

## ■ 第3章

---

プリアンブル遅延型ラベル処理を用いた光パケット転送

### 3.1 はじめに

本章では、光パケット転送のためのラベル処理方式について述べる。自己ルーティングスイッチを用いてマルチホップ転送を行うには、予め中継ノードにおける転送制御情報がパケットのヘッダに記されている必要がある。また、各中継ノードでは、自身における制御情報がラベル中のどの位置に記されているかを知る必要がある。これらの要求に対し、経路上の各中継ノードにおける制御情報を通過順に並べてラベルとし、プリアンブル信号をラベルの既読部分相当だけ遅延させる光領域ラベル更新を行うラベル処理方式を提案する。ラベル更新により各中継ノードで読み取るべき情報の位置が固定され、識別長も短いことから非常に単純な論理ゲートの組み合わせのみでラベル識別が実現される。そのため、処理時間の固定化のみならず、光領域でラベル識別を行うことによる高速化の可能性を持つ。また更新処理は固定長の遅延線とスイッチのみによって可能であり、現在の光デバイス技術で十分実現できる。

本方式に関する評価としては、自己ルーティングスイッチを用いることで問題となるヘッダ長の影響について着目する。シミュレーションにより、ラベル更新においてプリアンブルを再利用する本ラベル処理方式は他の自己ルーティングスイッチを用いた方法よりもヘッダ長を短く抑えられることを示す。最後に、中継ノードのプロトタイプ実装について報告する。実証実験において光領域で所望のラベル処理とパケット転送が行われたことを確認し、設計やコスト面もふまえて本方式の実現性について議論する。

### 3.2 自己ルーティングスイッチを用いた光パケット転送

光パケット転送におけるラベル処理では、どのようなラベルを用いて経路・宛先を示すか、どのようにラベルの読み取りを行うか、どのようにラベルを更新するか、の3点における実現方法が課題となる [18]。固定長のラベルや IP アドレスを参照する場合は、パケット中の決まった位置に決まった長さの宛先情報が記されているため、読み取り処理は単純である。しかし自己ルーティングに基づく場合は各ノード制御ビットが転送時に参照されるラベルの一部分に書かれており、各ノードで参照すべき制御ビットの位置を知るための仕組みが必要となる。

光領域でのラベル識別処理を行う場合、ヘッダ中に書き込まれた情報を解析して制御ビットの位置を知るといった複雑な処理を行うことは難しい。そのため、各ノードで参照すべき制御ビットの位置があらかじめ分かっている必要がある。以下、既存のアプローチである「ノード毎に読むべき位置を固定的に決め、ラベル更新を行わない方法」「全てのノードで読むべき位置が固定であり、ラベル更新により制御ビットの位置合わせをする方法」について説明する。

なお、表 3.1 に以下共通して用いる記号を示しておく。

#### 3.2.1 Self-Routing Addressing

Self-Routing Addressing (SRA) は、自己ルーティングの転送制御情報を用いてネットワーク全体における一意な宛先アドレスを構成する方法である [19]。SRA ではヘッダに  $N$  個のフィールドを予め確保しておき、ネットワーク内の各ノードにおける転送制御情報をそれぞれのフィールドに記述する (図 3.1(a))。図 3.1(b) に示すトポロジーにおいて、転送制御情報として各出力インタ

表 3.1: 記号の定義

$N$	ネットワークの総ノード数
$h$	経路のホップ数
$r(n_i, n_j, k)$	$n_i$ から $n_j$ への経路の $k$ 番目のノード
$d(n_i)$	ノード $n_i$ の出力ポート数
$I(n_i)$	ノード $n_i$ における転送制御情報
$p$	プリアンブル長 (時間単位)
$x$	オプション長 (時間単位)
$g$	セーフティギャップ長 (時間単位)
$b$	ビットレート

フェースに1ビットが割り当てられる場合のノードDのアドレスは、ノードA, B, C, EでD宛てのパケットを送出するインタフェースがそれぞれb,c,c,bであることから、01 0010 001 00 001となる(宛先ノードのフィールドは全て0とする)。他のノードのアドレスも同様に割り当てられ、図3.1(c)のようなアドレス表が構築される。

パケット転送時には、中継ノードはそれぞれに割り当てられた転送制御情報領域を読み取り、完全一致により識別される出力先インタフェースにパケットを転送する(図3.1(b))。

SRAではヘッダ更新の仕組みが不要である代わりにアドレス長が長くなる。プリアンブルなどの制御信号を含めたヘッダ全体の長さは、時間を単位として

$$H_{SRA} = \frac{1}{b} \cdot \sum_{i=1}^N d(n_i) + p \quad (3.1)$$

と表される。

アドレス長が長くなるという問題に対しては、以下のような圧縮方法が提案されている。例えば図3.1(c)のアドレス表では、1列目と8列目と11列目、3列目と7列目、5列目と9列目はそれぞれ同じ情報であるため、これらを1列にまとめることで圧縮が行われる。また全ビットが反転した関係にある列同士は、NOTゲートが利用可能であればやはり1列にまとめることができる。インターネットバックボーンのリンク特性を想定した環境では、この圧縮方法によりアドレス長がおよそ30~40%圧縮されることが確認されている[19]。

### 3.2.2 ARTEMIS

ARTEMIS (All-optical self-RouTer EMploying bIt and packet-level procesSing) は、経路上の各中継ノードにおける転送制御情報を通過順に指定してマルチホップ転送を行う機構である[20]。ARTEMISのパケットフォーマットは、プリアンブル、転送制御情報、セーフティギャップを1つの単位(タグ)とし、これを通過ノードの順に並べ、その後パイロードが配置されるという構成になっている(図3.2(a))。セーフティギャップは、ヘッダの更新を導波路型スイッチの切替えによって行うための安全帯として機能する。 $d(n_i)$ ポートの出力インタフェースを持つ中継ノード  $n_i$

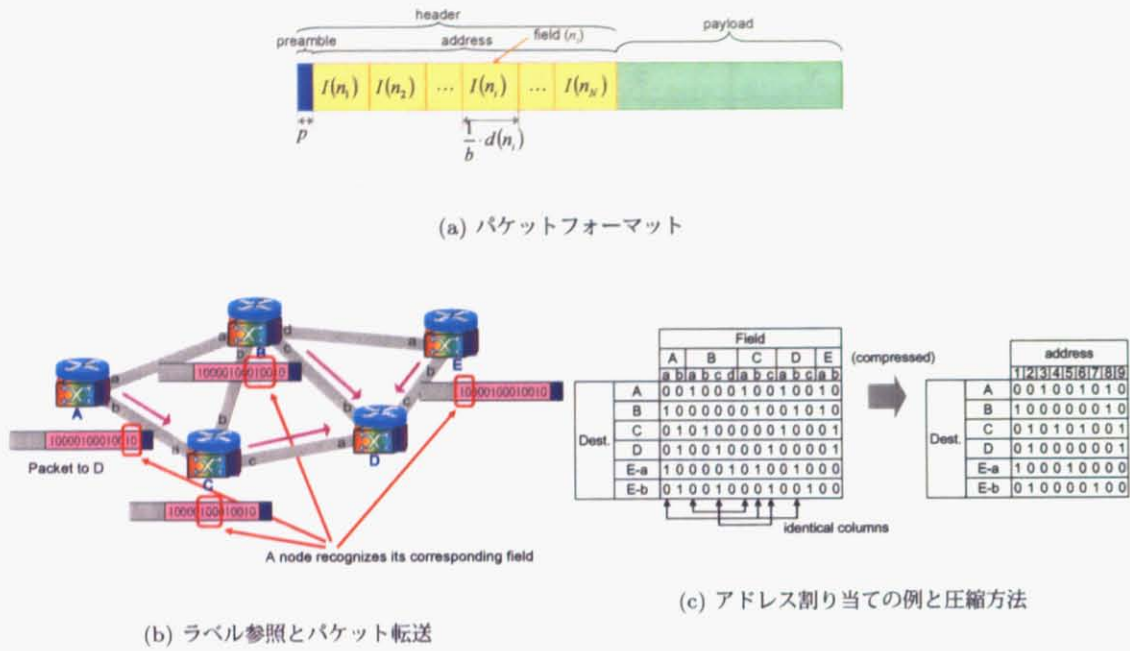


図 3.1: Self-Routing Addressing

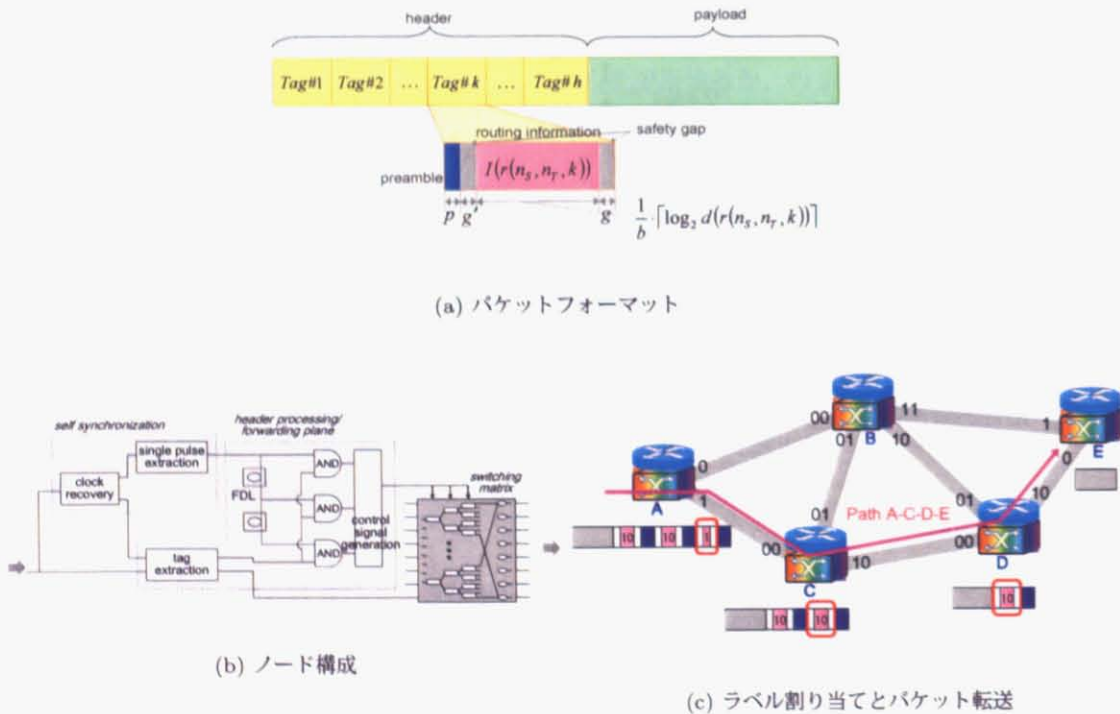


図 3.2: ARTEMIS

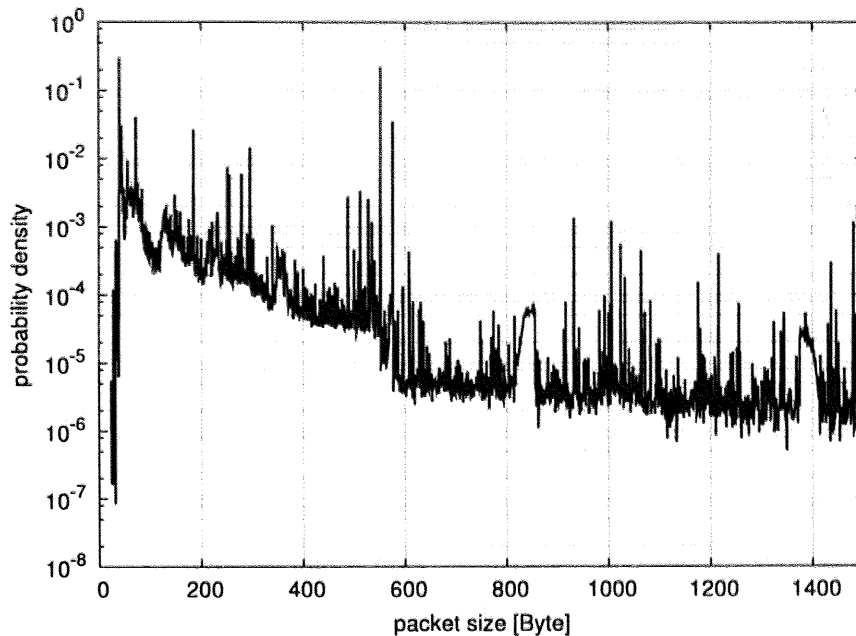


図 3.3: インターネットにおける IP パケット長分布 (1997 年)

における転送制御情報は  $\lceil \log_2 d(n_i) \rceil$  ビットの符号で表される。パケットが中継ノードに到着すると、パケットの先頭に配置されているタグとそれ以降の部分に分けられ、タグ内に記述された転送制御情報に基づきスイッチングマトリックスのスイッチが制御される (図 3.2(b))。スイッチングマトリックスは多段の  $1 \times 2$  スイッチによって構成され、各段のスイッチの制御信号が転送制御情報の各ビットとして用いられる。図 3.2(c) にパケット転送の様子を示す。各ノードにおける転送処理を制御するための情報が各ノードで削除されていくため、全てのノードで先頭のタグを識別処理することでパケットは宛先ノードまで届けられる。

ARTEMIS のヘッダはパケットの転送経路に関係のない不要な情報を含んでいないため、転送制御情報の合計ビット数は SRA に比べて短くなる。また、スイッチでタグを削除するという光領域で容易に実現できるヘッダ更新によりそれぞれの中継ノードでの読み取るべきタグが常にラベルの先頭に合わされているため、読み取り位置の検索が不要となる。ノード  $n_S$  から  $n_T$  の経路において、 $k$  番目のノード  $r(n_S, n_T, k)$  にパケットが到着した時点での時間単位のヘッダ長  $H_{ARTEMIS}$  は

$$H_{ARTEMIS} = \sum_{i=k}^h \left( \frac{1}{b} \cdot \lceil \log_2 d(r(n_S, n_T, k)) \rceil + p + g \right) \quad (3.2)$$

と表される。

### 3.2.3 既存手法のインターネットへの適用における問題点

式 (3.1)(3.2) より、自己ルーティングによる光パケット転送方式における各ヘッダ長は、ビットレートなどのデバイス性能やパケットフォーマットの他、ネットワークのトポロジ的な特性と大

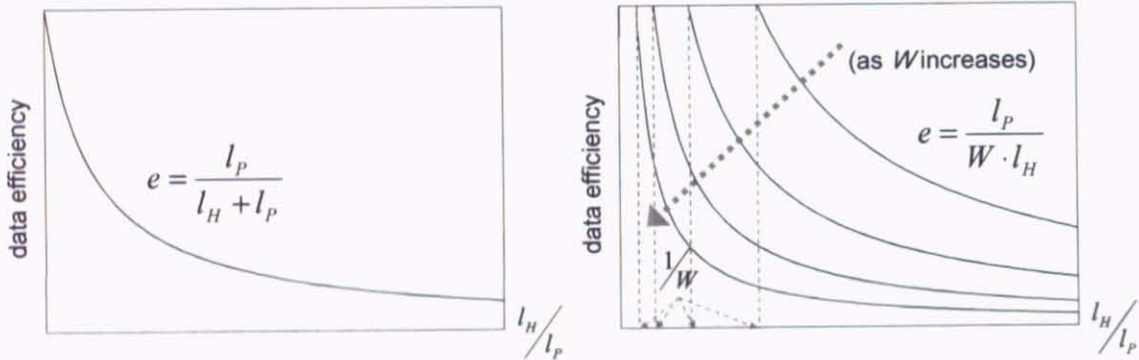
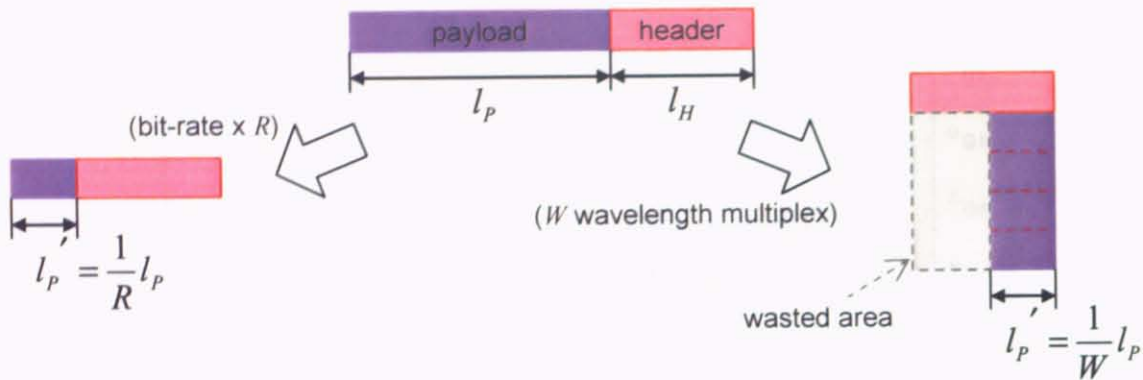


図 3.4: 光パケット構成とペイロードの効率

きな関連を持つことが分かる。具体的には、SRA ではネットワークの総ノード数に、ARTEMIS ではホップ数に比例してヘッダ長が長くなる。SRA のヘッダ長が長くなる原因は、宛先ノードへパケットを転送するための全ノードにおける転送制御情報をアドレスに記すため、中継ノードでは使われない不要な転送制御情報が含まれてしまうことにある。また、ARTEMIS においては、プリアンブルやヘッダの更新に必要なセーフティギャップといった転送制御情報以外の領域が大きくなってしまいう問題がある。転送制御情報自体はノードの出力インタフェース数が 1000 ポート以上の規模にならなければ数ビットに収まる。ヘッダのビットレートが 40Gbps である場合、これは 250ps 以内となる。それに対し、現在商用化されている光スイッチの切り替え速度は数 ns であり、図 3.2(a) よりタグの大部分はセーフティギャップということになる。たとえセーフティギャップが 1 ビットと同等の長さであったとしてもそのオーバーヘッドは小さくない。そのため、ホップ数の増加にしたがってヘッダが長くなる際の増加量が大きくなってしまいうことになる。

一方、インターネットにおけるパケット長分布は図 3.3 のようになっており、数十バイトという

短いパケットも多く含まれている [21]. さらに光パケット転送ネットワークではペイロード部分が各中継ノードで参照されないことから、ヘッダよりも高速に変調したり、また1つのデータパケットを分割し波長領域で多重したり [22] といったことにより転送の広帯域化が可能であり、さらにペイロードは短くなりうる. 図 3.4(a) にヘッダとペイロードが直列である場合と複数チャネルを利用して並列になる場合の概念図を示す (図中  $l_H, l_P$  はそれぞれヘッダ・ペイロードの長さ). ヘッダとペイロードが直列な場合は, リンク容量に対する実データの利用率が  $\frac{l_P}{l_H+l_P}$  を乗じたものになることから, ヘッダ長の増大に対し図 3.4(b) のように帯域利用率が減少してしまう. ヘッダとペイロードが並列な場合, ヘッダ長がペイロード長より長くなった場合にデータ信号を入れることができない領域がでるため, 帯域利用に無駄が生じる. ペイロードの波長多重数を  $W$  としたときの最大帯域利用率の傾向は図 3.4(c) に示すようになる.

このようにヘッダの長大化は光パケット転送ネットワークにおける帯域利用率に大きく影響することから, これを短く抑えられるラベル処理方式が求められる.

### 3.3 プリアンブル遅延型光ラベル処理による光パケット転送機構

本節ではプリアンブル遅延型ラベル更新を用いた光パケット転送機構 [61, 67] について述べる. 本機構では, 経路上の中継ノードにおける転送制御情報を通過順に並べることでその経路を示すラベルとする. そしてスイッチと遅延線のみを用いて, 先頭検出信号であるプリアンブルを遅延させることでラベルの更新を行う. プリアンブルを中継時に削除せず同じパケット内で再利用することで, プリアンブルとセーフティギャップがホップ数に比例して増加するのを防ぎ, ヘッダ長が短く抑えられる.

以下, 本機構の詳細について述べる.

#### 3.3.1 パケットフォーマット

パケットは, 中継ノードで転送制御のために参照されるヘッダと, 参照されることなく宛先ノードまで転送されるペイロードから構成される.

本機構におけるヘッダフォーマットを図 3.5 に示す. ヘッダはプリアンブル, オプション, セーフティギャップ, ラベルから成る. 本機構では, ヘッダの識別, 更新, スイッチ制御は全て, プリアンブルの検出時刻を基準に固定時間で行われる. そのため, プリアンブルとオプション, ラベル先頭のそれぞれの位置関係は固定とする. プリアンブルはパケットの先頭を識別するための信号である. オプションは転送に関わる補助的な情報を格納する領域であり, 必要な場合にはプリアンブルの直後に固定長の領域として確保する. オプションに記述される情報としては, 例えばパケット長や優先度クラスなどがある. セーフティギャップは, ラベル更新処理においてスイッチの切り替え制御によりプリアンブルとラベル以降を切り分けるための固定長の補助領域である. ラベルは各中継ノードにおける転送制御情報を通過ノード順に隙間なく並べたもので, 宛先までの経路を指定する識別子となる. 転送制御情報としては, ノードにおける出力ポートを識別可能な  $\lceil \log_2 d(n_i) \rceil$  ビットの符号を用いる.

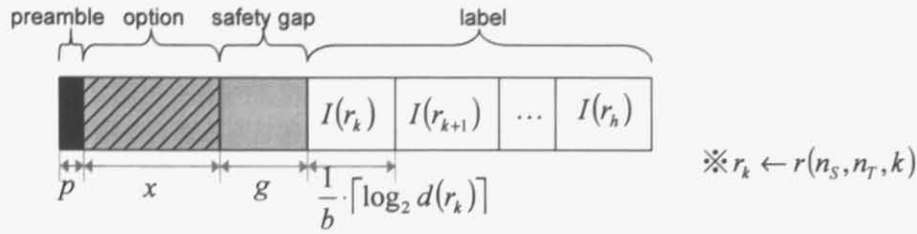


図 3.5: ヘッダフォーマット

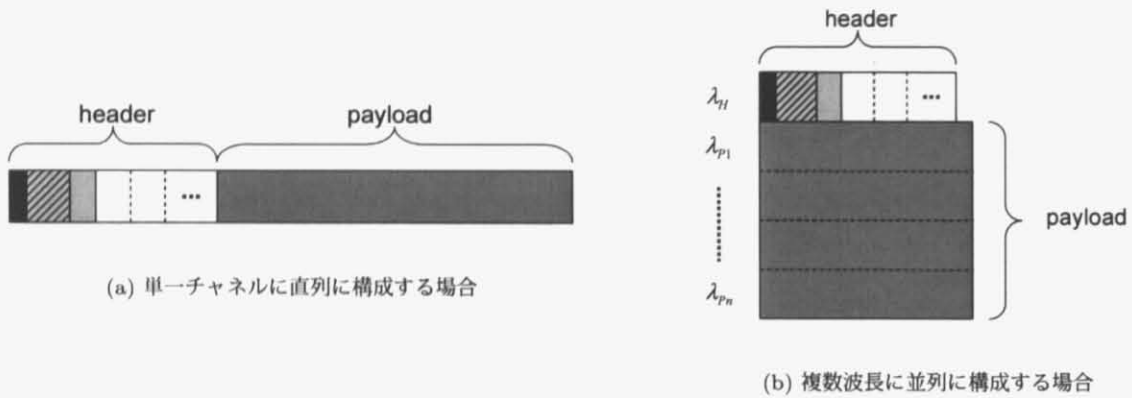


図 3.6: パケットの構成例

以上より、ノード  $n_S$  から  $n_T$  の  $h$  ホップの経路において、 $k$  番目のノード  $r(n_S, n_T, k)$  にパケットが到着した時点での時間単位のヘッダ長  $H$  は

$$H_{PDLU} = \sum_{i=k}^h \left( \frac{1}{b} \cdot \lceil \log_2 d(r(n_S, n_T, i)) \rceil \right) + p + x + g \quad (3.3)$$

と表される。

ペイロードは宛先ノードまで届けられるデータ本体で、ビットレート、符号化方法ともに任意な領域である。

本機構では、ヘッダの識別はプリアンブルを基準とした時間的に固定な処理として行われ、ヘッダの更新は遅延線を用いた時間領域での処理によって実現される。そのため、ヘッダとペイロードをどのようにパケットに構成するかは限定されない。パケットフォーマットの例として、ヘッダとペイロードを単一チャンネルに直列に並べて構成する場合を図 3.6(a) に、波長多重を用いて複数波長に並列に構成する場合を図 3.6(b) に示す。

### 3.3.2 ノード構成

中継ノードは各入力に対する転送処理機構 (forwarding plane)、出力毎の衝突回避機構 (contention resolution plane) によって構成される (図 3.7)。



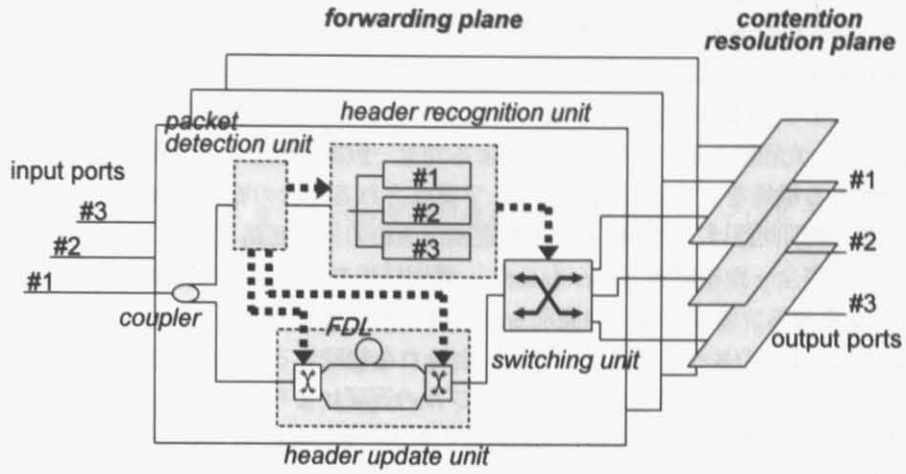


図 3.7: 中継ノードの構成

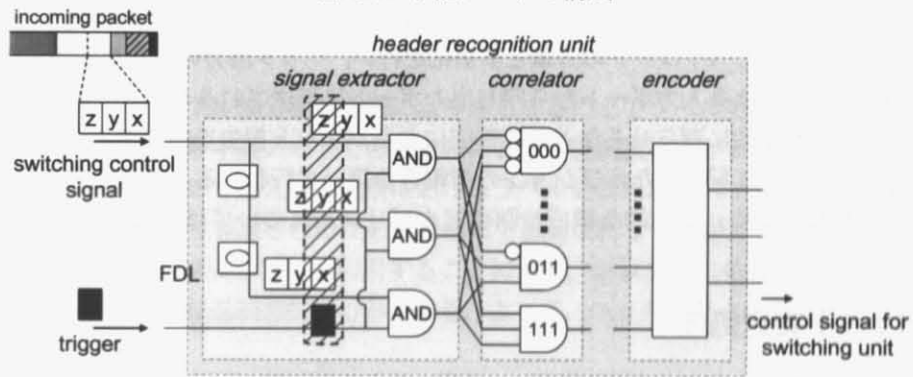


図 3.8: 論理ゲートによるヘッダ識別回路の構成例

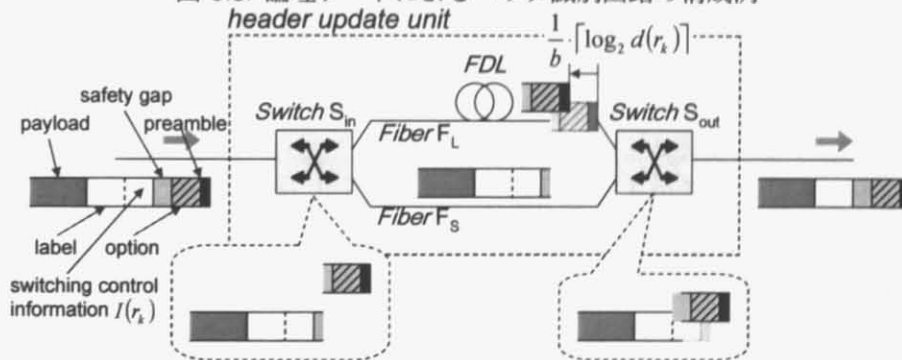


図 3.9: ヘッダ更新部の構成と更新処理

転送処理機構では入力パケットのヘッダの内容に基づいた出力ポートへの転送を行う。転送処理機構はパケット検出部 (packet detection unit), ヘッダ識別部 (header recognition unit), ヘッダ更新部 (header update unit), スイッチ部 (switching unit), それらを接続する光導波路と制御信号線からなる。いずれも RAM を用いずに構成される。

パケット検出部は、プリアンブルの識別によりパケットの到着を検出する機能と、ヘッダ識別部、ヘッダ更新部、スイッチ部の各部に対する駆動信号を生成する機能を有する。

ヘッダ識別部は、光回路あるいは電子回路による完全一致識別回路により転送制御情報およびオプションを識別する機能を有する。論理ゲートで構成される場合の転送制御情報の識別回路の例を図 3.8 に示す。識別回路は、読み取るべき転送制御情報の抽出回路 (signal extractor), 各出力ポートに対応した完全一致相関器 (correlator), 識別結果をスイッチへの制御信号に変換する回路 (encoder) から成る。オプションの識別も、同様の機構を用いる。

ヘッダ更新部は、2つのスイッチ ( $S_{in}$ ,  $S_{out}$ ) とそれを接続する長さの異なる光ファイバ導波路 ( $F_L$ ,  $F_S$ ) によって構成され (図 3.9), プリアンブルの遅延によりラベルの先頭を削除する機能を有する。導波路の光路長差がノードにおいて読み取る転送制御情報に相当する長さ  $c \cdot \frac{1}{b} \cdot \lceil \log_2 d(n_i) \rceil$  (ただし  $c$  はファイバ中の光の速度) となるよう、遅延線 (FDL) を調節する。入り口側スイッチ ( $S_{in}$ ) と出口側スイッチ ( $S_{out}$ ) は、パケット検出部からの駆動信号により切り替えが行われる。なお、波長多重パケット (図 3.6(b)) の場合は、更新部の前後でアレイ導波路回折格子や波長フィルタなどを用いてヘッダとペイロードの分離と多重化を行い、ヘッダ部分のみを更新部で処理する。

衝突回避機構では、異なる入力ポートから同じ出力ポートに転送されるパケット同士が衝突を起こす場合に、片方を時間的に遅らせるなどして衝突によるパケット損失を回避する処理を行う。衝突回避の制御は、ヘッダに記されたオプションの情報等を用いて行われる。衝突回避処理は転送処理とは独立に行われるため、具体的な構成や制御アルゴリズムについては本論文の対象外とする。

### 3.3.3 パケット転送処理

本機構における中継ノードでの転送処理は、以下に示す手順で行われる。

1. 入力パケットを光カプラで複製し、パケット検出部とヘッダ更新部に送る。
2. パケット検出部ではプリアンブルを検出し、ヘッダ識別部とヘッダ更新部に対する駆動信号を生成する。
3. ヘッダ識別部では、プリアンブルからそれぞれ固定の位置にある  $\lceil \log_2 d(n_i) \rceil$  ビットの転送制御情報と固定長のオプションを読み取り、スイッチ部を制御する。識別回路に入力されたパケットは、まず 1 ビット相当の遅延線を用いてシリアル・パラレル変換され、パケット検出部からの駆動信号によって抽出される。抽出された転送制御情報は各ビットパターンに相当する完全一致相関器により識別され、その結果に基づきスイッチ部への制御信号が生成される (図 3.8)。
4. ヘッダ更新部では、プリアンブルとオプションが FDL 側の導波路  $F_L$  に送られた直後にスイッチ  $S_{in}$  を切り替え、ラベル以降の部分の導波路  $F_S$  へ出力する。スイッチ  $S_{out}$  もプリアンブルとオプションが通過した直後に切り替えが行われ、プリアンブル、オプション、スイッ

チ切り替え後のラベル以降が出力される。プリアンブルとオプションが  $\lceil \log_2 d(n_i) \rceil$  ビット分遅延していることから、更新後のパケットはラベルの先頭にあった転送制御情報がちょうど削除された形となる (図 3.9)。なお、セーフティギャップ内には更新前の転送制御情報の一部が残存するなど任意の信号が残りうるが、この領域はラベル識別時には無視される領域であるので、転送処理上の問題はない。また、波長多重パケットの場合はヘッダから分離されたペイロードにプリアンブルおよびオプションと同じ遅延量を与え、ヘッダの更新後に合成する。

5. 更新済みのパケットを、すでにヘッダ識別部からの制御が完了しているスイッチ部経由で適切な出力ポートへ出力する。

各中継ノードで以上のような転送処理を繰り返していくことで、送信元ノードから送出されたパケットは徐々にラベル長が短くなりながら宛先ノードまで転送される。宛先ノードに到着した時点でラベルは全て削除されているため、ヘッダとペイロードが直列に構成されている場合 (図 3.6(a)) はセーフティギャップの後を抽出することでペイロードが復元される。ヘッダとペイロードが並列に構成されている場合 (図 3.6(b)) は、ペイロード部分の波長を電気変換すればよい。

### 3.3.4 特徴

本機構では、プリアンブル部分を遅延線により遅らせることでラベル更新を実現している。自己ルーティングの適用により中継ノードでの転送時における経路表検索が省略され、参照する転送制御情報も抑えられることにくわえ、本機構におけるラベル更新処理はスイッチと遅延線のみを用いた単純な構成と制御により実現されるため、光領域で容易に行うことができる [69]。また、プリアンブルおよびセーフティギャップを再利用することにより、ホップ数に依存して増加するヘッダの要素は必要最小限となる転送制御情報のみである。よって、ヘッダの長大化を抑えることができる。

## 3.4 他の自己ルーティングを用いた光パケット転送機構との比較

### 3.4.1 ラベル長の評価

本機構のヘッダ長が、適用ネットワークのトポロジ的特性やデバイス性能により受ける影響について、SRA および ARTEMIS と比較し評価した。ARTEMIS と本機構ではパケットが中継されるたびにヘッダが短くなるが、評価には送信元ノードからパケットが送出される際の初期ヘッダ長を用いた。

#### 想定環境

評価に用いた各パラメータの値を表 3.2 に示す。

トポロジ形状としてはべき法則トポロジと格子状トポロジの2種類を用いた。べき法則トポロジは、現状のインターネットにおける AS トポロジやルータトポロジを想定したもの [23] であり、リンクが集中する少数のハブ的なノードが存在する。格子状トポロジは、ポート数が少ない現状の

表 3.2: 評価に用いたパラメータ

トポロジ形状	べき法則 (Barabasi-Albert), 格子
経路	遅延をメトリックとする最短経路
ノード数	25, 100, 400, 2500, 10000
セーフティギャップ長	5ns, 0.5ns, 0.05ns
ヘッダビットレート	40Gbps, 80Gbps, 160Gbps
プリアンブル	2 ビット
オプション	なし
転送制御情報	$\lceil \log_2 d(n_i) \rceil$ ビット

光スイッチで構成されるネットワークの一例であり、特定のハブ的ノードが存在せず「ノードあたりのリンク数の平均と分散が小さい」という特徴を持つ。いずれも  $d(n_i)$  の平均は 4 である。ノード間の経路はリンク遅延をメトリックとする最短経路をダイクストラ法により求め用いた。

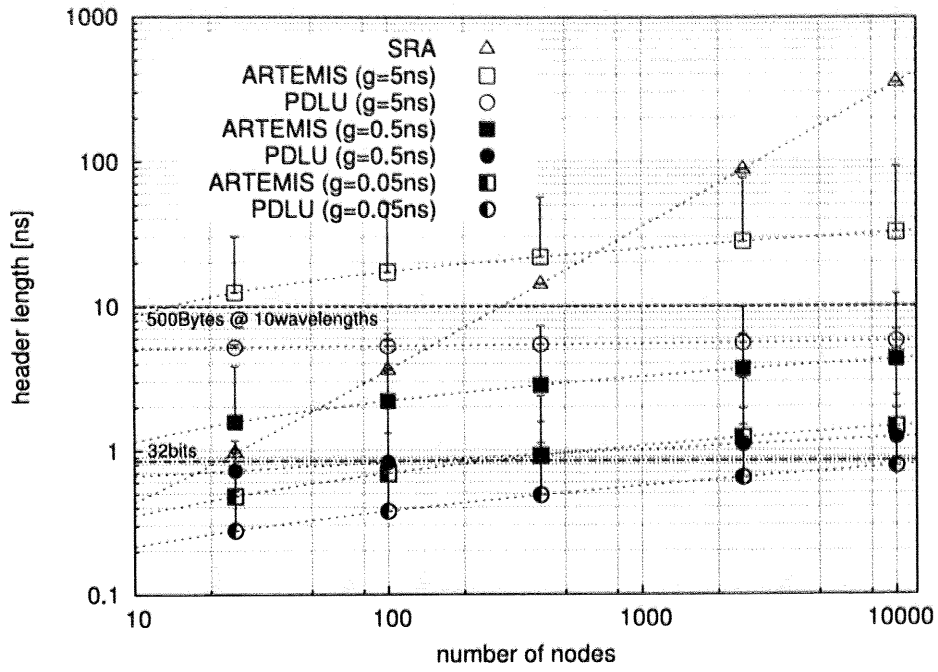
また、送受信デバイスやスイッチの性能の違いによる比較のため、セーフティギャップ長を 5ns から 0.05ns まで、ヘッダのビットレートを 40Gbps から 160Gbps まで変化させた。セーフティギャップ長は、実用化されている高速光スイッチの切替速度が数 ns、シミュレーションで達成されている切替速度が数十 ps [20] 程度であることから上記の範囲とした。また、ビットレートについては、オン・オフ変調による 1 波長あたりの伝送容量として実用化されているのが 40Gbps、実験的に達成されているのが 160Gbps [24] であることから上記の範囲とした。

ヘッダフォーマットに関して、プリアンブルは 2 ビット、オプションは 0 ビット、転送制御情報は出力ポートを識別する最小のビット長である  $\lceil \log_2 d(n_i) \rceil$  ビットに統一した。

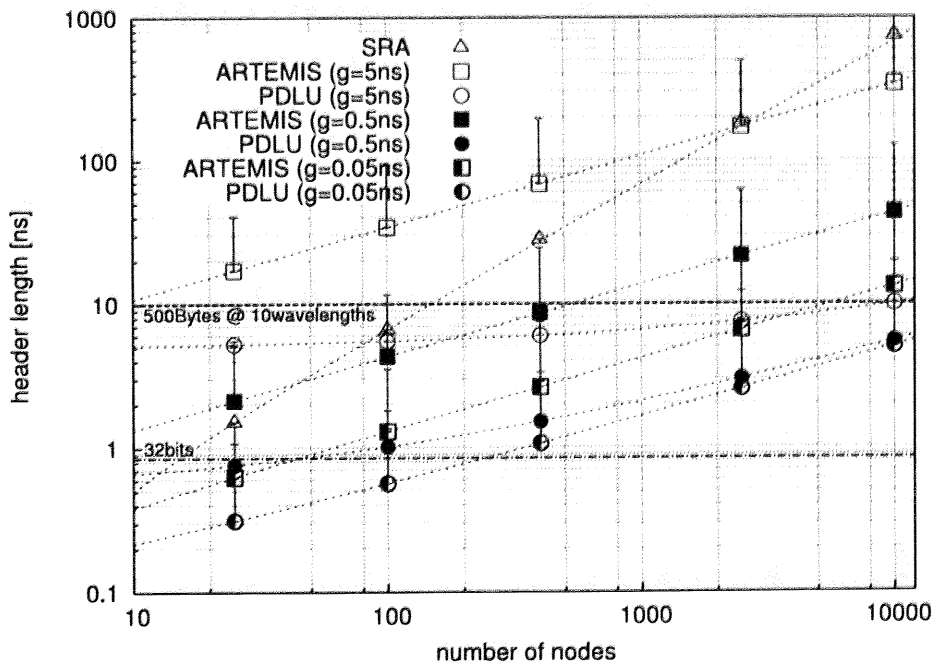
以上の条件においてシミュレーションを行い、各グラフに測定値をプロットし、最小二乗法により求めた近似曲線を点線で示した。また、各グラフには参考として、32 ビットのデータと 10 波長に分割多重された 500 バイトのデータを、それぞれヘッダと同ビットレートで変調した場合の時間を示してある。前者は、図 3.6(a) で IPv4 アドレスをラベルにした場合と比較して、本機構のヘッダが帯域利用においてどの程度のオーバーヘッドになるかの目安である。後者は、図 3.6(b) でペイロード長とヘッダ長の比較をする際の目安である。インターネットにおける平均パケット長が約 500 バイトであり、波長多重パケットの伝送実験が 10 波長程度で行われていることから [25, 73], これらの値を用いた。

### トポロジ特性とヘッダ長の関係

図 3.10, 3.11 にネットワークの総ノード数とヘッダ長の関係を、図 3.12, 3.13 に 10000 ノードのネットワークにおけるホップ数とヘッダ長の関係を示す。図 3.10, 3.11 では各ノード数における全経路の平均ヘッダ長を、図 3.12, 3.13 では各ホップ数毎の全経路の平均ヘッダ長を、それぞれ測定値としてプロットし、あわせて最大値を示している。なお、図 3.10, 3.11 で SRA のヘッダ長はセーフティギャップ長  $g$  によらないため、1 系列で表記している。また、図 3.12, 3.13 における SRA のヘッダ長はホップ数によらず、(a) では 353.52ns, (b) では 740.15ns である。他の手法

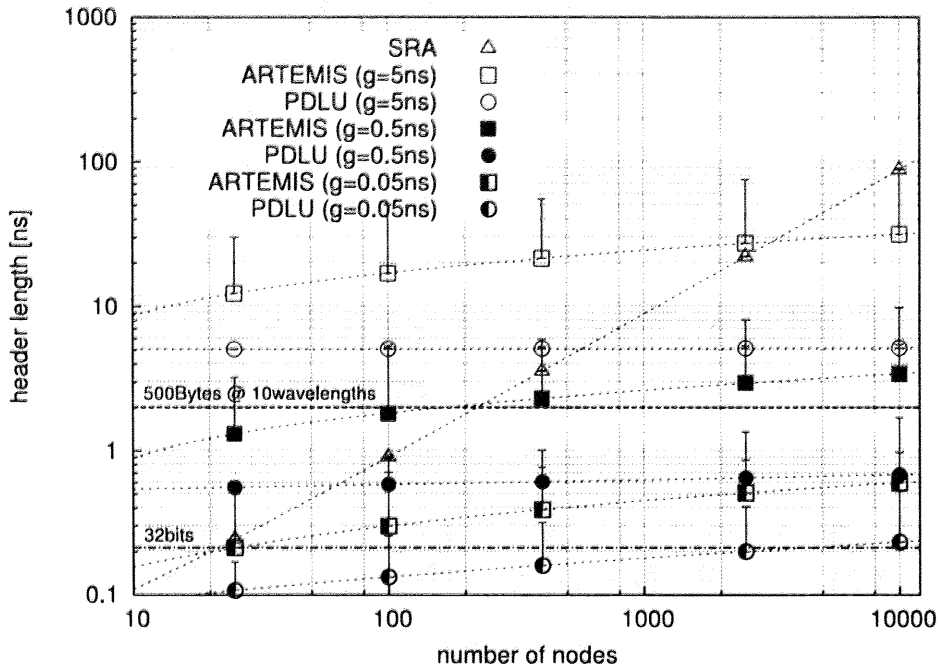


(a) べき法則トポロジ

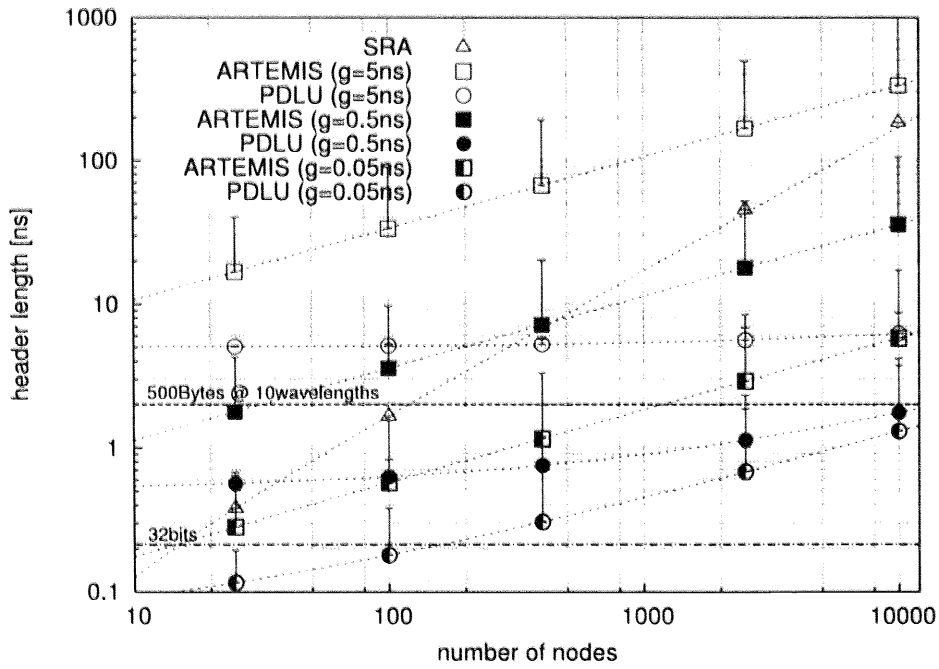


(b) 格子状トポロジ

図 3.10: ノード数とヘッダ長の関係 ( $b = 40\text{Gbps}$ )

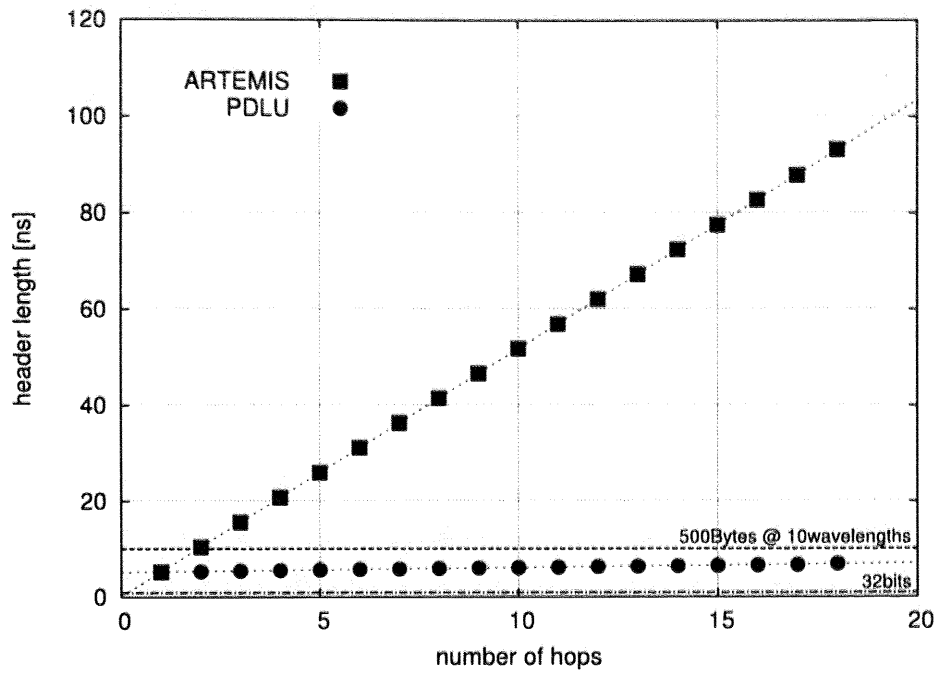


(a) べき法則トポロジ

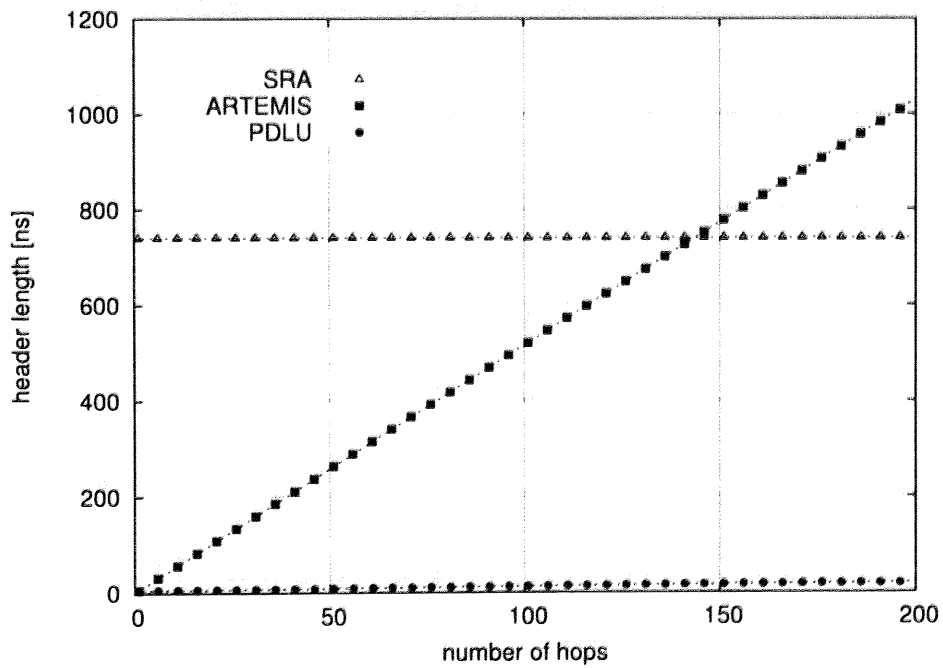


(b) 格子状トポロジ

図 3.11: ノード数とヘッダ長の関係 ( $b = 160\text{Gbps}$ )

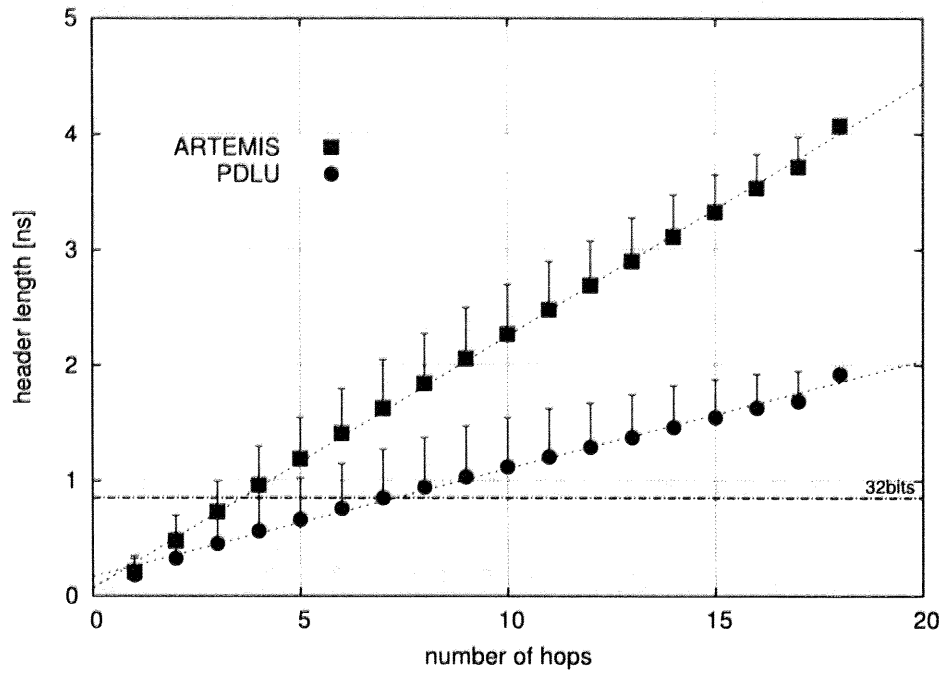


(a) べき法則トポロジ

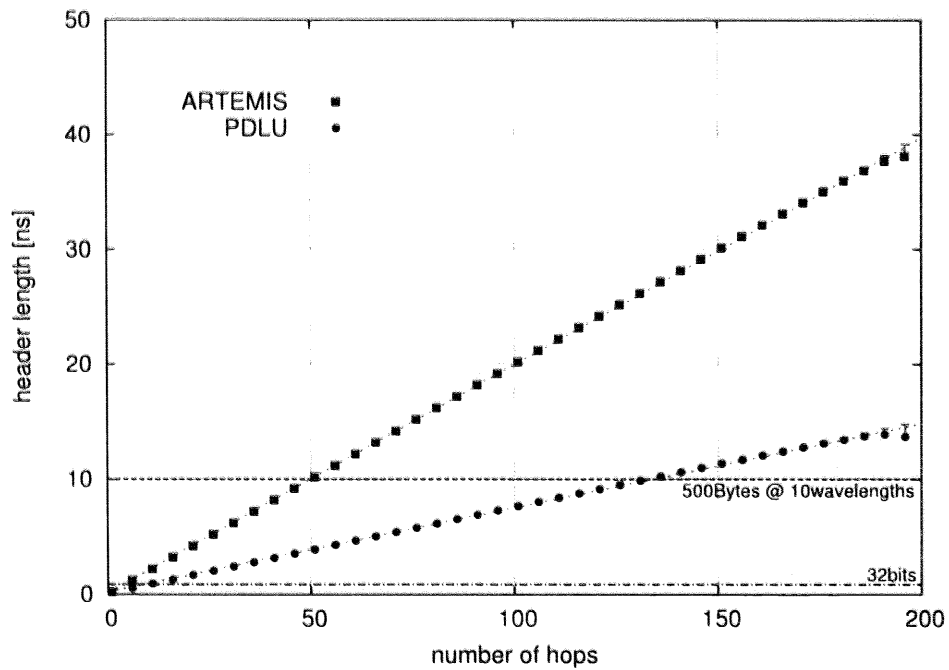


(b) 格子状トポロジ

図 3.12: ホップ数とヘッダ長の関係 ( $N = 10000$ ,  $b = 40\text{Gbps}$ ,  $g = 5\text{ns}$ )



(a) べき法則トポロジ



(b) 格子状トポロジ

図 3.13: ホップ数とヘッダ長の関係 ( $N = 10000$ ,  $b = 40\text{Gbps}$ ,  $g = 0.05\text{ns}$ )



に比べて十分大きい値であることから図 3.12(a) および図 3.13 においては表示を省略した。

それぞれのグラフから、SRA のヘッダ長がノード数に比例し、ARTEMIS と本機構 (PDLU) の初期ヘッダ長がホップ数に比例することが確認できる。また、ARTEMIS は本機構よりホップ数の増加に伴うヘッダ長の増加が大きい。図 3.10, 3.11 で (a) より (b) の方がヘッダ長が長くなっているのは、同じ規模のネットワークの場合に格子状トポロジの方が平均ホップ数が増えることに起因する。また、図 3.12, 3.13 で (a) の方が (b) よりヘッダ長が長くなるのは、べき法則トポロジにおける最短経路は接続が集中し  $d(n_i)$  が多くなっているハブノードを経由しやすいことに因る。

SRA と本機構のヘッダ長をノード数との関係で比較すると、SRA に比較して本機構のヘッダ長が最も長くなる { べき法則トポロジ,  $b = 160\text{Gbps}$ ,  $g = 5\text{ns}$  } の場合でも、564 ノード以上で本機構のヘッダ長の方が短くなる。ARTEMIS と本機構のヘッダ長をホップ数との関係で比較すると、最もセーフティギャップ長の影響が小さくなる { 格子状トポロジ,  $b = 40\text{Gbps}$ ,  $g = 0.05\text{ns}$  } の場合でも、本機構のヘッダ長は ARTEMIS の 37.5% に抑えられている。以上より、本シミュレーションで想定したパラメータの範囲では、特にネットワーク規模が大きくなる場合に、本機構のヘッダ長が他の手法に比べて短く抑えられることがわかる。

### デバイス性能とヘッダ長の関係

図 3.14 に、10000 ノードのネットワークにおけるセーフティギャップ長およびヘッダのビットレートとヘッダ長の関係を示す。測定値として、各ビットレートにおける全経路のヘッダ長の平均値をプロットしている。

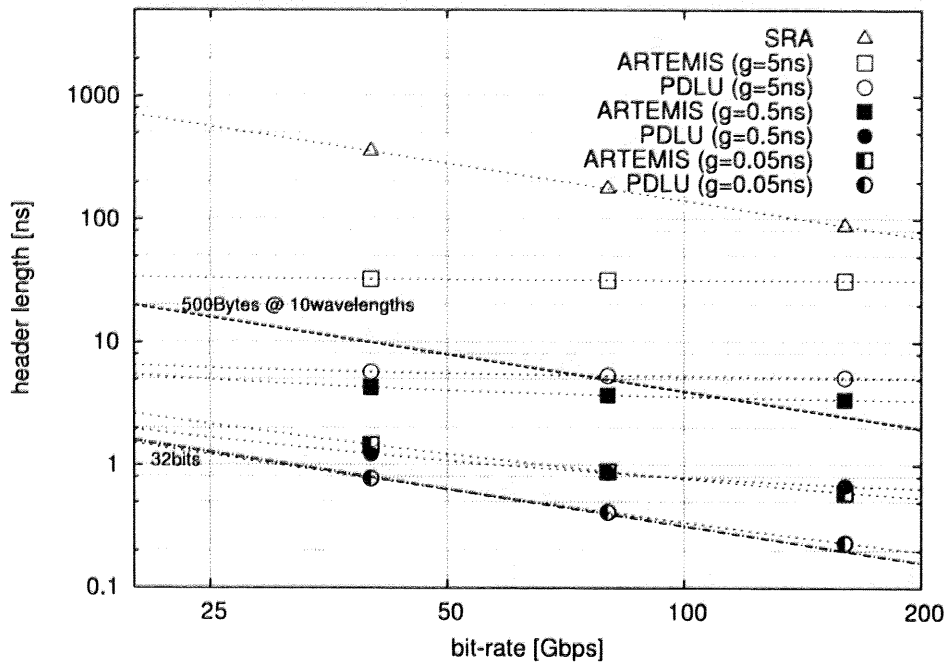
SRA はのヘッダ長はビットレートの向上により ARTEMIS や本機構よりも大きく改善されるが、それでもグラフに示したパラメータの範囲では本機構のヘッダ長を大きく上回る。

また、セーフティギャップの短縮によるヘッダ長の縮小効果は本機構より ARTEMIS の方が大きい。セーフティギャップ長が 1 ビットと同程度のオーダーに抑えられてもやはり本機構のヘッダ長の方が十分短く抑えられている。ARTEMIS においてプリアンブルやセーフティギャップの累積によるオーバーヘッドが無視できるためには、タグに占めるそれぞれの割合が十分小さくなる必要がある。しかし、転送制御情報は 1000 ポートのノードでも 10 ビットと大きくならず、現実的にはオーバーヘッドが無視できるほど小さく抑えられることはないと考えてよい。

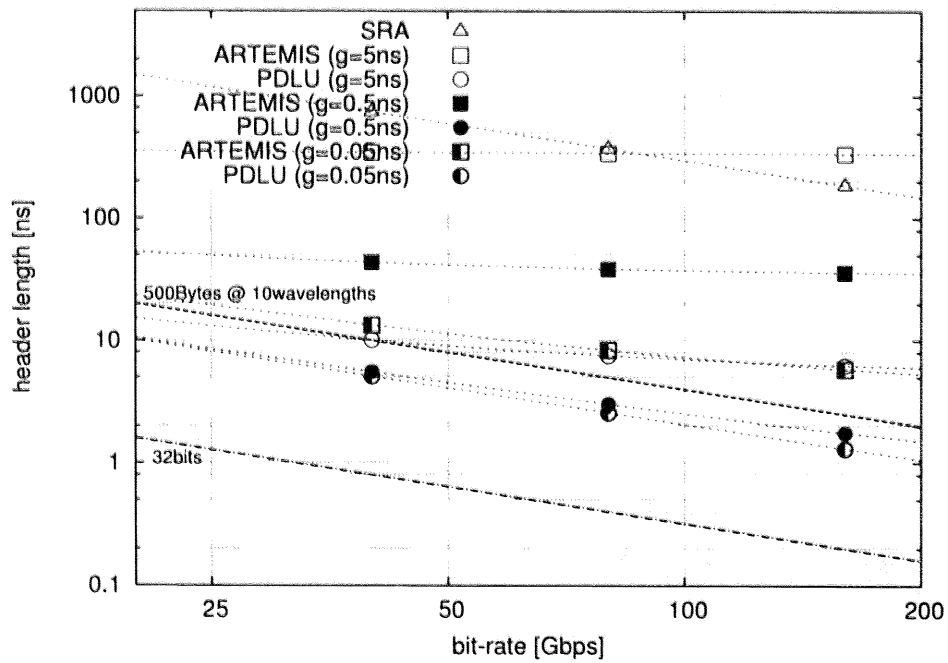
### 3.4.2 ラベル更新機構のハードウェア構成が伝送に与える影響

SRA, ARTEMIS, プリアンブル遅延型ラベル更新による光パケット転送機構のそれぞれについて、ラベルを読み取り解析する部分はいずれのラベル割り当て手法でもほぼ同等の構成で実現できる。自己ルーティングスイッチを用いていることから、各ノードで解析するビット数は出力ポートを識別できるビット数に収まる。読み取り位置も、必ずラベルの先頭であるか決まった位置 (読み取りノードに割り当てられたフィールド) に固定されているため、調整する必要がない。

各機構で違いがあるのはラベル更新を行う部分である。自己ルーティングアドレスを用いた光パケット転送では、ラベルの更新を行わない。それに対し、ARTEMIS では毎ノードでの転送時にスイッチにより既読のラベルを削除するという処理を行う。本機構では、ヘッダの一部を遅延させるために 2 つのスイッチを用いて信号の分解と合成を行う。さらに、遅延線によってパケットの一部



(a) べき法則トポロジ



(b) 格子状トポロジ

図 3.14: ヘッダ長におけるビットレート・セーフティギャップの影響 ( $N = 10000$ )

を遅延させるという処理を行っている。これらのデバイス構成的な特徴に起因する伝送信号への主な影響として、減衰による強度次元での信号劣化と、遅延量の誤差による時間次元での信号劣化が考えられる。これらは光領域をカットスルー転送されるパケット信号に対しノードを経由するごとに蓄積されていくものであるため、信号再生を行わない場合の転送可能ホップ数に影響する。

減衰の影響は、特に光スイッチの挿入損失が大きいことから光増幅器で信号強度を補償する必要が生じ、光信号雑音比 (OSNR: Optical Signal-to-Noise Ratio) が劣化することによるものである。これによりビット誤り率が上昇し、伝送品質が下がってしまう。スイッチの通過だけに着目すると、本機構ではラベル更新を行わない場合の3倍、ARTEMISの1.5倍のOSNR劣化が起こることから、それだけ転送ホップ数が限られる。

パケット信号の時間的なずれによる影響は、ラベル更新の際にプリアンブルを遅延させる回路の光路長差の誤差に基づくものである。プリアンブルの遅延量の誤差が大きくなると、ビットの消失や余分なビットの挿入によりラベルの解析やペイロードの復元が正しく行われなくなる。

これらラベル処理機構の構成に基づく信号劣化は本機構の大きな課題であり、デバイスコンポーネントの性能・品質向上や全光信号再生技術の発展による改善が求められる。

## 3.5 光パケット転送ノードの実装

本パケット転送機構の実現可能性を確認するため、ノードのプロトタイプ実装を行い、パケットの転送実験を行った。

### 3.5.1 ノードの設計

光パケット転送ノードおよび検証用サブシステムから構成される本実装環境の構成図を図 3.15 に示す。

#### 光パケット転送ノード

光パケット転送ノードはラベル解析部、ラベル更新部、スイッチング部の3つで構成される。

ラベル解析部は光検出器 (Photo Detector, PD) と FPGA で構成されている。PDにより光パケット生成部から到着するパケットを光信号から電気信号に変換し、FPGAへ送る。FPGAはパケットのプリアンブル信号を検出するとラベル更新部のスイッチにパケットの到着を報告する。そして自ノードで読み取るべき転送制御情報を読み取り、その結果を基にスイッチング部へ制御信号を送る。

ラベル更新部は、プリアンブルの分離と合成を行う2組のPLZT型光スイッチと長さの違う2本の光ファイバを基本的な要素として構成される。入り口側PLZT型光スイッチでは、ラベル解析部から送られてくるパケット到着の報告を基にプリアンブルの分離を行う。そしてプリアンブルのみ遅延量の多い方の光ファイバ遅延線 (Fiber Delay Line, FDL) に出力され、プリアンブル以下の信号は遅延させずに出口側のPLZT型光スイッチに送られる。FDLの長さは、その光パケット転送ノードにおける転送制御情報に相当する遅延を与えられるよう設定されている。出口側の

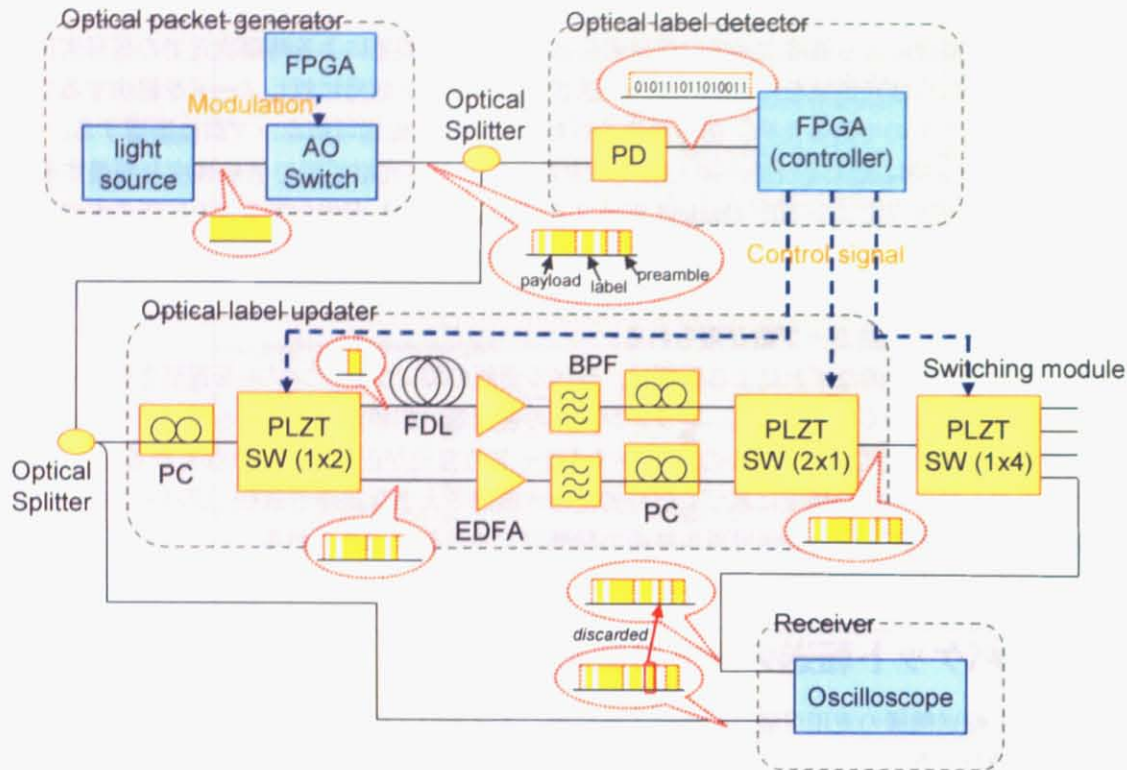


図 3.15: 実装環境の構成

PLZT 型光スイッチではプリアンブルとプリアンブル以下のパケットの合成を行い、合成された光パケットはスイッチング部に送られる。

さらに、パケット信号を調整するためのコンポーネントとして、光増幅器 (EDFA)、帯域通過フィルタ (BPF) および偏波コントローラ (PC) を実験系に含めている。光増幅器は PLZT 型光スイッチの通過より減衰した信号を増幅するため、帯域通過フィルタは PLZT 型光スイッチおよび光増幅器を通過する際に信号に乗るノイズを除去するために用いられている。また偏波コントローラは PLZT 型光スイッチが偏波依存することから信号の偏波調整のために用いられる。

スイッチング部は PLZT 型光スイッチで構成される。ラベル更新部から受け取った光パケットを、ラベル解析部の制御信号によって指定された出力ポートに出力する。

### 検証実験用サブシステム

サブシステムには光パケット生成部と検証部の2つがある。

光パケット生成部では、検証実験を行うための仮想的な光パケットを生成し、光パケット転送ノードへ光パケットを入力する。光パケット生成部は光源、Field Programmable Gate Array (FPGA)、Acoustic Optic Switch (AO スイッチ) で構成されている。まず光源で C バンド帯の単一波長の連続発振波を生成する。この連続発振波を FPGA で制御された AO スイッチを用いて On-Off 切

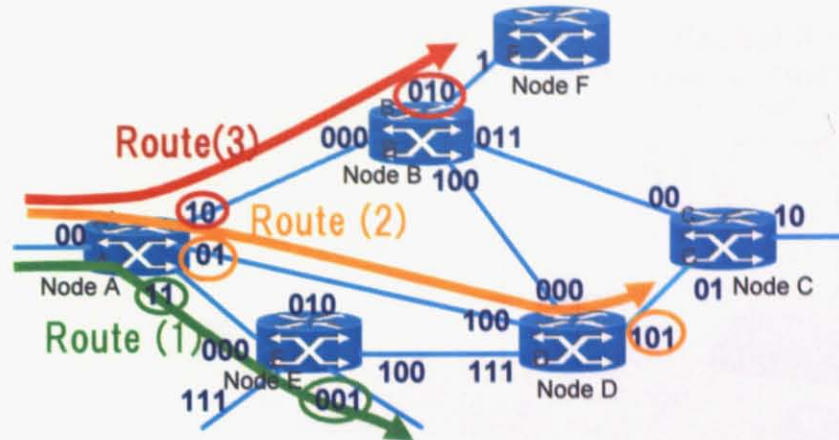


図 3.16: 実証実験の想定ノードトポロジ

替えし、強度変調された2値の信号を作り出す。その信号を光パケットとし、光パケット転送ノードに入力する。

検証部では、光パケット生成部で生成された光パケットと光パケット転送ノードから出力される光パケットを比較する。検証部はオシロスコープのみで構成される。

### 3.5.2 実験パラメータ

実験用光パケットには、2ビットのプリアンブル(“11”), 2ビット相当のセーフティギャップ、5ビットのラベル、1100ns~1300nsのペイロード(“11...1”)という書式を用いている。ラベルの5ビットの内、最初の2ビットが自ノードの転送制御情報、残りの3ビットが次ノードの転送制御情報である。光パケット転送は図3.16に示すモデルトポロジーのRoute(a), Route(b), Route(c)についてそれぞれ行った。ヘッダのビットレートは10Mbpsに設定した。ペイロードはノード内で参照されることなく転送されるため、ビットレートに関しては特に定めず連続発振波によって代用した。波長はCバンド帯単一波長(本実験では1554.13nm)を用いた。ラベル更新部内のFDLの長さは光路差が2ビット相当になるよう調節した。

### 3.5.3 実証実験

検証実験の様子を図3.17に、ノードAにおけるパケット転送処理の結果を図3.18に示す。

図3.18の各波形は、最上段が入力波形、以下それぞれの出力ポートにおける出力波形である。入力パケットは各転送制御情報に対応する出力ポートへ転送されていることが分かる。また出力パケットは、プリアンブルが2ビット分遅延させられたことによりラベルが更新され、セーフティギャップの直後が次ノードの転送制御情報(それぞれ“001”, “101”, “010”)となっている。

以上より、ノードにおけるラベル更新とパケット転送が正しく実現されていることが確認できた。

出力信号の波形に着目すると、特にセーフティギャップ部分の波形が乱れている様子が見取れる。本機構ではセーフティギャップの間でスイッチの切替えによりラベルの更新を行うため、更新

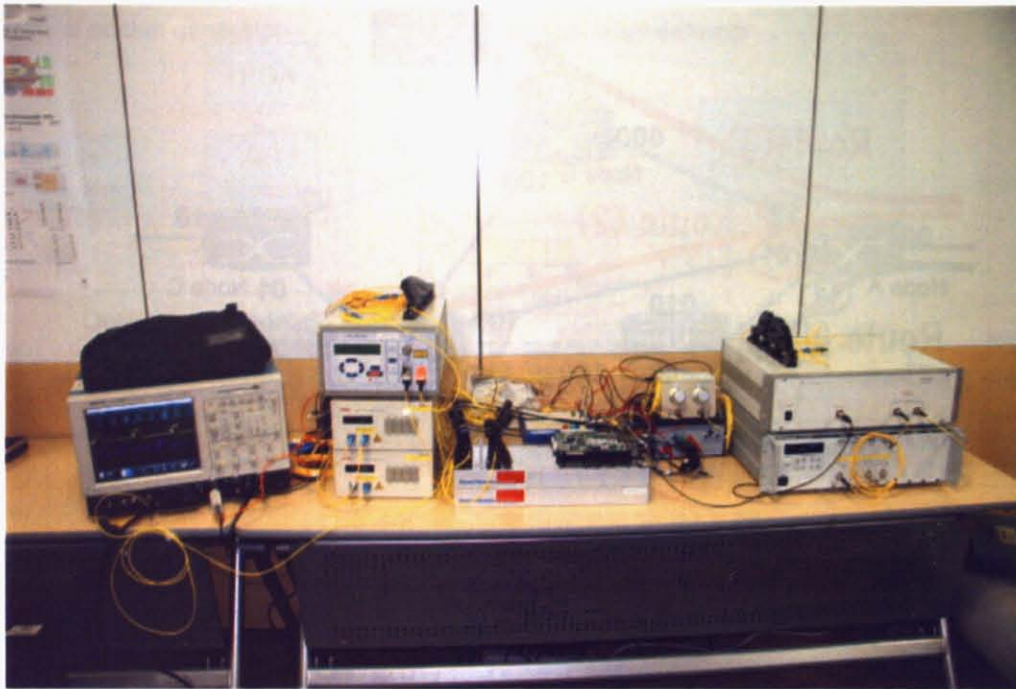


図 3.17: ラベル更新部の実験構成

前の信号の一部がセーフティギャップの部分に残存するためである。しかし、各中継ノードでのラベル参照時にはセーフティギャップの部分は無視されるため、ラベルの識別処理には影響しない。

本実験構成では、偏波を調整する必要があるものの、波長次元での処理を行わないため AWG の温度依存性や波長分散の影響を受けない。また、波長毎の処理コンポーネントが不要であるため、装置規模を十分小さく抑えることができている。動作環境への耐性や装置サイズといった光パケット交換ノードの実装における問題についても、本機構は十分現実的な環境で動作することが示された。

### 3.6 おわりに

本章では、プリアンブル遅延型ラベル更新を用いた光パケット転送機構について述べ、ラベル長とハードウェアコストに関する比較を行った。また、ラベル更新部の実証実験について述べた。

本機構は自己ルーティングスイッチによるマルチホップ転送を実現するために、中継ノードにおける出力インタフェース識別子を並べてラベルとする。そして、各ノードでの転送時にプリアンブル部分を固定長遅延させることによって既読部分を削除し、ラベルを更新する。これにより、各ノードでの出力インタフェース識別子の読み取り位置がプリアンブルに続く固定長の安全ギャップの直後に合わせられるため読み取り位置をパケット毎に調整することが不要となり、また遅延線とスイッチいう単純な構成のみでラベル更新が行われる。ラベルから出力インタフェースの識別子を削除する際、単純にスイッチの切り替えのみによりプリアンブルと既読部分を削除するのではな

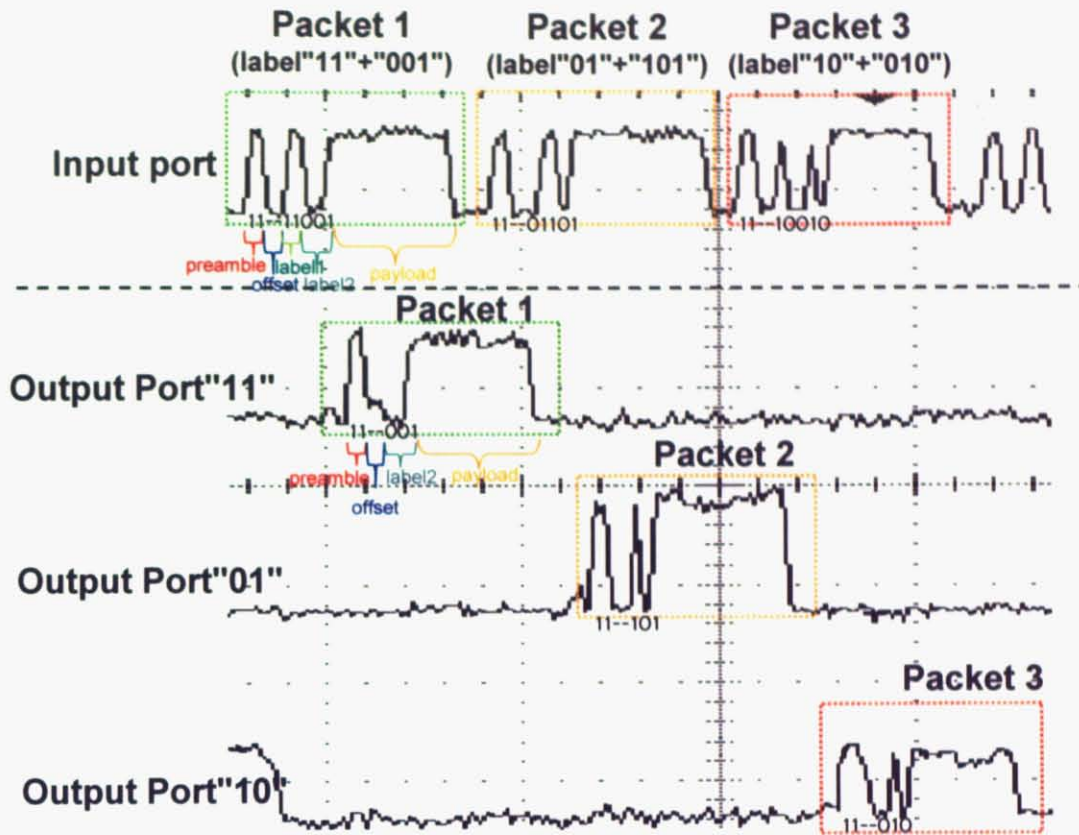


図 3.18: パケット転送動作時の信号波形

く、プリアンブルと安全ギャップを遅延線により遅延させ“再利用”することで、インタフェース識別子間のギャップを排除し、ラベル長を時間的に短く抑えられる。

既存の自己ルーティングスイッチを用いた光パケット転送機構である SRA および ARTEMIS との比較では、ラベル長に関して効率の向上をシミュレーションにより示した。

また、中継ノードのプロトタイプ実装を行い、特に光領域でのラベル更新が現実的なデバイス構成で正しく行われることを確認し、本パケット転送機構の実現性を検証した。

## ■ 第4章

---

# 光ファイバ遅延線を用いた衝突回避機構の評価



## 4.1 はじめに

パケット通信においては、ノードにおいて異なる入力パケットが同じポートに同時に出力されることによるパケット衝突が起こりうるため、その解決が不可欠である。しかし光 RAM がないことから、光領域では電気で行われているような柔軟なバッファリングを行うことは難しい。そのため、これまで光領域で可能な衝突回避処理に関する研究が進められてきた。例えば、波長変換、迂回ルーティング、遅延線バッファなどである。

本章では固定長の光ファイバ遅延線を組み合わせることで実現される遅延線バッファを用いた衝突回避機構について、その構成に着目した性能評価について述べる。遅延線バッファは固定的な物理構造を持つためバッファリング時間の制御が離散的な固定時間になることから、適用するネットワークのトラフィックパターンによってその性能が左右される。そこで、実トラフィックを含むいくつかのトラフィックモデルを適用したシミュレーションにおいて、遅延線バッファの構成（単位長・深度）が、パケット損失率や遅延などネットワーク特性にどのように影響するかを評価した。

以下、4.2 で既存の光領域衝突回避の実現方法を述べ、遅延線バッファを用いる方法の光パケット交換ネットワークに対する適性について説明する。4.3 では、遅延線バッファを用いた衝突回避機構の構成要素とバッファ性能への影響について述べる。そして4.4 でバッファの物理的構成とトラフィックパターンとの関連についてシミュレーション評価した結果を示し、現在のインターネットと同等なトランスポート性能を実現するための構成やコストについて論じる。最後に4.5 で本章をまとめる。

## 4.2 光パケット交換ネットワークにおける衝突回避機構

通常の電気処理によるスイッチやルータでは、入力パケットをキューに一時的に格納しておき、スケジューリング制御を行ったうえで順次出力処理が行われる。このようなストア・アンド・フォワード型の転送処理の性質上、パケット衝突の問題は特別な対処をしなくとも自然に解決される。

しかし、光領域でのパケット転送では RAM が無いため柔軟なバッファリングが困難であることから、異なる入力ポートからの到着パケットが同時に同じ出力ポートへ転送される場合にパケットの衝突が発生してしまう。非同期にデータを送受信するパケット交換ネットワークにおいてはこのパケット衝突は頻繁に起こるものであり、パケットの損失率を実用的な範囲にとどめるには、何らかの衝突回避機構が不可欠となる。

### 4.2.1 光パケット交換ネットワークにおけるノード内衝突回避処理

ノードでパケット衝突が発生した際にそれを解決するための基本的なアプローチとして、これまで「波長変換」「迂回ルーティング」「遅延線バッファ」の3つが検討されてきている（図 4.1）。

**波長変換** 衝突するパケットの片方の波長を別の空き波長に変換することで、波長次元での衝突回避を実現するものである（図 4.1(a)） [26]。遅延・ジッタ・逆転などトランスポートに対する影響が小さい、ノード内の制御のみで衝突回避処理が完了するといった利点がある。逆に欠点としてはリンクの波長多重度が上がるほど必要な波長変換器が増え装置コストが大きくなることがある。

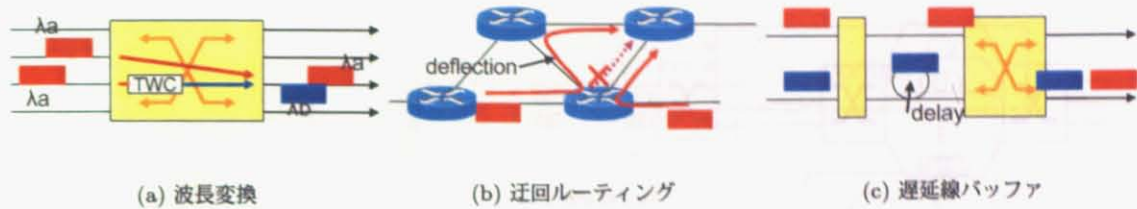


図 4.1: 光領域における衝突回避手法

**迂回ルーティング** 衝突するパケットの片方を別の隣接ノードへ転送することで、空間次元での衝突回避を実現するものである(図 4.1(b)) [27]. 迂回ルーティングはハードウェア的な変更を全く必要とせず、各中継ノードでの転送制御アルゴリズムを変更するだけで実現される. 欠点としては、効率的な迂回アルゴリズムがなくトラフィック負荷の上昇を起こすこと、遅延・ジッタなどパケットの伝送時間の変動がリンク伝播遅延規模で生じるため非常に大きくなることがあげられる.

**遅延線バッファ** 衝突するパケットの片方を時間的にずらすことで、時間次元での衝突回避を実現するものである(図 4.1(c)) [28-30]. 利点としては、波長変換と同様にノード内の制御のみで衝突回避処理が閉じること、また波長に依存しないため多波長パケットに適用可能といった点が挙げられる. 逆に欠点としては、迂回ルーティングほどではないが遅延・ジッタ・逆転などトランスポートに対する影響が起りうる.

これらの方式はそれぞれ異なる次元で衝突回避を行うものであるため、実際の衝突回避機構の研究では組み合わせによる機構が用いられることがほとんどである [31,32].

しかし、波長変換と迂回ルーティングには、光パケット転送における高速・大容量化の技術と相性が悪いという問題がある. 波長変換は空き波長があることが前提であるため、例えば多波長パケット [22] のようにパケットのエンコード時点で波長次元での多重化を行い大容量化を図ると、波長変換による衝突回避はできなくなる. また、中継処理の高速化のために自己ルーティングのような転送方式を用いている場合、各ノードがグローバルな識別子に基づいた転送を行っていることが前提の迂回ルーティングを適用することはできなくなる.

それらに対し遅延線バッファによる衝突回避は、波長多重された信号の一括処理に対応しているうえに、ノード内で衝突回避処理が閉じるため転送経路の変更を必要としない. 他の手法に比べて光ファイバの容積により装置サイズが大きくなる可能性があるが、光ファイバ導波路の小容積化技術も発達してきており、またバッファ対象がパケットという短いデータ単位であることから、装置サイズは特に問題にならなくなっている. 以上より、遅延線バッファを用いた衝突回避機構は、光パケット交換ネットワークへ最も汎用的に適用できるものであるといえる.

#### 4.2.2 遅延線バッファを用いた衝突回避機構

遅延線バッファを用いた衝突回避機構は、さらにコンポーネントの物理トポロジー形態から図 4.2 に示す 3 種類に大別される.

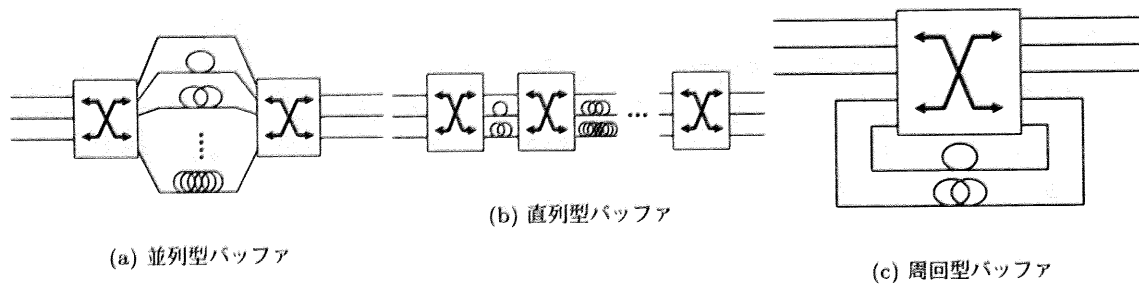


図 4.2: 遅延線バッファを用いた衝突回避機構

並列型バッファは長さの異なる遅延線を並べ、入り口側スイッチによりパケットに与える遅延量を制御するものである。遅延量は入り口側スイッチでの制御1回のみによって決定される。また制御が柔軟でない分、必要となるファイバ遅延線の量も他の方式に比べて多くなる。反面、制御は単純であり、構成としてもスイッチの経由数が決まっていることから信号の劣化に関して有利である。

直列型バッファ [33] はスイッチを長さの異なる遅延線で多段に繋いだ構成を持ち、各スイッチの切替えによりパケットに与える遅延量を制御するものである。上限はあるものの、可変時間の遅延を擬似的に実現できるものである。欠点としてはスイッチ数が多くなることから装置規模の増大や信号劣化の問題がある。

周回型バッファはスイッチングマトリックスの出力の一部を遅延線を経て入力に戻す構成により、衝突により出力ポートへ送出できなかったパケットを再び入力パケットとして扱えるようにするものである。ファイバ資源を最も有効活用できるが、スイッチの経由回数が増えることから信号劣化に対する制御が複雑になる。また、衝突回避処理されたパケットは再び入力パケットとなってスイッチに入力されることから、バッファ溢れを防ぐような制御も必要になる。さらに、制御対象となる高速スイッチも、ある程度の規模を求められることになる。

全光信号再生技術が十分でない現状では、光領域のパケット転送においては信号劣化をできるだけ抑えることが望ましい。スイッチを不定回通過する直列型および周回型は、制御の柔軟性はあるものの信号の減衰が大きくなってしまふ。そのため、本研究では並列型バッファによる衝突回避機構に注目し、以下で特性の評価を行う。

### 4.3 並列型バッファによる衝突回避機構の構成

遅延線バッファを用いた衝突回避機構を特徴付ける要素は、物理構成と制御アルゴリズムに大きく分けることができる。図 4.3 に非同期可変長パケットのための並列バッファ型衝突回避機構の概略を示す。

物理的構成は具体的には、バッファの単位長、バッファ長分布、バッファの深度などになる。

単位長は、バッファの遅延量の粒度を特徴付ける量である。単位長が大きすぎると遅延・ジッタの増大の原因になり、逆に小さすぎると衝突回避のために必要なバッファ深度の増加を引き起こす。

バッファ長分布は、バッファによる遅延量がどのようなパターンになるかを特徴づけるものであり、この分布とトラフィックパターンの関係によりバッファ資源の利用効率が決まる。

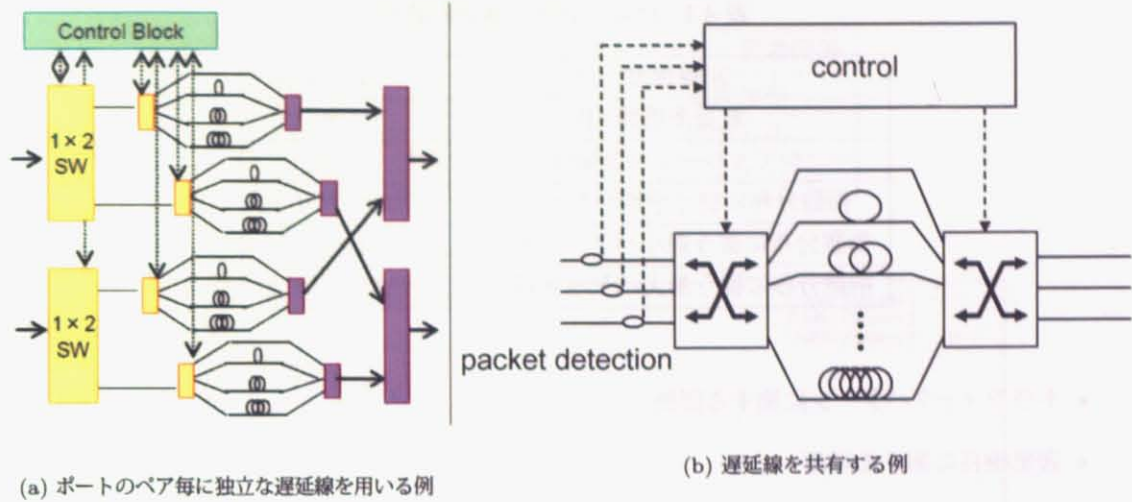


図 4.3: 光ファイバ遅延線による並列バッファ型衝突回避機構

深度はバッファの段数に相当し、総バッファ資源量に関する。深度が大きいほど損失を抑えることができるが、装置サイズの増大や遅延・ジッタの最大量の増加につながる。

アルゴリズムについては、どのような情報を用いるか、どのような規則でスケジューリングを行うかにより、その制御の特徴が決まる [34]。

具体的には、到着パケットの長さだけ取得し、到着順に既にスケジュールされているパケットと衝突が起こらない最も短い遅延線を選択してスケジュールをパケット毎に逐次確定していく方法 [35] や、一定期間将来までのスケジュールを取得し、衝突が起こらないスケジュールの空白を埋めるように遅延させる方法 (void filling [36,37]) といった方法が検討されている。前者はスケジュールの複雑さがなく、後者は効率的な資源利用を可能とすることが特徴である。

また、パケット長の取得方法についても、光領域の転送処理では電気とことなり柔軟な情報取得ができないことから、様々な方法が考えられる。粗い粒度でパケット長情報をヘッダに記述する、制御チャンネルにパケットの先頭と終端を示すなど、取得情報の粒度やタイミングも考慮する必要がある。

#### 4.4 インターネットトラフィックにおけるバッファ構成と衝突回避性能の関係

バッファの構成の違いが衝突回避性能にどのように影響するかについて、シミュレーションにより評価した。

##### 4.4.1 評価項目

ここで評価の対象とするのは、以下の4つの項目である。

表 4.1: シミュレーションの諸元

回線速度 ( $C$ )	10.0 Gbps
波長多重数 ( $W$ )	4
シミュレーション時間 ( $t$ )	60 s
指数分布に従う平均パケット長 ( $l$ )	512 バイト
指数分布に従う最小パケット長 ( $l_{min}$ )	64 バイト
指数分布に従う最大パケット長 ( $l_{max}$ )	1514 バイト

- トラフィックパターンに関する評価
- 遅延線長に関する評価
- 遅延線数に関する評価
- ポート数に関する評価

トラフィックパターンに関する評価では、入力トラフィックの特性の違いによって、どのような衝突回避性能やトランスポート性能の違いが生じるかを、バッファ構成との関連から明らかにする。

遅延線長に関する評価では、バッファにより生じる遅延の粒度が変わることによって、どのような性能差が生じるかを示す。

遅延線数に関する評価では、入力負荷とパケット損失率の関係を示し、所望のパケット損失率を得るために必要なバッファの総資源量について明らかにする。

ポート数に関する評価では、実際の高速光スイッチの制約に基づき実システムでの衝突回避性能を評価する。

#### 4.4.2 シミュレーション環境

表 4.1 にシミュレーションの諸元を示す。上記に示したパラメータ以外に遅延線の本数 ( $M$ )、到着負荷 ( $\rho$ )、ポート数 ( $N$ )、遅延線係数 ( $k$ ) がある。これらのパラメータを適宜変化させたときの遅延線バッファの特性を見る。なお、ポート数は特に断らない限り  $2 \times 2$  構成であるとする。

#### 4.4.3 トラフィックパターンに関する評価

表 4.2 に示す 3 つのトラフィックパターンの比較を行う。

##### パケット損失率

図 4.4 に負荷 20%、 $k = 0.1$  において遅延線数を変えた場合の遅延線数とパケット損失率の関係を示す。

表 4.2: 評価トラフィックパターン

	パケット長	パケット到着間隔
パターン1	指数分布	指数分布
パターン2	実トラフィック	指数分布
パターン3	実トラフィック	実トラフィック

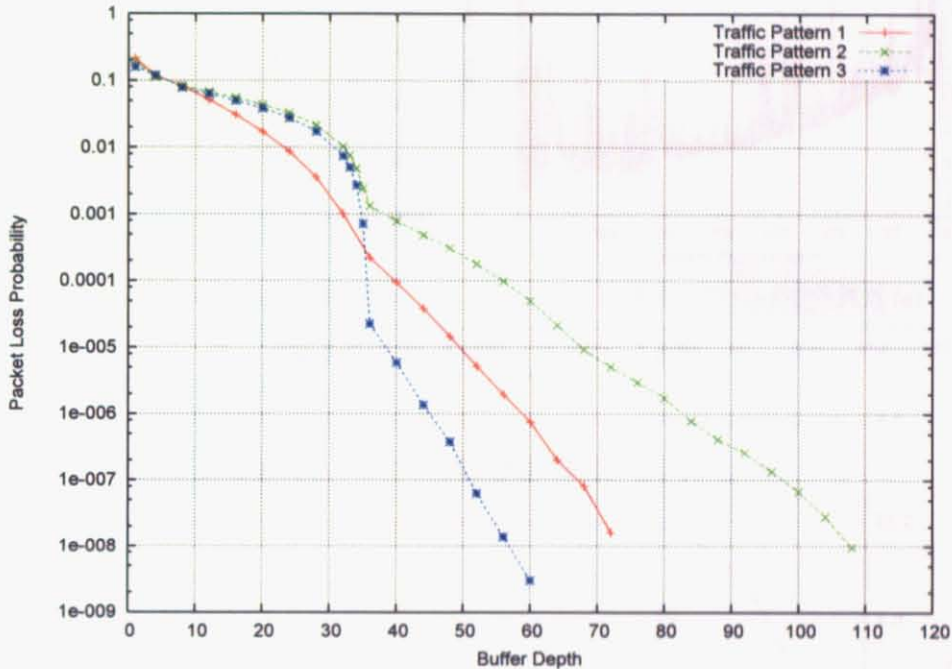


図 4.4: 遅延線数とパケット損失率の関係 ( $\rho=20\%$ ,  $k = 0.1$ )

負荷 20%は IP ネットワークのバックボーンにおけるリンクの平均使用率である [38]. まず, トラフィックパターン 2, 3 に関して, 30~36 段におけるパケット損失率が急激に良くなっていることが分かる. この傾向を分析するために, 図 4.5 に波長多重後の実トラフィックデータのパケット長分布を示す.

図 4.5 に示すように, Ethernet の MTU(波長ごとに 3,028 ビット) を格納できる遅延線の段数は 30 段目 (384 バイトを遅延可能) 以降である. そのため, Ethernet の MTU を格納できるだけのバッファ容量があればバッファ性能は向上することが分かる. なお, 実トラフィックの平均パケット長は 417.8 バイトであるが, 遅延線の単位長  $D$  は 512 バイトとしている. 100 バイト程度の遅延線長の違いによる遅延線バッファの性能差はほぼ無いことを確認しているため, この設定は評価には影響しない.

また, 段数が 12 段目から 36 段目の間では相対的に短いパケットが多く, 指数分布パケット長の方が損失しにくい傾向を示している. さらにパターン 2 と 3 がほぼ同じ値をとっていることから, MTU を格納できる段数が設けられるまではパケット長分布による影響に依存しやすいと考え

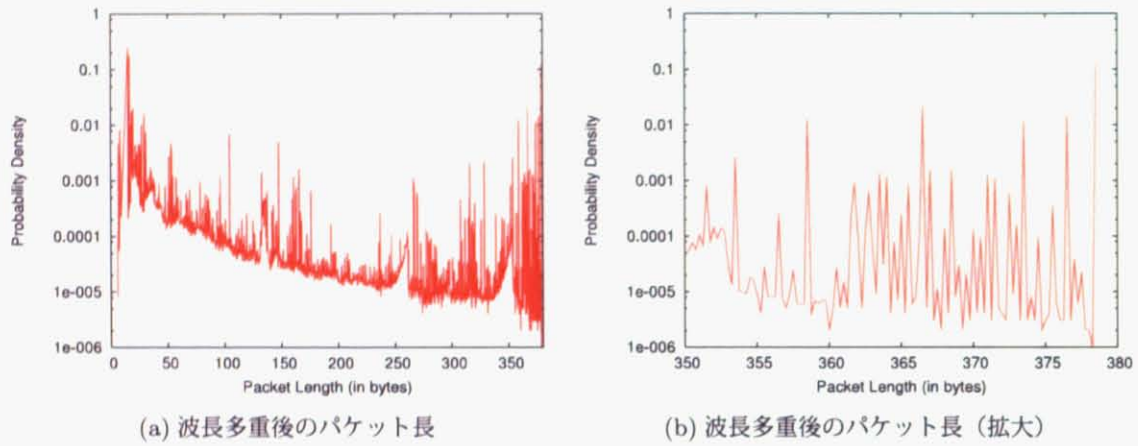
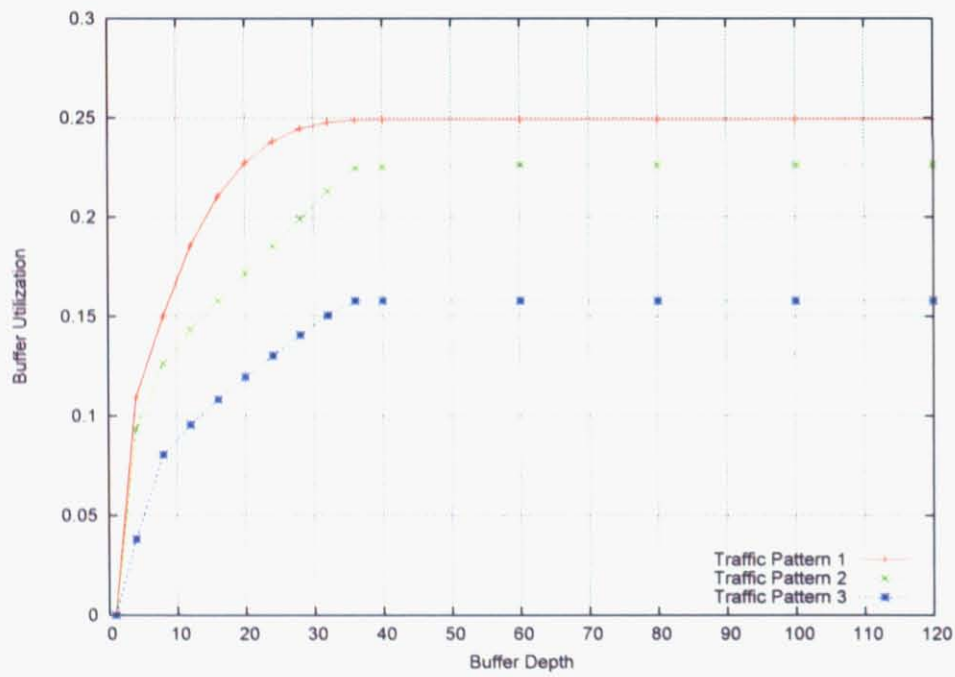


図 4.5: パケット長分布 (4 波長多重)

図 4.6: 遅延線数とバッファ使用率の関係 ( $\rho=20\%$ ,  $k=0.1$ )

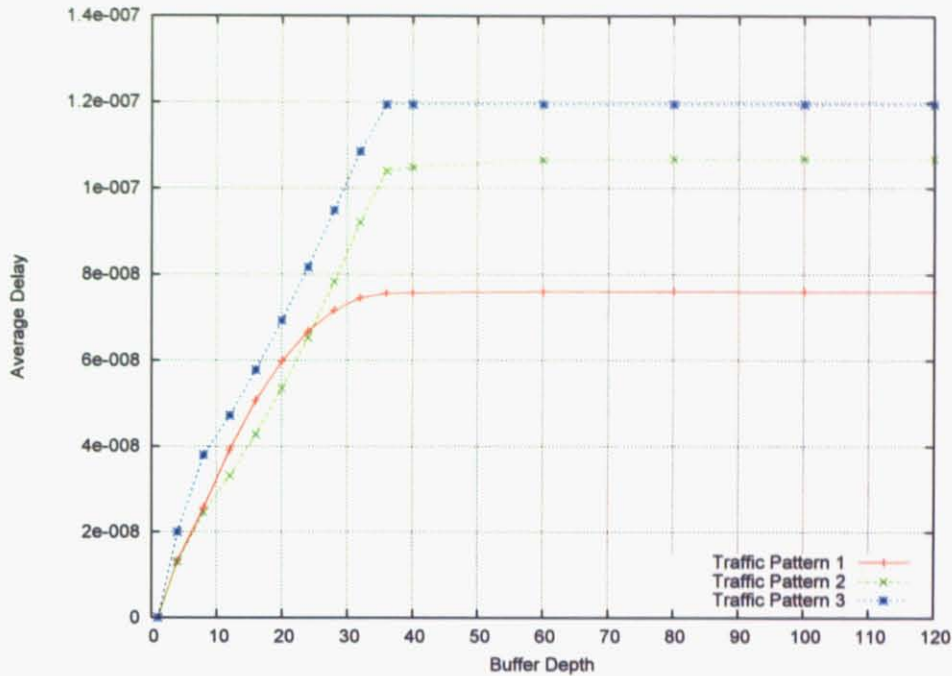


図 4.7: 遅延線数とバッファ平均遅延の関係 ( $\rho=20\%$ ,  $k=0.1$ )

られる。

#### バッファ使用率

図 4.6 にバッファ使用率を示す。ここで、バッファ使用率をノードから受信したパケットが  $M$  段構成のバッファにおいて、 $d_1$  から  $d_{M-1}$  のバッファに入った数の割合と定義する。これより、図 4.4 においてパケット損失率が急激に良くなっていた 36 段以降、バッファ使用率は一定となることが分かる。これは、MTU を満たす段数がある程度増やしていくと、衝突回避に十分な段数が確保されることを意味する。また、衝突回避アルゴリズムが遅延線が短いものから順番に選択していく方式をとっていることから、遅延線長が大きいものは比較的使用される可能性は低い。一定となるときにバッファ使用率は、パターン 1 は 25%、パターン 2 は 23%、パターン 3 は 16% であった。そのため、全送信パケット数のうち、大多数のパケットは  $d_0$  の遅延線を通過していることになる。バッファ使用率がパターン 1、パターン 2、パターン 3 の順番になっているのは、実トラフィックのパケット長分布とパケット到着間隔が偏った分布をしているため、バッファに蓄積する必要が比較的小さいためであると考えられる。

#### バッファ平均遅延

図 4.7 にバッファ平均遅延を示す。バッファ平均遅延はパケットが通過したファイバ遅延線長の平均に等しい。バッファ使用率と同様に、遅延に関しても 36 段以降は一定となることが分かる。



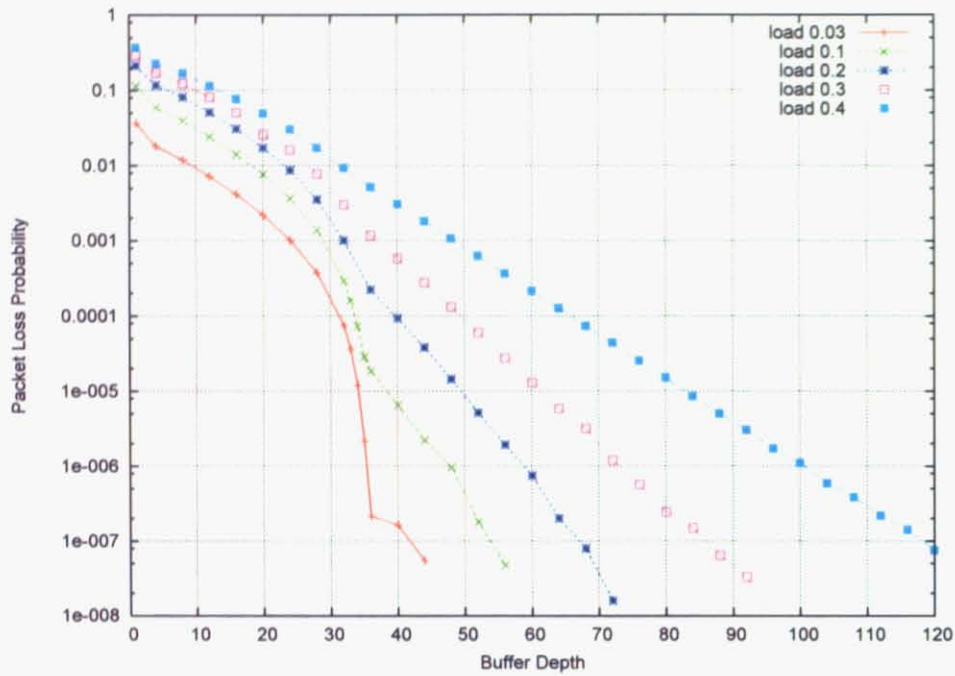


図 4.8: パターン 1 における遅延線数とパケット損失率の関係 ( $k = 0.1$ )

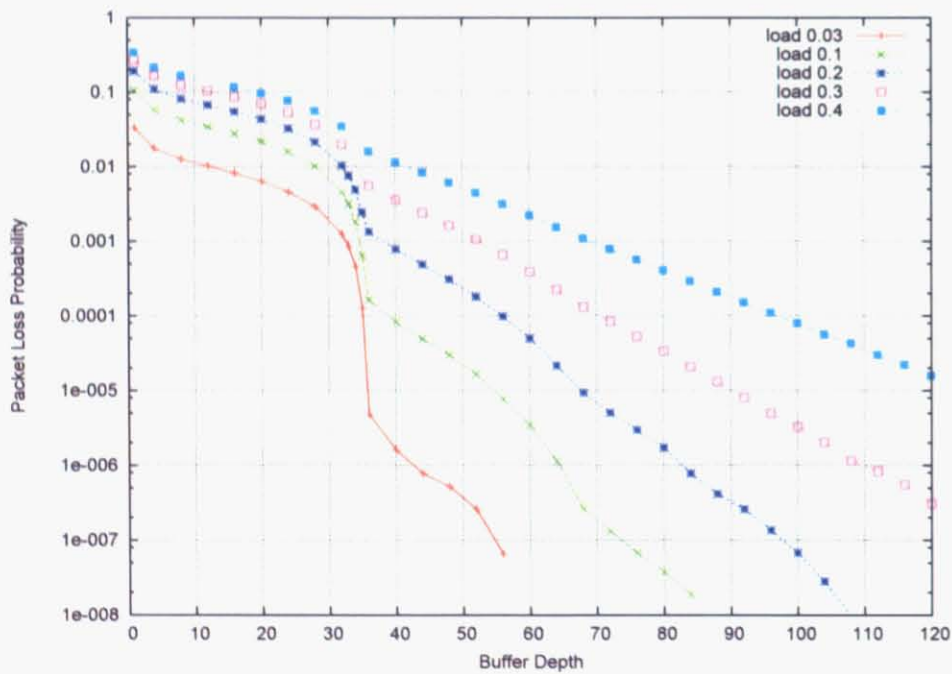


図 4.9: パターン 2 における遅延線数とパケット損失率の関係 ( $k = 0.1$ )

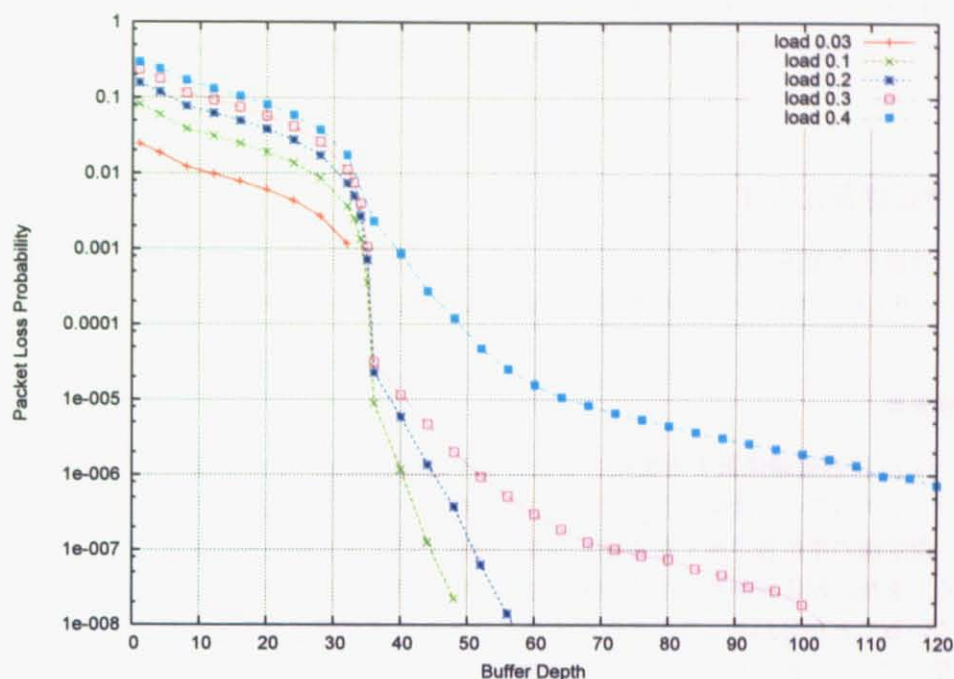


図 4.10: パターン 3 における遅延線数とパケット損失率の関係 ( $k = 0.1$ )

また、段数が増えるとパケット損失率は改善されるが、遅延線数がある一定以上の場合には遅延が一定となる。これより、平均遅延が一定になる時のパケット損失率が全体のパケット数との比率と考えるとほぼ無視出来る程度のパケット数であるため、1パケットが落ちたところで平均遅延には影響しないと考えられる。平均遅延がパターン3、パターン2、パターン1の順番になっているのは、実トラフィックではMTU付近のパケットが大多数を占めており、これらがバースト的に到着する傾向があるからであると考えられる。

なお、平均遅延が120ns程度であるので、バッファ内で発生する遅延はノード全体から見るとPLZTスイッチなどによる処理の遅延の方が影響が大きいため、無視できると判断出来る。但し、遅延線の単位長が大きくなり、ms単位の遅延が発生するような場合は無視出来なくなる。

#### トラフィックパターンごとのパケット損失率の比較

図 4.8, 4.9, 4.10 に遅延線数とパケット損失率の関係において、負荷を変化させた場合の特性を示す。いずれのトラフィックパターンにおいても、負荷が小さい場合はMTU付近において損失率が良くなる事が分かる。負荷が高くなる程、実トラフィックと統計分布モデルトラフィックの差が大きくなる事が分かる。特に、パターン3では負荷が30%を超えるとバッファの段数を増やしても損失率が改善されない事が分かる。

以上より、実トラフィックではトラフィックに特有のパケット長間隔、パケット到着間隔があることから、パケット損失率や遅延に関して統計分布モデルトラフィックと差が出る事が分かった。特に、近年においてトラフィック増加やアクセス速度が向上している現状を考慮すると、帯域の増

強に対してトラフィックの負荷が大きくなる傾向があるとも考えられる。そのような場合は統計分布モデルトラフィックとの差がさらに大きくなることから、実トラフィックによる性能分析は有用であると言える。

#### 4.4.4 遅延線長に関する評価

遅延線長は波長多重後の平均パケット長を  $D$  とし、遅延線の単位長  $kD$  とすると、 $d_1 = kD$  と表せる ( $k = 0.1, 0.2, 0.3, \dots$ )。すなわち、遅延線  $d_n$  の長さは  $n \cdot kD$  となる。

##### パケット損失率

図 4.11~4.13 に遅延線係数  $k$  を変化させたときの遅延線数とパケット損失率の関係を示す。なお、トラフィックパターンは実トラフィックによるものである。まず、全てのグラフに共通する特徴として、遅延線の単位長が長くなるほど少ない段数でパケット損失率を改善することができることが分かる。また、遅延線係数が小さい場合は MTU を格納できる段数になる前後でグラフの形状が異なっていることが分かる。この係数が大きくなるとほぼ直線になる。一方で、ある程度単位長が大きくなっていくと前の単位長とさほど性能が向上しないことが読み取れる。

次に、図 4.13 において、他のグラフと比べると形状が異なっている。負荷が高くなるほど、MTU を格納可能な段数を満たすと  $k = 0.3$  の時が最も性能が悪いことが分かる。

##### バッファ平均遅延

図 4.14~4.16 に遅延線係数  $k$  を変化させたときの遅延線数とパケット損失率の関係を示す（トラフィックパターンは実トラフィック）。負荷を 10% から 30% まで変化させたとしても、遅延はほとんど変化しないことが分かる。単位長に着目すると、0.1 倍から 1.0 倍まで変化させても 60ns 程度の遅延差であるため、ほぼ無視できると言えるが、ある程度単位長が大きくなると前の単位長との遅延差が大きくなっていくと考えられる。

##### 光デバイスの制約を考慮したシステム評価

コアノードの構成において、適切な遅延線への振り分けを行う  $1 \times M$  スイッチは高速な PLZT 光スイッチを想定している。PLZT 光スイッチは、高速なスイッチング速度と偏波無依存性等の利点がある一方、挿入損失が大きく規模性に問題がある。現状では 32 ポート程度が限界と考えられる。このようなスイッチの制約に着目すると、図 4.11, 4.12, 4.13 より、いずれの負荷についても平均パケット長の 0.2 倍の時が最適値をとることが分かる（パケット損失率は  $10^{-6}$  を基準とする）。

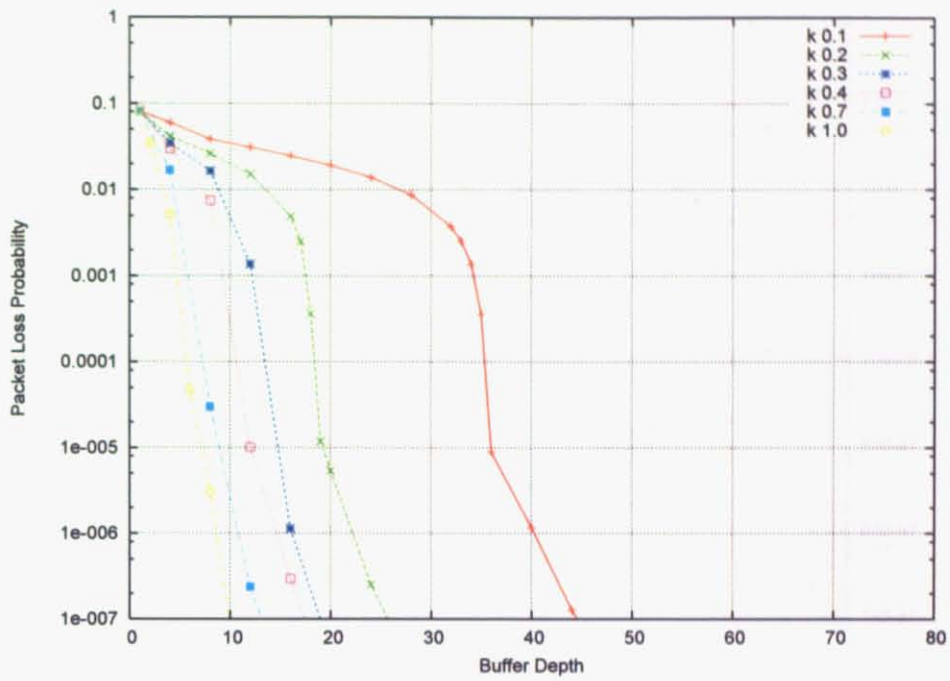


図 4.11: 遅延線数とパケット損失率の関係 ( $\rho = 0.1$ )

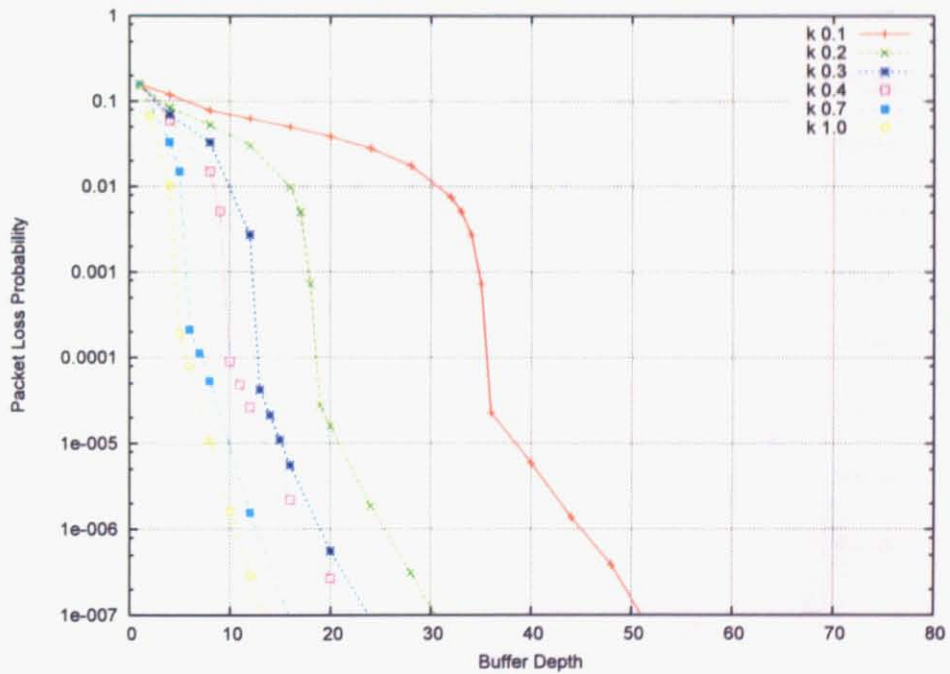
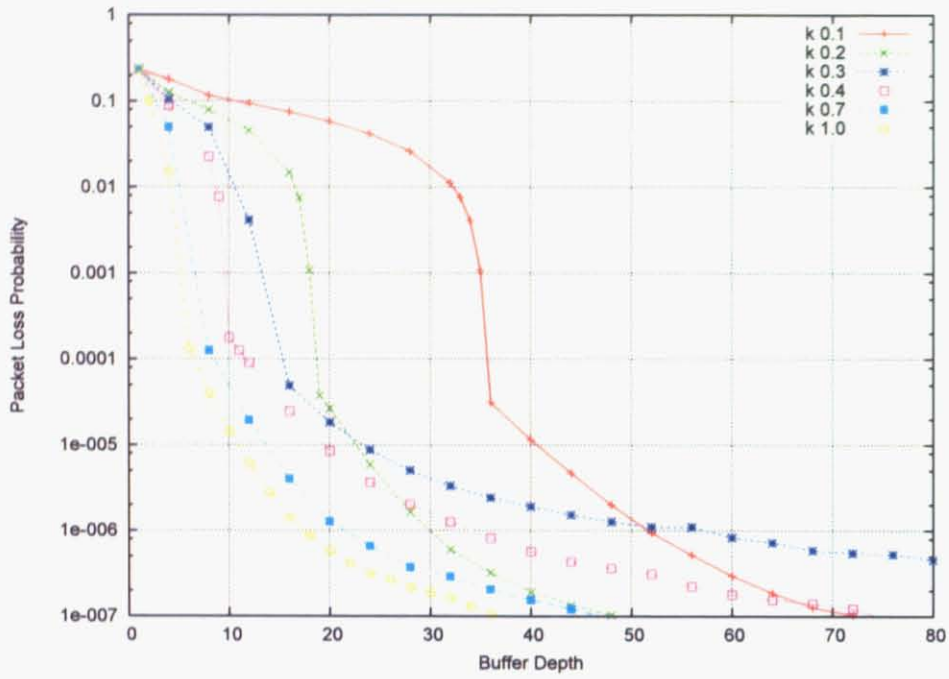
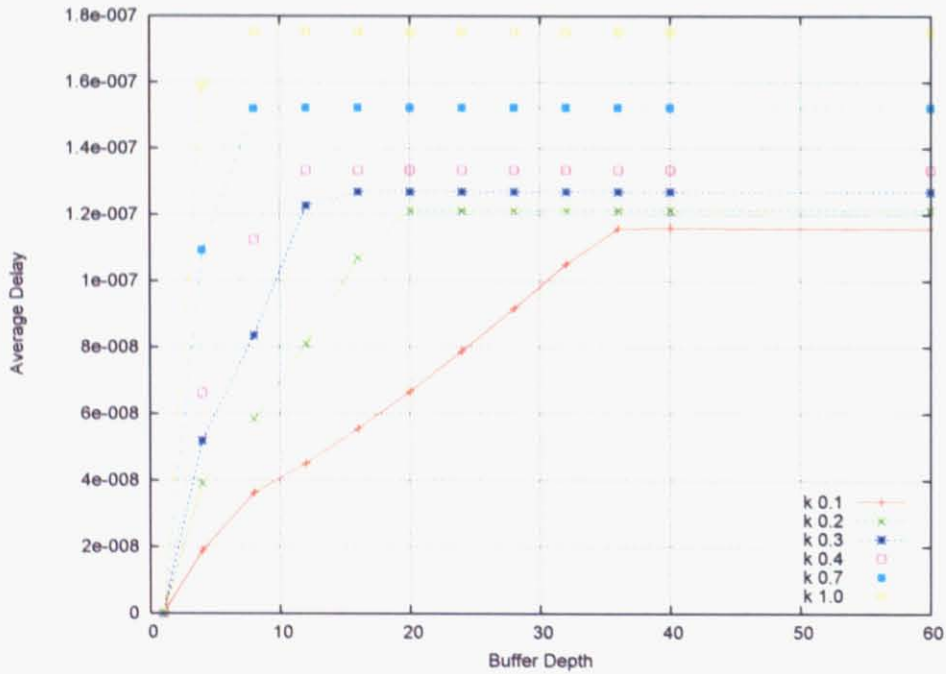


図 4.12: 遅延線数とパケット損失率の関係 ( $\rho = 0.2$ )

図 4.13: 遅延線数とパケット損失率の関係 ( $\rho = 0.3$ )図 4.14: 遅延線数とバッファ平均遅延の関係 ( $\rho = 0.1$ )

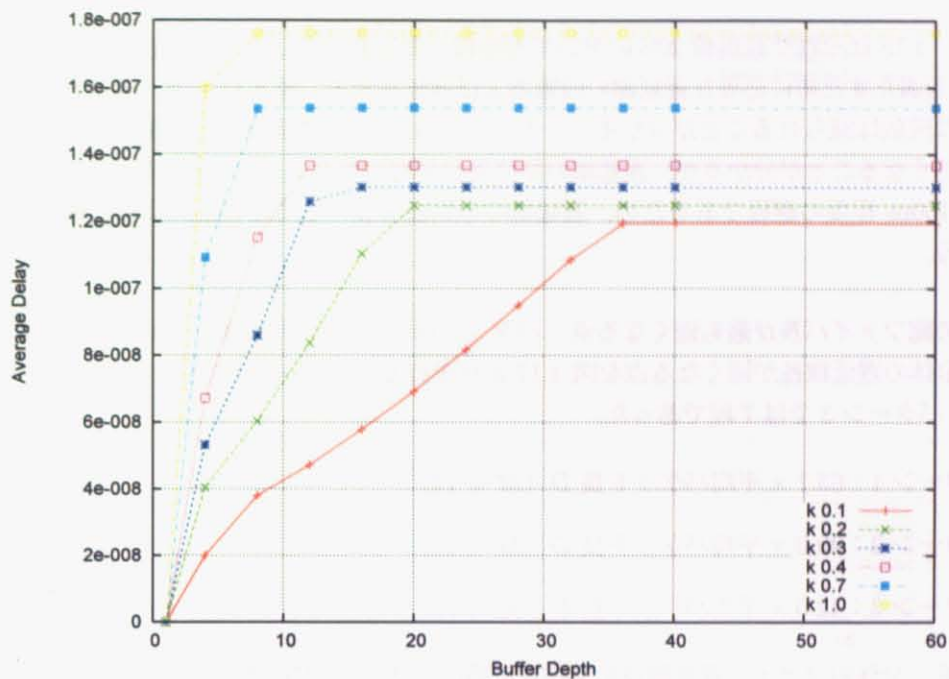


図 4.15: 遅延線数とバッファ平均遅延の関係 ( $\rho = 0.2$ )

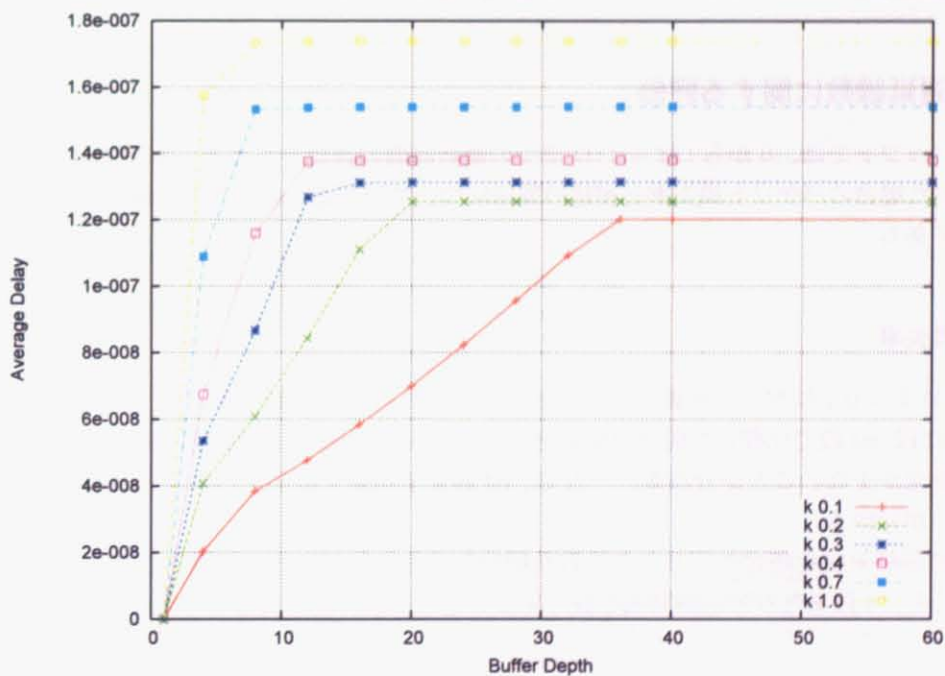


図 4.16: 遅延線数とバッファ平均遅延の関係 ( $\rho = 0.3$ )

### パケット損失率 $10^{-6}$ を実現するバッファ構成

図 4.17, 4.18 はそれぞれ負荷 20%において遅延線の単位長を変化させたとき、パケット損失率  $10^{-6}$  以下を満たすために必要な遅延線の段数および遅延を示す。図 4.17 より、単位長を大きくすれば必要な段数は減らせることが分かる。一方で、図 4.18 に示すように、単位長が大きくなると遅延は大きくなることが分かるが、遅延線の単位長を平均パケット長の 2 倍にしても 0.1 倍の時と比較して 150ns 程度の遅延であるため、遅延線の単位長を長くしても遅延による影響は無視できると言える。

**バッファの総ファイバ長が最も短くなる点** パケット損失率  $10^{-6}$  を実現するバッファ構成において、最も全体の遅延線長が短くなる点を図 4.17 より求める。パターン 1 では 14 段、パターン 2 では 10 段、パターン 3 では 7 段であった。

- パターン 1 :  $63.7 \times$  平均パケット長  $D$  ( $M = 14$ )
- パターン 2 :  $90.0 \times$  平均パケット長  $D$  ( $M = 10$ )
- パターン 3 :  $42.0 \times$  平均パケット長  $D$  ( $M = 7$ )

このとき、全体のファイバ遅延線長を求めるために、ファイバ内の光速を  $2.0 \times 10^8$  m/s とする。平均パケット長  $D$  が 128 バイト、回線速度が 10Gbps であるため、平均パケット長を遅延線長に換算すると約 20m となる。したがって、800m~1800m 相当の遅延線が必要になる。20%程度の負荷であれば遅延線バッファでも十分実用的な衝突回避性能が提供可能であることが示された。

#### 4.4.5 遅延線数に関する評価

遅延線バッファでは、0 から  $(M - 1)D$  までの離散時間の遅延を与える。本節では遅延線数  $M$  を変化させた場合のパケット損失率と遅延の関係について述べる。各グラフは実トラフィックを用いたものである。

#### パケット損失率

図 4.19 は  $k = 0.1$  において到着負荷を変化させた時のパケット損失率について、遅延線の段数を 16 段, 32 段, 64 段で比較した場合の結果を示している。16 段と 32 段では負荷 3%においても衝突回避の効果が少ないことが分かる。一方で、64 段まで用意すると、パケット損失率の改善効果が高いことが分かる。

また、負荷が 70%を超えた辺りから段数を増やしても効果がほとんど無いことが分かる。単位長を大きくし、1 段あたりの格納時間を長くすることで改善可能であると考えられる。

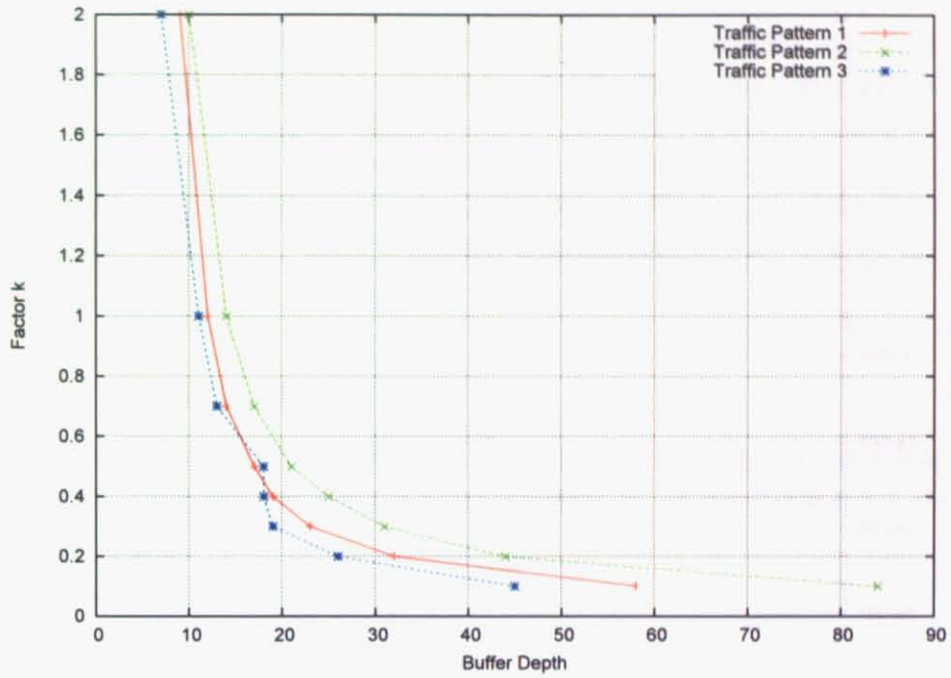


図 4.17: パケット損失率  $10^{-6}$  を満たす遅延線数と遅延線長係数の関係

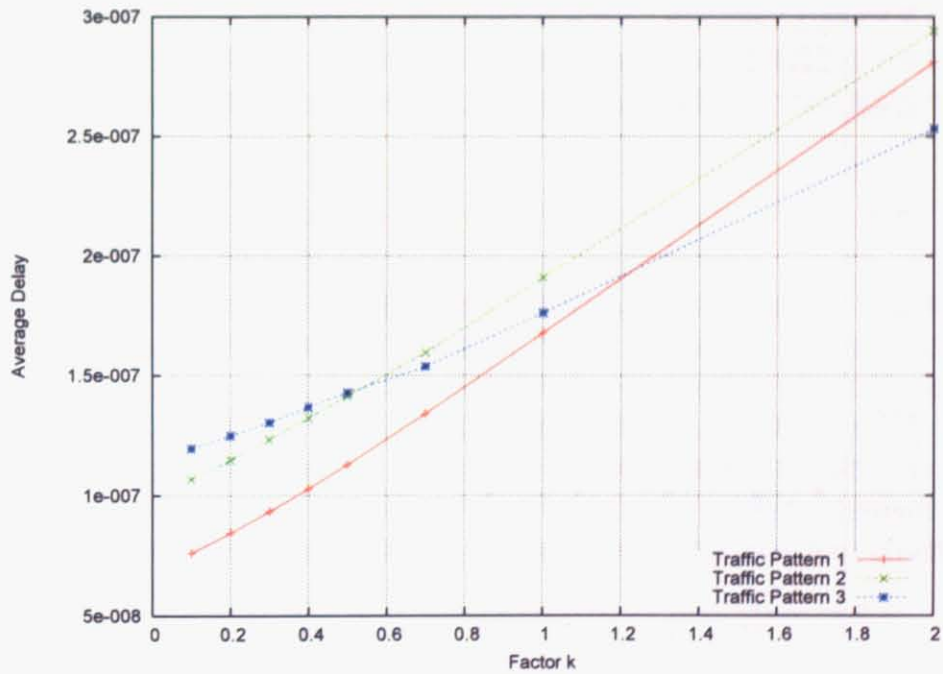
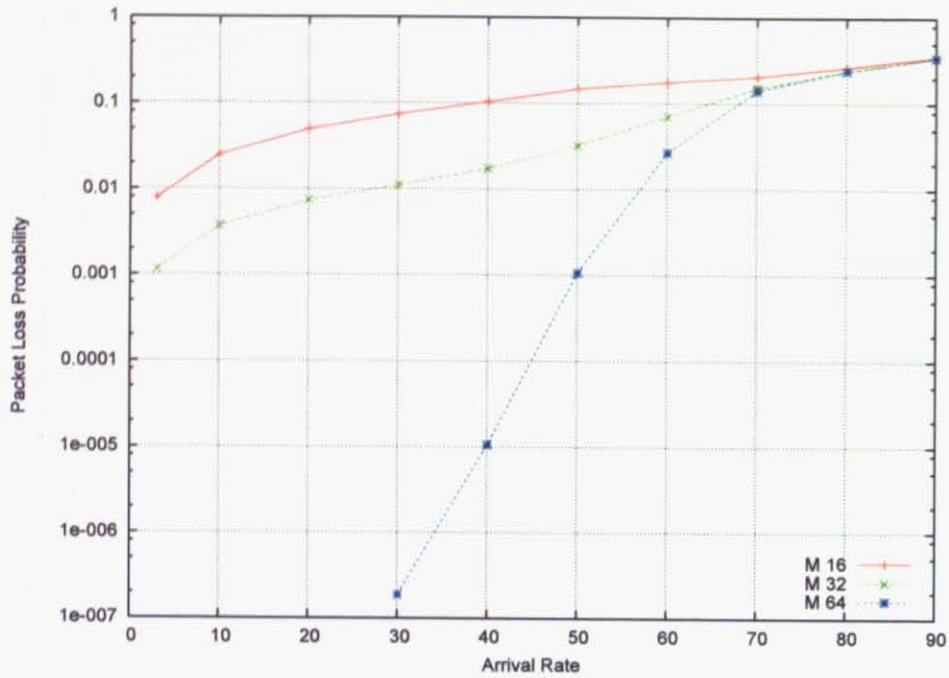
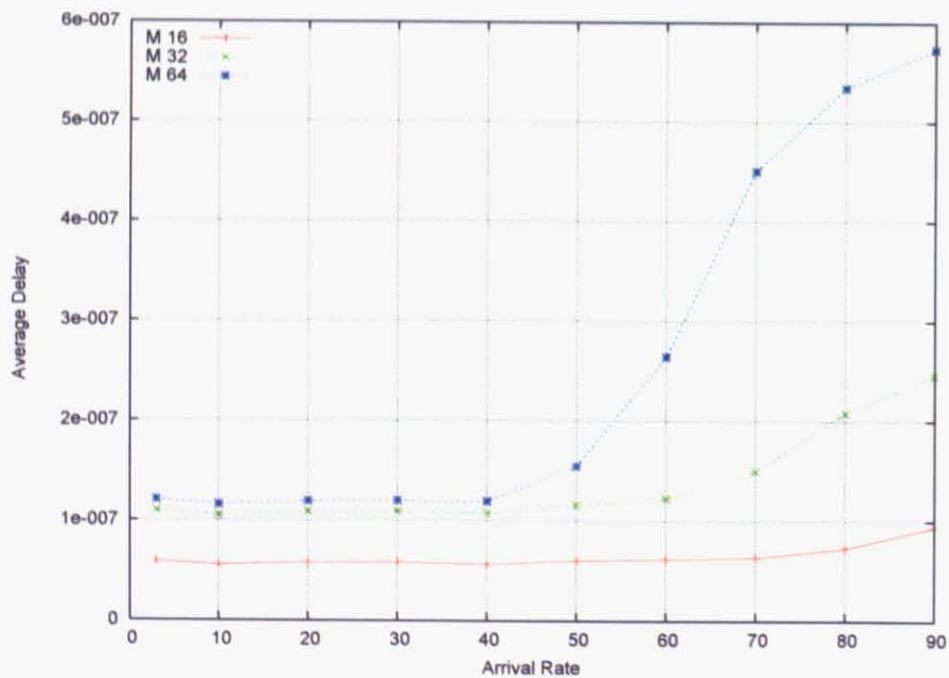


図 4.18: パケット損失率  $10^{-6}$  を満たす遅延線長係数とバッファ平均遅延の関係



図 4.19: 到着負荷とパケット損失率の関係 ( $k = 0.1$ )図 4.20: 到着負荷とバッファ平均遅延の関係 ( $k = 0.1$ )

### バッファ平均遅延

図 4.20 は  $k = 0.1$  において到着負荷を変化させた時のバッファ平均遅延について、遅延線の段数を 16 段, 32 段, 64 段で比較した場合の結果を示している。段数が少ない場合は負荷が高くなっても無視出来る程度の遅延であるが、段数が多くなると負荷が高くなるにつれて急激に遅延が大きくなる事が分かる。但し、インターネットの平均リンク利用率が 3% から 20% であることを考慮すると、遅延に関しては段数が変化してもさほど影響がないと言える。但し、年々アクセス速度が向上している現状を踏まえると、高負荷状態における遅延は段数とのトレードオフとなるため、設計に影響が出てくるものと考えられる。

#### 4.4.6 ポート数に関する評価

本節ではポート数を変化させたときのパケット損失率と遅延の関係を示す。到着負荷は 20% とする。各入力ポートの合計が 20% となるように負荷を調節した。各グラフは実トラフィックを用いたものである。

### パケット損失率

図 4.21 は  $k = 0.1$  においてポート数を変化させた時のパケット損失率について、遅延線の段数を 16 段, 32 段, 64 段で比較した場合の結果を示している。入力ポート側で同じ負荷を与えていてもポート数が増えるとパケット損失率が良くないことが分かる。これは、ポート数が多いほど各入力ポートから同時にパケットが到着する可能性が高くなるため、衝突が起こりやすくなるものと考えられる。一方で、バッファの段数が少ない場合はポート数の増減による損失率はあまり変わらないことが分かる。したがって、段数が大きくなるほどポート数による損失率の性能に差が出てくると言える。

### バッファ平均遅延

図 4.22 は  $k = 0.1$  においてポート数を変化させた時のバッファ平均遅延について、遅延線の段数を 16 段, 32 段, 64 段で比較した場合の結果を示している。段数が少ない場合は遅延はほぼ無視できる値であることが分かる。64 段ではポート数が 2 ポートから 8 ポートまでの間で 400ns の遅延差が生じているが、これについてもノード全体から見ると問題無い値であると言える。ポート数が多くなると遅延も大きくなる理由は、ノードにパケットが同時に到着する可能性が高くなり、使用中の遅延線が増えるため、空きの遅延線に回される可能性が高くなるからであると考えられる。特に、バッファ段数を増やすと衝突回避性能は上がるが、線形増加型のバッファでは段数が増えると遅延線長が大きくなるため、遅延に影響が出てくると言える。

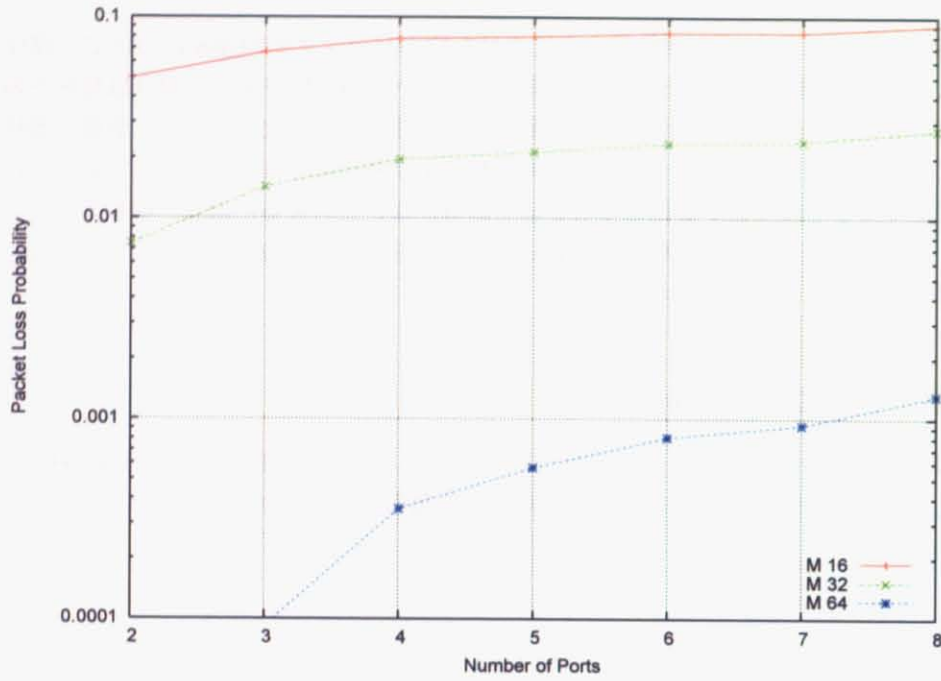


図 4.21: ポート数とパケット損失率の関係 ( $k = 0.1$ )

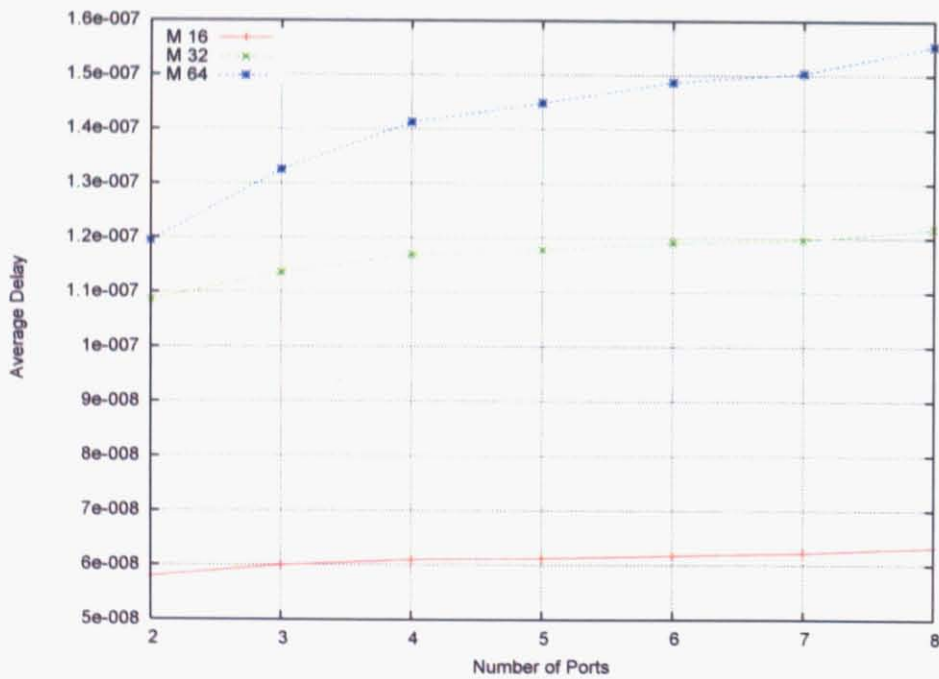


図 4.22: ポート数とバッファ平均遅延の関係 ( $k = 0.1$ )

## 4.5 おわりに

本章では、遅延線バッファの構成と衝突回避性能およびそのときのトランスポートへの影響について明らかにした。遅延線バッファではハードウェア的な構成により衝突回避時の遅延の粒度や大きさ、分布が変わることから、適用するトラフィックパターンによって性能が左右される。そこで実トラフィックを含む複数トラフィックモデルにおける評価を行った。

本シミュレーションにより、パケット損失率  $10^{-6}$  を実現するための並列型バッファによる衝突回避機構の構成が明らかになり、現状のトラフィックを想定すれば、インターネットと同等のパケット損失率が十分現実的な衝突回避機構の構成で実現されることが確認された。また、入力トラフィックとして実トラフィックを統計分布モデルトラフィックを用いた場合の性能の傾向の違いが示された。

今後の課題としては、より詳細な特性の評価を行う必要がある。例えば、入力トラフィックのデータパターンを増やし一般性のある結論を導き出すこと、バッファの構成パターンを増やしトラフィックパターンに対する適切なバッファ構成を導き出すこと、衝突回避の挙動をよりミクロに観察することで衝突回避において得られた特徴のメカニズムの解明をするといったことが考えられる。これらについて検討し、並列型バッファによる衝突回避機構の一般的モデルの導出を行うことが、実ネットワークへの適用時に適切なバッファ構成をとるために求められる。