z, n)$ 内のノードを 3 回以上通過したり、 $\Omega(Z, n)$ ($n > n$) 内のノードに転送されることから発生する。したがって、これを禁ずることにより、ループ現象を抑止することが可能である。

3. 適応ループ・フリー方式 (Adaptive Loop Free - ALF)

3.1 ルート選択のアルゴリズム

適応ループ・フリー方式を提示する。以下、ルートを選択すべきメッセージは、宛先ノードが区であり、 $\Omega(Z, n)$ 内のノード A に存在すると仮定する。

まず、 $Q(l)$ ($l=-1, 0, 1$) を定義する。

$$Q(l) \triangleq \min_{S \in \omega(Z, n+l, A)} \{ \text{ノード A より ノード S への転送遅れ} \} \quad (l=-1, 0, 1) \quad (5)$$

転送遅れは、ノード S へ送信するための出力キューに並んだメッセージの数により与えられる。 $Q(-1)$ ないし $Q(0)$ は、メッセージが次に送られるべきノードへの転送遅れである。前者は最短路、後者は迂回路に対応する。いずれを選択するかをの基準となるのが、式(6)である。図1を参照されたい。

$$\min \{ Q(-1), Q(0) + \hat{Q}(-1) \} \quad (6)$$

$Q(-1)$, $Q(0) + \hat{Q}(-1)$ は、それぞれ、最短路、迂回路を介して $\Omega(Z, n-1)$ に転送された場合の転送遅れを表わす。 $\hat{Q}(-1)$ は、式(5)において、ノード A のかわりに A の隣接ノード B を代入

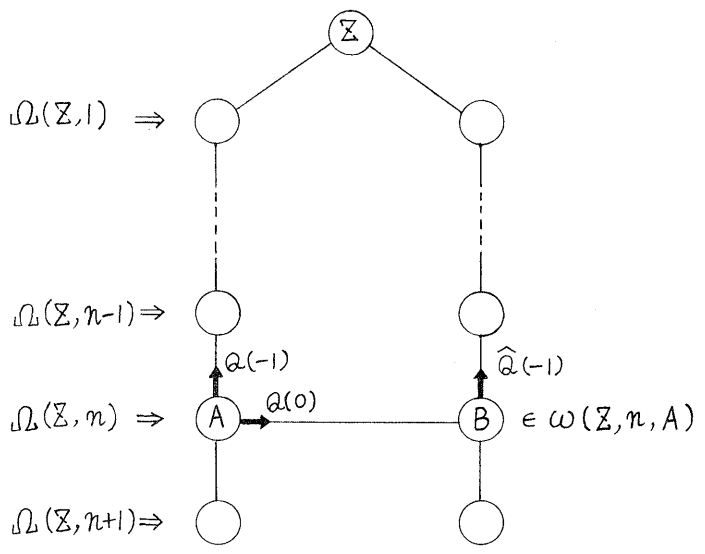


図1. 適応ループ・フリー方式におけるルート選択

し、 $l = -1$ においてえられる $Q(-1; B)$ の推定値を表わす。ただし、 B は A からの転送遅れが $Q(0)$ に対応する、 $\omega(\Sigma, n, A)$ 内のノードである。

なお、 $\omega(\Sigma, n, A)$ が空集合のときは、最短路がただちに選ばれる。

迂回路を介して到着したメッセージに対しては、無条件で最短路が選ばれる。この理由は、式(6)において、迂回路の直後には最短路をとることが仮定されているためである。この結果、 $\Omega(\Sigma, n)$ 内のノードから Σ に到るまでに通過するノードの数はたかだか $2n$ 個であり、ループ現象は完全に防止される。

以上のベタアルゴリズムは、 $d(n) \rightarrow \min$ をみたす最適解とは限らないが、実用上さしつかえない程度にこれと近い解を与えることができる。

3.2 $\hat{Q}(-1)$ の推定方式

式(6)における $\hat{Q}(-1)$ の推定法について以下のべる。隣接ノード B の状態にかかわらず、これを一定値とする方法と、 B の状態を B から常時知らせてもらう方法が考えられる。前者は Shortest Queue plus Bias 方式と類似の方法であるが、予測の精度が低下する。後者は Periodic Updating 方式と類似の方法であるが、トラフィック・オーバヘッドが問題となる。

本方式では、 B より A に迂回路をとるメッセージに、その時受における最短路の転送遅れ $Q(-1; B)$ を付加することにより、できるだけ小さいオーバヘッドで隣接ノード B の状態を知らせてもらう。いま、時受 t_j 、 t_{j-1} ($= t_j - \Delta t$) における $Q(-1; B)$ の推定値をそれぞれ $\hat{Q}_j(-1)$ 、 $\hat{Q}_{j-1}(-1)$ とし、 $t_{j-1} \sim t_j$ の間にノード B から到着した最後の迂回メッセージに付加されていた転送遅れを $Q_j(-1; B)$ とする。 $\hat{Q}_j(-1)$ は、 Δt ごとに次式により算出される。

$$\begin{cases} \hat{Q}_j(-1) = (1-\theta) \cdot Q_j(-1; B) + \theta \cdot \hat{Q}_{j-1}(-1) \\ \hat{Q}_0(-1) = 0 \end{cases} \quad (\theta: \text{const}, 0 < \theta < 1) \quad (j = 1, 2, 3, \dots) \quad (7)$$

式(7)を書き直すと、

$$\hat{Q}_j(-1) = (1-\theta) \cdot \sum_{u=0}^{j-1} \theta^u \cdot Q_{j-u}(-1; B) \quad (8)$$

であるから、結局、過去の重みつき平均で未来の値を予測することになる。

ノード B では、前回 A に迂回メッセージを送ってから Δt 経過したとき、 $Q(-1; B) \neq 0$ の場合に限り、この値を付加した特別のメッセージを A に送る。 $Q(-1; B) = 0$ の場合は何も送らない。したがって、 A では、 $t_{j-1} \sim t_j$ の間に $Q_j(-1; B)$ の情報をうけとらなければ、 $Q_j(-1; B) = 0$ とみなして $\hat{Q}_j(-1)$ を算出する。Periodic Updating 方式では、定期的に、すべての宛先ノードに関する転送遅れ予測値であるアップデート・ベクトルを送るのに対し、本方式で送られる情報は、1つの出力キューの待ち合わせ数にすぎない。したがって、迂回メッセージにたかだか 8 ビット程度の情報を付加すればよい。また、迂回メッセージが全く送られない場合は、 $Q(-1; B) = 0$ の確率が大きいと予想されるので、これを送るためのトラフィック・オーバヘッドは比較的小さいとみなすことができる。

3.3 ネットワーク・トポロジの変動に対する適応

回線やノードの障害発生にともなうネットワーク・トポロジの変動に対する適応性について述べる。適応ループ・フリー方式においては、各ノードはネットワーク・トポロジを知っている必要はなく、各宛先ノードに関して、自ノードならびに隣接ノードの属す集合 Ω を知っていればよい。これらは、ネットワークの初期設定時、回線やノードの障害発生/回復時、ネットワーク・トポロジ変更時等

に、動的に学習される。

学習のアルゴリズムは、ARPAネットワークにおいて、ノードの「disconnection」を感知するために行っている方法と全く同一である³⁾。ただし、ARPAネットワークでは、これを定期的に行うが、適応ループ・フリー方式においては、障害や変更など、必要が生じた時点で実行する。

4. シミュレーション評価

適応ループ・フリー方式の性能を、従来の代表的な適応型ルート選択方式であり、かつ実際に使用されている Periodic Updating 方式、Shortest Queue plus Bias 方式と比較した。(なお、以下、適応ループ・フリー方式をALF方式—Adaptive Loop Free方式、Periodic Updating 方式をPU方式、Shortest Queue plus Bias 方式をSQ+B方式と略記する。)図2に示す種類のネットワークについてシミュレーションを実行し、検討を行った。

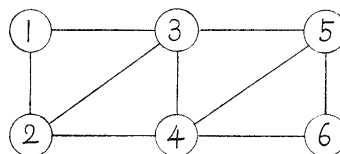
ALF方式のごとく d のかねりに $d(n)$ の最小化を目標とする方式が効果を發揮するのは、同じ d に属する各ノードどうしが互いに連結されている場合の多いネットワークである。逆に、これらが連結されていないネットワークにおいては、PU方式のような d の最小化を目標とする方式が有効であると考えられる。前者の例として、図2の梯形ネットワーク、後者の例として、図2の環形

ネットワークを採用する。また、現実的な場合として、図2の日本列島形ネットワークについて検討する。これは、日本の代表的な都市を結ぶネットワークを想定したものであり、ノード1は東京、ノード2は仙台、ノード3は名古屋、ノード4は金沢、ノード5は札幌、ノード6は大坂、ノード7は広島、ノード8は北九州市に対応する。

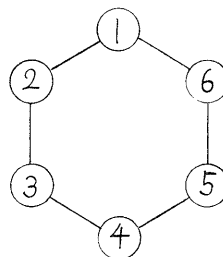
これらのネットワークにおいて、各ノードは、容量4800 bit/secの全2重回線が結ばれており、固定長500 bitのメッセージが、ポアソン分布にしたがってネットワークに流入すると仮定する。

性能評価実行に際し、各種のパラメータの値は、中程度の負荷のもとでメッセージの平均転送時間が最も小さくなるように設定した。これらのパラメータは、 θ 、 Δt 、 d_p バイアス¹⁾、Periodic Update Rate¹⁾などである。

(1) 梯形ネットワーク



(2) 環形ネットワーク



(3) 日本列島形ネットワーク

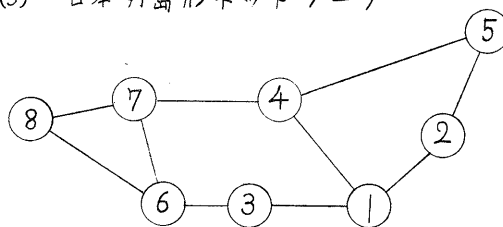


図2. 考察したネットワーク

4.1 梯形ネットワーク

(1) 負荷が一様な場合

梯形ネットワークの場合について、3種類のルート選択方式のもとでの平均転送時間(メッセージが発信元ノードを出発してから宛先ノードに到着するまでの平均時間間隔)を、負荷(ネットワークに単位時間当たりに入力するメッセージ量 bit/sec)の函数として表したのが図3である。ここで、負荷は「一様」であると仮定する。一様であるとは、ネットワーク内の任意の2ノード間のトラフィック量を表わす6行6列のトラフィック・マトリックスにおいて、対角要素(=0)以外の要素がすべて等しいことを意味する。

負荷が高くなるにつれ、PU方式は、ALF方式やSQ+B方式より、平均転送時間が増大する。高負荷状態(30,000 bit/sec)において、ALF方式、SQ+B方式の平均転送時間は、PU方式よりも11~12%小さい。この理由を示すのが表1である。

表1は、負荷30,000 bit/secの場合について、平均ホップ数と出力キュー平均滞在時間とを示している。平均ホップ数とは、発信元ノードから宛先ノードにたどりつくまでに経由した回線数の平均値である。出力キュー平均滞在時間とは、メッセージが出力キューに並んだ時分から転送終了するまでの時間の平均値である。ALF方式の場合、SQ+B方式と比べて、平均ホップ数が増加しているが、出力キュー平均滞在時間が減少しており、迂回路の選択による効果をおよぼしている。しかし、PU方式におけるように、両者がともにSQ+B方式より増加していることは、ループ現象もしくは不適切な迂回が行われたことを示唆している。

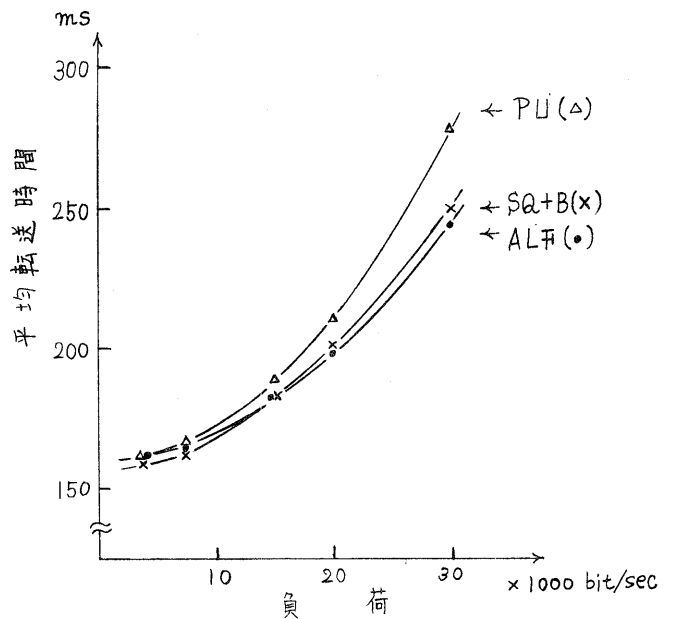


図3. 梯形ネットワークにおける平均転送時間

表1. 梯形ネットワークにおける性能比較

(1) 負荷が一様な場合 (30,000 bit/sec)			
	SQ+B	PU	ALF
平均ホップ数	1.75	1.89	1.82
出力キュー平均滞在時間 (ms)	142	147	135
(2) 負荷に偏りがある場合 (4→3 / その他 = 16/1)			
	SQ+B	PU	ALF
平均転送時間 (ms)	159	168	154
平均ホップ数	1.35	1.44	1.41
平均転送時間(4→3) (ms)	145	151	136
平均ホップ数(4→3)	1.10	1.20	1.21
出力キュー 4→3の平均滞在時間 (ms)	135	130	119

(2) 負荷に偏りがある場合

梯形ネットワークにおいて、トラフィックに偏りがある場合につき測定した結果を、表1にあわせて示す。6行6列のトラフィック・マトリックスにおいて、オ4行オ3列の要素、すなわちノード4からノード3に到るトラフィックのみ4メッセージ/sec、対角要素を除くそれ以外の要素は0.25メッセージ/secとする。

表1には、ネットワーク全体の性能データに加えて、ノード4からノード3に到るトラフィックに関するデータが示してある。これに注目すれば、ALF方式とその他の2方式との性能上の差異は明らかである。同トラフィックに関する平均ホップ数に示されているように、ALF方式とPU方式とは、迂回の発生率はほぼ等しい。しかし、前者は後者よりノード4→3の出力キューの平均滞在時間が小さく、この結果、同トラフィックに関する平均転送時間も減少している。また、SQ+B方式は、ALF方式に比べ、同トラフィックに関する平均ホップ数は小さく、ノード4→3の出力キュー平均滞在時間は大きい。したがって、迂回の効果はALF方式の方がSQ+B方式より大きいことがわかる。

4.2 環形ネットワーク

(1) 負荷が一様な場合

環形ネットワークの場合について、3種類のルート選択方式のもとでの平均転送時間を、一様な負荷の函数として表わしたのが図4である。総じて、3方式の間で顕著な差異はみられない。高負荷状態(18,750 bit/sec)のとき、全体的最適化を行うPU方式は、局所的最適化を行うALF方式やSQ+B方式に比べて、ややすぐれている程度である。

この場合の関連データを表2に示

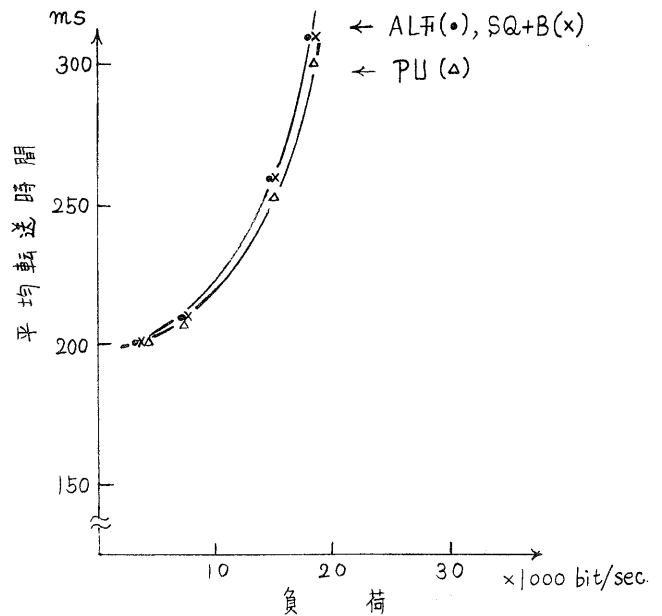


図4.環形ネットワークにおける平均転送時間

表2.環形ネットワークにおける性能比較

(1)負荷が一様な場合 (18,750 bit/sec)			
	SQ+B	PU	ALF
平均ホップ数	1.80	1.81	1.80
出力キュー平均滞在時間 (ms)	172	166	173
(2)負荷に偏りがある場合 (2→1 / その他 = 16/1)			
	SQ+B	PU	ALF
平均転送時間 (ms)	200	181	189
平均ホップ数	1.57	1.55	1.54
平均転送時間 (4→1) (ms)	367	341	355
平均ホップ数 (4→1)	3.06	3.01	3.00
出力キュー 2→1の平均滞在時間 (ms)	161	144	156

す。平均ホップ数は、すべて同様であり、ループ現象は生じていない。出力キュー平均滞在時間は、PI方式が、AL方式やSQ+B方式よりやや小さい。これは、たとえばメッセージがノード4からノード1に転送される場合、AL方式やSQ+B方式では、ノード4→3および4→5の出力キューのみの状態からルートを決めるのに対し、PI方式は、これに加えて3→2, 2→1, 5→6, 6→1の出力キューの状態をも考慮してルートを決めるためと考えられる。

(2) 負荷に偏りがある場合

環形ネットワークにおいて、局所的最適化と全体的最適化との差異は、各出力キューにおける待ちがほぼ同程度である場合は顕著ではないが、トラフィックに偏りがあるときは大きなものとなる。6行6列のトラフィック・マトリックスにおいて、才2行才1列の要素、すなわちノード2からノード1に到るトラフィックのみ4メッセージ/sec、対角要素以外のその他の要素はすべて0.25メッセージ/secとした場合の測定値を、表2にあわせて示す。

PI方式は、AL方式より4%、SQ+B方式より9%、それぞれ平均転送時間が小さい。ノード2→1の出力キュー平均滞在時間も、PI方式は、AL方式やSQ+B方式より減少している。ノード2→1の出力キューにおける待ちが多いとき、PI方式ではこれをノード4やノード3で知ることができ、ノード4からノード1に到るメッセージには4→5→6→1、ノード3からノード6に到るメッセージには3→4→5→6というルートを選択する。AL方式やSQ+B方式においては、このような迂回は期待できず、4→3→2→1ないし3→2→1→6のルートもとられるため、ノード2→1の出力キューにおける待ちが増大する。

AL方式とSQ+B方式との差異は、ノード4からノード1に到るトラフィックに関する平均ホップ数からも明らかかなように、後者ではループ現象が発生したためである。AL方式は、この点に関してSQ+B方式よりすぐれており、PI方式との差異は比較的小さい。

4.3 日本列島形ネットワーク

(1) 負荷が一様な場合

日本列島形ネットワークについて、3種類のルート選択方式のもとでの平均転送時間を、一様な負荷の函数として表わしたのが図5である。負荷が高くなるにつれ、PI方式やSQ+B方式は、AL方式より平

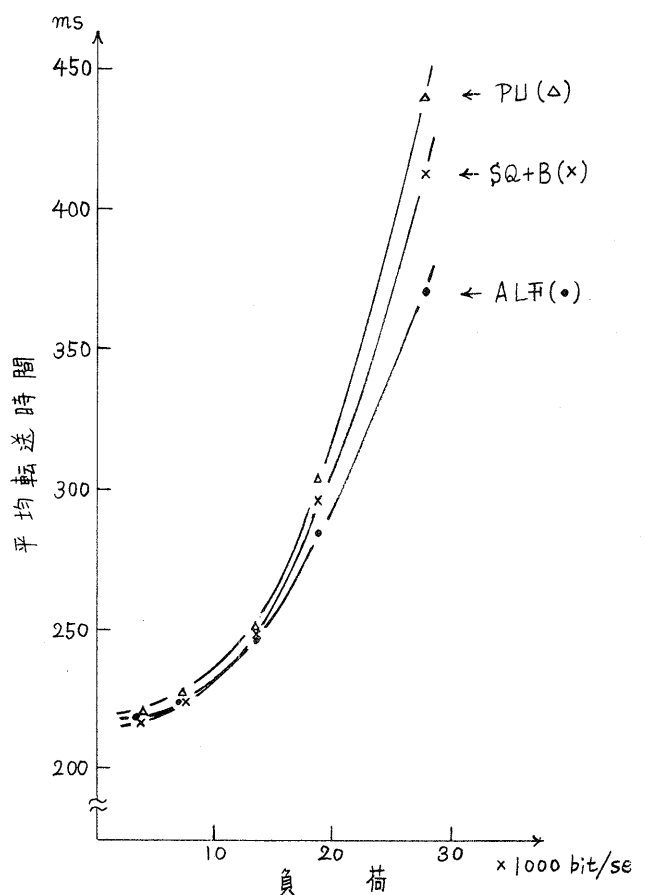


図5. 日本列島形ネットワークにおける平均転送時間

均転送時間が増大する。28,000 bit/sec の高負荷のもとで、ALF方式は、PU方式より16%、SQ+B方式より10%、それぞれ平均転送時間が減少している。

表3に、この場合の関連データを示す。PU方式は、平均ホップ数、出力キュー平均滞在時間とも、ALF方式より増大しており、ループ現象ないし不適切な迂回が発生したことを示唆している。PU方式においては、負荷が高く、ひんぱんに最適路が変化するようなとき、性能が低下する恐れがある。一方、SQ+B方式は、ALF方式に比べ、平均ホップ数はほぼ同様であるが、出力キュー平均滞在時間が増大している。これは、SQ+B方式においては、隣接ノードの出力キューにおける待ちの状態が全く不明なため、迂回効果が小さかったと考えることができる。

(2) 負荷に偏りがある場合

日本列島形ネットワークにおいて、トラフィックに偏りがある場合の測定結果を、表3にあわせて示す。8行8列のトラフィック・マトリックスにおいて、オ6行オ3列およびオ7行オ4列の要素、すなわちノード6からノード3に到るトラフィックと、ノード7からノード4に到るトラフィックのみ4メッセージ/sec、その他は0.25メッセージ/secとする(ただし、対角要素は0)。

ALF方式は、平均転送時間において、PU方式より12%、SQ+B方式より20%、それぞれ減少している。特に、ノード8からノード2に到るトラフィックに着目すると、3方式のうちでSQ+B方式の性能低下が著しい。同トラフィックに関する平均ホップ数、ノード6→3、7→4の出力キュー平均滞在時間、がすべて増大していることから、SQ+B方式のもとで、6→7→6のようなループ現象がひんぱんに発生したと考えることができる。一般に、SQ+B方式においては、トラフィックに偏りがあり、迂回先における最短路の待ちがいずれも大きいとき、しばしばループ現象を生ずる可能性がある。

5. おわりに

ループ現象の発生を抑止することにより、速く、確実なメッセージ転送を可能とする、適応ループ・フリー(Adaptive Loop Free)方式とよばれルート選択方式を提案した。シミュレーションにより、従来の代表的な適応型ルート選択方式であ

表3. 日本列島形ネットワークにおける性能比較

(1) 負荷が一樣な場合 (28,000 bit/sec)			
	SQ+B	PU	ALF
平均ホップ数	2.18	2.39	2.15
出力キュー平均滞在時間 (ms)	189	184	173
(2) 負荷に偏りがある場合 (6→3, 7→4 / その他 = 16 / 1)			
	SQ+B	PU	ALF
平均転送時間 (ms)	266	241	213
平均ホップ数	1.98	1.81	1.67
平均転送時間 (8→2) (ms)	586	544	499
平均ホップ数 (8→2)	4.55	4.37	4.20
出力キュー 6→3の平均滞在時間 (ms)	181	164	158
出力キュー 7→4の平均滞在時間 (ms)	184	177	162

る Periodic Updating 方式, Shortest Queue plus Bias 方式と比較したところ, 梯形および日本列島形ネットワークにおいて, すぐれた性能が観察された。

適応ループ・フリー方式は, ミニ・コンヒトAC-10II (16 K words) を蓄積交換機として, 実験的に実現されており, 今後機会があれば, 改良を加えていく予定である。

(参考文献)

- 1) G. L. Fultz, Adaptive Routing Techniques for Message Switching Computer-Communication Networks, UCL A-ENG-7252, UCL A, July 1972.
- 2) R. L. Pickholtz, et al., Improvements in Routing in a Packet-Switched Network, Proceedings of the ICC '74, T3230, 1974, 249-252.
- 3) F. E. Heart, et al., The Interface Message Processor for the ARPA Computer Network, Proceedings of the SJCC, 1970, 551-567.
- 4) T. Ito, et al., Packet Switching Subnet of JIPNET System, Proceedings of the 2nd UJCC, 1975, 459-463.