● 特集：生きたインターネット研究への取り組みと成果

3. インターネット・プロトコル・アーキテクチャの成果

山本和彦／奈良先端科学技術大学院大学
寺岡文男，出水法俊，長健二朗／ソニーコンピュータサイエンス研究所
伊藤純一郎／慶應義塾大学

インターネットに関する研究では，提案した方式の実装と運用を通じた評価が重要視される．WIDEプロジェクトは，この価値観に従ってさまざまな研究に取り組んできた．本稿ではその中でもインターネット・プロトコル・アーキテクチャに関係する分野を取り上げ，実際に作成したソフトウェアを交えて解説する．

はじめに

ネットワーク・プロトコルのモデルは，多くの文献で階層構造として説明されている．OSIがその顕著な例であり，同じホスト上での隣接する階層間の境界面や，異なるホスト上での同一レベルの階層間の境界面が厳密に定められていることが望ましいとされる．

しかし，インターネット・プロトコル（以下IPv4）は，このような窮屈な縦型のモデルに従わなかったからこそ実際に動き，そして世の中に普及したのではないか．TCP/IPの世界では，横方向にフラットなIPv4層があり，さまざまなホスト上にまずIPv4層を実現することに重点を置いている．そして，疑似ヘッダに象徴されるように，TCPやUDP，あるいはその他の上位のプロトコルは，IPv4層さえ実現できれば稼働するように設計されている．

また，さまざまなデータリンク層も，IPv4層が要求する機能を実現するよう構築される．BSDは早くからTCP/IPを実装したので，さまざまな実験のプラットフォームとして利用されてきた．BSDの構造をネットワークに焦点を当てて観察すると，やはりIPv4層が中心をなしている．

トランスポート層を見るとTCPやUDP，Raw IPv4などが，IPv4層に強く依存した形で実装されている．しかし，OSIに見られるこれより上位の層は存在しない．単純にAPIとして，これらのプロトコルを抽象化するソケットがあるのみである．もし，上位の層を実現する必要があるのであれば，カーネルではなくユーザ・プログラムで実装することになる．デバイスに目を向けると，デバイス・ドライバとIPv4層の間には，ネットワーク・インタフェース（以下IF）があり，同種のデバイスの特性を抽出し抽象化している．

TCP/IPやBSDは，多様化する利用形態への対応と普及に伴う広域性の確保のために発展してきたと言ってよい．利用の多様化としてデバイスに注目すると，従来固定的なEthernetやシリアルに限られていた需要が，動的なダイアル・アップや活線挿抜可能なネットワーク・カード，そして1394などに広がっている．

WIDEプロジェクトは初期に，X.25，ISDN，衛星（「衛星とインターネット」の節），無線（「無線とインターネット」の節）など，それまでIPv4が対応していなかったデータリンクに取り組んだ．また，活線挿抜可能なデバイスや電源の管理機構としてWildboar（「携帯計算機とインターネット」の節）を開発した．

利用の多様化への対応として強化されたIPv4の機能には，移動透過性，マルチキャスト，セキュリティ，リアルタイムがある．WIDEプロジェクトでは，移動透過性機能を実現するために，早くからVIP（Virtual Internet Protocol）を提案していた（「移動計算機とインターネット」の節）．また，セキュリティに関してはIP/Secure[1]の開発やswIPeの拡張などに取り組んできた．リアルタイムに関連する研究としては，汎用的なキューイングのプラットフォームを実装して配布している（「パケット・キューイング」の節）．

TCP/IPやBSDの構造に大きな変化をもたらした機能もある．よい例は，あるパケットを他のパケットにカプセル化して配送することで，仮想的なネットワークを構築するトンネル技術（「トンネル」の節）である．実装の視点からトンネルを見ると，関数呼び出しがループを起こすため，永久ループに陥らないための工夫が必要になる．また，複数のデータリンクを1つに抽象化するIF（「衛星とインターネット」の節）

や，ネットワーク層の機能を提供するATMとIPv4層の効率のよい対応付けであるCSR（Cell Switch Router）（「ラベル・スイッチング」の節）も，それまでBSDになかった構造を持ち込んだといってよい．

広域性の確保としては，アドレスの効率的な利用とそれに伴う経路制御の変化がよい例である．従来IPv4アドレスは，クラスという構造を持ち，経路制御もクラス構造を前提としていた．しかし，IPv4アドレスの枯渇により，効率利用が重視されるようになり，アドレス空間をアドレスとマスクで表現するようになった．このため新しい経路制御の実現とその運用技術の蓄積が重大な課題となった．この分野においてWIDEプロジェクトは，ルータ製品の相互接続性の検証に貢献したり，独自の経路表を実現し運用したりした（「クラスレス経路制御」の節）．

アドレス構造の著しい変化としては，クラスの消失に加えて，一意性の喪失があげられる．セキュリティ保全のために，多くのサイトがファイアウォールを構築し，インターネットのすべてのホストで到達性があるというそれまでの前提を覆した．これは，プライベート・アドレスの存在を可能とし，IPv4アドレスの延命に一役かっている．また，NAT（Network Address Translator）やマスカレードの出現を促した．

これらの集大成がIPv6である．IPv6では，移動透過性，マルチキャスト，セキュリティ，リアルタイムなどの機能があらかじめ盛り込まれているか，あとから追加できるように考慮されている．アドレス構造では，スコープという概念でプライベート・アドレスが一般化され，またマルチキャストやエニーキャストもあらかじめ取り込まれている．WIDEプロジェクトでもIPv6を大きな研究課題と捉え取り組んでいる（「IPv6」の節）．

本稿では，TCP/IPのアーキテクチャに関してWIDEプロジェクトが開発し運用を通じて評価したソフトウェアについてまとめる．その多くはBSD上で開発されている．紙面の関係上，本稿ではWIDEプロジェクトのメンバが執筆した論文のみを引用する．

WIDEプロジェクトの成果

この章ではインターネット・プロトコル・アーキテクチャに関しWIDEプロジェクトが上げた成果について述べる．以下順に，衛星，無線，携帯計算機，移動計算機，経路制御，キューイング，トンネル，ラベル・スイッチング，IPv6について説明する．

衛星とインターネット

衛星を用いた通信は，一方向性を利用することが多い，大規模なマルチキャスト通信が可能，遅延が大きいという特徴を持っている．WIDEプロジェクトは，このように地上網とは性質を異にする衛星を，インターネットのデータリンクの1つとして活用するための研究に取り組んできた．

1990年代の前期には，CS衛星で動画像を送る際の帯域のすき間を利用してデータを送る帯域内方式によるIPv4パケットの送信とその応用を研究した．利用可能帯域は160kbpsで当時としてはそこそこ広く，また大規模なマルチキャスト媒体としても期待できた．WIDEプロジェクトでは，市販のCS放送受信機器の出力を入力とし，そこからIPv4パケットを取り出してEthernetに流すアダプタを作成した．そしてこれを用いて衛星を利用したマルチキャスト・ファイル転送やマルチキャスト経路制御プロトコルの研究に取り組んだ．

1990年代の中頃からは，数Mbpsというより広い衛星の帯域をインターネットのバックボーンとして利用することに目的を移した．受信に加え送信も可能になったので，WIDEバックボーンのNOCに衛星通信の送受信設備を配置し，学会やコンサートの中継に用いたり，災害時の代替回線に利用したりした．それまで1対1の双方向通信は可能であったが，1対多の一方向通信を実現する枠組みはなかった．そこで，Ethernetフレームを利用して衛星リンクをブロードキャスト媒体に見せかけるJCというデバイス・ドライバをSunOSやBSD上で実装し用いた．

最近では，衛星を一方向性データリンクとしてバックボーンで利用するための研究を進めている．インターネットでは，リンクの双方向性を仮定している制御プロトコルが多く使われているため，一方向性データリンクと相性がよくない．この問題を解決するためにWIDEプロジェクトでは，NATを用いた方式とトンネリングを用いた方式を提案している．

トンネリングを用いた方式では，衛星リンクと地上網リンクを1個の仮想IFで抽象化することで双方向性を持たせる．このため，制御プロトコルの多くを修正なしで利用できる．WIDEプロジェクトは，この方式とトンネルを自動的に敷設するためのプロトコルDTCP（Dynamic Tunnel Configuration Protocol）をIETFのUDLR（Uni-Directional Link Routing）分科会に提案している．

無線とインターネット

広域無線パケット網は通信速度が低く遅延が大きいが誤りを回復するので，IPv4層からは低速・高遅延・高信頼性のデータリンクに見える．この性質は，LANや専用線と大きく異なっている．そこで，移動ホストから広域無線パケット網を介してインターネットに接続する環境では，この特性を活かした通信方式を確立する必要があった．

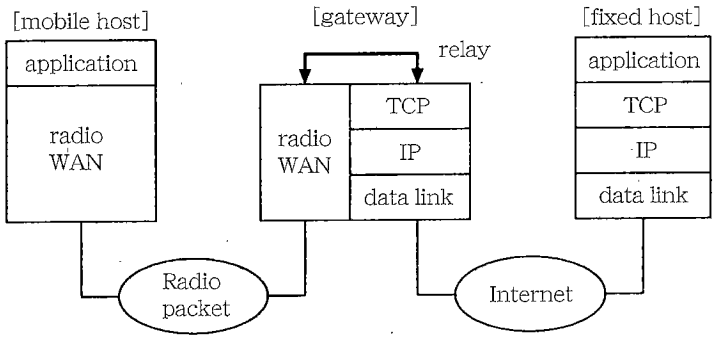

図-1　WRWアーキテクチャ

たとえば，信頼性のある広域無線パケット網の上に信頼性のないIPv4層を置き，さらにその上位のTCP層で信頼性を確保すると，オーバヘッドが大きくなる．これは低速・高遅延の媒体上の通信としては望ましくない．

1994年にWIDEプロジェクトは，このオーバヘッドを解決する枠組みとして**図-1**に示すWRW（WIDE Radio WAN）を提案した．移動ホスト側では広域無線パケット網に直接データを送り，固定ホストは今まで通りTCP/IPを利用する．そして，トランスポート・ゲートウェイを介して，広域無線パケット網とTCP層間でパケットを中継する．

携帯計算機とインターネット

ノートPCなどの携帯計算機では，これまでのBSDでは扱いづらいハードウェアが多数提供されており，重要な役割を占めている．たとえばノートPCでは，電源供給状況を細かに管理しなければ長時間のバッテリ動作は不可能である．また，Ethernetカードなどは PCMCIAインタフェースを介して接続される活線挿抜可能なデバイスであるので，カーネルはデバイスを動的交換する枠組みを提供しなければならない．我々の多くは研究活動のためにノートPCで稼働するBSDを望んでいたので，BSDに対してこのような枠組みを追加する必要があった．

この問題を解決するため，篠田，徳川らが中心となって1994年末頃からBSD/OS上でWildboarの開発に取り組んだ．Wildboarはカーネルに対する大規模な変更として実現されている．カーネル内には，電源管理用のドライバやPCMCIAインタフェースのための管理機構，そして，そこに接続されるハードウェアのためのドライバが追加される．現在のところEthernetカード，無線LANカード，モデム・カード，動画キャプチャ・カード，不揮発メモリ・カードなど多彩なハードウェアをサポートしており，現在販売されているノートPCのほとんどで安定して動作する．現在WildboarはBSD/OSに標準で組み込まれている．また，他のBSDへの対応，CardBus/ZV-portインタフェースへの対応などの作業も進められている．最近では他のBSDにもノートPC用の機能が追加されているが，Wildboarは安定性において群を抜いている．このため，移動透過な計算機利用の研究基盤として大きな役割を果たしている．

移動計算機とインターネット

TCP/IPでは，IPv4アドレスがホストの位置指示子と識別子の2つの意味を持っている．よって，位置が異なれば必然的にIPv4アドレスも変化するので，TCP/IPでは移動に対し透過な通信を実現できなかった．WIDEプロジェクトでは，1990年にホストの識別子と位置指示子の分離より移動透過性を実現するというアーキテクチャを提案した．これに基づいてIPv4を拡張したプロトコルがVIP[5),6)]である．

このアーキテクチャをプロトコル階層に当てはめると**図-2**のようになる．従来の上位層とIPv4層の間にVIP層が挿入されている．このVIP層では，識別子としてVIPアドレスを導入しており，上位層はVIPアドレスを使ってホストを指定する．VIP層がVIPアドレスを位置指示子であるIPv4アドレスに対応付ける．IPv4層はIPv4アドレスにしたがってパケットを配送する．IPv4アドレスはホストが移動すると変化するが，VIPアドレスは不変であるので，移動に透過な通信が実現できる．

WIDEプロジェクトはVIPバージョン1をNEWS-OS，BSD/OS，SunOS上で実装し，1994年6月に配布を開始した．バージョン1には送信ホストの認証機能がないため，悪意のあるユーザは容易に自分のホストのVIPアドレスを偽称できた．認証機構を組み込んだバージョン2をNEWS-OSとBSD/OSで開発し，1995年11月に配布しはじめた．さらに中間にファイアウォールがある場合でも通信できる機能を持つバージョン3を1996年5月に開発した．

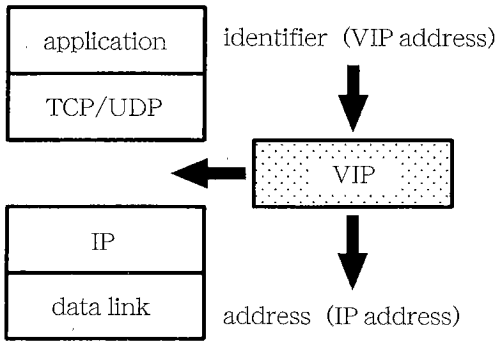

図-2　移動透過性を実現するプロトコル階層

IETFでは，1992年7月に開催された第24回の会合で，移動計算機のための経路制御に関する分科会が発足し，標準化作業を開始した．VIPを含む複数の案が提案されたが，最終的にMobile IPと呼ばれる新たなプロトコルが採択され，現在Proposed Standardになっている．

クラスレス経路制御

1993年から1994年にかけては，それまでのクラスに基づいた経路制御からクラスレスの経路制御への移行が，インターネットの存続に係わる重大な課題であった．この時点では日本サイトの大部分がクラスBを所有しており，サブネットを用いてサイト内のネットワークを構築していた．しかし，サブネットの技術は，段数は高々1段，かつマスク長は固定という制約を持っていた．サイト内でのホスト数の増加により，任意の長さのマスクを多段で利用する必要が生じ始めていた．このような環境を実現するには，ルータがVLSM（Variable Length Subnet Mask）を適切に実装すること，そして運用に耐えることが重要であった．

このような背景を基に，JEPG/IPに参加しているWIDEプロジェクトの加藤と吉村が中心となり，5種類のルータを検証した．この結果はIP Meeting '93で発表され[2]，サイトがVLSMに移行するための指標となった．どの製品にどいう欠点があるかは一般には公表されなかったが，この検証がルータ・メーカにVLSMを適切に実装させる強い動機付けとなったことは確かである．

また，当時のWIDEのバックボーンには，ルータとしてSunなどのWSが使われていた．たとえば，4.3BSD Tahoeは経路表に単純なクラス別のハッシュを利用しているため，これを基に実装されているSunOS 4.xはクラスの制約を受けていた．4.3BSD Renoではradixというクラスレスに対応した2分木の経路表が利用できた．しかし，このコードは難解であり，後からバグが見つかることも多かったので，SunOS 4.xへのradixの移植は適切ではないと考えられた．そこで，加藤，渡邊，山本は，Tahoeのハッシュベースの経路表に代えて2分木上で最長一致を実現する経路表Radish☆を実装し実際にWIDEバックボーンで運用した．

パケット・キューイング

インターネットにとって輻輳回避，公平な帯域割り当て，QoS保証などを実現するトラフィック制御技術の確立は重要な課題である．トラフィックはネットワーク階層の各層で，そのレベルに応じた粒度で制御される必要がある．IPv4層の下に位置するパケット配送レベルでのトラフィック制御は，パケットを粒度とした配送スケジューリング，パケット廃棄，バッファ管理などからなり，これらは総称してキューイングと呼ばれる．

現在までにさまざまキューイング方式が提案されているが，多くが理論の提示やシュミレーションによる評価であり，利用できる実装はほとんどない．また，BSDはFIFOキューイングを仮定して実装されているため，新たなキューイングを組み込むことは容易ではない．そこで，さまざまなキューイング機構を実装するための枠組みとしてALTQ（Alternate Queuing）[8]という代替キューイング機構をBSD上に作成した．

図-3に示すようにALTQでは，複数のキューイング機構から1つをIFごとに動的に選択できる．デバイス・ドライバに関連するコードを共有できるので，新しいキューイング方式の実装が容易となっている．

現時点で，CBQ（Class-Based Queueing），RED（Random Early Detection），WFQ（Weighted Fair Queueing）などのキューイング方式やECN（Explicit Congestion Notification）を実装済みである．また，CBQはRSVPと連動して，動的な資源予約によるQoS保証が可能である．これらの性能評価については，文献8）を参照されたい．

☆ http://www.mew.org/~kazu/radish.ps

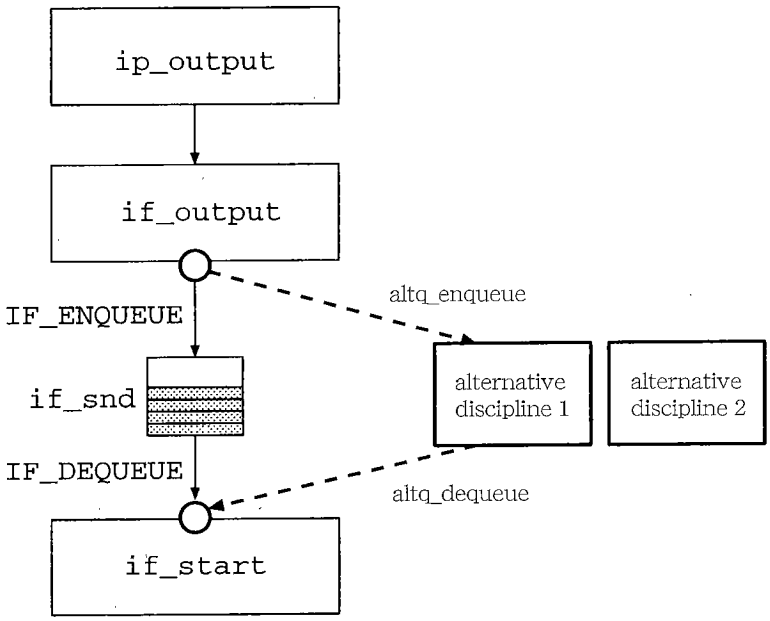

図-3　ALTQの構造

FreeBSD上のALTQの実装は，1997年3月から一般に公開している☆2．CBQ，RED，WFQ，ECNは実現済みであり，また，FreeBSDのほとんどのデバイス・ドライバで利用可能である．

現在ALTQは，WIDEプロジェクトのRT-Bone[7]の一部として運用されている．また，同様のフリーな実装が他に存在しないため，世界中の多くの研究機関でRSVP関連をはじめとした研究用プラットフォームとして利用されてきている．プロバイダが実際の運用に利用している例もある．

ALTQで実現している機能のいくつかが，IPv6への移行を契機にBSDなどに標準で組み込まれれば，移行を促進する大きな動機となるであろう．

トンネル

1992年から1993年にかけて，トンネル技術が発展した．トンネルを用いると仮想ネットワークを構築できるので，ポリシ経路制御がネットワークの物理的なトポロジから受ける制約を緩和できる．また，物理的に連続していない領域にあるプロトコルAを，中間に敷設されたプロトコルBに格納し配送すれば，プロトコルAの領域を仮想的に接続したことになる．マルチキャスト・ネットワークをマルチキャスト用の経路制御機能を持たないインターネット上に構築したMBoneがその代表例である．

UNIXに実装されたマルチキャスト用のトンネル機能は，特殊なAPIから制御するよう設計されていたため，MBone構築以外の目的への利用は難しかった．

そこで，出水は通常のIFとしてトンネルを実現するDDTを開発した．DDTの運用やトンネルに関する議論から，トンネル特有のループ問題や運用の難しさが明らかにされるという成果が得られた[3]．

その後トンネルの利用は多様化した．たとえば，MBoneと同様にIPv6の仮想ネットワークを構築するには，IPv6 in IPv4が必要である．IPsecや移動透過性の実現でもトンネルが利用される．DDTは，XNS用のコードも書かれたが，基本的にはIPv4 in IPv4のトンネル用に設計されていた．そのため，DDTの使いやすさを引き継ぎながら，さまざまな利用に耐え得るトンネルを開発する必要にかられた．

そこで，IPv6を実装中であった山本は，複数のプロトコルを柔軟に扱えるGIFを開発した[4]．GIFは後述する6boneを構築するために利用されており，実際に運用されている．

ラベル・スイッチング

近年，スイッチング技術に基づいた大容量なデータリンクを安価に利用できるようになってきた．その代表的な例としてはATMが挙げられる．しかし，単純にATM上にIPv4層を乗せると，各ルータにおいてIPv4層で転送されるパケットは，ATMセル，IPv4パケット，ATMセルのような変換を施される．この処理からセルとパケットの間の変換を省略できれば，ATMスイッチでのセルのスイッチング処理のみでIPv4パケットの転送が可能となる．そのため，ルータでのIPv4層のオーバヘッドを低減できる．

データリンク層とIPv4層にこのような今までにない関係を，特にATMに対して持ち込んだのがCSRで

☆2 http://www.csl.sony.co.jp/person/kjc/programs.html

ある．WIDEプロジェクトでは，1996年頃からCSR技術に関して検討を開始し，さらに1997年頃から実験的な運用を通じてCSRを評価してきた．

CSRのアーキテクチャ・モデルは，ATM以外のデータリンクにも拡張できる．これはラベル・スイッチングの名称で知られており，品質保証機能などをより容易にかつ効率的に提供できる技術として期待されている．現在IETFのMPLS（Multi-Protocol Label Switching）分科会では，ラベル・スイッチング技術に関する標準化が進められている．

WIDEプロジェクトでは，ATMスイッチを用いたMPLSルータがファイバなどで直接接続されていない場合（たとえばVP/SVC/PVC接続）でも，ラベル・スイッチング技術を適用可能にする方式をMPLS分科会に提案した．これらは，MPLSルータの仕様に対し正式に反映される予定である．

IPv6

WIDEプロジェクトではIPv6の研究に取り組むため，1995年にIPv6分科会を結成した．IPv6分科会では多数の独立したIPv6とIPsecの実装が開発された．その中でWIDE 6boneで標準としてよく利用され，また参照コードとしての役割を果たしているのがHydrangeaである．Hydrangeaは現在山本と伊藤を中心に開発されており，FreeBSDとBSD/OSに対しIPv6とIPsecの機能を拡張する形で提供されている．

IPv6の開発や運用からは，始点アドレス選択のアルゴリズム，mbufの効率的な管理，そして前述したGIFなどの成果が得られた．我々が提案した始点アドレス選択のアルゴリズムは，スコープ指向であり，リナンバーリングに対し頑健で，マルチホーム環境での経路制御の問題を部分的に解決する．

BSDではメモリの管理機構としてmbufがあるが，このmbufに対するIPv6からの要求を定義できた．具体的には，各ネットワーク・デバイスのドライバにおいて，mbufの作成方法を決定し，多くのドライバを書き換えた．また，IPv6の断片を再構成する際はmbufの構造をうまく利用するので，IPv4での実装よりも効率よくなっている．他のさまざまな部分でもIPv4のロジックから独立した実装になっている．

また，IPv6の長所の1つであるプラグ&プレイも適切に実装している．たとえば，Wildboarとの協調動作を実現しているので，ノートPCにEthernetカードを挿した時点でIPv6アドレスが自動的に生成される．

IPv6分科会で開発したIPv6ホストやルータは，実際に6boneというIPv6の実験ネットワークで運用されている．世界の6boneにおいて，WIDEの6boneは最も古く，しかも中心的な役割を担い続けている．規模や利用しているデータリンクの種類で比較すると，WIDEの6boneに肩を並べる実験ネットワークはない．

IPv6の研究分野では，IPv4からの緩やかな移行技術の開発が重大な課題である．IPv6分科会では，IPv4/IPv6ヘッダ変換方式，トランスポート・リレー方式，SOCKS拡張方式をすでに開発しており，実際に6boneで評価中である．トランスポート・リレー方式の実装は，Hydrangeaに含まれて配布されている．

おわりに

実用に耐えるプログラムの作成は多くの時間が必要であるし，実際のネットワーク上での評価には手間がかかる．しかし，机上でのアイディアには見落としがあったり，トイ・プログラムによる評価は現実と遊離している場合がある．また，新たな問題を発見する機会にも恵まれにくい．

インターネットは使われてこそ意味がある．よって，インターネットに関する技術は，実際に実装し運用を通じて評価する必要がある．そして，動かしてみてはじめて分かることも多い．これからインターネットの研究に携わる人がこの価値観を理解していただき，研究を進める上での参考になれば幸いである．

謝辞　初期の原稿に対し貴重な意見をくださった江崎浩氏と神明達哉氏に感謝します．

参考文献
1) Tanida, T. and Shinoda, Y.: IP/Secure: Providing Security On Datagram Delivery For Mobile Host Environment, Proceedings of INET '94 (1994).
2) 加藤　朗, 吉村　伸: Subnet Workshop報告, Proceedings of IP Meeting '93, pp.13-19 (1993).
3) Demizu, N. and Yamaguchi, S.: DDT－A Versatile Tunneling Technology, Proceedings of INET '94/JENC5, pp.661-1～661-9 (1994).
4) 山本和彦: IPトンネルのモデル化と実装, コンピュータソフトウェア, Vol.15, No.2, pp.38-47 (1998).
5) Teraoka, F., Yokote, Y. and Tokoro, M.: A Network Architecture Providing Host Migration Transparency, In Proceedings of SIGCOMM '91 (1991).
6) Teraoka, F., Uehara, K., Sunahara, H. and Murai, J.: A Network Architecture Providing Host Mobility, CACM, Vol.37, No.8 (1994).
7) 石井公夫, 塩野崎敦, 木幡康弘, 小林克志, 石田慶樹, 長健二朗, 寺岡文男: WIDEプロジェクトにおける実時間通信バックボーンの構築, 日本ソフトウエア科学会第14回全国大会 (1997).
8) Cho, K.: A Framework for Alternate Queueing: Towards Traffic Management by PC-UNIX Based Routers, To be appeared in Proceedings of USENIX 1998 Annual Technical Conference (1998).

（平成10年3月2日受付）