ＤＣＥ/ＤＦＳ 環境における分散型電子メールサーバの構築[1]

柴田章博、押久保智子、浜田 紀、佐々木節
高エネルギー加速器研究機構(KEK) 計算科学センター

概要

DCE/DFS 環境に分散型電子メールサーバを構築する。ユーザにはどのサーバにアクセスしているかを意識させない透過的な環境を実現する。日常的な研究や業務に不可欠なものとして停止時間の少ない安定的な運用や処理能力の高いシステムであり、利用率の向上に伴うシステムの拡張性やコストパフォーマンスを考慮して負荷分散、可用性、および拡張性のあるシステムであるよう設計する。

1. はじめに

インターネットの普及により、電子メールは日常的な研究や業務に不可欠なものとなった。研究所などにおける利用頻度の高いサーバは、年々増加する利用率に対応し、停止時間の少ない安定的な運用や処理能力の高いシステムとすることが求められる。利用率の向上に伴うシステムの拡張性やコストパフォーマンスを考慮して、SMP を利用した集中型のシステムとすることよりも分散型メールサーバのシステム開発を対象とし、様々なメールクライアントからの利用に対応できるインターネット標準のプロトコル、SMTP、 IMAP4、POP3 をサポートするシステムの構築を行う。

本研究では DCE/DFS で提供される分散環境に複数の IMAP、POP、SMTP の分散サーバを配置することにより、分散型電子メールサーバを構築する。ユーザに対してアクセスしているサーバを意識させない透過的な環境を実現し、負荷分散、可用性、および拡張性のあるシステムを構築する。

2. 分散型電子メールサーバ

分散型メールサーバの構成方法には、複数台の IMAP/POP サーバの配置によって、ユーザやリソースを各々のサーバに分割して配置する「ユーザ・リソース分散型」(図１)と、リソースを複数台共有し、透過的な利用環境を提供する「分散ファイル共有型」(図２)が考えられる。

ユーザ・リソース分散型では、各ホストにユーザを分割して割り振るため、それぞれのホストの設定は単独のものとして設定できるが、分割したユーザに共通の電子メールアドレスによる配送を行うためにメールハブなどのメールの転送機能が必要となる。また、IMAP/POP サーバへの総称ア

---

[1] “Building mail server on distributed computing system”, Akihiro Shibata, Tomoko Oshikubo, Osamu Hamada, Takashi Sasaki, Computing Research Center, High Energy Accelerator Research Organization (KEK)

Abstract: We build the distributed electronic mail system based on the DCE/DFS environment, which are symmetrically distributed and provide seamless access of SMTP and IMAP. We design the server, which are stable and highly available for aindispensable function of daily job.

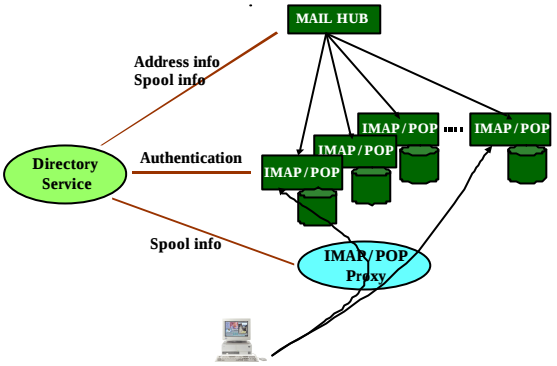

図　1　ユーザ・リソース分散型

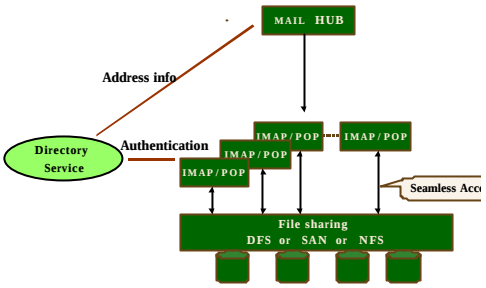

図　2　分散ファイル共有型

	ユーザ・リソース分散型	分散ファイル共有型
認証サーバ	統合のため必須	統合のため必須
総称アドレス	メールハブによる転送が必要 ユーザごとのディスパッチが必要	メールハブ、ディスパッチャーはオプション
IMAP サーバ障害	障害サーバに登録されているユーザは、障害復旧までメール機能は全く利用できない。	サービスするサーバは減るが、メール機能は接続し直すことにより利用できる。
可用性	サーバ毎にリソースを二重化しなくてはならず、コストが高い。	ファイルサーバのデータアクセスの二重化が必要。
開発の重要項目	ディレクトリサービスとメールハブやディスパッチャーの連携。	分散ファイルアクセスの同時アクセスの排他制御

表　1　分散型メールサーバの対比

ドレスによるアクセスを提供するために、ユーザ名を判断してメールフォルダのあるホストへ接続させるプロキシー(ディスパチャー)が必要である。これらの総称アドレスを管理し、ユーザに透過的で接続されているホストを意識させない大規模なシステムの運用を行うには、ディレクトリサービスとの連携が必須となる。

分散ファイル共有型では、ディスクのデータが各ホストから共通に参照できるため、メールの転送やディスパチャーの機能を使用することなく統一されたアクセス環境を構築することができる。しかし、電子メールシステムにおいては、メールのスプーリングと IMAP/POP のサーバアクセスなど、同一ファイルに対して異なるホストからの同時アクセスが発生するため、分散ファイルシステムでの排他制御を十分考慮した設計とすることが重要である。各システム構成の対比を表 1にまとめる。

3. DCE/DFS 環境におけるシステムの構成

本システムは、分散メールサーバを配置するには図２の分散ファイル共有型のモデルを採用する。分散環境を提供するミドルウェアとして、DCE 及び DFS を用いる。（DCE/DFS については、例え

ば参考文献[4][5]を参照のこと。）システムの設計全体は、電子メールのサービスを機能ごとに分散させ、かつ複数の同等構成のサーバを配置することで可用性と負荷の均衡を図るようにする。

ディレクトリサービスには、DCE の認証サーバやセルディレクトリサービスサーバ、及び DNS サーバ、LDAP サーバをそれぞれ複数配置する。DCE の認証サーバやセルディレクトリサービスにより、透過的ユーザ環境を提供する。DCE は Kerberos V のセキュリティ技術に基づいており、セルを構成するホスト間の通信は堅牢に守られる。DCE の認証サーバは、DCE のセルを構成するホストの認証情報とユーザのアカウント情報を保持する。セルディレクトリサービスは、DCE 環境のアクセスを統一的なディクトリとして参照するためのデータベースを保持する。これらのサーバはレプリカ作成によって負荷分散と可用性を保証している。

ファイル共有システムには、DCE 上の分散ファイルシステムである DFS を使用する。UNIX のシステムでは古くから NFS がファイル共有を行う手段として利用されているがメールサーバのような複数プロセスがランダムにアクセスする環境では、ファイルサーバの状態を通知する機能がないためフェイルオーバーの機能やファイルアクセスへの排他制御が不十分である。DFS は“トークン”を用いて、サーバの状態や排他制御の管理を行っており、大規模の分散環境において利用できるよう考慮されている。また、ファイルアクセスのセキュリティの高さ、異種 OS のサーバ・クライアント間のデータ共用などでも DFS を利用するメリットがある。DFS は、読取専用のレプリカ機能を有しており、統一されたディレクトリのアクセスを複数のサーバ間で負荷分散や読取専用のフェイルオーバーの機能がある。本システムでは、メール配信の可用性を向上させるため DFS サーバの高可用性（high availability）システム[2][3]を導入し、ファイルの書き込みに対してもダウンタイムを最小限とするよう構成する。

分散 IMAP/POP サーバは、DCE/DFS の提供する透過的な環境では、個々の IMAP サーバをあたかもひとつのサーバのシステムであるかのように設定できる。DCE 環境に対応するためには、IMAP/POP サーバの認証を受けるように認証方式の変更が必要であるが、ソースコードの 20－30 行程度の変更で対応できる。（DCE の認証に関しては文献[6]を参照のこと。）ファイルに対するアクセスに関しては、排他処理を行う為の特別な変更を行う必要はなく、メールエージェント間の排他処理を統一する一般的な処方を行えば、DFS のロック機構により分散ファイル上での同時アクセスは制御される。分散サーバのアクセスは総称のアドレスを設定し、DNS サーバのラ

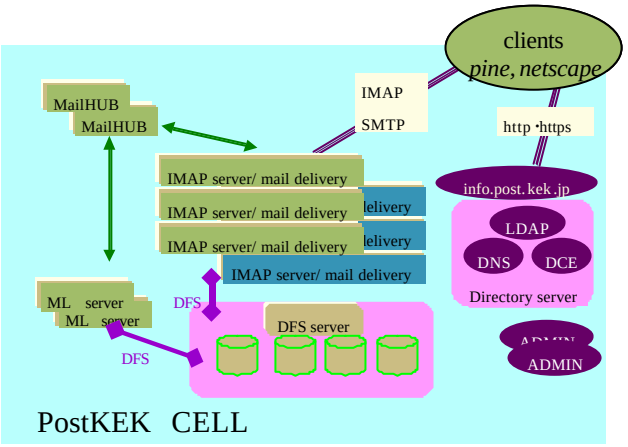

図 3　システム構成図

IBM RS/6000 モデル B50 375MHz (21)

IBM RS/6000 43P モデル 150　(4)

SSA 磁気ディスク装置 RAID5 255GB (2)

ウンドロビン機能を用いて、負荷の均衡化を図ることができる。

メールのスプーリングの実装は、DFS 環境下では IMAP/POP サーバと同様に透過的に構成できる。ただし､DCE 配下のリソースにアクセスするためには、すべてのプロセスが DCE で認証される必要があり、スプーラが DCE の認証を受け起動されるように変更を行う。

電子メール配送には、メールハブを導入しインターネットの通信に専念させる。このことにより複数あるサーバのセキュリティの対策や保守を簡便にすることができる｡メールハブも複数構築し、負荷分散と可用性を上げる。DNS の同一 MX レコードに対し、同列のプレファレンスホストを複数設定することで実装される。

図３に構成概念図をまとめる。システムの実装に用いたパラメータは次のようである。ユーザ数は 3,000 人以上で 10,000 以上のアドレスを管理できること。利用頻度はメールを時間あたり、10,000 通以上、IMAP/POP アクセスを時間あたり、12,000 以上、５０ユーザの同時接続を遅延なく処理できるものとする。メールフォルダは、１ユーザあたり 50MB～200MB、メールスプールは 15MB 以上とする。

4. 提供するサービスとユーザ利用環境

ユーザに提供するサービスは、ログインを廃止し UNIX 環境を意識させない利用環境として提供する｡即ち IMAP をベースとしたディスクトップ環境のメールソフトと web ブラウザを利用し電子メールシステムに関するすべての操作を実現できるよう「ユーザ情報ページ」と称する web サーバを新たに構築する。セキュリティ向上のため、IMAP４、POP3 サーバ及び、ユーザ情報ページは、SSL プロトコルにも対応した。

ユーザ情報ページでは、メールアドレス関係情報として、着信するアドレス一覧、メール自動転送の設定、メール送信時の発信者アドレス（From）のアドレスを統一メールアドレス[2]へ書き替える設定が、リソース情報としては、メールスプールとホームディレクトリのクォータサイズとその使用サイズ、メール件数、ディレクトリ数、ファイル数の情報の取得が行える。パスワードの変更、LDAP を利用したアドレス帳の検索、メール配信ログの確認などの機能も備えている。これら機能は、デイレクトリサービスとして統合された機能に対するインターフェースの実装で実現される。

また、メーリングリストのサービスを分散ファイルシステム上に構築する。IMAP/POP 同様に複数のサーバを配置し負荷分散や可用性のある構成をとることができる。メーリングリストの記事の web 参照や NAMAZU による検索システムも分散ファイルシステム上に構築し、これらの機能を用いたメーリングリストの運営が、各メーリングリストの管理者で行えるよう GUI ベースの管理者機能を実現した。

[2]機構職員にのみ発行されるフルネームからなるメールアドレス

5. システムの運用

ユーザ情報ページと連携した統一的な運用管理を実現した。DCE/DFS 環境下では、サーバ管理のコマンドが整備され管理者権限をもつユーザは、DCE セル内のどのホストにおいても、ユーザの作成からホームディレクトリの作成までが一元管理できる。また、ユーザごとのホームディレクトリやスプールをファイルセット[3]として作成することで、サーバを停止することなく、ユーザリソースのファイルサーバ間の移動や個々のユーザ毎にバックアップを行うことができる。

各種サーバが複数台ずつ稼動しているため、それぞれのマシンの稼動状態やサービスプロトコルの稼動状況の監視、ログ確認などを集約し集中管理を行わないとシステム管理は煩雑で負担の大きなものになる。各サーバのログを１台のホストに集約して管理するシステムを構築した。情報管理機能としては次のものを実装した。

1) 各メールサーバのアクセス数、処理数、負荷値のグラフ、 2)ユーザアカウント情報の表示、3)クォータ情報などの資源情報監視システム、4)全サーバのログ情報監視システム、5）メール配信ログ確認システム。これらを、グラフや表の可視化情報として web ベースの GUI 環境で操作することで、日常管理業務は簡便なものとなる。

本システムは現在、1,400 人以上のユーザが利用している。図４は IMAP/POP サーバのアクセス数を表す。４台の分散 IMAP/POP サーバへのアクセスは、ほぼ同じ曲線を描いており、DNS のラウンドロビンによる負荷均衡化がはかられている。図５は一ヶ月のメールプロセス数をグラフにしたものである。平日には 1 日あたり 60,000 から 80,000 プロセスを処理している。図６は一日における分散メールスプーラと分散メール発信サーバの処理件数をサーバごとにグラフにした

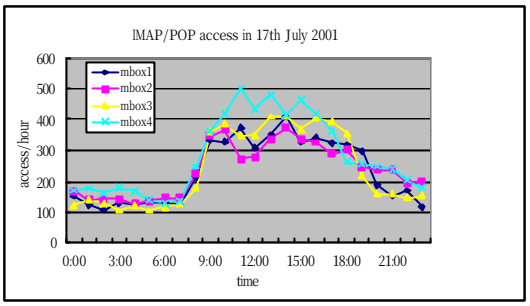

図 4 IMAP/POP アクセス

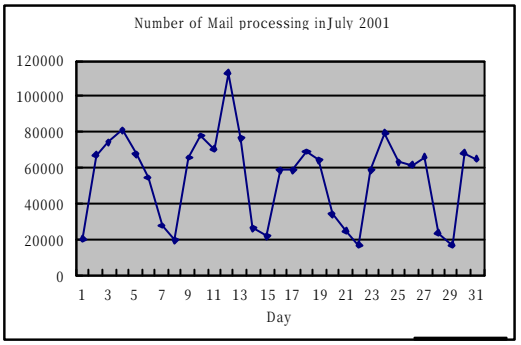

図 5 メールのプロセス数（/day）

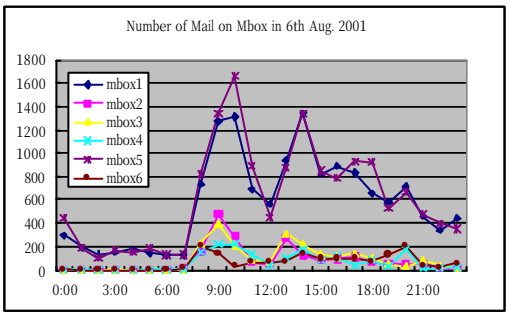

図6 分散スプーラと分散メール発信サーバ

mbox1,5 スプーラ
mbox2,3,4,6 発信サーバ

[3] DFS では、アグリゲートと呼ばれる UFS のパーティションの中に、ファイルセットと呼ばれるリソース単位が存在し、マウントなどのリソース管理がファイルセットに対しを行うことができる。

ものである。mbox1、mbox5 はスプーリングサーバであり、mbox2, mbox3, mbox4, mbox6, は SMTP メール発信サーバである。スプーリングサーバ及びメール発信サーバはそれぞれ DNS ラウンドロビンにより負荷分散を図っており、単純な負荷分散機能でも十分機能している。

透過的な分散メールサーバの構築により、障害時の対応や保守においても運用の面で効果を発揮した。DNS のラウンドロビンのエントリーから切り離すことで、ユーザアクセスを抑制し、運用を停止することなくサーバの切り離しや付加などのシステム構成の変更を行うことができる。ユーザにとっても、複数のサーバが存在するので再接続によって、障害のないサーバを使用することができるため、サーバのダウンタイムを最小限に抑えることができる。これまでにも、OS やサーバソフトのパッチ修正やバージョンアップ、システムパラメータの変更、機器構成の変更などを実施しているが、該当マシンのみを切り離して作業し、メールのサービス機能の停止をすることなく運用できた。

6. まとめと討論

DCE／DFS の分散環境に分散型電子メールサーバを構築した。DCE 及び DFS の提供する透過的な環境と代表サーバ名の利用で、ユーザには接続されたメールサーバを全く意識せずに電子メールの送受信を行うことのできる、負荷分散や可用性のあるシステム構成とできた。分散システムは、ホスト数が多いため管理面で負担増があるが、本システムで構築したような負荷分散や可用性などの面でメリットが大きいと考える。大規模な分散システムを構築するには、ミドルウェアが重要であり、DCE/DFS は大変役立つものであった。一方、DCE/DFS を理解し運用を行うに必要なシステム管理のノウハウの蓄積には、ハードルが高い面がある。日本における DCE/DFS の利用が少ないのもひとつの要因と考えられる。

参考文献

[1] Akihiro Shibata, Tomoko Oshikubo, Osamu Hamada and Takashi Sasaki “Buildin mail server on Distributed Computing system”, CHEP01 at Beijing, KEK preprint 2001-62.

[2] Akihiro Shibata. “High availability file service using DFS” (in preparation)

[3] Akihiro Shibata. “PostKEK – A new mail system using DCE/DFS－” Talk at HEPiX’99, RAL, 1999 April.

[4] Ward Rosenberry, David Kenney and Gerry Fisher. “Understanding DCE”, O’Reilly & Associates, Inc. ISBN 1-56592-005-8

[5] 土門哲也、高清水直美、佐々木節　著「DCE による分散ファイルアクセスの研究」 KEK Report 97-13 January 1998 D

[6] Jhone Shirley, Wei HU and David Magig, “Guide to writing DCE applications”, O’Reilly& Associaties, Inc ISBN 1-56592-045-7