

分散データベースにおける重複データ制御方式の評価

金子朝男 西原義文
(日本電気株式会社)鶴岡邦敏 服部光宏
(中央研究所)

1. はじめに

分散データベースシステムにおける重複データ制御の問題について、活発な研究が行われており、今日までに数多くの方式が発表されている[6]。先駆的のものとしては多数決方式[5]があり、また筆者らも論理時刻方式[3, 7]を提案した。

これまで発表された制御方式に関して定性的評価は行われているが[7など]、定量的評価は少ない[1, 2, 4]。重複データ制御方式の評価項目として

- ・ 応答時間/処理能力
- ・ 通信頻度/通信量
- ・ 制御の複雑性
- ・ 障害対策/リカバリ方式

等があり、最初の3項目は定量的評価が可能である。

筆者らはシミュレータDDACS (Distributed Database Access Control Simulator)を開発し、各方式の応答時間、通信量等を定量的に比較評価したので、ここに報告する。

重複データ制御方式は

- 1) 集中制御方式
- 2) 回覧型多数決方式
- 3) 分散型論理時刻方式

を評価対象として選択した。ネットワーク・トポロジ=星、分散管理を指向した完全グラフ形から、集中管理を指向した星形、木形までの5形態を選んだ。

本評価により、完全グラフ形の場合、論理時刻方式が他の方式に比べて応答時間が短く、処理能力が大きいことが明らかとなった。また一般的にネットワーク形態に近い形の場合、論理時刻方式は集中制御方式と比べて高負荷に

耐えられることが示された。

本資料では初めにシミュレータDDACSの説明を行ない、次に完全グラフ形における性能を考察し、最後にネットワーク・トポロジ=星の制御方式に対する影響を検討する。

2. シミュレータ

シミュレータDDACSおよびシミュレータ=シミュレーション環境について説明する。

(1) シミュレータ=シミュレーション: モデル

DDACSは図1に示すようなイベント駆動型のシミュレータである。通信回線の遅れを重視し、各回線ごとに待合せの制御を行っている。シミュレータ=シミュレーション時間を短縮するため、中央制御装置とデータベースでの特定の制御は行わない。更新要求の発生は指定された頻度のポアソン分布に従い、各ホストで独立に制御される。更新要求間の競合は考慮しない。

(2) 指定項目

DDACSは表1に示す12の入力項目を有する。制御方式およびネットワーク・トポロジ=星に関しては後述する。更新要求間隔は指数分布の平均値を指定する。日=カルデタ: データ入出力に要する時間である。データベース管理システム処理時間、ネットワーク管理システム処理時間は、管理システムが制御に要する時間である。以下のシミュレータ=シミュレーションにおいて指定項目が特別に記述されていない場合、その値は表1の既定値とする。

(3) 重複データ制御方式

集中制御方式，回覧型多数決方式，分散型論理時刻方式を対象とする。

どうかの判定を行ない，可能な場合，すべてのスレーブホストに対して，更新実行命令を送信する。

a. 集中制御方式

図2に示す集中制御方式を測定する。各更新要求は，その参照データとともに，マスターホストに転送される。マスターホストは更新可能か

b. 回覧型多数決方式

図3に示す回覧型多数決方式を採用する。更新要求が発生すると他ホストに回覧し，更新可能か否かの投票を求める。可能の投票が過半数に達した場

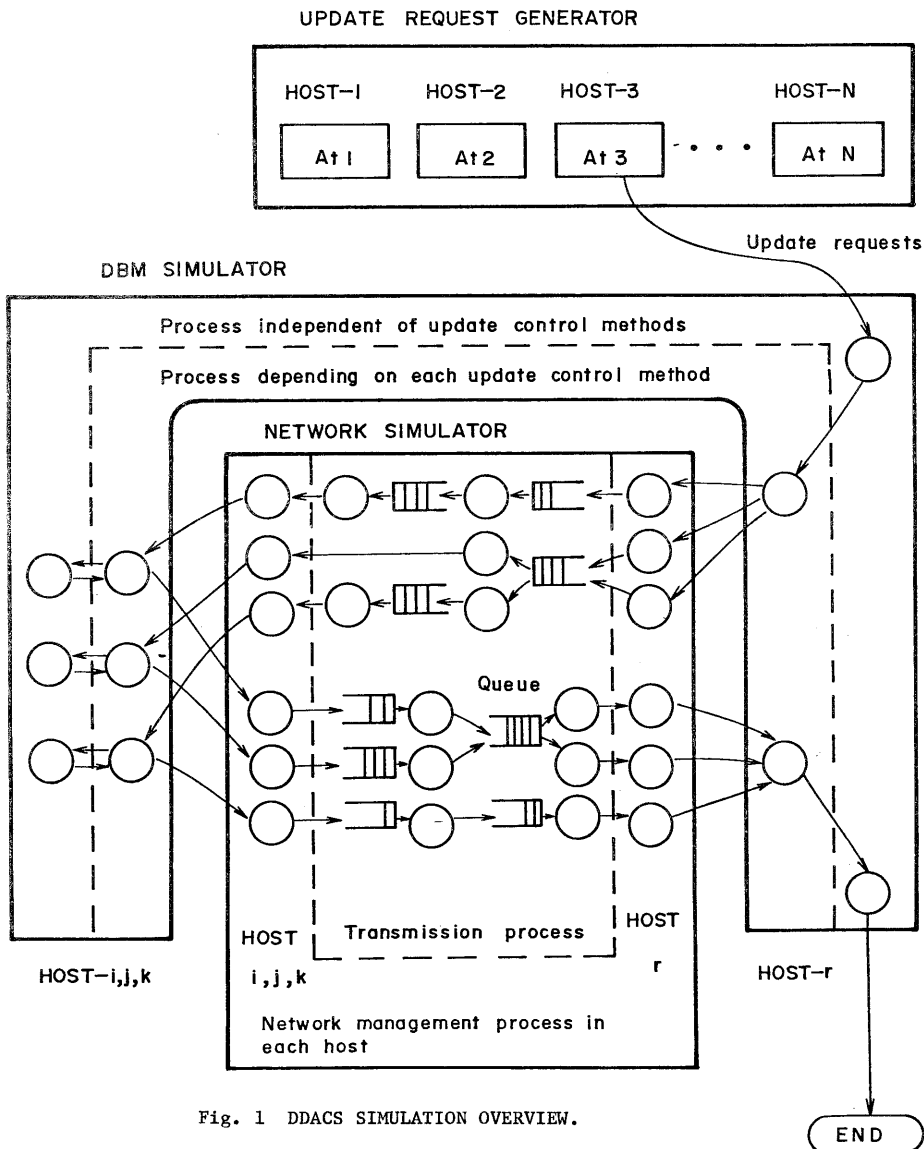


Fig. 1 DDACS SIMULATION OVERVIEW.

合、最後に投票したホストは他の全ホストに更新実行命令を送信する。以後この方式を単に多数決方式と呼ぶ。

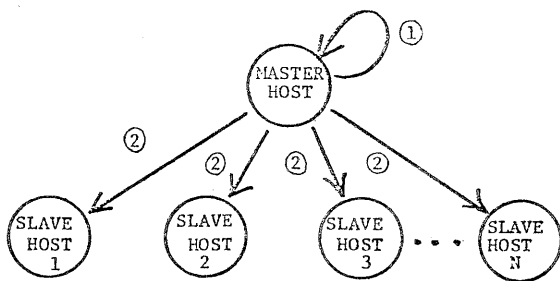
C.分散型論理時刻方式

論理的な時計を使用した方式にはSDD-1等の最遅時刻同期方式と筆者等が提案した分散型論理時刻方式とがある。これらの方式の違いは文献[6]で検討した。ここでは分散型論理時刻方式I[3]を代表として採用する。この方式では各ホストに論理的な時計が存在し、他ホストの論理時刻と同期をとりながら進行する。あるホストで発生した更新要求は他のホストに転送され、全ホストで同一の論理時刻に更新の可能

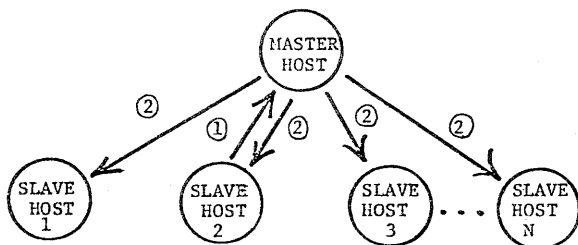
表1.DDAC Sの指定項目

No	指 定 項 目	既定値
1	制御方式名	—
2	ネットワーク・トポロジ・名	—
3	更新要求間隔	1秒
4	ローカルデータアクセス時間	50ミリ秒
5	データベース管理システム 処理時間	1ミリ秒
6	ネットワーク管理システム 処理時間	1ミリ秒
7	通 信 速 度	50 Kbps
8	更新要求メッセージ長	200バイト
9	制御メッセージ長	20バイト
10	重複データ保有ホスト数	—
11	更新要求発行ホスト数	—
12	シミュレーション時間	60秒

(1) Master host issues a request.



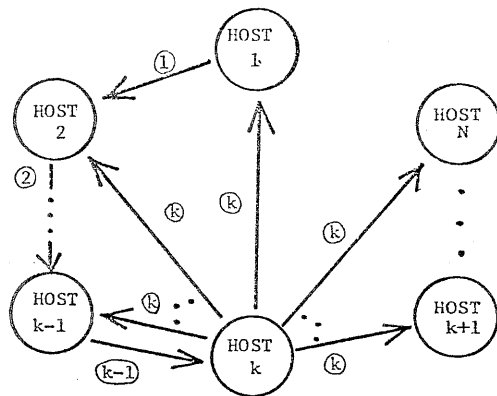
(2) Slave host issues a request.



[NOTES]

- ①: An update request message.
- ②: A control message which indicates acceptance of the request.

Fig.2 CENTRALIZED CONTROL METHOD.



[NOTES]

When an application process at host 1 issues an update request and host k detects a majority consensus.

- ①, ②, ..., (k-1): An update request message issued at host 1.
- (k): A control message which indicates acceptance of the request.

Fig.3 MAJORITY CONSENSUS METHOD.

性能判定され、更新が実行される。

(4) ネットワーク：トポロジ

DDACS に登録されている 20 個のネットワーク形態の中から、5 個の形態を選んで測定した。図 4 に、それらの形態と便宜上の名称を示す。完全グラフ形態では、ホスト数を変化させた時の結果も示す。

(5) 出力項目

DDACS は、シミュレ = ショシが終了すると以下に示す項目の平均および分布を出力する。分布は指定によりグラフ化して出力する。

- ・ 応答時間：システム全体、およびホスト単位。
- ・ 回線使用率：ホストの全三重回線を含むネットワークでは、各々の送受信による 20 個の回線使用率を測定する。
- ・ 更新要求発生個数：ホスト単位。
- ・ 発生したメッセージの個数、その長さ。
- ・ 論理時刻の時間的分布。
- ・ 1 論理時間内に処理された更新要求数。

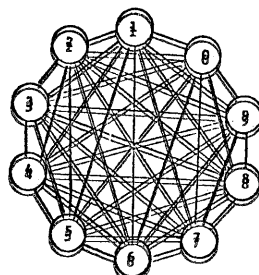
3. 完全グラフ形における性能

図 4 の (1) に示した完全グラフ形ネットワークにおける各制御方式の性能特徴を説明する。

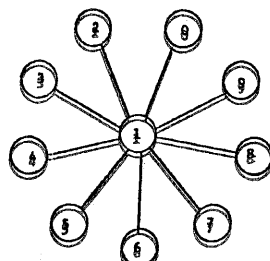
(1) 更新要求発生間隔と応答時間

ホスト数が 5 台の場合の更新要求発生間隔と応答時間の関係を図 5 に示す。論理時刻方式はすべての更新要求発生間隔において、他方式よりも応答時間が短い。集中制御方式と多数決方式を比較すると、負荷が小さい場合、集中制御方式が良いが、高負荷になると逆転する。ホスト数が 5 台の場合、集中制御方式では、更新が実行されるまでに 20 回の

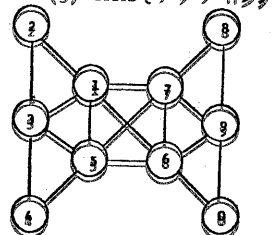
(1) COMPLETE (完全グラフ形)



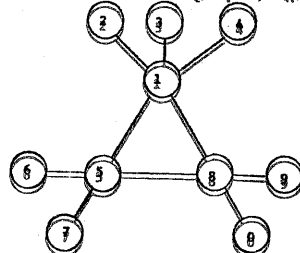
(2) STAR (星形)



(3) RING (リング形)



(4) TRIANGLE (トライアングル形)



(5) TREE (木形)

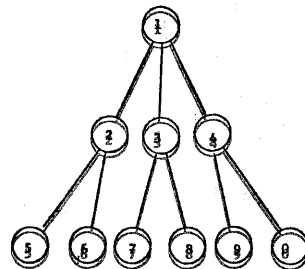


Fig. 4 NETWORK TOPOLOGY

通信が必要であるが（スレーブホストで更新要求が発生した時）、多数決方式では最低3回の通信が必要である。論理時刻方式では1回の通信時間を少し上まわる程度の時間で更新要求の実行可否が判定できる。従って低負荷の時の応答時間は論理時刻方式、集中型制御方式、多数決方式の順で大きくなる。

一部の回線に負荷が集中する集中制御方式では、他の方式よりも低負荷で回線が飽和する。これにより多数決方式と集中制御方式の応答時間の逆転が発生する。

図5よりホスト数5台の時の各制御方式の処理能力を概算すると、論理時刻方式：15要求数/ホスト秒，多数決方式：10要求数/ホスト秒，集中制御方式：7要求数/ホスト秒となる。

(2) 要求発生間隔と回線使用率

要求発生間隔と最大回線使用率の関係を図6に示す。集中制御方式と多数決方式では使用率が更新要求数に比例する。集中制御方式では一部の回線に負荷が集中するため常に多数決方式よりも高い値を示す。

DDACSは論理時刻方式を更新要求がネットワーク内に存在しない場合でも時刻情報を交換するようにシミュレートしている。従って論理時刻方式の最大回線使用率は低負荷時でも高い値を示す。しかしこれは論理時刻方式の本質的な性質ではなく、要求が存在する時のみ時刻情報を交換するように変更し、回線使用率を低くすることも可能である。

高負荷時の論理時刻方式の最大回線使用率は集中制御方式と多数決方式よりも低い。

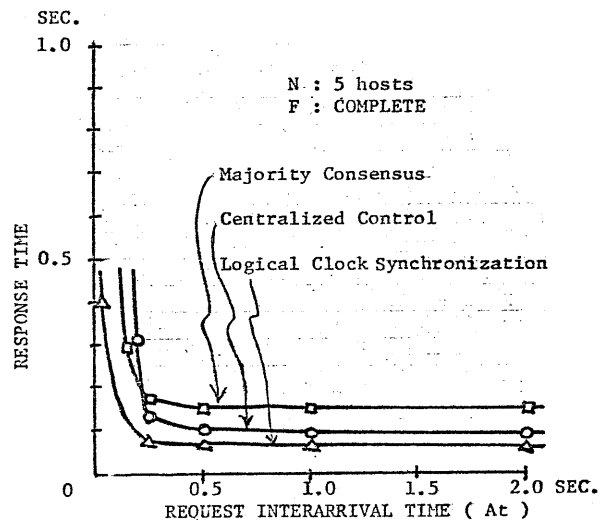


Fig.5 RELATIONSHIP BETWEEN INTERARRIVAL TIME AND RESPONSE TIME

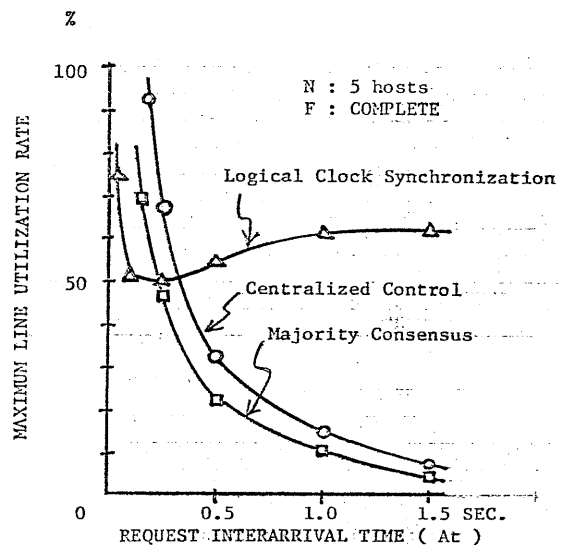


Fig.6 RELATIONSHIP BETWEEN INTERARRIVAL TIME AND MAXIMUM LINE UTILIZATION RATE

(3) ホスト数と応答時間

重複データを保有するホストの個数と応答時間の関係を図7に示す。各ホストは1秒間に1個の割合で更新要求を出すものとしている。集中制御方式と論理時刻方式はホスト数が増加しても応答時間はそれほど増加しない。多数決方式ではホスト数の増加とともに急激に増加する。これは更新要求数が増加して回線負荷が増すのに加えて過半数を得るまでに必要な通信回数が増加するからである。

(4) 応答時間の分布

更新要求発生間隔が500ミリ秒、100ミリ秒、50ミリ秒の時の論理時刻方式の応答時間分布を図8に示す。システムの処理能力の限界に近い要求発生間隔が50ミリ秒の場合を除けば、応答時間の分散は小さい。

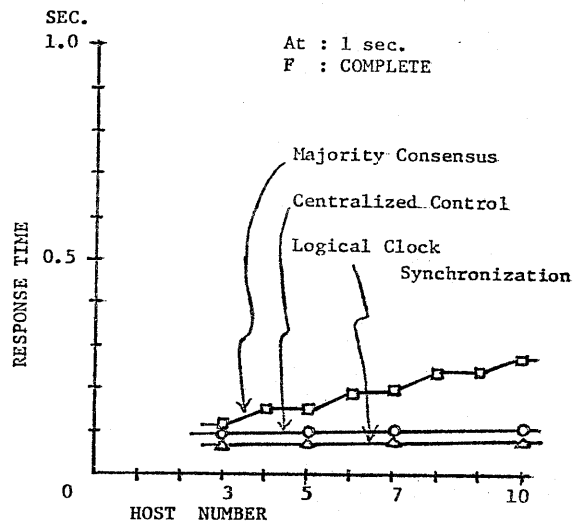


Fig.7 RELATIONSHIP BETWEEN HOST NUMBER AND RESPONSE TIME

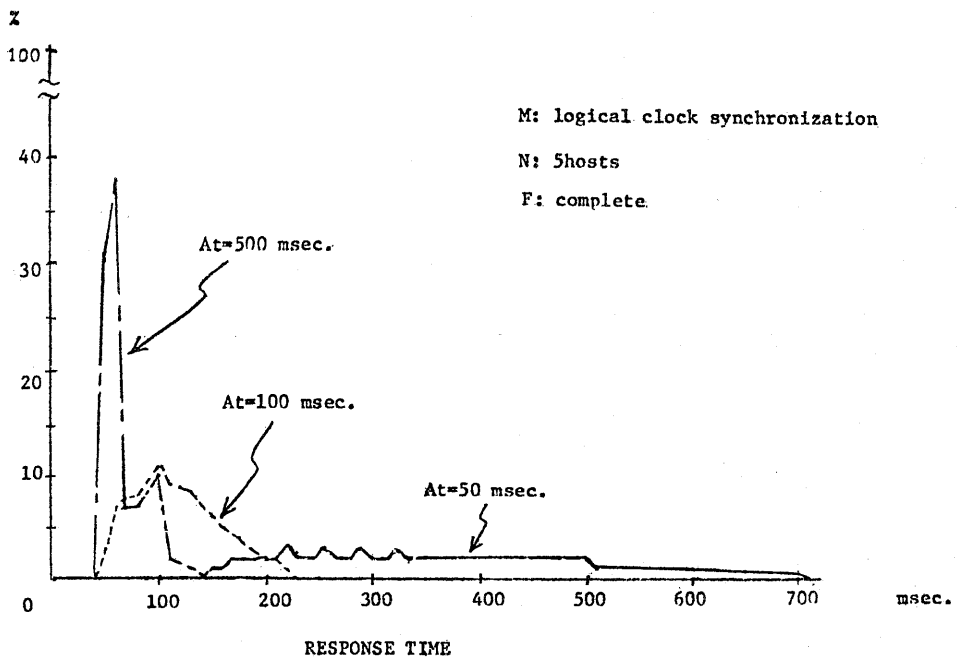


Fig.8 RESPONSE TIME DISTRIBUTION.

4. ネットワーク・トポロジーと性能

ネットワーク・トポロジーが重複データ制御方式の性能に与える影響について検討する。まずネットワークにおいてホストの密集度を示す指標としてホスト間平均距離を定義する。

マスター・ホストを有する集中制御方式では、ホスト間平均距離はマスター・ホストと各スレーブ・ホストの距離の平均であり、以下の式で表わされる。

$$\sum_{\substack{1 \leq i \leq N \\ i \neq j}} \frac{D_{ij}}{N-1}$$

ここで

N : ホスト数

D_{ij} : マスター・ホスト j とスレーブ・ホスト i の距離。

分散型論理時刻方式の場合、全ホスト間の組の距離の平均がホスト間平均距離である。

$$\sum_{\substack{1 \leq i, j \leq N \\ i \neq j}} \frac{D_{ij}}{N(N-1)}$$

ここで

N : ホスト数

D_{ij} : ホスト i とホスト j の距離。

図4のトポロジーに対するホスト間平均距離を表2に示す。マスター・ホストはホスト1と仮定する。集中制御方式では完全グラフ形と星形は同じ特性を示す。

集中制御方式と論理時刻方式の応答時間とネットワーク・トポロジーの関係を図9, 10に示す。

(1) 集中制御方式

応答時間は星形（完全グラフ形）が一番良く、以下クラブ形、トライアングル形、木形の順で続く。この順序はホスト間平均距離と一致する。

表2. ホスト間平均距離

	集中制御方式	論理時刻方式
完全グラフ形	1.0	1.0
星形	1.0	1.8
クラブ形	1.4	1.8
トライアングル形	1.4	2.1
木形	1.7	2.6

クラブ形、トライアングル形、木形の三形態は応答時間曲線がよく似ており、時間の差が小さい。これはホスト間平均距離が1.4～1.7とあまり差がなく、さらに最大回線使用率が、ほぼ同一の要求発生間隔で100%になるためと考えられる。

集中制御方式では星形（完全グラフ形）を除いて、負荷が高くなると応答時間が急激に変化する。

(2) 論理時刻方式

集中制御方式に比べてトポロジーの違いによる応答時間の差が大きい。これは表2のホスト間距離が集中制御方式の場合より大きいからである。

論理時刻方式の応答時間は低負荷時には、完全グラフ形、クラブ形、星形、トライアングル形、木形の順で増加し、高負荷時には完全グラフ形、クラブ形、トライアングル形、星形、木形の順で増加する。低負荷時の順序はホスト間平均距離とほぼ一致するが高負荷時には一致しない。高負荷時は、ホスト間平均距離よりも、回線使用率の高い回線の存在が応答時間に強く影響するた

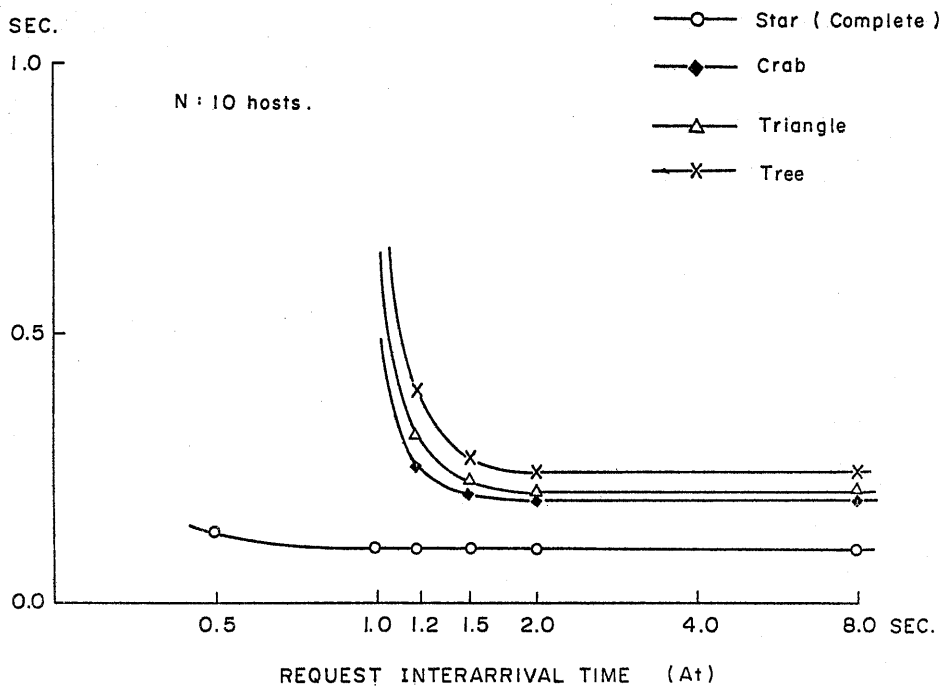


Fig. 9 AVERAGE RESPONSE TIME (CENTRALIZED CONTROL METHOD)

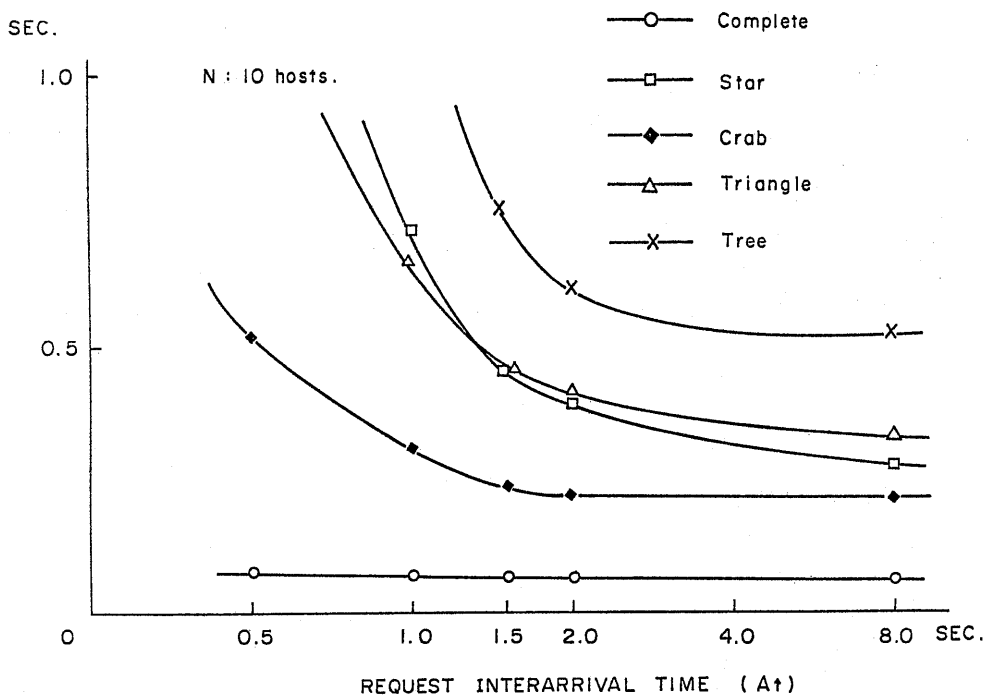


Fig.10 AVERAGE RESPONSE TIME (LOGICAL CLOCK SYNCHRONIZATION METHOD)

めと考えられる。

論理時刻方式では各ネットワーク・トポロジーとも高負荷時における応答時間の悪化はゆるやかであり、回線に対する負荷が分散していることがわかる。

(3) ネットワーク・トポロジーと重複データ制御方式

完全グラフ形の場合、論理時刻方式が集中制御方式より、よい応答時間を示す。

星形と木形の場合は、集中制御方式が論理時刻方式よりも良い。これらのネットワーク・トポロジーが集中型管理方式に合った形態であるのは明らかである。

クラブ形とトライアングル形の場合、低負荷時は集中制御方式が良いが、高負荷時は論理時刻方式が良い性能を示す。特にクラブ形では低負荷時の差は小さく高負荷時の差は大きい。クラブ形は集中制御方式よりも論理時刻方式に合ったトポロジーと言える。

(4) 回線使用率

論理時刻方式では全回線を均等に使用するが、集中制御方式では一部の回線を集中して使用する。図11にトライアングル形の場合の累積回線使用率分布を示す。この分布は、全回線に流れたメッセージ量に対する、各回線に流れたメッセージ量の比率を計算し、値の大きい順に加算したものである。図11により、集中制御方式では全体の1/6の回線で全体の50%の負荷を分担し、全体の1/2の回線で90%の負荷を分担することがわかる。論理時刻方式では全回線が均等に負荷を分担している。

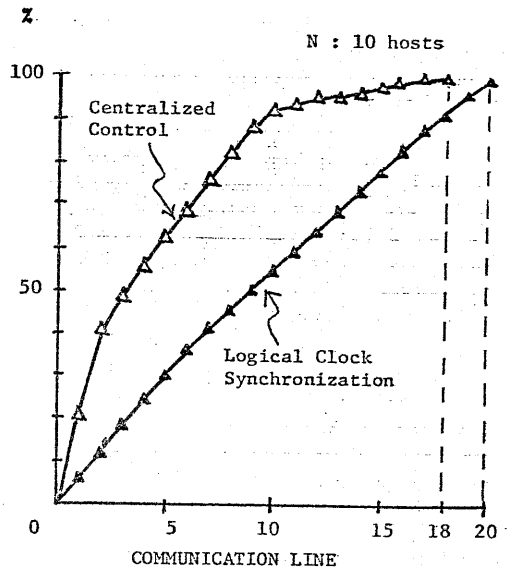


Fig.11 COMMUNICATION LINE LOAD CUMULATIVE DISTRIBUTION (TRIANGLE)

5. おわりに

今までに提案された分散データベースシステムにおける重複データ制御方式の中かゝる、代表的な方式、集中制御方式、多数決方式、論理時刻方式を選んで定量的評価を行なった。シミュレーションによる結果を要約すると次のようになる。

- ・完全グラフ形ネットワークの場合、論理時刻方式は他の方式よりも応答時間が短く、高負荷に対しても強い。
- ・星形・木形のネットワークの場合、集中制御方式は論理時刻方式よりも良い応答性を示す。
- ・クラブ形、トライアングル形の場合、低負荷の時は集中制御方式が論理時刻方式よりも応答性が良いが高負荷時には論理時刻方式の方が良い。特にクラブ形の場合には低負荷の差は小さく、高負荷時の差は大きい。

・集中制御方式では負荷が高くなると応答時間が急激に増加する。一方論理時刻方式の応答時間は負荷の増大にともなってゆるやかに増加する。

・集中制御方式では一部の回線に負荷が集中するが、論理時刻方式では全回線を均等に利用する。

・多数決方式はほとんどの環境下で他の方式よりも応答性が悪い。特にホスト数が増加すると応答時間は急激に増加する。

本資料ではある条件のもとに、5つのネットワーク・トポロジーと重複データ制御方式の性能の関係について検討した。今後はこの条件を更にゆるめた定量的評価が必要である。このためにはシミュレーションによる方法に加えて、他の効率良い評価方式が必要になるであろう。

最後に、本研究の機会を与えて下さった日本電気中央研究所コンピュータシステム研究部祢津部長に感謝いたします。

6. 参考文献

- [1] Garcia-Molina, H., "Performance Comparison of Update Algorithms for Distributed Databases," Technical Note No.143 (Progress Report #1-#5), June 1978.
- [2] Garcia-Molina, H. "Performance Comparison of Two Update Algorithms for Distributed Databases," Proc. 3rd Berkeley Workshop on Distributed Data Management and Computer Networks, Aug. 1978, pp.108-119.
- [3] Kaneko, A., Nishihara, Y., Tsuruoka, K. and Hattori, M., "Logical Clock Synchronization Method for Duplicated Database Control," Proc. 1st Int. Conf. on Distributed Computing Systems, Oct. 1979, to appear.
- [4] Lee, C., "A Regenerative Queueing Model for a Lockout Synchronization Scheme of Multiple-Copied Files," Trends and Applications: 1978, Distributed Processing, pp.63-68.
- [5] Thomas, R. H., "A Solution to the Concurrency Control Problem for Multiple Copy Data Bases," Proc. COMPCON 1978 Spring, pp.56-62.
- [6] 服部, 西原, 金子, 鶴岡「分散データベースにおける重複データ制御方式と論理時計」情報処理学会 第15回データベース管理システム・第2回分散処理システム合同研究会, 1979
- [7] 西原, 金子, 鶴岡, 服部「分散型データベースシステムにおける重複データ制御方式」電子通信学会 EC, 78-6 1978