マルチメディア通信と分散処理 51－4
（1991. 7. 5）

大規模分散環境WIDEのアーキテクチャ

村井 純* 慶應義塾大学
中村 修† 東京大学
楠本博之‡ 慶應義塾大学
加藤 朗§ 慶應義塾大学

大規模で広域にわたる分散環境を実現するには、通信技術、オペレーティングシステム技術、応用技術などに生じる課題を解決するとともに、運用技術において発生する諸問題も解決しなければならない。WIDEプロジェクト [1] では、これらの問題を解決するために実験基盤としてのWIDEインターネット [2] の構築を行ない、そのネットワークを国際規模で実際に運用し、諸問題点の解決のための研究実験を行なっている。本論文では、WIDEインターネット上での実験内容とその運用技術に関して報告し、今後の研究課題に関して考察する。

Architecture of WIDE Systems

Jun Murai
Faculty of
Environmental Information
Keio University
5322 Endo, Fujisawa,
Kanagawa 252, Japan

Osamu Nakamura
Computer Centre
Oparation Center,
The University of Tokyo,
2-11-16 Yayoi, Bunkyo-ku,
Tokyo 113, Japan

Hiroyuki Kusumoto
Faculty of
Environmental Information
Keio University
5322 Endo, Fujisawa,
Kanagawa 252, Japan

Akira Kato
Faculty of
Environmental Information
Keio University
5322 Endo, Fujisawa,
Kanagawa 252, Japan

Researches to achieve goals to realize a very large-scale distributed systems in wide-area environment have to be based on various studies, namely, communication technologies, operating system technologies, network operation/management technologies, and application technologies. WIDE research project has developed a testbed called WIDE Internet and is doing practical studies to achieve the above goals. This paper describes concepts on sytsem and network architecture which the researches and devlopments of WIDE project are based on, and reports current status of technologies employed to develop computing environment of WIDE.

*jun@sfc.keio.ac.jp
†osamu@cc.u-tokyo.ac.jp
‡kusumoto@sfc.keio.ac.jp
§kato@sfc.keio.ac.jp

1 はじめに

コンピュータ環境の技術の原点としては、いずれも1969年に独立して開発が開始された広域ネットワークであるARPAnetとUNIXオペレーティングシステムに見ることができる。これらが1970年代に独自に発展し、現在のネットワークとオペレーティングシステムに関する技術の基盤が形成された。1980年代には、ネットワークオペレーティングシステムとしてのUNIX 4.2BSD、Ethernet、そして、ワークステーションの登場と発展により、それまで独立していたコンピュータとコミュニケーションの技術が統合された。これが、局所的な分散システムの時代の幕開けとなる。

1990年代には、ラップトップコンピュータのような新しい形態によってもたらされる、人間の活動環境におけるコンピュータの新しい守備範囲、音声や画像などの新しいメディア、高性能な計算能力といったコンピュータ技術面での変化や発展が期待される。そして、さまざまな通信技術を利用した分散環境の発展と相互接続技術によって構築される大規模で広域にわたるコンピュータコミュニケーション基盤を確立する必要がある。そしてこれらを同時に考慮した開発を行なわなければ、次世代のコンピュータ環境を支える技術の確立は不可能である。

このような、背景から、JUNETの構築にたずさわってきた研究・技術グループは、大規模広域分散環境に関する総合的な研究・開発の実験を行ない、今後の計算機環境の基盤となる技術の確立をめざす研究プロジェクト、WIDE (Widely Integrated Distributed Environments) の研究活動を1987年から行なってきた。ここでは、実験環境を構築し、実証的に研究開発を行なう方針にした。これは、コンピュータ環境という成果の性質上、運用や利用を含めた臨床的な方法による研究活動が不可欠なためである。

大規模広域分散環境における課題はその大規模性から生じるスケールの問題、通信技術、そして、これらの基盤を形成するネットワーク技術に分類することができる。

2 基盤となるネットワーク環境

実証的な研究活動を行なうために、WIDEインターネットは、十分に大規模なネットワークを構築し、実用的な運用を行なう基盤を確立した。これがWIDEインターネットである。WIDEインターネットは仙台、藤沢、東京、京都、大阪、福岡をそれぞれ拠点とした研究者によって運用されている。これらの拠点はWNOC(WIDE Network Operation Center) として相互に接続され、WIDEインターネットのバックボーンを構成している。ここで用いられている回線は64Kbpsから192Kbpsまでのデジタル専用回線を中心に、ISDNによるバックアップとの組合せで実現されている。

各参加組織のネットワークは近隣のWNOCに接続され、これらを含んでWIDEインターネットが構成される。ここには、音声専用回線の非同期通信を用いた接続と192Kbpsまでのデジタル専用回線による同期通信が混在して用いられている [3]。WIDEインターネットは、理学系研究用運営ネットワーク であるTISN (Todai International Science Network)、科研費による研究活動として運用されるJAIN (Japan Academic Inter-university Network)、全世界規模の運用ネットワークに一部である日本BITNETとの協調運用を行なうことによりさらに現実的な技術要求を抽出することが可能となっている。また、WIDEインターネットは国際接続を行なっていて、これを通じて世界規模の研究基盤を研究活動に利用することができる。図1にこうしたWIDEインターネットの現状を示した。

3 大規模広域分散環境の背景

分散処理の技術が透過性の提供とその基盤となる通信技術とで発展してきた背景にはローカルネットワークに代表される比較的安定したブロードキャスト型の通信技術がある。一方、大

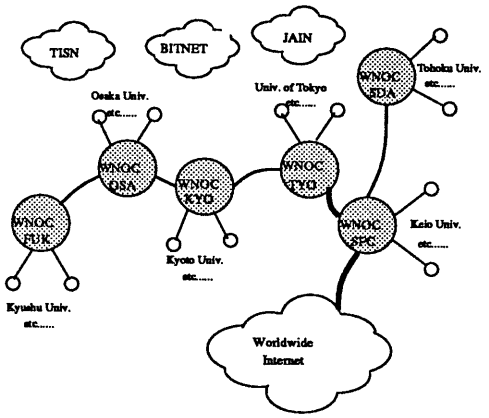

図 1: WIDE インターネットの現状

規模な分散環境を構築する基盤にはネットワーク間接続の概念に基づいたプロトコルアーキテクチャによる通信技術を利用する必要がある。この通信モデルの違いが大規模広域分散環境の諸問題点を引き起こす原点となっている。

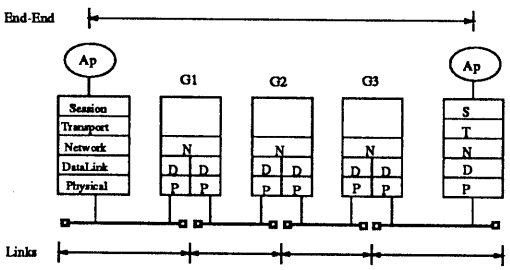

図 2: プロトコル構造とネットワーク間接続モデル

図 2 は階層型プロトコルアーキテクチャによるネットワーク間接続による通信形態を示している。ここで、応用モジュールはゲートウェイ間を接続する各ネットワークの相互接続によって構成される通信経路を用いて実際の通信を行なう。ローカルネットワークによる通信（この図では個々のネットワーク内の通信）と比較して、ネットワーク間接続のモデルでは考慮しなければならない点が存在する。

- 各ネットワークの運用はそれぞれのポリシによって行なわれている。利用の制限などを吸収する運用に関連する技術の開発が透過性を実現するために必要となる。
- セキュリティのレベルを含め、通信の質に対する保証は各ネットワークの協調によって提供されるために、このための総合的な技術を確立する必要がある。
- 各接続は異なる通信技術によって実現されている。これらの正確な認識と効率的な利用が行なわれないとエンド間での通信の透過性を提供できない。
- 中間のゲートウェイにおいては原則としてネットワーク層以下の通信を制御する。多数のゲートウェイにより構成される広域接続ではゲートウェイの状態がエンド間の通信に多大な影響を与える。

4 通信技術のアーキテクチャ

大規模広域分散環境における異種通信技術の利用に関してはネットワーク層において洗練されたデータリンク層との関係が実現される必要がある。ここには、各通信技術の属性がこの関係の中で明確に定義され、それに基づいたアーキテクチャが設計されなければならない。これに基づいた機能が図 2 におけるゲートウェイに含まれると通信経路上で可能経路を選択する際に要求に基づいた決定を行なうことができ、全体としての透過性が実現できる。

広域接続に利用される通信技術にはローカルエリアネットワークに用いられるブロードキャスト型のネットワークに加えて、一対一の通信を提供する専用線型の回線と、回線交換網やパケット交換網のような接続先の切替が可能な回線がある。これらの通信技術をコンピュータシステムのアーキテクチャとして実現されるプロトコルアーキテクチャの中でどのように取

り扱うかは未定義であり、確立された既存技術が存在しない。

そこで、WIDE では、TCP/IP プロトコル体系に基づいた BSD UNIX[4] のオペレーティングシステム構造の環境のなかで、これらの問題を解決する試行を進めている。BSD UNIX のモデルでは、利用者プロセスとのインターフェースとして `socket` の概念を、階層的プロトコル構造の実現に `protocol switch` の概念を、データリンクモジュールの抽象化に `if` の概念が用いられている。

この枠組に基づいて各種の通信技術を議論する際に、一対一の接続をどのようにとらえるかが問題となる。WIDE では従来考えられているような一対一接続を一つのネットワークとしてとらえるネットワーク間接続モデルと同時に、図 3 に示したような一対一接続の両端のゲートウェイを仮想的に一つのゲートウェイととらえるモデルを用いたアーキテクチャ構築も試行している。これにより全体のネットワークの数を抑え、制御のモデルを明確化するだけでなく、ネットワークの数に比例する経路情報に関するトラフィックを減少させることにより全体の効率をあげることにも貢献している。

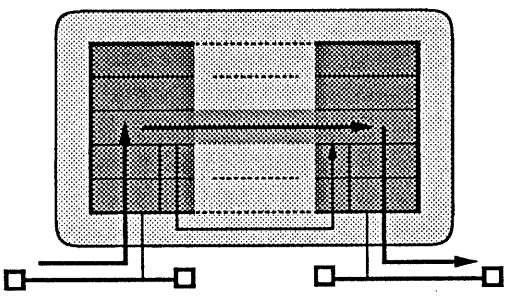

図 3: 1 対のゲートウェイによる仮想的なゲートウェイ

4.1 X.25 パケット交換網

WIDE のアーキテクチャの中では、X.25 を用いたパケット交換網は上記のデータリンクモジュールの抽象化の範囲での利用を試行している。データグラム系のネットワークプロトコルを基盤としたネットワーク構築においては、X.25 をデータリンクエンティティとして利用することにはいくつかの問題がある [5]。CSNET では、IP データグラムを X.25 を用いて転送するための技術が確立され、[6] として広く利用されている。

特にわが国での状況を考慮すると 分散計算機環境構築のためには、データグラム系のネットワークにおける X.25 の有効な利用技術が次のような理由により確立される必要がある：

- 学術情報網の X.25 ネットワークが全国の大学環境に提供されている。
- 信頼性の高い安定した公衆パケット交換網が全国で利用可能。
- ISDN のパケットサービスが開始され入手可能性が高まった。
- OSI 体系のプロトコル実験の基盤が必要である。

そこで、**WIDE プロジェクト** [7] では、広域分散環境の構築に有効な X.25 モジュールの開発を行なった。本モジュールの仕様は次のような点である：

i) 広く利用できるように、一般的なワークステーション上で稼働すること。

ii) 将来の実験開発が継続できるように、BSD 系の UNIX 上で稼働するソースコードを提供できること。

iii) 一般に用いられる 他の X.25 上の IP 機能 との接続が可能なこと。

iv) 低速でかつオーバヘッドが予想される X.25 上で可能な限りのスループットを実現すること。

実際に実現された **WIDE/X.25**[8,9] 。は、4.3BSD の標準モジュールとして実現されている。そのため、同期のシリアル回線のデバイス

ドライバが開発され得るシステムへの移植は容易になっている。また、CISCO 社のルータ、SUN LINK/X.25、及び、各種Σワークステーションの X.25 機能との相互接続が実証されている。さらに、このような相互運用性を保証しつつ、複数の X.25 論理チャネルを同時に並行利用する「マルチリンク」の機能を組み込むことにより資源が限定される網においても高いスループットを実現している。

WIDE/X.25 は IP[10] レイヤの下に位置し、次の4つのサブレイヤに分割することができる。

1. HDLC
2. X.25
3. マルチリンク
4. インタフェースレイヤ

HDLC レイヤは伝送制御手順を規定し、基本機能として、DTE - DCE 間の回線の接続、切断、データフレームの送信を行なう。

X.25 レイヤでは、通信するマシン間のコネクションを End-to-End で管理する。通常、このレイヤでは、通信したい相手に対して 1 対 1 で仮想回線を張ることになる。このため、仮想回線が設立されていない相手に対しては、通信に先だって呼の設定が必要となる。この時、相手に対して、ウィンドウサイズ、スループット、その他のファシリティなどを取り決めることができる。

マルチリンク レイヤは、一つの通信相手に対する複数本のコネクションを管理する役割を持つ。すなわち、上位レイヤに複数の X.25 の論理回線を 1 本の回線であるかのように見せる。従来の X.25 では、各通信相手に対して仮想回線は 1 本であったが、これを複数本確立する。そして、このレイヤで、あるマシンに向かう IP データグラムを効率良く振り分けることによって、X.25 の高速化を実現する。また、回線の使用状況を見て、動的に論理回線の本数を増減する機能を持つ。これにより、バンド幅を広げ、スループットを向上させることが可能となる。ここで、注意しなければならないのは、IP レイヤから見るとただひとつの仮想回線が設立されているように見え、マルチリンクを意識させないことである。つまり上位層からは、回線が太くなったようにしか見えない。

最上位の Interface レイヤは、IP レイヤとのインタフェースをつかさどる部分である。このレイヤでは、IP アドレスから X.121 アドレスへの変換 [6]、その逆変換を受け持つ。そのための情報をテーブルとして持ち、ユーザレベルから追加、削除が可能である。また、IP リンクを統括的に管理するため、論理回線の接続、切断などの制御を指示する。

通常、UNIX では、ネットワークに接続されているデバイスに対してインタフェースが割り当てられている。この観点からすると X.25 網を一つのネットワークと考え、ただ一つのインタフェースを割り当てるのが自然であるように思われる。この場合、X.25 網に対してもネットワークアドレスをただ一つ決めなければならないが、ルーティングに関して不都合が生じる。そこで WIDE/X.25 では、通信する相手ごとにインタフェースを割り当てる方式をとっている。

本モジュールは、学術情報網を利用した実験 IP ネットワーク、JAIN において実際に安定して稼働している。また、CISCO 社のルータ、SUNLINK/X.25、及び、各種Σワークステーションなどの他の X.25 機能を提供するものとの相互接続も確認されている。WIDE/X.25 は、SONY NEWS-800 シリーズ NewsOS3.3、Sun3 SunOS4.X、SparcStation SunOS4.X で稼働している。

4.2 ISDN

現在のわが国の ISDN サービスではパケット交換と回線交換が提供されている。オペレーティングシステム構造から見ると、これらの特徴をデータリンクモジュールとして抽象化し、上位層から透過的に取り扱う機構を開発する必要がある。ここで要求される機能として次のよ

うな点をあげることができる。

- 回線交換の制御（自動発着信、自動切断の機能）
- 複数の回線を同一相手と接続し通信行なうためのマルチリンク機能
- 上位アドレスと下位アドレス（電話番号やDTE アドレス）との変換
- 統一的な制御構造

広域大規模分散環境における ISDN の利用モデルとしては、i) 末端ノードの接続、ii) 主回線障害時のバックアップ、iii) 主回線、のそれぞれの目的を設定して設計と開発を行なった。

i) に関しては接続されているネットワークの一部としての取り扱いを行ない、自宅の計算機や移動計算機のために利用されている。ii) に関しては主回線の切断や混雑時に動的な回線確立のために用いることができる。この機構の設計は、実証的な大規模ネットワークの通信状態に基づいた機能が開発されなければならない。WIDE では [11] の結果に基づき、データ転送の遅延に関する要求を分類し、遅延の許されるものに関してまとめて公衆回線網を利用して転送をする実験を行なった。また、iii) に関しては上記の X.25 技術との統合的な機能の開発を行なうことで実現でき、現在設計を行なっている。

実際に構築した機能は ISDN のターミナルアダプタを X.21 インターフェースによって接続することによって実現されている。

5 ネットワーク間接続

大規模広域分散環境の通信はネットワーク間接続のモデルによって実現されている。各要素ネットワークは異なる運用母体によってそれぞれのポリシによって運用されるためにこれらを用いた透過的な通信基盤を構築するためには、i) 各ポリシの要求を実現可能なトポロジに基づいて構築する必要があり、さらに、ii) ポリシに基づいた運用技術を確立する必要がある。

i) の目的を達成するためには大規模分散環境全体の制御が統一的に行なわれる必要があり、ネットワーク間接続の制御が集約されることが望ましい。これを実現するトポロジのモデルが図 4に示されるネットワーク交換モデルである。ここではネットワーク間の接続を実際に行なう実体をネットワーク交換基地として集約し、そこで各ネットワークと接続されるルータによって各ネットワークのポリシに関する要求に基づく制御を実現する。現在の WIDE バックボーンはネットワーク交換基地として働いていて、経路制御情報交換の制御を行なうことにより統一的なポリシの制御を行なっている。

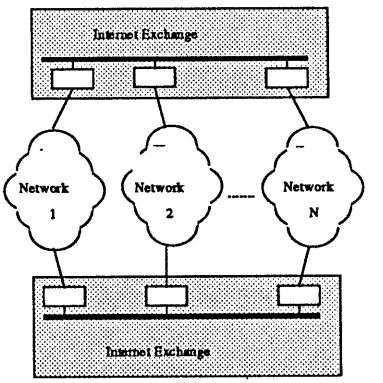

図 4: ネットワーク交換モデル

ii) の例として考えることができる制御項目として現在のわが国のインターネットの国際回線の制限がある。わが国の国際接続はいずれも米国インターネットと接続されている。米国インターネットとわが国の国際接続所有者にはそれぞれのポリシがありこれらの違いを吸収するために名前とネットワークアドレスの対応機能である DNS [12] の構造を確立し、背景となる各ネットワークのポリシ要求の実現をはかっている [13]。ここでは、米国インターネットへ

の接続が可能なアドレスの集合をAグループ、許可されていないアドレスの集合をBグループとわけている。これは名前サーバにより ii) の目的を実現している試行の例である (図5)。

わが国全体の名前サーバは、I) .jp の米国向け主サーバ、II) .jp の国内向け主サーバ、III) 米国 と II) の 副サーバ の3種類のサーバによって行なわれている。米国からは I) に、国内のグループ A は III) に、国内のグループ B は II) に、また、I) 自身は III) にそれぞれ照会することによりそれぞれのポリシに整合した制御が実現されている。この解決方法はこの問題の唯一の解ではないが、さまざまなポリシに関する要求に対応する技術の試行として重要である。

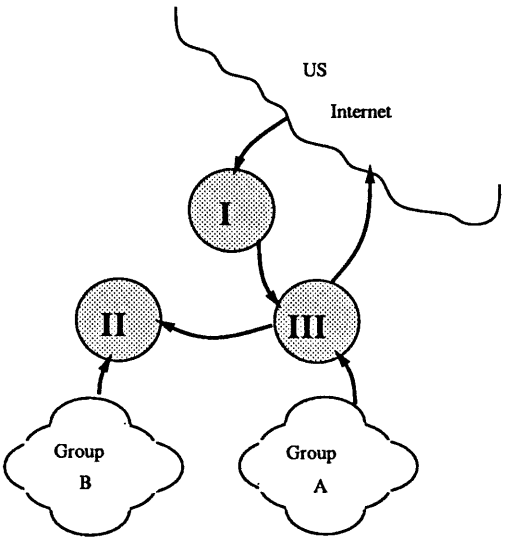

図 5: わが国の名前サーバの試行

6 応用技術

通信技術を性能面から考察すると、複数のことなる通信回線の集合によって提供される大規模広域分散環境のエンド間の通信性能を把握することは困難である。しかし、応用技術の構築基盤を構成するためにはこのような情報に動的に対応する知的なシステムアーキテクチャを実現する必要がある。この基盤となるのがネットワークマネージメントの技術である。大規模ネットワークにおける通信性能の認識が困難なことは良く知られている。図6は図2に対応した通信経路を示したもので、H で示したホスト間の通信は実際には G で示したゲートウェイを介したネットワーク間接続によって実現されていて、これらのホスト間の通信を把握するためには経路上のすべてのゲートウェイの通信状態を把握する必要がある。しかし、実際にはこれらのゲートウェイにはそれぞれに接続されているこの経路と関係のない通信が通過しているために、本質的に開放的なトラフィックが集約され、単純な理論的なモデルを用いての認識が不可能である。そのため、WIDE では実証的な手法を用いて各ゲートウェイにおける通信を概算し、全体のシステムに利用する技術を研究している [14]。

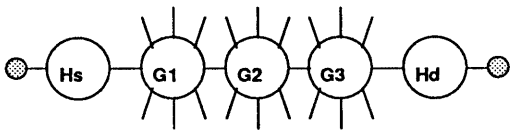

図 6: 端点システムとゲートウェイの関係

このようにして構築された応用技術構築基盤上では、広域分散ファイルシステムやマルチメディア通信などが、効率良く実現することができる。これらを始めとした各種の応用機能の設計を開発を現在行なっている。

7 まとめ

大規模広域分散環境の実現に向けて、極めて実証的な手法でそのアーキテクチャを決定し、各技術の実験試行を行なっている。WIDE インターネットは他の学術ネットワークの運用と協調し、各種の要求分析と問題解決を行なうための実験基盤として構築された。現在は、TCP/IP プロトコル体系上に既に実現されているような

応用技術が、大規模広域分散環境上に応用可能にするための基本的なアーキテクチャとそれに基づいた技術を開発し、試行運用している。今後はこの基盤の発展に加えて、これに基づいた応用技術の開発が課題である。

謝辞 研究を支援して戴いている共同研究各組織と議論をしていただいた WIDE プロジェクトのメンバに深謝する。

参考文献

[1] J. Murai, A. Kato, H. Kusumoto, S. Yamaguchi, and T. Sato, "Construction of the Widely Integrated Distributed Environment," in *Proceedings of TENCON '89*, IEEE, November 1989. Bombay, India.

[2] 村井純、中村修、加藤朗、森島晃年、山口英、平原正樹, "Wide インターネットの現状とその利用," in 情報処理学会第 *42* 回全国大会, 情報処理学会, November 1990. 東京.

[3] WIDE プロジェクト, "1990 年度 wide プロジェクト研究報告書," Tech. Rep., WIDE プロジェクト, July 1990.

[4] M. K. M. Samuel J. Leffler and M. J. Karels, *The Design and Implementation of the 4.3BSD UNIX Operating System,*. Addison-Wesley Publishing Company, Inc., 1989.

[5] D. Comer and J. Korb, "CSNET Protocol Software: The IP-to-X.25 Interface," in *SIGCOMM Symposium on Communications Architectures and Protocols*, March 1983.

[6] J. T. Korb, *A Standard for the Transmission of IP Datagrams Over Public Data Networks.* Purdue University, 9 1983. RFC 877.

[7] Jun Murai, Akira Kato, Hiroyuki Kusumoto, Suguru Yamaguchi and Tomomitsu Sato, "Construction of the Widely Integrated Distributed Environment," in *Proceedings of TENCON '89*, Nov 1989.

[8] WIDE プロジェクト, "X．２５グループによる研究報告," *WIDE* プロジェクト研究報告書, 1989.

[9] WIDE プロジェクト, "X.25 グループによる研究報告," *WIDE* プロジェクト研究報告書, 1990.

[10] J. Postel, *Internet Protocol - Darpa Internet Program Protocol Specification.* September 1981. RFC 791.

[11] WIDE プロジェクト, "Stat グループによる研究報告," *WIDE* プロジェクト研究報告書, 1990.

[12] J. Bloom and K.J.Dunlap, "Experiences implementing bind, a distributed name server for the darpa internet," in *Proceedings of the USENIX 1988 Summer Conference*, 6 1986.

[13] H. Takada, "Bind name server and its configuration," in *WIDE Project Report 1990*, pp. 487–498, WIDE Project, Jul 1991.

[14] 中村 修、北島 剛、村井 純, "大規模広域ネットワークにおける RTT に関する考察," in マルチメディア通信と分散処理研究会, 情報処理学会, Jul 1991.