

分散型電子メールアドレス検索システム

岸 健一† 木村 成伴, 海老原 義彦‡

†筑波大学 第三学群 情報学類 ‡筑波大学 電子・情報工学系

電子メールアドレス検索システムとしては WHOIS サービスなどが知られているが、これらのシステムでは電子メールなどの情報を保持するデータベースが検索サーバによって集中的に管理されるため、データベースの管理が容易でなくなる他、登録されているデータが必ずしも最新の情報を反映しているとは限らないという問題点がある。そこで本論文では、データベースを各ドメインで分散的に管理する電子メール検索システム DMS を提案する。分散化されたことにより、データベースの管理が簡単になる反面、検索のための通信量が増大することが問題となるが、各ドメインを属性情報の木構造によって関連付けることにより、通信量の軽減を図っている。

Decentralized Mail-address Search System

Ken-ichi Kishi† Shigetomo Kimura, Yoshihiko Ebihara‡

†College of Information Sciences, Third Cluster of Colleges, Univ. of Tsukuba

‡Institute of Information Science, Univ. of Tsukuba

There are many E-mail address directory services such as WHOIS service. Every server of this service has huge database including the information of E-mail addresses and gives users addresses searched from it. However centralized maintaining causes difficulty to manage the database. In addition, some informations in database may not be guaranteed to be up-to-date. This paper proposes the new E-mail address directory service DMS(Decentralized Mail-address Search), which distributes the database to each domain for simplification of the management and rapid renewal of the database. In the new system, the traffic for searching will be increased. But we can reduce the traffic by relating the database of each domain with tree structure.

1 はじめに

近年のコンピュータ・ネットワークの普及に伴い、様々なネットワークサービスが多くのユーザに利用されている。そのようなサービスの一つに電子メールサービス [1] が存在するが、電子メールを利用する際には、メールアドレスを送付

する相手の電子メールアドレスを事前に知る必要がある。アドレスを知らない場合や忘れてしまった場合に必要となるのが電子メールアドレス検索システムである。

電子メール検索システムとしては WHOIS システム [2] が最も良く知られている。これは TCP (Transmission Control Protocol) 上で動

作するシステムである。現在では、アメリカでは InterNIC[3] が、日本においては JPNIC (Japan Network Information Center) ¹がこのサービスを提供している。サービスを提供するサーバの最新の情報は MIT(Massachusetts Institute of Technology) より発行されている²。

WHOIS システムでは、各サーバが電子メールなどの情報をデータベースとして保持しており、ユーザの要求に応じてデータベースの検索が行われる。データは各ユーザが自主的に登録する方式を採用しており、データの改変、削除なども各ユーザが所定の手続きによって行う。

しかしながら、この WHOIS システムには問題点がいくつか挙げられる。まず、データベースが検索サーバによって集中的に管理されるため、データベースは膨大なものとなり、その管理が容易でなくなることである。また、データの登録が自己申告によるものであることから、全てのユーザが検索対象となるとはいえず、更に、登録されているデータが最新の情報を反映しているとは限らない。

後者の問題を緩和するため、WWW(World Wide Web) 上のシステムの一部³ ⁴や WHOIS システムの拡張⁵ ⁶では、自動的に電子メールアドレスの情報を収集するシステムが見られる。これらのシステムでは、ネットワークニュース等のネットワークサービスから得られるデータを参照したり、finger サービスを利用して各ホストからデータを集めるといった方法を採用している。しかしながら、この方式では本人の知らぬうちにデータベースにデータが登録されてしまうため、プライバシーの保護という観点からは好ましいとは言えない。

そこで本論文では、データベースを各ドメインで分散的に管理する電子メールアドレス検索システムを提案する。分散化されたことによりデータベースの管理が簡単になる反面、検索のための通信量が増大することが問題となるが、各ドメインを属性情報の木構造によって関連付けする [4]

¹<http://www.nic.ad.jp>

²<ftp://sipb.mit.edu/pub/whois/whois-server.list>

³internet address finder <http://www.iaf.net>

⁴Four11 Corp. <http://www.four11.com>

⁵Horton <ftp://ftp.inria.fr:/network/misc/horton-1.8.shar.Z>

⁶Name Lookup Service

<ftp://csus.edu:/pub/nls/nls-1.21.tar.Z>

ことにより、通信量の軽減を図っている。

このシステムの特徴は、

- データを各ドメイン(ホスト)が持つことにより、データ更新が迅速に行える。
- 個々のデータベースのサイズが小さくて済むため、管理が容易になる。
- データの内容を個人、あるいはホスト単位で決められるので、プライバシーの保護が容易である。

といったことが挙げられる。

2 分散型電子メールアドレス検索システム DMS

この節では、分散型電子メールアドレス検索システム DMS(Decentralized Mail-address Search) (図 1) について説明する。

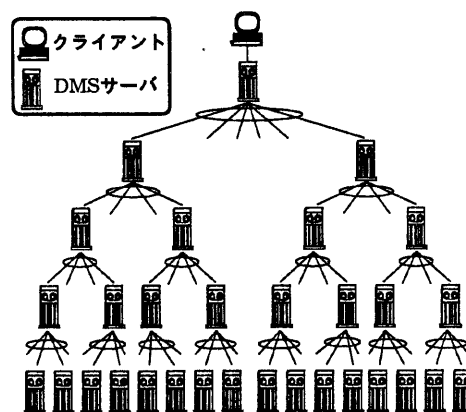


図 1 DMS システムの概要

2.1 DMS サーバ

DMS は図 1 に示されるように複数のサーバが木構造を形成して構成されている。各サーバは、ホストデータベース (以下 HDB)、ユーザデータベース (以下 UDB) の 2 つのデータベースを保持している。前者には木構造の自分の直属の子のサーバの情報が、後者にはそのサーバを持つドメイン検索者情報が格納されている。両者の詳細については次節に述べる。

クライアントはまず検索木の根のサーバに接続し、子のサーバを特定するためのホストキーと被検索者の名前を示すユーザキーを与える。そして、これらキーを基にした検索結果から該当のサーバに接続し、このサーバに対して同様の操作を繰り返す。

木構造を取るという性格上、上位のサーバは下位のサーバと比べて負担がかかりやすくなる。上位のサーバに過度の負担がかかり、システムそのものの破綻が起ころを未然に防止するために、まず上位のサーバは1つに限定せず、ミラーサーバを複数持つことにより上位サーバの負担を減らす。また、サーバ・サーバ間とサーバ・クライアント間の通信プロトコルを同一のものとし、クライアントは根のサーバのみならず、どのサーバにもアクセスできるようにする。これによって極力上位サーバへのアクセスの負担を軽減させる。

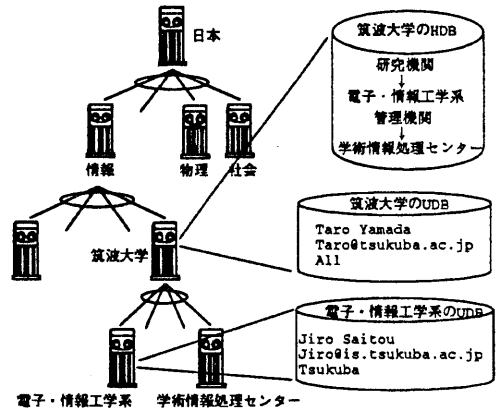


図2 DMSシステムのデータベース

2.2 データベース

HDB

HDBには自分の子のサーバのホスト名(ホストのアドレス)と、子のホストの属性が格納され、このデータベースはクライアントの検索者が与えたホストキーを基に参照される。ホストキーは検索者すなわち人間が自然言語によって与えるので、ミスタイプや自然言語の曖昧さによる検索洩れを防ぐために、検索エンジンに曖昧検索[5, 6]を導入する必要があるが、本論文ではこれに関する議論は行わない。

UDB

UDBにはユーザの氏名、電子メールアドレス、アクセス制限情報の3つのデータが格納されている。ここで、アクセス制限情報とは情報を提供しても良い相手かどうかを判定するための情報であり、これによってプライバシーの保護が行われる。また、検索の際には、検索者の与えたユーザキーに基づいてこのデータベースの氏名と電子メールアドレスを参照する。ここでもHDBと同様に曖昧検索を導入する必要がある。

図2では「筑波大学」のDMSサーバにHDBとUDBが格納されている。このサーバのHDBには「電子・情報工学系」と「学術情報処理センター」のサーバ情報とサーバのアドレスが格納されている。「筑波大学」のUDBには、「Taro

Yamada”の氏名とその電子メールアドレス、また、アクセス制限なしという情報が格納されている。「電子・情報工学系」はサーバ木の葉になるのでHDBは存在しない。UDBには「Jiro Saitou”の氏名とその電子メールアドレス、また、アクセスは「Tsukuba」に関係した人のみという制限情報が格納されている。

2.3 動作方法

クライアントは根のサーバに接続し、決められた手続きを踏んだ後、ホストキーとユーザキーをサーバに与える。サーバはホストキーを基にHDBを参照する(図3)。

HDBを検索した結果は、次の4つの場合に分ける事ができる。

1. 該当する子のサーバは1つだけであった。
2. 該当する子のサーバは複数あった。
3. 該当する子のサーバはなかった。
4. ホストキーが存在しない時。

1の場合、検索にマッチしたホストキーを削除し、残ったホストキーとユーザキーを子のサーバに渡して、検索を依頼する。

2の場合、検索された全ての子のサーバに対して1の場合と同様な操作を行う。ただし、このようなマルチキャストを許可するかないかの判断は各サーバ毎に決められる。このため、マルチキャストが許可されないサーバの場合はその旨を

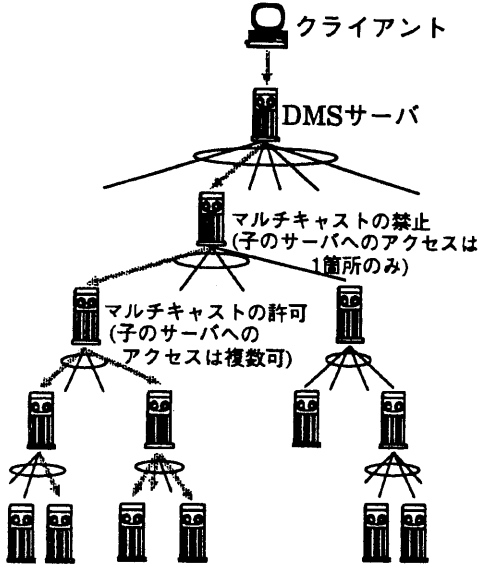


図3 DMSシステムの動作方法

親のサーバに伝える。これにより、検索データが木構造を成したサーバに過度に広がる事を防止する。

3と4の場合も2と同様である。ただし、3の場合は検索キーが削除されずに検索が続けられる点が、4の場合にはHDBの検索を行わないが、ユーザキーを用いてUDBの検索を行い、該当するデータを親サーバに返す点が異なる。

設定ミスにより木がループしている場合を考慮して、根のサーバから連続して検索の依頼が伝搬される階層の深さ(ホップ数)を制限する。この値を参照する事により検索依頼が永久に回り続ける事を防止する事ができる。

システムの動作例を図4と図5に示す。図4では全てのサーバがマルチキャストを許可するものとする。この時、クライアントがホストキーとしてサーバAを示すキー'a'とサーバCを示すキー'c'を根のサーバTOPに与えたとする。まずサーバTOPで'a'と'c'の2つのホストキーを基にHDBが検索される。ここでサーバAが見つかるのでホストキー'a'を削除し、ホストキー'c'をサーバAに渡す。サーバAはホストキー'c'を基に検索を行う。ここで、サーバCが見つかるのでホストキー'c'を削除し、サーバCに接続する。サーバCにはホストキーが与えられない

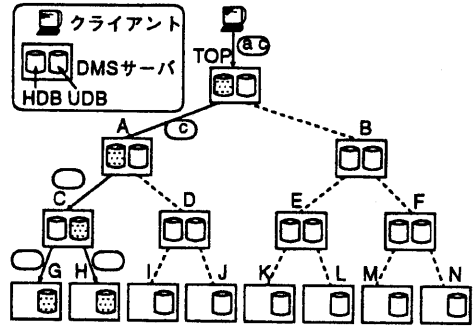


図4 DMSシステムの動作例1

ので、ユーザーキーを基にUDBを参照する。また、子のサーバであるGとHに検索を依頼し、GとHでもユーザーキーを基にUDBの検索が行われる。

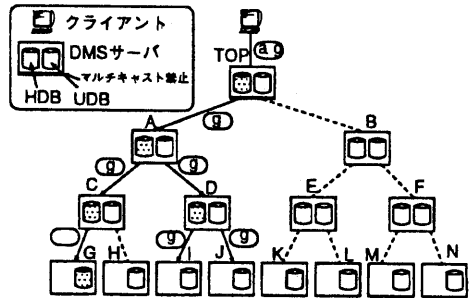


図5 DMSシステムの動作例2

図5ではサーバTOPのみがマルチキャストを禁止し、他のサーバがマルチキャストを許可するものとする。クライアントがホストキーとしてサーバAを示すキー'a'とサーバGを示すキー'g'を根のサーバTOPに与えたものとする。まずサーバTOPで'a'と'g'の2つのホストキーを基にHDBが検索される。ここでサーバAが見つかるのでホストキー'a'を削除し、ホストキー'g'をサーバAに渡す。サーバAはホストキー'g'を基に検索を行う。ここで、子のサーバが見つからず、かつマルチキャストが許可されているのでサーバCとサーバDにホストキー'g'を渡す。サーバCはホストキー'g'を基にHDBの検索を行うと、サーバGが見つかるのでホストキー'g'を削除してサーバGに接続する。サー

命令	引数	機能
HELO	検索者のホスト名	送信側 DMS サーバを受信側 DMS サーバに認識させる
DMSS	検索者のユーザ名	コマンド転送を開始する
UKEY	ユーザキー	ユーザキーを転送する
HKEY	ホストキー	ホストキーを転送する
EOKY	ホップ数	キーの終了を示す。
RSET	なし	以前に送ったキー転送の処理を中止する
HELP	なし	受信側 DMS サーバの HDB を転送する
QUIT	なし	接続を終了する。
NOOP	なし	何もしない。サーバの動作を確認するためのコマンド
RNUM	返値の数	検索されたユーザの数を転送する
RNAM	返値の名前	検索されたユーザの氏名を転送する
RADR	返値のアドレス	検索されたユーザの電子メールアドレスを転送する
ORCV	動作したサーバ名	動作したサーバの名前を転送する
NRCV	動作しなかったサーバ名	動作しなかったサーバの名前を転送する
DRCV	動作を拒否したサーバ	動作を拒否したサーバの名前を転送する

表 1 命令表

バ G にはホストキーがないので、ユーザキーを基に UDB を検索する。サーバ D でも同様にホストキー 'g' を基に HDB を検索するが、また子のサーバが見つからないので、その子のサーバ I と J にホストキー 'g' を渡して同様の操作を行う。サーバ I と J は検索木の葉のサーバであるので HDB が存在しない。そのため、I と J は検索に失敗したものとし、その旨を親のサーバ D に伝える。サーバ D は同様に検索に失敗したことをサーバ A に伝える。サーバ A はサーバ G の検索結果とサーバ I、J の検索失敗をサーバ TOP に伝える。

2.4 通信プロトコル

DMS のサーバ・サーバ間、サーバ・クライアント間で用いられる通信プロトコルは共通となっており、お互いがクライアントであるという事

コード	意味
214	HDB の一覧を転送します。
221	コネクションをクローズします。
250	リクエストされたコマンドを正常に終了しました。
421	このホストの DMS サービスは動作していません。
450	リクエストされたコマンドは実行できませんでした。サーバのビジーが原因です
451	ローカルエラーのため指定のコマンドは実行できませんでした。
452	リクエストされたコマンドは実行できませんでした。ファイルシステムの容量不足が原因です。
453	ホップ数が規定数を越えました。
500	コマンドの文法エラーです、またはコマンド・ラインが長過ぎます。
501	指定のコマンドのパラメータ・エラーです。
502	指定のコマンドはサポートされていません。
503	コマンドの実行順序に問題があります。

表 2 対応コード表

やサーバであるという事を意識せずに通信する事ができる。このことは、クライアントは木の根のサーバのみならず、木の任意の節のサーバに接続が可能であるという事を示している。本プロトコルの命令表を表 1 に、命令に対する対応コード表を表 2 に示す。このプロトコルは、SMTP(Simple Mail Transfer Protocol)のプロトコルに DMS で必要となる機能を加えたものである。表 2 のコードの付け方は [7] の APPENDIX E を参考としている。

次に、プロトコルの流れの例を図 6 に示す。

最初の HELO コマンドと DMSS コマンドは検索そのものには関係しないが、アクセス制限の認証のために用いられる。その後 UKEY コマンドと HKEY コマンドによって検索キーを与える。その後、クライアントは EOKY コマンドを発行して結果を待つ。DMS サーバ 1 はホストキー Information を基に HDB を検索し、DMS サーバ 2 を見つける。そして、DMS サーバ 2 に対して同様の通信を行う。この時、ホストキー Information はこれに対応するサーバがあったため削除され、DMS サーバ 2 には送られない。DMS サーバ 2 はホストキーが与えられなかったため、ユーザキー Yamanaka を基に UDB を検索する。その結果、Yamanaka Tadashi が見つかったので、上位のサーバにこの結果を返し、DMS

サーバ1はクライアントに結果を返し、QUITコマンドを送信して、通信を終了する。

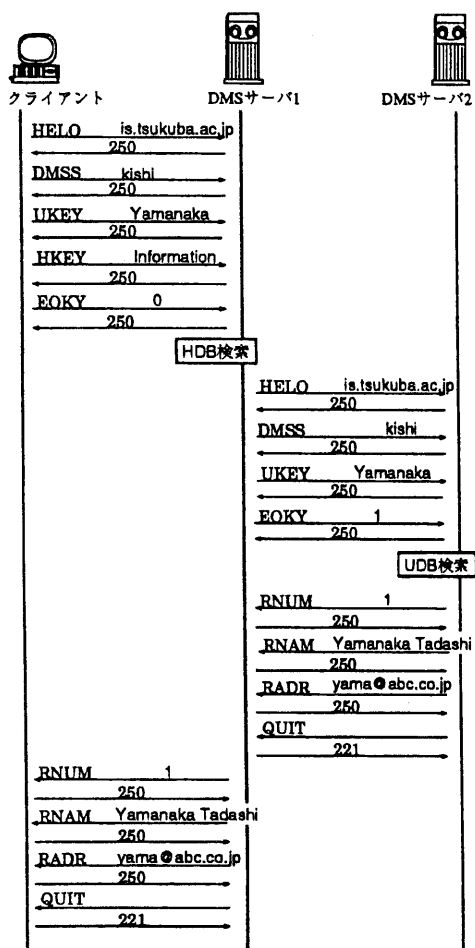


図6 プロトコルの実行例

3 まとめ

本論文では、電子メールアドレスなどの情報のデータベースを各ドメインに分散させた電子メールアドレス検索システムDMSを提案した。このシステムでは、各サーバを検索キーの関連付けによって木構造に構成する事で検索のための通信量を抑制し、また、各サーバ毎に下位のサーバに対するマルチキャストを行うかどうかを選択させる事で、検索の依頼が広範囲に広がる事を防いでいる。現在、本システムのプロトタイプが試作

されており、本システムの評価を行っている。

本システムでは検索対象は電子メールアドレスのみであったが、WHOISサービスなどで提供されている住所や電話番号などの情報を提供する機能はUDBを変更する事で容易に実現される。また、X.500シリーズ[8]などのディレクトリーサービスに対応する事も検討している。

参考文献

- [1] S. Bradner: "Source Directed Access Control on the Internet" RFC-2057, Network Information Center, Harvard University (1996).
- [2] K. Harrenstien, M. Stahl and E. Feinler: "NICNAME / WHOIS," RFC-954, Network Information Center, SRI International (1985).
- [3] G. Malkin: "Internet Users' Glossary," RFC-1983, Network Information Center, Xylogics (1996).
- [4] J. Postel and C. Anderson: "WHITE PAGES Meeting Report," RFC-1588, Network Information Center, Information Sciences Institute (1994).
- [5] T. Okuda, E. Tanaka and T. Kasai: "A Method for the Correction of Garbled Words Based on the Levenshtein Metric," *IEEE Trans. Comp.*, Vol. C-25, No. 2, pp. 172-178 (1976).
- [6] 楊 林: "分散電子メールアドレス宛先情報探索における平均探索長の評価," 情報処理学会論文誌, Vol. 33, No. 11, pp. 1431-1436 (1992).
- [7] J. Postel: "Simple Mail Transfer Protocol," RFC-821, Network Information Center, Information Sciences Institute (1982).
- [8] CCITT: "The Directory: Overview of Concepts, Models and Services," CCITT Recommendation X.500 (1988).