動的資源予約システムの運用実験

森島直人[†] 小川晃通[††] 嶺原桂二郎[††]
染川隆司[†] 山口英[†]

技術の発展によってインターネットは着実に安定性を増し, 人々の興味は到達性から通信の品質へと移ってきた. 現在, この要求を満たすためにDiffservが標準化されている. しかし, 標準化は優先配送制御を実現するための機構のみにとどまり, インターネット全体での運用モデルについては触れられていない. そのため, 必要なアドミッション制御の機構を決定することができず, 現在のところ人間の介在する静的な契約のみを想定したもののみが考えられている. 我々は, インターネットにおけるDiffservの階層化された運用モデルを提案している. そのモデルでは, ネットワークの末端部において人間の介在しない動的な契約システムが必要となる. そこで我々は, 動的な帯域割り当てシステムの実装および合宿における運用実験を行った. 本稿ではその報告および結果と評価について述べる.

Experiment of Dynamic Resource Reservation System

Naoto Morishima,[†] Akimichi Ogawa,[††] Keijiro Ehara,[††]
Ryuhji Somegawa[†] and Suguru Yamaguchi[†]

The stability of the Internet shift the interest of users from reachability to quality. Diffserv is under standardization to meet this requirement. However, the main focus of the standardization process is on technical network mechanisms. The absense of consideration on operational model is causing delay to admission architecture discussion. Thus, current Diffserv model assumes static SLA, ie, human interaction is required for every network bandwith agreement. We are proposing a hierarchial operational model for Diffserv. In this model, a dynamic mechanism to make an agreement without human interaction is required. We designed and implemented a dynamic resource reservation system. We evaluated the system by operating a live Diffserv ready network with dynamic SLA. In this paper, we present design, evaluation and conclusion of the operational experiment.

1. はじめに

インターネットの黎明期には, ユーザの最大の関心事は到達性であった. 初期の経路制御技術は未熟なものであり, 到達性の失われることがしばしば発生したためである. 今日では動的経路制御をはじめとするさまざまな技術の発展により, インターネットは着実に安定性を増し, 到達性が失われることは稀となった.

このような状況にしたがい, ユーザの関心事は到達性から通信そのものの品質へと移ってきた. この背景には, インターネットの商用化によるトラヒックの増加に起因した通信品質の悪化や, VoIP(Voice over IP)などの実時間通信を要求するアプリケーションの擡頭がある.

このような要求を満たすため, インターネットにおける通信品質を保証のための仕組みとしてDiffserv(Differentiated Services) [1]が標準化されている. DiffservはISP(Internet Service Provider)が提供する限定されたサービスをDSCP(Diffserv codepoint) [2]と呼ばれる識別子によって分類し, 優先制御によって通信品質を統計的に保証するための枠組みである.

IETFのDiffserv WGではサービスモデルに関する議論が繰り返されてきたが, 『ポリシとメカニズムの分離』という方針のもとに多様なポリシを実現するメカニズムだけが規格に盛り込まれている. しかし, 課金システムやアドミッション制御の機構を議論する上で, サービスモデルの議論は必要不可欠である.

我々は, Diffservの運用モデルが階層的なものであるモデルを考えている. このモデルでは, 末端のネッ

† 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology
†† 慶應義塾大学 政策メディア研究科
Graduate School of Media and Governance, Keio University

トワークにおいては個々のユーザが直接ネットワークに対して資源予約を要求する．しかし，このような動的資源予約に関する実験はほとんど行われておらず，ユーザの挙動やシステムに必要な機構，動的資源予約の妥当性に関してはほとんど明らかになっていない．

本研究では上記の点を明らかにするため，ユーザからの動的な要求に対するネットワークの資源予約システムを構築し，運用実験を行った．

2. Diffserv

Diffservは，インターネットにおける通信品質を保証のための仕組みとして標準化された．Diffservでは優先制御によって通信品質を統計的に保証するための枠組みである．顧客はISPとのサービスレベルの契約(SLA; Service Level Agreement)を結ぶ．これらのサービスはスループットや遅延，遅延の揺らぎなどの統計的な指標によって表される．

同一のポリシによって運営されるDiffservネットワークをDSドメインと呼ぶ．DSドメインの境界には入口/出口境界ノードと呼ばれる機器が設置され，DSドメインに流入するフローはすべて入口境界ノードを通過する．入口境界ノードではSLAにしたがって分類され，現在のフローの状態を勘案してIPヘッダにDSCPを付加する．DSドメインの内部では，同一のDSCPを持つすべてのフローは集約され，ホップ毎にDSCPに関連づけられたキュー制御などの処理をされる．この集約をBA(Behavior Aggregate)，ホップ毎の処理をPHB(Per-Hop Behavior)と呼ぶ．

Diffservの枠組みでは，PHBに利用するキュー制御アルゴリズムやアドミッション制御(ポリシの適用や資源配分機構等)の方式を限定していない．このため，

サービス提供者であるISPは，様々なキュー制御アルゴリズムとアドミッション制御を組み合わせることにより特色あるサービスクラスを定義し，品質保証サービスを構築・提供することができる．また，複数のフローをDSCPごとに集約することにより，より多くのフローが集中するDSドメイン内部のスケーラビリティを実現している．

しかし，インターネット全体でのDiffservの運用やそれに必要なアドミッション制御の方式を考えていく上で，DSドメイン間の関係を考慮した運用モデルは必要不可欠である．例えば，現在Diffservネットワークのアドミッション制御にはCOPS(Common Open Policy Service) [3]のCOPS-PR(COPS Policy Provisioning) [4]の利用が考えられている．これは，単一のDSドメイン内における，SLAの締結に人間の介在

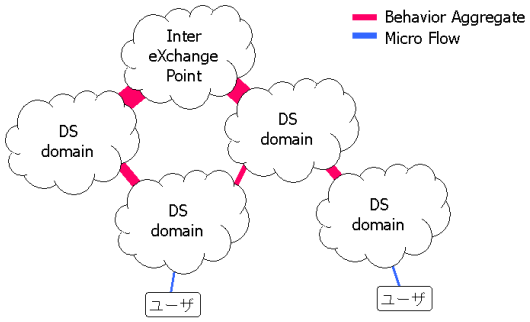

図1 Diffservの運用モデル

を前提とする静的なアドミッション制御のための仕組みであるが，個々のユーザからの動的なサービス要求に対する人間の介在しない動的なアドミッション制御については言及されていない．動的なアドミッション制御の必要性やその方式などについては，DSドメインの範囲やDSドメインの種類といった，前提となる運用モデルを欠いたままで議論を進めることが難しい．

Diffservの運用モデルを考える場合，以下の条件を考慮しなければならない．

C1. **現在のインターネットモデルとの親和性**　現存のネットワークが円滑にDiffservを利用できるネットワークに移行するためには，Diffservの運用モデルと現在のインターネットの運用モデルが大きく異なるものであってはならない．

C2. **スケーラビリティ**　前述のように，Diffservは同一のDSCPを持つフローを集約し，DSドメイン内部のスケーラビリティを確保している．Diffservの運用モデルがこの利点を失うようなものであってはならない．

C3. **柔軟なポリシの実現**　Diffservはその必要性や背景から，複雑なポリシの適用や処理が必要になると考えられる．したがって，DSドメイン内や複数のDSドメイン間でのフローに対するポリシや契約の柔軟性を確保できるようなものでなくてはならない．

これらの条件を考慮し，我々はインターネットにおけるDiffservの運用モデルとして，階層化モデルを提案している(図1)．具体的には，IX(Internet eXchange)を頂点とした階層化モデルであり，現在のインターネット上で容易にDiffservを運用することが可能なものとなっている．IXでは，DSドメイン間での帯域取引が行われると考えられる．

また，RFC2475では入口境界ノードでの分類はBA(Behavior Aggregate)およびMF(Multi Field)で行うことになっているが，これをBAのみによる分類

とすることによってスケーラビリティを確保する．たとえば, 下流のDSドメインBが上流のDSドメインAとサービスクラスαをn Mbpsで契約しているとする．このとき, DSドメインBはさらに下流のDSドメインに対してサービスクラスαを再販する場合, 合計でn Mbpsを越えないようにしなければならない．DSドメインBがサービスクラスαに属するトラヒックをn Mbps以上再販した場合, 上流のDSドメインAの入口境界ノードで契約外として処理される．このため, DSドメインBは下流のDSドメインとのSLAを満たすことができず, 契約違反となる．このような仕組みでDSドメイン間の関係を構築することにより, 各入口境界ノードでのMFによる分類が不要となり, スケーラビリティを確保することができる．

また, 組織の境界点とDSドメインの境界点を一致させることで, 柔軟な運用ポリシの設定が可能である．たとえば, ISP自体のネットワークとその顧客のネットワークの境界はDSドメインの境界となる．また, DSドメインは組織内で細分化されることもある．たとえば, ある組織内の部署ごとにDSドメインを構成する場合もある．このような末端のDSドメインでは, 人間の介在しない動的な契約システムが必要となる．

本研究では, 上記のモデルのうち, 末端のDSドメインに動的な契約システムを導入することの妥当性について検討を行う．また, 動的資源予約が提供された環境におけるユーザの挙動を明らかにする．これらの目的のため, 動的な帯域割り当てシステムの実装およびWIDE Projectの合宿において運用実験を行った．

3. WIDE合宿での運用実験

WIDE Project[5]はインターネットとその関連技術に関する研究と開発を行なう非営利団体である．WIDE Projectで は春と夏の毎年2回, メンバによる4日間の合宿を行なっている．WIDE合宿では “WIDE Camp-net” と呼ばれる一時的な運用ネットワークを構築し, インターネット との接続性を確保するとともにさまざまな実験を行なう．例えば, IPv6や MPLSのような最新のネットワーク技術を合宿参加者に提供している．

我々は, WIDE Projectの2000年春の合宿においてユーザから動的に帯域要求を行なえるDiffservネットワークを運用・実験した．この合宿は山梨県石和町で行なわれ, 236人が参加した．

3.1 実証実験の概要

WIDE Camp-netは, インターネットへの接続にATM over T1をもちい, ユーザからの要求によるトラヒック制御はこの両端で行った．インターネットへの接続性はIPv4およびIPv6で提供され, ATMで複数のPVCを設定することによって石和を含むループが作成された．このため, Diffservでのマーク付加を複数のルータで行う必要があった．運用におけるこのような負荷を軽減するため, ATM PVCブリッジングの技術をもちい, 一つのルータにATMレベルですべてのトラヒックを集中させた．これによって, DSCPのマーク付加やトラヒック制御を行うルータの数を減少させた．

本実験では, ルータへの品質制御パラメータの供給にCOPSを利用した．品質制御パラメータの管理およびアドミッション制御を行うPDP(Policy Decision Point)は石和のWIDE Camp-net内で運用した．また, PDPの決定にしたがって品質制御パラメータを受け取り, 実際に品質制御を行なうPEP(Policy Enhancement Point)は対外線の両端に設置したルータ内で運用した．

また, ユーザが帯域を予約するためのクライアントも提供した．これにより, ユーザはDiffservネットワークに対して直接予約要求を出すことができるようになる．このクライアントはIPv4およびIPv6の両方に対応した．

特定のユーザによる帯域の占有を防止するため, 長期にわたる帯域予約を抑制するための仕組みが必要である．本実験では, 合宿参加者全員にアカウントを発行し, 一定量の仮想通貨を配布した．この仮想通貨は帯域の予約時間に応じて減少し, 仮想通貨がなくなると予約ができなくなる．この仮想通貨はPDPによって管理した．また, ユーザはWebを通して現在の帯域予約状況および仮想通貨の残量を知ることができるようにした．

3.2 ネットワークトポロジ

本節では実験に利用したWIDE Camp-netのネットワークトポロジについて説明する．WIDE Camp-netのネットワークトポロジを図2に示す．前述のように, 今回のWIDE Camp-netはATM over T1(1.5Mbps)でインターネットとの接続性が確保された．この対外線上には3つのPVCが設定され, それぞれ以下の用途で利用された．

(1) IPv4によるインターネットとの接続
(2) 慶應義塾大学湘南藤沢キャンパスを経由したIPv6によるインターネットの接続
(3) 奈良先端科学技術大学院大学を経由したIPv6によるインターネットとの接続

この図に示すように, ATM over T1の対外線を一つのDSドメインとして扱った．対外線の両端に設置

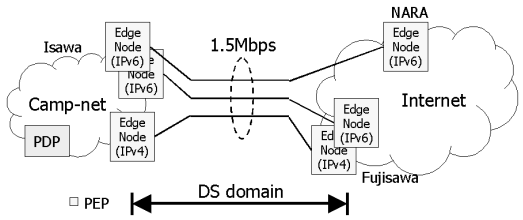

図2　WIDE Camp-net のトポロジ

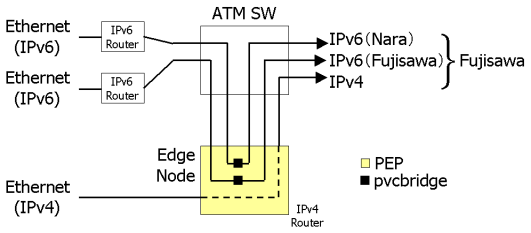

図3　石和における第2層の設定

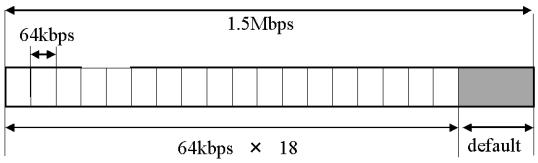

図4　対外線の分割

されたルータを PEP として設定し, これによってマーク付加とトラヒック制御を行った. また, アドミッション制御や課金管理等を行う PDP は WIDE Camp-net 内に設置した.

WIDE Camp-net でもちいた PEP の構成を図3に示す. 石和では IPv4 および IPv6 のインターネット接続性は合計3つの PC ルータによって提供された. ATM スイッチには対外線, IPv4 用 PC ルータおよび2つの IPv6 用 PC ルータが収容されている.IPv6 の2つの PVC は ATM スイッチから一度 PEP である IPv4 ルータの PVC ブリッジを経由し, ATM のままでそれぞれの対地へ接続された. これにより, IPv4 および2つの IPv6 の対外線を1つの PEP に集約し, すべてのトラヒックを集中的に制御することができる.

また, インターネット側には慶應大学湘南藤沢キャンパスに PEP 兼用の IPv4 用 PC ルータおよび IPv6 用 PC ルータ, 奈良先端科学技術大学院大学には IPv6 用 PC ルータを設置した. これらはすべて ATM の PVC によって慶應義塾大学に集められ, 石和と同様にすべて PEP に集約することにより管理の負荷を減少させた.

本実験では, PEP 内で ALTQ[6] を利用してトラヒックの制御を行った. WIDE Camp-net またはインターネットから DS ドメイン (対外線) へのトラヒックには, 石和または藤沢に設置した PEP の入力インターフェイスのキューでマーク付加を行った. また, 出力インターフェイスのキューでは, 入力側で付加された DSCP にしたがい, HFSC[7] と RIO[8] を利用したスケジューリングとキュー制御を行った.

各 PEP と PDP の間の TCP コネクションは, サービス提供期間中は常に維持された.

3.3　サービスクラス

本節では, WIDE Camp-net において提供したサービスのモデルについて説明する.　帯域予約サービスは WIDE Camp-net 内のユーザに対して提供された.　このサービスは, 64kbps の帯域をある始点と終点の組で表されるすべてのフローの集合に割り当て, AF(Assured Forwarding) [9] 方式で優先制御されるというものであった.

まず, 予約のために対外線の T1(1.5Mbps) の帯域を19ブロックに分割した (図4). このうち, 18ブロックはユーザからの予約に割り当てられ, 1ブロックは COPS の制御メッセージや経路制御に利用された. また, 残りの帯域は予約をしていないトラヒックに割り当てられた. また, 輻輳が発生していないときには, 使われていないブロックは他のトラヒックで共有される.

分割されたそれぞれのブロックは 64kbps の帯域をもつ. そのため, ユーザは帯域を予約することにより 64kbps の帯域を確保することができる. ユーザに割り当てた帯域は 18 ブロックであるため, 同時に予約可能な最大値も 18 であった.

ユーザがブロックを予約すると, TRTCM(Two Rate Three Color Maker) [10] によってマーク付加される. ユーザのフローが 64kbps 以内の場合は青色にマークされ, 優先的に処理される. また, 64kbps 以上 96kbps 未満のものについては黄色にマークされる. さらにそれ以上については赤色にマークされ, 予約を行っていないその他のトラヒックと同様に扱われる.

また, 特定のユーザによる帯域の占有を防止するため, 仮想通貨を導入して課金を行った. 仮想通貨の単位を WU(WIDE Unit) とし, 合宿の参加者には, 合宿開始時に一律 2000WU を配布した. また, 1ブロックの予約を1分間行うためには 10WU 必要であるとした.

ユーザが予約を要求するときの様子を図5に示す. 予約要求は次の順で行われる.

(1)　ユーザは予約要求を最寄りの PEP に対して送信する.

(2)　要求を受け取った PEP は, PDP に対して COP-

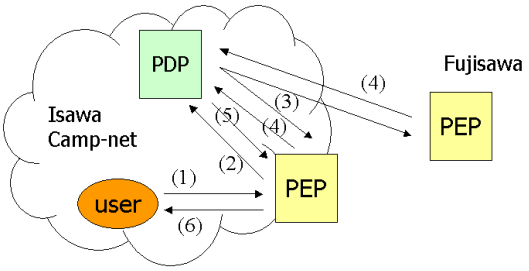

図5　ユーザからの予約要求の様子

SREQ(Request) メッセージを送信してアドミッション制御を求める.

(3)　PDP は要求が正当であるか判断をし, COPS DEC(Decision) メッセージを送信する. 判断基準は, 残余ブロック数, ユーザに仮想通貨の残高などがあげられる.

(4)　PDP からの DEC メッセージの内容が予約を許可する物である場合, PEP は COPS DEC メッセージに記述されているパラメータで帯域の予約を実現する. PEP は, 設定を行なった後に COPS RPT(Report) メッセージを利用し, その結果を PDP に報告する.

(5)　PDP は, 処理中の要求に関連する全ての PEP から COPS RPT メッセージを受け取ると, 要求を送信してきた PEP に対し, COPS DEC メッセージ送信する.

(6)　ユーザは PEP から要求に対する結果報告を受け, 予約要求が終了する.

3.4　予約クライアント

実験では, 予約要求を PEP に送信するためのクライアント「PEPe」をユーザに対して配布した. PEPe から PEP に送信される予約要求メッセージを図 6に示す. ユーザの認証情報や予約フローに関する詳細情報は PEPe メッセージに含まれる.

Diffserv ネットワークでの予約プロセスが終了すると, PEPe は PEP から予約要求に対する結果を受け取る. これには予約の成功・失敗および失敗した場合にはその原因が含まれている.

4.　結果と考察

今回の WIDE Camp-net の DS ドメインは T1(1.5Mbps) の対外線であり, 参加ユーザ数と比較して高帯域回線であった. このため, 各ユーザはそのままでも十分な帯域を利用することが可能であり, このままでは帯域を予約したユーザが優先制御の利益を得ることができない. そこで, 本実験では合宿参加者の帯域予約サービスの利用を促進するために, 対外線上に人為的な輻輳を発生させた.

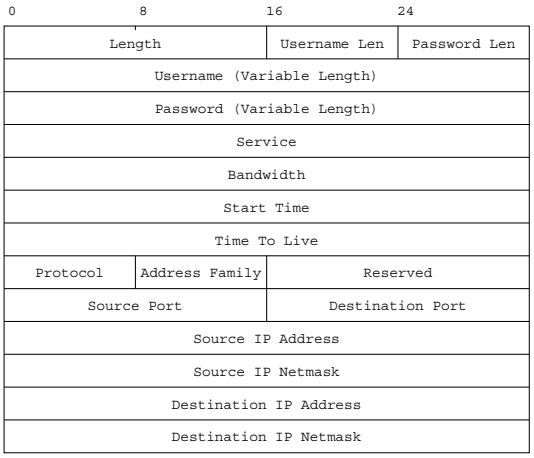

図6　PEPe message

表1　人為的輻輳の発生時間

開始時刻	終了時刻
3/15 18:14:50	3/15 18:19:05
3/16 12:33:37	3/16 13:39:28
3/16 13:42:54	3/16 13:49:03
3/16 16:19:47	3/16 19:39:47
3/16 20:55:27	3/16 23:49:24

輻輳は大量の UDP トラフィックをインターネット側から石和の WIDE Camp-net に対して送信することによって発生させた. UDP トラヒックの送信開始時刻と終了時刻を表 1に示す.

輻輳が発生していないときの対外線のトラヒックを図 7に示す. また, UDP によって輻輳を発生させた時のトラヒックを図 8に示す. これらはインターネット側から WIDE Camp-net 内部へのトラヒックを表している. 輻輳中は対外線上でパケットの喪失が発生した.

次に, ユーザからの予約要求の数を図 9に示す. 対外線が輻輳状態の時に予約要求が最も出されている.

また, 予約エラーの数を図 10に示す. 予約エラーは PDP で予約が拒否されたときに発生する. 表 1より, 予約エラーの多くは対外線の輻輳時に発生している. 本実験ではユーザに提供する帯域を 18 ブロックに限定したが, 図の予約エラーのほとんどは予約用の帯域ブロックの枯渇により発生したものである.

このように, ユーザからの予約の大部分は輻輳時に集中した. 逆に対外線に輻輳が発生していない時には, ユーザからの予約はほとんどなかった. 予約エラーの結果は, ユーザが予約の有用性を理解して輻輳時に予約するという行動を表している.

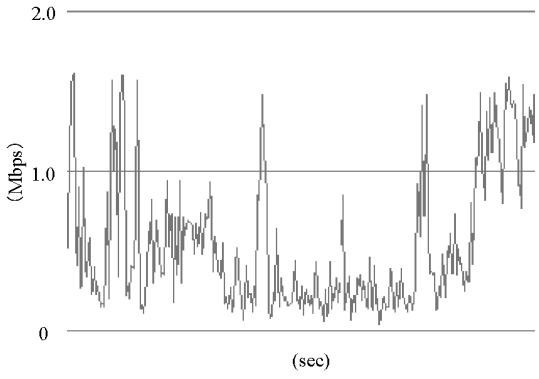

図7 通常時のトラヒック

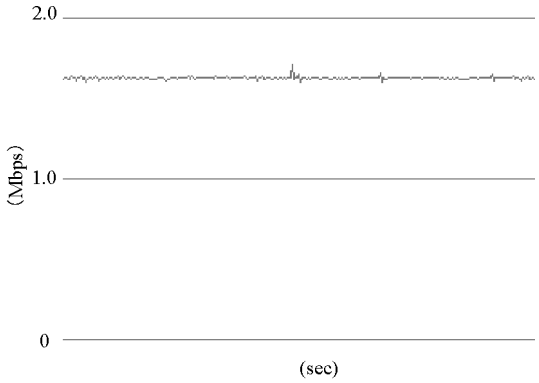

図8 輻輳時のトラヒック

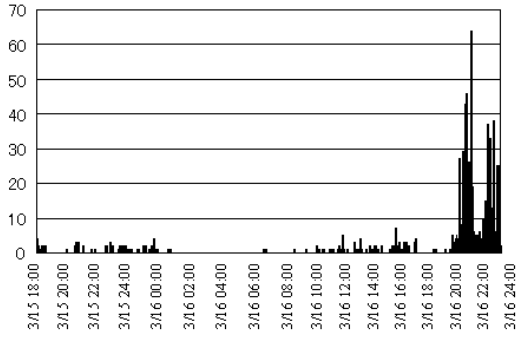

図9 予約要求数

また, ユーザからの予約要求は合宿が進むにつれ増加している. これは, 動的帯域予約がほとんどの合宿参加者にとって初めてのものであり, 始めのうちは予約に対する躊躇があったためであると考えられる. しかし, 帯域予約の有用性がわかると, 合宿参加者のほどんどは頻繁に予約した. これらの結果は, ユーザが輻輳時の通信品質を高く評価すること, そのような状況において優先制御機構が有用であることを示している. 逆に輻輳していないネットワークの通信品質に対しては低く評価され, 優先制御機構はほとんど利用されない.

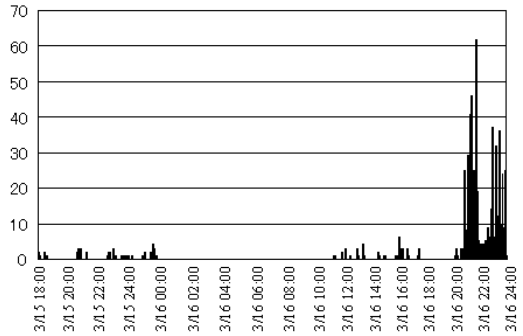

図10 予約エラーの数

インターネットで突発的な輻輳が発生する時間を予測することは困難である. そのため, 上記のようなユーザの挙動を考えれば, 末端のユーザにとっては静的なSLA よりも動的なSLA の方が有用である.

以上から, 我々の提案するDiffserv の階層化モデルにおける末端部分のネットワークにおいて動的なSLA を提供することは, ユーザにとっても有益であり, 妥当なものであると考えられる.

5. おわりに

Diffserv は, VoIP などの新しいインターネットアプリケーションの登場とともに高まる通信品質向上への要求を満たすために標準化された. しかし, その運用モデルは明らかではなく, そのため現在のDiffserv の運用は単一のDS ドメイン内における静的なSLA しか考えられていない.

インターネットのトラヒック状況は時間とともに変化するため, 静的なSLA のみでは限られた状況でしか利用できず, 動的なSLA は必須である. 我々の提案するDiffserv の階層化モデルでは, 末端部分のユーザは動的に品質保証サービスを要求することができる.

このモデルの末端部分が妥当であることを検証するため, 本研究ではWIDE Project の合宿において動的なSLA に対応したDiffserv ネットワークの設計・実装および運用を行った. これにより, 動的な品質保証サービスが提供された環境下におけるユーザのとり得る挙動を理解することができる.

実験の結果, 動的なSLA が提供されるDiffserv ネットワークでは, ユーザは品質保証サービスを頻繁に利用することがわかった. ただし, そのためにはユーザがその有用性を認識する必要がある. また, ユーザは輻輳に対して敏感であり, 輻輳時の通信品質を高く評価していることがわかった.

今後は, 動的SLAを提供することの妥当性についてさらに研究を進めていくとともに, DSドメイン内でのトラヒックエンジニアリング, IXにおける帯域取引等, 我々の提案するモデルの妥当性の検証を進めていく予定である.

参 考 文 献

1) S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. *An Architecture for Differentiated Service*, December 1998. RFC 2475.
2) U. Blumenthal and B. Wijnen. *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)*, April 1999. RFC 2574.
3) J. Boyle, R. Cohen, D. Durham, S. Herzog, R. Rajan, and A. Sastry. *The COPS (Common Open Policy Service) Protocol*, January 2000. RFC 2748.
4) at.el R. Yavatkar, Keith McCloghrie. *COPS Usage for Policy Provisioning*, July 2000. Work in Progress, draft-ietf-rap-pr-03.txt.
5) Wide project. http://www.wide.ad.jp/.
6) K.Cho. A framework for alternate queueing: Towards traffic management by pc-unix based routers. In *USENIX*, 1999.
7) I.Stoica, H.Zhang, and T.S.Eugene Ng. A hierarchical fair service curve algorithm for link-sharing. In *SIGCOMM '97*, 1997. Real-Time and Priority Service.
8) D.D.Clark and W.Fang. Explicit allocation of best effort packet delivery service. *the IEEE ACM Transactions on Networking*, 6(4):362–373, August 1998.
9) J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. *Assured Forwarding PHB Group*, June 1999. RFC 2597.
10) J.Heinanen and R.Guerin. *A Two Rate Three Color Marker*, September 1999. RFC 2698.

著　　者

森島直人
所属　奈良先端科学技術大学院大学 情報科学研究科
住所　〒630-0101 奈良県生駒市高山町 8916-5
TEL　0743-72-5216
FAX　0743-72-5219
Email　moley@kyoto.wide.ad.jp

小川晃通
所属　慶応義塾大学 政策・メディア研究科
住所　〒252-8520 神奈川県藤沢市遠藤 5322
TEL　0466-49-1100
FAX　0466-49-1101
Email　akimichi@sfc.wide.ad.jp

頴原桂二郎
所属　慶応義塾大学 環境情報学部
住所　〒252-8520 神奈川県藤沢市遠藤 5322
TEL　0466-49-1100
FAX　0466-49-1101
Email　popo@sfc.wide.ad.jp

染川隆司
所属　奈良先端科学技術大学院大学 情報科学研究科
住所　〒630-0101 奈良県生駒市高山町 8916-5
TEL　0743-72-5216
FAX　0743-72-5219
Email　somegawa@nara.wide.ad.jp

山口　英
所属　奈良先端科学技術大学院大学 情報科学研究科
住所　〒630-0101 奈良県生駒市高山町 8916-5
TEL　0743-72-5216
FAX　0743-72-5219
Email　suguru@wide.ad.jp