分散処理システム３－９
（１９７９．１１．２２）

分散データベースにおけるディレクトリ管理方式の評価

Cost Evaluation of Directory Management Schemes for Distributed Database Systems

山崎 晴明　足田 定幸　吉田 勇　松下 温
Haruaki YAMAZAKI　Sadayuki HIKITA　Isamu YOSHIDA　Yutaka MATSUSHITA
沖電気工業株式会社
OKI Electric Industry Co., Ltd.

1　はじめに

分散データベースシステムは地理的に分散配置されたデータベースを利用者に対して論理的に１つのデータベースとして提供する特色をもっている[1][2][3]。データベースの利用者はネットワークのどこからでも地理的に分散配置されたデータベースへアクセスすることができ、利用者はあたかもデータベースが自分のサイトにあるようにデータベースを検索または更新することができる。すなわち分散データベースシステムはその利用者に対してネットワーク内のデータベースの共用を実現させることができる。このようなサービスを実現するため、分散データベースシステムはデータがネットワーク内のどのサイトに保持されているかという対応表すなわちディレクトリ情報を持っている[7][8]。データベースのアクセスに対してアクセスすべきサイトの決定は、このディレクトリによってなされることになる。

分散データベースシステムは様々なアプリケーションに適用されると考えられるため、アプリケーションごとに異ると予想される様々な特性を持つトラヒックを効率良く処理できねばならない。すなわちアプリケーションによっては検索要求と更新要求の割合が大きく変化したり地域によってアクセスされるデータに偏りがあったりする（トラヒックのローカリティ）。分散データベースシステムの設計のため、この様なディレクトリ管理方式が報告されている[1][2][3][4]。しかしながらこれらの報告では、その評価のパラメタとして検索要求と更新要求の割合または更新要求に必要とするプロトコルのオーバヘッド等を使った評価しか行っていない。

本稿では分散データベースシステムでは重要と考えられるトラヒックのローカリティを使って最適なディレクトリ管理方式を評価する。第２節では分散データベースシステムにおけるディレクトリ管理方式を整理するとともに評価モデルを示し、第３節でそのモデルの通信コストをディレクトリアクセスの通信回数で解析するとともに数値例を示し、最適なディレクトリ管理方式について議論する。

2　ディレクトリ管理方式

本節では分散データベースシステムにおけるディレクトリ管理方式を整理し、本論文で比較検討する３つのモデルを示す。

2.1　ディレクトリ

分散データベースシステムでは、その利用者はデータのネットワーク上の物理的位置を知らなくてもデータベースへアクセスすることができる。このため分散データベースシステムはディレクトリ管理が必要になる。図１に分散データベースシステムの概念構成を示す。

例えばサイト１のユーザは目的とするデータを格納しているサイトのサイト番号を知らなくてもシステムへアクセスすることができる。このときのデータとその格納サイト番号との対応

づけはディレクトリによってなされる。このようにデータベース利用者が、データの物理的な格納場所を知ることなくデータへアクセスできることを"Invisibility"または"Location Transparency"と呼ぶ[1]。

ディレクトリ管理方式を議論するため、用語の定義をリレーショナルモデルに基づいて行う。リレーションの分割単位をフラグメントと呼ぶ。これはタプルの集合であり分散データベースのサイトへ格納される単位となる。あるサイトが格納しているフラグメントを示すテーブルを"ローカルディレクトリ(local directory)"と呼ぶ。従って"ローカルディレクトリ"には他サイトのフラグメント情報は含まれない。分散データベースを分割する、地理的に隣接したサイトの集合をゾーン(zone)と呼ぶ。分散データベースシステムは複数のゾーンにより構成することができる。このゾーンに存在するすべてのサイトと、そのサイトが保持するフラグメントの対応を示すテーブル情報を"ゾーンディレクトリ(zone directory)"と呼ぶ。"ゾーンディレクトリ"には他ゾーンの情報は含まれていない。分散データベースシステムに存在する"ゾーンディレクトリ"の数は、ゾーン数と一致する。またゾーン数が1である分散データベースシステムの"ゾーンディレクトリ"を、特に"ネットワークディレクトリ"と呼ぶ。さらにゾーン内に収容されているすべてのフラグメントとゾーンとの対応を示す情報を"グローバルディレクトリ(global directory)"と呼ぶ。分散データベースシステムにおいて相異なる"グローバルディレクトリ"は2つ以上存在しない。以上の用語を使用してディレクトリ管理方式を表1に示す。

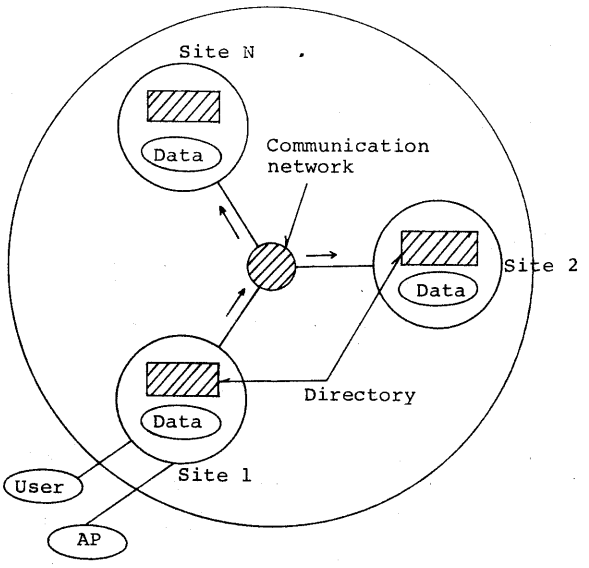

図1. A general description of a distributed database

ディレクトリ管理方式は大別すると、ディレクトリを冗長に持つか否かで冗長型,非冗長型に分けられる。例えば冗長型集中制御方式では各サイトが保持する"ローカルディレクトリ"とセンタサイトが保持する"ネットワークディレクトリ"とでは情報の重複がみられる。これに対し非冗長型の集中制御方式では、システムに"ネットワークディレクトリ"のみが存在し重複するディレクトリ情報はない。

通信コストという面から考えると非冗長型のアプローチでは良い評価を得られないのは明らかである。例えば非冗長型集中制御方式では、ユーザからの検索要求のたびにディレクトリを参照する通信が発生する。また非冗長型分散制御方式では、自身が保持していないフラグメントに対するアクセスはすべてブロードキャストによって処理されることになる。非冗長型ゾーン制御方式でも"グローバルディレクトリ"がないため、自ゾーンにないフラグメントを参照する場合はゾーン間でブロードキャストを行わねばならない。もち論、非冗長型ではシステム全体のディレクトリ容量(メモリコスト)は冗長型に対して小さくてすむが一般的にメモリコストは通信コストに比べて小さく、さらにこの傾向はメモリ技術の進歩により年々助長されると予測される。従って本稿では焦点を通信コストの節減に絞り、最適なディレクトリ管理方式を評価するものとする。このため非冗長型のア

表1　Classification of directory management schemes

Classification			Management schemes
Redundant type	Centralized control		Each site has its own "local directory". A unique center site within a network has the "network directory".
	Distributed control		Every site has the "network directory". [1] [3] [4]
	Zoned control	A	Each site has its own "local directory". Each unique center site within a zone has its own "zone directory" and the "global directory.
		B	Each site has its own "local directory". Each unique center site within a zone has "zone directories" of all zones. [4]
non-redundant type	Centralized control		No site has a "local directory". A unique center site within a network has the "network directory". [4]
	Distributed control		Each site only has its own "local directory".
	Zoned control		No site has a "local directory". Each unique center site within a zone has its own "zone directory".

プロトコルは評価の対象からはずす。また集中型のアプローチは、通信コストの面では良い結果を生ずることが予想されるが、信頼性の低下、トラヒックの過度の集中といった欠点があり分散システム特有の長所を持つことができない。従って本稿では、これを評価の対象から除外して話を進める。

本論文では表1における冗長型制御方式のうち、分散制御方式，ゾーン制御方式A，ゾーン制御方式Bの3つの方式について評価する。以下に3つの方式の概要を述べる。

(1) 冗長型分散制御方式

ネットワーク内のすべてのサイトが、すべて同一の"ネットワークディレクトリ"を持っているため、ディレクトリ参照に伴う通信は生じない。しかしながらディレクトリの更新が発生した場合は、更新要求が全サイトにブロードキャストされる。

(2) 冗長型ゾーン制御方式A

ディレクトリへの参照要求は次の3つに分類される。

(a) アクセス要求を受け付けたサイトの"ローカルディレクトリ"だけ参照することにより処理できる要求。

(b) ローカルディレクトリでは処理できず、アクセス要求を受け付けたゾーンの"ゾーンディレクトリ"を参照することにより処理できる要求。

(c) 要求を受け付けたサイトの"ローカルディレクトリ"および"ゾーンディレクトリ"では処理できず"グローバルディレクトリ"および他ゾーンの"ゾーンディレクトリ"を参照することにより処理できる要求。

これら3種類の参照要求を図2に示す。

一方ディレクトリ更新要求は次の2つのタイプに分類される。

(a) "ローカルディレクトリ"と"ゾーンディレクトリ"を更新する要求で、通信はゾーン内に閉じて行われる。

(b) "ローカルディレクトリ","ゾーンディレクトリ"および"グローバルディレクトリ"を更新する要求で、通信はネットワーク全体で行われる。

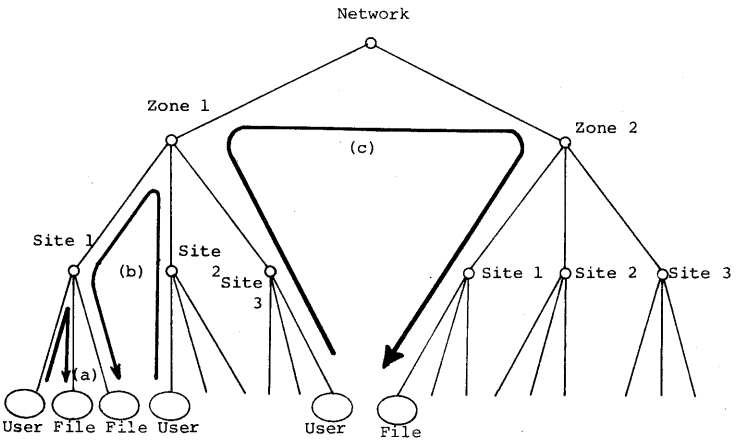

図 2 Three types of directory retrieval in a zoned control scheme-A

(3) 冗長型ゾーン制御方式B

この方式では各ゾーンのセンタサイトが他ゾーンの"ゾーンディレクトリ"を持っているため、ディレクトリの参照はすべて自ゾーンに閉じた通信で行われる。ただし、この方式では各ゾーンが持つ"ゾーンディレクトリ"が共有データとなっているため"ゾーンディレクトリ"の更新はすべてブロードキャストによって処理されることになる。

ディレクトリ管理に要する通信コストは、ディレクトリ要求の発生頻度、ディレクトリ要求のうち更新要求と検索要求の割合およびディレクトリ要求トラヒックのローカリティ等に強く依存する。ディレクトリ管理に伴う通信コスト、処理コストおよびメモリコストを含めたトータルコストの評価はChu[4]によって報告されている。しかしながら、この報告では分散データベースシステムで通信コストに大きな影響を与えるトラヒックのローカリティの考慮が払われていない。本稿ではトラヒックのローカリティに着目し、これがディレクトリ管理に要する通信コストにどのような影響を与えるかを考察している。

なお通信コストを適切に反映するものとして本稿では通信回数を評価値として選んでいる。これは多くのパケット交換網での料金体系では通信コストが伝送したパケット数にほぼ比例するという事実に基づいているからである[1][5][6]。

3 ディレクトリ管理方式の解析

本節では前節で紹介した3つのディレクトリ管理方式について通信コストを評価し、最適なディレクトリ管理方式について議論する。

3.1 パラメタの定義

通信コストの解析を行うため以下に示すパラメタを定義しておく。

n：分散データベースシステム内のサイト数
m：分散データベースシステム内のゾーン数
k："グローバルディレクトリ"または"ネットワークディレクトリ"を更新するために必要な通信回数
p：ディレクトリアクセス要求が更新要求である確率
α：要求するディレクトリが、要求が発生したサイトで得られ他サイトとの通信を必要としない確率
β：ディレクトリ検索要求に対して、ディレクトリが要求が発生したゾーンで得られ他ゾーンとの通信を必要としない確率

γ：ゾーン制御方式でディレクトリ更新要求が"グローバルディレクトリ"を更新しない確率

ここでαおよびβは、それぞれサイトおよびゾーンのローカリティを表わしている。q_{ij}はサイトiで発生した要求がサイトjのディレクトリを要求する確率とする。サイトiでのローカリティα_iは次式で表わされる。

$$\alpha_i = \frac{q_{ii}}{\sum_{j=1}^{m} q_{ij}}$$

本解析では、すべてのiについて$\alpha_i = \alpha$と仮定する。同様にゾーンにおけるローカリティもβと仮定する。

3.2 通信コストの解析

(1) 冗長型分散制御方式（方式D）

(a) 更新

kの値は分散データベースの様々な同期制御プロトコルに従い$1 < k \leq 5$の値をとる。この方式では、全データベースサイトが"ネットワークディレクトリ"を重複して持つため、更新に要する通信回数は$k(m-1)$である。

(b) 検索

一方ディレクトリの検索では、全サイトが"ネットワークディレクトリ"を持つため通信回数は0である。

(c) 通信回数の期待値（E_d）

ディレクトリの検索には通信を必要としないため、E_dは次式で表わされる。

$$E_d = k(m-1)p \qquad (1)$$

(2) 冗長型ゾーン制御方式A（方式A）

(a) 更新

ゾーンディレクトリ更新時のセンタサイトと要求発生サイト間の必要な通信回数をlとする。

もし"グローバルディレクトリ"が更新されるとすればセンタサイト間では$k(m-1)$回の通信が必要である。従って更新要求を受け付けたサイトがセンタサイトなら$k(m-1)$、そうでなければ$k(m-1)+l$の通信が必要である。

ディレクトリ更新に必要とする通信回数の期待値A_uは次式で表わされる。

$$A_u = \frac{m}{n}k(m-1)(1-\gamma) + (1-\frac{m}{n})\left[l\gamma + \left\{k(m-1)+l\right\}(1-\gamma)\right] \qquad (2)$$

ここで$\frac{m}{n}$は要求を受け付けるサイトがセンタサイトである確率、$(1-\frac{m}{n})$はセンタサイトでない確率とする。

(b) 検索

もし検索要求が"ローカルディレクトリ"に限られた場合通信回数は0である。一方検索要求がゾーンに限られた場合、通信回数は2である。これは要求発生サイトとセンタサイト間の検索要求と応答である。もし検索要求が"グローバルディレクトリ"をも必要とするなら、さらに2回の通信が必要になる。これは2つのセンタサイト間の検索要求とその応答である。従って検索要求に対する通信回数の期待値A_rは次式で与えられる。

$$A_r = 2\cdot\frac{m}{n}(1-\beta) + (1-\frac{m}{n})\left\{4(1-\beta)+2(\beta-\alpha)\right\} \qquad (3)$$

(c) 通信回数の期待値

通信回数の期待値E_Aは次式で与えられる。

$$E_A = \left[\frac{m}{n}k(m-1)(1-\gamma) + (1-\frac{m}{n})\left[l\gamma + \left\{k(m-1)+l\right\}(1-\gamma)\right]\right]\cdot P + \left[2\frac{m}{n}(1-\beta) + (1-\frac{m}{n})\left\{4(1-\beta)+2(\beta-\alpha)\right\}\right](1-P) \qquad (4)$$

(3) 冗長型ゾーン制御方式B（方式B）

(a) 更新

すべてのセンタサイトはすべての"ゾーンディレクトリ"を重複して持っている。ディレクトリ更新時の通信回数の期待値B_uは(2)式のγを0にして次式で与えられる。

$$B_U = \frac{m}{n}k(m-1) + \left(1-\frac{m}{n}\right)\left\{k(m-1)+l\right\} \quad (5)$$

(ℓ)検索

検索要求に対する通信回数の期待値 B_r は(3)式においてβを1として次式で与えられる。

$$B_r = 2\left(1-\frac{m}{n}\right)(1-\alpha) \quad (6)$$

3.3 数値例

本節では各パラメタに以下に示す条件を与えて評価する。

(a) $n=12$

(b) $k=5$　図3にプロトコル例を示す。

(c) $l=4$　図4にプロトコル例を示す。

(d) すべてのしおよびよ(しきよ)について確率 $\beta_{ij}=\beta$ としてβを以下のように近似する。

$$\beta = \alpha + \frac{\frac{m}{n}-1}{n-1}(1-\alpha)$$

この式で最初の項は要求するディレクトリが要求を受け付けたサイトにある確率であり、第2項目はディレクトリが要求を受け付けたサイトにはないがそれが属するゾーンにはある確率である。

(e) γは次式で表わされるものとする。

$$\gamma = \frac{\frac{m}{n}-1}{n-1}$$

もし $n=m$ なら各サイトは"グローバルディレクトリを持つことになりγ=0となる。$m=1$ ならネットワークは1つのゾーンから成り"グローバルディレクトリ"は存在しない。このときγ=1となる。

図5に $n=4$, $\alpha=0.3$ および0.8における更新確率Pに対する通信回数の期待値を示す(E_d, E_a, E_ℓ)。図5よりα=0.3のとき E_dとE_a は $P=0.06$, E_dとE_ℓ は $P=0.02$, E_aとE_ℓ は $P=0.34$ で交わる。従って方式Dは $0 \le P \le 0.02$ で, 方式Bは $0.02 < P \le 0.34$ で, 方式Aは $0.34 < P \le 1$ でそれぞれ最も効率が良い方式であると言える。同様にα=0.8について、方式Dは $0 \le P \le 0.01$, 方式Bは $0.01 < P \le 0.14$, 方式Aは $0.14 < P \le 1$ でそれぞれ最も効率が良いと言える。

分散制御方式(方式D)は更新確率Pが極めて小さい場合に有利であると言える。これは各サイトがネットワークディレクトリを保持しているため、要求を受け付けたサイトは他サイトへ問合せをする必要がないためである。しかしながら方式Dは、更新確率Pの値が増大するとサイト間の通信が急激に増大し、他方式に比べ不利となる。一方、方式Aおよび方式Bはディレクトリ更新時、方式Dより通信回数が少いため、更新確率が増大するほど方式Aおよび方式Bが方式Dより有利となる。方式Aと方式Bを比較すると、Pの小さい値に対しては方式Bが

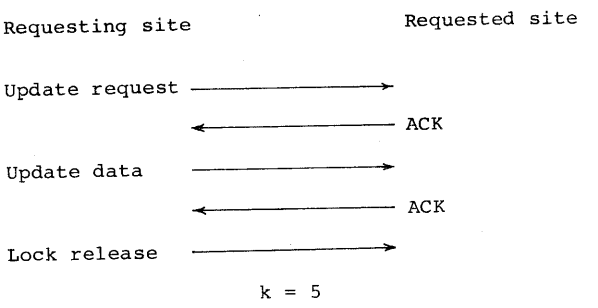

図3 The procedure for updating the global directory or the network directory

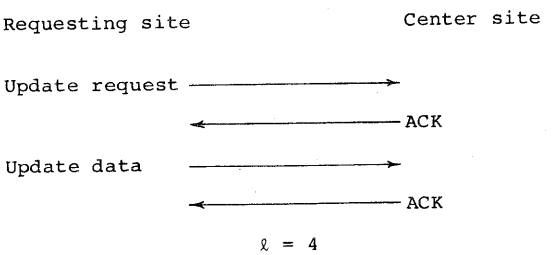

図4 The procedure for updating the zone directory

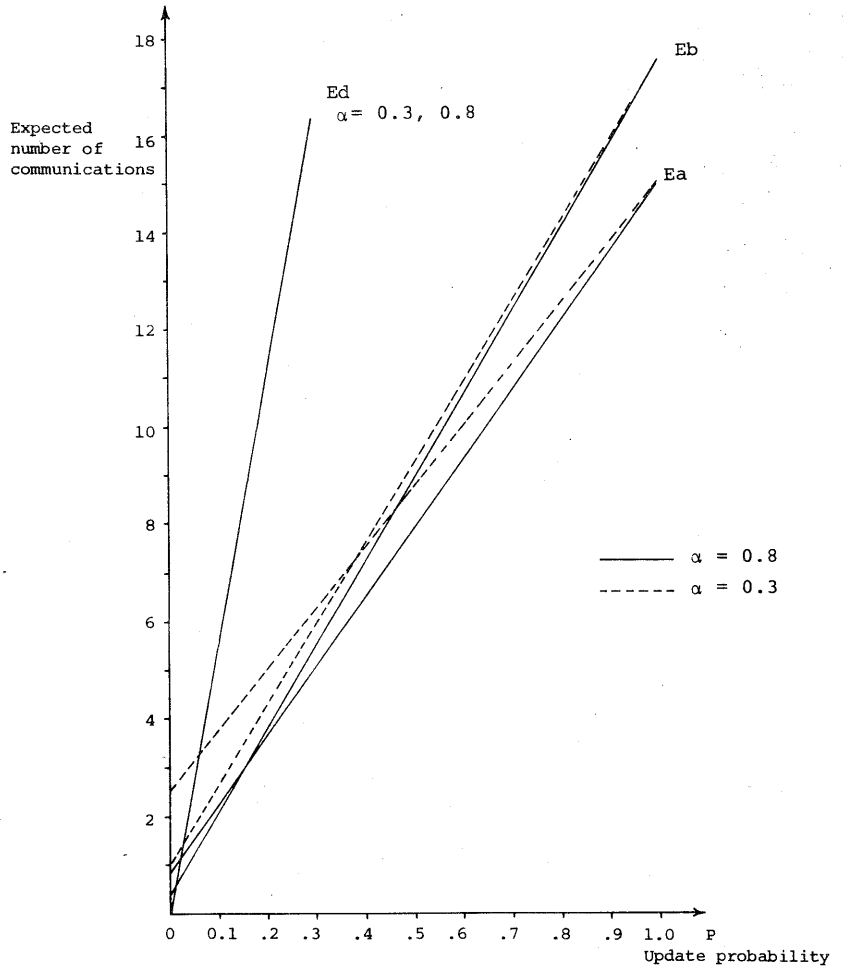

図5 The expected number of communications v.s. the update probability for m = 4

有利である。これは方式Bでは通信がゾーン内に閉じており、方式Aではゾーン間の通信が必要となるためである。一方更新確率Pの大きな値に対して方式Aが有利である。また方式Aおよび方式Bではαの値が大きくなると通信回数は減少する。さらに方式Aは方式Bに比べαの値にセンシティブであることがわかる。

図6はα=0.3のとき $m=3, 4$ についてのPと通信回数の期待値の関係を示す。E_aに関する$m=3$と$m=4$の2つの直線はP=0.04で交わる。E_dに関する2つの直線はP=0.03で交わる。従って小さいPの値に対してネットワークは3つより4つのゾーンに分割した方が良いことになる。一般的な特性として、センタサイトを増加させると検索要求における通信回数は減少する。一方ゾーン数を増加させると更

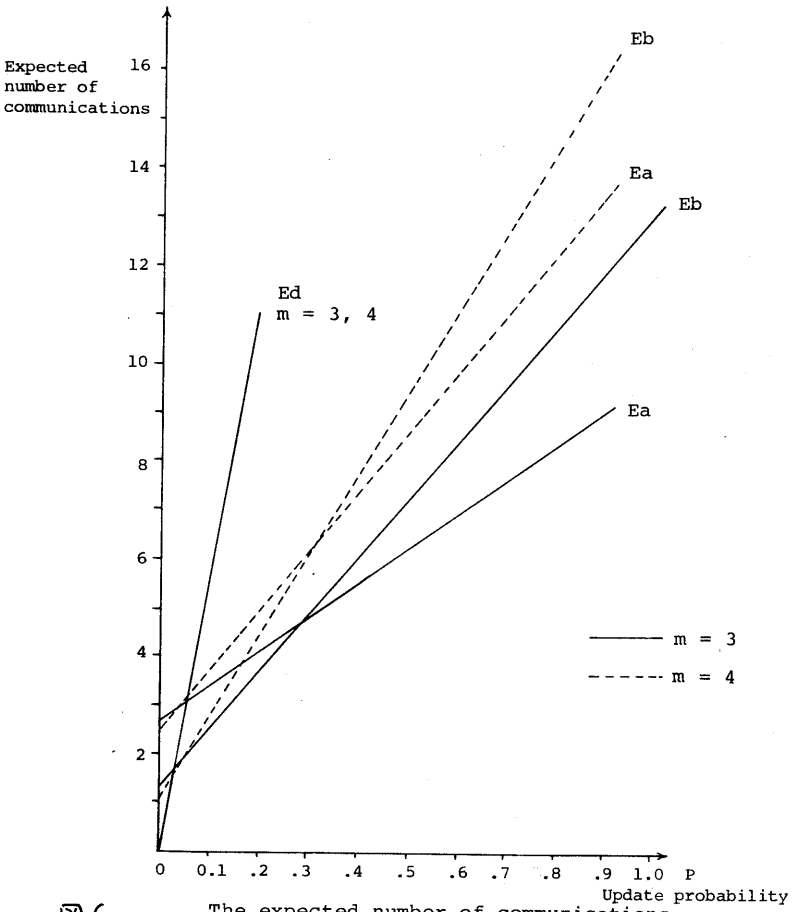

図6 The expected number of communications v.s. the update probability for $\alpha = 0.3$

新要求における通信回数は増大する。図6で示した特性は、ゾーン数を増加させても検索による通信回数の減少が更新による通信回数の増大より影響が大きいことを示している。

図7では$\alpha=0.3$における更新確率Pの0.3および0.4について、ゾーン数と通信回数の期待値の関係が示されている。方式Aと方式Bでは、更新確率Pが大きくなると方式Aが方式Bより有利となる。すなわちゾーン数mが3, 4, 6では更新確率Pが0.3や0.4でも方式Aが方式Bよりすぐれているということが言える。しかしながらm=2では更新確率Pが0.4において方式Bがすぐれていると言える。このようにゾーン数が少なくなると、方式Bが有利である更新確率Pの領域が広がると言える。

4 結論

本論文では分散データベースシステムにおいて考えられる3種類のディレクトリ管理方式について、通信コストから評価した。この結果、一般的傾向としてディレクトリ更新確率がごく小さい範囲では冗長型分散制御方式(方式B)が有利であり、更新確率がある程度以上になると冗長型ゾーン制御方式A(方式A)が最も有利となり、両方式の中間に冗長型ゾーン制御方式B(方式B)が有利となる更新確率の領域が存在することがわかった。さらにトラヒックのローカリティが大きくなるほど方式Aが有利となる更新確率の範囲が広がることがわかった。

分散システムの大きなねらいは、あるサイトで発生するアクセス要求はできるだけそのサイ

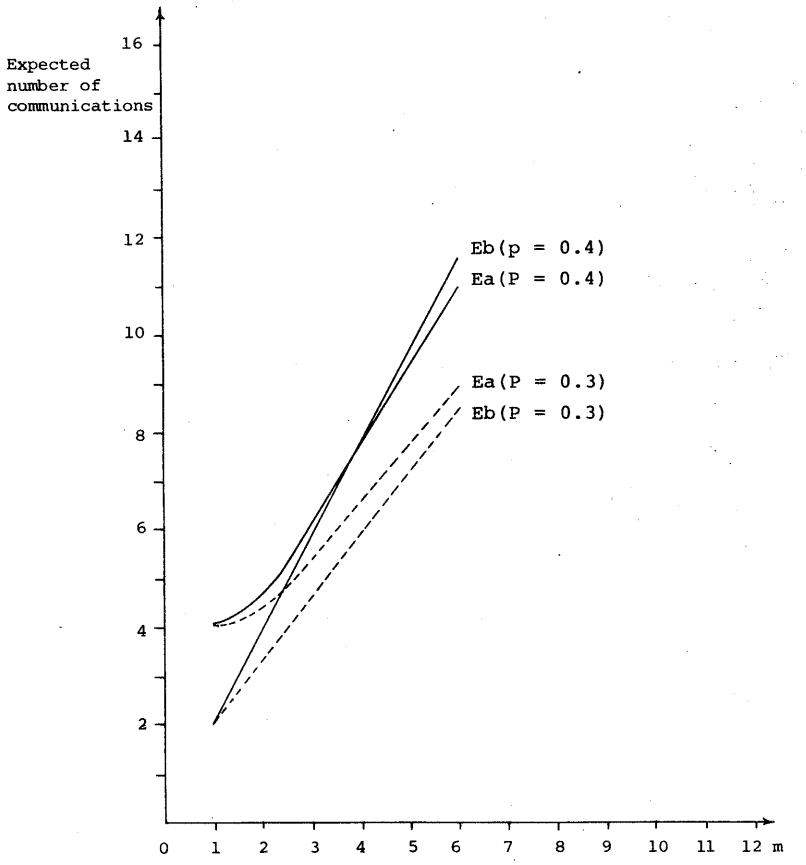

図7 Expected number of communications v.s. number of zones for α = 0.3

ト内で処理できるようにデータを配置して通信コストを低減させ、レスポンスタイムを向上させることにある。分散データベースシステムのアプリケーションを考える場合ディレクトリアクセスのローカリティ利用も重要な条件であり、その値も比較的大きな値を持つものと期待できる。従って"ローカルディレクトリ","ゾーンディレクトリ"および"グローバルディレクトリ"の3種のディレクトリを使用するゾーン制御方式Aは分散データベースシステムにおいて最も広い適応領域のある方式と言える。

なお今回の解析では通信コストに焦点をあてたが、処理コストやメモリコストをも含めた総合的評価およびレスポンスタイム等の評価も必要と考えられる。今後はこれらについて検討していく予定である。

[参考文献]

[1] Rothnie JB, Goodman N,
"A Survey of Research and Development in Distributed Database Management"
Proc. Int. Conf. Very Large Data Bases, 1977, PP. 48-62.

[2] Stonebraker M, Neuhold E,
"A Distributed Data Base Version of INGRES"
Proc. Second Berkeley Workshop, 1977, PP. 17-36.

[3] Rothnie JB, Goodman N,
"An Overview of the Preliminary Design of SDD-1: A System for Distributed Data Bases"
Proc. Second Berkeley Workshop, 1977, PP. 39-57

[4] Chu W W ,
"Performance of File Directory Systems for Data Bases in Distributed Networks",
Proc. AFIPS 1976 NCC, Vol. 45, PP. 577-587.
[5] Thomas RH ,
"A Solution to the Concurrency Control Problem for Multiple Copy Data Bases",
Proc. Spring COMPCON ,1978 , PP. 56-62.
[6] Thomas RH,
"A Majority Consensus Approach to Concurrency Control",
ACM TODS 1979, Vol. 4 No. 2, PP. 180-209.
[7] Yamazaki, Hikita, Yoshida, kawakami, Matsushita,
"A Hierarchical Structure for Concurrency Control in a Distributed Database System",
Proc. 6th Data Communications Symposium,
1979 , (to be published).
[8] 山崎, 疋田, 松下他,
"分散データベースにおける同期制御のための階層型プロトコル",
情報処理学会 . 分散処理システム研究会, 1979,
第2回研究会資料.