

博士論文

インターネットにおける  
要求駆動型品質制御サービスに関する研究

森島 直人

平成 15 年 2 月 7 日

奈良先端科学技術大学院大学  
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に  
博士(工学)授与の要件として提出した博士論文である。

論文番号: DT9961027

報告者: 森島 直人

審査委員: 山口 英 教授  
山本 平一 教授  
砂原 秀樹 教授  
門林 雄基 助教授

提出日: 平成 15 年 2 月 7 日

# インターネットにおける 要求駆動型品質制御サービスに関する研究\*

森島 直人

## 内容梗概

インターネットが誕生してから 30 年以上の月日が経過した。黎明期には研究への利用が主な目的であったインターネットは、伝送技術や制御技術の発展、さらには、1990 年代初頭における商用化により、日々の生活や経済産業活動を営む際のさまざまなシーンで利用されるようになった。今日では、インターネットはひとびとのコミュニケーションを支える社会基盤として、確固たる地位を確立している。

インターネットが通信する手段として重要な位置を占めるにともない、ひとびとのインターネットに対する要求も大きく変化し、コミュニケーション基盤としての機能に注目が集まるようになった。たとえば、P2P 技術を中心としたあたらしい通信形態や、ストリーミングやアプリケーション等を中心としたコンテンツの多様化、それにともなう通信品質の向上などをその例として挙げることができるだろう。

通信品質になんらかの制御を加えるサービスを実現するためには、制御対象となるインターネット上の特定領域において、資源管理や制御対象機器が示す挙動の一貫性を保持した制御が要求される。また、一般の利用者から発行される通信品質制御への要求は、時間的に粒度の細かな要求として現れることが多い。しかし、自律分散協調性の概念を設計の根底に持つインターネットでは、一貫性を保ちながら断続的に制御することは極めて難しい。本研究は、利用者からの要求によって駆動する通信品質制御サービスの実現を目標としている。特に、インターネットにおける制御層と転送層の分離モデルを立脚点とし、それぞれの層でサービスを実現するために必要な枠組みを検討し、設計と開発、あるいは運用を通じて検証と評価をおこなった。

本論文では、まず、インターネットの階層化構造における制御層と転送層の分離モデルについて述べる。従来の階層化構造では、通信のデータ配送に関する制御機構と転送機構が、ともにインターネット層へ組み込まれて密接に結び付けられている。そのため、制御と転送を分離することが難しく、結果として制御の自律分散協調性と集中性の共存を困難としている。これに対し、制御

層と転送層の分離モデルでは、インターネット層以下の転送機構と制御機構を分離して扱う。これによって、制御機構としてインターネット層以外が提供する機構を準用でき、集中性の概念の導入による一貫性を保持した制御が可能になる。

上記のモデルのもと、制御層と転送層のそれぞれにおいて、要件に関する議論やシステムの設計・実装と評価をおこなった。まず、実利用者を収容する仮設ネットワークを対象としたシステム運用に関して報告する。運用は3回にわたっておこなわれたが、特に最初のものは、実利用者へ要求駆動型通信品質制御サービスを提供した、世界で初めての例として大きな意味を持つ。さらに、この運用経験を通じて、特に利用者の視点から要求駆動型通信品質制御サービスの可能性と問題点を検討した。また、NATなど経路上に介在するエンティティへが優先制御を実現するにあたり問題となることを明らかにし、運用実験においてその対応を行った。

制御層から制御要求を受けた転送層は、サービス提供を実行するために制御層の決定を反映しなければならない。要求駆動型の通信品質制御サービスを考えた場合、要求は時間的に細かな粒度で発行されるため、反映遅延の短縮はサービス実現のための非常に重要な要素となる。本研究では、転送層のひとつとして注目されている MPLS を取り上げ、制御パラメータの反映遅延を縮小する手法を提案した。また、プロトタイプ実装による実験を通じて、この手法の有効性を示した。

制御層からの制御要求が反映されると、転送層の機器群では流入するデータから制御対象となるデータを識別し、対応する処理を決定して適用する。このとき、制御層から転送層の機器群に発行された制御要求がすべて同一であったとしても、それぞれの機器が識別にもちいる手法によって、導出される結果の異なる可能性がある。このような場合、通信品質制御サービスの実現に不可欠な通信経路上における処理の一貫性を保証できない。本研究では、サービス提供者の定義する条件を具現化する部分を通信の識別行程から分離する、二段階分類方式を提案する。これにより、一貫性の喪失などの原因箇所の特定が容易になることが期待される。

本論文では、最初に制御層と配送層の分離モデルについて述べる。次に、要求駆動型通信品質制御サービスを念頭においた、制御システムの実装と運用実験について述べる。さらに、配送層として MPLS における制御パラメータの反映遅延の縮小手法と、通信の二段階分類方式について述べる。

キーワード: 通信品質, ネットワーク制御, インターネット

# Studies on On-demand Quality of Communication Service in The Internet<sup>†</sup>

Naoto Morishima

## Abstract

Thirty years have passed since the Internet was born. At the dawn of the Internet, we put the Internet to good use for some research areas. With the passing of the years, the technologies in the information science has been developed more and more, and in the early years of 1990's, the Internet has been commercialized. Thus, the Internet have come into wide use on not only commercial business, but also daily lives including the economic activities. Up to the present, the Internet had been established a position as the social infrastructure for the global communications.

As technology advance and the status as the principal in these communication infrastructures, the requirements for the Internet has been undergone a sea change. Some concepts and formations for the new communication such as P2P, the diversification of contents like streaming or applications, and the improvement of the qualities of services.

For the realization of the service which was kept the communication quality under control, we need to keep consistency with the resource management and the actions from targeting machines in the specific domain. The other hand, requests demanded from users are closed to granularity of the series of time. However, it is too hard to control them intermittently since the Internet has been designed based on the concept of autonomous, distributed, and cooperative.

The aim at this research is to realize the quality of communication services which are driven with requests from users. As a start point, I have proposed for the separation model which divide data flow on the Internet as forwarding layer and control one. To put these services into practice on the each layer, I have designed and developed the framework, and also verified and evaluated the availability of this proposal model through the network operations in practice.

First of all, I describe the proposal model, that divide data flow into forwarding and control layers. On the current hierarchical structure, the mechanism of forwarding and control for a data flow is closely related, and integrated into the internet layer. It is too difficult to separate them, and this incurs the difficulty of the symbiosis property, that is, autonomous, distributed, and cooperative. On the other hand, I propose the separate model that the forwarding mechanism and the control one is addressed dividing under the internet layer. That is why, we can provide the control mechanism without on internet layer, and also be able to keep concentrate control consistence.

From discussion on the requirements on each layer, I have designed and implemented the concentrated network control systems. This system is designed on a premise that the request driven quality of services is provided on the control layer. It will be necessary to describe the restricts flexibility as the direction for control requests.

On the other hand, I have reported the system operations for the large-scale temporary network with real users. These operation were quite meaningful, since they are the first trial to provide quality of services which drive with requests in the world. Through the 3 times operations, I considered some problems and possibilities of the request driven quality of services from the point at users gaze.

Forwarding layer which accepte control requests from control layer must be reflected the decision to provide services. In the case of request driven quality of services, to reduce the reflection delay is essential topic for providing the services since these requests are brought out in close granularity of the series of time. In this research, I have proposed the method to reduce delay for the reflection of control parameters using MPLS which is widely noticed as forwarding layer. Furthermore, I have verified and evaluated the availability of this proposal method through experimentations using the implemented prototype system.

After the reflection of control requests on the control layer, equipments on the forwarding layer will identify the target data, decide the corresponding actions and apply them. Although all of control requests which will be delivered from control layer to forwarding layer are the same, there is a possibility to be worked out the difference result since the methods of identification data on each equipment are different. In the mentioned case, we cannot guarantee the consistency of process on the communication path which is essential for realization of quality of services. In this study, I proposed the 2-phase classification method. The definite condition is defined by service providers. With this method, I had separated the part for definite condition from the

process for identification of communications, and it will be easy to specify the problems such as the lost point of consistency.

In this paper, I have mentioned the model, the separation layers as control and forwarding one. Then, I had described the system implementation and operation though the request driven quality of services, moreover, I had referred the delay compaction method which is required to establish the LSP with explicit route, and the 2-phase classification method.

**Keywords:** quality of communication, network control, the Internet





# 目次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	人類とコミュニケーション	1
1.2	インターネットとサービス品質	2
1.3	本研究の取り組みと位置付け	4
1.3.1	制御層の取り組み	4
1.3.2	制御層の位置付け	5
1.3.3	転送層の取り組み	6
1.3.4	転送層の位置付け	6
1.4	本論文の構成	7
<b>2</b>	<b>インターネットと要求駆動型品質制御サービス</b>	<b>9</b>
2.1	インターネットの拡大	9
2.2	自律分散協調性	9
2.3	サービスの多様化と通信品質への要求	11
2.4	サービスの品質	11
2.5	静的 SLA と動的 SLA	13
2.6	要求駆動型制御と固定型制御	14
2.7	要求駆動型通信品質制御実現への課題	14
<b>3</b>	<b>制御と送層の分離</b>	<b>17</b>
3.1	自律分散協調性と制御・転送	17
3.2	制御層と転送層の分離構造	19
3.3	制御・転送の分離と要求駆動型制御	19
3.3.1	要求駆動型制御に対する制御層	21
3.3.2	要求駆動型制御に対する転送層	24

3.4	本研究での取り組み	25
<b>4</b>	<b>関連技術</b>	<b>27</b>
4.1	Differentiated Services	27
4.1.1	絶対的な QoS と相対的な QoS	27
4.1.2	DS ドメイン	28
4.1.3	DiffServ CodePoint	28
4.1.4	流入境界ノード	30
4.1.5	中間ノードと Per-Hop Behavior	31
4.2	Multi Protocol Label Switching	34
4.2.1	ラベル・スイッチングと MPLS	34
4.2.2	転送等価クラス	35
4.2.3	ラベル	35
4.2.4	データ転送とラベル操作	36
4.2.5	ラベルスイッチ・パス	39
4.2.6	ラベル配布	40
<b>5</b>	<b>通信品質制御サービス提供実験</b>	<b>47</b>
5.1	実験の概要	47
5.1.1	個別実験の概要	48
5.2	輻輳するリンク上での優先制御サービス	49
5.2.1	camp-net の概要	49
5.2.2	提供サービスの詳細	51
5.2.3	制御層・転送層の設計と運用	53
5.2.4	優先制御の流れ	58
5.2.5	輻輳の発生	58
5.2.6	結果と考察	59
5.3	異なる特性を持つ複数リンクの選択的利用サービス	60
5.3.1	camp-net の概要	60
5.3.2	提供サービスの詳細	64
5.3.3	転送層の設計と運用	67
5.3.4	制御層の設計と実装	68
5.3.5	優先制御の流れ	69

5.3.6	結果と考察	71
5.4	協調しない複数の転送機構による優先転送サービス	76
5.4.1	camp-net の概要	76
5.4.2	提供サービスの詳細	78
5.4.3	転送層の設計と運用	80
5.4.4	制御層の設計と運用	83
5.4.5	NAT への対応	86
5.4.6	仮想専用線利用の手順	88
5.4.7	結果と考察	89
5.5	実験のまとめ	91
<b>6</b>	<b>MPLS における LSP 確立の高速化</b>	<b>93</b>
6.1	LSP 確立の高速化の重要性	93
6.2	要求駆動と事前予測	94
6.3	制約経路独立ラベル配布プロトコル (CR-iLDP)	96
6.4	数値解析	100
6.5	実装と評価	101
6.5.1	設計と実装	101
6.5.2	実験と結果	103
6.6	障害復旧への適用に関する検討	104
6.6.1	事前予測戦略との融合	104
6.6.2	障害通知	105
<b>7</b>	<b>パラメータ・フィルタの 2 段階分離</b>	<b>107</b>
7.1	パラメータ・フィルタ	107
7.2	2 段階分離モデル	108
7.3	実装	110
7.3.1	パラメータの型	110
7.3.2	ステージ 1	111
7.3.3	ステージ 2	111
7.4	応用実装による動作確認	111
7.4.1	応用実装例	111
7.4.2	動作確認	112

7.5	考察と課題	113
7.5.1	ポリシ実現や矛盾検出の検証	113
7.5.2	抽出の高速化	114
8	研究の総括	117
8.1	研究を通して得られた知見	117
8.1.1	全体を通じて得られた知見	117
8.1.2	個別の研究に関する知見	118
8.2	今後の課題	119
8.2.1	通信制御要求発行のインタフェイス	119
8.2.2	個別の細目条件の検討	120
8.2.3	中間エンティティへの対応	120
8.2.4	ドメイン間の制御	120
8.2.5	制御層-転送層間インタフェイスの定義	121
8.2.6	実装品質の向上	122
8.3	結論	122
A	研究業績	131
A.1	本研究に関連する業績	131
A.1.1	学術論文誌	131
A.1.2	国際会議 (査読あり)	131
A.1.3	国内会議 (査読あり)	131
A.1.4	国内研究会等 (査読なし)	132
A.1.5	解説記事	132
A.1.6	講演	132
A.2	その他の業績	133
A.2.1	国際会議 (査読あり)	133
A.2.2	国内研究会等 (査読なし)	133
A.2.3	講演	133
B	用語集	135

# 図一覽

2.1	ホスト数の増加にみるインターネットの拡大 . . . . .	10
2.2	インターネットの基幹部における一日のトラヒックの遷移 . . . . .	13
3.1	従来のインターネットの構造における“点”の自律分散協調 . . . . .	18
3.2	従来のインターネットの構造における制御と転送 . . . . .	18
3.3	制御単位空間の拡張と制御/転送の分離 . . . . .	20
3.4	制御と転送の分離構造における処理のようす . . . . .	20
4.1	DS ドメイン . . . . .	28
4.2	DS フィールド . . . . .	29
4.3	境界流入ノードの構成 . . . . .	30
4.4	Class Selector PHB の DSCP の構造 . . . . .	32
4.5	Assured Forwarding PHB グループの DSCP の構造 . . . . .	33
4.6	MPLS ヘッダとラベル . . . . .	36
4.7	インターネット層での転送処理のようす . . . . .	37
4.8	MPLS での転送処理のようす . . . . .	37
4.9	経路上で見たインターネット層での転送処理のようす . . . . .	38
4.10	LSP でみた MPLS での転送処理のようす . . . . .	39
4.11	LDP におけるラベルの配布方向 . . . . .	41
4.12	LDP の順序ラベル配布制御 . . . . .	43
4.13	TLV エンコーディング . . . . .	44
4.14	CR-LDP による LSP 確立のようす . . . . .	45
5.1	2000 年春の研究会における camp-net の全体構成 (IPv4) . . . . .	50
5.2	2000 年春の研究会における camp-net の全体構成 (IPv6) . . . . .	50
5.3	石和 (研究会開催地) におけるデータリンク層の構成 . . . . .	54

5.4	藤沢におけるデータリンク層の構成	54
5.5	対外リンクの分割	55
5.6	優先制御の手順	57
5.7	輻輳時の往復所要時間	60
5.8	対外リンクのトラヒック (通常時)	61
5.9	対外リンクのトラヒック (輻輳発生時)	61
5.10	通信品質制御要求の発行数	62
5.11	通信品質制御要求の発行数 (エラー分)	62
5.12	2000 年秋の研究会における camp-net の全体構成	63
5.13	AF クラス 1 に対する単価の変動規則	66
5.14	ルータに実装されたリンク選択器の構造	68
5.15	優先制御の手順	70
5.16	利用者からみた優先配送制御の効果	71
5.17	AF クラス 1 における要求継続時間の分布	72
5.18	AF クラス 1 における要求帯域の分布	72
5.19	AF クラス 1 における要求継続時間の変化	74
5.20	AF クラス 1 における要求帯域の変化	74
5.21	AF クラス 1 のサービス単価の変化	75
5.22	AF クラス 1 による対外リンクの帯域利用の変化	75
5.23	2001 年春の研究会における camp-net の全体構成	77
5.24	仮想専用線申請インタフェース	80
5.25	契約履行情報	80
5.26	2 種類の転送機構による仮想専用線の実現	81
5.27	地上回線の両端におけるパケット・スケジューラの共有	82
5.28	FEC とフィルタ規則の分離	84
5.29	制御機構の構成	85
5.30	NAT の動作例	86
5.31	NAT 配下の利用者との契約	87
5.32	写像 $f$ の通知による識別	88
5.33	仮想専用線利用の手順	89
5.34	インターネットから camp-net へのトラヒック流量)	90
5.35	camp-net からインターネットへのトラヒック流量	90

6.1	事前予測戦略に基づく障害復旧 (Protection 1+1) . . . . .	95
6.2	事前予測戦略に基づく障害復旧 (Protection 1:1) . . . . .	95
6.3	CR-LDP による LSP 確立の高速化 (往路) . . . . .	97
6.4	CR-LDP による LSP 確立の高速化 (復路) . . . . .	98
6.5	CR-iLDP による LSP 確立のようす . . . . .	99
6.6	経路断片通知プロトコルの基本フォーマット . . . . .	102
6.7	経路断片通知プロトコルの経路断片通知フォーマット . . . . .	102
6.8	実験環境の構成 . . . . .	104
6.9	CR-iLDP の実験結果 . . . . .	105
7.1	規則の競合 . . . . .	108
7.2	応用実装例の構成 . . . . .	112
7.3	動作確認実験の環境 . . . . .	114





# 表一覧

4.1	DSCP の割り当て規則	29
4.2	Assured Forwarding PHB グループの DSCP	33
5.1	制御対象機器 (1)	53
5.2	制御対象機器 (2)	53
5.3	輻輳の発生時間	58
5.4	輻輳時の往復所要時間	60
5.5	camp-net における ATM 回線と衛星回線の特性の違い	64
5.6	サービス・クラスの比較	64
5.7	サービス・クラス別の往復所用時間	65
5.8	利用者収容ネットワークの種類	78
6.1	経路断片通知プロトコルにおけるデータの種類の補助種類	102
6.2	実験機材の仕様	104
7.1	プロトタイプ実装におけるパラメータの型	110
7.2	フィルタ規則のデータ構造 (IPv4 用)	113
7.3	フィルタ規則のデータ構造 (IPv6 用)	113
7.4	動作確認実験に利用した PC の仕様	113
7.5	分類レコードの抽出に要する時間	114



## 1.1. 人類とコミュニケーション

地球上に初めて生命が現れたのは、約 40 億年前、地球誕生から 6 億年ほど経った頃であるといわれている。それから今日に至る進化の過程で、生命はさまざまな能力を身につけていった。そのなかでも、個体間で情報を伝達する能力の会得は、特筆すべきことであろう。それまで、個体が単体で活動を営んできた生命は、他の個体とコミュニケーションを取り合うことで、ひとつの作業を集団でおこなうことをおぼえた。たとえば、群れで狩りをおこなって獲物を追い詰めるようなことも、互いのコミュニケーションがあつてこそ可能となった。また、世代間における技術の継承も、コミュニケーションによって可能となったもののひとつである。親は行動を子に見せ、子は親の真似をすることによって、生きるために必要な技術の多くを体験的に受け継ぐようになった。このように、コミュニケーションという能力により、生命の活動は、より複雑でより効率的な方向に向かった。

人類は、言語というさらに複雑で効率的なコミュニケーション能力を持つ。体験的なコミュニケーションが非常に粗粒度な伝達手段であるのに対し、言語はそれ自体が非常に限定された意味を持つために、細粒度で正確な伝達が可能であるという特徴を持つ。体験的なコミュニケーションでは難しい抽象的な概念の伝達も、言語を使うことによって可能となる。そのため、言語は体験的なコミュニケーションを補足する形で発展してきたと考えられる。

その後、文字の発明によって人類のコミュニケーションは大きく変化した。文字を利用することで、書類や書籍を通じて後世にさまざまな知識を継承したり、書簡や電報を通じて遠隔地にいるひととのコミュニケーションも実現できる。このような、時間や空間を越えたコミュニケーションはさまざまな制約から言語を中心としたものになり、結果として体験的な伝達のほとんどは失われた。しかし、言語の持つ細粒度な伝達という特徴と時間や空間の制約から解放される利便性から、文字による言語を中心としたコミュニケーションは広く利用されるようになった。

遠隔地とのコミュニケーションが発達してくると、その伝達システムの中に優先度という概念が現れた。その代表的な例は、郵便の配送システムに見ることができる。速達は郵便の配送システムに対し、通常の郵便物と比して優先的な配送を要求するものである。優先配送は伝えたい内容を早く、あるいは正確に伝えたい、と云う要求から生まれたものであろう。

そして現代、人類のコミュニケーションは新たな時代を迎えている。電話やラジオのように音声を利用したりテレビに代表される映像をもちいたものなど、時間や空間を越えるコミュニケーションの中に、言語以外のメディアを利用したものが普及してきたのである。これは、言語を中心としたコミュニケーションの中に、体験的なコミュニケーションを呼び戻そうと云う動きにはかならない。

今日、あらたなコミュニケーション基盤として、インターネットが注目されている。誕生当初は文字によるコミュニケーションが中心であったインターネットにおいても、伝送技術をはじめとするさまざまな技術の進歩によって音声や映像を利用したコミュニケーションが普及し始め、さまざまな用途に応用されるようになった。これにともない、インターネットにおける優先的な配送サービスの提供に対する要求が高まりつつある。このようなサービスによって、動画の再生を滑らかにしたり、より対話的な作業や会話を楽しむことができる。

これらに共通することは、自然であることへの要求——すなわち、あたかもそこにいるかのような臨場感を得たり、機器を直接操作しているかのような自然さに対する要求である。これは、速く安定して伝えることによって体験的なコミュニケーションをより実体験に近づけたいという、無意識のうちに発せられた要望の現れであろう。

## 1.2. インターネットとサービス品質

インターネットは研究用のネットワークとして設計・配備された。初期のインターネットの利用者は研究者であり、その利用目的も研究者の情報共有や計算機の利用がほとんどであった。また、利用者自体が運用者でもあり、ネットワークや各自が利用する計算機、サービスを提供するサーバ等を管理していた。そこでは、サービスの可用性といった運用に関するサービス品質や、遅延や帯域などの通信品質を制御するサービスは提供されていなかった。インターネットにおける通信品質制御の概念は設計当初から考慮されていたわけではなく、商用化や実時間メディアの登場、社会基盤としての成熟などにもともない市場の要求として現れたものである。

インターネットの商用化は、企業の商取引戦略を大きく変えるものとなった。従来からの EDI (Electronic Data Interchange) や紙の伝票による企業間取引の多くは、インターネット上でおこなわれるようになった。また、インターネットの個人利用者の増加により、Web を利用した個人向けの情報発信なども活発になっている。さらに、今まではほとんど見られなかった企業と個人の直接取引や、個人間での取引なども見られるようになった。前者はオンライン・ショッピングモールなど、後者はオンライン・オークションなどに代表される。このようなインターネット上の電子商取引は急激な拡大を見せており、経済産業省は平成 16 年度の国内市場規模が 141 兆円規模に

まで拡大すると予想している [1]。その一方、遅延や経路上のデータ喪失といった通信品質の低下により、多大な損失の発生する可能性が指摘されている。たとえば、インターネットの個人利用者が Web ページを閲覧する場合、そのページが表示されるまでに 8 秒以上を要すると、他のページに移動する傾向のあることが指摘されている [2]。さらに、この理由によってページの閲覧を中止した利用者が、再びそのページを訪れる割合も極めて低いと云われている。このため、取引の機会を逸する理由に、通信品質がひとつの可能性として加えられるようになった。利用者への快適な情報提供を実現するためには、情報を発信するサーバなどの性能や効率の向上とともに、利用者からサーバまでの経路上における通信品質制御が必要となりつつある。

インターネットが商用化によって拡大し、社会基盤として成熟するにしたがい、遠隔医療や工場の遠隔監視などのさまざまな目的で利用されるようになった。遠隔医療は医師の地理的位置に依存することなく高度な診療を提供するもので、医療の地域格差を解消する技術として盛んに研究がおこなわれている。特に、在宅医療は患者や家族の負担を軽くするものとして期待されている。また、工場の遠隔監視は地理的に分散した工場を少数の拠点で監視するものであり、人件費等の削減手法として採用されることが多い。このようなシステムは、一般的に高精度な動画配信や低遅延での情報配信を要求する。遠隔医療では、患者のもとで撮影された高精度の医用動画を医師のもとまで劣化させずに配信しなければならない。また、医師の指示に基づいて医療器具や撮影方法を変更しなければならないため、配信には低遅延であることが求められる。一方、工場の遠隔監視では異常を即座に検出するため、低遅延でのデータ配信が要求される。さらに、データ中の異常が発見された場合、即座に監視カメラの映像を高精度で配信する必要もある。このような、在宅医療や異常時だけといった、突発的に発生する要求のために高品質の通信路を用意するため、要求によって駆動する通信品質制御サービスの実現が必要になっている。

また、近年では緊急時における連絡用通信網として、インターネットを利用する方法についても検討されはじめた [3]。これは、地震等の災害時や国家の緊急時にインターネットを連絡網として利用する試みである。この場合、緊急を要する通信の他の通信に対する優先的な配送を考えるだけでなく、すでに優先制御を受けている通信に対しても割り込み処理をおこない、要求される帯域を確保する等の処理が必要になる。

このように、インターネット上ではさまざまな場面で通信の優先制御が求められている。また、インターネットの利用方法は多岐にわたり、それぞれの利用方法で異なる制御が必要とされている。しかし、先に述べたように、インターネットの設計に通信品質制御の概念が含まれていなかったため、多くの要求があるにも関わらず、現在においてもなお、広域ネットワークにおいて柔軟な通信品質制御を実現するには至っていない。

インターネットは、パケット・スイッチングに基づく最大努力型のネットワークである。すな

わち、通信の主体間に介在する個々のエンティティは、原則的に全てのデータを公平に扱い、それぞれに対して遅延や喪失を最小限にとどめようとする。その一方、その品質に対しては一切保証がなく、輻輳時には急激な性能悪化・通信品質劣化をもたらす可能性がある。そのため、さまざまな要求に対応できる柔軟な通信品質制御技術の確立が求められている。

## 1.3. 本研究の取り組みと位置付け

インターネットが社会において果たす役割が大きくなるにともない、多種多様な目的と手段で利用されるようになり、その結果としてさまざまな通信品質が求められるようになった。本研究では、利用者が発行する要求によって、即座に駆動する通信品質制御サービスの提供を目標とした。

本論文では、まず、要求駆動型制御や多彩なサービスを実現するために、インターネットの制御機構と転送機構の分離について論じる。この考え方と枠組みによって、あたらしい制御機構の導入や、特定領域に対する整合性を維持した資源管理が可能になる。次に、このモデルを立脚点として、要求駆動型通信品質制御サービス提供の実現に必要な要素技術の検討や、試験運用による問題点の洗い出しをおこなった。

### 1.3.1. 制御層の取り組み

制御層では、人間の介在しない要求駆動型制御において、多彩なサービスを実現するための制御構造について検討した。また、その制御構造を導入することで、実利用者を収容する仮設ネットワークにおいて、要求駆動型通信品質制御サービス提供の実証実験に取り組んだ。

従来のインターネットでは、サービスとして接続性を提供することが一般的であった。契約は実社会で取り交わされている契約書と同様のものに基づいて締結される。また、契約の有効期間も月単位から年単位と、人間の日々の活動サイクルと比較して長期にわたることが多い。ここでは接続性の維持が最重要サービスであり、サービス提供者の観点からは、他のサービス提供者との差別化を図るために接続の可用性の向上が重要な課題であった。また、障害時の復旧に要する時間等を明示することも、可用性の向上に関連する取り組みのひとつであると云える。

近年では、接続性に加えて付加価値サービスの提供が普及の兆しを見せている。仮想専用線の構築やそれにとまなう通信品質の保証、サービス提供者の管理領域内における伝送遅延の保証などである。しかし、これらの多くは既存の接続性提供サービスに対する、いわば付帯サービスの形で提供され、契約は接続性サービスと同様に、契約書に準ずる形態で締結されることがほとんどである。このような契約では、その処理に人間が介在する。したがって、契約処理からネットワークの制御に至るまでの過程の多くに、なんらかの形で人間が携わることが一般的である。ま

た、このような形態をとることから、契約処理の開始からネットワーク制御の結果によるサービスの開始までが、数時間から数日、あるいは数ヵ月といった、ある程度の期間を要することも多く見受けられる。

一方、上記のようなサービスよりも遥かに細粒度のネットワーク制御に対する要求の存在も明らかとなってきている。人間の日々の生活には「今日だけ」「今から」などという、極めて短時間、かつ、突発的な要求が多く見られる。個人利用者がインターネットで情報を収集する場合、その多くは計画されていない突発的なものであろう。また、動画や音楽配信を楽しむ場合にも、その通信が1日以上継続することは稀である。このような個人利用者の短時間で突発的な通信に対して発生する品質制御要求は、その通信の粒度と同様に細粒度であることが予想される。

細粒度の品質制御を実現するためには、人間が介在しない機構によって処理を完結させることが望ましい。サービス提供者の観点からは、要求処理のための人件費が莫大になること、利用者の観点からは、要求から契約の締結、サービスの開始までにある程度の時間を要してしまうこと、などが理由として考えられる。通信品質要求に対する処理は、固定制御では人間がおこなっていた多くの判断をとまなう。このような個別の条件を細目条件と呼ぶ。

現在まで、QoS 経路制御をはじめとするいくつかの細目条件が研究されてきたが、これらの細目条件を組み合わせるサービスを提供する実験はおこなわれてこなかった。そのため、実利用者に対してサービスを提供するために不足している細目条件などは検討されていない。また、このような実験がおこなわれてこなかったことから、提供されるサービスに対する利用者の挙動なども明らかになっていない。さらに、現状の接続性を提供するサービスとは利用方法が異なるために、インターネットのトラフィック流量の変化が観測される可能性や、それにとまなうネットワーク再構成の必要性があるかもしれない。

一般の利用者に対してサービスを開始するためには、このようなサービスの実利用者に対する試験的な提供をおこなうことで上記のような不明点を明らかにし、多くの経験を蓄積する必要がある。そこで、実利用者を収容する仮設ネットワークにおいて、要求駆動型通信品質制御サービスの提供実験をおこなった。実験は3度にわたり、さまざまなサービス内容を定義した。これらの実験から、定義されたサービスと利用者の挙動の関係や、中間エンティティへの対応の必要正当が明らかになった。

### 1.3.2. 制御層の位置付け

要求駆動型通信品質制御サービスの提供が困難な理由のひとつには、制御機構の不在があげられている。転送層はハードウェア・機構ともに高速化と高機能化を実現しつつある。その一方、人間の介在しない機構で複雑なサービスを定義することは、サービスに対する無限の可能性から、汎

用的な枠組みの構築は難しい。

このようなサービス定義の困難さと実証実験の困難さゆえに、要求駆動型品質制御サービスの運用評価に関しては、実績報告がなかった。したがって、サービスに対する利用者の挙動や、実現に向けての不足技術要素などが明らかになっていなかった。本研究では、実利用者を収容するネットワークでは初めてとなる、サービス提供実証実験をおこなった。これにより、サービスの可能性を確認するとともに、利用者の挙動や必要技術要素の一部を洗い出すことができた。このような経験の蓄積の第一歩を踏み出したことは、サービス実現に向けた大きな成果であろう。

### 1.3.3. 転送層の取り組み

転送層の取り組みとしては、MPLSにおけるLSP確立の高速化と、パラメータ・フィルタの二段階分離モデルの検討に取り組んだ。

一般的に利用者からの制御要求は、なんらかの形で契約が締結され、制御が転送層に反映されることでサービスが開始される。人間が介在しない要求駆動型制御でも、この過程は大きく変わることはない。しかし、転送層に制御内容を反映させるまでの遅延に対する要求は大きく異なる。前述したように、契約書に基づく契約は時間的な粒度が粗いため、人間の生活サイクルと比較しても、日や週単位での遅延が許容されることが多い。それに対し、要求駆動型制御では、要求に対する時間粒度が極めて細かいために、反映時間の短縮が要求される。

本研究では、転送層としてMPLSを取り上げ、明示的経路によるLSP確立の高速化手法を提案した。さらに、提案手法のプロトタイプシステムを実装し、実験を通じてその有効性を確認した。

また、転送層における通信の優先制御は、対象となる通信を識別することから始まる。この機構を、パラメータ・フィルタと呼ぶ。パラメータ・フィルタは、識別するための情報の設定と、それらの情報から最適な情報を抽出する過程からなる。しかし、同一の識別情報を設定したとしても、抽出の過程の違いによって抽出結果が異なる可能性があり、この場合には通信経路上における優先制御の一貫性が失われる。

本研究では、このような抽出結果を二段階に分割することで、競合の発生個所を特定する手法を提案した。

### 1.3.4. 転送層の位置付け

MPLSは転送機能を制御層から分離する機構として注目されている。VPNやトラフィック・エンジニアリングなど、制御から分離することによる利点を活かしたサービスが始まっている。したがって、要求駆動型通信品背筆制御サービスにも、MPLSが転送層の中心となっていくことが予想



される。このような状況で、明示的に指定した経路で LSP の確立するのに要する時間を短縮できたことは、要求駆動型サービスだけではなく、MPLS の障害復旧の観点からも重要な成果である。

また、パラメータ・フィルタが抽出する結果の矛盾に言及した先行研究はない。しかし、特定領域内で転送機構が示す挙動の一貫性は、通信品質制御の根底を支える重要な前提である。したがって、曖昧な部分の切り出しによって矛盾の検出ができる可能性を示したことは、転送層の基礎技術のひとつであるパラメータ・フィルタの設計の参考になるであろう。

## 1.4. 本論文の構成

本章に続き、2章ではインターネットの自律分散協調性について述べる。また、サービスレベル契約について概観し、要求駆動型通信品質制御の概念について説明する。3章では本研究の立脚点となる、インターネットの階層構造における制御層と転送層の分離について論じる。また、この概念を要求駆動型通信品質制御に適用し、サービスを実現するための検討項目と本論文での取り組みについて述べる。4章では、通信品質制御を実現するために標準化された、DiffServ と MPLS の技術について概説する。

5章では、要求駆動型通信品質制御サービスの実証実験について述べる。この実験では、実利用者を収容する仮説ネットワークにおいて、利用者から発行される要求に対する通信品質制御サービスを提供するものである。この実験を通じて明らかになった中間エンティティへの対応やパケット・スケジューラの共有の必要性、また、現在まで報告例のない、サービス・クラスの定義と課金に対する利用者の挙動などについて述べる。6章では、転送機構への制御パラメータ反映の高速化に関し、MPLS における LSP 確立の高速化手法を提案する。また、提案手法の実装と実験による評価について述べる。7章では、パラメータ・フィルタの二段階分離について述べる。

最後に8章では、得られた知見と今後の課題を述べ、本研究を総括する。



インターネットは1990年代の商用化にともなって爆発的に普及し、社会基盤としてその地位を確固たるものにした。また、普及にともなって通信品質制御への要求が表面化した。インターネット上での通信品質制御には多種多様な形態があり、その一部はすでに実サービスとして提供されているものもある。本章では、インターネットの拡大、および、拡大を可能にしたインターネットの自律分散協調性について述べる。また、通信品質に関する契約や制御方式をその特徴に応じて分類し、本研究の前提となる要求駆動型通信品質制御サービスのモデルを明らかにする。さらに、自律分散協調性という特性を持つインターネットでは、このようなサービスの実現が困難であることを述べ、要求を実現するために直面している課題について論じる。

## 2.1. インターネットの拡大

インターネットは、1969年に米国防総省のARPA(高等研究計画局)によって配備されたデータ・ネットワークに端を発する。このネットワークはARPANETと呼ばれ、導入当初はいくつかの大学や研究所を相互に接続する小規模なものであった。その後、技術の進歩やネットワーク研究の拡大とともに、接続拠点の増加や他のネットワークとの相互接続を繰り返すことになる。

1980年初頭、ARPANETでTCP/IPを利用した通信が始まった。いわゆるインターネット(The Internet)の幕開けである。接続拠点が緩やかに増加していたARPANETへの接続ホスト数も、これを期に大幅な伸びをみせるようになる。特に、1990年代半ばにはインターネットの急速な商用化が進み、規模の拡大はさらに加速度を増すことになった。現在でもインターネットの拡大の勢いは収まることなく、接続ホスト数も指数関数的な増加を続けている(図2.1)。

## 2.2. 自律分散協調性

インターネットがこのような拡大を見せた大きな要因のひとつに、自律分散協調性の概念を根底において設計されていることがあげられる。自律分散協調性とは、あるシステムを構成する要素が、

- 自律的に動作する

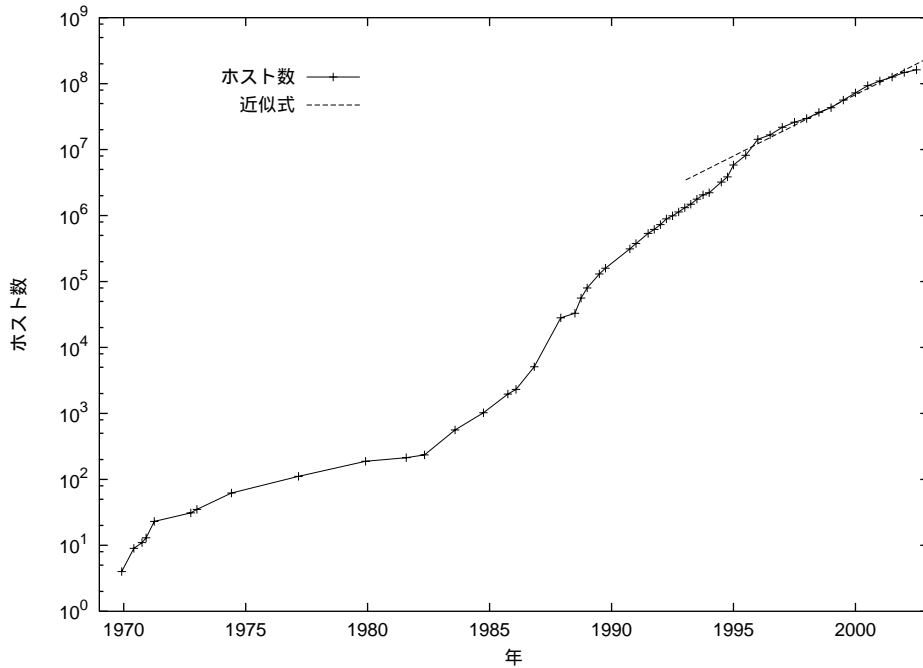


図 2.1: ホスト数の増加にみるインターネットの拡大

出展: 『Hobbes' Internet Timeline v5.6』 (Copyright (c) 1993-2002 by Robert H Zakon)[4]。インターネット (The Internet) 時代の幕開けとなる 1980 年代初頭から、急激な拡大が見られる。商用化が進んだ 1990 年代半ばから現在までの増加は、ほぼ  $8E + 06 * exp(0.43 * (year - 1995))$  と一致する。

- 空間的に分散している
- 相互に作用することでシステム全体の機能を成立させる

という性質を持つことをいう。したがって、自律分散協調性を持つシステムには、全体を統括・制御するエンティティは存在しない。システムを構成する要素が相互に機能を提供し、全体の機能が形成される。

ARPANET が設計された当初、ネットワークを取り巻く技術の主流は中央集中型であった。それにもかかわらず、ARPANET が自律分散協調システムとして構築されたのは、局所的な軍事攻撃を受けてもシステム全体が停止しないためであるといわれている。中央集中型システムでは、特定部分に制御を司る機能が集中している。このため、制御機能を持つ一部分の障害が全体の機能低下に直結する恐れがある。これに対し、自律分散協調システムでは、障害箇所の機能を他の部分で代替することが可能である。このため、一部分の障害でシステム全体に大きな機能低下はもたらされにくく、結果として高い可用性を維持することができる。

また、自律分散協調性を持つシステムは、可塑性・対応性に優れているという特徴も持つ。構成要素の変化や増加に対しても、その事象を他の構成要素に伝搬する必要がなかったり、伝搬する範囲を特定領域に限定することができるため、システム全体の拡大にも柔軟に対応できる特徴がある。インターネットが今日の規模にまで拡大したのも、この性質に因るところが大きい。

## 2.3. サービスの多様化と通信品質への要求

インターネットの黎明期には、利用者の多くが研究者であった。各組織に点在するスーパーコンピュータの共同利用や、研究者間のコミュニケーションを主な目的としていたのである。また、ネットワーク技術そのものも研究の対象であった。当時のインターネットでは、経路制御技術が未熟であったため、到達性の喪失がしばしば発生していた。そのため、利用者のインターネットに対する最大の関心事は、通信の到達性であった。

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になった。さらに、電算機技術や伝送技術等の発展は、インターネットを実用的な社会基盤へと変貌させた。利用者の大半は研究やインターネットの運用に関わりがなく、メールによるコミュニケーションや WWW(World Wide Web) を介しての情報収集を主な目的としている。

利用者や利用目的の変化にともない、インターネットに要求されるサービスも大きく変化してきた。たとえば、ストリーミング等を中心としたコンテンツや、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などは、これらの要求に応じて登場したものであるといえよう。

このような新しいサービスの多くは、なんらかの形で通信品質を要求する傾向にある。ストリーミングのなめらかな再生には安定した帯域を確保できることが重要であるし、VoIP では遅延の抑制によって自然な会話をすることができる。インターネットにおける通信は、到達性のみの提供から、到達性を前提とした通信品質の提供を求められるようになりつつある。

## 2.4. サービスの品質

インターネット上における一部のサービス提供者は、利用者に対する契約の一環としてサービス品質保証を提供している。これらの提供者の多くは、サービス品質保証を履行できなかった場合、利用者に対してサービスに対する料金の一部を払い戻している。このような、サービス品質に関する契約を、サービスレベル契約 (Service Level Agreement; SLA) という。

SLA におけるサービス品質の要素としては、次のようなものを例としてあげることができる。

#### サービス開始遅延

提供者と利用者間で契約が締結されてから、サービスが開始されるまでの期間が一定以下であることを保証するもの。

#### 可用性

一定期間内でサービスが利用できる時間の割合を保証するもの。

#### 障害復旧時間

利用者へのサービスに障害が発見された場合、事前に契約した時間内に復旧することを保証するもの。

#### 障害通知時間

提供者が利用者へのサービスに障害を発見した場合、事前に契約した時間内に利用者へ障害内容等の事実を通知することを保証するもの。

#### 網内伝送遅延

提供者が指定する領域内の任意の2点間の通信に対し、その往復所要時間(あるいはその平均値)が事前に契約した値を下回ることを保証するもの。

#### 帯域保証

提供者が指定する領域内の任意の2点間の通信に対し、その通信が常に占有できる帯域を保証するもの。

このような要素と、それに付随するパラメータの集合によって定義されるサービスの単位を、サービス・クラスと呼ぶ。

上であげた要素の多くは、計測によって定量的に得られる絶対的性能を保証するものである。現在、サービス提供者によって提供されているサービス・クラスの多くは、このような絶対的性能を保証する要素から構成されることが多い。一般的にこのようなサービスを、Quality of Service(QoS)と呼ぶ。インターネットにおける通信品質のQoSを実現するために、Integrated Service (IntServ) [5]が標準化されている。

それに対し、近年ではサービス・クラス間の相対的な品質差・性能差によってサービスを定義することがある。このようなサービスを Class of Service(CoS)と呼ぶ。インターネットにおける通信品質制御サービスとしての CoS は、絶対的な性能を保証するものではないため、QoS に比べて従来のインターネットの枠組みに近く、実現しやすいという特徴をもつ。その一方、サービス・クラスの厳密な定義が難しく、理解しやすい形で提示することが困難である。通信品質の CoS を実現する枠組みとしては、後述する Differentiated Services (DiffServ) が標準化されている。

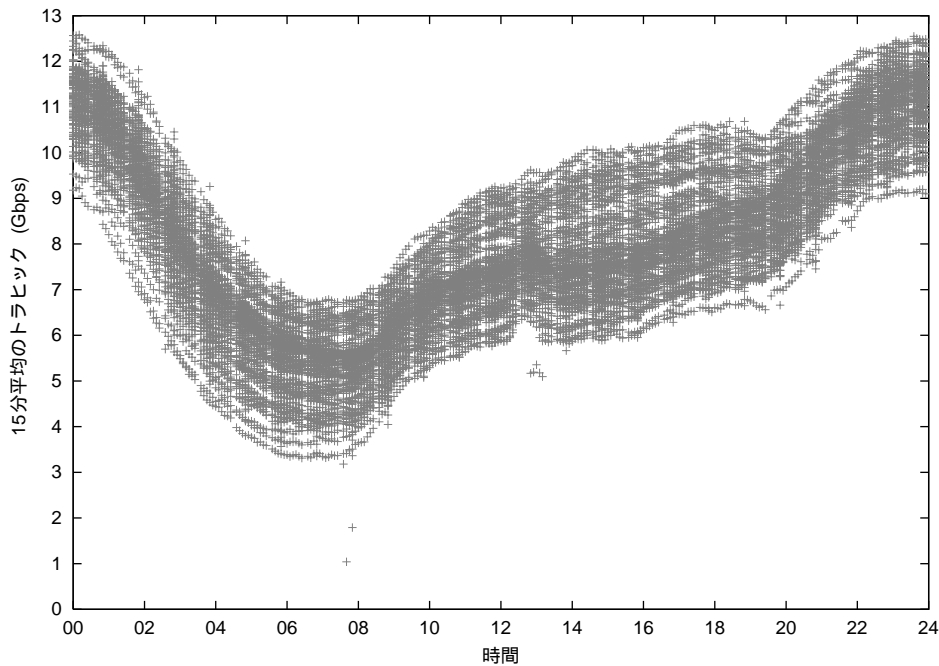


図 2.2: インターネットの基幹部における一日のトラフィックの遷移

日本におけるインターネットの中核、NSPIXP2 [6] における一日のトラフィックの遷移を示す。規則正しく遷移していることが観測できる。

## 2.5. 静的 SLA と動的 SLA

インターネット上における個々の通信に着目した場合、その挙動には非常にバースト性があることが知られる。個々の通信はその発生から収束が突発的に発生し、その挙動を予測することはできない。一方、多くの通信が集約されたトラフィックに着目すると、バースト性が消滅して予測可能となる。図 2.2 は、日本のインターネットの基幹部における、一日のトラフィックの変動を表す。この図に見るように、非常に規則正しい挙動を示していることがわかる。

インターネットの基幹部では、このような現象に着目し、時系列で制御内容を変更するような SLA が考えられる。たとえば、トラフィックの少ない夜間は特別な制御をおこなわず、昼間の混雑時に帯域を確保しながら優先制御をすることで、一日の通信品質を一定に近づけるようなことが考えられる。また、データ同期のため特定の時間帯だけに通信品質が要求される場合にも有効であろう。このような、時系列で内容が変化する SLA を動的 SLA と呼ぶ。一方、内容に時間的要素を含まない契約を静的 SLA と呼ぶ。

動的 SLA はその性質上、継続期間が長期にわたる契約に現れる。また、ある限定的な用途を念頭においた契約である場合が多い。逆に、静的 SLA はさまざまな用途に利用する VPN や、短期間で終了する契約などに現れる。

## 2.6. 要求駆動型制御と固定型制御

現在のSLAの締結は、社会一般で取り交わされる契約と同様に契約書に基づいておこなわれている。そのため、契約内容はある程度包括的になる傾向にあり、対象となる制御もマクロな指標に基づいたものになることが多い。さらに、契約書類の物理的な移動や人間の介在する処理に要する時間により、契約期間も数ヵ月から数年といった粒度の粗いものになる。この例としては、仮想専用線 (Virtual Private Network; VPN) があげられる。VPNは企業における事業所間をネットワークで直結するコストを削減するため、インターネット上に帯域の確保や暗号化などの特殊な処理を施した経路を設定し、仮想的な専用線として利用するものである。したがって、VPNは物理的な専用線と同様に長期にわたる契約であることが多い。このような、契約書によって締結される、時間的・空間的・品質指標的に粒度の粗い制御を固定型制御と呼ぶ。

その一方、固定型制御と比較して遥かに細かな粒度のサービスに対する需要も、明らかに存在する。たとえば、インターネットの個人利用者が使用するアプリケーションは、必要帯域や単位時間あたりの総トラフィック量が少なく、通信の継続時間も短い傾向にある。また、利用者が通信をおこなうにあたり、その通信が計画的なものであることは少なく、したがって特定の通信に対する品質要求も突発的に発生しがちである。そのため、個人利用者の発する通信品質要求は、さまざまな側面で粒度の細かなものであることが多い。

このような要求に対する制御を実現するためには、人間の介在しないエンティティによって通信品質要求を受理し、そのエンティティの判断に基づいて、即座に品質制御をネットワークに反映させるための仕組みが必要である。このような、突発的に発生する品質要求に応じてネットワークの挙動に変更を加える、粒度の細かな制御を要求駆動型制御と呼ぶ。

通信品質制御を実現する場合、対象となる通信にネットワークが保有するなんらかの資源を割り当てる。一般的に資源は有限であるから、制御対象となるネットワーク上の特定領域における資源は、全体で一貫性を持って管理されなければならない。固定型制御では、契約や運用に携わる人間が資源管理をおこなうことで、資源管理の一貫性を確保してきた。一方、要求駆動型制御では人間が介在せずに資源消費をとまなう制御がおこなわれるため、全体の資源管理を司る機構が必要である。

## 2.7. 要求駆動型通信品質制御実現への課題

利用者の要求に応じてネットワーク上の特定機器の挙動に変更を加え、その結果として通信品質を制御するような機能は、インターネットが設計された時点では想定されていないものであった。



通信品質を制御する場合、インターネット上のある領域に対して、一貫性のある制御をおこなうことが要求される。たとえば、通信経路上の機器は、品質制御の対象となる通信を等しく識別しなければならない。また、識別した通信を同一に処理する必要がある。さらに、品質制御のために確保した資源は、制御対象領域で矛盾なく管理されなければならない。

自律分散協調性を持つシステムでは、構成要素間で状態の同期をとることが容易ではなく、上記のような一貫性を要求する制御や判断を効率的に処理することができない。そのため、分散した構成要素に同一の動作を強制したり、一貫性を保持したまま分散したデータに処理を加える場合には、集中システム概念を導入した制御をおこなうほうが効果的である場合が多い。このようなことから、インターネットにおいて通信品質の制御をサービスとして提供するためには、インターネットが動作する仕組みになんらかのパラダイムシフトが必要であろう。



既存のインターネットでは、その構造上の制約からあらたな制御構造を組み入れることは困難であった。これに対し、より柔軟な制御を実現する試みのひとつとして、制御と転送の分離構造が議論されるようになった。本章では、制御層と転送層の分離について述べ、それぞれの層に対応する要求駆動型制御サービスの検討項目について論じる。

### 3.1. 自律分散協調性と制御・転送

2.2 節で述べたように、既存のインターネットは自律分散協調性を基本的な概念として構築されている。すなわち、制御単位となる“点”が互いに協調して必要な情報を交換しながら、それぞれが自律的に動作を決定・実行している。“点”の自律的な動作の概念を、図 3.1 に示す。たとえば、インターネットの中核を構成するルータは、このような“点”のひとつである。ルータはそれぞれが保持する経路情報を交換する。その情報から転送のために必要な次ホップや出力インタフェースを自律的に算出し、その結果に基づいて流入するデータを転送する。

制御単位の内部構造、および、その連携のようすを図 3.2 に示す。一般に制御単位の内部では、インターネット層以下が制御と転送を密接に組み合わせて動作するように設計されている。これらは一体化した機構として動作するため、インターネット層における転送処理を、同じ層に位置する経路選択から分離して実行することはできない。

現在、これまでに述べてきたような多様なサービスを実現するために、インターネットには高性能性と高機能性が求められている。高性能性は処理や転送に要する時間を短縮する高速化であり、個々の分散したモジュールの構造改善によって実現できると期待される。一方、高機能性は利用者の多彩な要求に対して即座に対応できる柔軟性であり、実現にはネットワーク内の特定領域において、さまざまな資源管理や挙動の一貫性を維持した制御が要求される。このためには、個々の制御単位が緩やかに協調しながら独立して動作するのではなく、複数の制御単位をひとつの集合として扱い、より包括的な制御をおこなわなければならない。すなわち、ネットワークの制御単位空間を“点”から“線”または“面”へと拡張する必要がある。しかし、制御と転送が一体化されたモジュールで実現されている現在の構造を維持したまま、制御単位空間だけを拡張することは困難であり、後の拡張性に関しても大きな問題を残すことは容易に予測できる。

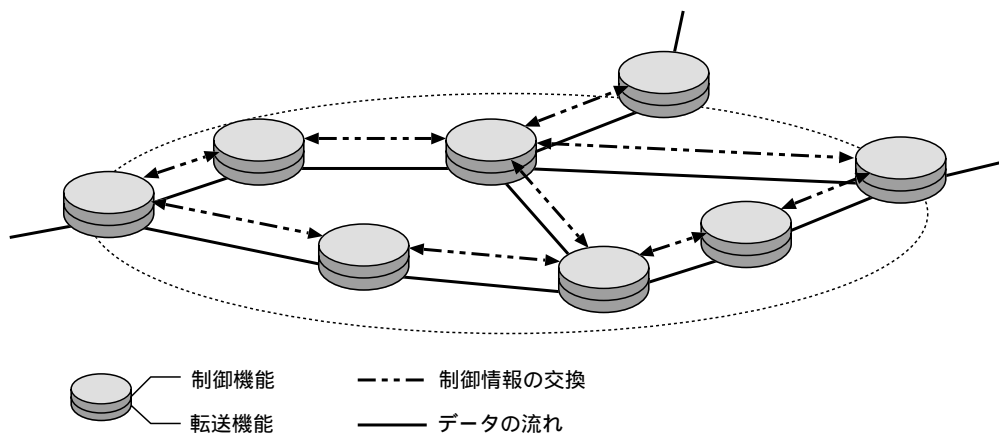


図 3.1: 従来のインターネットの構造における“点”の自律分散協調

従来のインターネットでは、制御単位が点在する構造になっている。それらの“点”は協調して互いに必要な情報を交換し、自律的な意思決定をおこなう。

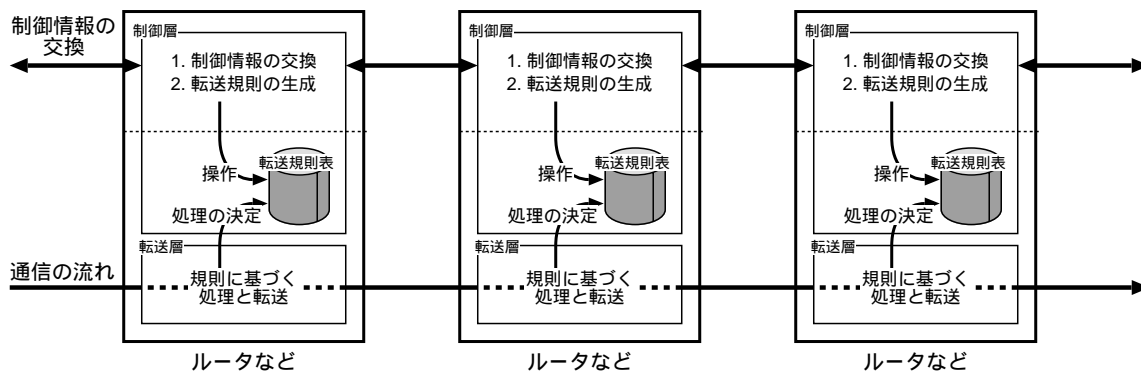


図 3.2: 従来のインターネットの構造における制御と転送

制御と転送の機能は、インターネット上に分散しているそれぞれのエンティティに組み込まれている。隣接するエンティティは制御に必要な情報を交換し、そこから制御の詳細を各自で導出、転送処理に反映させる。

## 3.2. 制御層と転送層の分離構造

前節で述べた問題を解決するためのあたらしい概念として、制御と転送の分離モデルが議論されるようになった。このモデルでは、経路制御に代表されるネットワーク制御機構とトラヒックの転送に携わる機構を分離して扱うことで、ネットワークを面として扱うことができ、より柔軟なネットワーク制御が可能になる。制御と転送の分離の概念を図 3.3 に示す。制御と転送の対応関係を切り放すことで、転送機構の制御にインターネット層以外が提供する機構を利用することができ、制御と転送の柔軟な組み合わせで多彩なサービスを展開できることが期待される。

制御と転送を分離するためには、いくつかの方式が提案されている。しかし、現在のインターネット層に対して、その整合性を維持したまま構造を改変し、制御と転送を分離することは困難である。また、インターネット層の枠組み自体を根本から見直し、大幅な変更を加えることは、機構の導入と展開の観点から避けられるべきである。このことから、インターネット層の構造を変更することなく、外部からインターネット層の決定に介入する手法、あるいは、インターネット層と独立した構造を導入する手法がとられる。特に後者は、インターネット層の直下に sub-IP と呼ばれる層を構成し、転送機構をインターネット層から分離する手法が注目されている。この機構の代表的な例としては、MPLS をあげることができる。また、転送時の処理に特化した DiffServ も、制御層を分離した転送の補助機構として利用される。

いずれの場合でも、転送機構は制御機構からあらかじめ制御内容を通知される。IP データグラムが流入してきた際には、その内容にしたがって固定的な処理と転送をおこなう。ある制御領域における転送のようすを図 3.4 に示す。

## 3.3. 制御・転送の分離と要求駆動型制御

要求駆動型制御による、利用者からの通信品質要求に対するネットワーク挙動の変更を考える。利用者の管理領域内であれば、機器の直接的な設定変更などにより、ある程度の要求を満たすことができる。しかし、多くの通信はインターネット上の複数の管理領域を横断するものであり、利用者の設定による解決法は一般的ではない。このような場合には、要求を受理するネットワーク側のサービスと対応が必須である。

広域ネットワークにおいて通信品質制御を実現するためには、ネットワークが保持する資源の管理が必要である。このような場合、独立分散制御では領域内での資源管理の整合性を維持することが難しい。また、利用者の要求に応じて特性の異なる経路を選択的に提供する必要がある場合には、インターネット層の経路制御と独立した形で転送層に制御を加える必要がある。このよ

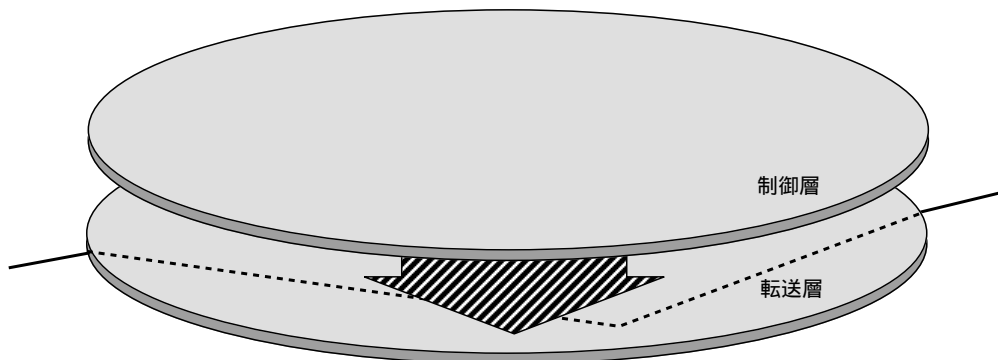


図 3.3: 制御単位空間の拡張と制御/転送の分離

制御と転送の分離モデルでは、転送層を制御から分離した面として扱い、制御機構の包括的な制御によって一貫性を維持した動作を示す。

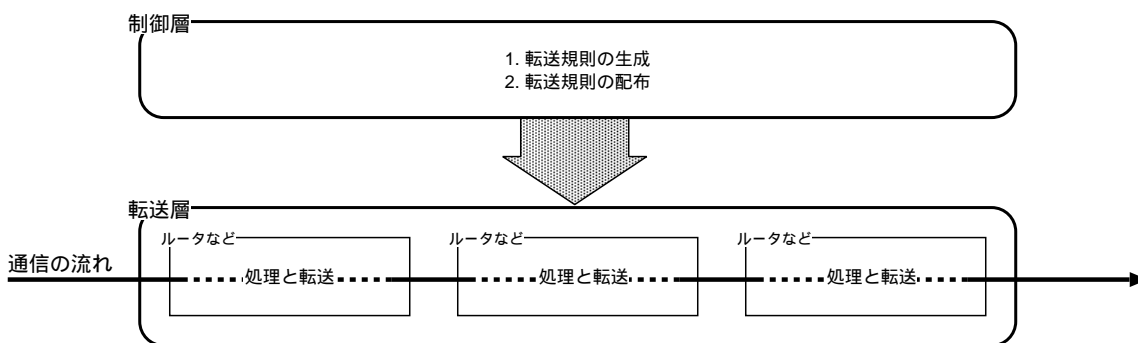


図 3.4: 制御と転送の分離構造における処理のようす

既存のインターネットにおける制御と転送の概念図を示す。制御と転送の機能は、インターネット上に分散しているそれぞれのエンティティに組み込まれている。隣接するエンティティは制御に必要な情報を交換し、そこから制御の詳細を各自で導出、転送処理に反映させる。

うなサービスの実現にも、前節で述べた制御と転送の分離が有効である。利用者が制御機構に対して制御要求を発行、SLA 締結にともなって転送層を包括的に制御すればよい。

### 3.3.1. 要求駆動型制御に対する制御層

要求駆動型の通信品質制御サービスでは、利用者やそれにかわるなんらかのエンティティからの品質制御に対する要求が発生してから、転送機構に反映されるまでの過程は、以下の4つの部分に分類することができる。

- 利用者とシステムの折衝
- アドミッション制御と制御パラメータの導出
- 複数の管理主体間における折衝
- 制御パラメータの配布と反映

#### 利用者とシステムの折衝

利用者と要求駆動型サービスの関係では、そのサービスが提供される対象によってさまざまな状況が考えられる。たとえば、加入者とISPを結ぶ接続網上では、加入者が選択したISPまでを仮想的な専用線で接続といった、非常に限定されたサービス提供が考えられる。一方、網に対して利用者から通信品質制御が要求される場合には、利用者や使用するアプリケーションに依存したサービス、つまり、特性や帯域、遅延、端点などの違いなどに対応した細粒度でのサービス提供が必要となる。

サービスの適用領域によって検討すべき項目は大きく異なることが予想される。たとえば、優先制御に対する課金方式は、前者の例ではISPの接続料金に包含する形で回収することが可能であるが、後者の例ではサービスの自由度が高いために、課金を決定するサービスの属性やその回収方法が非常に複雑になることが考えられる。また、サービス・クラスの定義による利用者の挙動や、要求発行のインタフェースなども検討しなければならない。

#### アドミッション制御と制御パラメータの導出

利用者から受け取った通信品質制御要求は、そのままネットワークに対して反映できるわけではない。多くの場合、通信品質制御にはなんらかの資源を利用する必要があるが、あらかじめ用意される資源は有限であるため、適切な配分方法を検討しなければならない。また、必要な資源を確保できなかった場合は要求を拒否しなければならない。さらに、サービスに対する課金や特定のポリシーなど、資源以外の条件によって要求を拒否する場合も考えられる。このような、利用

者からの要求に対して応諾の可否を決定することを、アドミッション制御と呼ぶ。また、要求が許可された場合、制御パラメータ導出として、転送層に反映するための具体的なパラメータを生成する必要がある。

アドミッション制御や制御パラメータの導出をおこなうためには、さまざまな条件を設定しなければならない。このような、細分化された条件を細目条件と呼ぶ。アドミッション制御と制御パラメータの導出の際に利用する細目条件は、多くの場合、互いに密接に関係しており、明確に分離することが難しい。細目条件としては、主に次のようなものがあげられる。

#### 認証

要求を発行した利用者の有効性を検証する。また、認証にもちいた利用者の識別子は、他の判断に流用されることもある。

#### ポリシー

発行された要求が、サービス提供者のポリシーとして応諾可能であるかを判断する。たとえば、ある品質制御サービスを特定の組織に所属する利用者にものみ許可する場合、利用者の識別子と要求、および、サービス提供者があらかじめ設定したポリシー・データベースを比較し、応諾の可否を判断しなければならない。

#### 資源確保

要求を実現するためには、ネットワーク上のリンク帯域や通信経路、機器のバス帯域などを適切に確保する必要がある。これらの資源は有限であるから、必要となる資源が確保可能であることを検査しなければならない。また、転送機構上で具体的に資源を確保するためのパラメータを導出する。

#### 課金

利用者の要求に対してサービスを提供するためには、いくらかのコストが発生する。サービスを長期にわたって継続的に提供するためには、このようなコストを回収する必要があるため、多くの場合、サービス提供にともなって課金処理がおこなわれる。この場合、利用者の支払い処理が正規に完結したことを確認しなければならない。

#### 中間エンティティへの対応

特定の通信に対して制御を加える場合、その通信を識別する必要がある。しかし、現在のインターネット上には、通信を識別するための要素を、経路上で変換するようなエンティティが多数存在している。たとえば、IP アドレスやトランスポート層のポート番号を変換する NAT は、その代表的な例である。通信に対して一貫した制



御を適用するためには、制御内容の決定時にこのようなエンティティの存在を考慮し、通信の識別を矛盾なくおこなえるようにしなければならない。

#### 明示的経路の算出

網の状態によっては、通常のインターネット層による経路制御とは異なる手法によって経路を算出し、それらを優先転送の対象となる通信などに適用することで、要求を実現することがある。

#### 複数の管理主体間における折衝

前項の仕組みで決定・導出できるパラメータは、単一の管理主体からなる特定領域のネットワークの優先制御に対するものである。しかし、広域ネットワークにおける通信では、通信の端点が異なる主体によって管理されるネットワークにあることが一般的である。また、端点間を結ぶ経路が通過するネットワークも、複数の管理領域を横断するものであることが多い。提供するサービスの内容によっては、単一管理主体下のネットワークのみに対する優先制御だけで十分な場合もあるが、インターネットにおける一般的な通信に対する端点間での優先制御を行う場合には、それぞれの端点が属するネットワーク、および、それらの中間に位置するネットワークの管理主体間において、優先制御内容の折衝をおこなう機構が必要不可欠である。

上述したように、現在は単一管理主体化におけるネットワーク内での制御機構も十分に動作しておらず、要求駆動型通信品質制御サービスの提供には至っていない。一部では管理主体間での折衝も検討されているが、具体的なサービスを想定した枠組みの提案や、折衝に利用するプロトコルの策定、実装による検証などはなされていない。

#### 制御パラメータの配布

制御パラメータが導出されると、制御層はその制御内容を転送層に反映しなければならない。従来のインターネットの構造では、制御と転送が制御単位内で密接に組み合わせられているため、反映処理は特に必要とされていないか、あるいは、転送層の管理するデータ領域に必要なパラメータを追加すればよい。したがって、この処理が注目されることはなかった。

これに対し、制御層と転送層の分離モデルにおいては、制御パラメータを導出するエンティティと、そのパラメータの反映対象となるエンティティが物理的・論理的に別である。また、多くの場合、制御層が単一の機構から構成されるのに対し、転送層は複数の機構から構成されることもあり、広範囲に分散している。したがって、制御パラメータの反映には制御層から転送層への伝達が必須である。このような処理を想定したプロトコルには、COPS (Common Open Policy Service) [7] や SNMPConf などが存在する。これらは、すでに標準化されていたり、現在標準化の途上にある。

しかし、配布するための制御パラメータを導出するための機構が欠落しているため、実運用には利用されていない。

### 3.3.2. 要求駆動型制御に対する転送層

転送層では、制御層で決定されたパラメータにしたがってサービスを実現する役割を持つ。提供者によって定義される複雑なポリシーを構成する必要のある制御層に対し、転送層には配布されたパラメータの反映のみが要求されるため、IntServ (Integrated Service) や MPLS (Multi Protocol Label Switching)、DiffServ (Differentiated Services) など、さまざまな機構の標準化と設計開発が進められている。

転送層でサービスを実現するためには、主に以下の2点の要素技術が重要となる。

- パケット・スケジューリングによる優先配送
- インターネット層の機構と独立した経路制御

前者は、ネットワーク上での各転送エンティティにおいて、データ転送の際の出力インタフェイスのバッファでなんらかの制御を加え、特定の通信を優先的に転送する枠組みである。既存のルータでは、出力インタフェイスのキュー制御に最後尾廃棄 (Tail Drop) を採用している。この方式では、出力キューの許容量をこえるデータが到着したとき、キューの最後尾から廃棄処理をおこなうものである。処理が単純であるためパケット・スケジューリングにかかる負荷が小さいという利点があるが、特定の通信に対する優先制御などは考慮されていない。これに対して、近年では優先キュー (Priority Queueing; PRIQ) や RIO (Random early detection with In/Out) [8]、WFQ (Weighted Fair Queueing)、CBQ (Class-Based Queueing) [9]、H-FSC (Hierarchical Fair Service Curve) [10] など、優先転送を実現するためのパケット・スケジューリング方式が数多く提案されている。これらの方式は、各種 BSD システム上におけるパケットスケジューラを実現する枠組みとして、ALTQ [11] によって実現されている。また、近年では商用ルータでもこれらの方式の一部を実装するものが一般的となっている。

一方、後者は複数の経路にトラフィックを分散させることによって輻輳を回避し、通信品質を向上させる技術であり、トラフィック・エンジニアリングと呼ばれる。インターネット層と独立して経路を選択できる転送機構がない場合には、経路を分散させるために、ネットワークの運用管理者によって、インターネット層の経路制御機構の設定を調整したり、静的経路を設定するなどの手法によって実現されてきた。しかし、これらの手法は管理者の経験に極めて強く依存しており、広く導入することのできるものではなかった。また、既存の経路制御機構の枠組みで実現するために自由度が低いうえ、誤った設定による他のトラフィックへの影響が大きいなど、要求駆動型のサー

ビスに適用できるものではない。このような状況の中、トラフィック・エンジニアリングは MPLS の標準化によって注目されるようになった。MPLS ではインターネット層とは独立に経由する経路を決定できる。また、明示的に経由する経路を外部から指定することができるため、柔軟な制御が可能となる。

### 3.4. 本研究での取り組み

現在、いくつかの ISP によって提供されている固定型制御サービスでは、細目条件による判定を人間が介在することで処理している。そのため、柔軟な判断と制御が可能であり、細目条件の複雑な組み合わせによって特色あるサービス・クラスを定義できる。しかし、要求駆動型制御では要求に対して瞬時に応答する必要があるため、人間が介在して細目条件を処理することができない。これらの細目条件は、QoS 経路制御をはじめとして理論的な研究が進められている。しかし、細粒度な品質制御サービスを実際に提供するという視点からは、実装による実証実験がおこなわれておらず、サービスの提供を開始するために不足している要素技術や検討項目等が明らかになっていない。

このように、さまざまな場面で要求駆動型通信品質サービスに対する要求が高まりつつあるにも関わらず、そのための制御機構は運用経験の蓄積と十分な検証ががおこなわれているとは云えない。そこで、本研究ではネットワーク利用者に対してサービスの実際的な提供を目指し、3 度にわたる実証実験をおこなった。この実験では、利用者から網に対して通信品質に関する要求を直接発行する形態をとった。対象となるネットワークは 200 人以上の参加者が利用する研究会の仮設ネットワークで、参加者の要求によって駆動するさまざまな通信品質制御サービスを提供した。この実験を通じて、利用者の挙動を予測し、うまく制御するためには提供するサービス・クラス間の整合性が重要であることが明らかになった。また、NAT などのように通信の識別子を変換する中間エンティティが存在する場合、優先転送の対象となる通信を経路上で同定することができない。このため、制御対象領域内で一貫した優先転送を実現するためには、中間エンティティから変換に関する情報を取得し、制御機構でその情報を考慮した制御パラメータを生成しなければならないことが明らかになった。

このように、制御層ではまだまだ多くの課題や不明な点が残されており、個々の課題に対する研究と運用実験による経験の蓄積が重要であることが再確認された。

一方、転送層は前述したように多くの機構が標準化されており、その実装や実運用ネットワークへ適用する動きも広まりつつある。しかし、これらの転送層は固定型制御を前提として策定されたものであるため、要求駆動型制御に応用するためにはいくつかの問題がある。たとえば、要

求駆動型制御では要求の時間的粒度が極めて細かいために、制御層からの制御パラメータを反映するために要する時間が問題となる場合がある。特に、MPLS では制御を反映するために LSP と呼ばれるパスを確立する必要があり、要求許諾時の LSP の確立や、障害時における LSP の再確立など、LSP 確立に要する時間を短縮することが重要となる。本研究では、MPLS 網における LSP 確立時間の高速化手法を提案した。また、プロトタイプ実装をもちいた実験を通じてその有効性を示した。

多くの転送機構が存在する場合、それらの転送機構をサービスに応じて同時に利用することが考えられる。一般的に、これらの異なる転送機構は互いに協調せずに動作する。そのため、制御対象となるネットワーク資源、および、各転送エンティティ内における固有の資源に競合が発生する可能性がある。本研究では、要求駆動型通信品質制御サービスの実証実験において、ふたつの協調動作しない転送機構を利用したサービス提供をおこなった。ここでは、単一制御層によるネットワーク帯域資源の競合の調停、および、各転送エンティティ内のパケット・スケジューラの共有による出力キュー資源の競合回避を実現し、複数の転送機構を並行して運用できることを示した。

また、通信の優先制御を実現するためには、転送層において対象となる通信を識別する必要がある。このような機構をパラメータ・フィルタと呼び、識別する通信に関する情報の設定と、識別情報の集合から通信を識別する過程から構成される。このうち、識別情報は制御機構で生成されるため、制御対象領域内での一貫性を保証することができる。しかし、後者の識別過程は各転送エンティティで実行されるため、その方式の違いによって識別結果が異なる場合がある。この場合、制御対象領域内での通信に対する優先転送処理の一貫性が保証できなくなる。本研究では、提供者が定義するサービスに依存する部分を識別過程から分離し、パラメータ・フィルタを 2 段階に分離する手法を提案する。これにより、一貫性を喪失する原因となる個所の特定が容易になることが期待できる。

近年、通信品質制御を実現するため DiffServ や MPLS といった転送機構や、それに関連するいくつかの技術が標準化されている。本研究では、これらの技術の改善や技術を利用した実証実験をおこなっている。本章では、このような標準化された関連技術について説明する。

## 4.1. Differentiated Services

Differentiated Services (DiffServ) は、通信をいくつかのクラスと呼ばれる集合に分類し、そのクラスごとに転送時の処理をおこなうものである。それまでの通信品質制御の考え方は、IntServ に代表されるような絶対的かつ厳密なものであり、現実と著しく乖離したものであった。それに対し、DiffServ は厳密な制御を要求しない、より緩やかな通信品質制御を目指した枠組みである。

### 4.1.1. 絶対的な QoS と相対的な QoS

Differentiated Services (DiffServ) [12] は、サービス提供者である IPS や市場からの要求により、比較的容易に実現が可能である通信品質制御機構として標準化が進められた技術である。DiffServ は比較的ゆるやかな制御による相対的なサービスを提供するもので、従来のインターネットの枠組みとの親和性が高く、多くの注目を集めている。

DiffServ は、1997 年 8 月の IETF で IntServ ワーキング・グループが開催した DiffServ BoF<sup>1</sup> に端を発する。1994 年から標準化が進められてきた Integrated Service (IntServ) と、その枠組みを利用してサービスを実現するための RSVP (Resource ReSerVation Protocol) は、絶対的な QoS、すなわち、絶対的性能の保証を前提に制定されていることや、経路上のルータが通信ごとの状態を保持する必要があるなど、大規模なネットワークでは運用が困難なものであった。これに対し、DiffServ BoF では相対的な QoS の概念に基づいた議論がなされた。相対的な QoS とは、通信をいくつかのクラスに分類し、クラス間でサービス品質に差をつける考え方である。2.4 節で述べたように、このような考え方は Class of Service (CoS) と呼ばれる。

DiffServ では、大きな粒度で通信をとらえることによって、経路上のルータが通信ごとの状態をもつ必要がないように設計されている。また、当初の DiffServ ではクラス間の相対的な QoS の

<sup>1</sup>Birds of a Feather の略で、開放的なオープン・ディスカッションを表す。

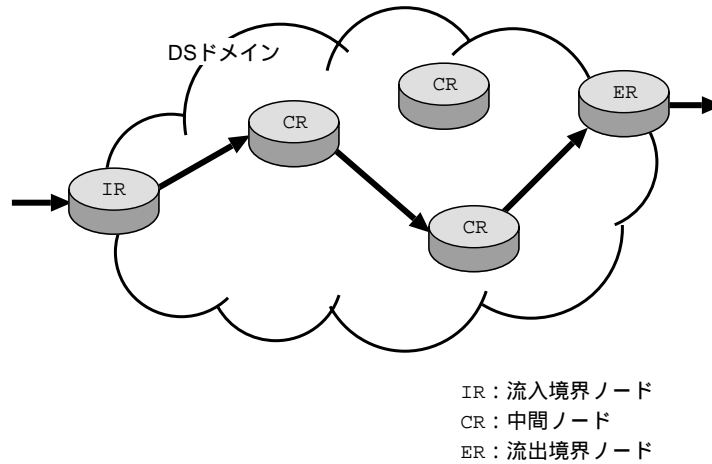


図 4.1: DS ドメイン

DS ドメインの概念図を示す。DS ドメインは、その境界上に位置する境界ノードと、それ以外の中間ノードから構成される。

みの定義であり、細かな粒度の厳密な制御をしなくとも、実用的な範囲で動作することを前提としている。このため、DiffServ 対応機器と非 DiffServ 対応機器が混在する環境においても、ある程度の動作を期待することができる。

なお、現在の DiffServ の仕様では、その枠組み内において絶対的な QoS も実現できるように設計されている。また、絶対的な QoS と相対的な QoS は共存して提供することができる。

#### 4.1.2. DS ドメイン

DiffServ では、インターネット上の特定の領域に対し、その境界上においてトラヒックをの制御を実現するための枠組みを提供する。このような領域は、同一のポリシーに基づいてサービス提供や運用がおこなわれ、領域内のノードが同一の転送規則に従うような制御対象ネットワークであり、DS ドメインと呼ばれる。DS ドメイン内のノードは、ドメインの境界上に位置する境界ノードと、ドメイン内部に位置する中間ノードに分類される。特に境界ノードは、DS ドメインに流入するトラヒックを基準にした場合には流入境界ノード、逆に DS ドメインから流出するトラヒックを基準にした場合には流出境界ノードと呼ばれる。DS ドメインの概念を図 4.1 に示す。

#### 4.1.3. DiffServ CodePoint

DiffServ における優先制御は、IP データグラム中に埋め込まれたマークのみを参照しておこなわれる。このマークを DiffServ CodePoint (DSCP) [14] と呼ばれ、通常は流入境界ノードにおいて IP ヘッダの DS フィールドに納められる。DS フィールドは、IPv4 では従来の ToS フィールド

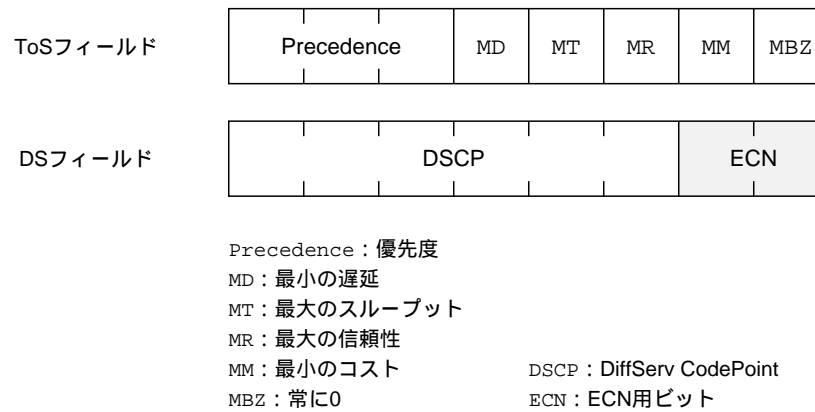


図 4.2: DS フィールド

IPv4 ToS フィールドと DS フィールドを示す。IPv4 では従来の ToS フィールドを、IPv6 ではトラフィック・クラスを、DS フィールドとして再定義している。なお、下位の 2 ビットは DiffServ が標準化された当初は未定義であった。その後、ECN (Explicit Congestion Notification) [13] 用のビットとして利用されることになった。

表 4.1: DSCP の割り当て規則

範囲番号	DSCP の空間	割り当て規則
1	xxxxx0b	標準的な PHB
2	xxxx11b	実験的あるいは局所的な利用
3	xxxx01b	実験的あるいは局所的な利用*

\*必要に応じて、標準的な PHB 用の空間として利用される可能性がある。

DSCP の割り当て規則を示す。1 番の範囲は標準的な DSCP として、後述する AF や EF などに割り当てられている。また、2 番と 3 番の範囲は、実験への利用や、サービス提供者が独自に定義することが想定されている。

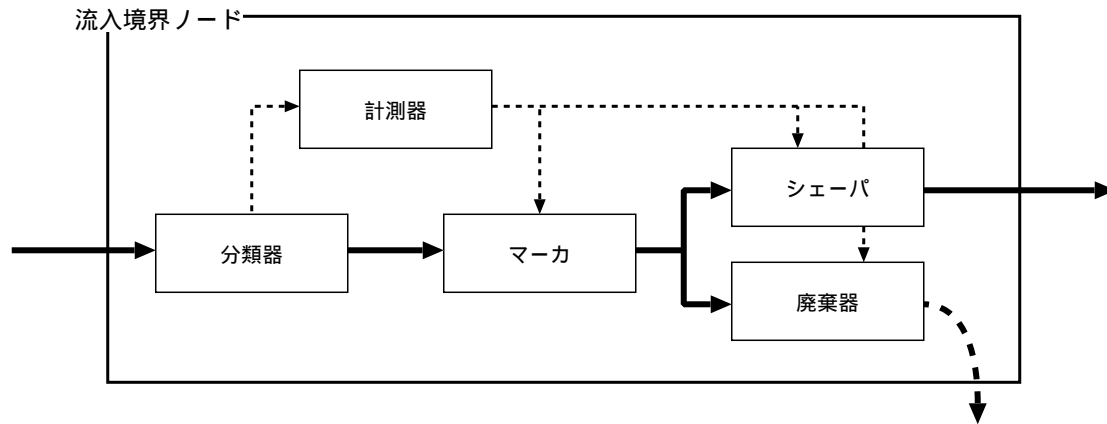


図 4.3: 境界流入ノードの構成

境界流入ノードを構成するコンポーネントを示す。流入したトラフィックは分類器によってサービスの種類が識別され、計測器の結果を加味しながらマーカによって適切な DSCP が付加される。また、必要に応じてデータの廃棄やシェーピングをおこなう。

を、IPv6 ではトラフィック・クラスを再定義した 1 オクテットの領域で、DSCP は DS フィールドの上位 6 ビットに書込まれる。従来の IPv4 ToS フィールドと DS フィールドの関係を図 4.2 に示す。

DSCP の値は優先制御の方式によって異なるものが定義され、このうちのいくつかは互換性をもつように標準の値が推奨されているが、DS ドメインの運用ポリシーによってべつの値を使うことも許可されている。また、実験用や局所的な利用のための空間も用意されており、個性あるサービスを柔軟に展開できるよう配慮されている。DSCP の割り当て規則を、表 4.1 に示す。

#### 4.1.4. 流入境界ノード

流入境界ノードでは、DS ドメインへ流入するトラフィックに対し、利用者との SLA に基づいて DSCP を付加していく。図 4.3 に境界流入ノードの構成を示す。この図に含まれるコンポーネントは、それぞれに以下のような機能をもつ。

##### 分類器 (classifier)

DS ドメインへ流入するトラフィックを、利用者と締結した SLA に基づいて対応するクラスに分類する機能をもつ。流入時にあらかじめ付加されている DSCP に基づいて分類する方法を BA (Behavior Aggregate)、IP データグラムヘッダ等を検査することによって分類する方法を MF (Multi-Field) と呼ぶ。

##### 計測器 (meter)

分類器によって分類されたトラフィックの流量を計測し、対応する契約に適合して



いるか否かを判断する。計測には、流量の平均値を計算する方式や、あらかじめ指定されたトークンバケットに対する適合を検査する方式などがある。計測器による判断は、マーキングや廃棄、シェーピングを行う際の指標になる。

#### マーカ (marker)

分類器および計測器の判断に基づき、DS フィールドに適切な DSCP を設定する。中間ノードでは、ここで設定された DSCP に基づいてデータを処理する。

#### シェーパ (shaper)

計測器の判断結果を受けて、分類器によって導出された契約に適合させるため、データの転送を遅延させる働きをもつ。この機能をシェーピングと呼ぶ。

#### 廃棄器 (dropper)

計測器の判断に基づき、データを破棄する働きをもつ。シェーパとともに、DS ドメインに流入するトラフィックを契約に適合させる役割がある。

### 4.1.5. 中間ノードと Per-Hop Behavior

流入境界ノードが前節で説明したような非常に複雑な処理をおこなうのに対し、中間ノードの処理は次に上げるような非常に簡素なものである。まず、DS ドメイン内の全ての中間ノードに、DSCP と具体的な転送処理の対応表が用意される。中間ノードに IP データグラムが到着すると、その IP ヘッダ中の DSCP から対応する転送処理が導出され、そのまま適応される。DiffServ では、DS ドメイン内での処理をこのように単純化することによって、IntServ の問題点であった対規模性を解決している。

中間ノードでおこなわれる具体的な処理を Per-Hop Behavior (PHB) と呼ぶ。PHB は処理方式や実装、およびそのパラメータなどを表す概念ではなく、外部から観測できるデータ転送の挙動を指すものである。具体的な PHB の実装には、スケジューリングやバッファ管理をとともなうことが多い。

現在、標準的な PHB として Default と Class Selector [14]、Assured Forwarding (AF) [15]、および、Expedited Forwarding (EF) [16] が定義されている。それぞれの PHB のために推奨されている DSCP は、表 4.1 における 1 番の範囲となり、したがって最下位ビットが 0 である。

#### Default PHB

この PHB は、最善努力型の転送を表す。対応する DSCP は 000000b である。流入境界ノードでは、分類器での分類処理の前に DSCP を消去することが一般的である。また、一致する契約が規則がない場合は、あたらしい DSCP がマークされない。

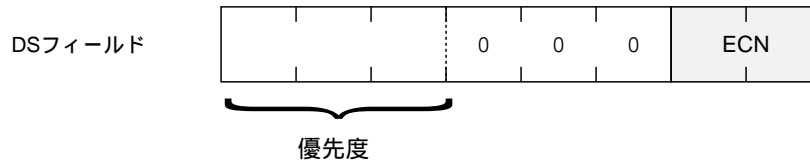


図 4.4: Class Selector PHB の DSCP の構造

Class Selector PHB における DSCP の構造を示す。上位 3 ビットが ToS フィールドの優先度に対応することで、従来の ToS フィールドを利用したアプリケーションに対する下位互換性を確保している。

したがって、このようなトラヒックは中間ノードにおいて Default PHB として扱われる。

### Class Selector PHB

この PHB は、DSCP の上位 3 ビットを利用して優先度を表すものである (図 4.4)。上位 3 ビットは、図 4.2 で示した ToS フィールドの優先度を表す領域と一致していることに注目する必要がある。これにより、従来の ToS フィールドを利用したアプリケーションに対し、一定の下位互換性を確保している。なお、下位 3 ビットは常に 0 となる。

### Assured Forwarding PHB グループ

この PHB は、相対的な品質をもつ複数の PHB からなるグループを定義し、計測器の結果によって PHB グループのなかから適切な PHB を使い分けることで、最低帯域を保証した最善努力型の転送を提供するものである。したがって、最低帯域を要求しながらも、実際に利用できる帯域に柔軟に適應できるアプリケーションなどに適している。

AF PHB グループを利用する場合、流入境界ノードでは、SLA で定められた帯域を超過したトラヒックに印をつける。一方、中間ノードでは、輻輳が発生した場合に超過の印がついた IP データグラムから優先的に廃棄する。このとき、契約に適合したトラヒック、すなわち、超過の印のついていないトラヒックは廃棄しない。これによって、最低帯域を保証している。

標準化されたものは、4 つの独立した PHB グループを定義しており、各 PHB グループには廃棄優先度の異なる 3 段階の PHB が含まれている。このような定義から、AF PHB グループに対応する DSCP は図 4.5 に示すような、比較的複雑な構造を持つ。この図のように、上位 3 ビットがクラスを、続く 2 ビットが廃棄優先度を

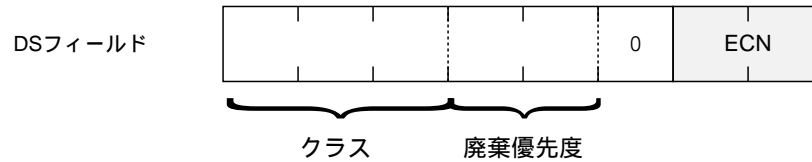


図 4.5: Assured Forwarding PHB グループの DSCP の構造

Assured Forwarding PHB グループにおける DSCP の構造を示す。上位 3 ビットがクラスを、続く 2 ビットが廃棄優先度を表している。

表 4.2: Assured Forwarding PHB グループの DSCP

クラス	廃棄優先度		
	低	中	高
クラス 1	001010b (AF11)	001100b (AF12)	001110b (AF13)
クラス 2	010010b (AF21)	010100b (AF22)	010110b (AF23)
クラス 3	011010b (AF31)	011100b (AF32)	011110b (AF33)
クラス 4	100010b (AF41)	100100b (AF42)	100110b (AF43)

AF PHB グループに推奨される DSCP の値を示す。4 つのクラスと 3 段階の廃棄優先度により、合計で 12 個の DSCP が定義されている。4 つのクラス間には相対的な関係は定義されていないことに注意が必要である。

表す。

なお、便宜上、クラスが  $i$  で廃棄優先度が  $j$  であるような PHB に対応する DSCP を、 $AF_{ij}$  と表記する。ここで、廃棄優先度は大きいほど優先的に廃棄されることを表す。この表記法によって表した、AF PHP グループに対して推奨される DSCP の値を、表 4.2 に示す。

### Expedited Forwarding PHB

この PHB は、通信主体間での帯域保証から発展した、DS ドメイン内における絶対帯域を保証するものである。対応する DSCP としては、101110b が推奨されている。

流入境界ノードでは契約を越えるトラフィックは全て廃棄する。その一方、DS ドメイン内では全契約で必要な帯域の総和以上の帯域を確保しておく。これにより、中間ノードでのデータの廃棄が発生せず、契約帯域を保証することができる。また、中間ノードにおける転送制御の実装としては、一般的に Priority Queue を利用する。この場合、他の PHB に割り当てられたトラフィックを圧迫しないために、無制限な優先制御を抑制するための機構を導入しなければならない。

## 4.2. Multi Protocol Label Switching

Multi Protocol Label Switching (MPLS) は、sub-IP 層での転送機構として注目されている技術である。標準化が開始された当初はラベル・スイッチングによる転送の高速化が期待されたが、現在ではインターネット層から独立した転送層としての機能が重要視されている。

### 4.2.1. ラベル・スイッチングと MPLS

Multi Protocol Label Switching (MPLS) [17] は、インターネット層の下部に介在する転送機構として注目されている技術である。また、階層構造的にはデータリンク層の上位にあたるため、一般に sub-IP と呼ばれる層に位置づけられる。sub-IP 層による独立した転送機構の確立は、インターネット層からの転送機構の分離を考えた場合、既存のインターネット層の枠組みを大きく改変する必要がないことから、非常に効果的な手法であるといえる。

MPLS は、1990 年代半ばから特にさかんに研究されるようになった、ラベル・スイッチングと呼ばれる技術から発展した [18]。ラベル・スイッチング技術としては、シスコ・システムズ社の Tag Switching や、東芝の Cell Switching Router (CSR) 等が開発された。これとほぼ同時期に開発されたイブシロン・ネットワークス社の IP Switching は、MPLS の原型ともいえる技術である。

ラベル・スイッチング技術は、“ラベル”という概念を利用してデータを転送する技術の総称である。ラベルとは、ある程度短い固定長のビット列で表されるデータの識別子である。ルータやスイッチは、転送をおこなうべきデータから、ラベルのみを転送のための情報として利用し、出力インタフェースや転送先を決定する。また、一般的なラベル・スイッチング技術では、転送の際にデータ内のラベルをあたらしい値に書き換える場合がある。

ラベル・スイッチングの概念をとりいれて開発された技術の多くは、ラベルとして IP アドレスよりも短い長さのものを採用し、ラベルの検索とスワッピング処理を軽減することで、ルータやスイッチの処理負荷の軽減と高速な転送を最大の目的としていた。しかし、近年の転送処理機構のハードウェア化や並列化によって転送処理効率は向上し、ラベル・スイッチングの当初の目的と利点は薄れつつある。現在、MPLS が注目される大きな目的とその利点は、冒頭で述べたように、インターネット層からの転送機構の独立による、既存のネットワーク制御機構からの分離である。このため、より柔軟で拡張性の高いネットワーク制御を必要とする場合、MPLS の導入は非常に重要な選択肢のひとつとなっている。

#### 4.2.2. 転送等価クラス

MPLS では、対象となるデータがある粒度で分類して転送する。この分類を“転送等価クラス (Forwarding Equivalent Class; FEC)”と呼ぶ。それぞれの FEC には、転送先 LSR (Label Switching Router) やラベル操作などの処理が対応づけられ、NHLFE (Next Hop Label Forwarding Entry) と呼ばれる経路情報として MPLS の経路表に格納される。データ転送時には、MPLS の経路表から該当する NHLFE を検索し、転送処理をおこなう。したがって、同一の FEC に分類されたデータは、すべて同一に処理・転送される。

データを FEC に分類する際には、データの受信者アドレスや転送時の制御など、さまざまな粒度の項目をもちいることができる。このため、FEC の定義によって特定の通信の品質を制御することもできる。

#### 4.2.3. ラベル

MPLS によって転送されるデータには、MPLS ヘッダが付加される。図 4.6 に、MPLS ヘッダが挿入された IP データグラムの例を示す。このように、MPLS ヘッダは L2 ヘッダと IP ヘッダの間に挿入される<sup>2</sup>。MPLS ヘッダは 32bit で、20bit のラベル、3bit の実験用フィールド (Exp)、ラベル・スタック (後述) の最下層を示す 1bit のフラグ、そして MPLS 網内での生存時間を示す 8bit の TTL (Time To Live) で構成される [19]。通常、MPLS ヘッダはデータに対してひとつである

<sup>2</sup>データリンクによっては、この図と異なる場合がある。

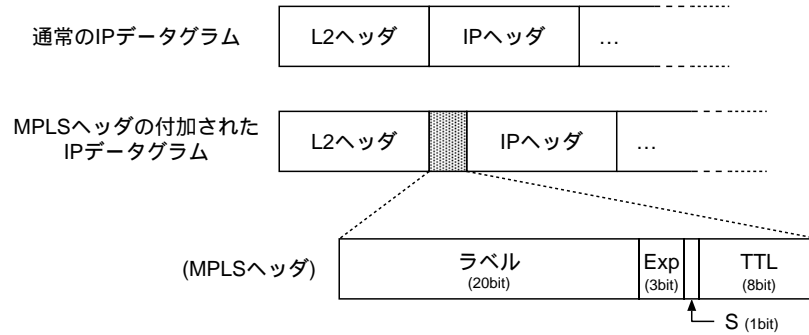


図 4.6: MPLS ヘッダとラベル

IP データグラムに MPLS ヘッダが挿入されたようすを示す。MPLS ヘッダの先頭 20bit には、FEC の識別子となるラベルが配置される。この例は、ひとつ MPLS ヘッダが挿入されたものであるが、一般的には複数の MPLS ヘッダがスタック構造をもちながら挿入され、ラベル・スタックを構成する。

が、MPLS 網上に積み重なる形で別の MPLS 網を構築する場合などには、複数の MPLS ヘッダが付加されることがある。これらヘッダはスタック構造になっていることから、“ラベル・スタック”と呼ぶ。データ転送時のラベル操作は、ラベル・スタックに対しておこなわれる。

ラベルは FEC の識別子であり、転送の対象となるデータに埋め込まれたラベルは、そのデータが分類されるべき FEC を示している。そのため、データを転送しあう隣接 LSR との間で、ラベルと FEC の対応は一致していなければならない。

#### 4.2.4. データ転送とラベル操作

インターネット層の転送には、受信者アドレスのプレフィックスとプレフィックス長、および、このふたつの組に対応する出力インタフェースからなる経路表を利用する。転送対象となる IP データグラムが流入すると、その受信者アドレスに対応する項目を経路表から検索し、その項目で指定された出力インタフェースを介して転送する。この検索は、IP データグラムの受信者アドレスに対して、経路表に保持された受信者アドレスのプレフィックスとプレフィックス長の最長一致検索である。インターネット層における IP データグラムの転送のようすを、図 4.7 に示す。

一方、MPLS での転送ではラベルと NHLFE の対応を表す MPLS 経路表を利用する<sup>3</sup>。転送対象となるデータが流入すると、データに含まれるラベル・スタックの最上位のラベルに対し、MPLS 経路表から対応する NHLFE が検索される。4.2.2 節で述べたように、NHLFE は FEC と転送先

<sup>3</sup>正確に記述すると、MPLS には ILM (Incoming Label Map) と FTN (FEC-to-NHLFE) の 2 種類の経路表が存在する。前者は流入データにラベルが付与されていた場合の経路表であり、後者は流入データがラベルを持たない場合の経路表である。いずれの経路表も、転送時に適用する NHLFE を検索するために利用する。

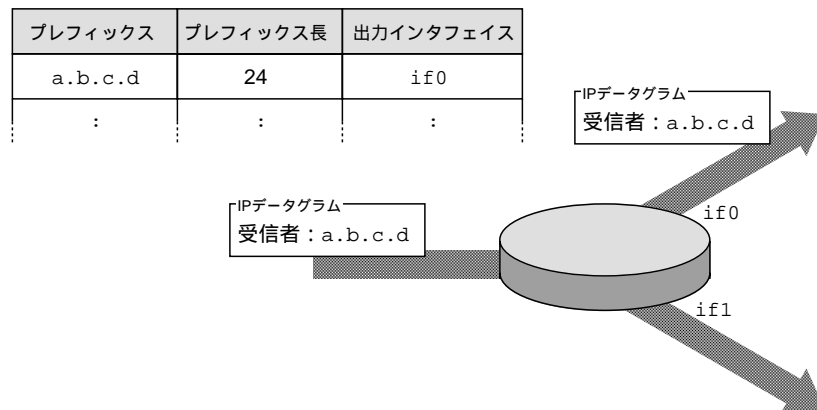


図 4.7: インターネット層での転送処理のようす

インターネット層での転送処理のようすを示す。受信者アドレスが a.b.c.d であるような IP データグラムがルータに到着した。これに対し、プレフィックス/プレフィックス長が a.b.c.d/24 である経路が採用され、インタフェース if0 に出力されている。

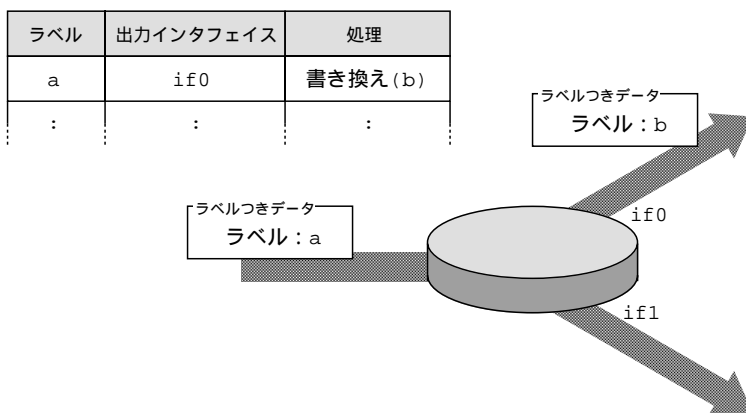


図 4.8: MPLS での転送処理のようす

MPLS での転送処理のようすを示す。ラベル a を持つデータが LSR に到着した。ILM と NHLFE を検索した結果、ラベルを b に書き換えてインタフェース if0 から出力した。

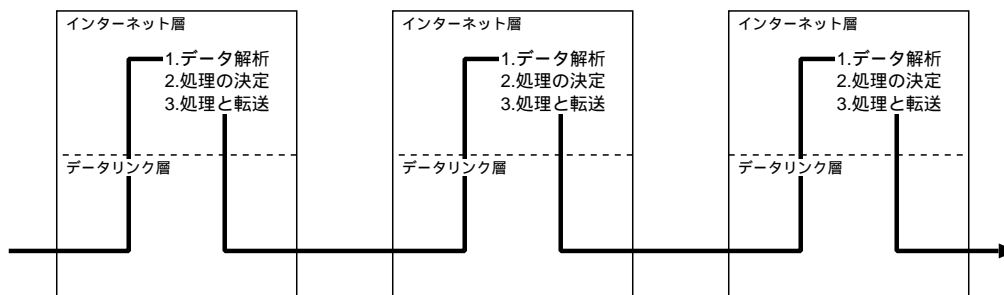


図 4.9: 経路上で見たインターネット層での転送処理のようす

通信経路上の一部分で見た、インターネット層での転送処理のようすを示す。データの扱いや転送先を決定するために、全てのホップでデータ解析をおこなわなければならない。

LSR やラベル操作等の対応関係を表す。データは NHLFE で指定された処理にしたがって転送される。

LSR においてデータが転送されるとき、対応する NHLFE の指示に従って、ラベル・スタックに操作が加えられる。ラベル操作は以下の 3 種類のいずれかとなる。

#### 積み上げ (Push)

ラベル・スタックの最上位に、あたらしいラベルを積み上げる

#### 書き換え (Swap)

ラベル・スタックの最上位に位置するラベルを、あたらしいラベルと置き換える

#### 取り除き (Pop)

ラベル・スタックの最上位に位置するラベルを取り除く

通常の MPLS 転送では、MPLS 網の内部ではラベルの書き換えだけがおこなわれる。しかし、4.2.3 節で述べたような、複数の MPLS 網が積み重なる形で構築される場合には、MPLS 網内でラベル・スタックに対する積み上げや取り除きの操作が加えられる。また、MPLS 網と非 MPLS 網の境界上に位置する LSR では、非 MPLS 網から MPLS 網へ流入してくるデータにはラベルが積み上げられ、MPLS 網から非 MPLS 網へ流出するデータからはラベルが取り除かれる。MPLS による転送のようすを、図 4.8 に示す。



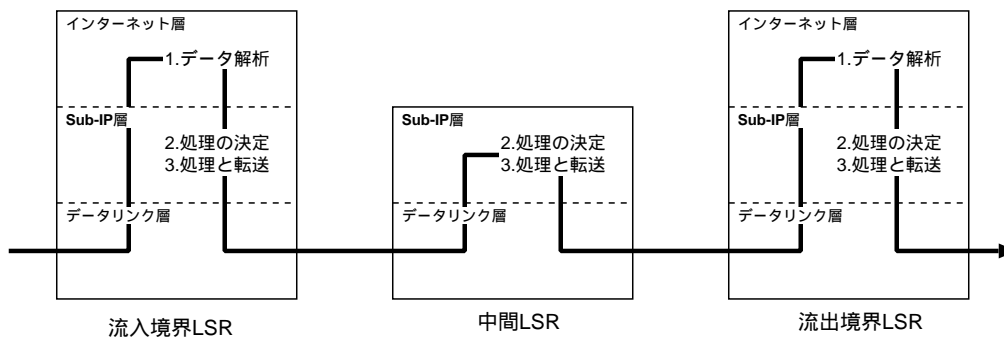


図 4.10: LSP でみた MPLS での転送処理のようす

LSP 単位で見た、MPLS での転送処理のようすを示す。流入境界ノードでデータ解析をおこなったあとは、ラベルのみで処理と転送先を決定する。

#### 4.2.5. ラベルスイッチ・パス

4.2.4 節で述べたように、インターネット層における転送では、ホップ毎に以下の処理が必要となる (図 4.9)。

1. データの解析と転送処理の決定に必要な情報の取得
2. 転送先と転送時の処理の決定
3. 実際の処理と転送

これに対して、MPLS 転送では MPLS 網の境界上の LSR でのみが上記の全ての処理を実行する (図 4.10)。境界 LSR は MPLS 網へ流入するデータを分類し、データの属すべき FEC を決定する。同時に、データにラベルを挿入して次 LSR に転送する。網内の LSR は前述のようにラベル値から FEC を導出し、その内容に基づいた転送をおこなう。このような MPLS による転送を MPLS 網全体で俯瞰した場合、データが MPLS 網に流入してから非 MPLS 網へ流出するまでの道筋は、データが対応づけられた FEC に対応する、ひとつ以上の LSR からなる連続した論理パスを構成する。このパスを“ラベルスイッチ・パス (Label Switched Path; LSP)”と呼ぶ。また、LSP を構成する LSR のうち、トラヒックが流入する境界 LSR を流入境界 LSR、トラヒックが流出する境界 LSR を流出 LSR、それ以外の LSR を中間 LSR と呼ぶ。LSP を確立して MPLS 網内で MPLS 転送による通信を有効にするためには、なんらかの仕組みによって FEC とラベルの対応を MPLS 網内に浸透させなければならない。

MPLS では FEC という単位で通信の扱いを決定する。したがって、非 MPLS 網におけるデータ単位での通信品質制御と FEC との対応づけを定義することで、非 MPLS 網と MPLS 網の境界においても、透過的で連続的な通信品質制御を提供することができる。さらに、上述のように LSP

は FEC に対応した論理パスであるから、LSP の制御によって通信を制御することができる。これは、MPLS では網の境界および内部の LSR の集合に対する包括的な制御によって、網全体に通信品質制御を適用することができること意味する。

このような特徴をうまく利用したサービスの例として、仮想専用線 (Virtual Private Network; VPN) やトラフィック・エンジニアリングをあげることができる。前者は、従来のインターネット上に一定の帯域を確保した LSP を確立することによって、複数の拠点間を仮想的な専用通信路で接続するものである。一方、後者は通信に利用される経路を明示的に指定するものである。このことにより、異なる品質特性をもつ経路を使い分けることで通信品質に差をもたせたり、ある経路に集中している通信の一部を他の経路に移すことにより、網内の帯域利用効率をあげるといったことが可能になる。

#### 4.2.6. ラベル配布

前節で述べたように、MPLS 転送による通信をおこなうためには、FEC とラベルの対応を MPLS 網内に浸透させなければならない。この手順を“ラベル配布”と呼ぶ。

現在、ラベル配布方式には多くの方式が提案されている。この中には、既存の制御プロトコルにラベル情報を付加して配布する方法と、ラベル配布を目的としたプロトコルを利用する方法がある。前者は、別のプロトコルを利用しないためにシステムが単純になること、インターネット層の経路情報をラベル配布に利用する場合に親和性が高いことなどから、利用が可能であれば望ましい方式とされている。ラベル配布に利用できる既存の制御プロトコルとしては、BGP (Border Gateway Protocol 4) [20, 21]、PIM (Protocol Independent Multicast) [22, 23]、RSVP (Resource ReSerVation Protocol) [24, 25] などがある。

一方、代表的なドメイン内経路制御プロトコル (Interior Gateway Protocol; IGP) である OSPF (Open Shortest Path First 2) [26] は、リンク状態型経路制御プロトコルである。このため、明示的な次ホップの値などが経路情報に含まれておらず、ラベル配布に不向きである。したがって、インターネット上の IGP を利用する領域では、多くの場合、ラベル配布に利用できる既存の制御プロトコルが存在しない。このような領域に対してもラベル配布を可能にするため、ラベル配布を目的としてプロトコルとして、ラベル配布プロトコル (Label Distribution Protocol; LDP) [27]、および、制約経路ラベル配布プロトコル (Constraint-based Routed Label Distribution Protocol; CR-LDP) [28] が標準化された。LDP はインターネット層における情報交換から導出された経路に一致するような LSP を確立するプロトコルであり、CR-LDP はインターネット層の経路から独立して指定された、明示的な経路によって LSP を確立するものである。

また、上記で述べた方式のほかに、経路上の LSR に対して手動で静的にラベルを配布する方法

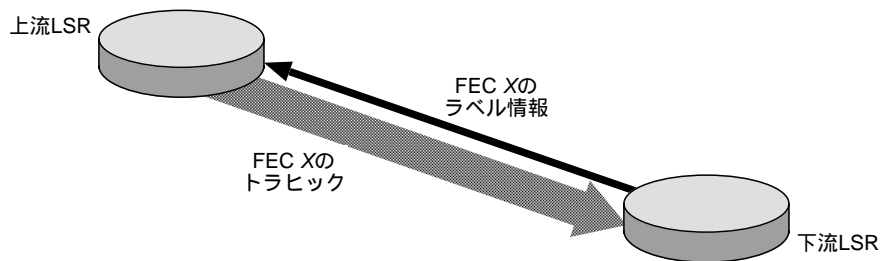


図 4.11: LDP におけるラベルの配布方向

LDP におけるラベル配布のようすを示す。ある FEC  $X$  に対応するトラヒックが流れる方向に対して、下流側の LSR がラベルの生成と配布をおこなう。

もある。

### ラベル配布プロトコル (LDP)

前節で述べたように、ラベル配布プロトコル (LDP) は、インターネット層で計算された経路情報を MPLS 網状に写像するためのラベル配布方式である。すなわち、インターネット層での経路制御の結果を元に、経路表に格納されたそれぞれの受信者アドレスを FEC とした LSP を確立する。

LDP では、ラベル配布を始める前に近隣の LSR と隣接関係を確立する。隣接関係の確立には、まず、近隣 LSR 発見機構をもちいて隣接する LSR を発見する。この機構には、リンクローカル・マルチキャストを利用している。次に、発見した隣接 LSR と TCP によって接続し、隣接関係を確立するために各種の折衝をおこなう。折衝が終了すると隣接関係が確立され、ラベルが配布される。図 4.11 に示すように、ラベルの生成と配布はトラヒックの流れに対して逆方向、すなわち、下流に位置する LSR から上流の LSR に対しておこなわれる。

LDP におけるラベル配布を LSR 間の近隣関係という局所的な観点で見た場合、その配布方式には要請駆動と自発のふたつのモードがある。

#### 要請駆動モード (Downstream-on-Demand; DoD)

このモードでは、上流の LSR は下流の LSR に対して明示的にラベル情報を要請する。

#### 自発モード (Downstream Unsolicited; DU)

このモードでは、下流の LSR があたらしい FEC に対してあらたなラベルを割り当てた場合、上流からの要請の有無に関わらず自発的に広告する。ただし、上流から明示的な要請があった場合には、対応するラベル情報を広告する。

これらのモードのいずれを利用するかは、隣接関係を確立する際の折衝において決定される。

また、インターネット層の経路表から FEC を導出し、ラベルの生成・配布による LSP 確立に至るまでの過程を、MPLS 網全体という帯域的な観点から見た場合には、その制御方式には順序ラベル配布制御と独立ラベル配布制御の 2 種類がある。

#### 独立ラベル配布制御

独立ラベル配布制御では、それぞれの LSR が独立して経路表からの FEC 抽出とラベルの生成をおこなう。ラベルの生成に関して前後の LSR との動作に束縛されることがないため、隣接 LSR からラベル情報を要請された場合にも、即座に返答することができる。

また、隣接 LSR との間が DU モードで動作している場合、インターネット層での経路表に対応するラベルが即座に配布される。したがって、経路制御が収束した直後に LSP を確立することが可能である。このように、DoD モードで独立ラベル配布制御がおこなわれた場合、インターネット層での経路制御に近い挙動となる。

#### 順序ラベル配布制御

順序ラベル配布制御では、LSR は以下のいずれかの条件を満たした場合にのみラベルを生成する。

- その LSR が対象となる FEC の流出境界 LSR である場合
- 対象となる FEC の次ホップへのラベルを保持している場合

すなわち、順序ラベル配布制御に基づくラベル配布では、流出境界 LSR のみがラベル配布を開始することができる。また、中間 LSR は次ホップとなる LSR からラベル情報を受け取っている場合にのみ、上流に対してラベル情報を広告することができる。順序ラベル配布制御によるラベル配布のようすを、図 4.12 に示す。

順序ラベル配布制御によって確立された LSP は、流入境界 LSR から流出境界 LSR までとぎれることなく連続であることが保証される。これに対し、独立ラベル配布制御を利用する場合には、いくつかの LSR が LSP 確立前に転送処理を開始する場合がある。このため、LSP 上の LSR において通信に品質制御などの特殊な操作を加えたい場合には、通過する LSR を保証するために順序ラベル配布制御を利用する必要がある。

以上で述べた LDP の動作をもとに、ラベル配布の過程を例示する。

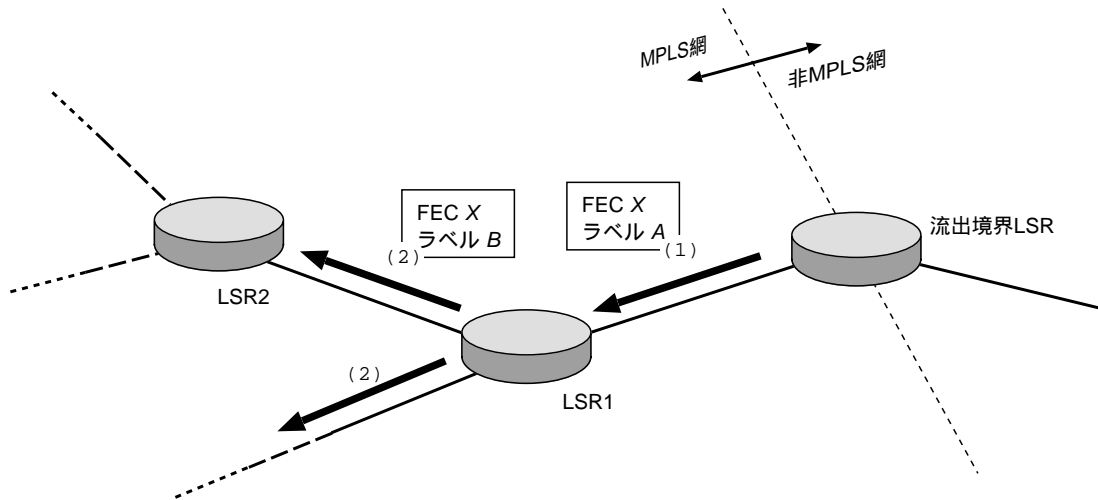


図 4.12: LDP の順序ラベル配布制御

順序ラベル配布制御に基づく LDP のラベル配布のようすを示す。(1) 境界 LSR が FEC X に対応するラベルを生成し LSR1 に配布する。(2) LSR1 は FEC X に対するラベルを受け取ったので、対応するラベルを生成して配布している。LSR1 と LSR2 の間の動作モードによっては、自発的な配布をしない場合もある。

#### 1. 隣接 LSR の発見

リンクローカル・マルチキャストを利用して、隣接 LSR を発見する。

#### 2. 隣接関係の確立

発見した隣接 LSR との間に TCP を利用した接続を確立し、隣接関係を確立するための折衝をおこなう。折衝内容としては、上記で述べたような DoD/DU などがあげられる。折衝が終了すると隣接関係が確立される。この際、TCP 接続はそのまま維持され、今後の通信に引き続き利用される。

#### 3. ラベルの生成と配布

独立ラベル配布制御の場合には、インターネット層の経路表などに基づいてラベルを生成する。また、下流から順序ラベル配布制御によってラベル情報が広告された場合には、対応するラベルを生成する。

いずれの場合にも、上流の LSR との間が DU モードで動作している場合は、生成したラベルを配布する。

#### 4. ラベルの要請と返答

下流の LSR に対し、必要に応じてラベル情報を要請する。また、上流の LSR からラベル情報を要請された場合には、該当する情報を広告する。

LDP で交換される情報の多くは、TLV エンコーディングと呼ばれる形式で表現される。この形

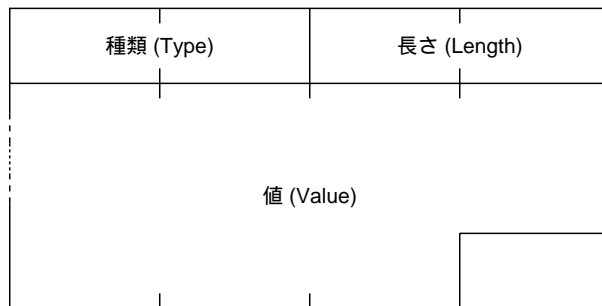


図 4.13: TLV エンコーディング

LDP で利用される TLV エンコーディングの形式を示す。種類と長さは固定長、値は長さに応じた可変長である。なお、TLV の終端が 4 オクテット境界にない場合も、パディングはおこなわれない。

式は、情報の種類 (Type)、情報の長さ (Length)、および、情報の内容を表す値 (Value) から構成される。TLV エンコーディングを図 4.13 に示す。この図のように、種類と長さは 2 オクテットの固定長であり、これによって認識できない種類の情報は容易に無視できる。また、あらたな種類を定義することで LDP の拡張も容易におこなえるように配慮されている。次節で説明する CR-LDP は、この柔軟性を活かした LDP の拡張である。

#### 制約経路ラベル配布プロトコル (CR-LDP)

MPLS では、FEC と LSP が流入境界 LSR で対応づけられ、MPLS 網内ではラベル値のみによって転送先が決定される。これは、流入境界 LSR が MPLS 網内におけるデータの転送経路を決定していると言える。したがって、MPLS 網内に任意の制約に基づく LSP を確立することができれば、インターネット層の経路制御と独立した経路選択が可能になる。本章の冒頭でも述べたように、このようなコスト以外の制約に基づく経路選択パラダイムをサポートすることは、MPLS の中核ともなる非常に重要な機能としてとらえられている。

前節で述べた LDP は、インターネット層における経路制御を MPLS 網上に写像するものである。したがって、LSP の経路はインターネット層でのコストに基づく制御に束縛され、自由な制約による経路選択を実現することはできない。制約経路ラベル配布プロトコル (Constraint-based Routed-Label Distribution Protocol; CR-LDP) は、外部から与えられた明示的な経路に基づく経路選択を実現するために、LDP を拡張したものである。

経路を明示的に指定するため、明示的経路 TLV (Explicit Route TLV; ER-TLV) と呼ばれるありがたい種類の TLV が定義されている。ER-TLV には、経路する LSR の順序が明示的に指定される。また、ER-TLV で指定した経路上の資源を確保することもでき、そのために数種類の TLV が

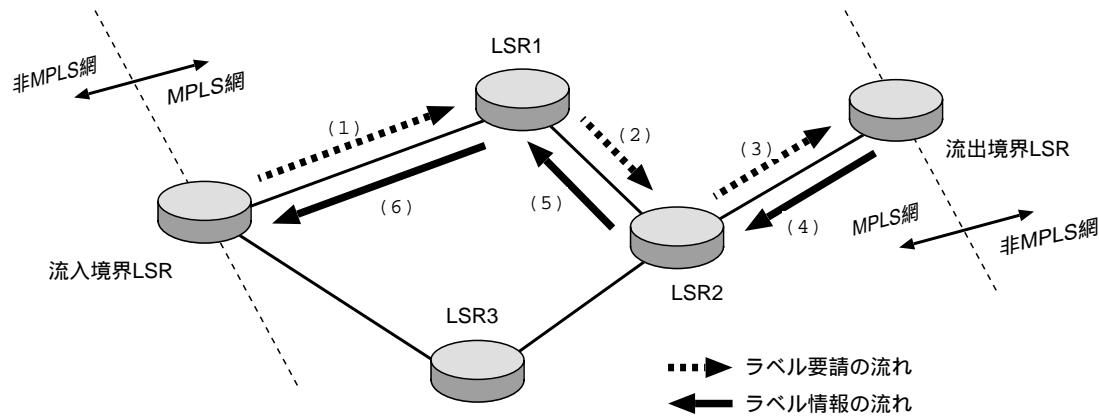


図 4.14: CR-LDP による LSP 確立のようす

CR-LDP により明示的に指定された経路に沿って LSP が確立するようすを示す。処理は流入境界 LSR から開始され、各 LSR は DoD モードで下流 LSR にラベル情報を要請する。これによってラベルは流出境界 LSR から配布され始めるため、順序ラベル配布制御によって LSP が確立されることになる。

あたらしく定義された。前節でも述べたように、明示的な経路指定によって LSP が経路する LSR を保証するために、CR-LDP では順序ラベル配布制御を利用する。また、LSP の確立は流入境界 LSR から開始するため、経路上のすべての LSR 隣接関係において、DoD モードでラベル配布がおこなわれる。

図 4.14 に、CR-LDP を利用した LSP 確立の過程を示す。この図で、明示的経路として

流入境界 LSR    LSR1    LSR2    流出境界 LSR

が指定されていることを想定している。まず、流入境界 LSR は明示的経路として (LSR1, LSR2, 流出境界 LSR) を含む ER-TLV を生成し、(1) LSR1 に対して ER-TLV とともにラベル要請を送信する。LSR1 では、受信した ER-TLV から自身を取り除き、(2) 再び LSR2 に対して ER-TLV とともにラベル要請を送信する。この処理は LSR2 でも同様におこなわれる (3)。流出境界 LSR では、受信したラベル要請に含まれる ER-TLV を検査することで、自身が明示的経路の終点であることを知る。これを受け、(4) 流出境界 LSR は順序ラベル配布制御にしたがってラベル情報を返答する。ラベル情報を受信した LSR2 は ILM および NHLFE に情報を反映し、(5) さきほど受信したラベル要請に対するラベルを生成して配布する。この処理は LSR1 でも同様におこなわれる (6)。最終的に流入境界 LSR がラベル情報を受信し、LSP が確立する。

以上のように、CR-LDP では流入境界 LSR によって LSP の確立が開始されたあと、ラベル要請が明示的経路に沿って流出境界 LSR に達し、そこから順序ラベル配布制御にしたがってラベルが配布される。すなわち、流入境界 LSR と流出境界 LSR の間を、ラベル要請とラベル情報が往

復して LSR の確立が完了する。



要求駆動型の細粒度な契約に基づく通信品質制御は、既存のインターネットの構造では扱いつらく、サービスの実提供には至っていない。今後、要求駆動型通信品質制御サービスを利用者に提供していくためには、多くの研究や運用経験、あるいは、利用者の挙動などを含めた市場調査が必要である。本章では、仮設ネットワークという実運用環境下における、要求駆動型通信品質制御サービスの提供を目指した実証実験について述べる。

## 5.1. 実験の概要

要求駆動型の細粒度な契約に基づく通信品質制御は、人間の介する固定型制御の場合と異なり、既存のインターネットの問題点や実現に際する不明な点が多い。特に、制御と転送を分離した際の問題点や、制御機構に必要なコンポーネントおよび要素技術群、サービス・クラスと利用者の挙動など、商用の提供に至るまでには多くの研究と運用経験が必要である。

試験的運用とサービスの提供は、サービスの性質上、ある程度の基礎的な試験を除けば、実験の対象は実利用者を収容するネットワークであることが望ましい。また、広域ネットワークの転送機構を構成する多くの機器に、試験的な要素を含む機能を追加しなければならない。したがって、試験運用は多くの利用者に大きな影響を与える可能性があり、商用ネットワークで実現することは困難である。そこで、実利用者を収容する実験的仮設ネットワークである、WIDE プロジェクト研究会の camp-net において、要求駆動型通信品質制御サービスの提供実験をおこなった。

WIDE プロジェクト [29] は、インターネットおよびその関連技術に関する研究と開発を行う、産学共同の研究コンソーシアムである。WIDE プロジェクトでは、隔月の研究会に加えて、毎年2回(春: 3月/秋: 9月)、各4日間におよぶ合宿形式の研究会を開催している。この研究会では200~300名のネットワーク研究者が一堂に会し、インターネットに関するさまざまな研究課題に関する集中的な議論が行われる。

この研究会の会場では、短期間ではあるが camp-net と呼ばれる仮設ネットワークが構築される。camp-net にはおおまかに2つの重要な役割がある。ひとつは、実運用ネットワークとしての役割である。近年、参加者の多くは各自のラップトップ型コンピュータを持参するが、これらの参加者は camp-net を通じてインターネットへの接続性を確保する。もうひとつは、ネットワーク

に関するさまざまな先進的技術の実証実験を行う、テストベッドとしての役割である。

camp-net の構成は、参加者の規模や会場の物理的構成、導入される実証実験の要求等、多くの制約条件に応じて毎回異なった形で設計される。特に、外部接続性を提供するリンクは、camp-net のサービスや構成に極めて強い影響を持つが、開催地の回線事情等に著しく依存するため、全体の構造を大きく左右する一因となっている。

camp-net のネットワークが仮設であるという性質は、ネットワーク資源に対する制約に大きく現れる。特に、対外リンクは臨時線による敷設であるため、一般的に予想される利用帯域より狭帯域であり、RTT も大きくなる傾向にある。そのため、camp-net を利用したときの通信品質の向上は、研究会の参加者の潜在的な要望となっていた。このような状態のネットワークを利用した際に、ネットワーク利用者に対する通信品質の向上は、現在のネットワークにおける課題のひとつである。

実証実験は、2000 年春/2000 年秋/2001 年春に開催された WIDE プロジェクト研究会の camp-net において、

- 利用者からの要求によって駆動する、通信品質制御システムの検討
- サービスの提供や通信品質の変動にともなう利用者の挙動の観察
- 提供するサービス・クラスとポリシーの関係

などを検証するためにおこなわれた。

### 5.1.1. 個別実験の概要

上述のように、3 回にわたる実証実験は camp-net の制約などから、全て異なる構成でおこなわれた。また、提供したサービス・クラスなども全て異なるものであった。以下に、それぞれの研究会における実験の概要について述べる。

#### 輻輳するリンク上での優先制御サービス (2000 年春)

2000 年春の WIDE プロジェクト研究会の camp-net における実験では、輻輳が発生する単一リンク上において優先転送サービスを提供した。輻輳は帯域を占有するトラフィックを人工的に生成することによって発生させた。制御層としては仮想通貨による課金や COPS による制御パラメータ配布などを、転送層としては DiffServ を利用してサービスを実現した。ここでは、輻輳の発生とともに利用者が優先転送を求める傾向が観測され、サービスを理解するとともに優先転送に価値を見いだす研究会参加者の姿が見られた。

なお、この研究会の camp-net における実証実験は、実利用者を収容するネットワークにおいて要求駆動型通信品質制御サービスを最初に実現したものであり、サービス提供に必要な要素機能の洗い出しや利用者の挙動の理解という点で、非常に大きな意義を持つものである。

#### 異なる特性を持つ複数リンクの選択的利用サービス (2000 年秋)

品質特性の異なる経路を選択的に利用する方式に関する研究は、従来より継続的におこなわれている。通信品質制御に経路選択の手法を組み合わせることによって、さらに多様なサービスを提供できる。しかし、このような経路制御は実運用にはまったく利用されておらず、実験的な運用に関する報告も見られない。このような状況から、2000 年秋の研究会の camp-net における実験では、帯域と往復所要時間というふたつの特性が異なる回線を利用者が選択的に利用できるサービスを提供した。

#### 協調しない複数の転送機構による優先転送サービス (2001 年春)

制御層と転送層の分離モデルでは、あるネットワークに実装される制御層と転送層が 1 対 1 に対応するとは限らない。実際のサービス提供の観点から予測すると、単一の制御層で管理される領域内に複数の協調動作しない転送機構が共存する可能性が高い。2001 年春の研究会では、このような場合に転送機構間の資源競合を調停するため、単一制御機構で包括的に制御することで、利用者に転送機構に依存しないサービスの提供を目指した。

また、経路上で通信の識別子を変換するエンティティが存在した場合に、変換の前後で通信の識別を一貫しておこなう手法についても検証をおこなった。

## 5.2. 輻輳するリンク上での優先制御サービス

この実証実験は、2000 年春の WIDE プロジェクト研究会で実施された。この研究会は 2000 年 3 月 14 ~ 17 日に山梨県石和町で開催され、参加者は 236 名であった。

### 5.2.1. camp-net の概要

この研究会の camp-net は、インターネットとの接続点として慶應大学湘南藤沢キャンパス (以下、藤沢と記す)、および、奈良先端科学技術大学院大学 (IPv6 のみ; 以下、奈良と記す) が設定された。対外接続用の回線としては、ATM over T1(1.5Mbps) が用意され、JGN (Japan Gigabit

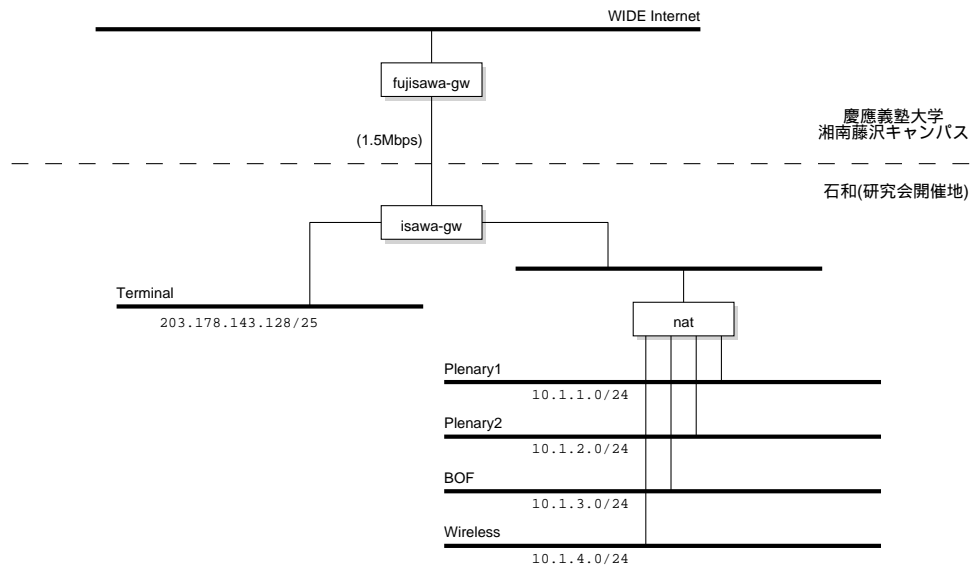


図 5.1: 2000 年春の研究会における camp-net の全体構成 (IPv4)

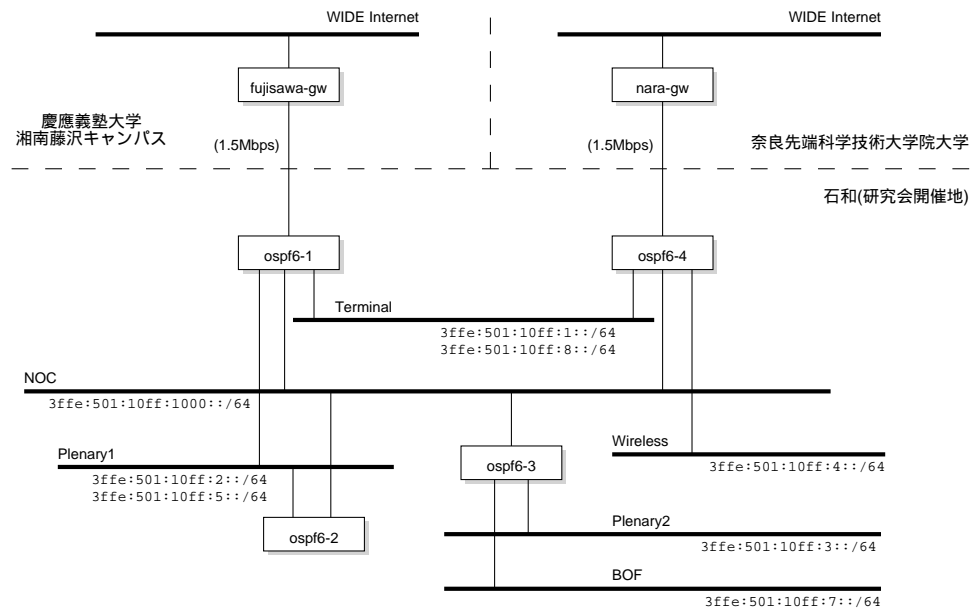


図 5.2: 2000 年春の研究会における camp-net の全体構成 (IPv6)

2000 年春の研究会における camp-net の全体構成を示す。OSPF version 3 の実験の要求により、IPv4 と IPv6 で完全に異なるネットワークが形成された。それぞれの構成図において、上部は慶應大学湘南藤沢キャンパスと奈良先端科学技術大学院大学 (IPv6 のみ) からインターネット全体を、下部は石和 (研究会開催地) を表す。石和からインターネットへの対外リンクは 1.5Mbps の ATM が用意された。図中の対外リンクはすべて対外リンク上に設定された VC であり、合計して 1.5Mbps の帯域であったことに注意が必要である。なお、図中では他の実験や DNS 等のサービスに利用されたサーバ、および、緊急用の冗長系に関しては適宜省略している。

Network)<sup>1</sup>を経由してインターネットとの接続性を確保した。この対外リンク上には3本のPVCが設定され、それぞれ以下の用途で利用された。

1. 慶應義塾大学湘南藤沢キャンパスを経由した、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との接続
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

研究会開催地のネットワークは、他の実験の要求により、インターネット層の視点からIPv4とIPv6が完全に異なるものとして設計された。この研究会のcamp-netの全体構成を図5.1、および、図5.2に示す。

### 5.2.2. 提供サービスの詳細

本節では、研究会参加者に提供したサービスの詳細について述べる。

#### サービス・クラス

研究会参加者がおこなう通信の多くは、camp-net外のインターネットへ向かうものである。したがって、camp-netから発せられる通信の多くは、必然的にcamp-netとインターネットを接続する対外リンクへと集中する。5.2.1節でも述べたように、この対外リンクはcamp-netを構成するネットワークの中でも、もっとも狭帯域なものである。そこで、この対外リンクを通信品質制御サービスの対象とし、研究会開催地の境界からインターネットとの接合点までをDSドメインとして定義した。

サービス・クラスとしては、最善努力(Best Effort; BE)と優先転送の2種類を用意した。研究会参加者は、ネットワークに対して通信品質制御に対する要求を発行することにより、これらのいずれかのサービスを楽しむことになる。

最善努力クラスに関連付けられた通信に対しては、特別な処理は一切おこなわない。一方、優先転送に関連付けられた通信には、あらかじめ定義された固定帯域のロットが割り当てられ、AF(Assured Forwarding) [15]方式で転送する。ロットに関しては、後節で詳説する。ふたつのサービス・クラスに対する通信の関連付けは、

- 参加者から要求が発行された場合、指定された通信を優先転送
- 上記に含まれない通信は最善努力

<sup>1</sup>TAO(放送通信機構)との共同研究による利用。

とした。

参加者が優先転送要求を発行する際には、対象となる通信の識別情報と、優先転送の時間を指定する。通信の識別情報としては、送信者と受信者の IPv4/IPv6 アドレス、トランスポートプロトコルの種別、および、送信者と受信者のポート番号を指定することができる。したがって、TCP であれば単一の接続単位の粒度で指定することが可能である。優先転送の継続時間は分単位で指定する。サービス中での優先転送要求の取り下げには対応しなかった。

また、すべての優先転送要求に対しては、一様にひとつのスロットが割り当てられる。したがって、要求時にスロット数を指定することはできない。スロットは後述するように有限であるため、すべてのスロットが割り当てられている場合には、優先転送要求は拒否される。また、転送要求の割り込み (プリエンプト) には対応していない。

## 課 金

限られた資源である対外リンクの帯域を扱うために、wideunit という仮想通貨の概念を導入した。また、“wideunit” は仮想通貨の単位でもある。研究会参加者は、サービスに対する対価として wideunit を支払う。この疑似課金処理により、参加者の意思による利用要求の反映と、有限である資源の平等な利用の両立を実現することができる。特に、資源利用の平等性は、少数の参加者による帯域の独占を抑止するために重要である。

仮想通貨は、研究会初日に参加者に対して 2,000wideunit 一様に配布した。疑似課金処理をおこなうため、優先転送要求に対する単価を設定しなければならない。単価の導出には市場 (研究会参加者) の優先転送要求に対する需要を見積もる必要があるが、初の実証実験であったため、予測が困難であった。そのため、研究会期間を通じて参加者ひとりにつき 1 スロット 200 分を仮定した。また、ひとつの優先転送要求 (対外リンクの固定帯域 1 スロットを占有) を 1 分間継続するために、10wideunit が必要であると設定した。

## 要求発行インタフェース

研究会参加者には、予約発行をネットワークに送信するためのクライアント “PEPe” を配布した。参加者の持参する端末は、Windows や UNIX の亜種など多岐にわたるため、C 言語、および、Java 言語で実装することにより対応した。

優先転送要求を発行するためには、参加者の ID とパスワードが必要となる。それに加え、5.2.2 節で述べた、優先転送の対象となる通信の識別情報、および、優先転送の継続時間を指定する。

要求に対する処理が終了すると、ネットワークから結果と必要に応じてその理由が伝えられる。

また、優先転送中の通信に関する情報と残り継続時間、および、wideunit の残高を確認できる

表 5.1: 制御対象機器 (1)

	石和	藤沢/奈良
IPv4	isawa-gw	fujisawa-gw
IPv6(藤沢)	ospf6-1	
IPv6(奈良)	ospf6-4	nara-gw

表 5.2: 制御対象機器 (2)

	石和	藤沢/奈良
IPv4	isawa-gw	fujisawa-gw
IPv6(藤沢)		
IPv6(奈良)		

表 5.1 は、対外リンクの両端に位置する第 3 層の機器を制御する場合を表す。藤沢の IPv4/IPv6 ルータが同一である以外はすべて異なる機器であり、5 台のルータを制御・運用しなければならない。一方、表 5.2 は、PVC ブリッジング機能を利用して制御機器を集約した場合を表す。camp-net とインターネットの間を流れるトラフィックは、すべて isawa-gw と fujisawa-gw を通過するため、制御が必要な機器はこの 2 台のみに減少する。

システムを、Web インタフェイスで提供した。

### 5.2.3. 制御層・転送層の設計と運用

本節では、前述のサービスを実現するための、camp-net の制御層と転送層の設計について述べる。

#### PEP の集約

5.2.2 節で述べたように、DS ドメインとして優先転送制御を行う対象は、camp-net とインターネットとを接続する対外リンクである。したがって、対外リンクの両端に位置する機器を PEP として設定し、優先転送をおこなわなければならない。この位置の第 3 層の機器 (IPv4/IPv6 ルータ) を PEP とする場合、5.2.1 節に示したトポロジから、表 5.1 で示したように、制御対象機器は 5 台となる。この台数は仮設ネットワークの迅速な立ち上げの圧迫と、運用上における負荷の増加をもたらす。このような負荷を軽減するため、PVC ブリッジング技術を利用して制御対象機器の台数を減少させた。

PVC ブリッジング技術によって集約した PEP のデータリンク層の構成を、図 5.3 および図 5.4 に示す。この図のように、対外 ATM リンク上に設定された 1 本の IPv4 用 PVC と 2 本の IPv6 用 PVC は、石和/藤沢においてそれぞれ isawa-gw/fujisawa-gw に集約された。

石和では、isawa-gw が IPv4 用ルータとして設定された。すなわち、camp-net からインターネットへ向かう IPv4 用対外リンクは、一度 isawa-gw で終端されることになる。一方、IPv6 用対外リンクは、IPv6 ルータである ospf6-1 および ospf6-4 から ATM を経由し、一度 isawa-gw に収容される。isawa-gw では PVC ブリッジング技術を利用することで、インターネット層を経由することなく、ATM のままそれぞれの対地へと接続された。藤沢では、fujisawa-gw が IPv4/IPv6 ルー

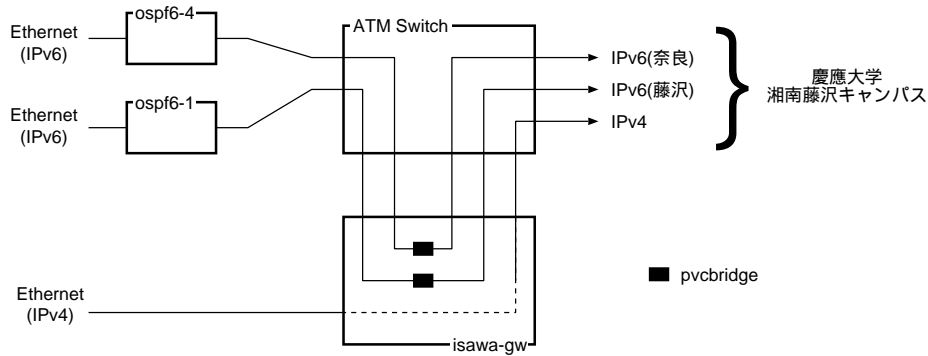


図 5.3: 石和 (研究会開催地) におけるデータリンク層の構成

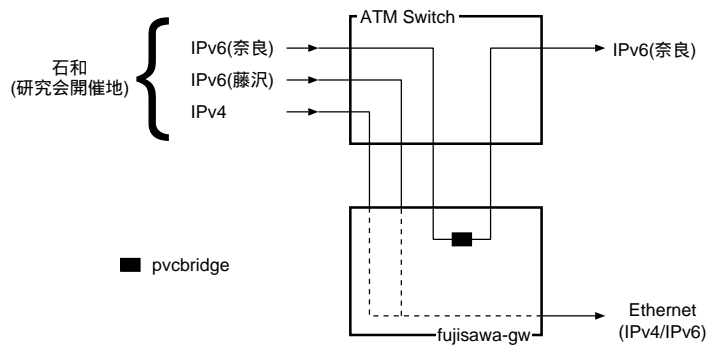


図 5.4: 藤沢におけるデータリンク層の構成

運用上の付加を軽減するため、PVCブリッジング技術を応用して、camp-net 側およびインターネット側の PEP を集約した。図中の黒い四角形で表されたものが PVC ブリッジであり、インターネット層を経由せずに転送処理がおこなわれる。また、点線で通過しているものは、インターネット層における転送であることを表す。これにより、camp-net とインターネットとの間に発生するトラヒックは、すべて isawa-gw および fujisawa-gw を通過することになり、制御対象となる機器を約 1/3 に圧縮することで、運用上の負荷や、仮設ネットワークの設置負荷を軽減することができた。



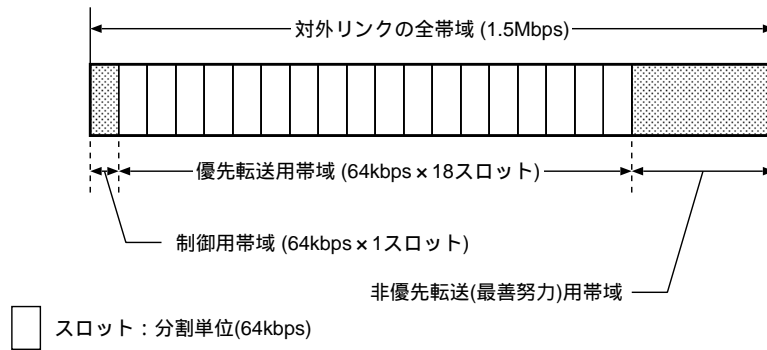


図 5.5: 対外リンクの分割

1.5Mbps の帯域を持つ対外リンクを 20 区画に分割した。このうち 18 区画は 64kbps の帯域を持ち、スロットと呼ばれた。スロットは優先転送要求に対して割り当てる帯域の単位となる。また、320kbps の帯域を持つ区域は、非優先転送トラフィック用として利用された。

タとして設定された。camp-net とインターネットを接続する対外リンクのうち、IPv4 用に設定された PVC、および、IPv6 用に設定された PVC の一方は、fujisawa-gw のインターネット層で終端される。IPv6 用に設定されたもう一方の PVC は、PVC ブリッジング技術によってインターネット層で終端されることなく、奈良に設置された IPv6 ルータである nara-gw へと接続された。

このように、優先転送の制御対象となる PEP を減らすことにより、仮設ネットワーク敷設と運用時の負荷を軽減することができた。また、IPv4/IPv6 の全ての対外リンクを isawa-gw/fujisawa-gw に集約することで、画一的な帯域資源管理が可能となった。

### 対外リンクの区画分割

要求駆動型品質制御サービスの提供は、この研究会における camp-net が初めてである。研究会参加者がサービスを直感的に理解できるようにするため、優先要求に対しては固定帯域を割り当てた。

まず、対外リンクの ATM over T1(1.5Mbps) の帯域を 20 区画に分割した。このようすを図 5.5 に示す。20 区画のうち、64kbps の帯域を持つ 1 区画は経路制御や優先転送制御などの運用管理を目的として利用された。この区画を利用するトラフィックは、全てのトラフィックの中でもっとも優先される。残りの 19 区画は研究会参加者の通信によって発生するトラフィックのために利用される。このうち 18 区画はスロットと呼び、それぞれが 64kbps の帯域を持つように設定した。スロットは研究会参加者からの優先転送要求に対して、割り当てられる帯域の単位となる。ひとつの優先転送要求に対しては一律に 1 スロットが割り当てられるため、優先転送される通信には 64kbps の帯域を確保することになる。また、スロットの制限から同時に許可される優先転送要求の最大値

は 18 となる。最後の 1 区画は 320kbps の帯域を持つ。この区画は優先制御とならないトラフィックに割り当てられた。

それぞれの区画の帯域は、その区画に割り当てられた通信が完全に占有するわけではない。また、利用されていない帯域は、他の区画に割り当てられた通信に貸し出される場合がある。すなわち、優先転送の対象となった通信は、対外リンクが輻輳していない場合には、64kbps 以上の帯域を利用することもできる。

### 通信の分類と優先制御

利用者の要求に対してスロットが割り当てられると、PEP は境界流入ノードとして動作し、AF PHB にしたがってマーキングおよび優先転送処理がおこなわれる。この処理には ALTQ [11] を利用した。また、AF PHB グループとしてはクラス 1 を利用した。

まず、camp-net あるいはインターネットから、DS ドメインである対外リンクへ流入する IP データグラムの DSCP に対し、isawa-gw および fujisawa-gw の入力インタフェースでマークを付加した。マーキングの方式は TRTCM (Two Rate Three Color Maker) [30] を採用した。あるスロットに対応づけられた通信の廃棄優先度は、64kbps 未満のものは低 (AF11) に、64kbps 以上 96kbps 未満のものは中 (AF12) に、そして 96kbps 以上のものは高 (AF13) にマークされる。それぞれのマークに対する DSCP は、表 4.2 で示した AF PHB の DSCP の推奨値にしたがった。

次に、出力 (対外リンク側) インタフェースのバッファで、入力インタフェースで付加された DSCP にしたがって転送処理される。AF11 にマークされたものは優先的な転送処理がおこなわれるのに対し、AF13 にマークされたものは非優先転送トラフィックと同様に、輻輳時には最善努力方式と同様の確率で廃棄される。このような出力インタフェースにおけるスケジューリングには、H-FSC (Hierarchical Fair Service Curve) [10] 方式を、AF を実現するための機構としては、RIO (Random Early detection with In/Out) [8] 方式を利用した。

### 細目条件と制御パラメータの配布

本実験における制御層の処理手順は、基本的に COPS [7] で定義されたものにしたがった。したがって、制御パラメータの配布は COPS を利用して配布され、各 PEP で ALTQ の設定を通じて反映される。PDP は camp-net 内に設置され、研究会期間を通じてふたつの PEP との TCP コネクションが常に維持された。また、COPS のクライアント種別としては、独自に定義した widcamp を利用した。

サービス・クラスを定義する各種の細目条件は、単一のモジュールとして PDP に組み込まれた。このモジュールには以下の細目条件が組み込まれた。

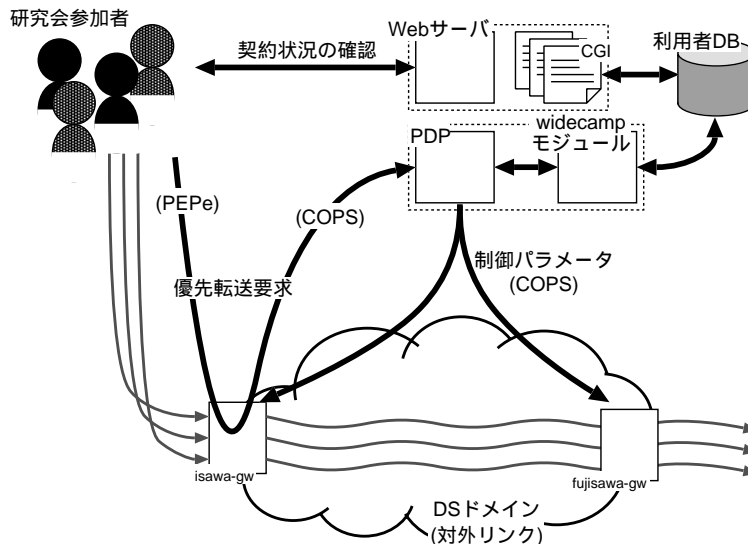


図 5.6: 優先制御の手順

優先制御要求の発行から要求の転送層への反映までを示す。研究会参加者は、あらかじめ配布した専用クライアント“PEPe”を利用して優先制御要求を発行する。要求は PEP から PDP に転送されて応諾の可否が決定され、必要に応じ COPS を利用して転送層に制御要求を発行する。

### 認証/課金

PEP を通じて送信された優先制御要求に含まれる認証情報により、利用者の認証をおこなう。また、利用時間に応じて課金処理もおこなう。

### 資源管理

既に述べたように、64kbps の帯域を持つスロットを資源として定義した。優先転送に利用する 18 スロットの割り当ての管理をおこなう。

### 契約管理

要求には優先転送の希望継続時間が含まれている。継続時間が経過すると、占有していたスロットを解放しなければならない。このような契約管理をおこなう。

### 制御パラメータの導出

利用者から指定された通信を表す属性や空き資源などから、PEP に配布するための制御パラメータを導出する。

上記の処理において必要となるデータは、すべて PostgreSQL [31] を利用して管理された。

表 5.3: 輻輳の発生時間

	1 回目	2 回目	3 回目	4 回目	5 回目
開始時刻	15 日 18:14:50	16 日 12:33:37	16 日 13:42:54	16 日 16:19:47	16 日 20:55:27
終了時刻	15 日 18:19:05	16 日 13:39:28	16 日 13:49:03	16 日 19:39:47	16 日 23:49:24

明らかな輻輳によって研究会参加者が示す挙動を観察するために、人為的な輻輳を発生させた。輻輳は研究会期間中に 5 回発生させた。この表は輻輳を発生させた時間を示している。

#### 5.2.4. 優先制御の流れ

利用者の優先制御要求の発行から、要求への応諾、さらに転送層への反映までの流れを図 5.6 に示す。

1. 利用者は PEPe を利用して優先制御要求を発行する。要求は最寄りの PEP(この camp-net では isawa-gw) によって受理される。
2. PEP は PDP に対して REQ メッセージを発行、優先制御を要求する。
3. PDP は要求の内容を確認、利用者を認証したのち、利用者の wideunit の残高や対外リンクの残余スロット数を考慮して、要求の応諾の可否を決定する。
- 4a 優先転送要求を許可する場合 ...PDP は制御要求を DEC メッセージで PEP に送信、PEP は機器に制御要求を反映して成否を RPT メッセージで PDP に返答する。RPT メッセージを受け取った PDP は、その内容に応じて要求発行元の PEP(isawa-gw) に DEC メッセージを送信する。
- 4b 優先転送要求を拒否する場合 ...要求発行元の PEP(isawa-gw) に、拒否理由を含む DEC メッセージを送信する。
- 5 PEP(isawa-gw) は利用者に結果を報告し、優先転送要求に対する処理を終了する。

#### 5.2.5. 輻輳の発生

camp-net の DS ドメインとなる対外リンクは帯域が 1.5Mbps の回線であり、研究会参加者が通常利用する限りは輻輳はほとんど発生しない。輻輳発生時の研究会参加者の挙動を観察するため、DS ドメイン上で大量のトラヒックを生成して人為的な輻輳を発生させた。

人為的なトラヒックの発生には UDP を利用し、camp-net とインターネットの接続点から camp-net に向けて送信した。輻輳を発生させた時間を表 5.3 に示す。

### 5.2.6. 結果と考察

通常時の対外リンクにおける帯域利用のようすを、図 5.8 に示す。稀に回線帯域である 1.5Mbps に達することもあるが、全体として継続的な輻輳は発生していない。一方、人為的な輻輳を発生させたときの帯域利用のようすを、図 5.9 に示す。対外リンクの帯域をほぼ完全に消費していることがわかる。この状態ではインターネットから camp-net へ流入する IP データグラムの喪失が継続的に発生した。また、図 5.7 や表 5.4 に示すように、camp-net とインターネットの間の往復所要時間の急激な増加が見られた。このように、輻輳を発生させた期間中は通信品質の急激な悪化が見られた。

研究会参加者 236 名のうち、通信品質制御要求を一度以上発行した参加者は 41.5%にあたる 98 人であった。図 5.10 に研究会参加者からの要求の発行数の変化を示す。夜間を除くと、発行数が徐々に増加しているようすが見られる。また、網をかけた部分は人為的輻輳を発生させた時間を表す。利用者が研究会が進むにつれ、輻輳に対して敏感に反応するようになっている。図 5.11 は、発行された要求のうちエラーになったものを示す。この値はアドミッション制御において要求が許可されなかった数を示している。5.2.2 節や 5.2.3 節で述べたように、本実験では同時に許可できる優先転送要求が 18 個に制限されていた。上記のエラーは、そのほとんどが資源枯渇によって要求が拒否されたものである。

以上のように、研究会参加者からの品質制御要求は輻輳時に集中した。通常時は輻輳時に比べてそれほど多くの要求は見られない。通信品質要求の発行数やそれに対するエラー数の結果は、利用者が輻輳時における通信品質制御の有用性を理解し、要求を発行する行動を表している。また、前述したように利用者からの制御要求数は研究会の進行とともに、わずかながらに増加している。これは、要求駆動型帯域制御サービスが初めての試みであったため、研究会開始時には多くの参加者が要求の発行に対する躊躇があったためであるとみられる。しかし、優先転送の有効性を理解した参加者は繰り返し要求を発行するようになった。

これらの結果は、輻輳時の優先転送制御が利用者にとって有用であり高く評価されることを示している。逆に、通常時には優先転送クラスと最善努力クラスの通信品質に大きな差はなく、相対的に優先転送クラスの価値が低く評価される。インターネット上での突発的な輻輳の発生や、高品質通信を要求するアプリケーションの利用は、予測することが困難である。そのため、個人利用者にとっては固定的制御よりも要求駆動型制御によって、自らの品質を柔軟に指定できるほうが有用であると云えるだろう。

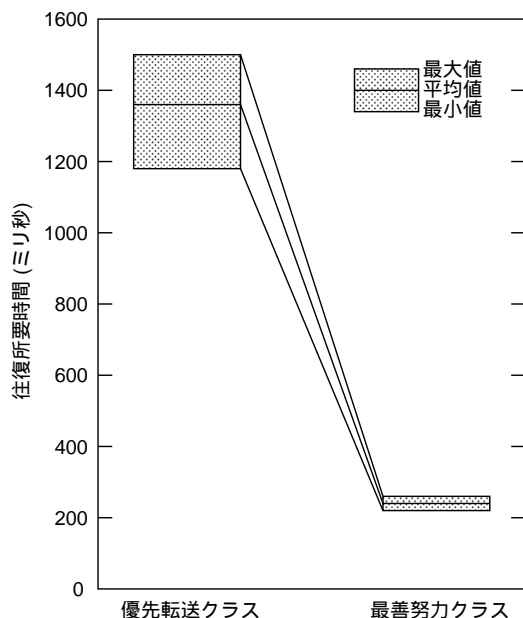


図 5.7: 輻輳時の往復所要時間

表 5.4: 輻輳時の往復所要時間 (ミリ秒)

	最小	最大	平均
優先転送クラス	223	259	241
最善努力クラス	1194	1503	1347

右図 5.7、および、表 5.4 は、輻輳時における camp-net からインターネットへの往復所要時間を表す。計測は isawa-gw と fujisawa-gw の間でおこなっている。このように、輻輳時の最善努力クラスには通信品質の急激な悪化が見られるのに対し、優先転送クラスでは対話的なアプリケーションの利用が可能な品質が保たれている。

### 5.3. 異なる特性を持つ複数リンクの選択的利用サービス

本実験は、2000 年秋の WIDE プロジェクト研究会で実施された。この研究会は 2000 年 9 月 11 ~ 14 日に長野県茅野市で開催され、参加者は 239 名であった。

#### 5.3.1. camp-net の概要

この研究会で設置された camp-net の全体構成を、図 5.12 に示す。インターネットへの接続点として慶應義塾大学湘南藤沢キャンパス (以下、藤沢と記す) が設定された。藤沢へは、茅野と直接結ばれる経路のほかに、奈良先端科学技術大学院大学 (以下、奈良と記す) を経由するものが用意され、ともに IPv4/IPv6 によるインターネットへの接続性を提供した。このふたつの経路の選択は、OSPFv2(IPv4) と OSPFv3(IPv6) を利用した動的経路制御によっておこなった。OSPF のコストを調整することで、藤沢に直接向かう経路が優先的に選択され、障害時に奈良を経由する経路が選択されるようにした。

camp-net から藤沢、あるいは、奈良へ接続するための対外リンクとしては、JGN<sup>2</sup>を経由する ATM over T1(1.5Mbps)、および、衛星回線<sup>3</sup>(2Mbps)<sup>4</sup>が用意された。ATM 回線上には、256kbps

<sup>2</sup>TAO(放送通信機構) との共同研究による利用。

<sup>3</sup>(株)JSAT との共同研究による利用。

<sup>4</sup>片方向リンクとして利用した場合の帯域。この帯域を複数に分割して、両方向リンクのように利用したり、別拠点との通信に利用することができる。

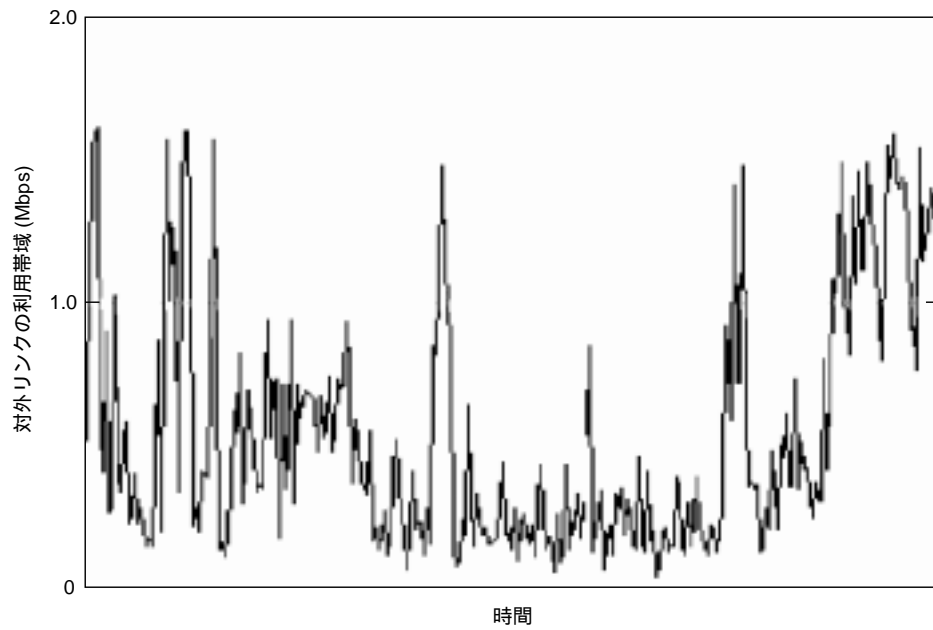


図 5.8: 対外リンクのトラヒック (通常時)

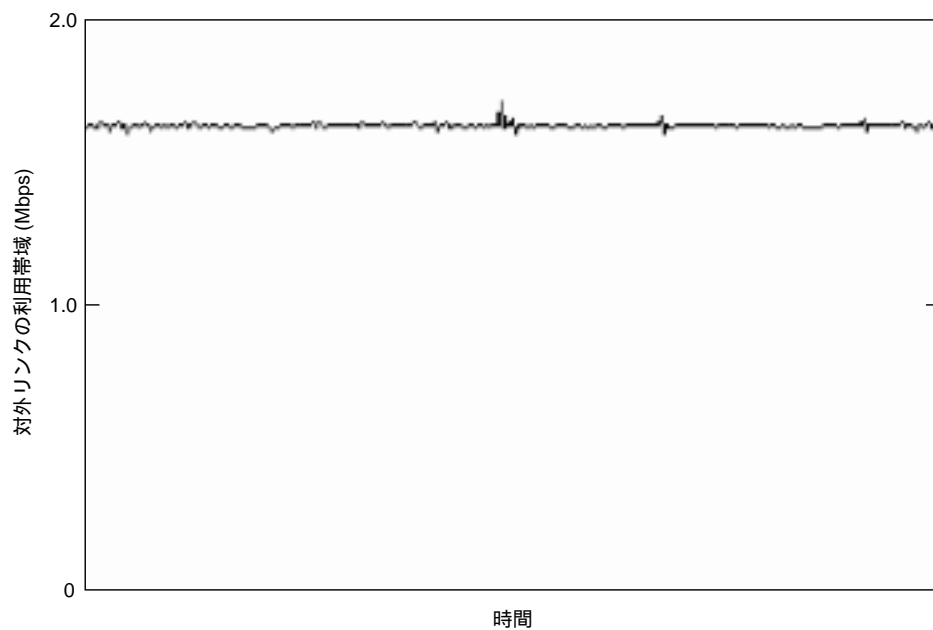


図 5.9: 対外リンクのトラヒック (輻輳発生時)

対外リンクにおける帯域利用のようすを示す。対外リンクの帯域は 1.5Mbps である。図 5.8 は通常時のようすであり、稀に上限に達する状態も観測できるが、継続的な輻輳は見られない。図 5.9 は輻輳時のようすである。通常二の場合に対して、輻輳発生時には帯域を完全に消費する状態が継続的に発生している。

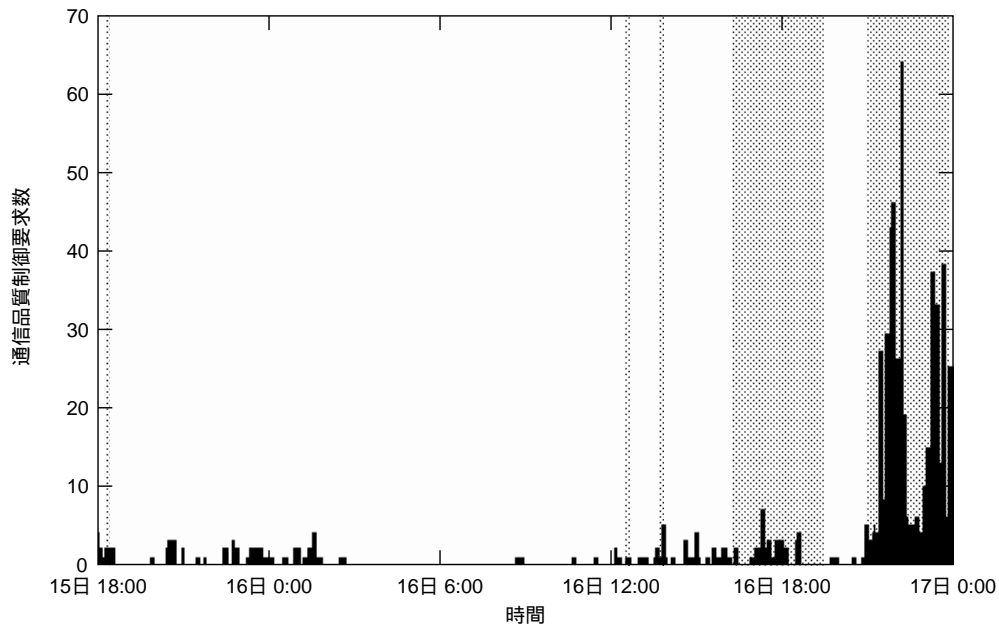


図 5.10: 通信品質制御要求の発行数

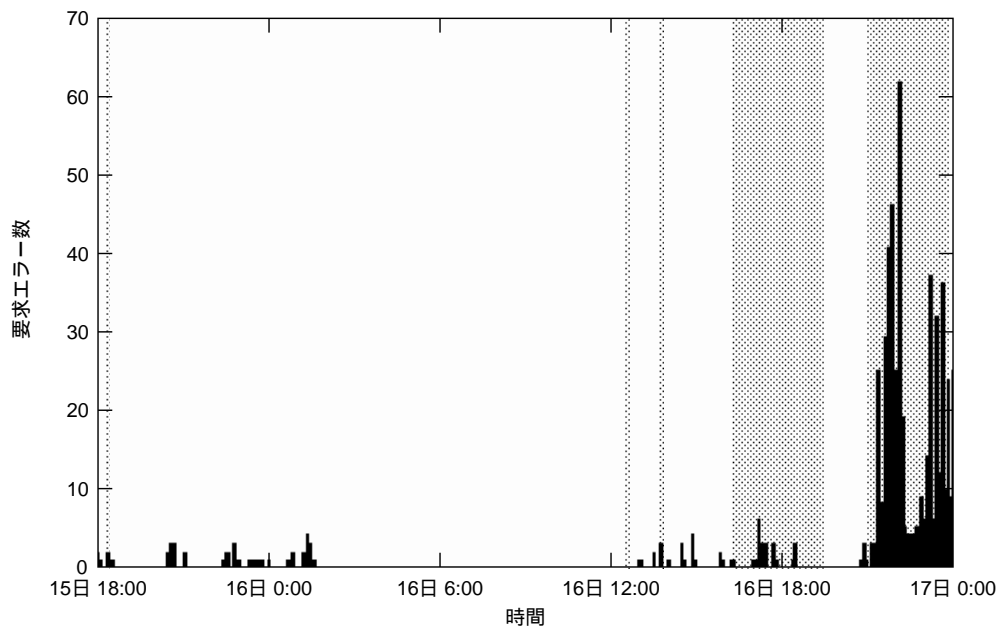


図 5.11: 通信品質制御要求の発行数 (エラー分)

研究会参加者からの通信品質制御要求の発行数と、それに対するエラーの発生数を示す。



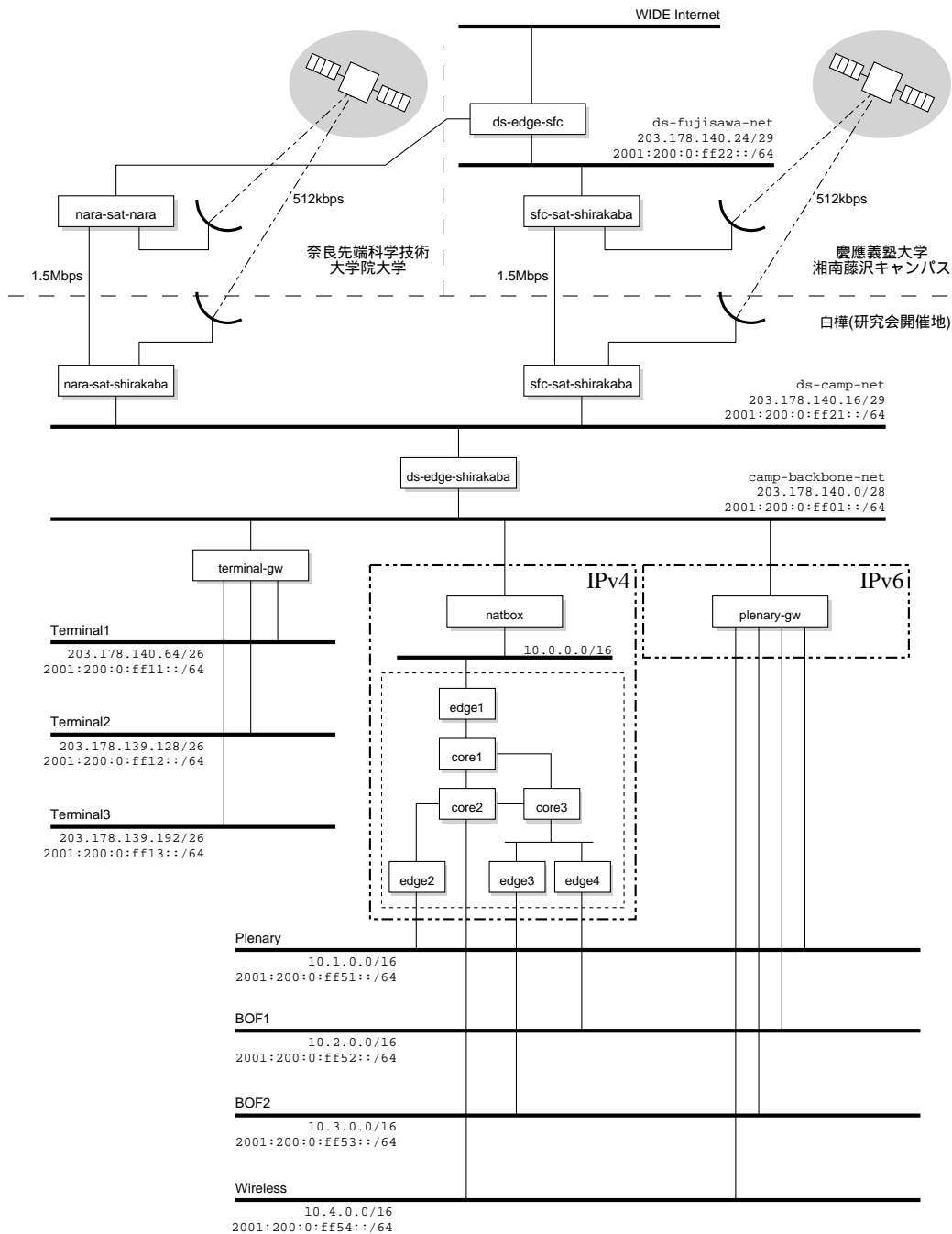


図 5.12: 2000 年秋の研究会における camp-net の全体構成

2000 年秋の研究会における camp-net の全体構成を示す。上部は慶應大学湘南藤沢キャンパスと奈良先端科学技術大学院大学からインターネット全体を、下部は茅野(研究会開催地)を表す。図中では理解を容易にするために衛星をふたつ示しているが、実際には同一の衛星である。なお、図中では他の実験や DNS 等のサービスに利用されたサーバ、および、緊急用の冗長系に関しては適宜省略している。

表 5.5: camp-net における ATM 回線と衛星回線の特性の違い

	帯域	遅延
ATM 回線	狭 (256kbps)	小 (50 ミリ秒程度)
衛星回線	広 (512kbps)	大 (500 ミリ秒以上)

camp-net の対外リンクである ATM 回線と衛星回線の特性を示す。藤沢/奈良の両拠点に対する PVC は 256kbps で設定されたため、ATM 回線より衛星回線が広帯域であるのに対し、衛星回線の特性より遅延時間は ATM 回線が大幅に小さい。

表 5.6: サービス・クラスの比較

	利用回線	優先制御
AF クラス 1	ATM 回線	あり
AF クラス 2	衛星回線	あり
最善努力クラス	衛星回線	なし

提供した 3 種類のサービス・クラスを示す。AF クラス 1 および AF クラス 2 は、それぞれ ATM 回線および衛星回線における優先転送である。優先制御要求の対象となっていない通信は、特別な処理をされずに衛星回線を通じてインターネットに転送される。

の帯域を持つ 2 本の PVC と 512kbps の帯域を持つバックアップ用の PVC が設定された。このうち、藤沢および奈良へは 256kbps の帯域を持つ PVC によって接続された。一方、衛星回線は 512kbps の帯域を持つ 2 本の双方向リンクとして、camp-net とそれぞれの拠点を接続した。

### 5.3.2. 提供サービスの詳細

本節では、研究会参加者に対して提供したサービスについて述べる。

#### サービス・クラス

この camp-net で用意した ATM 回線と衛星回線の遅延特性を表 5.5 に示す。この表にみるように、2 種類の回線は著しく異なる特性を持つ。したがって、利用者は通信の特性に応じてこれらのリンクを選択できることが望ましい。そこで、これらのリンクを包含する対外リンク部分を DS ドメインと定義し、利用者からの要求にしたがって 2 種類のリンクを選択的に利用できるサービスを提供した。DS ドメインは、図 5.12 では ds-edge-shirakaba と ds-edge-sfc で挟まれた領域になる。

サービス・クラスとしては、2 種類の優先転送と非優先転送となる最善努力クラスを提供した。2 種類の優先転送クラスは AF クラス 1/AF クラス 2 と呼ばれる。これらのクラスに対応づけられ

表 5.7: サービス・クラス別の往復所用時間 (ミリ秒)

	最小	最大	平均
AF クラス 1(ATM 回線)	36.063	515.888	48.419
最善努力 (衛星回線)	552.909	904.997	568.590

camp-net からインターネットとの接合点までの往復所要時間を、ping コマンドを利用してサービス・クラスごとに計測した。それぞれの条件で ICMP ECHO REQUEST は 800 回送出され、ともに損失なく ICMP ECHO REPLY を受信した。

た通信は、それぞれ ATM 回線および衛星回線で要求した帯域が確保され、優先的に転送される。一方、最善努力クラスに対応づけられた通信の転送には、衛星回線が利用される。この際、帯域確保や優先転送等の処理は加えられない。

研究会参加者のインターネットに対する通信は、特別な要求を発行しない限り最善努力クラスとして扱われる。また、参加者から要求が発行された場合、その通信は要求にしたがって 2 種類のいずれかの優先転送クラスに対応づけられる。

インターネットとの通信における AF クラス 1 と最善努力クラスの往復所要時間を表 5.7 に示す。この表は、ping コマンドを利用して計測した往復所要時間の最小値/ 最大値/ 平均値を示している。これによれば、AF クラス 1 と最善努力クラスの最小往復所要時間の間には、衛星回線の伝送遅延によって生じる 500 ミリ秒の差があることがわかる。また、往復所要時間の最大値においても 400 ミリ秒近い差がある。

## 課 金

2000 年春の研究会における実験と同様に、優先転送に公平性を持たせるため、“wideunit” と呼ばれる仮想通貨を導入した。この仮想通貨は、研究会開催時にすべての参加者に対して一様に 10,000wideunit が配布された。

課金額は次の 4 種類の要素によって決定した。

- サービス・クラス
- 占有する帯域
- 優先転送の継続時間
- 回線占有率

まず、AF クラス 1(ATM 回線) のサービスにおいて、回線帯域を 1kbps 占有した優先転送を 1 分間継続する場合に必要な額を、1wideunit と設定した。さらに、優先転送要求により占有される帯

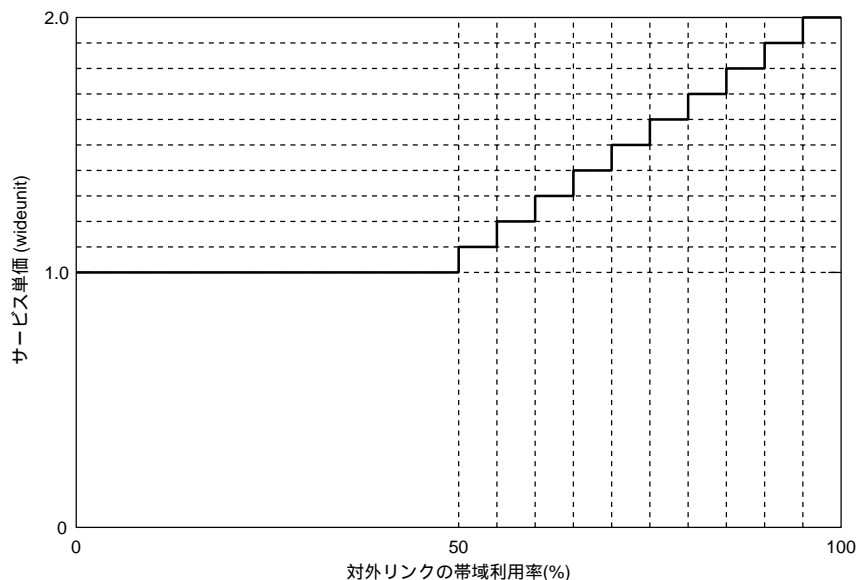


図 5.13: AF クラス 1 に対する単価の変動規則

AF クラス 1 に対するサービス単価の変動を示す。たいがいリンクの帯域利用率が 50%以下の場合には一定であるが、50%を上回ると徐々に単価が上昇する設定とした。95%以上では、単価は 50%以下の場合の 2 倍に達する。なお、AF クラス 2 の場合も同様の変動規則で 1/4 の単価とした。

域に応じて、回線の混雑を緩和する目的で単価を上昇させた。この上昇の規則を図 5.13 に示す。この図のように、占有帯域が 50% を越える状態では、5% ごとに単価を 0.1wideunit 上昇させた。したがって、回線の帯域の 95% 以上が優先転送によって占有された場合、その単価は初期単価の 2 倍に達する。

AF クラス 2(衛星回線) に関しても、AF クラス 1 と同様の規則にしたがって課金をおこなった。ただし、往復所要時間の差を反映して AF クラス 1 の 1/4(小数点以下繰り上げ) とした。

## 要求発行インタフェース

優先転送要求を発行するためのインタフェースとして、専用の Web ページを提供した。Web ページを閲覧するためには HTTP に基づく認証がおこなわれ、要求を発行する利用者の識別子とパスワードが求められる。

要求を発行する際には、利用する対外リンクの種別 (ATM 回線または衛星回線) と占有する帯域幅、優先転送の継続時間、そして、優先転送の対象となる通信の識別情報を入力する。識別情報は、送信者および受信者の IP アドレスとトランスポート層プロトコルの種別、および、そのポート番号で構成される。また、受信者アドレスに関してはアドレスのプレフィックス長を指定するこ

とにより、通信先のネットワークを表現することもできる。

利用者は、優先転送契約の継続期間中であれば、同じ Web ページ上から契約を更新することができる。更新できるのは、要求帯域の増加と契約継続期間の延長である。

### 5.3.3. 転送層の設計と運用

転送層では、要求に応じた通信の分類、および、それに基づく DSCP マーキングと優先制御、さらには、DSCP に基づく利用対外リンクの選択がおこなわれる。

#### 通信の分類と優先制御

AF クラス 1/2 および最善努力クラスには、それぞれ AF PHB グループ 1/2 および Default クラスを利用した。また、DS ドメインの境界ノードとなる ds-edge-shirakaba と ds-edge-sfc では、ALTQ を利用して通信の分類とマーキングをおこなった。

DS ドメインに流入するトラフィックに対しては、まず、すべての処理の前にあらかじめ付与されている DSCP を消去した。次に、分類器が制御層からの要求にしたがい設定されている規則に基づいて分類し、その結果と計測器の判断から適切な DSCP が付与される。AF PHB グループに対する廃棄優先率の設定は、利用者から要求された占有帯域までが低 (AFi1)、占有帯域以上から占有帯域の 1.2 倍以下は中 (AFi2)、それ以上は高 (AFi3) にマークされる。AF クラス 1/2 のそれぞれのマークに対する DSCP の値は、表 4.2 で示した AF PHB の DSCP の推奨値にしたがった。また、最善努力クラスに対応する DSCP は Default PHB を表す 000000b とした。

マーキングが終わると、DSCP の値にしたがってシェーピングや廃棄処理が適用される。出力インタフェースのキュー管理には RIO をもちいた。RIO のパラメータは AF PHB グループの廃棄優先度に従うように設定された。

#### リンク選択器

DiffServ の概念では、DSCP は PHB を選択するためのものであり、経路やリンクの選択への利用は考慮されていない。ここでは、DSCP の意味を拡張し、PHB に加えてリンクの選択もおこなうこととした。前節で述べたように、DSCP は利用者からの融点転送要求に応じてマークされる。したがって、上記の拡張により SLA に応じたリンクの選択が実現できる。

リンク選択器は、複数の物理リンクを集束し、単一の仮想的な Point-to-Point リンクとして扱うものである。仮想リンクに対してデータが送出されると、あらかじめ決められた規則にしたがって実際に送出する物理リンクが選択される。たとえば、IP データグラムのヘッダにおける送信者や受信者アドレス、IP プロトコル番号、DSCP、トランスポート層プロトコルやそのポート番号

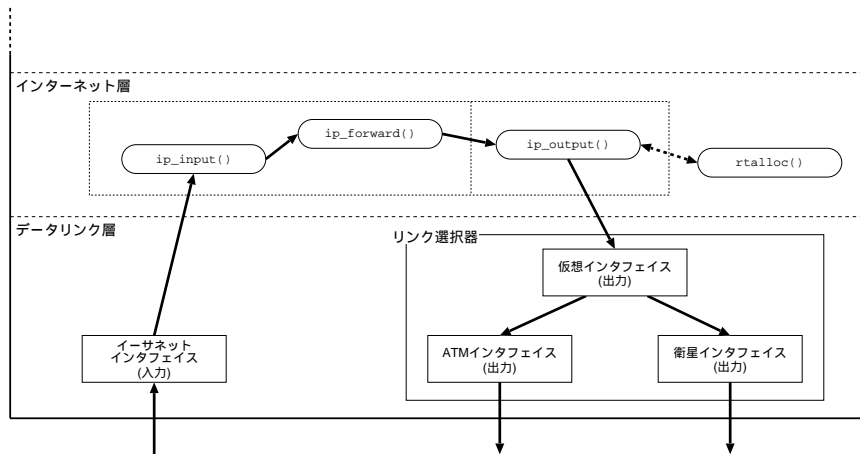


図 5.14: ルータに実装されたリンク選択器の構造

実験に利用したリンク選択器の構造を示す。ip\_output() から仮想インタフェースに出力された IP データグラムは、DSCP の値によって適切な物理インタフェースから出力される。

などを、物理リンク選択のための判断要素として利用することができる。また、これらの判断要素と物理リンクを選択する規則の対応は、柔軟に設定できる必要がある。また、仮想リンクはアプリケーション層からは従来の Point-to-Point リンクとして扱われるため、既存の経路制御プロトコル等を改変せずに利用することができる。

この camp-net では、リンク選択器の非常に限られた機能のみを、ATM 回線と衛星回線を収容するルータに実装した。この実装では、物理リンクとしては ATM 回線と衛星回線のみを収容することができる。また、物理リンクを選択するための判断要素は DSCP のみである。ルータに実装されたリンク選択器の構造を図 5.14 に示す。

### 5.3.4. 制御層の設計と実装

#### PDP 主導パラメータ配布

本実験では、研究会参加者からの優先転送要求は Web インタフェースを通じて発行される。発行された要求は PDP へ通知され、アドミッション制御や制御パラメータの導出がおこなわれる。利用者からの要求は PEP を経由することはない。したがって、要求に対する一連の制御は PDP 主導で開始される。制御パラメータの配布に COPS を利用する場合、PDP 主導の制御は COPS-PR [32] 拡張の利用が想定されている。しかし、COPS-PR では PEP が生成したクライアント・ハンドルを変更せずに使い続けるため、契約単位で制御パラメータの変更や破棄といった処理の実現が困難である。

今回の実験では、利用者が要求を発行することで、契約内容を更新することが可能である。したがって、契約単位でクライアント・ハンドルが割り当てられ、それをもちいて契約更新を転送層に反映できることが望ましい。そのためには、PDP においてクライアント・ハンドルを生成できる必要がある。そこで、本実験で利用した COPS では、PDP でクライアント・ハンドルを生成できるような拡張を加えた。

PDP でもクライアント・ハンドルを生成する場合、ある PEP と PDP の組み合わせにおいて、両者で同一のクライアント・ハンドルが生成される可能性がある。この競合を回避するために、COPS の共通ヘッダに次のフラグを追加した。

#### 0x08

##### PDP 生成クライアント・ハンドル

PDP で生成したクライアント・ハンドルによって処理をおこなう場合、同時にこのフラグを指定する。これにより、PEP と PDP のクライアント・ハンドル空間を分離することができる。

#### アドミッション制御と制御パラメータの導出

前節で述べたように、制御パラメータの配布には COPS を利用した。アドミッション制御や制御パラメータの導出を構成する細目条件は、COPS PDP の単一モジュールとして設計・実装した。組み込まれた細目条件を以下にあげる。なお、それぞれの細目条件において利用される認証情報は、要求発行に利用する Web ページを閲覧する際の情報を利用した。

#### 資源管理

地上回線および衛星回線の占有帯域を管理する。要求に対して十分な帯域が確保できない場合は、要求を拒否する。

#### 課金

優先制御に対する課金をおこなう。課金が対外リンクの残り帯域と連動しているため、資源管理と連携してサービス単価を算出する。

#### 契約管理

契約継続時間経過後の優先制御解除をおこなう。また、契約更新要求にともなう情報の管理もおこなう。

### 5.3.5. 優先制御の流れ

優先転送要求から転送層に反映されるまでの流れを、図 5.15 に示す。

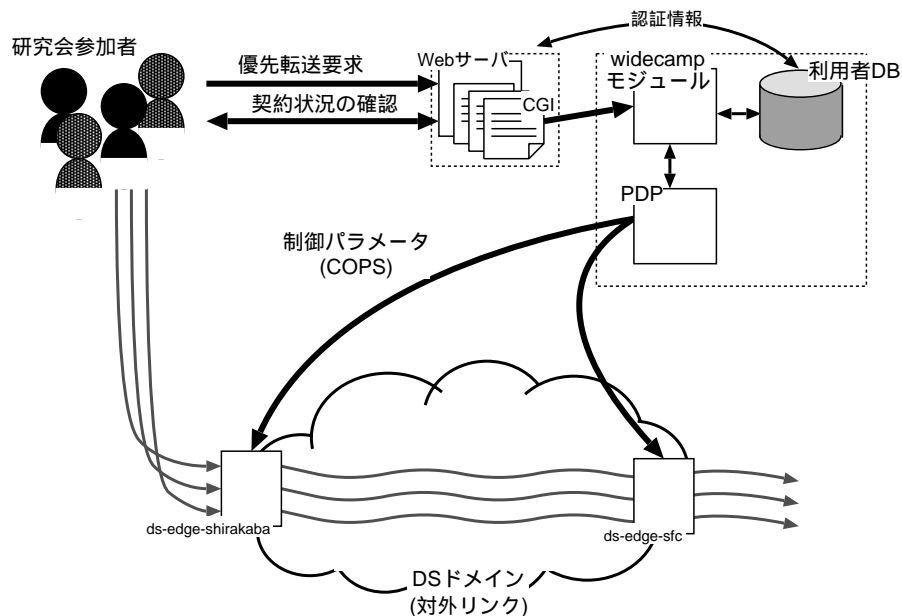


図 5.15: 優先制御の手順

利用者は、優先制御要求を Web インタフェース経由で発行する。要求の発行を受けた PDP モジュールは、アドミッション制御や制御パラメータの導出をおこなひ、COPS PDP を通じて転送層に反映する。なお、認証は Web ページを閲覧する際におこなう。この認証情報は、その後のアドミッション制御等にも利用される。

1. 利用者は適当な Web ブラウザから優先転送要求ページに接続する。接続にはあらかじめ配布された利用者 ID とパスワードが要求される。認証はバックエンド DB のデータを照会して処理される。なお、この認証情報は PDP に引き継がれ、後のアドミッション制御などに利用される。
2. 優先対象となる通信の識別情報など、必要な情報を入力して優先転送要求を発行する。
3. 要求の発行を受けた PDP モジュールは、前述のアドミッション制御をおこなう。要求が許可される場合には、制御パラメータを導出し、PDP に対して転送層に反映するよう要求する。
4. PDP モジュールは PEP からの報告を集計し、処理の完了を利用者に報告する。
5. 利用者は Web インタフェースを通じて、契約の更新や内容変更、現在の契約状況の閲覧などをおこなうことができる。



```

% ping -n www.wide.ad.jp
PING www.wide.ad.jp (203.178.136.57): 56 data bytes
64 bytes from 203.178.136.57: icmp_seq=0 ttl=250 time=529.435 ms
64 bytes from 203.178.136.57: icmp_seq=1 ttl=250 time=528.705 ms
64 bytes from 203.178.136.57: icmp_seq=2 ttl=250 time=528.708 ms
64 bytes from 203.178.136.57: icmp_seq=3 ttl=250 time=528.710 ms
64 bytes from 203.178.136.57: icmp_seq=4 ttl=250 time=528.701 ms
64 bytes from 203.178.136.57: icmp_seq=5 ttl=250 time=528.715 ms
64 bytes from 203.178.136.57: icmp_seq=6 ttl=250 time=528.967 ms
64 bytes from 203.178.136.57: icmp_seq=7 ttl=250 time=528.939 ms
64 bytes from 203.178.136.57: icmp_seq=8 ttl=250 time=38.286 ms
64 bytes from 203.178.136.57: icmp_seq=9 ttl=250 time=37.789 ms
64 bytes from 203.178.136.57: icmp_seq=10 ttl=250 time=38.788 ms
64 bytes from 203.178.136.57: icmp_seq=11 ttl=250 time=38.783 ms
64 bytes from 203.178.136.57: icmp_seq=12 ttl=250 time=37.548 ms
64 bytes from 203.178.136.57: icmp_seq=13 ttl=250 time=37.801 ms
64 bytes from 203.178.136.57: icmp_seq=14 ttl=250 time=38.795 ms
64 bytes from 203.178.136.57: icmp_seq=15 ttl=250 time=38.292 ms
^C
--- www.wide.ad.jp ping statistics ---
16 packets transmitted, 16 packets received, 0% packet loss
round-trip min/avg/max/stddev = 37.548/283.560/529.435/245.300 ms

```

図 5.16: 利用者からみた優先配送制御の効果

AF クラス 1 の優先配送制御による、camp-net 内のホストからインターネット上のサーバに対する往復所要時間の減少を示す。ping コマンドで往復所要時間を観測しながら、優先配送要求を発行した。

### 5.3.6. 結果と考察

研究会の期間中、多くの参加者から優先転送要求を受けた。ここではその結果を報告するとともに、結果に基づく考察をおこなう。

#### 利用者からみた優先配送制御の効果

図 5.16 に、利用者からみた優先配送サービスの効果を示す。ping コマンドを利用して、camp-net 内部のホストからインターネット上のサーバまでの往復所要時間を計測しながら、AF クラス 1 による優先配送制御要求を発行した。icmp\_seq=7 と icmp\_seq=8 の間で往復所要時間が著しく減少しており、5.3.2 節で述べたサービス・クラスが提供できていることを確認できる。

#### 利用者の挙動

この研究会の参加者 252 名のうち、優先配送要求を発行したのは 126 名であった。また、発行された優先配送要求の総数は 605 であり、そのうちの 78.8%にあたる 477 が AF クラス 1 での要求であった。

図 5.17 は、AF クラス 1 に対する要求継続時間の分布を示す。48.2%が 6–30 分である一方、1 時間を越えるものも 23.3%あった。図 5.18 は、AF クラス 1 に対する要求帯域の分布を示す。1kbps とい

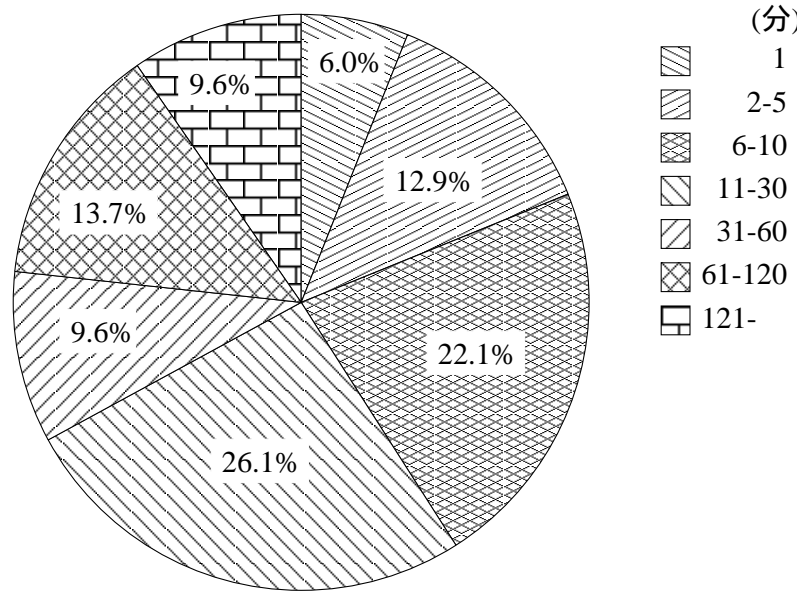


図 5.17: AF クラス 1 における要求継続時間の分布

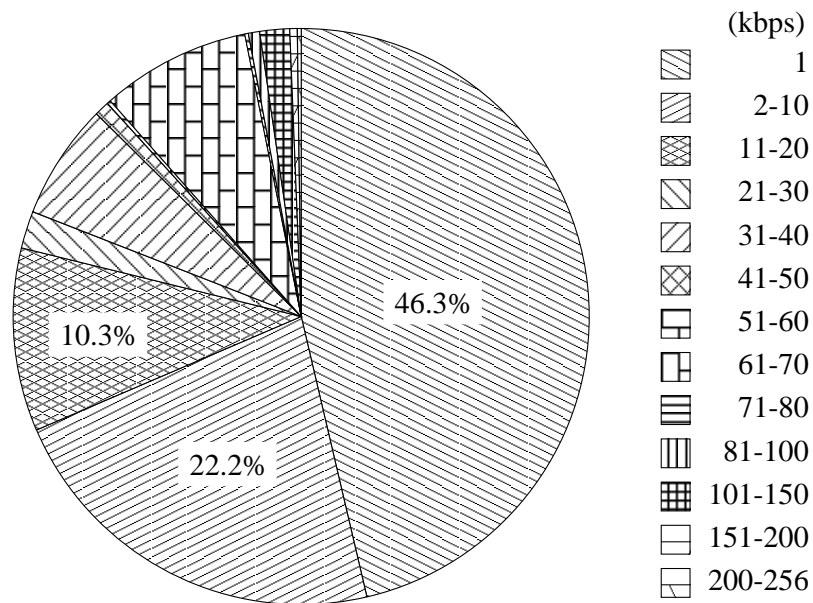


図 5.18: AF クラス 1 における要求帯域の分布

AF クラス 1 における要求内容の分布を示す。図 5.17 と図 5.18 は、それぞれ要求継続時間と要求帯域を表している。要求継続時間は短時間から 2 時間を越える長期にわたるものまで幅広い要求がみられるが、要求帯域は 10kbps 以下のものだけで 70%弱を占める、非常に偏った分布になった。

う最小単位での要求が 46.3%ともっとも多く、続いて 2–10kbps が 22.2%であった。一方、100kbps を越える要求を発行した利用者はごく少数であった。すなわち、AF クラス 1 での優先配送要求の大半が狭帯域での要求であった。

図 5.19 は、研究会期間中の要求継続時間の変化を表している。これによれば、要求継続時間は時間の経過とともに増加していることがわかる。特に、研究会 3 日目の夜から最終日にかけては、研究会終了時まで優先制御を継続するような要求が観測できる。図 5.20 は同じく研究会期間中の要求帯域の変化を表している。要求継続時間が増加傾向にあるのに対し、要求帯域はあまり変化しておらず、大部分が 10kbps 以下の要求であった。

図 5.21 と図 5.22 は、それぞれ AF クラス 1 に対する研究会期間中のサービス単価と、対外リンクの利用帯域の変化を表している。この図のように、研究会期間中を通じて AF クラス 1 の全ての帯域が確保されることはなかった。

### 考察と評価

すでに述べたように、利用者は長期間にわたる AF クラス 1 の契約を 1kbps で締結することがもっとも効果的であると判断した。この情報は研究会参加者の間に流布し、結果として多くの参加者が長期にわたる狭帯域な AF クラス 1 の優先転送要求を発行することになった。これは、提供するサービス・クラスの構成に整合性がなかったためであると考えられる。

まず、AF PHB グループを利用したサービスでは、利用者の挙動を予測・制御することは困難であることがわかった。AF では、対象となるリンクが極度の輻輳状態でない場合、確保する帯域による通信品質の差が明確に現れるわけではない。先に述べたように、今回の実験では AF の実現のために RIO をもちいた。RIO では、キューの長さが十分に長くない限り、廃棄優先度が高に設定された IP データグラムもほとんど廃棄されない。図 5.22 で示したように、ATM 回線の帯域は低遅延のサービスを求める参加者の数に比べて十分に広く、したがってキューはほとんど長くならなかった。この結果、AF クラス 1 で優先転送を要求した利用者は、要求したよりも実質的に多くの帯域を利用することが可能であった。

次に、ATM 回線の帯域が衛星回線とくらべて極端に狭くなく、結果として衛星回線の利点が強調されなかったことがあげられる。ふたつのサービス・クラスは、ATM 回線が対話的な低遅延を要求するアプリケーションでの、衛星回線が広帯域を要求する大容量転送アプリケーションでの利用を想定して提供された。しかし、このふたつの回線がもつ特性の違いのうち、伝送遅延の差が帯域幅の違いよりも大きく、ATM 回線の低遅延という利点によって衛星回線の広帯域という利点が半減してしまった。

特性の異なるリンクや経路を選択的に利用することで、通信品質の異なるサービスを提供でき

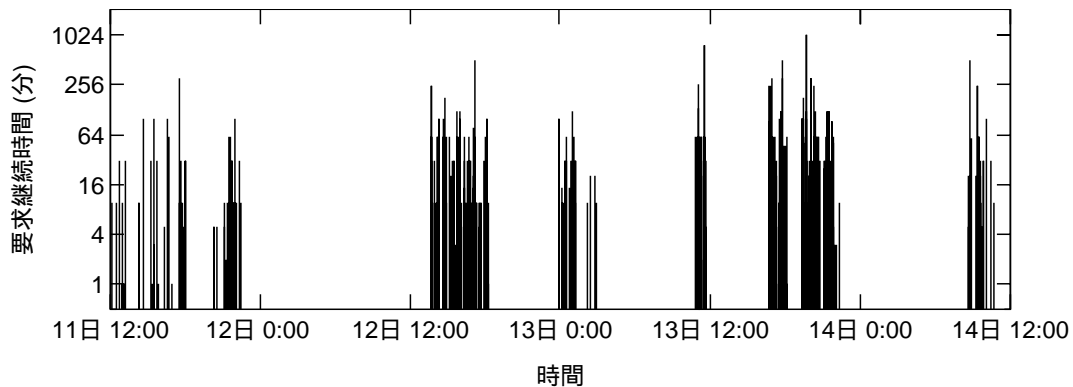


図 5.19: AF クラス 1 における要求継続時間の変化

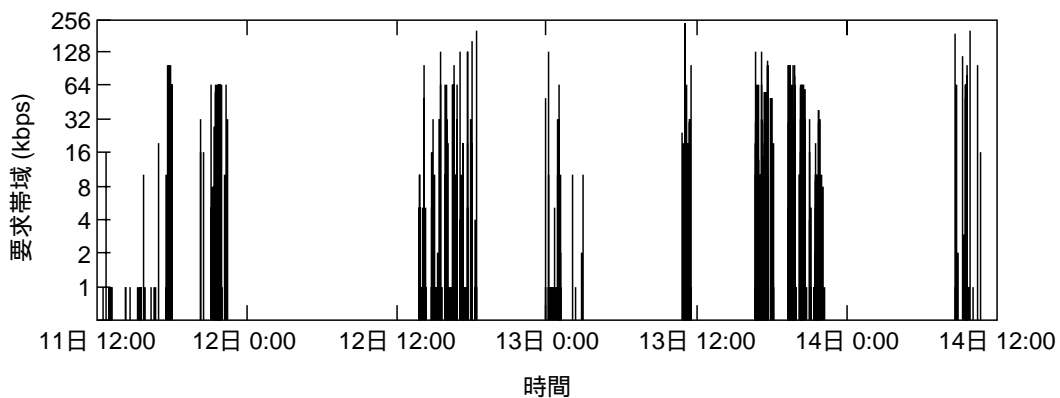


図 5.20: AF クラス 1 における要求帯域の変化

AF クラス 1 の研究会期間中における要求内容の変化を示す。図 5.19 と図 5.20 は、それぞれ要求継続時間と要求帯域の変化を表している。要求継続時間は、3 日目の夜に研究会終了時までの要求を発行する利用者が多く見られる。一方、要求帯域は研究会期間中はあまり変化がなく、ほとんどが 10kbps 以下のものであった。

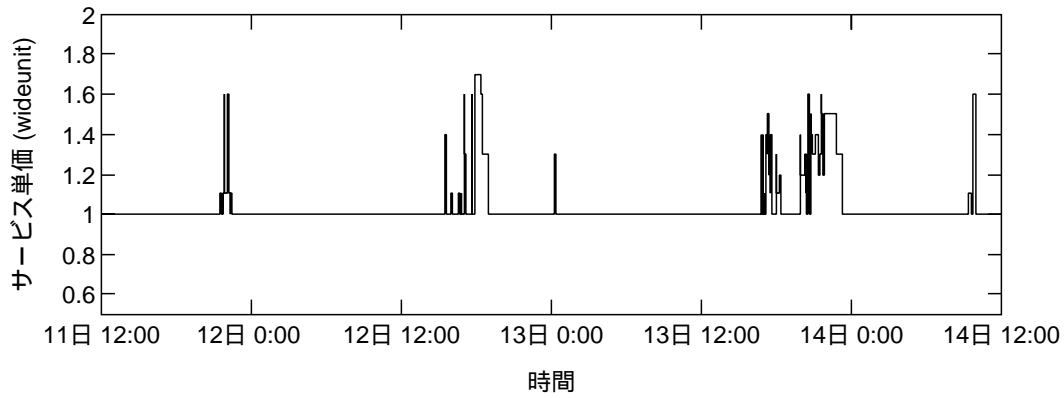


図 5.21: AF クラス 1 のサービス単価の変化

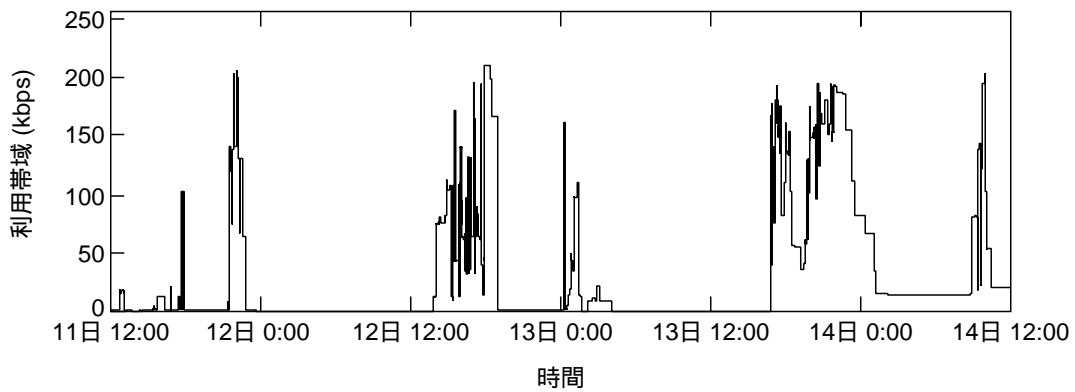


図 5.22: AF クラス 1 による対外リンクの帯域利用の変化

図 5.19 と図 5.20 は、AF クラス 1 に対する研究会期間中のサービス単価の変化、および、対外リンクの確保帯域の変化を表す。

る。しかし、それらのリンクや経路がいくつかの特性において異なる場合、最も差の大きな特性によって他の特性の利点を消し去る可能性がある。したがって、このようなサービスを提供する際には、提供するサービスの釣り合いが全体としてとれるように、利用者が選択可能な特性を注意深く選択する必要がある。

第3に、帯域に基づく単価設定が利用者を効果的に誘導できなかった。単価は対外リンクの帯域に対する確保帯域の割合で変動したが、多くの利用者が狭帯域での契約を締結したため、単価はほとんど変動せず、安価にAFクラス1サービスを楽しむことができた。その結果、研究会期間中にわたって1kbpsの契約を締結することが可能となった。すでに述べたように、利用者からみたATM回線の利点は低遅延特性にあることから、選択したリンクに基づく課金をおこなう必要があったと考えられる。このように、極度に異なる特性を持つ複数のサービス・クラスを提供する場合、課金の対象となる特性は慎重に選択しなければならないだろう。

一方、前章で述べた2000年春の研究会と比較すると、一度以上の要求発行をおこした参加者は増加していることがわかる。要求発行をおこなった参加者の数は、2000年春の研究会では参加者236名のうち41.5%にあたる98人であったのに対し、今回の研究会では総参加者の50.0%にあたる126人が要求発行をおこなった。研究会参加者がみずから要求することによって求める通信品質を楽しむことが、WIDE研究会参加者の間にわずかながらも浸透していることを示している。これは、一般の利用者に対して要求駆動型品質制御サービスを開始した際にも、浸透するにはある程度の時間が必要である可能性を示唆している。

## 5.4. 協調しない複数の転送機構による優先転送サービス

本実験は、2001年春のWIDEプロジェクト研究会において実施された。この研究会は2001年3月4～7日に、愛知県西浦町で開催され、参加者は270名であった。

### 5.4.1. camp-net の概要

この研究会のcamp-netは、研究会開催地におけるネットワーク、および、インターネットとの接続点となる慶應大学湘南藤沢キャンパス(以下、藤沢)のネットワークの一部から構成された。camp-netの全体構成を図5.23に示す。

西浦から藤沢までは、以下の対外リンクで接続された。

- 地上回線(128Kbps専用線×5をYAMAHA RT300にて集束)

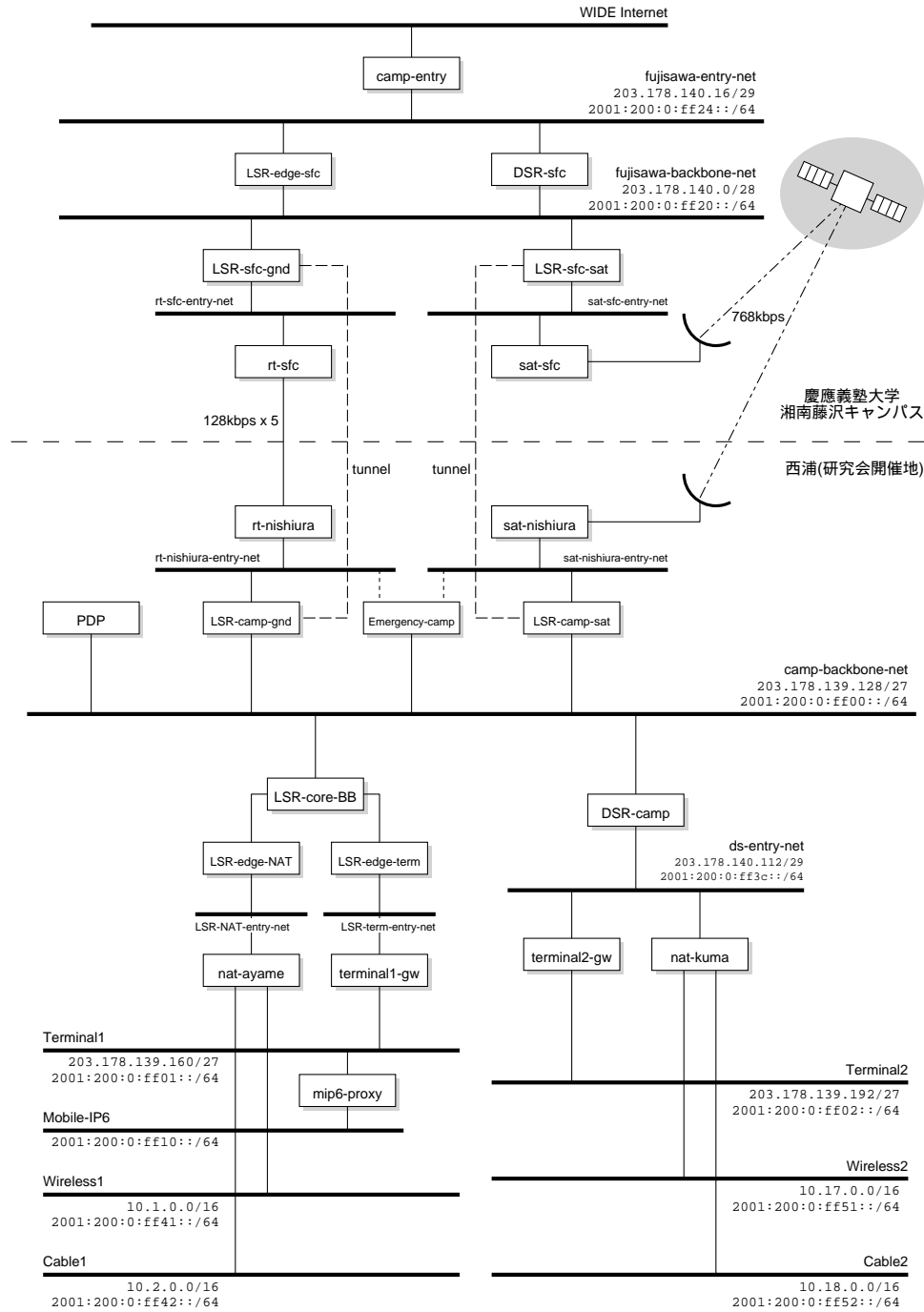


図 5.23: 2001 年春の研究会における camp-net の全体構成

2001 年春の研究会における camp-net の全体構成を示す。上部は慶應大学湘南藤沢キャンパスからインターネット全体を、下部は西浦 (研究会開催地) を表す。西浦から湘南藤沢キャンパスまでは、128kbps × 5 を集束した地上回線と、双方向 768kbps の衛星回線によって接続された。なお、図中では他の実験や DNS 等のサービスに利用されたサーバ、および、緊急用の冗長系に関しては適宜省略している。

表 5.8: 利用者収容ネットワークの種類

		NAT	
		あり	なし
転送機構	MPLS 系	nat-ayame	terminal1-gw
	L3TE 系	nat-kuma	terminal2-gw

利用者収容ネットワークの種類と収容ルータを示す。仮想専用線を申請した利用者が接続しているネットワークの種類により、制御層および転送層での処理内容が異なる。しかし、利用者はいずれのネットワークに接続していても、完全に同一のサービスを楽しむことができる。

- 衛星回線<sup>5</sup>(768Kbps の双方向対称リンク)

このうち、地上回線は 5 本の専用線を集束することで、論理的に 1 本の回線のように扱った。

対外リンクを含む camp-net は、IPv4/IPv6 とともに同じ構成でインターネットへの接続性を提供した。地上回線と衛星回線は OSPFv2(IPv4) および OSPFv3(IPv6) をもちいた動的経路制御により選択的に利用される。それぞれの回線のコストを調節することにより、衛星回線に障害が発生しない限り、通常のトラフィックは衛星回線を経由する設定とした。

研究会参加者は camp-backbone-net より下側のネットワークに収容される。このネットワークは表 5.8 に示すように、提供する転送機構やアドレス空間によって 4 種類に分類できる。通信品質制御要求を発行した参加者が収容されているネットワークの種類によって、制御層や転送層での処理が異なる。これについては後の節で詳しく述べる。

#### 5.4.2. 提供サービスの詳細

研究会 camp-net で提供した通信品質制御サービスは、MPLS と L3TE という 2 種類の異なる転送機構で提供した。しかし、利用する転送層の枠組みに関わらず、参加者は同一のサービスを楽しむことができる。本節では、研究会参加者に提供したサービスについて述べる。

#### サービス・クラス

今回の camp-net の構成は、2000 年秋の研究会と同様に、地上専用回線と衛星回線という特性の大きく異なる 2 種類の対外リンクが含まれている。したがって、研究会参加者に提供する通信品質制御サービスは、対外リンクの選択的利用に基づいておこなうことにした。また、利用者を

<sup>5</sup>(株)JSAT との共同研究による利用



収容するネットワークの出口ルータから、camp-net とインターネットの接合点までを通信制御対象領域とし、この領域を単一管理による DS ドメインとして設計した。

サービス・クラスとしては、最善努力と仮想専用線の 2 種類を定義した。仮想専用線は、地上回線上に要求帯域に応じた仮想的な専用線が確保され、経由するトラヒックの優先的な転送がおこなわれる。要求帯域は 4kbps/8kbps/16kbps/32kbps/64kbps から選択できる。また、確保された仮想専用線は、契約締結時に指定された継続時間が経過すると解放される。利用者は、仮想専用線契約の継続期間中であれば、その仮想専用線に収容する通信の加除を要求することができる。また、継続期間中の契約を強制的に破棄することもできる。

仮想専用線を経由しない通信は最前努力クラスにとして扱われる。このクラスに収容された通信には、特別な処理は適用されない。また、インターネット層の経路制御にしたがって、衛星回線を経由した通信となる。

### 要求発行インタフェース

サービスとのインタフェースには Web をもちいた。Web ページの閲覧には HTTP での認証がおこなわれ、研究会開始時に配布された利用者 ID とパスワードが求められる。仮想専用線は図 5.24 のページから申請することができる。申請には仮想専用線が通過する物理リンク (今回は地上回線のみ)、契約の継続時間、仮想専用線に収容する通信の識別子の入力が必要される。通信の識別には送信者 IP アドレス、受信者 IP アドレス、受信者 TCP ポートをもちいた。受信者 IP アドレスはプレフィックス長を指定することでネットワークを表現することもできる。

図 5.25 に契約履行情報のページを示す。このページでは、契約継続中の仮想専用線の情報や wideunit の残額を確認することができる。また、それぞれの仮想専用線に収容されている通信も表示されており、収容解除の要求を発行することもできる。さらに、仮想専用線の契約自体を破棄要求も発行可能である。

### 課 金

過去 2 回の実験と同様に、wideunit と呼ばれる仮想通過を導入し、サービスに対する疑似課金をおこなった。仮想通過は、研究会開催時にすべての参加者に対して 5,000wideunit が配布された。また、研究会期間中の 2:00AM に、全ての参加者に対して 5,000wideunit(最終日は 2,500wideunit) が追加された。また、サービス単価は 4kbps の仮想専用線を 1 分利用するために、10wideunit が必要であると定義した。仮想専用線に収容されている通信の数や種類は課金対象とならない。

仮想専用線の契約を破棄した場合は、契約の残り継続時間に基づいて wideunit が返金される。

新しい予約

新しい予約をします。

現在の WIDE UNIT は 10000000 です。

帯域 (kbps)

時間(分)

サービスタイプ

アドレスファミリー

送信元アドレス

送信先アドレス

ポート

プロトコル

送信先ポート

図 5.24: 仮想専用線申請インターフェイス

予約状況

現在の WIDE UNIT は 9961890 です。

予約は2件です。

取消	帯域	回線	状態
<input type="checkbox"/>	16 kbps	地上回線	有効
	3月5日 7時51分47秒 → 3月5日 8時21分47秒		
<input type="checkbox"/>	16 kbps	地上回線	有効
	3月5日 7時51分26秒 → 3月5日 8時21分26秒		
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80
<input type="checkbox"/>	PROTO SRC: 203.178.140.115/32	DST: 203.178.132.123/32	DPORT: 80

図 5.25: 契約履行情報

研究会参加者に提供したサービスへのインターフェイスを示す。図 5.24 は新規の仮想専用線を申請するためのインターフェイスであり、図 5.25 は継続中の契約の確認と破棄をおこなうためのインターフェイスである。

### 5.4.3. 転送層の設計と運用

仮想専用線の実現には、協調動作しないふたつの転送機構が利用されている。ここでは、それぞれの転送機構について説明する。

#### 2 種類の独立した転送機構

仮想専用線サービスを実現するの転送機構には、MPLS と第 3 層トラフィック・エンジニアリング (Layer 3 Traffic Engineering; L3TE) のふたつを利用した。図 5.23 における利用者収容ネットワークのうち、左側に位置する nat-ayame と terminal1-gw で収容されるものは、MPLS によって仮想専用線を実現する。また、右側の nat-kuma および terminal2-gw で収容されるネットワークは、L3TE による仮想専用線となる。それぞれの利用者収容ネットワークを、MPLS 系および L3TE 系と呼ぶ。図 5.26 に、これらの転送機構によって仮想専用線を構築するようすを示す。

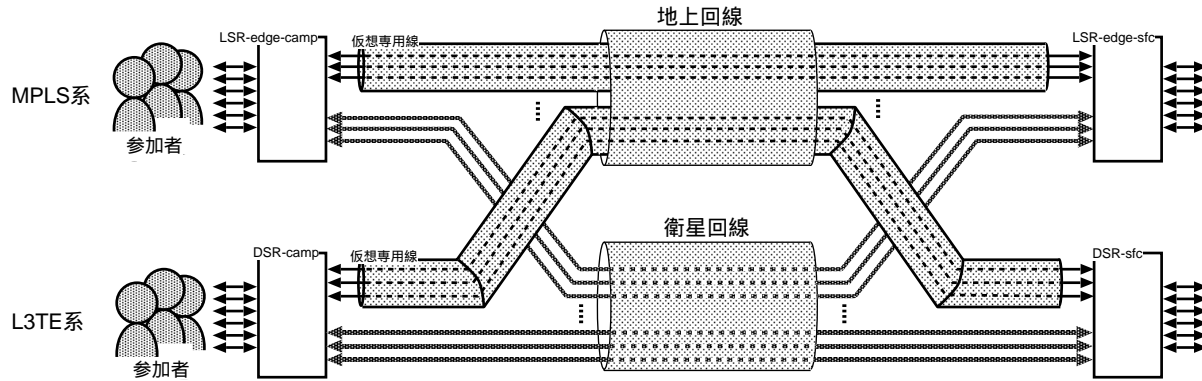


図 5.26: 2 種類の転送機構による仮想専用線の実現

2 種類の転送機構による仮想専用線確立のようすを示す。上側が MPLS 系、下側が L3TE であり、このふたつの転送機構は互いに協調して動作しない。

### MPLS

MPLS はラベル・スイッチングによって転送層から制御層を分離するための技術であり、あやめプロジェクトの開発する MPLS 環境 AYAME [33] を利用した。利用者収容ネットワークから藤沢に至る領域を MPLS 網として構築し、LSP 確立のためのラベル配布機構として明示的経路が指定可能な CR-LDP をもちいる。これにより、MPLS 系利用者収容ネットワークから発行された仮想専用線要求に対して、経由する対外リンクを動的に選択できる。

### L3TE

L3TE はインターネット層での経路決定機構に介入することによって、従来の受信者に加え、送信者やトランスポート層のポート番号などを加味した経路制御を実現するための技術である。対外リンクに流入する L3TE 系の前ホップとなる DSR-camp および DSR-sfc で運用し、通信の識別子と契約に応じて経路を選択した。

これらの転送機構は相互に協調せずに動作する。それぞれの転送機構内における資源管理の一貫性は維持できるが、網全体での資源管理の一貫性は保証できない。そのため、上位層で資源割り当てを解決する機構が必要となる。本実験では、ふたつの異なる転送機構が単一の制御層を共有し、サービス領域内の資源を包括的に管理することで、この問題を解決している。制御層に関しては後節で詳述する。

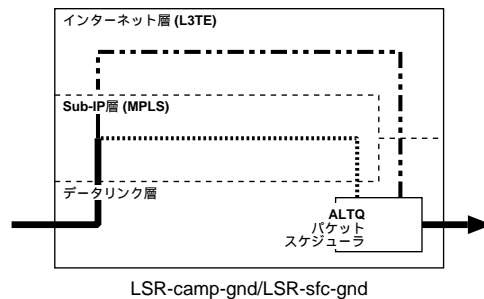


図 5.27: 地上回線の両端におけるパケット・スケジューラの共有

LSR-camp-gnd および LSR-sfc-gnd における、MPLS と L3TE によるパケット・スケジューラの共有のようすを示す。それぞれの転送機構で独自に処理されたデータは、出力インタフェイスのパケット・スケジューラで単一のクラスとして扱われる。

### パケット・スケジューラの共有

仮想専用線を構築するためには、制御層から指定される以下の 2 点の要求を満足する通信路を確立しなければならない。

- 明示的な経路
- 帯域の確保

転送機構として取り入れた MPLS と L3TE は前者の要求を満足できるが、それぞれ単独では後者の要求を満たすことができない。しかし、MPLS では DiffServ 拡張 [34] を利用することにより、L3TE はデータリンク層で独立して DiffServ を併用することにより、帯域を確保した通信路を構築できる。

一般に、通信の品質に対してなんらかの制御を加えるルータは、制御対象となる IP データグラムに対応づけられたクラスに応じて、入出力インタフェイスでパケット・スケジューリングによる処理をおこなっている。4.1 章で述べたように、DiffServ では出力インタフェイスでシェーピングや廃棄処理を実施するが、実装には特定のスケジューリング方式を利用している。

複数の転送機構が共存する環境では、出力インタフェイスでのパケット・スケジューリング要求が、それぞれの機構で独立して発生する。したがって、物理的に単一のインタフェイスの視点で見た場合、IP データグラムが転送機構の期待した挙動を示すためには、転送機構の種別に依存しないパケット・スケジューリング処理をおこなう統合された制御が必要となる。

本実験の camp-net では、地上回線の両端となる LSR-camp-gnd および LSR-sfc-gnd で、複数の転送機構からのパケット・スケジューリング要求が発生する。MPLS および L3TE の Diffserv 対応では、ALTQ をもちいたパケット・スケジューリングをおこなっている。ALTQ は、異なるプロトコル上で定義された通信の集合を、パケット・スケジューラ上で同一のクラスとして処理す

ることが可能な構造をもっている。これにより、MPLS および L3TE で期待されたパケット・スケジューリングが実現できる。LSR-camp-gnd および LSR-sfc-gnd の構成を図 5.27 に示す。

#### 通信の分類と優先制御

前節で述べたように、優先制御には DiffServ の枠組みを利用した。仮想専用線サービスを実現するために、優先制御には EF PHB を採用した。また、DS ドメインの境界ノードでは、ALTQ を利用した通信の分類とマーキングをおこなった。

MPLS 系では、境界ノードは LSR-edge-camp および LSR-edge-sfc となる。仮想専用線収容として分類された通信は E-LSP にマーキングされ、契約締結時に確立された LSP による転送がおこなわれる。また、LSP の中間 LSR では、E-LSP に基づく優先制御を適用した。

一方、L3TE 系では DSR-camp と DSR-sfc が境界ノードとなる。ここでは、分類器の結果にしたがって DSCP にマーキングされる。さらに、DSCP に基づいて経路が決定される。中間ノードでは、DSCP に基づく優先制御が適用される。

#### 5.4.4. 制御層の設計と運用

転送層が 2 種類の独立した機構から構成されているのに対し、制御層は単一の機構から構成される。ここでは、制御層における資源の管理や、中間エンティティへの対応について述べる。

##### 単一の制御層による資源管理

ネットワーク資源を利用する複数の転送機構が共存するための必要条件は、転送機構に対する資源の配分に関する調停作業がおこなわれることである。このため、転送機構間で相互に協調して動作するか、あるいは調停を担当する機構の介在が必要となる。

本実験で扱うサービスは仮想専用線であり、そこで必要とされる資源は明示的経路と地上回線の帯域である。これらの資源は転送機構に依存しない抽象的なものとして扱うことができるため、転送層から分離して管理することができる。camp-net で利用する MPLS と L3TE は互いに協調して動作しないが、資源を抽象化して単一の制御層で包括的に管理することで、上記の条件を満足することができる。

##### FEC とフィルタ規則の分離

FEC の定義はその FEC に収容する通信の識別子を含んでいることがある。特に要求駆動型の品質制御サービスに発せられる要求は、特定の通信に限定したものであることが一般的であり、そのため FEC の生成と同時に通信が対応づけられることが多い。しかし、本来の FEC の概念は必

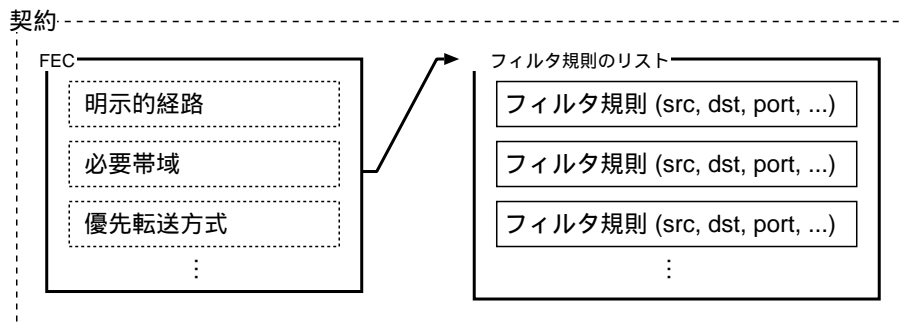


図 5.28: FEC とフィルタ規則の分離

本実験では、通信の品質特性を表す FEC と、FEC に収容される通信を表すフィルタ規則が明確に分離された。これにより、仮想専用線に収容する通信を自由に加除できるようになる。

ずしも特定の通信を束縛するものではなく、FEC の持つ通信品質特性とその FEC が収容する通信群は、独立した管理対象として扱うべきである。

本実験では、契約における FEC の通信品質特性を示す属性と、FEC に対応づけられる通信を示すフィルタ規則を分離して個別に管理した。契約と FEC、フィルタ規則の関係を図 5.28 に示す。これによって FEC を単独で適用して、その FEC に収容する通信をあとから自由に加除することが可能となり、仮想専用線を構築することができる。

#### 制御パラメータ配布

制御パラメータの配布には、COPS に基づくプロトコルを利用している。ただし、利用者からの要求は Web インタフェイスを通じ、直接 PDP に対して発行されるため、PDP 主導での制御パラメータ配布をおこなわなければならない。

このため、2000 年秋の研究会における実験で利用した、PDP によるクライアント・ハンドル生成の拡張を利用した。クライアント・ハンドルは仮想専用線ごとに個別のものが割り当てられた。仮想専用線に収容する通信の加除要求は、このクライアント・ハンドルを利用しておこなわれる。

#### アドミッション制御と制御パラメータの導出

サービス・クラスを定義する各種の細目条件は、それぞれが単独のサービスとして実装した。アドミッション制御と制御パラメータの導出は、細目条件統合機構がそれぞれの細目条件の結果を統合しておこなう。また、アドミッション制御・制御パラメータの導出と制御パラメータ配布の役割を明確するため、細目条件統合機構は COPS PDP と独立して実装した。細目条件は、課金、資源管理、契約管理から構成されている。図 5.29 に制御機構の構成を示す。

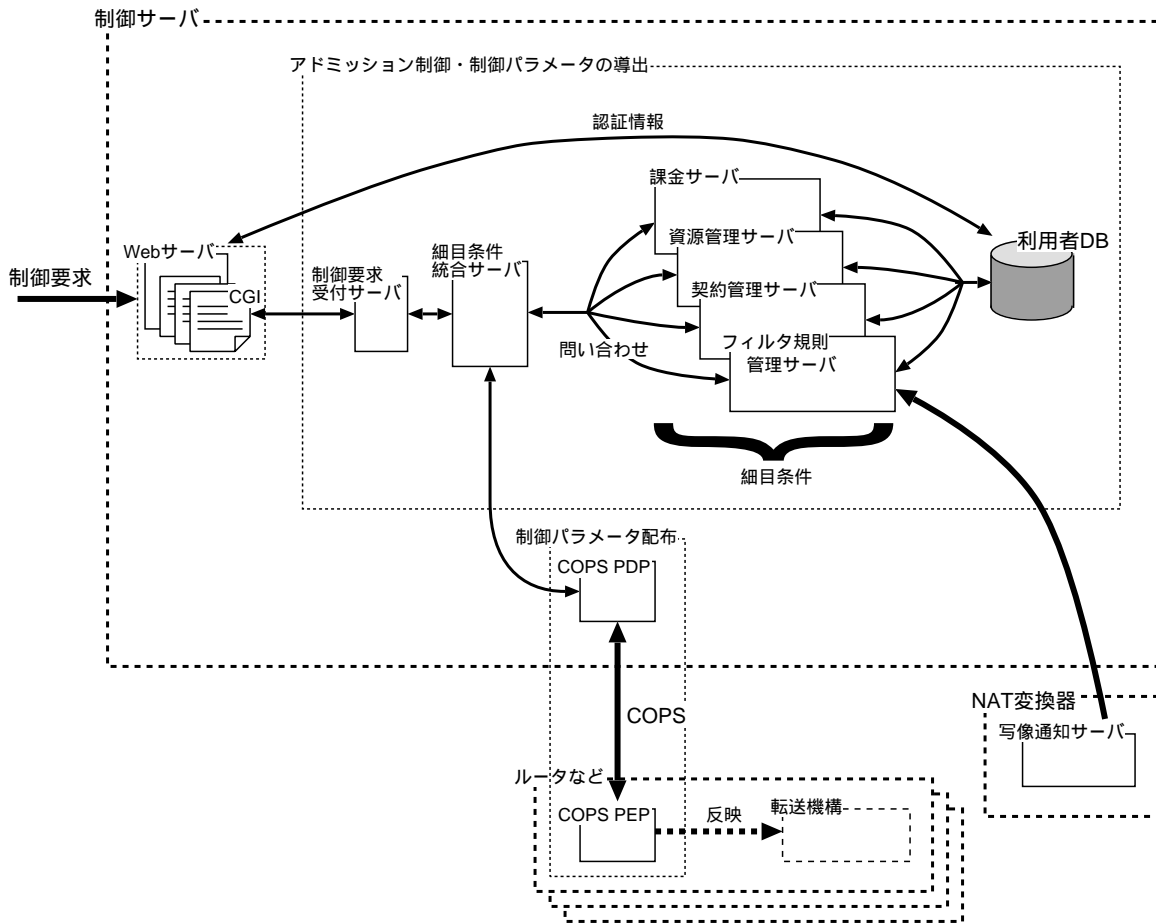


図 5.29: 制御機構の構成

制御機構は、細目条件の判断を担当する各種サーバと、それらを統合してアドミッション制御や制御パラメータの導出をおこなう細目条件統合機構からなる。

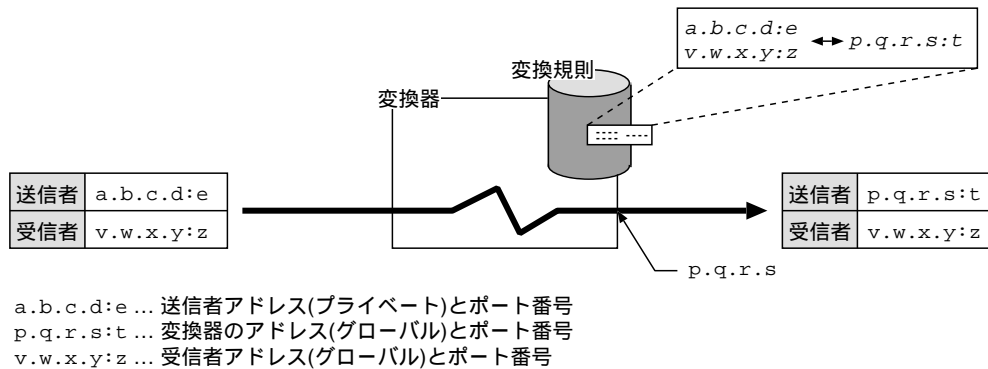


図 5.30: NAT の動作例

本実験では、通信の品質特性を表す FEC と、FEC に収容される通信を表すフィルタ規則が明確に分離された。これにより、仮想専用線に収容する通信を自由に加除できるようになる。

#### 5.4.5. NAT への対応

camp-net で利用できる IPv4 アドレス空間は、仮説ネットワークであるという特性からそれほど広くない。250 名を越える参加者が、ほぼひとりに一台以上のラップトップ・コンピュータを持参することに加え、対外リンクやサービス用、実験用にもアドレスを消費することから、そのままでは必要な IPv4 アドレスの個数を満たすことができない。そのため、利用者収容ネットワークの一部に NAT を導入し、プライベート・アドレス空間を利用することでこの問題を解決している。

#### NAT と通信の識別

NAT (Network Address Translator) [35] はおもに IPv4 アドレスの枯渇を一時的に回避するために利用される。プライベート・アドレス空間を利用するネットワーク (以下、プライベート・ネットワークと呼ぶ) とインターネットの境界上に変換器を設置し、IPv4 アドレスなどを変換することで、それらの間の透過的な通信を可能にする。IPv4 データグラムが変換器を通過する際、変換器は内部に蓄積した写像関係に従い、IP アドレスとトランスポート層のポート番号を書き換える。また、写像の定義域に対象となる IPv4 データグラムが存在しない場合、あらたな写像関係を動的に生成し、内部に蓄積する。ただし、あらたな写像関係は、プライベート・ネットワークからインターネットへの通信に対してのみ生成される。したがって、NAT を利用した通信の透過性は、通信がプライベート・ネットワーク内部で開始された場合のみ実現される。動的に生成された写像関係は、対応する通信の終了から一定時間が経過すると自動的に消去される。NAT を通じた通信のようすを図 5.30 に示す。

一方、利用者から指定された通信を仮想専用線に収容するためには、対象となる通信を識別す



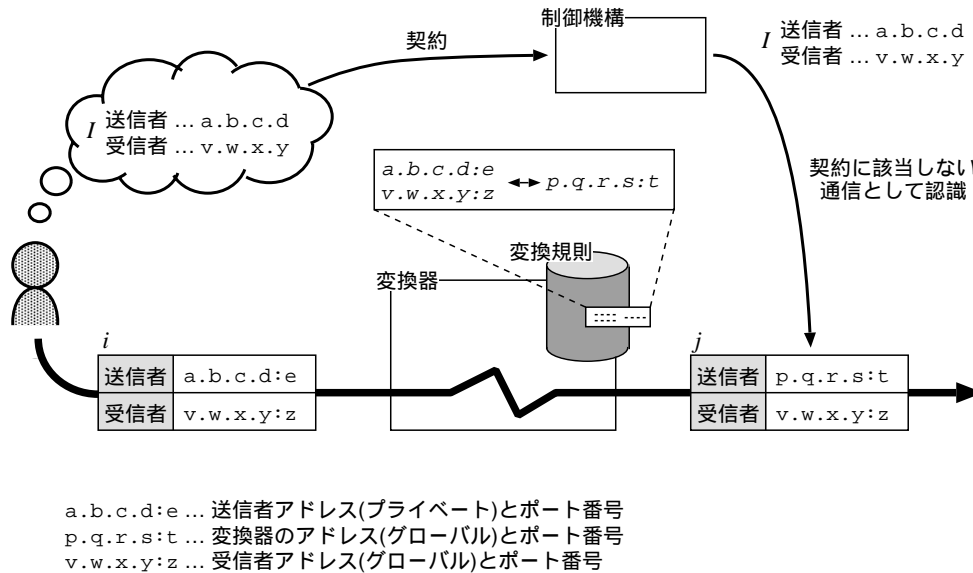


図 5.31: NAT 配下の利用者との契約

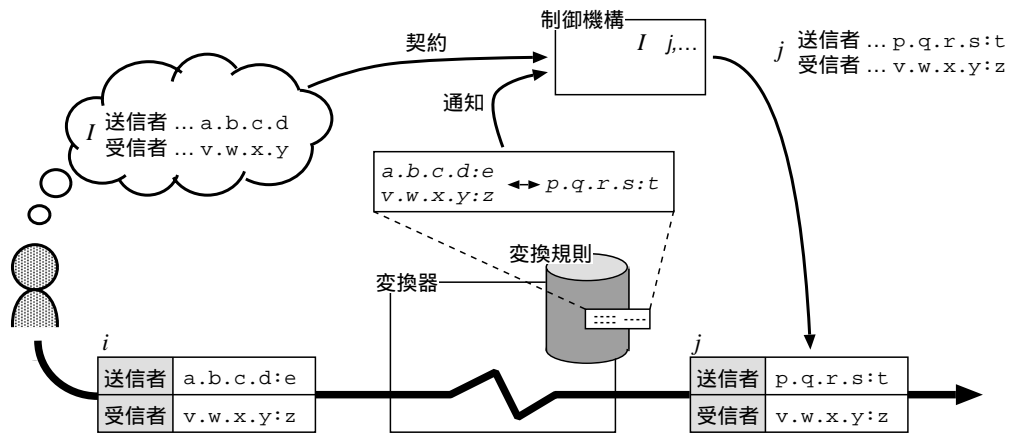
NAT を通した通信では、通信主体である利用者の視点からみた識別子と、変換器を経由したあとの識別子が異なる。したがって、利用者の申告した識別子だけでは通信が識別できず、処理を適用できない。

必要がある。通信の識別には、一般に送信者/受信者アドレスやトランスポート層プロトコル、およびそのポート番号など、IP データグラム中の通信主体を表すパラメータの集合を利用する。前述したように、NAT を利用する場合には、変換器で IPv4 アドレスや TCP/UDP ポート番号が変換される。これらの情報は通信を識別するためのパラメータとして利用されるため、通信主体間で同一の通信が変換器の前後で異なる通信として識別される。したがって、NAT が介在する環境は、通信の識別を前提とするサービスとの親和性が極めて低い。

#### 識別子の写像 $f$ の通知

図 5.31 は、NAT 配下のネットワークに收容されている利用者が、仮想専用線へのあらたな通信の收容を要求しているようすを表している。要求では、利用者の視点からみた通信の識別子  $I$  が指定されている。利用者は、これから開始する識別子  $i$  の通信が  $I$  で分類されることを期待している。それに対し、仮想専用線の入り口は NAT 変換器を通過したあとにあるため、通信の識別子は  $j$  に変更されている。したがってこの通信を  $I$  で識別できず、適切な処理を加えることができない。

変換器を経由したあとの通信を識別するためには、識別子  $j$  が必要である。一方、上で述べたように、制御機構は利用者からの要求で指定された識別子  $I$  を保持している。ここで、変換器が



a.b.c.d:e ... 送信者アドレス(プライベート)とポート番号  
 p.q.r.s:t ... 変換器のアドレス(グローバル)とポート番号  
 v.w.x.y:z ... 受信者アドレス(グローバル)とポート番号

図 5.32: 写像  $f$  の通知による識別

利用者から指定された識別子  $I$  と NAT 変換器の写像  $f$  から、変換後の識別子 (群) を導出している。これにより、変換器を経由したあとの通信の識別が可能となる。

内部に保持する通信識別子の写像を  $f$  とすると、識別子  $i$  と  $j$  は  $j = f(i)$  を満たす。したがって、制御層が写像  $f$  とその定義域  $X$  を知ることができれば、

$$j \in \{f(x) | x \in X, x \in I\}$$

であるから、変換後の通信を識別することができる。このようすを図 5.32 に示す。

そこで、本実験では制御パラメータ導出のための細目条件のひとつとして、NAT の写像  $f$  を取り入れた。NAT 変換器上で写像  $f$  の監視するサーバを構築し、変化を発見すると制御機構へ差分を送信する。制御機構では、写像  $f$  の差分を受信すると NAT 配下からの契約に対応するものがあるかを調べる。対応するものがある場合、識別子情報を変換して転送機構に反映する。また、NAT 配下のネットワークに収容された利用者から新規の通信収容要求を受けた場合、定義域  $X$  から対応する識別子をすべて選択し、変換した識別子群をもちいて転送機構を制御する。

今回、設計・実装した監視器を `natmapd` と呼ぶ。camp-net における NAT 変換器は `nat-ayame` と `nat-kuma` であるから、このふたつの変換器上で `natmapd` を動作させた。

#### 5.4.6. 仮想専用線利用の手順

利用者の仮想専用線の申請からその確立、契約の更新などの手順について述べる。概要を図 5.33 に示す。

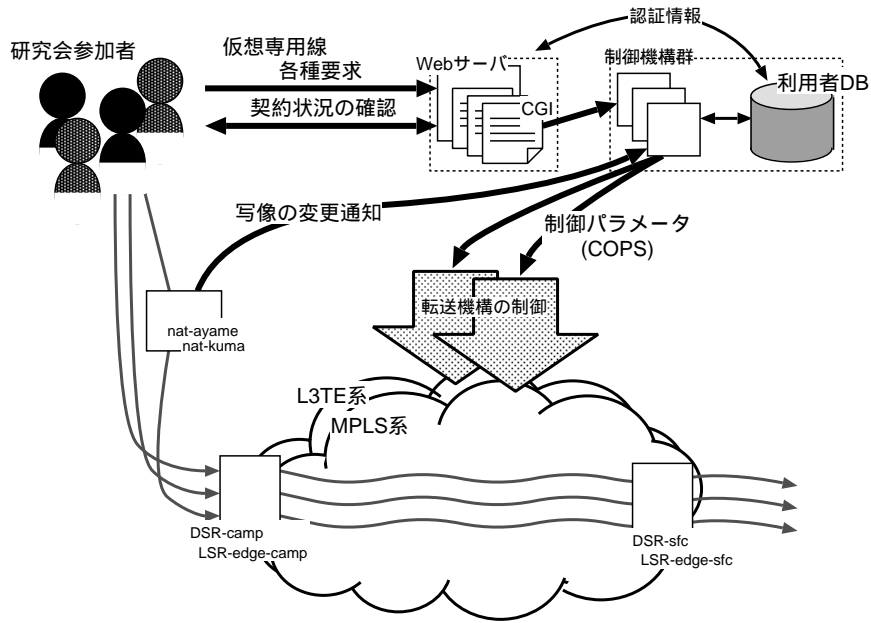


図 5.33: 仮想専用線利用の手順

利用者からの要求をうけた制御機構は、細目条件を担当するサーバの判断を統合して、アドミッション制御と制御パラメータの導出をおこなう。その結果に従い、適切な PEP に対して制御を適用する。

1. 利用者は適当な WWW ブラウザから camp-net の仮想専用線申請ページに接続する。ページを閲覧するためには、あらかじめ配布された利用者 ID とパスワードが要求される。認証はバックエンド DB の利用者情報を照会して処理される。
2. 図 5.24 にしたがって、仮想専用線の申請処理をおこなう。
3. 制御機構において、各細目条件の処理がおこなわれる。アドミッション制御としては、課金と余剰資源が検査される。また、制御パラメータの導出としては、明示的経路 (静的設定) と、natmapd からの情報を考慮したフィルタ規則の生成をおこなう。
4. 申請を許可できる場合 ...課金をおこなうと同時にネットワークに反映する。  
申請を拒否する場合 ...Web を通じてその旨を通知する。
5. natmapd より写像の更新通知を受信した場合、必要に応じてフィルタを生成し、ネットワークに反映する。
6. 契約状況を図 5.25 の Web を通じて利用者に提示する。

#### 5.4.7. 結果と考察

実験を通じて得られた結果を述べ、考察を加える。

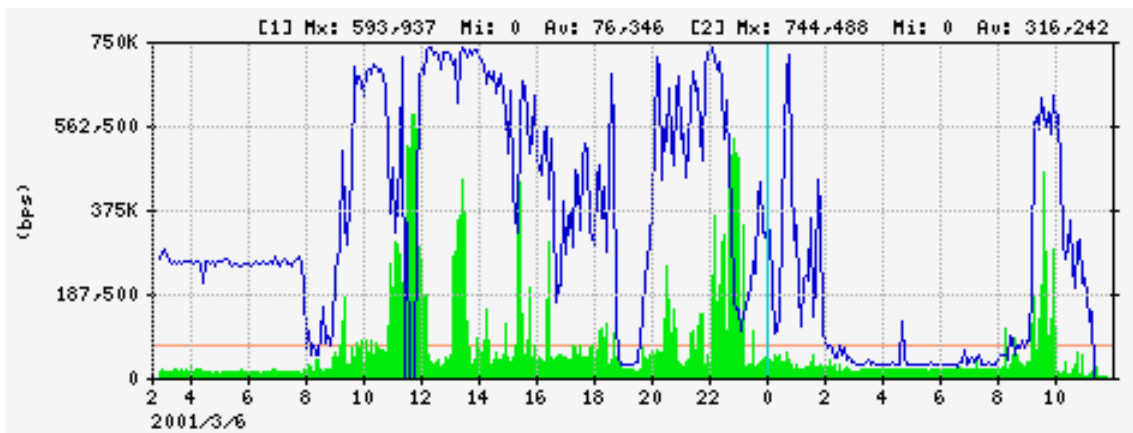


図 5.34: インターネットから camp-net へのトラフィック流量)

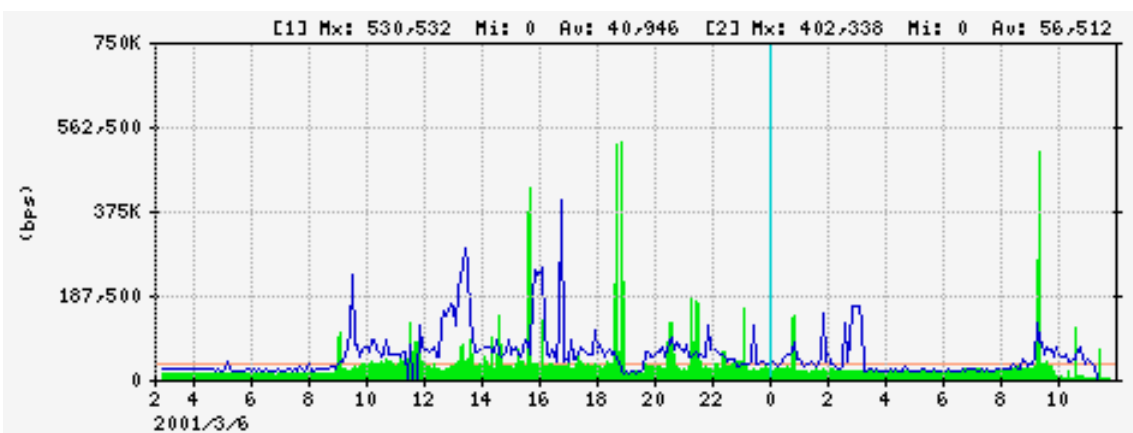


図 5.35: camp-net からインターネットへのトラフィック流量

濃い色の線は衛星回線のトラフィック流量を示し、灰色の部分が地上回線でのトラフィック流量を示している。

## 対外リンクのトラフィック

campnet への入出力トラフィックの時間毎の変化を示したグラフを図 5.34 および図 5.35 に示す。これらの図からデフォルト経路である衛星回線の混雑に応じて、地上回線のトラフィックが増加していることがわかる。これは利用者がネットワークの状況に応じて、仮想専用線を申請・利用していたことを意味している。このことから、優先制御の選択的利用に対する要求駆動型制御が、期待通りに動作していたことが確認できた。

## NAT 対応に関する考察

NAT の写像情報を制御機構に通知することで、変換器を越えて一貫した通信識別が可能になる。この手法の有用性は、本実験を通じて確認することができた。しかし同時に、いくつかの改善すべき点も明らかになった。

まず、NAT の利用では通信のたびにあたらしい写像関係が生成されるため、制御機構への通知頻度が高いことがわかった。現在の設計では、変換器が保持する写像の差分を、制御機構にとって必要であるかに関わらず通知する。しかし、差分の更新に関しては、継続中の契約に関連するものだけを通知すれば十分である。このため、無駄な通知が頻繁に発生してしまう。この問題を解決するためには、制御機構側から `natmapd` に対して必要な写像のフィルタを通知すればよい。同時に、新規の契約にも対応できるように、必要なものを要求に応じて取得できる必要がある。

次に、さまざまな変換器へ対応できる、柔軟な構造が必要である。`natmapd` の設計は、IPv4 アドレス/ポート番号変換に特化したものであった。しかし、インターネット上に介在する変換器は NAT だけではなく、IPv4/IPv6 変換器や SOCKS、Web 代理サーバなど、さまざまである。このような変換器では、識別情報を構成するパラメータは、IPv4 アドレスやポート番号などに限定されるものではない。したがって、写像情報の通知にあたり、さまざまな種類のパラメータに柔軟に対応できる構造をもつことが望ましい。また、制御機構側から事前に必要なパラメータを通知し、変換器が指定された情報のみの差分を通知する機構が必要であろう。

## 5.5. 実験のまとめ

インターネットが情報基盤として定着するにともない、要求に応じた通信品質制御サービスが求められるようになりつつある。このようなサービスを提供するためには、通信品質制御のための高性能化とサービス多様化のための高機能化という、いわば背反した要求を同時に実現していかなければならない。それに加えて、サービスの運用実績に乏しいため、サービスの実現に必要なコンポーネントや利用者の挙動、サービスと課金の関係など不明な点が多く残されており、実証実験を通じた経験の蓄積が必須であると言われていた。

本研究では、まず、制御層における要件を明らかにし、制御機構の設計と実装について論じた。要求駆動型のサービスでは、さまざまな意思決定や機器の設定が人間を介さずにおこなわれる。サービスは付加価値をつけるためにますます多様化していくと予測されることから、自動的に処理されるこれらの過程を柔軟かつ容易に構成できるような枠組みが必要である。そこで、意思決定や制御内容の導出に關与する条件を細分化し、その細目条件群を相互に組み合わせてサービス定義を構築する、制御機構の細目条件統合モデルを提案した。このモデルに基づく実装は、2001

年春の WIDE 研究会での要求駆動型通信品質制御サービス提供実験において、制御機構として導入された。この実装は機能を簡略化したものであったが、いくつかの細目条件を組み合わせた形で、人間を介さずに意思決定がなされていることを確認した。

次に、サービスを実現するために不足しているコンポーネントや、特性の異なるリンクを選択的に利用する場合の問題点、輻輳時における利用者の挙動などを明らかにするため、WIDE プロジェクト研究会の仮設ネットワークでの実証実験をおこなった。実証実験は3回の研究会にわたっておこなわれ、研究会の参加者にさまざまなサービスを提供した。いずれの研究会でも、ボトルネックが対外リンクのみのネットワーク構成において、そのリンク上での資源に関する要求駆動型制御をおこなっている。ネットワークや利用者の構成に一部特殊な部分もあるが、多くの利用者を収容する実運用ネットワークとしての役目を果たしており、その意味で実環境に近い現象を観測できたと考えられる。

3回の実験を通じて、利用者は輻輳時の品質制御サービスを高く評価し、実際にこのようなサービスが必要とされていることがわかった。また、サービス実現に際して、以下のような点を明らかにした。

- AF によるサービスは、非輻輳時のときにはパラメータの値による違いが明確に現れない。そのため、AF による利用者の制御や確保帯域に基づく課金は困難である。
- 複数の極度に異なる特性を持つサービス・クラスを提供する場合、価格設定による利用者の制御をおこなうためには、課金の対象となる特性の選択が重要となる。

ADSL や FTTH など加入者線の高性能化にともない、インターネットはますます高機能化を求められている。要求駆動型の通信品質制御も、そのサービスが期待されるようになった。このような状況において、契約の締結からネットワーク資源の管理、さらには、転送機構への反映に至る一貫したシステムを構築し、実利用者が接続するネットワークにおいて細粒度の通信品質制御サービスの可能性と有効性の実証と問題点の洗い出しをおこなったことは、大きな成果であろう。

MPLS では、特定の通信の転送路としてあらかじめ LSP と呼ばれるパスを確立する。そのため、MPLS を要求駆動型通信品質制御サービスにおける制御層として利用する場合、要求許諾時の LSP の確立や、障害時における LSP の再確立など、LSP 確立に要する時間を短縮することが重要となる。本章では、あたらしいラベル配布の仕組みを提案し、それによる LSP 確立時間の高速化について述べる。

## 6.1. LSP 確立の高速化の重要性

MPLS 網内において MPLS による転送を利用するためには、なんらかのラベル配布方式を利用して網内にラベル情報を浸透させ、LSP を確立する必要がある。すでに見てきたように、特性の異なる複数の経路やリンクの選択的な利用は、通信品質制御の実現という観点から非常に重要である。したがって、MPLS を利用した通信品質の制御を考える場合、LSP を構成する LSR とその経路が既存の枠組みから独立して指定できるという特徴を、どのように利用していくかが大きな要点となる。

このような MPLS の特色を効果的に利用するため、CR-LDP と RSVP-TE [25] が標準化された。CR-LDP は、4.2.6 節で説明したように、明示的に指定した経路に沿って LSP を確立するためのラベル配布方式である。また、RSVP-TE も同様の機能を持つ。これらのプロトコルは、流入境界 LSR において明示的な経路を指定することで LSP を確立する。この際、流入境界 LSR から流出境界 LSR までの間を、指定された経路に沿って制御情報が往復する。このため、制御対象となる領域の規模に比例して LSP 確立に要する時間が増大する。

このような LSP 確立の遅延は、利用者との間で締結される SLA を履行する上で問題となる場合がある。たとえば、2.4 節で述べたサービス開始遅延に関する SLA を例としてあげることができるだろう。LSP 確立に要する時間は、制御要求が発行されてから実際に制御が反映されるまでの時間を増大させる。人間が介在する従来の契約書に基づく制御要求では、その時間的粒度が LSP 確立に関わる遅延を十分無視できる程度に粗いため、問題となることはない。しかし、要求駆動型通信品質制御サービスの提供を考えると、通信の発生と同時に制御要求が発行されることも予想され、このような場合には LSP 確立の遅延によって SLA が履行できない可能性がある。

また、この遅延は可用性や障害復旧時間に関する SLA にも影響がある。MPLS では LSP 上で障害が発生した場合、その影響を回避してサービスを継続するためには、べつの経路で LSP を再確立しなければならない。CR-LDP や RSVP-TE によって明示的な経路で LSP を確立した場合は、流入境界 LSR に対してあらたな経路を明示的に指定して再確立するため、制御情報が流入境界 LSR と流出境界 LSR 間の往復する時間が、サービスの中断時間に大きく影響することになる。したがって、LSP 確立に要する時間によっては、可用性や障害復旧時間に関する SLA を履行することができなくなる。

このように、明示的経路に沿った LSP 確立を高速化することは、契約に基づくサービス品質を提供するために重要である。

## 6.2. 要求駆動と事前予測

明示的経路に沿った LSP の確立には、要求駆動と事前予測の 2 種類の戦略がある。要求駆動は、LSP 確立を要求する事象が発生してから実際の確立作業を開始する戦略である。たとえば、5.4 節で報告した 2001 年春の研究会におけるサービスでは、利用者からの要求によって LSP の確立を開始した。また、障害発生を認識してから LSP を再確立する方式も、要求駆動戦略のひとつである。要求駆動戦略による障害復旧を Reroute と呼ぶ。前節で述べたように、要求駆動戦略に CR-LDP または RSVP-TE を利用する場合、LSP が要求されてから確立が完了するまでの遅延が問題となる。

一方、事前予測は将来的に必要となるであろう LSP を予測し、あらかじめ確立しておく戦略である。たとえば、明示的経路によって LSP を確立する際に、障害に備えてあらかじめ代替用の LSP を確立しておく方式は、事前予測戦略の一例である。事前予測戦略に基づく障害復旧方式を Protection と呼ぶ。

Protection には、事前に準備された代替 LSP への切り替え方式によって、“1+1” と “1:1” の 2 種類の方式がある。Protection 1+1 は、要求により確立した主 LSP を通じて転送する全てのデータを複製し、代替 LSP へも送信するものである (図 6.1)。このため、どちらか一方の LSP に障害が発生した場合でも、流出境界 LSR では主/代替いずれかの LSP からデータを受信することができる。したがって、障害による通信の中断は理論上 0 になることが期待される。しかし、主/代替の両 LSP に同一のトラヒックを収容するため、そのトラヒックの消費する帯域が 2 倍になる。また、すべての LSP に対して Protection 1+1 で代替経路を用意する場合、MPLS 網で用意しなければならない帯域は、利用者に対して実質的に提供する帯域の 2 倍になる。また、あらかじめ代替 LSP を確立するために、ラベル空間の一部が消費されることにも注意する必要がある。



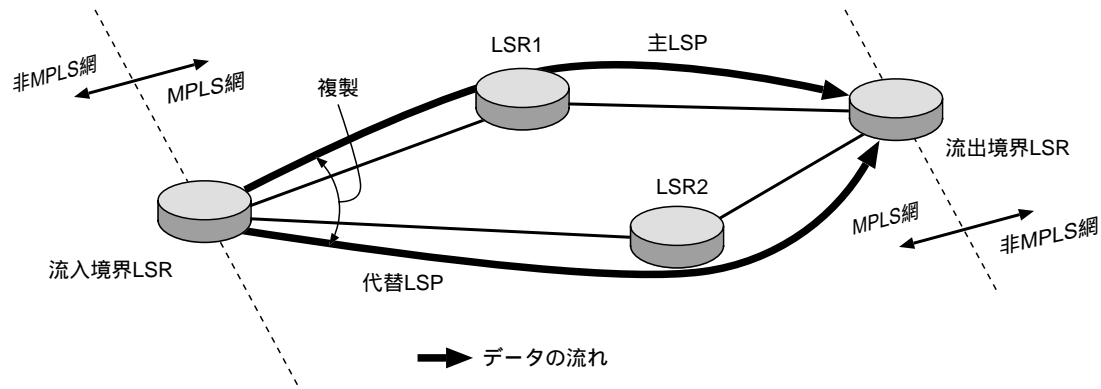


図 6.1: 事前予測戦略に基づく障害復旧 (Protection 1+1)

事前予測戦略に基づく障害復旧方式のうち、Protection 1+1 を示す。この方式では、主 LSP を通じて転送される全てのデータを複製し、代替 LSP にも送信する。これにより、どちらか一方の LSP に障害が発生した場合でも、流出境界 LSR では主/代替いずれかの LSP からデータを受信できるため、障害によるサービス停止時間は理論上 0 である。

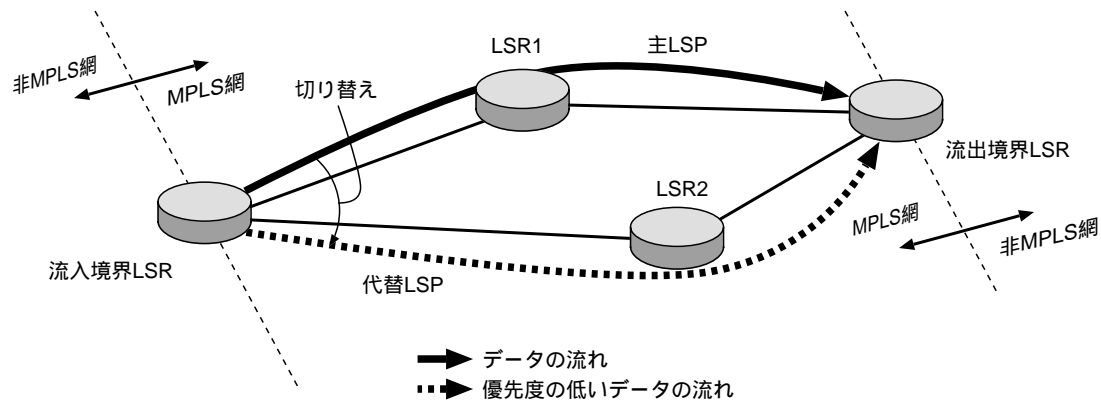


図 6.2: 事前予測戦略に基づく障害復旧 (Protection 1:1)

事前予測戦略に基づく障害復旧方式のうち、Protection 1:1 を示す。この方式では、代替 LSP を優先度の低いトラフィックの転送に利用することで、Protection 1+1 と比較して帯域の利用効率を向上させている。主 LSP に障害が発生した場合には、代替 LSP の優先度の低いトラフィックを廃棄し、主 LSP のかわりとして利用する。

これに対し、Protection 1:1 は代替 LSP に優先度の低いトラフィックを収容する。障害時には代替 LSP への低優先度トラフィックを全て廃棄し、主 LSP から切り替えて利用する (図 6.2)。通常時は代替 LSP を他のトラフィックの転送に利用するため、Protection 1+1 よりは帯域利用効率が向上している。しかし、障害の発生から発見まで、および、切り替えの動作中は、通信が中断する。特に切り替えの時間は、障害を発見した LSR から LSP に沿って障害情報が伝搬し、流入境界 LSR に到達してから切り替え動作が実行されるため、障害発生位置と流入境界 LSR の距離に依存して差を生じる。また、Protection 1+1 の場合と同様にあらかじめ代替 LSP を確立するため、ラベル空間の一部が消費される。

このように、MPLS 網内の資源を十分に取れる場合は、障害復旧の手法に事前予測戦略を採用することは選択肢の一つとして考えられる。しかし、MPLS 網内の帯域が十分でない場合には、Protection 1+1 は利用できない。また、要求駆動型通信品質制御サービスのように、品質制御要求の発行と取消しの間隔が細粒度でおこなわれる場合、ラベルを余分に消費することはラベル空間の枯渇につながる恐れがある。

要求駆動型のサービスにおける新規要求に対しては、利用者からの要求が発生すると予測される区間に、あらかじめ LSP を確立することになる。しかし、利用者の増加や ISP 間の相互接続点の増加にともない、考えうる流入/流出境界 LSR の組み合わせは急激に増加する。また、要求発行の間隔が細粒度で発生する場合には、予測が困難であること、予測と利用の管理が複雑になること、ラベルの浪費による枯渇が懸念されることなどから、事前予測戦略は現実的であるとは云えない。

以上のことから、本研究が対象とする通信品質制御サービスにおける LSP 確立には、要求駆動戦略を採用するべきであると云える。

### 6.3. 制約経路独立ラベル配布プロトコル (CR-iLDP)

前節で、事前予測戦略による LSP の確立は、要求駆動型通信品質制御サービスに適合しないことを述べた。そこで、要求駆動戦略による LSP 確立の高速化として、CR-LDP の拡張による LSP 確立遅延の改善を図った。

6.1 節で述べたように、CR-LDP を利用した場合の LSP 確立遅延は、流入/流出境界 LSR 間を制御情報が往復するために要する時間が大半を占める。この制御情報の往復は、本質的に次のふたつの理由に起因する。

#### 明示経路の告知方法 (往路)

CR-LDP では、ER-TLV を流入境界 LSR から明示経路に沿って流出境界 LSR ま

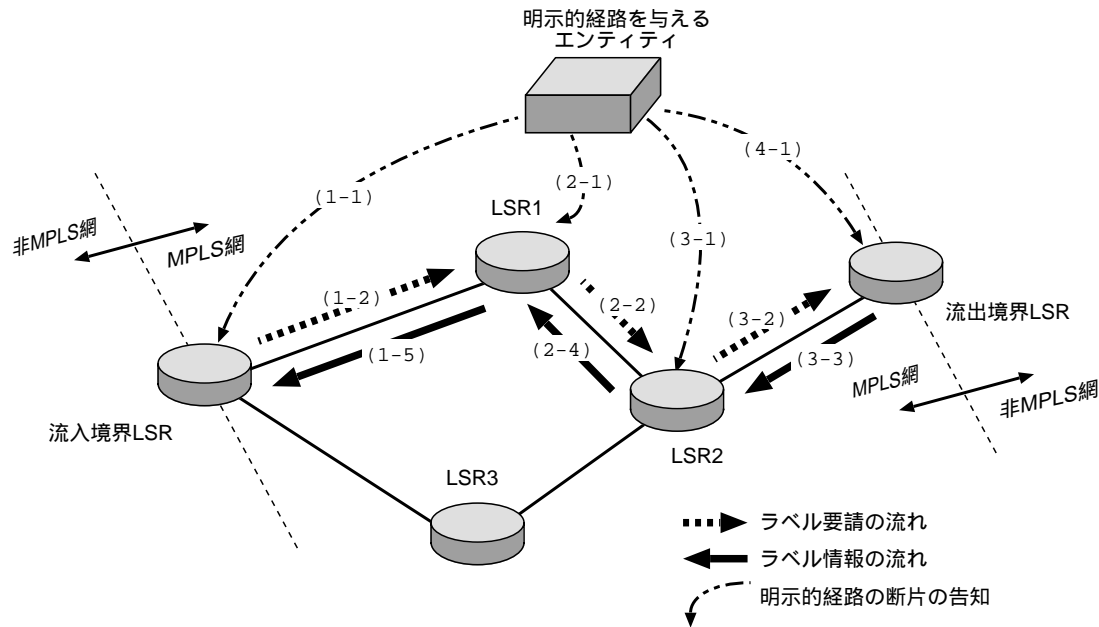


図 6.3: CR-LDP による LSP 確立の高速化 (往路)

明示的経路を与える外部のエンティティから、経路上の LSR に経路の断片を同時に告知することによって、往路に要する時間を短縮する。図中の表記  $(x-y)$  において、 $x$  は一連の情報交換を、 $y$  は全体での動作の順番を表す。 $(*-0)$  および  $(*-1)$  は経路の断片の告知とそれに対するラベル要請であり、それぞれは同時におこなわれる。一方、 $(3-3)$ 、 $(2-4)$ 、 $(1-5)$  は、順序ラベル配布制御にしたがうラベル配布である。

で伝達することで、経路上の LSR に対して明示経路を告知する。

#### 順序ラベル配布制御の利用 (復路)

4.2.6 節で述べたように、明示的経路を指定する LSP や LSR 上で特殊な処理を加える LSP を確立する場合には、LSP が流入境界 LSR から流出境界 LSR までがとぎれることなく連続であることを保証するため、順序ラベル配布制御を利用する必要がある。

往路の遅延に関しては、明示経路の告知を経路に沿って順番におこなうことによって発生する。ところで、CR-LDP による LSP 確立では、流入境界 LSR に明示的経路を指定する枠組みは決っていない。LDP の外部から明示的経路を与えるエンティティの存在を仮定すると、このエンティティは経路上の全ての LSR を認識していることから、経路上の全ての LSR に対して明示的経路を同時に直接告知することができる。これにより明示的経路が往路を伝達していくのに要する時間の短縮が期待できる。図 6.3 に、エンティティからの明示的経路の告知とラベル配布のようすを示す。

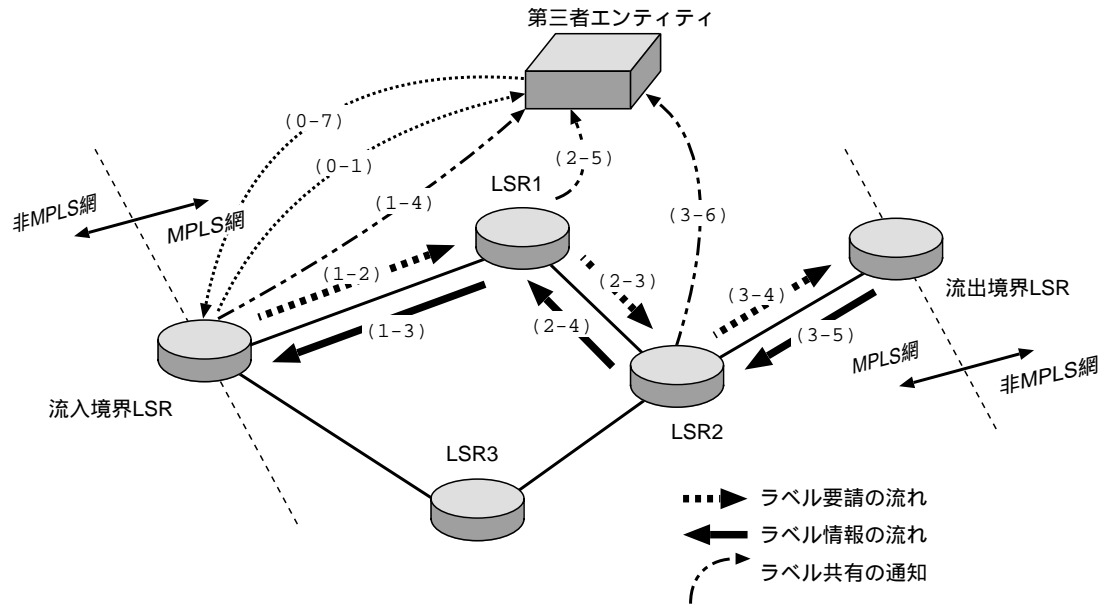


図 6.4: CR-LDP による LSP 確立の高速化 (復路)

MPLS 網から見て第三者のエンティティを仮定し、そのエンティティに対してラベル共有の完了を通知することで、LSP の完全性を保証し、往路の時間を短縮する。図中の表記  $(x-y)$  は図 6.3 と同様である。(1-2)、(2-3)、(3-4) は明示的経路の告知とラベル要請を表す。それぞれの LDP は独立ラベル配布制御で動作するため、要請の到着後、すぐにラベルの割り当てと配布をおこなっている。また、ラベルを要請した LSR は、ラベルの共有とともに第三者エンティティに通知する。なお、第三者エンティティが LSP の完全性を検証するために、流入境界 LSR から (0-1) で明示的経路を通知している。また、(0-7) で流入境界 LSR に LSP の完全性を通知し、転送の開始を促す。

なお、明示的経路  $(LSR_1, LSR_2, \dots, LSR_n)$  に対し、ある  $LSR_i$  に告知される内容はその全てである必要はなく、 $(LSR_{i-1}, LSR_i, LSR_{i+1})$  が含まれていればよい。このうち、 $LSR_{i+1}$  は経路上の次ホップを示している。また、 $LSR_{i-1}$  は上流からのラベル要請に対し、返答すべきかを判断するための要素として利用される。

往路の遅延を短縮する方法としては、明示的経路を流出境界 LSR から流入境界 LSR へ伝達する方法も考えられる。しかし、経路上における全ての LSR の隣接関係が UD で動作している必要があるため、ここでは除外している。

一方、復路が経路に沿って順番にラベル配布をおこなわなければならないのは、順序ラベル配布制御を利用する必要があるためである。この制約は CR-LDP によって確立された LSP の連続性保証のためであるから、なんらかの方法によって LSP の連続性を保証できれば、CR-LDP を独立ラベル配布制御で動作させることができる。独立ラベル配布制御は個々の LSR が任意の時期に

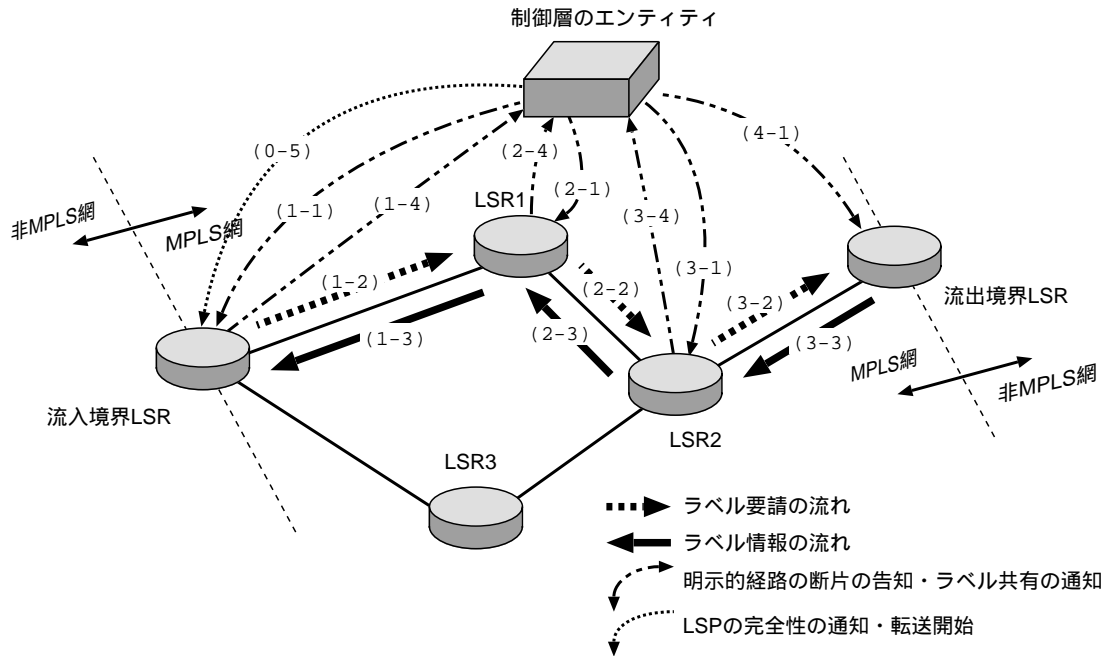


図 6.5: CR-iLDP による LSP 確立のようす

CR-iLDP による LSP 確立のようすを示す。明示的経路配布サーバから経路断片を通知された LSR は、独立してラベル情報を生成し、上流からのラベル要請に返答する。同時に、下流となる LSR にラベル要請を発行する。ラベル共有が完了すると制御層エンティティに通知する。制御層エンティティは LSP 確立の完了を確認すると、流入境界 LSR に転送開始を促す。

ラベルの割り当てをおこなうことができるため、ラベルが復路を伝達していくのに要する時間の短縮が期待できる。ここでは、MPLS 網から見て第三者のエンティティの存在を仮定し、そのエンティティにラベル共有を通知することにより、全体として LSP が確立できていることを保証することとした。図 6.4 に、第三者エンティティの存在を仮定した復路の高速化のようすを示す。

上記の往路と復路の高速化は併用することができる。併用による LSP 確立のようすを図 6.5 に示す。この図では、明示的経路の計算および配布をおこなうエンティティと、LSP の完全性を保証する第三者エンティティを同一のものと仮定し、明示的経路配布サーバとしている。この仮定は、エンティティを制御層でアドミッション制御等を司るサーバであるとすると、サービス提供の形態やその運用を考慮しても無理のないものである。図 6.5 のラベル配布方式を、制約経路独立ラベル配布プロトコル (Constraint-based Routed-independent Label Distribution Protocol; CR-iLDP)<sup>1</sup>と呼ぶ。上記のように、CR-iLDP には経路断片の配布とラベル共有の通知を実現するための機構が不可欠である。そのため、CR-iLDP 単独では LSP を確立することができない。

<sup>1</sup>Rapid-CR-LDP と呼ばれることもある。

以下に、CR-iLDP による明示的経路 ( $LSR_1, LSR_2, \dots, LSR_n$ ) に沿った LSP 確立の過程を示す。

#### 1. 明示的経路配布サーバから経路断片を配布

明示的経路配布サーバは、確立する LSP の明示的経路から、それぞれの LSR へ告知すべき経路の断片を導出する。LSR<sub>*i*</sub> に告知する経路断片は ( $LSR_{i-1}, LSR_i, LSR_{i+1}$ ) となる。ただし、LSR<sub>1</sub> および LSR<sub>*n*</sub> に対しては、それぞれ上流および下流の LSR の指定が省略される。また、FEC としては明示的経路配布サーバによって生成された LSP 識別子が指定される。

#### 2. ラベル情報の生成とラベル要請の発行

経路断片を受け取った LSR<sub>*i*</sub> は、上流に配布すべきラベル情報を生成する。また、LSR<sub>*i+1*</sub> に対し、LSP 識別子を FEC としてラベル要請を発行する。

#### 3. ラベル情報の配布

ラベル要請を受けた LSR<sub>*i*</sub> は、ラベル要請を発行した LSR が LSR<sub>*i-1*</sub> であることを確認し、ラベル情報を配布する。

#### 4. ラベル共有の通知

LSR<sub>*i+1*</sub> とラベル情報を共有した LSR<sub>*i*</sub> は、明示的経路配布サーバに対してラベルを共有した旨を通知する。

#### 5. LSP 確立の通知

それぞれの LSR からの返答によって LSP が確立したことを確認した明示的経路配布サーバは、流入境界 LSR に対して LSP 確立を通知し、転送の開始を促す。

## 6.4. 数値解析

実装による評価の前に、簡単な数値解析によって LSP 確立に要する時間の最大値を見積もる。以下では、( $LSR_1, LSR_2, \dots, LSR_n$ ) で明示的に指定される LSP を構成するものとする。また、LSR<sub>*i*</sub>-LSR<sub>*i+1*</sub> 間の伝送遅延を  $LLD_{(i,i+1)}$  CR-iLDP における明示的経路配布サーバと LSR<sub>*i*</sub> の間の伝送遅延を  $SLD_i$  とする。なお、各 LSR 内におけるラベル生成などの処理時間は、各種の伝送遅延に比べて無視できるほど小さいため、ここでは考慮していない。

#### CR-LDP での LSP 確立時間

すでに述べたように、CR-LDP では流入境界 LSR と流出境界 LSR の間を、ER-TLV をともなったラベル要請とラベル情報が往復する。したがって、LSP の確立に

要する時間  $E_{\text{CR-LDP}}$  は以下ようになる。

$$E_{\text{CR-LDP}} = 2 \sum_{i=1}^{n-1} LLD_{(i,i+1)}$$

### CR-iLDP での LSP 確立時間

CR-iLDP では、LSP の断片の告知とラベル要請/ラベル情報の配布、および、ラベル共有の通知処理が、LSR ごとに独立して動作する。 $LSR_i$  に対する一連の動作に要する時間  $LD_i$  は、図 6.5 より

$$LD_i = 2SLD_i + 2LLD_{(i,i+1)}$$

となる。明示的経路配布サーバは、全ての LSR からのラベル共有通知によって LSR の完全性を確認してから、流入境界 LSR に転送の開始を促す。したがって、LSP の確立に要する時間  $E_{\text{CR-iLDP}}$  は以下ようになる。

$$\begin{aligned} E_{\text{CR-iLDP}} &= \max_i(LD_i) + SLD_1 \\ &= \max_i(2SLD_i + 2LLD_{(i,i+1)}) + SLD_1 \end{aligned}$$

すべての伝送遅延が  $d$  であるとき、すなわち、 $LLD_{(i,i+1)} = SLD_i = d$  とした場合、LSP 確立に要する時間は、上記より CR-LDP では  $2(n-1)d$ 、CR-iLDP では  $5d$  となる。したがって、このようなネットワークにおいては、4 つ以上の LSR から構成される LSP を確立する場合に CR-iLDP のほうが高速となる。

## 6.5. 実装と評価

上記の CR-iLDP のプロトタイプを実装し、実験を通じて評価をおこなった。

### 6.5.1. 設計と実装

6.3 節で述べたように、CR-iLDP は単独では LSP を確立できない。そこで、評価のために簡単な経路断片通知プロコルを定義した。このプロコルは、明示的経路配布サーバと LSR との間で経路断片等の情報交換に利用するものである。

図 6.6 にプロコルの基本フォーマットを示す。16 ビットのヘッダを含むデータ長のあとに 8 ビットの種類と補助種類が続く。種類と補助種類は、その組み合わせによってさまざまな操作内容やデータの意味を表すことができる。プロトタイプ実装で定義した種類と補助種類を表 6.1 に示す。また、それぞれの意味を以下に示す。

データ長	種類	補助種類
値		

図 6.6: 経路断片通知プロトコルの基本フォーマット

経路断片通知プロトコルの基本フォーマットを示す。ヘッダを含むデータ長に続き、データの種類と補助種類を指定する。値には種類/補助種類に依存したデータが挿入される。たとえば、Keep Alive の場合には値はない。

表 6.1: 経路断片通知プロトコルにおけるデータの種類と補助種類

種類		補助種類	
値	利用目的	値	利用目的
1	Keep Alive	0	—
3	LSP 操作	1	LSP 確立・経路断片通知
		2	LSP 解放

経路断片通知プロトコルの種類と補助種類を示す。種類の値によって補助種類の意味も変化する。Keep Alive は補助種類を利用せず、常に 0 である。

16(=データ長)	3(=種類)	1(=補助種類)
0	LSP 識別子	
上流 LSR の識別子		
下流 LSR の識別子		

図 6.7: 経路断片通知プロトコルの経路断片通知フォーマット

経路断片通知に利用されるデータ・フォーマットを示す。16 ビットの LSP 識別子と 32 ビットの上流/下流 LSR の識別子が含まれている。なお、プロトタイプ実装では上流 LSR の識別子は無視された。



### Keep Alive

LSR-サーバ間の接続性を確認するために、10 秒間隔で送信する。接続性が持続している間は、その LSR をサーバの管理するリストに保持する。また、続けて 40 秒間受信しなかった場合にはリストから削除する。Keep Alive はヘッダのみの交換になる。

### LSP 操作

LSP の確立や解放に利用する。操作内容は補助種類によって指定する。確立の際には LSP 識別子や経路断片情報を含む。図 6.7 に、経路断片情報の配布に利用するフォーマットを示す。

実装は新規に設計・実装した部分と既存実装の拡張から構成される。新規部分は明示的経路配布サーバであり、BSD 上のユーザ空間で動作するサーバとして設計・実装した。ここでは、LSR からの接続の待ち受け、接続 LSR の管理、明示的経路の断片情報の生成と配布などをおこなう。

既存実装の拡張は、あやめプロジェクト [36] によって NetBSD [37] 上で開発されている MPLS 実装、AYAME [33] を対象としている。ここでは、AYAME の LDP を拡張し、サーバに接続するための機能、Keep Alive によって接続を持続する機能、および、経路断片情報による LSP 確立のための機能を追加した。また、CR-iLDP による LSP 確立の際には、独立ラベル配布制御を利用するようにした。

## 6.5.2. 実験と結果

6.4 節で、CR-iLDP による LSP 確立に要する時間が LSP を構成する LSR の数に大きく影響されないことを示した。この正当性を検証するため、上記のプロトタイプ実装を利用して実験をおこなった。

実験に利用した機器構成を図 6.8 に示す。この構成で LSP 確立に要する時間の変化を計測した。LSP の確立には、明示的経路配布サーバに対して LSP 確立を要求する。計測する時間は、要求を発行した時点から LSP が確立するまでとした。実験では、確立する LSP を構成する LSR の数を 2 台から 7 台に変化させ、それぞれの台数に対して 100 回の試行をおこなった。この実験に利用した機器の仕様を表 6.2 に示す。

実験結果を図 6.9 に示す。この結果から、LSP を構成する LSR の数が増加しても、LSP 確立に要する時間は大きく影響を受けないことがわかる。6.4 節で述べたように、CR-iLDP による LSP 確立の時間は LSR-LSR 間、および、サーバ-LSR 間の伝送遅延に依存する。前者の伝送遅延は同一リンク上の LSR 間の遅延であるから、網の構成を変更しても大きな変化はない。一方、後者は

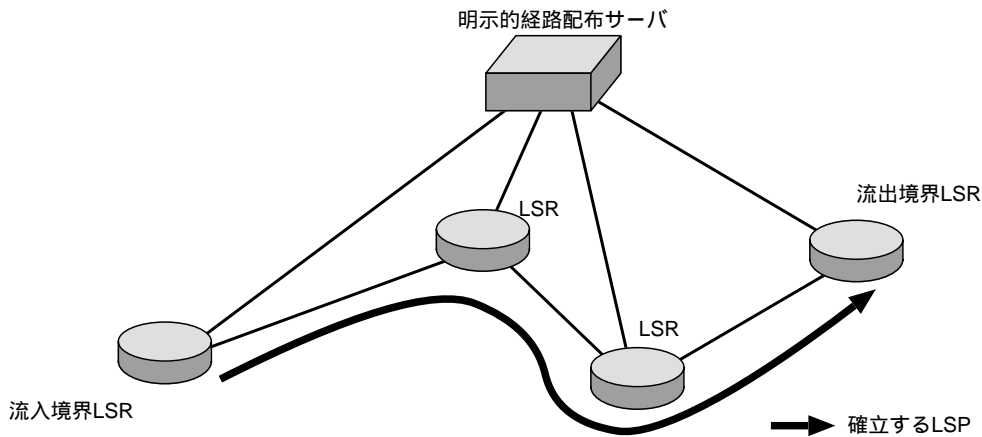


図 6.8: 実験環境の構成

実験環境は 1 台の明示的経路配布サーバと LSR から構成した。2 台から 7 台まで変化させた。また、サーバと LSR は直接通信することができる。

表 6.2: 実験機材の仕様

モデル	IBM PC/AT 互換機
CPU	Intel Pentium III 866MHz
メモリ	512MB
ネットワーク	Intel i82557 Ethernet 100baseT
OS	NetBSD 1.5.2 (AYAME によって MPLS に対応したもの)

明らかにサーバを設置する位置に大きく影響を受ける。したがって、運用において実験結果と同様の効果を得るためには、サーバから MPLS 網内のすべての LSR への伝送遅延の最大値が最小になるような位置に、サーバを設置しなければならない。

## 6.6. 障害復旧への適用に関する検討

ここでは、CR-iLDP の障害復旧へ適用に関して検討する。

### 6.6.1. 事前予測戦略との融合

6.2 節で述べたように、事前予測戦略はあらかじめ必要とされる LSP を予測し、前もって確立しておく戦略である。この事前予測戦略は、障害の仮復旧に取り入れることはそれほど困難ではない。新規に LSP を確立する場合と異なり、障害復旧では事前予測の対象となる LSP の流入/流出境界 LSR が固定される上、主 LSP と異なる LSR を経由する必要があることから、ある程度の

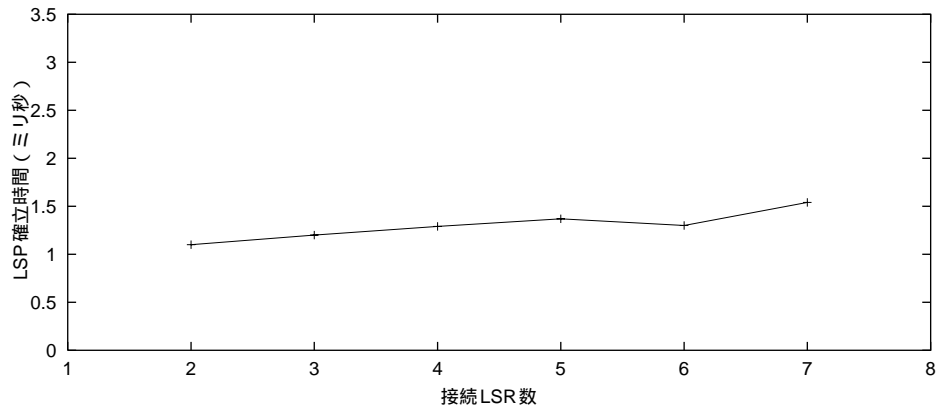


図 6.9: CR-iLDP の実験結果

実験結果を示す。結果は 100 回の計測の最大値を表している。このように、CR-iLDP を利用した場合は、LSP を構成する LSR 数が増加しても LSP の確立に要する時間はあまり増加しない。

制約によって事前の LSP を決定することができるからである。同節で述べた Protection は、事前予測戦略に基づく障害復旧方式であり、障害時にトラヒックを代替 LSR に切り替えることによって高速な復旧を実現している。しかし、Protection はリンク帯域やラベル空間など、MPLS 網の資源を大量に消費することも述べた。

CR-iLDP を障害復旧に利用する場合にも、上記と同様の理由により、代替 LSP の明示的経路をあらかじめ算出することが可能である。したがって、明示的経路配布サーバで代替経路を保持しておくことにより、障害通知の受信に対して即座に代替 LSP の確立を開始できる。また、代替 LSP を事前に確立しないことから、MPLS 網の資源を消費することもない。

ただし、要求駆動型通信品質制御サービスのように、細粒度での LSP 確立要求が発生する場合には、経路計算にかかるコストに留意する必要がある。事前予測を行う場合、要求を受信するとともに主 LSP/代替 LSP の 2 経路を算出する必要があり、それに要する時間も 2 倍近くになることが予測される。このため、要求の到着間隔よりも経路算出に要する時間のほうが長くなる可能性がある。したがって、高負荷の場合に SLA の内容も加味しながら代替 LSP の計算を遅延させたり、経路算出のために特殊なサーバを設置するなどの処置が必要であろう。

### 6.6.2. 障害通知

Protection や Reroute の場合、障害の発生は LSP に沿って流入境界 LSR まで伝達される。したがって、障害復旧が開始されるまでの時間は、LSP のどの位置において障害が発生するかに依存する。特に、流出境界 LSR 近傍において障害が発生した場合には、流入境界 LSR までの伝達

遅延が問題となる。

一方、CR-iLDP の場合、代替 LSR の確立は明示的経路配布サーバによって開始されることから、明示的経路配布サーバに対して障害が通知されればよい。このため、障害発見から代替 LSP の確立開始までの時間は、障害の発生位置にあまり依存せずに予測できる可能性がある。これを実現するためには、サーバの設置位置や網構成を適切に選択しなければならないだろう。したがって、CR-iLDP を高速障害復旧に適用するためにはこれらの項目の適切な選択手法に関して研究を進める必要がある。

制御層からの制御要求が反映されると、転送層の機器群では流入するデータから制御対象となるデータを識別し、対応する処理を決定して適用する。このとき、制御層から転送層の機器群に発行された制御要求がすべて同一であったとしても、それぞれの機器が識別にもちいる手法によって、導出される結果の異なる可能性がある。このような場合、通信品質制御サービスの実現に不可欠な、通信経路上における処理の一貫性を保証できない。本章では、このような曖昧性の発生原因箇所を特定するために、サービス提供者の定義する条件を具現化する部分を通信の識別行程から分離する、パラメータ・フィルタの2段階分離モデルについて述べる。

## 7.1. パラメータ・フィルタ

インターネットで提供されるサービスの多様化にしたがって、通信の識別・分類する機構を必要とする場面が多く見られるようになった。このような機構は、従来よりIPデータグラムの転送処理で利用されている。インターネット層では、流入するIPデータグラムの受信者アドレスを検査し、転送の際に経路表から適切な転送先の検索をおこなう。近年、多くの組織で導入が進んでいるファイアウォールでも、通信の識別・分類が必要とされている。ファイアウォールは流入・流出するIPデータグラムを検査し、識別・分類することによって通過の可否を決定する。

通信品質制御サービスもまた、このような通信の識別と分類を必要とするサービスのひとつである。特に、配送層では制御の対象となる通信を識別し、IPデータグラムを関連づけられた規則にしたがって転送しなければならない。このような、通信の識別・分類とそれに続く一連の処理をおこなう機構を、“パラメータ・フィルタ”と呼ぶことにする。

多くの場面では、パラメータ・フィルタには複数の規則が定義される。一般的に、規則は互いに独立して定義されるため、ひとつのIPデータグラムが複数の規則によって識別された場合、そこから導出される処理に矛盾が生じる可能性がある。たとえば、組織における透過Webキャッシュの導入と、特定の利用者に対する専用線の割り振りを考えてみる。このようなネットワークは、以下の規則によって制御される。

- (1) TCPの受信者ポートが80番であるものは、透過Webキャッシュへ転送
- (2) 送信者アドレスがa.b.c.dであるものは、専用線へ転送

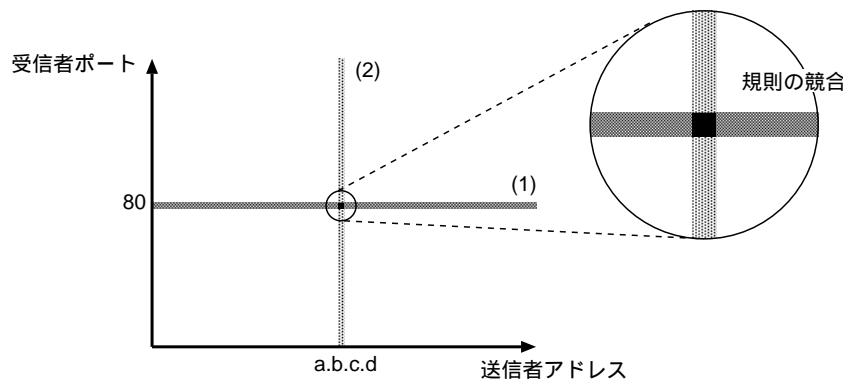


図 7.1: 規則の競合

(1) と (2) はそれぞれ、「TCP の受信者ポートが 80 番 (HTTP) であるものを、透過キャッシュサーバへ転送する」および、「送信者アドレスが a.b.c.d であるものを、専用線へ転送する」という規則で識別される通信の集合を表している。この 2 本の直線の交点は、(送信者アドレス, TCP の受信者ポート) が (a.b.c.d, 80) であるような通信を表しており、ふたつの規則が競合する。

図 7.1 は、それぞれの規則が適用される通信の集合を図示したものであり、水平の直線と垂直の直線は規則 (1) と規則 (2) が適用される通信の集合を表している。この中で、送信者アドレスが a.b.c.d であり、TCP の受信者ポートが 80 番であるような通信の集合は、2 本の直線の交点で表される。このことは、この通信に対しては規則 (1) と (2) がともに適用される可能性を持つことを示す。しかし、これらの規則は互いに矛盾するため、同時に実行することはできない。また、これらの規則だけでは、どちらの規則を選択すべきか、あるいはともに破棄すべきであるかを一意に決定することはできない。このような場合、局所的な規則によって強制的に一意性を確保することが多く、経路上での通信に対する制御の一貫性が失われる可能性がある。

本章では、上記のようなパラメータ・フィルタにおける曖昧性の排除を容易にするために、2段階パラメータ・フィルタ方式を提案する。この方式では、パラメータ・フィルタの分類過程において、一意に決定できる部分と潜在的に曖昧性を持つ部分を分離している。これにより、新たなパラメータ・フィルタを構成する際には、曖昧性を排除することが容易になる。

## 7.2. 2段階分離モデル

パラメータ・フィルタは、以下のふたつの処理がおこなわれる。

1. あらかじめ定義された規則に基づく通信の識別と分類
2. 分類結果に基づく、通信に関連づけられた処理の導出

このうち、あらかじめ定義された規則を“フィルタ規則”と呼ぶ。フィルタ規則は、通信を識別するための必要な項目と条件から構成される。多くの場合、フィルタ規則を構成する項目は、送信者及び受信者の IP アドレス、トランスポート層のプロトコルおよびポート番号がもちいられる。

フィルタ規則には、そのフィルタ規則を満足する通信に適用する処理が関連づけられる。この処理を“アクション”と呼ぶ。具体的なアクションは、パラメータ・フィルタが応用される場面に大きく依存する。たとえば、7.1 節でのファイアウォールの例では、「通過を許可する」「通過を拒否する」「結果を記録する」などが考えられるであろう。また、複数のアクションが同時に適用されることもありうる。

フィルタ規則とアクションの対を“分類レコード”と呼ぶ。分類レコードが設定されると、パラメータ・フィルタは流入する IP データグラムを検査し、適応する分類レコードの集合を抽出する。IP データグラムには、抽出された分類レコードのアクションが適用される。

以下では、パラメータ・フィルタに設定されている分類レコード全体を  $R_{all}$ 、流入する IP データグラム  $d$  に適用する分類レコードの集合を  $R_{result(d)}$  とする。また、ある分類レコード集合  $R$  を構成する分類レコードのフィルタ規則、または、アクションのことを、単に『 $R$  のフィルタ規則』『 $R$  のアクション』などと呼ぶ。

このとき、 $R_{result(d)}$  はある手続き  $f$  によって

$$R_{result(d)} = f(R_{all}, d) \quad (7.1)$$

と表すことができる。ここで、 $R_{result(d)}$  のフィルタ規則は、少くとも流入する IP データグラム  $d$  を満たす必要がある。したがって、 $d$  を満足するフィルタ規則を含む分類レコードの集合を  $R_{mid(d)}$  とすると、手続き (7.1) は以下のようにふたつの段階に分割することができる。

$$R_{mid(d)} = f_1(R_{all}, d) \quad (7.2)$$

$$R_{result(d)} = f_2(R_{mid(d)}, d) \quad (7.3)$$

手続き 7.2 と手続き 7.3 を、それぞれステージ 1、ステージ 2 と呼ぶ。また、パラメータ・フィルタをステージ 1/ステージ 2 に分離したモデルを、2 段階分離モデルと呼ぶ。

ステージ 1 は、対象となる IP データグラム  $d$  に対し、全てのフィルタ規則を適用して検査する。その結果、 $d$  を満足する分類レコードの集合を  $R_{mid}$  とする。この過程は、フィルタ規則の構成にのみ依存するため、パラメータ・フィルタが利用されるサービスに独立して、自動的に処理することができる。

一方、ステージ 2 は著しくサービスに依存する。たとえば、インターネット層の経路制御では、流入する IP データグラムの受信者アドレスに対し、最長一致の分類レコードが選択される。また、

表 7.1: プロトタイプ実装におけるパラメータの型

型		比較一致条件	
名称	説明 (例)	名称	説明
整数型 (符合あり/なし)	整数列およびそのデータ長 (TCP/UDP のポート番号)	完全一致	整数列とデータ長の完全一致
ビット列型	ビット列およびそのビット長 (IPv4 アドレス (32 ビット) / IPv6 アドレス (128 ビット))	完全一致	ビット列とビット長の完全一致
		プレフィックス長付一致	指定されたプレフィックス長でのマスク演算結果に対する完全一致
文字列型	文字列および文字列長 (URL など)	完全一致	文字列と文字列長の完全一致
特殊型	ワイルド・カード	無条件一致	全ての型・全てのデータと無条件に一致

ファイアウォールや優先転送の転送機構では、最初に一致したもの、あるいは最後に一致したものを採用することが多い。

このように、パラメータ・フィルタの2段階分離モデルは、分類レコードの抽出をサービスから独立した部分と、サービスに依存する部分に抽象化して分離している。したがって、複数の機器間で発生したパラメータ・フィルタの抽出結果の矛盾を解決したり、あたらしいサービスのためにパラメータ・フィルタを構成する際にも、ステージ2の構造のみに着目すればよい。

## 7.3. 実 装

本節では、前節で提案した2段階パラメータ・フィルタのプロトタイプ実装について述べる。実装は、C言語をもちいて NetBSD 1.5.2 および NetBSD 1.6 上でおこなわれた。また、2段階パラメータ・フィルタは、ユーザランドにおけるライブラリ、および、カーネルの一部として実装された。このプロトタイプ実装は KUPF と呼ばれる。

### 7.3.1. パラメータの型

フィルタ規則を表現するためのパラメータの型として、整数型、ビット列型、文字列型、および特殊型の4つの基本型と、それぞれに対応する基本的な比較一致条件を定義した。これらを表



7.1 に示す。フィルタ規則は、これらの型を複数個組み合わせたデータ構造と、それらに対する比較一致条件から構成される。

また、ここにあげた以外の型に関しても容易に追加できる枠組みを用意した。

### 7.3.2. ステージ 1

ステージ 1 では、フィルタ規則に含まれる全ての項目に対して、その項目の型と比較一致条件にしたがい、候補となる分類レコード群を抽出する。フィルタ規則のデータ構造と具体的な項目が定義され、対象となるデータを入力すると、抽出は機械的な比較一致作業の連続で終了する。

### 7.3.3. ステージ 2

ステージ 2 はサービスに依存する部分を集約したものであるため、一般的な実装をおこなうことはできない。次節では、具体的なサービスの定義に基づいて、ステージ 2 の評価実装をおこなっている。

## 7.4. 応用実装による動作確認

プロトタイプ実装の動作を確認するため、KUPF を応用した簡単なサービスを構築した。ここでは、応用実装と動作確認について述べる。

### 7.4.1. 応用実装例

KUPF の動作確認をおこなうため、具体的なサービスを想定して KUPF を利用した応用実装をおこなった。サービスは ALTQ [11] を利用した DiffServ によるマーキングとし、ALTQ の分類器に KUPF を適用した。応用実装の構成を図 7.2 に示す。この応用実装は以下のような動作を示す。

1. ALTQ に IP データグラムが流入する。
2. IP データグラムから必要な情報を取得し、KUPF へ入力する。
3. ステージ 1 では、入力された情報から最終的なアクションの候補となる分類レコード群が抽出される。
4. ステージ 2 では、分類レコード群から後述する規則にしたがってアクションを決定する。
5. KUPF によって決定されたアクションを IP データグラムに適用し、出力する。

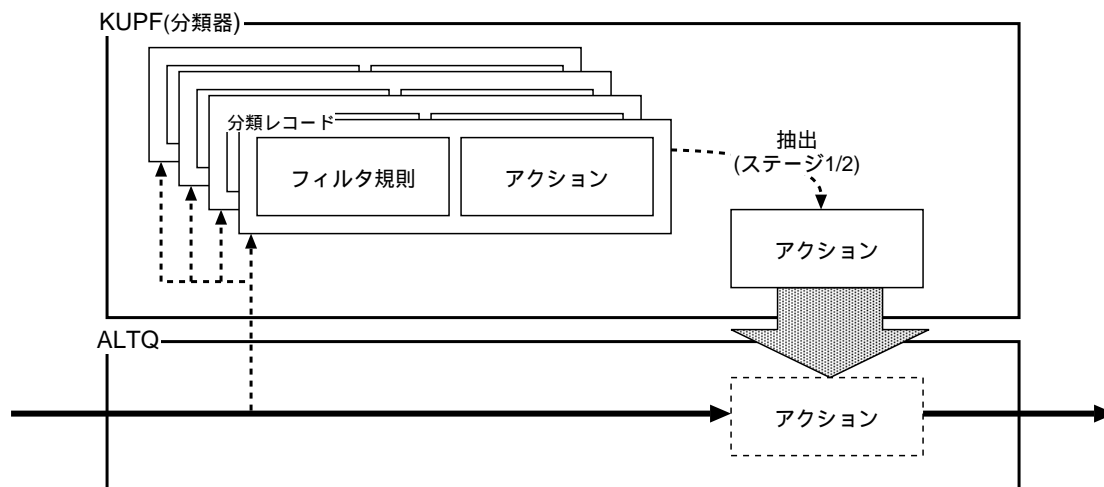


図 7.2: 応用実装例の構成

KUPF は ALTQ に流入した IP データグラムから、適用すべきアクションを抽出する。アクションには廃棄や計測などが含まれる。なお、ALTQ に含まれるパラメータ・フィルタとの互換性のために、ステージ 2 はフィルタ規則に含まれる優先度のみによる選択とした。

フィルタ規則のデータ構造として、表 7.2 および表 7.3 を定義した。フィルタ規則の項目のうち、優先度以外はステージ 1 において IPv4/IPv6 データグラムを分類するためのものである。また、KUPF による分類器の挙動を ALTQ が備える分類器と互換性のあるものにするため、フィルタ規則の項目に優先度を加えた。ステージ 1 で複数の分類レコードが抽出された場合、ステージ 2 ではこの優先度によって唯一の分類レコードを選択する。

#### 7.4.2. 動作確認

前節で述べた応用実装によって KUPF の動作を確認するため、DiffServ の流入境界ノードにおけるマーキングの試験をおこなった。実験のネットワーク構成を図 7.3 に示す。この図のように 3 台の PC を直列に接続した。3 台の PC の仕様を表 7.4 に示す。3 台のうち、中央の PC はルータとして機能している。また、この PC では応用実装した ALTQ+KUPF カーネルで動作しており、KUPF には 1,000 個のフィルタ規則が設定されている。この構成において、送信端末から受信端末に向けて IP データグラムを送信し、意図した通りのマーキングが加えられていることを確認した。

表 7.2: フィルタ規則のデータ構造 (IPv4 用)

属性	パラメータの型 (長さ)
プロトコル	符号なし整数型
TOS	ビット列型 (8 ビット)
受信者 IP アドレス	ビット列型 (32 ビット)
送信者 IP アドレス	ビット列型 (32 ビット)
受信者ポート	符号なし整数型
送信者ポート	符号なし整数型
IPsec 情報	符号なし整数型
優先度	符号なし整数型

表 7.3: フィルタ規則のデータ構造 (IPv6 用)

属性	パラメータの型 (長さ)
プロトコル	符号なし整数型
トラヒック・クラス	ビット列型 (8 ビット)
フロー・ラベル	符号なし整数型
受信者 IP アドレス	ビット列型 (128 ビット)
送信者 IP アドレス	ビット列型 (128 ビット)
受信者ポート	符号なし整数型
送信者ポート	符号なし整数型
IPsec 情報	符号なし整数型
優先度	符号なし整数型

応用実装で利用したフィルタ規則のデータ構造を示す。ALTQ による DiffServ を提供するサービスと仮定し、KUPF を分類器として組み込んだ。表 7.2 および表 7.3 は、それぞれ IPv4/IPv6 の通信を識別するためのデータ構造である。ステージ 1 での分類レコードの抽出には、IPv4/IPv6 のどちらも優先度以外の項目が利用された。一方、ステージ 2 では優先度が利用された。

表 7.4: 動作確認実験に利用した PC の仕様

モデル	IBM PC/AT 互換機
CPU	Intel Pentium III 1GHz
メモリ	512MB
ネットワーク	Intel i82557 Ethernet 100baseT
OS	NetBSD 1.6 (ルータは ALTQ+KUPF カーネルで動作)

## 7.5. 考察と課題

パラメータ・フィルタの 2 段階分離モデルに関する研究は、まだ始まったばかりである。現在までの結果をふまえ、考察と課題について述べる。

### 7.5.1. ポリシ実現や矛盾検出の検証

本研究では、パラメータ・フィルタの 2 段階分離モデルの提案と、プロトタイプ実装による動作確認をおこなった。しかし、さまざまなステージ 2 を定義することによる多様なサービスの実現や、パラメータ・フィルタ間での矛盾検出等の検証はおこなっていない。提案モデルの有効性を示すためにも、これらの検証を早期におこなわなければならない。

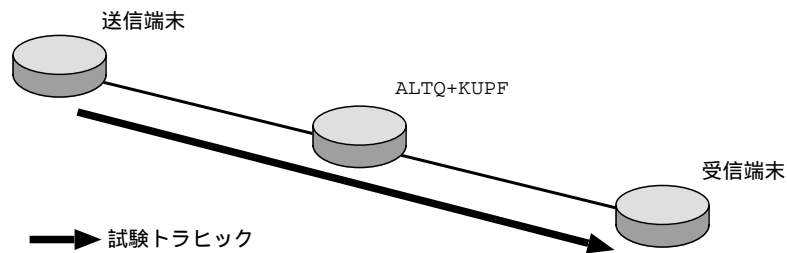


図 7.3: 動作確認実験の環境

動作確認実験におけるネットワーク構成を示す。3 台の PC を直列に接続した。また、中央の PC は応用実装したカーネルにより、DiffServ の流入境界ノードとして動作させた。この構成で送信端末から受信端末に向けてトラフィックを発生させ、中央の PC においてマーキング処理をおこなった。

表 7.5: 分類レコードの抽出に要する時間

フィルタ規則	処理速度 (ミリ秒)	
	ALTQ	KUPF
IPv4 (ホスト)	0.002	0.65
IPv4 (ネットワーク <sup>a</sup> )	0.021	0.63
IPv6 (ホスト)	0.25	0.63
IPv6 (ネットワーク <sup>b</sup> )	0.25	0.50

<sup>a</sup>28 ビットのプレフィックス長をもちいた。

<sup>b</sup>64 ビットのプレフィックス長をもちいた。

ALTQ に備わっている分類器と KUPF で、分類レコードの抽出にかかる時間を比較した。

### 7.5.2. 抽出の高速化

前節の動作確認の際、参考として処理速度を計測した。その結果を表 7.5 に示す。実験にはランダムな受信者アドレスの IP データグラムをもちいたため、キャッシュなどの特殊な条件はない。この表にみるように、KUPF の現在の実装は ALTQ にくらべてかなり低速であることがわかる。この結果は、実験前に以下の理由から予測されていた。

#### 検索方法

ALTQ がハッシュを利用した検索であるのに対し、KUPF のステージ 1 では逐次検索を利用している。

#### KUPF の抽象化

分類レコードの抽出を 2 段階に抽象化したことによって、抽出のオーバーヘッド

が増加する。

検索方法については、プレフィックス長や複数のパラメータ、範囲を持つ値などを扱う必要があるために、単純なハッシュや B-tree をもちいることはできない。空間的なデータの検索手法としては R-tree [38] やその変種が利用されているが、それらの応用によってステージ 1 の高速化が期待できる。

また、一般にパラメータ・フィルタを利用するサービスでは、ステージ 2 の規則を頻繁に変更することはないと考えられる。そのため、サービスに導入する際の実装が抽象化されている必要はない。そこで、具体的なステージ 2 が定義されたのち、ステージ 1 と統合して高速化するような、汎用的な手法の研究が必要であると考えられる。



本研究では、要求駆動型通信品質制御サービスの提供を実現するために、制御層・転送層の要素技術をさまざまな視点から検討した。また、設計や実装、またそれらを利用した実証実験等を通じて、いくつかの知見を得ることができた。しかし、商用サービスとして利用者への提供を開始するためには、まだまだ多くの課題が残されている。本章では、本研究で得られた知見と今後の課題について述べ、研究を総括する。

## 8.1. 研究を通して得られた知見

### 8.1.1. 全体を通じて得られた知見

#### 運用実験の困難さ

要求駆動型通信品質制御サービスを実現するためには、制御の対象領域内の機器が一貫性を保持しながら協調して動作する必要がある。したがって、運用実験をおこなうためには、実験の対象となる領域のネットワーク機器を、転送層として採用した機能を備えたものに置き換える必要があり、対象となるネットワークに対して非常に大きな影響を与える。一方、実利用者に対してサービスを実験的に提供するためには、その利用者が接続するネットワークを対象領域としなければならない。

このことから、サービスの運用実験をおこなう際には、利用者の従来の利用に対して影響を与えないよう、転送層の配置に十分な配慮を加える必要がある。本研究では実利用者が接続する仮設実験ネットワークを利用したが、商用ネットワークにおいて試験的なサービス提供をする場合には、転送層の機器とソフトウェアを綿密に検証するなど、十分な配慮が必要となる。

#### インターネットを取り巻く環境の変化

インターネットを取り巻く環境の変化は目覚ましい。ネットワークの基盤を支えるデータリンク層の帯域が高速化の一途を辿る一方、機器や接続にかかるコストは低下し続けている。また、インターネット上で提供されるアプリケーションやコンテンツは、多様化し続けている。このため、利用者が享受または必要とする通信品質も刻々と変化している。このような状況を正しく分析し、

変化する環境に応じたサービス・クラスを提供するため、転送層の高機能化によるサービスの多様化と、制御層での柔軟な対応が必要である。

### サービス・クラスの属性

現在のインターネットの通信サービスに対する課金は、おもに提供されるリンクの帯域幅に基づいている。一般の専用回線では、利用者はその帯域を重視するため、利用者が重視する特性と価格決定の対象が一致している。この場合、サービス価格によって利用者の挙動を制御することができる。たとえば、狭帯域のサービスを廃止したい場合には、サービスの価格を引き上げることで、利用者が契約を安価で広帯域なサービスに変更することが予想できる。

しかし、特性の異なる複数の有為な回線を提供する場合には、価格決定の対象とした回線特性が、利用者の重視する回線特性とは異なる可能性がある。本研究では、帯域と往復所要時間の特性が大きく異なる回線を、個人が選択的に利用できるサービスを提供した。この実験では帯域に基づく価格設定をおこなったが、利用者が重視した特性は往復所要時間であった。このため、価格の変動による利用者の制御ができなかった。

このように、極度に異なる特性を持つ複数のサービス・クラスを提供する場合には、価格設定の対象となる特性は慎重に選択しなければならない。

## 8.1.2. 個別の研究に関する知見

### 制御層の障害に対する配慮

通信品質制御サービスでは、現在広く提供されている接続性の提供とくらべて高い可用性が求められる。特に、契約継続中の通信に対する優先転送は、障害に対する高い耐性が必要であろう。優先転送の経路中で転送層に障害が発生した場合に関しては、高速に代替経路を確立するせざるを得ない。しかし、すでに確立されている転送層での優先制御は、制御層の障害から独立しているべきである。すなわち、制御層の障害を検知しても、再び復旧するまでは優先転送を継続するべきである。

また、制御層の二重化の方式を検討するなど、制御層の可用性を高めることも重要である。

### 非同期処理の困難さ

通常、導出された制御パラメータを転送機構に適用するには、複数の転送層機器に制御を加える必要がある。要求駆動型サービスの場合には、要求が転送層に反映されるまでに要する時間を短縮するため、これらの機器に対する制御は並行しておこなわなければならない。この場合、並



行設定の過程でいずれかの機器に異常が発生すると、他の機器の扱いが困難となる。設定過程では、いかなる状態においても矛盾が発生しないようするため、それぞれの機器に対する状態を管理して複雑な例外処理をおこなうなど、非常に高度なサーバ設計が必要である。

### 制御の完了と通信の開始

本研究では、利用者が通信を開始する前に Web インタフェイスや専用クライアントをもちい、優先転送要求を発行するという形態をとった。そのため、利用者が優先転送を希望する通信を開始するときには、すでに優先制御の準備が完了している。しかし、NAT 変換器の写像情報通知は、利用者が通信を開始し、変換情報を取得してから制御が開始される。そのため、通信開始直後には制御を完了することが難しい。

このように、通信の開始とともに優先制御要求が発行される場合には、通信開始から制御が完了するまでの間の通信品質に関して、利用者とサービス提供者の間でなんらかの合意が必要であろう。

## 8.2. 今後の課題

### 8.2.1. 通信制御要求発行のインタフェイス

近年、CATV や ADSL(Asynchronous Digital Subscriber Line)、さらには FTTH(Fiber To The Home) といった技術が注目されている。こからは主に収容局や ISP から利用者までを接続する、ラスト・ワンマイルと呼ばれる区間の回線として利用される。これらの技術を利用したインターネット接続サービスの普及にともない、一般家庭のインターネットへの接続形態は、間欠接続から常時接続へ、狭帯域から広帯域へと、大きな転換期を向かえた。

この現象は、広帯域ストリーミング等の通信品質を要求するコンテンツや、遅延にきびしい制約条件を持つアプリケーションを、一般家庭に浸透させる原動力になっている。したがって、今後、一般家庭における利用者からも、通信品質制御に対する突発的な要求が寄せられるようになるであろう。

本研究の制御層に関する実験では、通信品質制御の要求を発行するために、利用者に対して専用クライアントや Web インタフェイスを提供した。これらのインタフェイスでは、優先制御の対象とする通信を識別するために、送信者および受信者の IP アドレスやトランスポート層のプロトコル、送信者および受信者のポート番号等を指定する必要がある。

このような、通信に関する詳細な属性の指定は直感的ではない。特に、ネットワークの専門知識を持ち合わせていない一般家庭の利用者には、これらの属性を要求されるインタフェイスは、非

常に難解なものとして受け止められる可能性が高い。したがって、通信品質制御サービスの提供を開始し、一般家庭の利用者に普及させていくためには、利用者の視点に立った直感的で容易なインタフェースを用意する必要がある。たとえば、端末内の通信を監視しながら、利用者に要求の有無を問い合わせるようなミドルウェアの開発が必要であろう。

### 8.2.2. 個別の細目条件の検討

本研究で利用した細目条件は、実証実験に特化したものであった。たとえば、明示的経路を与える細目条件は、あらかじめ指定された経路を出力するものであった。また、認証や課金のシステムは PostgreSQL を利用した独自のものを実装・利用した。

しかし、広域ネットワーク上でサービスを展開する際には、資源の状態によって動的に経路を計算する必要があるだろう。また、課金はさまざまなサービスで共有するために、RADIUS [39] 等の標準的な仕組みを利用することが望ましい。このような個々の細目条件の詳細な検討が必要であろう。

### 8.2.3. 中間エンティティへの対応

本研究では、制御対象となる通信の識別子変換器として、NAT への対応をおこなった。しかし、識別子を変換することは、インターネットのあらゆる階層で考えられる。たとえば、ファイアウォールなどで利用される SOCKS [40] は、トランスポート層におけるセッションの中継をおこなうため、通信の識別子を変換される。WWW キャッシュサーバはアプリケーション層による通信識別子変換の代表的な例である。また、IPv6 の普及過程においては、IPv4/IPv6 プロトコル変換器も頻繁に利用されると考えられる。

このようなことから、識別情報を構成するパラメータは、送信者/受信者の IP アドレスやポート番号のみに限定するのではなく、より多くの項目の取り扱いが必要である。たとえば、サービス・クラスの定義から必要となる項目を導出し、中間エンティティへ通知することによって、指定された情報を与える機構の実現などが考えられる。

### 8.2.4. ドメイン間の制御

インターネットにおいて、ある共通のポリシーの元に管理される領域のことを、ドメインと呼ぶ。たとえば、多くの会社や大学などの組織は、ある共通のポリシーの元で運用されるため、ドメインを構成しているといえることができる。また、自律システム (Autonomous System; AS) は複数の組織を内包することが一般的であるが、それらの接続組織を含めた共通の (あるいは接続組織との契

約上で合意された) ポリシが存在する。この意味から、自律システムもドメインを形成しているといえる。

本研究における通信品質制御の枠組みは、主に単一ドメイン内で完結する制御として取り扱った。この場合、要求に対する応諾の可否は、そのドメインにおけるサービス提供ポリシのみを判断基準として決定すればよい。また、ネットワークは単一の運用主体の元に管理されるため、転送層における挙動の一貫性も確保しやすい。

しかし、インターネット上の通信の多くは、複数のドメインが関与するものである。したがって、利用者の通信品質制御に対する要求もまた、複数のドメインに渡るものである可能性が高い。異なるドメインでは、サービス提供ポリシが異なるため、単一のドメインのポリシのみで要求の応諾可否を決定することはできない。また、ネットワークの運用主体も異なることから、転送層での挙動も一貫性を確保することが難しくなることが予測される。

このような、複数ドメインが関与する通信品質制御を実現するためには、ドメイン間におけるポリシの折衝や資源の確保、課金の調整等が必要になる。今後、ドメイン間制御の枠組みを検討・確立させることは、商用サービスとして要求駆動型品質制御を提供していく上で、非常に重要な位置を占めると考えられる。

#### 8.2.5. 制御層-転送層間インタフェイスの定義

現在、通信品質に対してなんらかの制御を加える機構は、転送層における枠組みの標準化が独立して先行している。これらの枠組みにおいては、転送層が機構内部に閉じた形で利用されることを想定している。そのため、機構の外部で生成された制御要求を、転送層に反映するための仕組みに関しては考慮されていない。

制御層と転送層の分離モデルでは、それぞれの層に属する機構が独立して動作する。ここでは、制御層における決定を転送層に対して反映し、サービスを実現するという構造をとっている。このため、系全体として通信品質制御サービスを提供するためには、それぞれの制御層と転送層のそれぞれに属する機構を、適切に接合して協調させる必要がある。

また、通信品質制御の結果として表面に現れる効果は、転送層での通信に対する処理である。このため、さまざまな特色を持つサービス・クラスを提供するためには、転送層の多様化と高機能化をともなうであろう。

このようなことから、制御層と転送層の単一的なインタフェイスを定義する必要がある。さらに、このインタフェイスは、将来的に改良された、あるいは、新しく登場する転送機構に対して、制御層の決定をうまく反映できるような、拡張性を持った形で設計する必要がある。

### 8.2.6. 実装品質の向上

現在の制御機構の実装は、一部に研究会での実験用に特化した部分が含まれており、不安定要素も確認されている。また、CR-iLDPの実装は性能検証のための簡易なものである。より多くの実証実験をおこなうためにも、より頑健で抽象度の高い実装へと改善を続けていく必要がある。

## 8.3. 結 論

1990年初頭から始まった商用化により、研究を目的として構築されたインターネットは、社会を支えるコミュニケーション基盤としての役割を担うようになった。ひとびとがインターネットを通じて利用できるコンテンツやアプリケーションは、日を追うごとにその選択の幅を広げている。現在、ひとびとの興味は“インターネットを利用できること”ではなく、“インターネットをいかに快適に使うか”ということに移りつつある。ひとびとが利用するコンテンツやアプリケーションにとって最適な通信品質を、必要なときに必要なだけ提供できる環境の整備が必要とされる時代は、もう目の前に来ている。本研究では、利用者の要求によって駆動する通信品質制御サービスの実現を目標とし、そのために必要な技術の設計・開発と実利用者への試験提供をおこなった。

通信品質制御をおこなうためには、制御処理を加えるネットワーク領域内で、処理内容や資源管理の一貫性を保持して制御する必要がある。また、要求駆動型の通信品質制御サービスにおいては、要求は時間的粒度の細かなものとして現れる場合が多いと考えられる。しかし、自律分散協調性の概念のもとに設計されたインターネットでは、一貫性を保ちながら断続的な制御をおこなうことは困難である。本研究は、インターネット階層構造における制御層と転送層の分離という視点から進められた。この分離によって、制御層で特定領域内の一貫性を保持しながらさまざまな判断をおこない、また、転送層では制御層の決定に基づいて自律分散的に処理をおこなうことが可能となる。

通信品質制御サービスでは、サービス提供者ごとに独自色の豊かなサービス・クラスを用意することが予想される。そのため、制御要求に対する制約条件を柔軟に構成するための枠組みが重要である。本研究では、制約条件をより細かな単位の細目条件に分割し、細目条件間に依存関係を定義することで、制約条件の構成を容易にすることを試みた。

さらに、実利用者が接続する仮設ネットワークにおいて、上記の枠組みを利用した要求駆動型通信品質制御サービスを実験的に提供した。3回にわたるサービス提供の内、特に最初のものは実利用者へ要求駆動型通信品質制御サービスを提供した、世界で初めての例として大きな意味を持つものであった。実験では、制御対象となる通信の特性に対して課金対象を注意深く選択しなければならないことや、通信品質制御に価値を見いだした利用者は繰り返しサービスを利用するこ

となどが明らかになった。また、独立した複数の転送層によって同様のサービスを展開する場合にも、単一の制御層によって制御することで、対象領域内での資源管理等の一貫性が維持できることを示した。さらに、通信経路上に介在する通信の識別子変換器に対して、その変換情報を取得することで、変換器の前後で一貫した制御が可能であることを示した。

転送層は、制御層で決定された制御内容を受け取り、処理に反映しなければならない。要求駆動型のサービスを考えた場合、利用者からの要求は対話的に発行されることが多いため、制御要求の反映に必要とする時間の短縮は重要である。本研究では、転送層のひとつである MPLS に対し、制御要求の反映遅延の縮小を試みた。MPLS において、明示的に指定された経路で LSP を確立する場合、CR-LDP と呼ばれるラベル配布方式を利用する。しかし、この方式では LSP の端点間をラベル要求とラベル対応が往復するため、LSP の確立に端点間の RTT 以上の時間が必要であった。これに対し、本研究で提案する CR-iLDP では、導出された経路上の LSR へ前後の LSR を並行して通知することで、ラベルの折衝に必要な時間を短縮することができる。プロトタイプ実装を利用した実験によって、これらの有効性を示した。

制御要求を反映した後の転送層では、制御対象となる通信を識別し、適用すべき処理を決定する。しかし、対象領域内に等しく同一の制御要求がおこなわれたとしても、それぞれの機器における通信の識別手法が異なる場合、ひとつの通信に対して異なる処理が適用される可能性がある。本研究では、上記のような不整合が発生する原因の特定を容易にするため、二段階分類方式を提案した。この方式では、通信から処理を導出する過程を、フィルタ規則から必然的に定まるステージ 1 と、サービス提供者のポリシーによって決定するステージ 2 に分離する。不整合はステージ 2 のみで発生するため、原因箇所の特定が容易となる。

以上のように、本研究では要求駆動型通信品質制御サービスの実現を目標とした、さまざまな要素技術に関する検討と開発をおこない、実験運用を通じてその可能性を明らかにした。しかし、商用サービスとして一般の利用者へ提供するためには、まだまだ多くの課題がある。今後も、継続して研究と開発をおこない、時代の要求に応じた通信制御を実現できるシステムを確立していく必要がある。





# 謝 辞

---

本研究を進めるにあたり、多くの人々のご協力と助言を賜りました。これらのご支援なくして、この研究は成し得ませんでした。ここに感謝の意を表します。

WIDE プロジェクト研究会でのサービス提供実験は、研究会の参加者各位のご理解と協力のもとで実現できました。WIDE プロジェクト関係者諸氏に深く感謝致します。また、実験に利用した制御層および転送層の一部は、慶應義塾大学との共同プロジェクトによる成果です。制御層のサーバ群は慶應義塾大学政策メディア研究科の小川晃通氏、2001年春の研究会の実験におけるL3TEは同所属の頼原桂二郎氏、同実験のnatmapdは奈良先端科学技術大学院大学インターネット工学講座の横山輝明氏の協力によって実現されました。ともに研究を進めた共同プロジェクトの関係者諸氏に深く感謝致します。

2001年春の研究会における転送層は、北陸先端科学技術大学院大学のあやめプロジェクトのご協力を戴きました。あやめプロジェクトは同じ目標に向かって研究を進める仲間であり、ときに励まし合い、ときに激しく議論を交し、そして、たびたび勇気づけられました。あやめプロジェクトの宇夫陽次郎氏、宇多仁氏、小柏伸夫氏に深く感謝致します。

MPLSにおけるLSP確立の高速化のプロトタイプ実装は、奈良先端科学技術大学院大学インターネット工学講座に所属されていた越智一敦氏に協力して戴きました。本文中では述べていませんが、氏は提案した手法の有効性をシミュレーションおよび数値解析によっても示されました。また、パラメータ・フィルタにおける二段階分類手法の実装は、奈良先端科学技術大学院大学情報科学センターの垣内正年氏によるものです。両氏に深く感謝致します。

東京大学の江崎浩助教授および加藤朗助教授、ソニーコンピュータサイエンス研究所の長健二郎氏には、共同プロジェクトの運営と研究を進めるにあたり、多大なご支援と数多くの貴重なご意見を戴きました。また、松下通信工業株式会社の鈴木良宏氏、および、石田寛史氏には、市場の要求や制御層における制約条件の構成に関してさまざまな示唆を戴きました。本論文執筆にあたっては、奈良先端科学技術大学院大学附属図書館研究開発室の羽田久一助手、情報科学センター和泉順子女史に多大なる支援を戴きました。深くお礼申し上げます。

ご指導戴いた奈良先端科学技術大学院大学インターネット工学講座の山口英教授、門林雄基助教授、飯田勝吉助手、奥田剛助手に感謝致します。そして、あたたかな助言を戴いた奈良先端科学技術大学院大学情報コミュニケーション講座の山本平一教授、情報科学センターの砂原秀樹教授、九州工業大学電子情報工学科の尾家祐二教授に深く感謝致します。







## 参考文献

---

- [1] 経済産業省, 電子商取引推進協議会, and 株式会社 NTT データ経営研究所. 平成 13 年度電子商取引に関する市場規模・実態調査, February 2002. <http://www.meti.go.jp/kohosys/press/0002379/0/020218ec.pdf>.
- [2] Zona Research, Inc. The Economic Impacts of Unacceptable Web-site Download Speeds. Write Paper, April 1999.
- [3] Internet Emergency Preparedness. <http://www.ietf.org/html.charters/ieprep-charter.html>.
- [4] Robert H'obbes' Zakon. Hobbes' Internet Timeline v5.6, April 2002. <http://www.zakon.org/robert/internet/timeline/>.
- [5] R. Braden, D. Clark, and S. Shenker. *Integrated Services in the Internet Architecture: an Overview*, June 1994. RFC 1633.
- [6] NSPIXP - Network Service Provider Inter eXchange Point. <http://nspixp.wide.ad.jp/>.
- [7] D. Durham, Ed., J. Boyle, R. Cohen, S. Herzog, R. Rajan, and A. Sastry. *The COPS (Common Open Policy Service) Protocol*, January 2000. RFC 2748.
- [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] S.Floyd and V.Jacobson. Link-sharing and resource management models for packet networks. *IEEE/ACM Transactions on Networking*, 3(4), August 1995.
- [10] 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.
- [11] K.Cho. A framework for alternate queueing: Towards traffic management by pc-unix based routers. In *USENIX*, 1999.

- [12] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. *An Architecture for Differentiated Service*, December 1998. RFC 2475.
- [13] K. Ramakrishnan, S. Floyd, and D. Black. *The Addition of Explicit Congestion Notification (ECN) to IP*, September 2001. RFC 3168.
- [14] K. Nichols, S. Blake, F. Baker, and D. Black. *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, December 1998. RFC 2474.
- [15] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. *Assured Forwarding PHB Group*, June 1999. RFC 2597.
- [16] B. Davie, A. Charny, J.C.R. Bennet, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu, and D. Stiliadis. *An Expedited Forwarding PHB (Per-Hop Behavior)*, March 2002. RFC 3246.
- [17] E. Rosen, A. Viswanathan, and R. Callon. *Multiprotocol Label Switching Architecture*, January 2001. RFC 3031.
- [18] Bruce S. Davie and Yakov Rekhter. *MPLS: Technology and Applications*. Networking Series. Morgan Kaufmann, May 2000.
- [19] E. Rosen, D. Tappan, G. Fedorkow, Y. Rekhter, D. Farinacci, T. Li, and A. Conta. *MPLS Label Stack Encoding*, January 2001. RFC 3032.
- [20] Y. Rekhter and T. Li. *A Border Gateway Protocol 4 (BGP-4)*, March 1995. RFC 1771.
- [21] Y. Rekhter and E. Rosen. *Carrying Label Information in BGP-4*, May 2001. RFC 3107.
- [22] G. Gross, M. Kaycee, A. Li, A. Malis, and J. Stephens. *PPP Over FUNI*, July 1998. RFC 2363.
- [23] D. Ooms, B. Sales, W. Livens, A. Acharya, F. Griffoul, and F. Ansari. *Overview of IP Multicast in a Multi-Protocol Label Switching (MPLS) Environment*, August 2002. RFC 3353.
- [24] R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, and S. Jamin. *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*, September 1997. RFC 2205.

- [25] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. *RSVP-TE: Extensions to RSVP for LSP Tunnels*, December 2001. RFC 3209.
- [26] J. Moy. *OSPF Version 2*, July 1997. RFC 2178.
- [27] L. Andersson, P. Doolan, N. Feldman, A. Fredette, and B. Thomas. *LDP Specification*, January 2001. RFC 3036.
- [28] B. Jamoussi, Ed., L. Andersson, R. Callon, R. Dantu, L. Wu, P. Doolan, T. Worster, N. Feldman, A. Fredette, M. Girish, E. Gray, J. Heinanen, T. Kilty, and A. Malis. *Constraint-Based LSP Setup using LDP*, January 2002. RFC 3212.
- [29] WIDE Project. <http://www.wide.ad.jp/>.
- [30] J. Heinanen and R. Guerin. *A Two Rate Three Color Marker*, September 1999. RFC 2698.
- [31] Postgresql. <http://www.postgresql.com/>.
- [32] K. Chan, J. Seligson, D. Durham, S. Gai, K. McCloghrie, S. Herzog, F. Reichmeyer, R. Yavatkar, and A. Smith. *COPS Usage for Policy Provisioning (COPS-PR)*, March 2001. RFC 3084.
- [33] Yojiro Uo, Zin Uda, Nobuo Ogashiwa, Satoshi Ohta, and Yoichi Shinoda. AYAME: A design and implementation of the CoS capable MPLS layer for BSD Network stack. In *INET2000*, July 2000.
- [34] F. Le Faucheur, L. Wu, B. Davie, S. Davari, P. Vaananen, R. Krishnan, P. Cheval, and J. Heinanen. *Multi-Protocol Label Switching (MPLS) Support of Differentiated Services*, May 2002. RFC 3270.
- [35] P. Srisuresh and K. Egevang. *Traditional IP Network Address Translator (Traditional NAT)*, January 2001. RFC 3022.
- [36] AYAME Project. <http://www.ayame.org/>.
- [37] The NetBSD Project. <http://www.netbsd.org/>.
- [38] Antonin Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In *ACM SIGMOD International Conference on Manegement of Data*, pages 47–57, November 1984.
- [39] C. Rigney. *RADIUS Accounting*, June 2000. RFC 2866.

- [40] M. Leech, M. Ganis, Y. Lee, R. Kuris, D. Koblas, and L. Jones. *SOCKS Protocol Version 5*, March 1996. RFC 1928.

## A.1. 本研究に関連する業績

### A.1.1. 学術論文誌

1. Naoto Morishima, Akimichi Ogawa, Hiroshi Esaki, Osamu Nakamura, Suguru Yamaguchi, and Jun Murai. Preliminary Field-Trial for QoS Routing and Dynamic SLA. *IEICE: Special Issue on Internet Technology*, Vol. E84-B, No. 8, pp. 2039–2047, August 2001.

### A.1.2. 国際会議 (査読あり)

1. Naoto Morishima, Akimichi Ogawa, Keijiro Ehara, , and Youki Kadobayashi. Field-Trial of dynamic SLA in Diffserv-capable network. In J.Crowcroft, J.Roberts, and M.I.Smirnov, editors, *Quality of Future Internet Services*, No. 1922 in Lecture Notes in Computer Science, pp. 117–128, Berlin, Germany, September 25-26 2000.
2. Naoto Morishima, Akimichi Ogawa, Hiroshi Esaki, Osamu Nakamura, Suguru Yamaguchi, and Jun Murai. Field-Trial of Dynamic SLA in Diffserv-capable Networks. In *International Workshop on Next Generation Internet and its Applications*, pp. 157–163, Tokyo, Japan, February 21-23 2001.
3. Masato Kakiuchi, Naoto Morishima, Yutaka Nakamura, Kazutoshi Fujikawa, and Hideki Sunaraha. KUPF: 2-Phase Selection Model of Classification Records. In *The Symposium on Applications and the Internet*, Orlando, Florida, USA, January 27-31 2003. (to be appeared).

### A.1.3. 国内会議 (査読あり)

1. 森島直人, 小川晃通, 額原桂二郎, 染川隆司, 山口英. 動的資源予約システムの運用実験. インターネット・コンファレンス, pp. 39–45, November 2000.

**A.1.4. 国内研究会等 (査読なし)**

1. 森島直人, 小川晃通, 額原桂二郎. 2000 年春の WIDE 合宿における Diffserv 実験について. 第二回 インターネットにおける QoS 提供に関するワークショップ, No. 10 in ITRC Technical Report, April 2000.
2. 森島直人, 小川晃通, 染川隆司, 額原桂二郎, 横山輝明. Moon bear project – くまぶろじえくと –. 九州インターネットプロジェクト 第三回 WG 会議, 九州インターネットプロジェクト 事業報告書, pp. 185–191, March 2001.
3. 森島直人. くまあやめプロジェクト wide 合宿実験概要. 九州インターネットプロジェクト 第 3 回 WG 会議, 九州インターネットプロジェクト 事業報告書, pp. 192–202, March 2001.
4. 横山輝明, 森島直人, 小川晃通, 宇夫陽次郎, 宇多仁, 江崎浩, 山口英. NAT ゲートウェイを通過するフローの網内での識別手法の提案. 情報処理学会研究報告, 第 2001 巻, pp. 95–100, June 2001.
5. 森島直人, 中村修. 招待講演: End-to-End QoS に向けて. 電子情報通信学会技術研究報告, pp. 31–36, October 2001. IN2001-94.
6. 越智一敦, 森島直人, 飯田勝吉, 山口英. 高速障害復旧が可能な MPLS ラベル配布機構の提案及び実装と評価. 情報処理学会研究報告, 第 2002 巻, February 2002. 2002-QAI-2.
7. 垣内正年, 森島直人, 宇多仁, 中村豊, 砂原秀樹. 汎用的なパラメータフィルタの設計と実装. 電子情報通信学会技術研究報告, 第 102 巻, pp. 7–14, June 2002. IA2002-2.

**A.1.5. 解説記事**

1. 森島直人, 小川晃通, 額原桂二郎, 染川隆司. 解説・次世代インターネット研究開発の最前線, 4. bandwidth broker system の開発と運用実験. 情報処理, Vol. 41, No. 10, pp. 1134–1138, October 2000.

**A.1.6. 講演**

1. 森島直人. QoS/CoS 技術. CKP セミナー, March 2001.

## A.2. その他の業績

### A.2.1. 国際会議 (査読あり)

1. Osamu Nakamura, Suguru Yamaguchi, Noriyuki Shigechika, and Naoto Morishima. ShowNet on INTEROP Tokyo 2002 – First Largest Demonstration Network of IPv4/IPv6 Dual Stack. Orlando, Florida, USA, January 27-31 2003. (to be appeared).

### A.2.2. 国内研究会等 (査読なし)

1. 高村真俊, 河部展, 森島直人, 門林雄基, 山口英. シームレスな IPv4/IPv6 の相互通信方式. 情報処理学会研究報告, 第 96 巻, pp. 79–84, September 1996. 96-DPS-78-14.
2. 森島直人, 小林和真, 山口英, 尾家祐二. 自律系における経路情報の管理方法の提案. 情報処理学会研究報告, 第 97 巻, pp. 201–206, November 1997. 97-DPS-85-28.
3. 宮地利幸, 宇夫陽次郎, 森島直人, 篠田陽一. N\*(NStar): ns-2 の real external interface の構想. 情報処理学会研究報告, 第 2001 巻, pp. 113–118, June 2001. 2001-DPS-103.
4. 瀧智博, 宇夫陽次郎, 宇多仁, 森島直人, 篠田陽一. Path Snapper: MPLS 検証システムの構想. 情報処理学会研究報告, 第 2001 巻, pp. 119–124, June 2001. 2001-DPS-103.
5. 衛藤将史, 森島直人, 門林雄基. 自律システム相互接続点における経路制御運用の自動化の為の技術要素の提案. 電子情報通信学会技術研究報告, pp. 15–20, July 2002. IA2002-11.

### A.2.3. 講演

1. 森島直人. TCP/IP とトラブルシューティング. Networld+Interop Tokyo ワークショップ, June 2000.
2. 森島直人. NSPIXP – 現状と今後. NEC ユニバーシティ技術研修, August 2001.





品質制御を実現する技術は、従来のインターネットの枠組みでは利用されない固有のものが多く、そこで使われる用語は一般的でないものが多い。そこで、本節では本文中で利用した一般的でないと思われる用語について説明する。

#### AF PHB グループ (Assured Forwarding PHB Group)

複数の PHB からなるグループを定義し、計測器の結果によって適切な PHB を使い分けることにより、最低帯域を保証した最善努力型の転送を提供するもの。RFC2597 で標準化されている。

#### ALTQ

BSD 系 UNIX 上で実装された、さまざまなパケット・スケジューラを管理するための枠組み。

#### CoS (Class of Service)

サービス・クラス間の通信品質や性能の相対的な差によって定義したサービス。インターネットにおける通信品質制御サービスとしての CoS は、絶対的な性能を保証するものではないため、QoS に比べて実現性が高いが、厳密な定義や容易に理解できる形でのサービス提示が困難である。

#### CR-LDP (Constraint-based Routed Label Distribution Protocol)

外部から与えられた明示的な経路に基づく経路選択を実現するために拡張された LDP。経路を明示的に指定するため、明示的経路 TLV (Explicit Route TLV; ER-TLV) と呼ばれる TLV を定義している。

#### DiffServ (Differentiated Services)

インターネットにおける通信品質の CoS を実現する枠組み。RFC2475 で標準化されている。

#### DSCP (DiffServ CodePoint)

DS ドメインの流入境界ノードにおいて IP ヘッダの DS フィールドに設定される、通信の

クラスを表す識別子。DiffServ の優先制御は、IP データグラム中に埋め込まれた DSCP のみを参照しておこなわれる。

#### DS ドメイン

同一のポリシーに基づいてサービス提供や運用がおこなわれ、領域内のノードが同一の転送規則に従うような DiffServ ネットワーク。

#### DS フィールド

IP ヘッダ内の DSCP を納める領域。IPv4 では従来の ToS フィールド、IPv6 ではトラフィック・クラスを再定義した 1 オクテットの領域のうち、上位 6 ビットが DS フィールドとして DSCP を格納する。

#### EF PHB (Expedited Forwarding PHB)

最小限の遅延と低い損失率で転送する PHB。帯域を保証するため、仮想専用線を実現するために利用される。RFC3246 で標準化されている。

#### E-LSP (EXP-Inferred-PSC LSPs)

MPLS 網内で DiffServ による優先転送サービスを実現するために、MPLS ヘッダ内の Exp フィールドに DSCP を写像する手法。RFC3270 で標準化されている。

#### FEC (Forwarding Equivalent Class)

MPLS 網における、同一の転送処理対象となる通信の集合。MPLS による転送の対象となった通信は、流入 LSR において特定の FEC に対応づけられる。MPLS 網内では、同一の FEC に分類されたデータはすべて同一に処理・転送される。

#### IntServ (Integrated Service)

インターネットにおける通信品質の QoS を実現するための枠組み。RFC1633 で標準化されている。帯域と最大遅延を保証するサービス、および、自らが流量を制御する適応型のアプリケーションを想定したサービスの、ふたつのモデルを想定している。

#### LDP (Label Distribution Protocol)

MPLS におけるラベル情報を配布するためのプロトコル。インターネット層で計算された経路情報を MPLS 網上に写像し、経路表に格納された各受信者アドレスを FEC とした LSP を確立する。RFC3036 で標準化されている。

#### LSP (Label Switched Path)

MPLS におけるデータ転送時の、データが MPLS 網に流入してから非 MPLS 網へ流出す

---

るまでの道筋。MPLS 網内で定義されている FEC のひとつに対応し、ひとつ以上の LSR からなる連続した論理パスを構成する。

#### LSR (Label Switching Router)

MPLS でラベルに処理を加えながらデータを転送するルータ。

#### MPLS (Multi Protocol Label Switching)

ラベルの概念を利用してデータを転送するラベル・スイッチング技術から発展した、インターネット層の下部に介在する転送機構。RFC3031 で標準化されている。階層構造的にはデータリンク層の上位にあたるため、sub-IP 層に位置づけられる。インターネット層での経路制御と独立した転送が可能のため、より柔軟な制御を実現するための仕組みとして注目されている。

#### PHB (Per-Hop Behavior)

DiffServ 網の中間ノードでおこなわれる具体的な処理。処理方式や実装、および、そのパラメータなどを表す概念ではなく、外部から観測できるデータ転送の挙動を表す。標準的な PHB としては、Default、Class Selector、Assured Forwarding (AF)、Expedited Forwarding (EF) が定義されている。具体的な実装には、なんらかの packets・スケジューリング方式をとまなうことが多い。

#### QoS (Quality of Service)

定量的に得られる絶対的性能を保証するサービス。

#### アドミッション制御

利用者からの要求に対して応諾の可否を決定すること。認証や課金、資源確保、ポリシーなど、さまざまな細目条件から構成される。

#### 境界 LSR

MPLS 網と非 MPLS 網の境界上に位置する LSR。MPLS 網に流入するトラフィックを基準にした場合には流入境界 LSR、MPLS 網から流出するトラフィックを基準にした場合には流出境界 LSR と呼ばれる。

#### 境界ノード

DS ドメインと非 DS ドメインの境界上に位置するノード。DS ドメインに流入するトラフィックを基準にした場合には流入境界ノード、DS ドメインから流出するトラフィックを基準にした場合には流出境界ノードと呼ばれる。

### サービス・クラス

サービス品質を構成する要素や、それに付随スフパラメータの集合によって定義されるサービスの単位。

### サービスレベル契約 (SLA; Service Level Agreement)

サービス品質に関する契約。現在、いくつかの ISP によって提供されているサービスでの SLA は、定量的に計測できる品質を要素とするものが多い。そのような要素としては、サービス開始遅延、可用性、障害復旧時間、障害通知時間、網内伝送遅延、帯域保証などが挙げられる。

### 中間 LSR/中間ノード

DS ドメインや MPLS 網の内部に位置する LSR やノード。