卒業論文　　2004年度 (平成16年度)

LOLCAST: Layered Overlay Multicast Protocol

慶應義塾大学 環境情報学部

氏名：小椋 康平

指導教員
慶應義塾大学 環境情報学部
村井 純
徳田 英幸
楠本 博之
中村 修
南 政樹

平成17年2月15日

卒業論文要旨 2004 年度 (平成 16 年度)

LOLCAST: Layered Overlay Multicast Protocol

本論文では, 階層構造を持つデータの配信に適したオーバーレイ・マルチキャストプロトコルである LOLCAST を提案し, その設計, 実装を行った. オーバーレイ・マルチキャスト技術は IP マルチキャスト技術の代替技術として研究が行なわれており, その適用性の高さが注目されている. しかし, 既存のオーバーレイマルチキャストプロトコルでは, 単一的なデータの配信しか考えられていないため, 受信者の品質に対する多様な要求に応えることができない. 本研究では, この問題を解決するために, 階層構造を持つデータの配信に適した新たなオーバーレイマルチキャストプロトコルによりこの問題を解決した.

キーワード
1. オーバーレイ・マルチキャスト, 2. 映像配信, 3. 資源環境

慶應義塾大学 環境情報学部

小椋 康平

Abstract of Bachelor's Thesis

Academic Year 2004

LOLCAST: Layered Overlay Multicast Protocol

This paper proposes a new overlay multicast protocol for delivering hierarchical structured data and discussed on capability of the protocol. Overlay multicast has been proposed for alternative technology to IP multicast. Existing overlay multicast protocols are only designed for delivering single quality data. Therefore, existing overlay multicast protocol could not offer a service to the recipients, which could apply to the request on quality of data. This paper proposed overlay multicast protocol for delivering hierarchical structured data to solve this problem.

Keywords :
1. overlay multicast, 2. video transfer, 3. resource environment

Keio University, Faculty of Environmental Information

Kohei Ogura

目 次

第 1 章　序論 **1**
1.1　背景 1
1.2　目標環境の必要要件 2
1.3　論文の構成 2

第 2 章　既存配信技術の問題点 **3**
2.1　サーバクライアント型モデル 3
2.2　CDN 4
2.3　IP マルチキャスト技術 4
2.4　既存配信技術の限界 5

第 3 章　オーバーレイ・マルチキャスト技術 **6**
3.1　オーバーレイ・マルチキャスト技術の特徴 6
3.2　既存オーバーレイ・マルチキャスト技術のアプローチ 7
3.2.1　参加ノードの体系化 7
3.2.2　マルチキャストツリーの構成手法 8
3.3　既存オーバーレイ・マルチキャスト技術の問題点 9
3.3.1　受信者の品質に対する要求 9
3.3.2　複数品質の提供に伴う問題 9

第 4 章　LOLCAST の提案 **11**
4.1　階層的な構造を持つデータ 11
4.1.1　階層符号化 11
4.1.2　その他の階層的な構造を持つデータ 12
4.2　LOLCAST 概要 13
4.3　マルチキャストツリーの構成 13
4.3.1　レイヤ数をメトリックとしたマルチキャストツリー 13
4.4　実現する機能 15
4.4.1　ノード障害時におけるデータの保証 15
4.4.2　要求品質の幅を利用した輻輳回避 17

第 5 章　LOLCAST の設計 **18**
5.1　想定利用環境 18
5.2　LOLCAST 設計概要 18
5.3　ノードの定義 18
5.4　各ノードの保持する情報 19

5.4.1 ノード情報 19
5.4.2 Source Node 固有情報 20
5.4.3 Recipient Node 固有情報 21
5.5 メッセージ群 21
5.5.1 Rendezvous Message 21
5.5.2 PPL Message 22
5.5.3 Join Message 22
5.5.4 Data Message 23
5.5.5 Leave Message 23
5.5.6 Prune Message 24
5.5.7 Notify Message 24
5.6 メッセージフォーマット 25
5.7 マルチキャスト・ツリーの構成 26
5.7.1 パラメータの定義 26
5.7.2 PPL Extraction 26
5.7.3 Rendezvouz Procedure 28
5.7.4 Join Procedure 28
5.7.5 Leave Procedure 32

第 6 章 実装 35
6.1 実装概要 35
6.2 実装環境 35
6.2.1 CTmpLib 36
6.3 データ構造 36
6.3.1 Source Node 36
6.3.2 Recipient Node 37
6.4 LOLCAST プロトコル処理モジュール 37
6.4.1 ツリー操作関数群 37
6.4.2 メッセージ処理関数群 41
6.4.3 モジュール結合部 42
6.5 シミュレータモジュール 42

第 7 章 評価 44
7.1 評価の目的 44
7.2 定性評価 44
7.3 定量評価 44
7.3.1 実験環境 45
7.3.2 実験結果 45
7.4 評価結果 47

第 8 章 結論 51
8.1 まとめ 51
8.2 今後の課題 51

8.2.1 PPL Extraction の性能改善 51
8.2.2 フレームワークの提供 51
8.2.3 アプリケーションによる評価 51

図 目 次

2.1 サーバクライアント型モデルにおける問題点 3
2.2 IP マルチキャストにおける問題点 4

3.1 オーバーレイ・マルチキャストと IP マルチキャストの比較 6
3.2 オーバーレイ・マルチキャストにおけるノードの体系化 7
3.3 既存配信サービスにおける映像音声の品質制御 9
3.4 既存オーバーレイ・マルチキャスト技術を利用した配信に関わる問題 10

4.1 階層的な構造を持つデータ 11
4.2 階層符号化方式を利用した際の映像の変化 12
4.3 DV フォーマットの映像フレームを利用したデータ / 既存メディアの複合データ 12
4.4 マルチキャストツリーの一例 14
4.5 Host Cast における親ノードの冗長化手法 15
4.6 Base レイヤを利用した Data Topology での冗長化 16
4.7 要求品質の幅を利用した輻輳回避 17

5.1 LOLCAST におけるノードの定義 19
5.2 メッセージフォーマット 25
5.3 PPL Extraction の流れ 27
5.4 Rendezvous Procedure 28
5.5 Join Procedure 1/3 29
5.6 Join Procedure 2/3 30
5.7 Join Procedure 3/3 31
5.8 Join Procedure におけるメッセージパッシング 31
5.9 Leave Procedure 1/2 32
5.10 Leave Procedure 2/2 33
5.11 Leave Procedure におけるメッセージパッシング 34

6.1 LOLCAST システム図 36
6.2 Source Node の保持するデータ構造 38
6.3 Recipient Node の保持するデータ構造 39
6.4 ツリー構造関数一覧 40
6.5 disttree_changeparent() 40
6.6 disttree_getparentlist() 41
6.7 メッセージ処理関数一覧 42
6.8 シミュレータ関数群 43

6.9 マルチキャストツリー出力の例 43

7.1 シミュレータ実行例 45
7.2 Join Procedure 所要時間 (10,000 nodes / Fixed layer / Skip List) 48
7.3 PPL Extraction 所要時間 (10,000 nodes / Fixed layer / Skip List) 48
7.4 Join Procedure における PPL Extraction 所要時間 (10,000 nodes / Fixed layer / Skip List) 49
7.5 Join Procedure 処理時間 (10,000 nodes / Random layer / Skip List) 49
7.6 PPL Extraction 処理時間 (10,000 nodes / Random layer / Skip List) 50
7.7 Join Procedure における PPL Extraction 所要時間 (10,000 nodes / Random layer / Skip List) 50

表 目 次

5.1 ノードとパラメータの定義 26

6.1 実装環境 36

7.1 定性評価 44
7.2 評価環境 45
7.3 同一のレイヤ数の値を指定した際の処理時間平均 46
7.4 ランダムにレイヤ数を指定した際の処理時間平均 46

第1章　序論

1.1　背景

インターネットは社会インフラとして成り立ち, その上で無数の人々が様々な活動を行なっている. これらの人々は活動の表現の場としてインターネットを用い, 絵画や音楽など様々な創作活動を行ない, その成果を個々に発信している. 本研究では, これら「インターネット上で共通の興味・目的を持ち活動している個人あるいは集団」を, インターネット・コミュニティ (以後, コミュニティと記述) と定義する.

既にこれらのコミュニティは, インターネットを表現の場として利用し, 多くの作品を発信している. それは詩や小説などの文字メディア, 音楽や落語などの音声メディア, 映画や演劇などの映像メディアまでに及ぶ. 将来的には 3 次元映像 [1] を利用した創作活動も登場すると考えられる.

しかし現状では, これらのコミュニティが映像中継等のコンテンツ配信を行なうことは困難である. 多くの場合, 個々のコミュニティを構成する個人は一般利用者であり, 利用可能な計算機資源やネットワーク帯域資源に限界がある. そのため, 提供可能な受信者数やコンテンツ品質の種類には限界があり, 多くの受信者を対象とすることができない. したがって現状において, 少年野球の試合中継や地元のお祭りの中継といった, 地域に根差した報道活動等を個人レベルで成り立たせることは非常に難しい.

これらの問題点の本質には二つの側面がある. 一つは上述の通り, 発信者側の計算機資源・ネットワーク帯域資源に限界があることである. これによって, 受信者の数や提供可能なコンテンツの品質が限定されてしまう. もう一方は, 受信者側の計算機資源と通信路の品質である. 発信者からインターネット上に分散する多くの受信者までの通信路はそれぞれ異なる性質をもつ. したがって, 単一的な品質のサービスでは, 受信者までのネットワーク環境あるいは受信者の計算機資源の多様性に適応できない.

本研究では, 「多くの受信者を対象としたコンテンツ配信を行なうコミュニティの活動を支援すること」を目標とし, 上述した問題点を階層構造を持つデータの配信に適した新たなオーバーレイ・マルチキャストプロトコルである LOLCAST により解決した. 本研究により「発信者が一般利用者にとって現実的な資源で, 多くの受信者に対し, 個々の受信者が要求する品質でサービスを提供できる環境」という理想的な環境を実現する.

1.2 目標環境の必要要件

1.1節で示した理想環境を実現するために必要な要件を以下に示す．コンテンツ配信を行なう人々を発信者, コンテンツを受信し視聴を行なう人々を受信者とする．

1. **一般的なネットワーク, 計算機資源で配信網を構築**
   個人の発信者であっても, 一般利用者が持ちうる現実的な資源環境で配信網を構築できる．資源環境とは, ネットワーク資源, 計算機資源を指す．発信者の資源環境は般利用者が容易に保持できるものを対象とする．また, 発信者の運用コストに関しても一般利用者にとって許容できる範囲で運用できる．

2. **自由な参加, 脱退**
   受信者が自由に配信網に参加, 脱退できる．これは受信者が, 自分のネットワーク的な位置, ネットワーク資源, 計算機資源の環境に関わらず配信サービスを受けられることを意味する．

3. **受信者の要求に基づくコンテンツ品質**
   それぞれの受信者は, 所有する資源の範囲内であれば, 望んだ品質のコンテンツを受信できる．受信者の品質に対する要求は, 受信者の持つネットワーク資源, 計算機資源, もしくはコンテンツへの関心度等により常に変化する．例として受信者が他のアプリケーションでネットワーク資源を消費することを予め認知している場合には, 受信者自ら品質を下げることができる．このように受信者は, 自らの要求に基づく品質でコンテンツを視聴できる．

1.3 論文の構成

本論文は 8章から構成される．2章で目標環境の実現における既存配信技術の問題点を述べる．3章で, 既存の配信技術にかわるマルチキャスト技術であるオーバーレイ・マルチキャスト技術と, その問題点を述べる．4章で新たなオーバーレイ・マルチキャストプロトコルである LOLCAST の提案を行う．5章で LOLCAST の設計を行い, 6章で LOLCAST を用いたシステムの実装を行う．7章では, 設計・実装を行った LOLCAST の評価を行う．最後に 8章において, LOLCAST の総括と今後の課題を述べる．

第2章　既存配信技術の問題点

本節では, 既存の放送型の通信を実現する技術である, サーバークライアント型モデル, CDN, IP マルチキャストを取り上げる. それぞれの技術において, 1.1節で述べた目標環境の実現にあたって発生する問題を述べる.

2.1　サーバクライアント型モデル

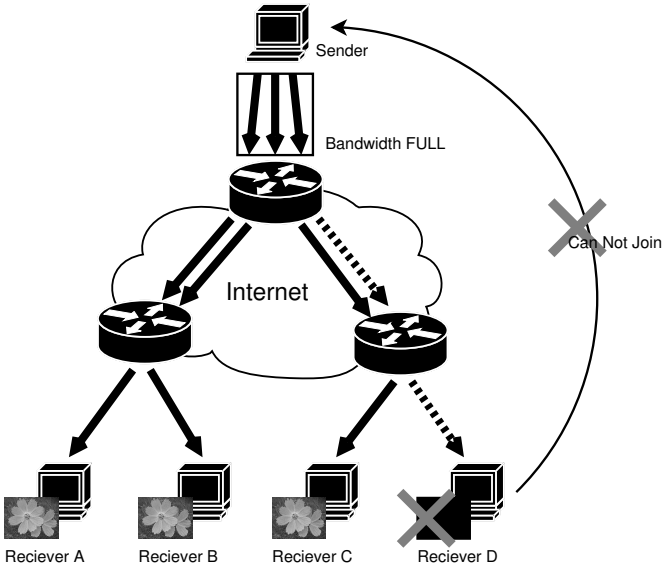

図 2.1: サーバクライアント型モデルにおける問題点

一般的に利用されている配信手法として, 複数のユニキャストにより通信を行うサーバクライアント型モデルが挙げられる. 特殊な配信環境を必要としない点から, 多くの映像配信サービスで利用されている. 図 2.1に示すようにサーバクライアント型モデルでは, 発信者からそれぞれの受信者に対し, ユニキャスト通信を用いてデータの配信が行われる. この際, 発信者が一般利用者である場合に問題となるのが, ネットワークの帯域的資源が限られている点である.

図 2.1において, 発信者は受信者 A,B,C に対し配信を行っている. 発信者の所持するネットワーク資源は受信者 A,B,C に配信を行っている状態ですでに枯渇しており, 新たに配信を希望する D は配信を受ける事ができない. これにより, 対象とできる受信者の数が限られてしまい, 多くの受信者を対象とした配信ができない.

上述したように, サーバークライアント型モデルを利用した配信では, 受信者の数に比例したネットワーク資源が必要である. よって, 現状で行われているサーバクライアント型モデルを利用した個人による配信は, 十数人を対象とした小規模なものや, 配信データの利用帯域を制限し

たものが殆どを占める.

2.2 CDN

CDN (Contents Distribution Network) は 2.1節で述べたサーバクライアントモデルにおける発信元を複数箇所に分散させることにより実現される. 発信元の分散により発信者にかかる負荷が軽減され, さらに受信者に安定したコンテンツ配信が可能となる. 具体的には, 発信元に存在するコンテンツを各 ISP に分散させたサーバに対してキャッシュし, 受信者に最適なサーバを伝達することによって配信の効率化を行なう. しかし, 一般利用者である一個人が気軽に利用できる CDN サービスは提供されていない. また, 各 ISP が CDN 網 に参加していることを前堤としており, 広域に分散している受信者に対して映像配信を行なうことは困難である.

2.3 IP マルチキャスト技術

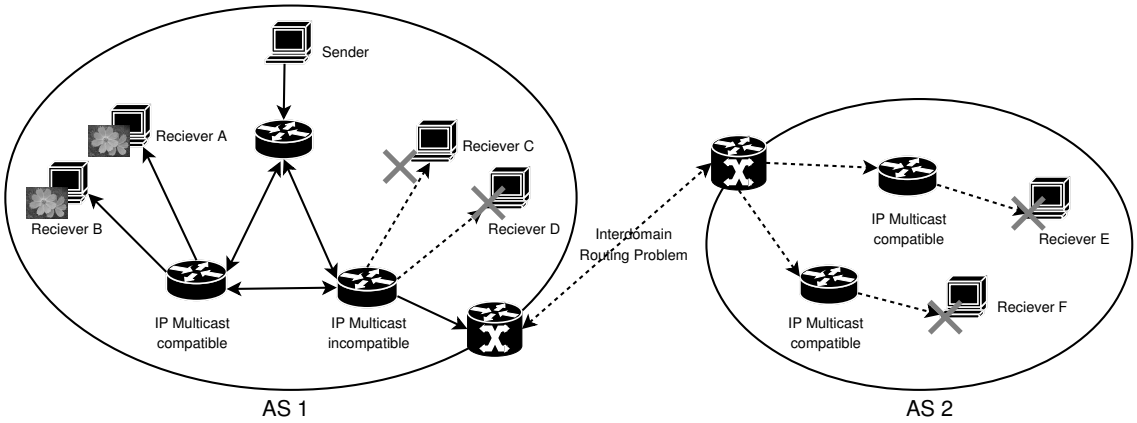

図 2.2: IP マルチキャストにおける問題点

従来考えられてきたグループ通信の手法として IP マルチキャスト技術 [2] がある. ネットワーク層でグループ通信を実現する IP マルチキャストを広域に適用するためには, マルチキャストアドレス予約, グループの管理, ISP 間のポリシの違い, AS 間マルチキャスト経路制御, 全ルータのマルチキャスト対応といった諸問題を解決しなければならない [3].

図 2.2のように, 発信者は受信者 A,B,C,D,E に対し配信を行う例を考える. 受信者 C,D は IP マルチキャスト非対応なルータに接続しているため, 配信を受けることができない. さらに, 受信者 E,F は AS 間のポリシの違いにより, AS1 からの IP マルチキャストのトラフィックが流れず, 配信が行われない.

上述した理由から発信者は広域に分散して存在する受信者に対し自由にグループ通信を行えない. このため, 現在 IP マルチキャスト技術は, 特定 ISP 内等の限られた範囲への適用に留まっている.

2.4 既存配信技術の限界

上述したようにインターネットにおいて放送型の通信を実現する技術はこれまでも研究・開発が盛んに行なわれてきた. しかしこれらの技術は一般利用者が手軽に利用できるものではない. サーバクライアント型モデルは発信者として一般利用者を対象とできるが, 発信者側の帯域的資源の限界から受信者の数が制限され, 自由な配信網の構築は不可能である. CDN はサービスとして一個人が気軽に利用できるものが提供されていない. また, CDN に参加していない ISP に接続する受信者を対象とできない. IP マルチキャスト技術はグループ通信を実現する技術として従来より研究が進められているが, 前述した技術的な問題, 運用上の問題からインターネットへの広域な適応がなされていない.

第3章　オーバーレイ・マルチキャスト技術

近年, IP マルチキャストの代替技術として, オーバーレイ・マルチキャスト技術が登場し, 多くの研究が進められている [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. 本章ではオーバーレイ・マルチキャスト技術について詳しく解説する. はじめにオーバーレイ・マルチキャスト技術の特徴を述べる. 次にオーバーレイ・マルチキャストを構成する要素である, ノードの体系化, マルチキャストツリーの構成手法について述べる. 最後に既存オーバーレイ・マルチキャスト技術の問題点を取り上げる.

3.1　オーバーレイ・マルチキャスト技術の特徴

マルチキャストを行なうためのボトルネックとなる処理としてデータの複製が挙げられる. IP マルチキャストではデータの複製をルータが行なう. これに対し, オーバーレイ・マルチキャストでは配信網に参加するそれぞれのエンドホストがデータの複製を行なう. データの配信元となるエンドホストは子となるエンドホストにデータを複製してもらうことにより配信に伴う負荷を抑えることができる. このデータの配信網をマルチキャストツリーと呼ぶ.

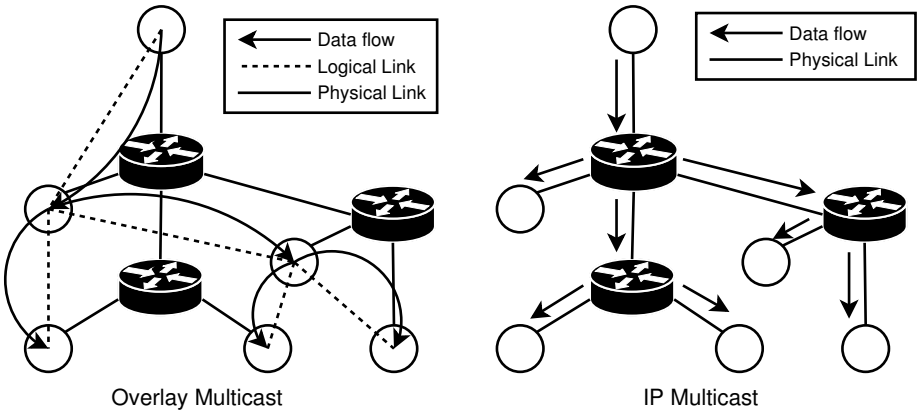

図 3.1: オーバーレイ・マルチキャストと IP マルチキャストの比較

図 3.1に示すように, オーバーレイ・マルチキャスト技術は, 実ネットワークの上で構築された論理的なトポロジであるオーバーレイ・ネットワーク上で, マルチキャストを実現する. このオーバーレイ・ネットワークはエンドホスト間のユニキャスト通信によって構築され, 通信基盤として利用される. これに対し, IP マルチキャストではルータを基礎とする IP ネットワークを通信基盤として利用する. オーバーレイ・マルチキャスト技術の特徴は, ネットワーク管理機能, データの複製, グループの管理機能の全てをアプリケーションレイヤなどの上位レイヤに任せる点にある. このようにオーバーレイ・マルチキャスト技術はエンドホストを通信基盤として利用することにより, IP マルチキャストのインターネットへの適用に関わる問題を解決した.

一方で, 実ネットワーク上に論理的なトポロジを形成したことによるオーバーヘッドも存在

する．まず, 実リンク上で重複したデータが転送される可能性がある．これは実ネットワーク上に被さる形で構築されたオーバーレイ・ネットワークからは実ネットワークの状態を考慮できないことに起因する．また通信基盤として利用されるエンドホストは, その稼働状況に関して保証ができないため, 予期せぬマルチキャストツリーの分断が起こりうる．

効率的なグループ通信を行うためにオーバーレイ・マルチキャスト技術では, 上述したオーバーヘッドをできる限り解消する必要がある．オーバーレイ・マルチキャスト技術の配信における性能は, 構築されたそれぞれのマルチキャストツリーから決定される．よって, 効率的であり耐障害性に優れたマルチキャストツリーの構築がオーバーレイマルチキャスト技術の最も重要なポイントとなる．

3.2 既存オーバーレイ・マルチキャスト技術のアプローチ

オーバーレイ・マルチキャスト技術では, それぞれのオーバーレイ・マルチキャストプロトコルによってマルチキャストに関わる機能が提供されている．本節では既存のオーバーレイ・マルチキャストプロトコルのアプローチについて述べる．はじめにそれぞれのプロトコルの達成目標を述べた上で, オーバーレイ・マルチキャストプロトコルの要素である, ノードの体系化とそれぞれのプロトコルのマルチキャストツリーの構成手法を述べる．オーバーレイ・マルチキャストプロトコルについて述べる上で用語を以下の様に定義する．マルチキャストツリーに参加しているエンドホストをノード, 配信元となるノードをソースノードとする．

既存研究として様々なオーバーレイ・マルチキャストプロトコルが提案されている．主要な例として, ALMI[14], Narada[6, 7], Scattercast[15], Overcast[16], HMTP[10], Yoid[4], Hostcast[11], CAN[8], NICE[9] 等が挙げられる．各々のオーバーレイ・マルチキャストプロトコルは異なる達成目標に基づき設計されている．これらの設計は, 対象とするアプケーションの特性に依存し決定され, それぞれの特性に合わせたものとなる．また, それぞれのプロトコル毎にマルチキャストツリーの構成手法も異なる．

ALMI では構築したマルチキャストツリー上の各ノード間の遅延が最小となるマルチキャストツリーの構成を目標としている．これに対し Narada, Scattercast では, 各ノードとソースノード間の遅延が最小となるようにマルチキャストツリーを構成することを目標とする．遅延以外のメトリックとして帯域が挙げられるが, Overcast ではソースノードから各ノードまでの利用可能な帯域が最大となるマルチキャストツリーの構成を目標としている．Hostcast では, マルチキャストツリーの自体の安定性を高めることを目標としている．

3.2.1 参加ノードの体系化

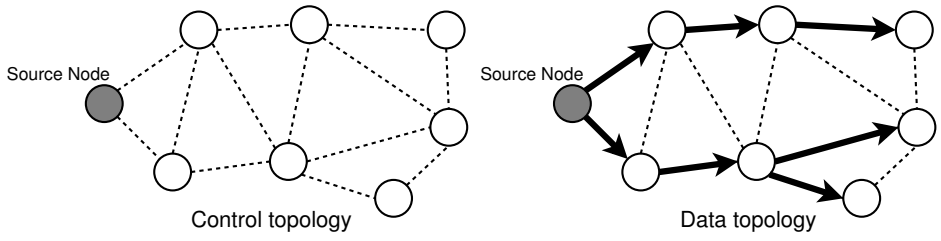

図 3.2: オーバーレイ・マルチキャストにおけるノードの体系化

オーバーレイ・マルチキャストプロトコルは 2 種類の論理的なトポロジをもつ [5]. 各ノードの管理を行なう Control Topology と実際のデータを送信するデータ転送のための Data Topology である. 図 3.2に両者の一例を示す. Control Topology の主な目的は, 各々のノードの状況を把握し, ノードの予期しない切断への対処を行なうことである. ノードの予期しない切断とは, マルチキャストツリーに参加するノードが, 計算機やネットワークの障害等によってマルチキャストツリーから決められた切断手順を踏まずに切断されてしまうことを表す. Data Topology は, Control Topology の一部である場合が多く, 実際のデータの流れを規定するために存在する. Control Topology は, その形態から Mesh と呼ばれることが多く, Data Topology は, Tree と呼ばれることが多い.

3.2.2 マルチキャストツリーの構成手法

オーバーレイ・マルチキャストプロトコルは, 3.2.1節で述べた二つのトポロジの構成手法から 3 種類に大別可能である [5, 12, 17]. メッシュ状の制御トポロジを最初に形成する Mesh-first 型, 分散的にデータ転送の為のツリーを最初に形成する Tree-first (Direct) 型, 制御トポロジを何らかのメトリックに沿って形成する Implicit 型である.

Mesh-first 型

Mesh-first 型では, メッシュ状の Control Topology を最初に形成する. この際, 対となるノード同士に複数のパスが存在していてもよい. それぞれのノードは Control Topology 上でルーティングを行い, 分散的に一意なパスを決定することにより Data Topoloy を形成する. 多くのプロトコルにおいて, ソースノードから RPF[18](Reverse Path Forwarding) を行う事により Data Topology を形成している. Mesh-first 型のプロトコルのとして, Narada, Scattercast が挙げられる.

Tree-first 型

Tree-first 型では最初に分散的に Data Topology の形成を行う. 次にそれぞれノードが Data Topology 上で対となっていないノードを発見し追加的な制御用のパスを張ることにより Control Topology を形成する. Tree-first 型のプロトコルのとして, ALMI, Overcast, HMTP, Yoid, Hostcast が挙げられる.

Implicit 型

Implicit 型のプロトコルでは, ある特定のメトリックを用い Contorl Topology を形成する. Data Topology は規定された Control Topology により暗黙的に決定づけられる. そのため, Implicit 型のプロトコルでは Contorl Topology と Data Topology が同時に規定され, 追加的な処理を行わずにそれぞれのトポロジが形成される. Implicit 型のプロトコルとして, CAN, NICE が挙げられる.

3.3 既存オーバーレイ・マルチキャスト技術の問題点

3.2節で述べたように, 現在多くのオーバーレイ・マルチキャストプロトコルが提案されている. しかし既存オーバーレイ・マルチキャストプロトコルでは単一的なデータの配信しか考えられていないため, 本研究の目標である受信者の要求に基づくコンテンツ品質を提供する上で問題がある. 以下でこの問題を詳細に論じる.

3.3.1 受信者の品質に対する要求

図 3.3: 既存配信サービスにおける映像音声の品質制御

オーバーレイ・マルチキャスト技術を用いて受信者の要求に基づくコンテンツ品質を提供するためには, 発信者が複数品質での配信を行う必要がある. 配信を行う主なコンテンツとして映像や音声が考えられるが, これらのコンテンツは品質を制御可能である. 品質とは, コンテンツの情報量や, 解像度, 画質等を指す. 現状のコンテンツ配信においても, 受信者の品質に対する様々な要求に応じるために複数の品質でこれらのコンテンツを提供している. この例を図 3.3に示す. 映画の予告編を配信している Web サイト [19] では ”Mid”, ”High”, ”光” のように受信者の資源環境を回線の種類によって抽象化し提供を行っている. インターネット・ラジオのような音声配信を行うサービスでは, ”24kbps”, ”56kpbs”, ”96kbps”のように受信者の帯域を指標として利用し, 複数品質でのサービスを行っている. ここでの品質に対する要求は, 受信者の計算機環境, ネットワーク環境や, 受信者の品質要求に影響される. このように, 受信者のコンテンツの品質に対する要求は様々である.

3.3.2 複数品質の提供に伴う問題

現状のオーバーレイ・マルチキャストプロトコル技術を利用し 3.3.1節で述べたような複数品質のサービスを提供する場合, 以下の 2 点が問題となる.

- 品質毎のコンテンツの生成
  発信者は受信者の品質に対する様々な要求に応じるために, 品質毎のコンテンツを提供する必要がある. 特に本研究の対象である中継等の映像配信を行おうとした際には常に複数の映像を生成する必要があり, 多くの計算機資源を必要とする. よって発信者への負荷が非常に大きい.

- 複数のマルチキャストツリーの管理
  発信者は品質毎のコンテンツを配信するために複数のマルチキャストツリーを管理する必要がある. それぞれの品質毎にマルチキャストツリーを構築することにな, これには発

信者への帯域的な負担, 複数のマルチキャストツリーの管理に対するオーバーヘッドの増加が伴う.

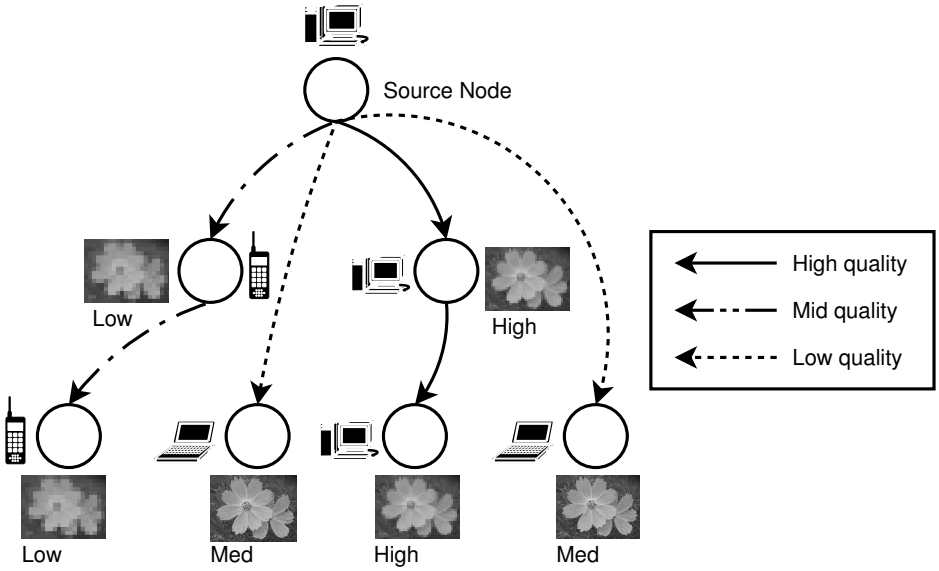

図 3.4: 既存オーバーレイ・マルチキャスト技術を利用した配信に関わる問題

図 3.4のようなマルチキャストツリーを用い, この問題の例を示す. ソースノードは無線端末のような狭帯域なノードから FTTH を利用しているような広帯域なノードまで, 多種多様なな資源環境に適応した配信を行わねばならない. この例では, ”Low”, ”Med”, ”High” の 3 種類の品質での配信を行っている. ソースノードは 3 種類の品質の映像の生成と, さらに 3 種類のマルチキャストツリーを個別に管理する必要がある. 以上のように, 既存のオーバーレイ・マルチキャストプロトコルでは単一なデータの配信しか考慮されていない. よって発信者はただ一つのコンテンツに対し, 品質毎のコンテンツの生成, 複数のマルチキャストツリーの管理に非常に多くの資源環境が要求され, 現実的ではない.

第4章　LOLCASTの提案

本章では 3.3節で述べた問題点を解決し目標環境を実現するため, 階層的な構造を持つデータの配信に適した新たなオーバーレイ・マルチキャストプロトコルである LOLCAST (Layered Overlay Multicast Protocol) の提案を行う.

はじめに LOLCAST で用いる階層的な構造を持つデータの定義とその特徴を述べる. またその主な例として, 階層符号化されたデータを取り上げる. 次に, このような構造を持つデータの配信を行うマルチキャストツリーの構成手法について述べ, 最後に LOLCAST の応用例を取り上げる. LOLCAST では階層的な構造をもつデータを用い, 受信者の資源環境をレイヤという単位で抽象化を行う. これにより, 受信者の資源環境や要求に応じた配信が可能となる.

4.1　階層的な構造を持つデータ

階層的な構造をもつデータは複数のレイヤから構成される. このレイヤには Base レイヤと Enhanced レイヤの二種類がある. それぞれのレイヤは一定の品質を提供しており, Base レイヤはそのデータの最低限の品質を提供する. Base レイヤの上位に複数の Enhanced レイヤを重ねることにより品質を向上できる. ここでの品質とは, 解像度, 情報量等を指す. 受信者は受信レイヤ数の選択により, 資源環境や要求に応じたデータを選択できる.

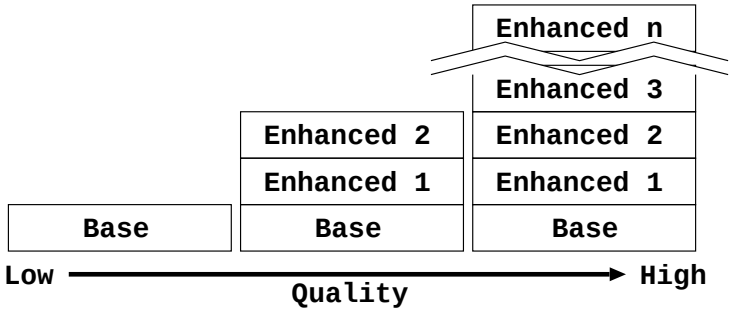

図 4.1: 階層的な構造を持つデータ

4.1.1　階層符号化

階層符号化とは, 解像度等を変化させた画像を階層的に複数用意し, 画像の階層数に応じて品質を選択できる符号化方式である. それぞれのレイヤごとに符号化を行ない, 下位レイヤの差分に対し符号化を行ったものがひとつ上位のレイヤとなる. つまり, 下位レイヤを補完する形で上位レイヤが存在する. 図 4.2 に示すようにレイヤの数によって LOW から HIGH にかけて, 映像品質が向上する. 階層符号化を利用した映像メディアの例として MPEG2 SNR (Signal to Noise Rate) Scalable/Spatial Scalable Profile[20], JPEG2000 EBCOT (Embedded Block

Coding with Optimized Truncation)[21, 22], MPEG4 AAC-SSR (Advanced Audio Coding Scalable Sample Rate) の 3 つが挙げられる.

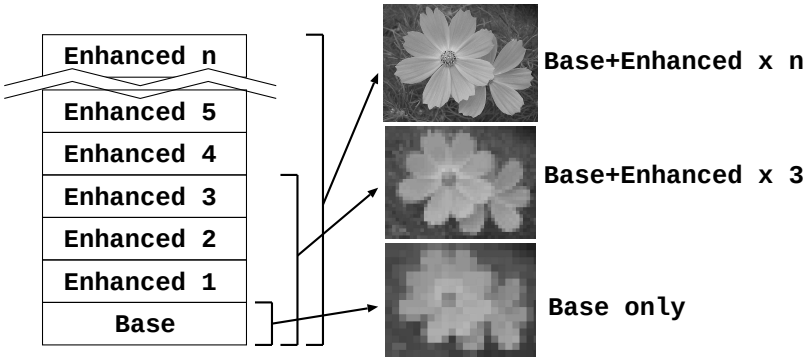

図 4.2: 階層符号化方式を利用した際の映像の変化

4.1.2 その他の階層的な構造を持つデータ

階層符号化以外にも, 階層的な構造を持つデータは構成できる. 例として, DV (Digital Video)[23] フォーマットの映像フレームを用いたデータと, 既存メディアを複合したコンポジットデータの二つを取り上げる.

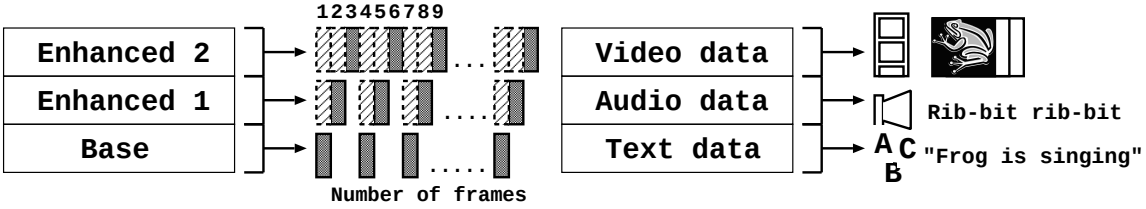

図 4.3: DV フォーマットの映像フレームを利用したデータ / 既存メディアの複合データ

はじめに DV フォーマットの映像フレームを用いたデータについて述べる. このデータでは, DV フォーマットで圧縮されたデータが複数の映像フレームに分割されているという特徴を利用する. DV データの映像フレームをレイヤに対して割り当てを行うことで, 映像のフレーム数を選択できるデータを構成する. 図 4.3の左図に示すように, 上位レイヤの映像フレームにより下位レイヤが補完される. 図で示した例では, 1 階層のみで 1/3 のフレームレート, 全階層でフルフレームの映像を提供できる.

この他にも既存メディアを利用したコンポジットデータが考えられる. 図 4.3の右図に示すように, Base レイヤに文字データを, Enhanced レイヤに音声データ, 映像データを重ねる事により複数のメディアを一つのデータに内包できる. このような構造を持つデータにより, 文字データのみ, 文字と音声データ, 文字と音声と映像データ, のように閲覧するメディアを選択できる.

4.2 LOLCAST 概要

4.1節で述べたように, LOLCAST により提供されるデータの品質はレイヤという概念に抽象化され, 各サービスはそのレイヤ数によって表現される. 本プロトコルではこのようなレイヤに分割されたデータの配信を行う. 発信者であるソースノードは最高品質のデータ, つまりレイヤ数が最大となるデータを配信する. このソースノードを頂点とするマルチキャストツリーを構成し, ツリーに参加するノードはレイヤ数を用い資源環境に応じたデータの要求を行う. さらに各ノードはデータを受信すると共に, ツリー上で下位に位置するノードに対しデータを中継する. これにより, 多くのレイヤ数を要求するノードが上位に位置するようなマルチキャストツリーを構築する.

本プロトコルの特徴は, 単一なデータの配信によって, 各ノードの受けるサービスの品質が異なる点である. 階層的な構造を持つデータの利用により, ソースノードは品質毎のマルチキャストツリーを管理する必要がない. また, ソースノードは最高品質のデータを最低一ストリーム, 下位ノードに配信すればよい. よってそれぞれの品質に対してデータを冗長に送る必要が無く, 既存技術と比べネットワークの帯域的に有利な配信が行える. さらにこの特徴を利用することで新たに, Base レイヤを用いたマルチキャストツリーの効率的な冗長化やレイヤの間引きによる輻輳回避の機能を提供できる.

本プロトコルにより, 「各ノードの要求した, あるいはそれの資源環境に適応した複数の品質」を単一のマルチキャストツリーで提供できる.

4.3 マルチキャストツリーの構成

本プロトコルの最も重要な機能は, ツリーに参加する各受信ノードが自由に受信するデータの品質を要求できることである. 本プロトコルは Tree-first 型のマルチキャストツリー構成手法を用い, 各ノードの要求するレイヤ数をメトリックとしたマルチキャストツリーを構築する.

4.3.1 レイヤ数をメトリックとしたマルチキャストツリー

図 4.4に, 本プロトコルを利用したマルチキャストツリーの一例を示す.

Source Node と記されている A がソースノードである. ノード内に記されている値は各ノードが所持するデータのレイヤ数である. 値が大きい程レイヤ数が多く, 高品質を求めるノードにデータを中継できる. ノード A の 10 レイヤの品質を持つデータが配信される最高品質のデータである. ソースノードである A は, 受信ノードの要求にしたがって配信を行なう.

マルチキャストツリーに参加を行うノードは, 自らの要求するレイヤ数以上のレイヤ数を保持しているノードからの配信を受ける. 各ノードは親と子の関係を持ち, 親は子に, 子の要求する品質を持つデータを提供する. また, 各ノードは親となるノードから受信している品質までのデータを配信できる. この際, マルチキャストツリーに変更を加える必要は無い. 例として, E は 親となる B から 6 レイヤの品質を持つデータを受信している. よって, 子となる H, I には 6 レイヤまでの品質でデータを送信できる. H や F の下にはさらに伸びるツリーがあると仮定する. G や I は, 2 や 1 レイヤの品質を持つデータを要求するノードが存在しないためツリー上の末端のノードとなる. 無線端末等の狭帯域のネットワークしか持たないノードは, ノード G, I の様に末端のノードになると考えられる.

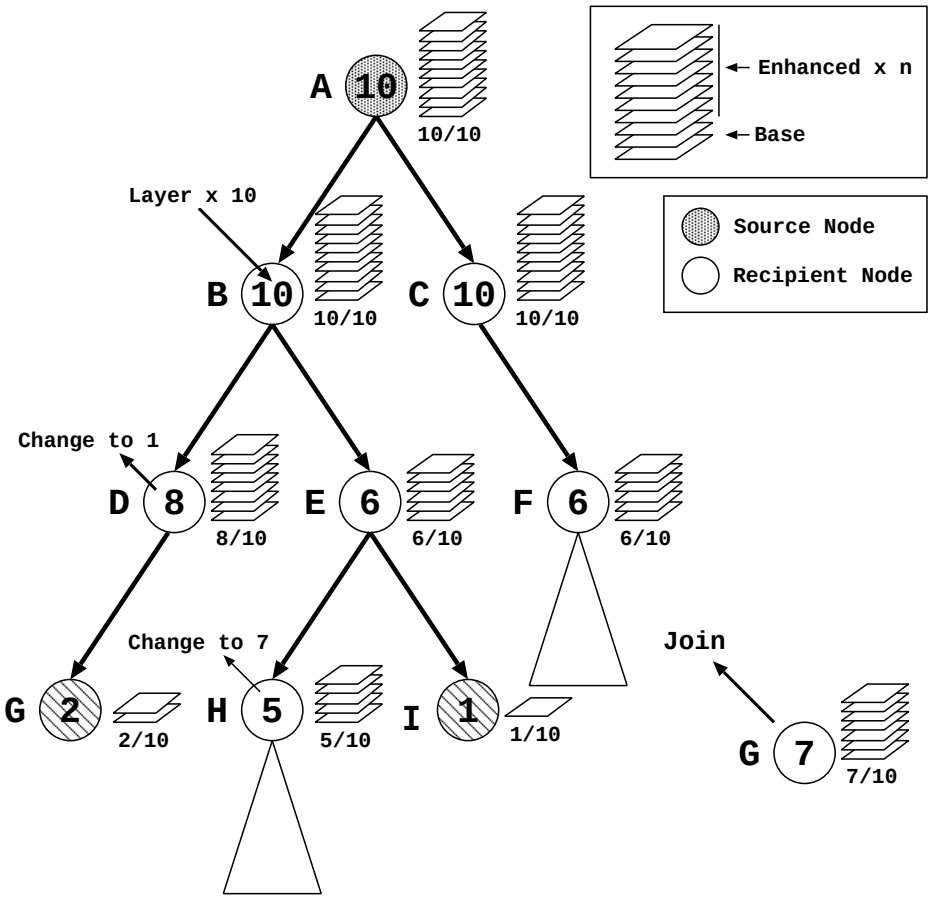

図 4.4: マルチキャストツリーの一例

4.4 実現する機能

LOLCAST により新たにレイヤという単位で配信が行える．これにより, 発信者は個々の受信者の資源環境を意識すること無く適切な配信を行える．この機能の他に, レイヤに分割したことによる新たな機能を提供できる．

4.4.1 ノード障害時におけるデータの保証

不安定な通信基盤上で動作するオーバーレイ・マルチキャストにおいては, 予期せぬノードの障害からマルチキャストツリーを復旧することが必要である．本節では既存のオーバーレイ・マルチキャストプロトコルが用いる復旧手法を紹介する．また, 本プロトコルの階層符号化の利点を活かした新たなノード障害時の効率的な復旧手法を提案する．

ノードの予期せぬ切断であっても, 切断の行なわれたノードの下位に位置するノードは, 継続してサービスを受けられる必要がある．ALMI[14] では, 特にノードの障害時における効果的な手法は存在せず, マルチキャストツリーから分断されたノード群は参加の手続を再度行ないマルチキャストツリーへの復旧を行なう．Narada[6] は Mesh-first 型のプロトコルであり, Contorl Topology 上でリンクを付け足すことによって復旧を行っている．

Host Cast のパス冗長化手法

これに対し, 予期しない切断に対する対処法として, 図 4.5 に示す Host Cast[11] のパス冗長化手法がある．この手法では, 親ノードの冗長化により, 複数の配信パスを事前に確保する．

図 4.5 において, B が予期せぬ切断によって, マルチキャストツリーから切断した際の E での処理の例を示す．定常状態においては, E に対して $A-B-E$ というパスを利用して配信が行なわれている．この時, 冗長化を行なうパスとして, E の親である B と対等な位置関係を持つ C を経由する $A-C-E$, E の親である B のさらに上位ノードである A を経由する $A-E$ を Control Topology 上に確保している．E は B の切断時にこの冗長パスに切り替えることによって, 予期しない切断時における Data Topology の復旧までの時間を短縮する．これにより, ノードの予期しない切断の影響を受けるノードのマルチキャストツリーからの切断時間を抑えることができる．しかしながら, Host Cast の手法は, Control Topology 上で冗長化を行なっているため, マルチキャストツリーが収束するまでの時間が発生する．

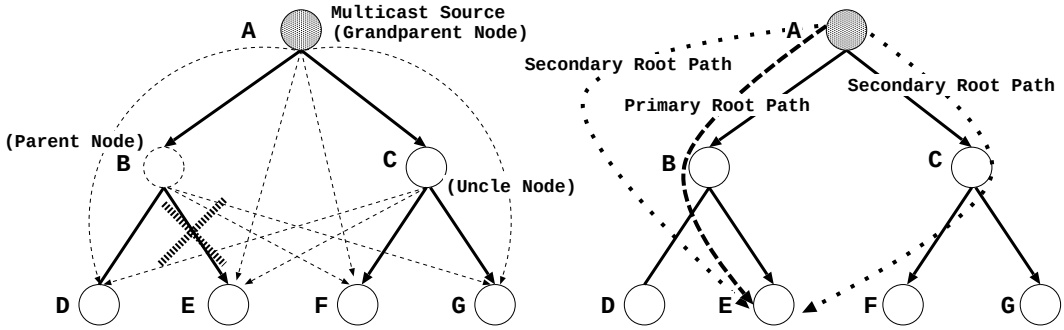

図 4.5: Host Cast における親ノードの冗長化手法

Base レイヤを利用したパス冗長化手法

本プロトコルでは, 階層符号化の利点を活かし, Base レイヤを利用したノード障害時における効率的な保証手法を提案する. 4.1 節で述べたように, 階層符号化において Base レイヤとは最低限の品質を保証するレイヤを指す. Host Cast の手法では Control Topology 上でパスの冗長化を行なっているが本手法では Data Topology 上でパスの冗長化を行なう.

図 4.6において, B が予期しない切断によって, マルチキャストツリーから切断した際の D での処理の例を示す. 定常状態では, $A-B-D$ というパスを利用し D に配信が行なわれている. この時, D は A, C から Data Topology 上で冗長的に Base レイヤのみを受け取っている. B の切断が起こると $A-B-D$ というパスが無効となり, D への配信が停止する. D は B の切断を検知すると即座に A もしくは C から最低限の品質を持つ Base レイヤのデータに切り替え, 情報の損失を防ぐ. 冗長化されたパスから Base レイヤの供給を受けつつ, 上述した HostCast 等の既存のオーバレイ・マルチキャストプロトコルの手法を利用しマルチキャストツリーの復旧を行なう. マルチキャストツリーの復旧を終えると, D は新たな親ノードとなる C から要求した品質でのデータを受信する.

既存のオーバーレイ・マルチキャストプロトコルでは単一的なデータの配信しか考えられていない. そのため, Data Topology 上で冗長化を行なうにはフルレートのデータをメッシュ状に提供する必要がある. これは各ノードへの帯域的な負荷が非常に大きいため現実的でない. 本プロトコルでは Base レイヤのみを利用するため各ノードへの影響を抑え, データの復旧を効率的に行なえる.

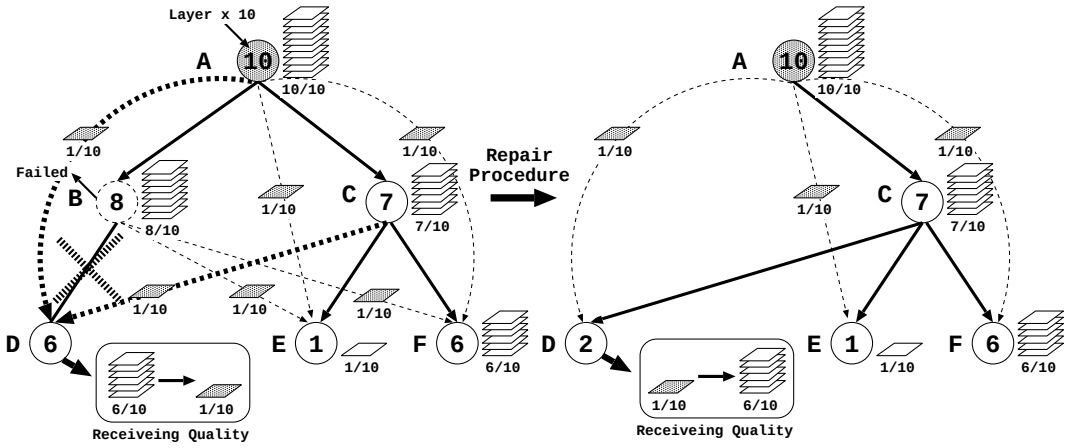

図 4.6: Base レイヤを利用した Data Topology での冗長化

4.4.2 要求品質の幅を利用した輻輳回避

LOLCAST においてデータの品質を自由に制御できる特徴を活かし, 輻輳回避を行なうことができる. 各ノードの要求する品質, つまりレイヤ数に幅を持たせ, 輻輳が発生した際に優先的に上位レイヤのデータを破棄することで, 輻輳の回避を行なう. この例を図 4.7 に示す.

B は A からのデータを E に中継している. B から E への通信路に輻輳が起きていない場合, E は要求した最高品質のデータが得られる. しかし, B が何らかの方法で輻輳を検知した場合, B は上位レイヤから優先的に中継を止め, 輻輳を避ける. 輻輳の検知手法に関しては, 既存研究 [24] を参考にする.

以上のように, レイヤ数の幅によって各ノードからの品質要求に柔軟性を持たせることができる. 要求レイヤ数の幅が狭い場合は要求レイヤ分のサービスを受けられるが, これを満たせない場合には配信が受けられない. 要求レイヤ数の幅が広い場合は, マルチキャストツリーへの変更は発生しにくく, 安定したサービスを受けられる. ただし, 受信品質が上下する可能性がある.

要求レイヤ数の幅を狭く取ることで, 「切断が発生しても要求品質でのサービスを受けたい」と考える利用者を対象とし, 要求レイヤ数の幅を広く取ることで, 「品質が落ちても継続してもサービスを受けたい」と考える利用者を対象とできる.

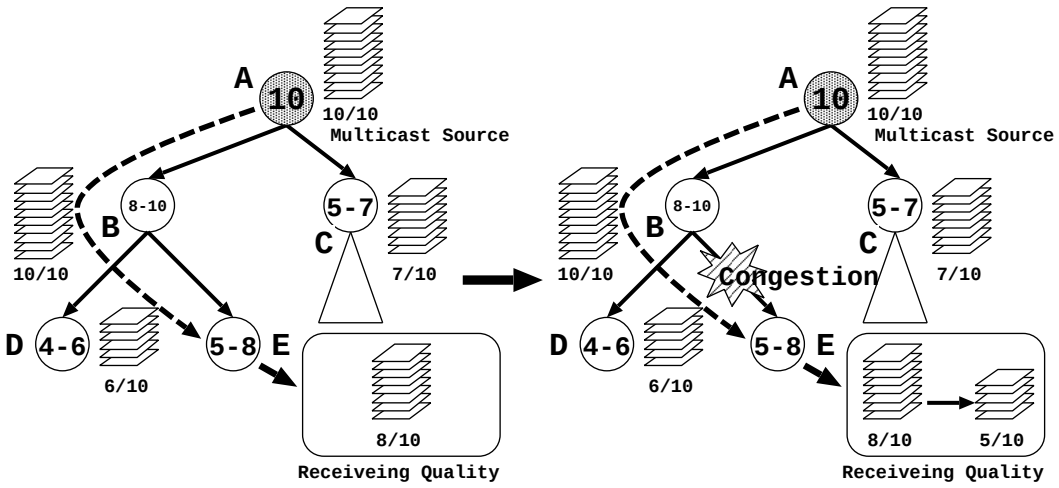

図 4.7: 要求品質の幅を利用した輻輳回避

第5章　LOLCASTの設計

本節では階層的な構造を持つデータの配信に適したオーバーレイ・マルチキャストプロトコルである LOLCAST の設計を行う．はじめに, LOLCAST の想定利用環境を述べる．次に各ノードの定義を行い, それぞれの保持する情報を挙げる．LOLCAST でノード同士の情報交換に用いられる各メッセージとそれを格納するメッセージフォーマットについて述べる．最後にノードの参加や離脱等, マルチキャストツリーの構成を行うための処理である, PPL Extraction, Join Procedure, Leave Procedure について詳細に述べる．

5.1　想定利用環境

LOLCAST の想定利用環境を述べる．本プロトコルの対象とする規模は数千人とした．これは, 個人レベルで運営される最大のコミュニティでも数千人程度であると考え, この規模を想定した．参加する発信者, 受信者の所持するネットワーク資源や計算機資源は, 多種多様であり一般利用者にとって現実的にに入手できるものとする．

5.2　LOLCAST 設計概要

LOLCAST では想定規模が数千人と小さいため, ソースノードが全ノードの情報を保持しマルチキャストツリーの管理を行う手法を取った．これに対し受信ノードはデータの受信, 中継に最低限必要な情報を保持し, ソースノードの指示に従ってマルチキャストツリーを構成する．ノード間の情報のやり取りはメッセージパッシングにより行われ, メッセージを受け取ったノードは自身の保持するデータ構造に更新を加え, 必要であれば返信となるメッセージを送る．

5.3　ノードの定義

本節では, LOLCAST におけるノードとその関係を定義する．図 5.1に一例を示す．

マルチキャストツリーに参加するノードは 2 種類に分けられる．Source Node と Recipient Node である．Source Node は図中の Node$A$ にあたり, データの配信元となるノードを指す．Recipient Node は図中の NodeB, C, D にあたり, 配信されたデータを受け取り中継するノードを指す．

図 5.1においてノード D がマルチキャストツリー上で上位に位置するノード B からデータを受け取っていた際, ノード D はノード B の子ノードであり, Child Node と定義される．これに対応してノード B はノード D の親ノードであり, の Parent Node と定義される．マルチキャストツリーに新たに参加するノード E は New Node と定義され, ノード E の Parent Node となる条件を満たしているノード A, B, C を Potential Parent Node と定義する．最後に, マルチキャ

ストツリーから離脱を行うノード B を Leave Node と定義する.

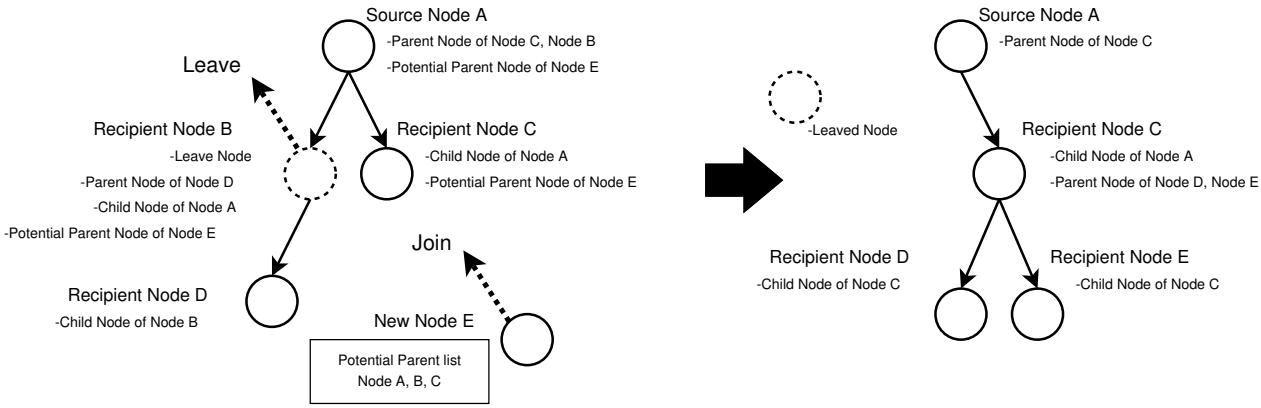

図 5.1: LOLCAST におけるノードの定義

5.4 各ノードの保持する情報

5.3節で述べたように, マルチキャストツリーに参加するノードは Source Node と Recipient Node の 2 種類に分けられる. 本節ではそれぞれのノードの保持する情報を述べる. はじめに各ノードの保持する情報を述べる上で必要となる, ノード情報について述べる. ノード情報は各ノードの基本情報を格納しており, 自身や他ノードの保持に用いられる. 次に Source Node の保持する情報について述べる. Source Node の保持する情報にはマルチキャストツリーの管理を行う, ツリー構造が含まれる. 最後に Recipient Node が保持する情報を述べる. Recipient Node はデータの受信・転送に最低限必要となる情報を保持している.

5.4.1 ノード情報

ノード情報は LOLCAST における各ノードの基本情報を格納している. ノード情報により, Source Node や Recipient Node は自身や他のノードの情報を保持している. また, ノード情報は LOLCAST で行われるメッセージ交換においてノードを特定する情報として利用される. 本節ではノード情報に含まれるそれぞれの情報について述べる.

- **ノード識別子**
  ノードに対して Source Node により一意に割り振られる識別子. ツリー構造におけるノードの検索の際にキーとして用いられる.

- **要求・保持レイヤ数**
  ノードの要求・保持するデータのレイヤ数を示す. Source Node の保持する値が最大となり, 1 レイヤが最小値となる.

- **ツリーにおける深さ**
  ノードのマルチキャスト上で位置する深さを示す. Source Node の値が最小の 1 となる.

- **対応するネットワーク層プロトコル**
  ノードの利用可能なネットワーク層プロトコルを示す．例として IPv4 や IPv6 等が挙げられる．

- **アドレスのリスト**
  ノードの利用可能なネットワーク層プロトコル毎のアドレスが格納されているリスト．メッセージの送受信やデータの送受信の際に用いられる．

5.4.2 Source Node 固有情報

本節では Source Node の保持する情報を述べる．Source Node は自身のノード情報, Child Node のリスト, ツリー構造を保持しており, それぞれについて述べる．この中のツリー構造はマルチキャストツリーの状態保持を行っており, これに含まれる情報とそのエントリとして用いられるツリーノード情報について述べる．

- **自身のノード情報**
  自身のノード情報の保持を行う．

- **Child Node のリスト**
  自身の管理する Child Node のノード情報のリスト．

- **ツリー構造**
  マルチキャストツリーの状態保持を行う．

ツリー構造

ツリー構造はマルチキャストツリーの状態保持を行い, Source Node によって管理される．Recipient Node からマルチキャストツリーの更新が伝えられると, Source Node はツリー構造に対して変更を行う．

- **全ノードを含むテーブル**
  全ノードのツリーノード情報が登録されているテーブル．ノード識別子をキー, ツリーノード情報をデータとして持つ．主にノードの検索に利用される．

- **Leave Node のリスト**
  離脱処理中である Leave Node のノード識別子のリスト．Leave Node の状態をロックするために用いられる．

ツリーノード情報

ツリーノード情報はツリー構造の各エントリとして利用される．ツリーノード情報には, 基本情報であるノード情報の他に, マルチキャストツリーの保持を行う際に必要となる Parent Node や Child Node の情報が含まれる．

- **自身のノード情報**
  ツリーノードの基本情報としてノード情報が格納される．

- **Parent Node のツリーノード情報**
  自身の Parent Node のツリーノード情報が格納される．ツリー構造で Parent Node を参照する際に利用される．

- **Child Node のリスト**
  自身の管理する Child Node のノード識別子のリスト．ツリー構造で Child Node を参照する際に利用される．本リストに含まれるノード識別子をキーとして Source Node の保持する全ノードを含むテーブルから Child Node のツリーノード情報が検索, 参照される．

5.4.3 Recipient Node 固有情報

Recipient Node は自身の子ノードや親ノードの情報等, データの送受信に最低限の情報を保持している．Recipient Node の所持する情報を以下に示す．

- **自身のノード情報**
  自身のノード情報の保持を行う．

- **Source Node のノード情報**
  Source Node のノード情報の保持を行う．

- **Parent Node のノード情報**
  Parent Node のノード情報の保持を行う．

- **Child Node のリスト**
  自身の管理する Child Node のノード情報のリスト．

- **Potential Parent のリスト (PPL)**
  自身の Parent Node となりうるノードのリスト．Source Node によって生成され, Recipient Node はこの PPL を利用して Parent Node に対しマルチキャストツリーへの参加要求を行う．

5.5 メッセージ群

LOLCAST はメッセージパッシングにより, ノード間の情報のやり取りを行う．メッセージはその機能別にいくつかのメッセージ群に分けられる．本節では LOLCAST で利用するメッセージ群について述べる．メッセージは, Rendezvous Message 群, PPL Message 群, Join Message 群, Data Message 群, Leave Message 群, Prune Message 群, Notify Message 群の 7 つの群に分けられる．

5.5.1 Rendezvous Message

Rendezvous Message 群は New Node がマルチキャストツリーに参加するために必要となる情報を Source Node に対し要求するためのメッセージ群である．

- **Rendezvous Request (Recipient Node to Source Node)**
  Recipient Node が Source Node に対し, マルチキャストツリーへの参加に必要となる情報を要求するメッセージ. メッセージには, Recipient Node のノード情報が含まれる. メッセージを受けた Source Node は Recipient Node と自身の対応するネットワーク層プロトコルを比較し, 満たしていれば Rendezvous Accept, 対応していなければ Rendezvous Reject を Recipient Node に返す.

- **Rendezvous Accept (Source Node to Recipient Node)**
  Source Node が Recipient Node に対し, Rendezvous Request の承認を伝えるメッセージ. メッセージには, 新たに生成された Recipient Node のノード識別子と自身のノード情報が含まれる. メッセージを受けた Recipient Node はノード識別子と Source Node のノード情報を格納する.

- **Rendezvous Reject (Source Node to Recipient Node)**
  Source Node が Recipient Node に対し, Rendezvous Request の拒否を伝えるメッセージ. メッセージを受けた Recipient Node は, マルチキャストツリーへ参加できない.

5.5.2 PPL Message

PPL Message 群は, Potential Parent Node のリスト構造である Potential Parent List (PPL) の送受信を行う. PPL については, 5.2.5 節にて述べる.

- **PPL Request (Recipient Node to Source Node)**
  Recipient Node が Source Node に対し PPL を要求するメッセージ. メッセージには Recipient Node のノード情報が含まれる. メッセージを受けた Source Node は, PPL Data メッセージにより PPL を Recipient Node に対し送る.

- **PPL Data (Source Node to Recipient Node)**
  Source Node が Recipient Node に対し, PPL を送信するメッセージ. Source Node は PPL Request によって取得した Recipient Node のノード情報を用い PPL を自身の保持するツリー構造を利用し生成する. この際 Recipient Node の要求レイヤ数, ツリー上での深さ, Child Node の数を利用し優先度順に並べられた PPL が作られる. メッセージを受けた Recipient Node は PPL を自身のノード情報に格納する.

5.5.3 Join Message

Join Message 群は PPL Message 群により PPL を入手したノードが Potential Parent Node に対し, 自身を Child Node として受け入れる要求を行う.

- **Join Request (Child Node to Potential Parent Node)**
  Child Node が Potential Parent Node に対し, 自身を子ノードとして受け入れるよう要求するメッセージ. この際の Potential Parent Node は PPL に格納されている中で最も優先度の高いものが選ばれる. メッセージには Child Node のノード情報が含まれる. Child Node はメッセージ送信後に PPL から Potential Parent Node のエントリを取り除く. メッ

セージを受けた Potential Parent Node は, 自身のノード情報と Child Node のノード情報から保持レイヤ数が要求レイヤ数を満たしているか, 管理子ノード数に空きがあるかの比較を行う. 要求を満たしていれば, Child Node のノード情報が Potential Parent Node の子ノード情報のリストに追加され, Join Accept を Child Node に返す. 参加が不可能であれば Join Reject を Child Node に返す.

- **Join Accept (Potential Parent Node to Child Node)**
  Potential Parent Node が Child Node に対し, 子ノードとして受け入れたことを伝えるメッセージ. メッセージには Potential Parent Node のノード情報が含まれる. メッセージを受けた Child Node は Potential Parent Node のノード情報を自身のノード情報に格納する.

- **Join Reject (Potential Parent Node to Child Node)**
  Potential Parent Node が Child Node に対し, 子ノードとしての受け入れの拒否を伝えるメッセージ. メッセージを受けた Child Node は次に優先度の高い Potential Parent Node に対し Join Request を送る.

5.5.4 Data Message

Data Message 群は自身の Parent Node に対してデータの送信開始要求, 送信停止要求を行う.

- **Data StartRequest (Child Node to Parent Node)**
  Child Node が Parent Node に対し, データの送信開始を要求するメッセージ. メッセージには Child Node のノード情報が含まれる. メッセージを受けた Parent Node は Child Node のノード情報のアドレスリストに含まれるアドレスに対しデータの送信を開始する.

- **Data StopRequest (Child Node to Parent Node)**
  Child Node が Parent Node に対し, データの送信停止を要求するメッセージ. メッセージには Child Node のノード情報が含まれる. メッセージを受けた Parent Node は Child Node に対するデータの送信を停止する.

5.5.5 Leave Message

Leave Message 群は主にマルチキャストツリーから離脱を行うノードが自身の Child Node に対し, 離脱の広告を行う.

- **Leave Request (Parent Node to Child Node)**
  Parent Node が Child Node に対し, 自身のマルチキャストツリーからの離脱を広告し, 他の Parent Node への移動を要求するメッセージ. メッセージは Parent Node の管理する全 Child Node に対し送信される.

- **Leave Complete (Child Node to Parent Node)**
  Child Node 以前の Parent Node に対し, 他の Parent Node への移動の完了を伝えるメッセージ. メッセージには Child Node のノード情報が含まれる. メッセージを受けた Parent Node は, 子ノードのリストから Child Node のノード情報を取り除く.

5.5.6 Prune Message

Prune Message 群は Child Node が Parent Node に対して, 自身をマルチキャストツリーから切り離す要求を行う.

- **Prune Request (Child Node to Parent Node)**
  Child Node が Parent Node に対し, 自身の切り離しを要求するメッセージ. この際, Child Node は管理する子ノードが存在しない必要がある. メッセージには Child Node のノード情報が含まれる. メッセージを受けた Parent Node は子ノードのリストから Child Node のノード情報を取り除き, Prune Complete を Child Node に返す.

- **Prune Complete (Parent Node to Child Node)**
  Parent Node が Child Node に対し, Child Node の切り離しの完了を伝えるメッセージ. メッセージを受けた Child Node は, マルチキャストツリーの更新を伝えるため Notify LeaveComplete を Source Node に送信する.

5.5.7 Notify Message

Notify Message 群は Source Node に対し, ノードの参加や離脱によって発生したマルチキャストツリーへの変更を更新するよう要求を行う.

- **Notify JoinComplete (Recipient Node to Source Node)**
  Recipient Node が Source Node に対し, マルチキャストツリーへの参加の完了によるツリーの更新を伝えるメッセージ. メッセージには Recipient Node とその Parent Node のノード情報が含まれる. メッセージを受けた Source Node は Recipient Node の Parent Node をツリー構造から検索し, その子ノードとして Recipient Node を追加する.

- **Notify LeaveProgress (Recipient Node to Source Node)**
  Recipient Node が Source Node に対し, 自身がマルチキャストツリーからの離脱処理中であることを伝えるメッセージ. メッセージには Recipient Node のノード情報が含まれる. メッセージを受けた Source Node は Recipient Node のノード情報を離脱処理中であるノードのリストに格納する.

- **Notify LeaveComplete (Recipient Node to Source Node)**
  Recipient Node が Source Node に対し, マルチキャストツリーからの離脱の完了によるツリーの更新を要求するメッセージ. この際 Recipient Node は管理する子ノードや親ノードが存在しない必要がある. メッセージには Recipient Node のノード情報が含まれる. メッセージを受けた Source Node は, ツリー構造から Recipient Node を取り除く.

- **Notify ParentChanged (Recipient Node to Source Node)**
  Recipient Node が Source Node に対し, 親ノードの変更によるマルチキャストツリーの更新を伝えるメッセージ. メッセージには, Recipient Node と新たな Parent Node のノード情報が含まれる. メッセージを受けた Source Node は Recipient Node のノード情報を用い, 以前の Parent Node から新たな Parent Node へ Recipient Node の位置をマルチキャストツリー上で更新する.

5.6 メッセージフォーマット

LOLCAST で用いられる各メッセージは図 5.2に示すフォーマットに格納される．メッセージフォーマットの各フィールドは大きく 4 つに分けられる．本節ではそれぞれのフィールドについて述べる．

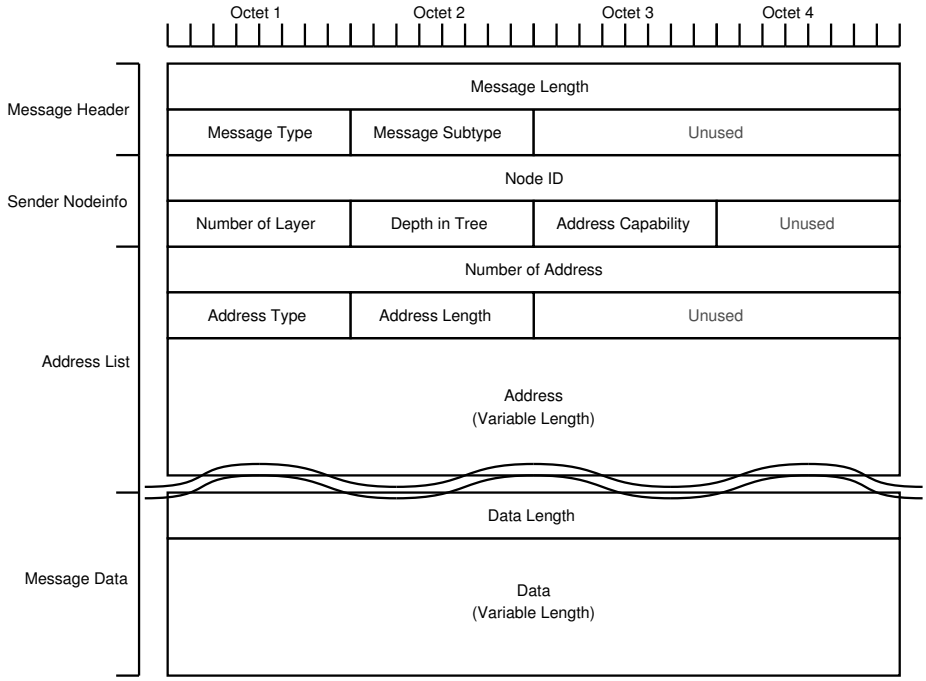

図 5.2: メッセージフォーマット

- **Message Header**
  Message Header にはメッセージの自体の情報を示すフィールドが格納される．**Message Length** フィールドにメッセージ長が格納され，**Message Type**, **Message Subtype** フィールドに 5.5節で述べたメッセージの種類が格納される．

- **Sender Nodeinfo**
  Sender Nodeinfo は送信ノードのノード情報を格納するフィールドである．5.3節で述べたノード情報に含まれるノード識別子, 所持・要求レイヤ数, ツリー内で位置する深さ, 対応するネットワーク層プロトコルがそれぞれ, **Node ID**, **Number of Layer**, **Depth in Tree**, **Address Capability** フィールドに格納される．

- **Sender Address List**
  Sender Address List にはメッセージの送信ノードのアドレスリストが格納される．**Number of Address** にメッセージ中に含まれるアドレスの数が格納される．**Address Type** フィールドに IPv4 や IPv6 等のネットワーク層プロトコルの種類が格納される．**Address** フィールドは可変長であり, アドレスが格納される．

- **Message Data**
  Message Data には，メッセージにより送信されるデータである PPL 等が格納される．

Data Length フィールドにデータ長が格納される．**Data** フィールドは可変長でありデータが格納される．

5.7 マルチキャスト・ツリーの構成

本節では, LOLCAST のマルチキャストツリーの構成手法について詳細に述べる．はじめにマルチキャストツリーの構成について述べる上で必要なパラメータの定義を行う．次に, PPL の抽出を行う PPL Extraction, 新たなノードがマルチキャストツリーに参加する際に必要となる Rendezvous Procedure について述べる．最後にノードがマルチキャストツリーに参加・離脱する際の処理である Join Procedure, Leave Procedure について述べる．

5.7.1 パラメータの定義

マルチキャストツリーの構成手法について述べる行う上で利用する各ノードとパラメータの定義を行う．図 5.1にパラメータとその定義を示す．Source Node と Recipient Node はそれぞれ S, R_n と定義する．次にノードの要求・保持レイヤ数を L_n, ツリーにおける深さを D_n と定義する．最後にノードの管理する Child Node の数を C_n と定義する．

表 5.1: ノードとパラメータの定義

Source Node	S
Recipient Node	$R_1, R_2, R_3, ...R_n$
Number of Layer	L_n
Number of children node(s)	C_n
Depth	D_n

5.7.2 PPL Extraction

Source Node が全ノードの情報を保持し, マルチキャストツリーを管理する LOLCAST において, Source Node による Parent Node の決定が大きくマルチキャストツリーの品質を左右する．本節では, 最適な Potential Parent Node の選択をし, Potential Parent List の抽出を行う PPL Extraction を図 5.3を用いて解説する．

PPL Extraction は 2 つのフェーズにより構成される．Temporary PPL Extraction と PPL Extraction である． Temporary PPL Extraction により, 一定量のノードを抜き出し, PPL Extraction により抜き出されたノードの中から最適なノードを選出する．比較を行うノード数を示すパラメータを設け, その変更が可能である．Potential Parent Node は以下の条件を満たす必要がある．

- Child Node の要求レイヤ数を満たしている
- Leave Node のリストに含まれていない
- 管理 Child Node 数に空きがある

以上の条件を満たすノードから, マルチキャストツリーにおける深さが最小となるノードから優先度が高い Potential Parent Node として, PPL に格納される.

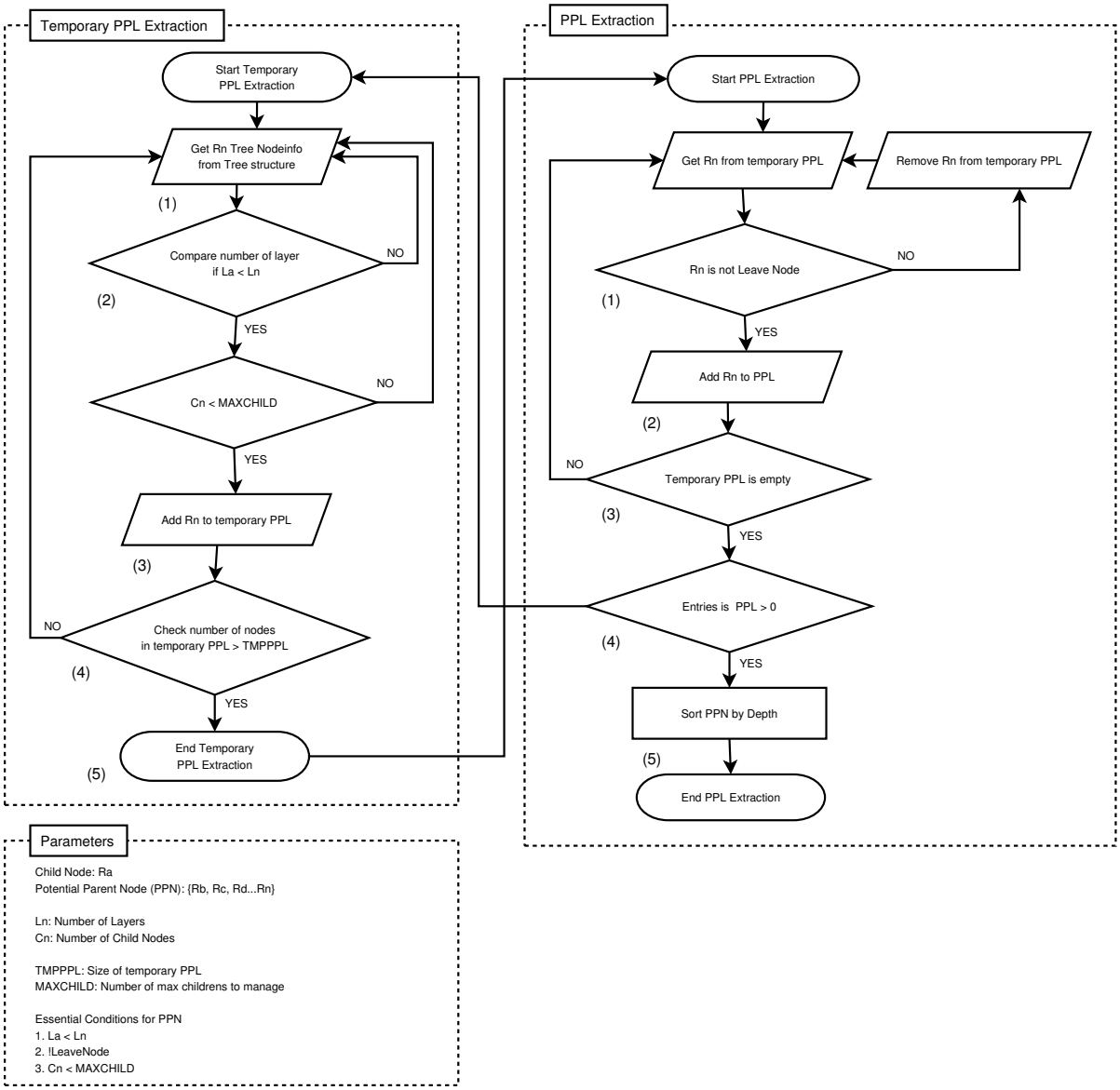

図 5.3: PPL Extraction の流れ

Temporary PPL Extraction

1. Source Node の保持する, ツリー構造内の全ノードを含むテーブルよりツリーノード情報を取り出す
2. Child Node の要求レイヤを満たしているか, Child Node 数に空きがあるかの比較を行う
3. 条件を満たしたツリーノード情報を Temporary PPL に追加する
4. 条件を満たしたノードが一定量に達するまで 1-3 を繰り返す

5. 抽出された Temporary PPL を PPL Extraction に引き渡す

PPL Extraction

1. Temporary PPL に格納されているノードが Leave Node のリストに含まれていないか比較を行う
2. 含まれていないノードを PPL に追加する
3. Temporary PPL のエントリが空になるまで 1-2 を繰り返す
4. Temporary PPL にエントリが存在しない状態で, PPL に一つもエントリが存在しない場合には Temporary PPL Extraction が再度実行される
5. PPL に含まれるエントリをマルチキャストツリーにおける深さが浅い順に並び替えられ, PPL Extraction が完了する

5.7.3 Rendezvouz Procedure

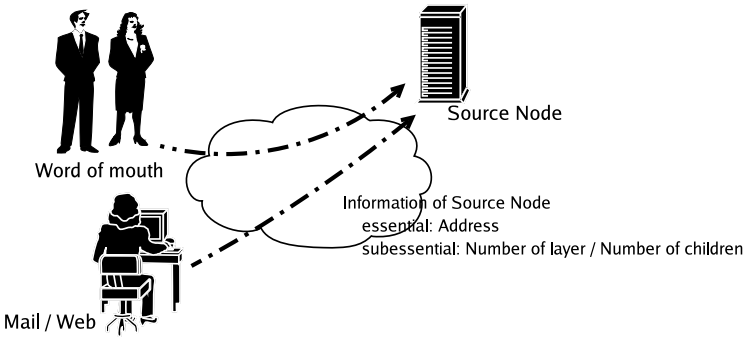

図 5.4: Rendezvous Procedure

New Node はマルチキャストツリーに参加するために Source Node の情報が必要となる. Source Node の情報は図 5.4に示すように, Web ページ/メールや人づてに伝えられる. Source Node の情報の入手は以上のような手段によって行うことを想定し, プロトコルの処理には含まれない.

5.7.4 Join Procedure

本節では, New Node がマルチキャストツリーに参加する際の手順を図 5.5, 5.6, 5.7, その際行われるメッセージパッシングの様子を図 5.8を用い解説する. 例で用いるマルチキャストツリーには S_a, R_b, R_c, R_d の 4 ノードが参加しており, S_a が Source Node となる. このようなマルチキャストツリーに, New Node R_e が新たにマルチキャストツリーに参加するすることを想定する. また, 各ノードの管理する Child Node の最大数を 2 として進める.

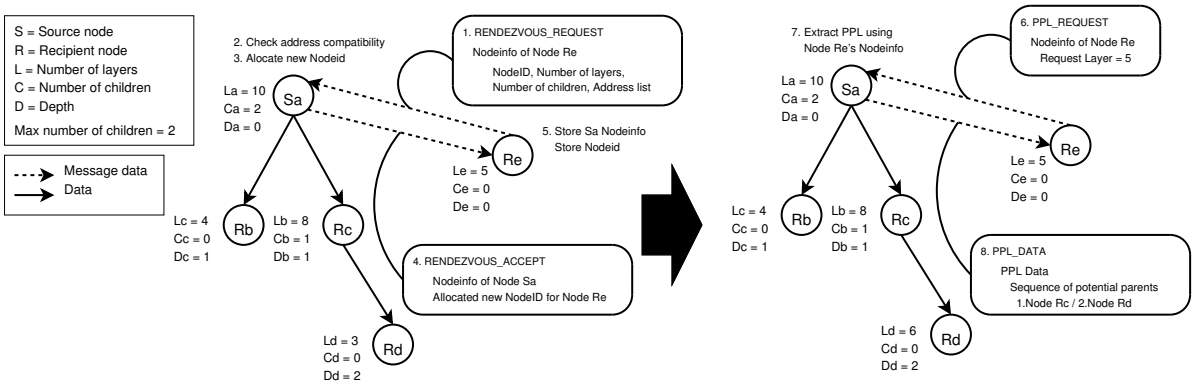

図 5.5: Join Procedure 1/3

1. 新たにマルチキャストツリーに参加する R_e は 5.7.3節で述べた Rendezvous Procedure により取得した Source Node S_a のノード情報を元に **Rendezvous Request** を S_a に対し送信する.

2. **Rendezvous Request** を受けた S_a はメッセージに含まれる R_e の対応するネットワーク層プロトコルを自身のそれと比較を行う.

3. S_a の対応するネットワーク層プロトコルに適合していれば, 新たに R_e のノード識別子を生成し割り当てる.

4. 割り当てを行った R_e のノード識別子をメッセージに格納し, **Rendezvous Accept** を R_e に対し送信する.

5. **Rendezvous Accept** を受けた R_e は, S_e のノード情報を格納し, さらにノード識別子を自身のノード情報に格納する.

6. 次に R_e は自身の親ノードとなるノードを探すために S_a に対して **PPL Request** を送信する. メッセージには R_e の要求レイヤ数である L_e が含まれる.

7. S_a は 5.7.2節で述べた手法を用い, PPL の生成を行う. この際, メッセージに含まれる R_e の要求レイヤ数がパラメータとして渡される. この例では R_c, R_d が PPL に優先度順に格納される.

8. 生成した PPL をメッセージに格納し, **PPL Data** を R_e に対し送信する.

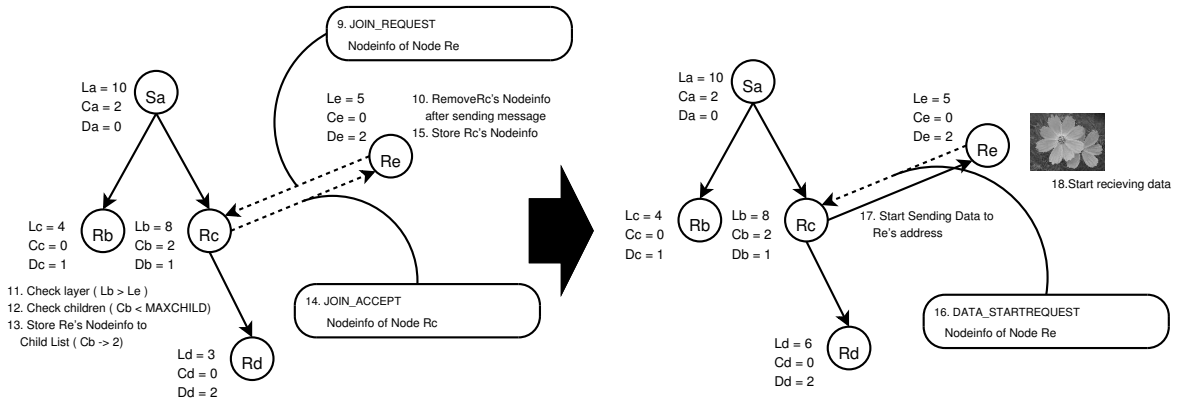

図 5.6: Join Procedure 2/3

9. メッセージを受信した R_e は, PPL から最も優先度の高い R_c のノード情報を元に, **Join Request** を R_c に対し送信する.

10. **Join Request** を送信し終えた R_e は R_c のノード情報を PPL から削除する.

11. R_c は **Join Request** に含まれる R_e のノード情報を用い, Child Node として参加が可能であるかを判断する. まず R_e の要求レイヤ数が R_c の保持レイヤ数を満たしているか, 比較を行う ($L_c > L_e$). この場合の R_e の要求レイヤ数は 5, R_c の保持レイヤ数は 8 であるため参加可能である.

12. 次に, R_c の管理子ノード数に空きがあるかの比較を行う ($C_c <$ MAXCHILD). この際の最大管理子ノード数は 2 であるため, 参加可能である.

13. R_e を子ノードとして受け入れ可能と判断した R_c は R_e のノード情報を子ノードのリストに格納する.

14. R_c は子ノードとして受け入れ可能である事を R_e に伝えるため, **Join Accept** を R_e に対し送信する.

15. メッセージを受信した R_e は, R_c のノード情報を親ノードのノード情報に格納する.

16. 子ノードとしての受け入れが完了した R_e はデータの要求を行うために, R_c に対して **Data StartRequest** を送信する.

17. **Data StartRequest** を受けた R_c は要求レイヤ分のデータの送信を R_e に対し開始する.

18. R_e はデータの受信を開始する.

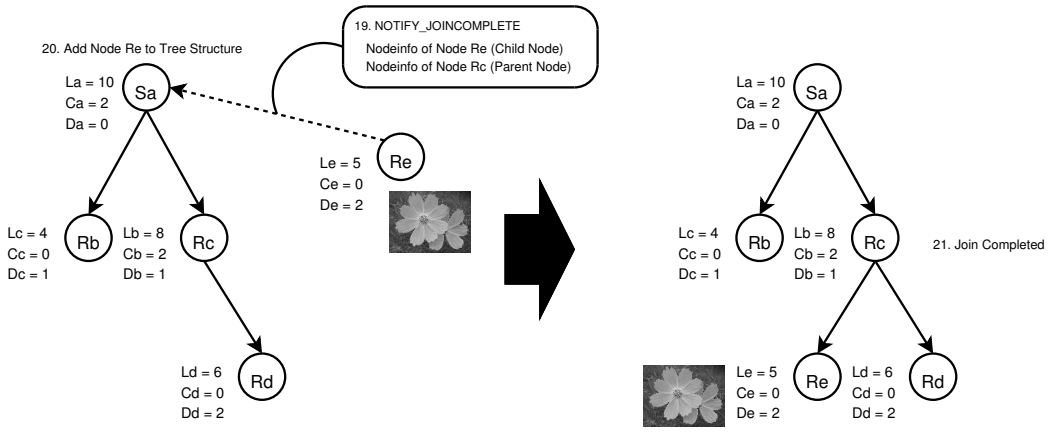

図 5.7: Join Procedure 3/3

19. R_e はマルチキャストツリーが更新された事を S_a に伝えるために **Notify JoinComplete** を S_a に対し送信する．このメッセージには, 親ノードと子ノードの対, R_c と R_e のノード情報が含まれる．

20. **Notify JoinComplete** を受けた S_a は, まず R_c のツリーノード情報を R_c のノード識別子を用い検索する．R_c のツリーノード情報に子ノードとして R_e を追加する．以上で R_e のマルチキャストツリーへの参加が完了する．

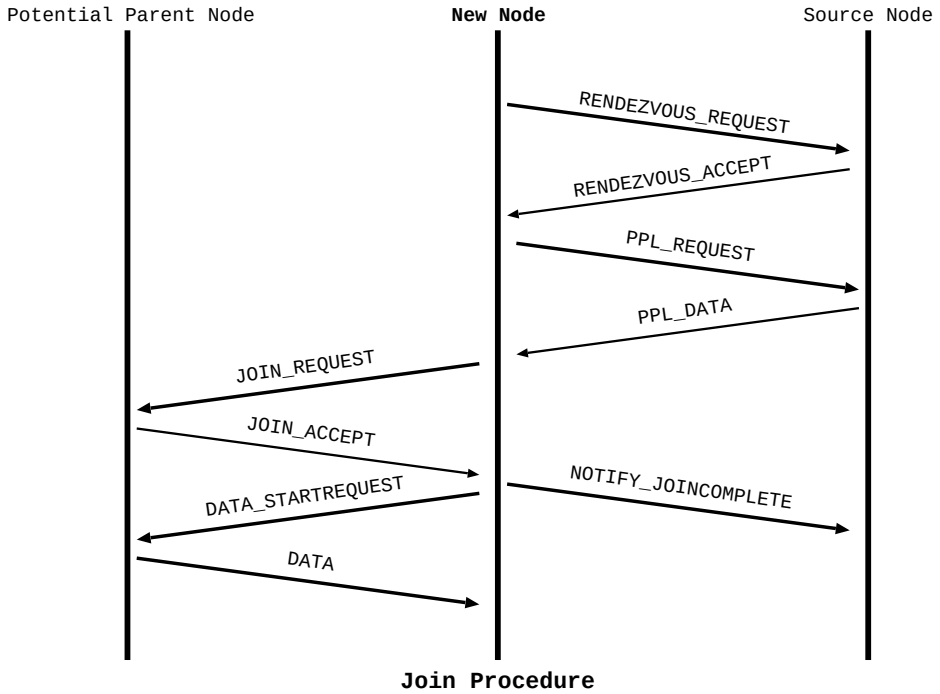

図 5.8: Join Procedure におけるメッセージパッシング

5.7.5 Leave Procedure

本節では, Leave Node がマルチキャストツリーからの離脱を行う際の手順を図 5.9, 5.10を用い解説する. この際のメッセージパッシングの様子を図 5.11に示す. 例で用いるマルチキャストツリーには S_a, R_b, R_c, R_d, R_e, R_f の 6 ノードが参加しており, S_a が Source Node となる. このようなマルチキャストツリーより, Child Node R_f を管理している R_e がマルチキャストツリーからの離脱を行った場合を想定する. また, 各ノードの管理する Child Node の最大数は 2 とする. Leave Procedure において Leave Node は, 自身の Child Node 以下のノードに影響を与えない離脱を行うことがが重要である.

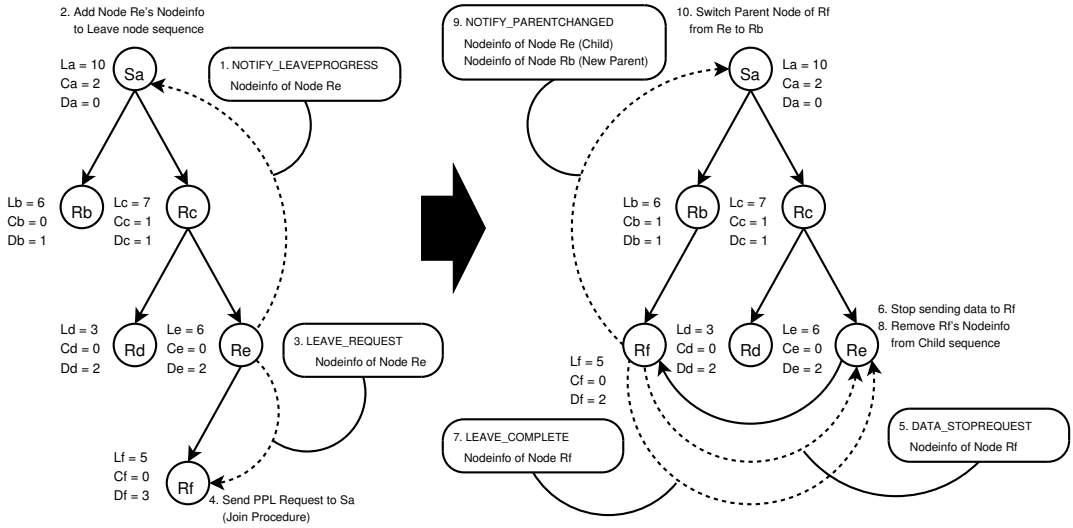

図 5.9: Leave Procedure 1/2

1. マルチキャストツリーからの離脱を行う R_e は, S_a に対し, **Notify LeaveProgress** を送る. **Notify LeaveProgress** は自身が Leave Procedure を実行中であることを Source Node に伝え, PPL Extraction の際に PPL に自身を含めない様, 伝える役割がある.

2. **Notify LeaveProgress** を受けた S_a はツリー構造内の離脱処理中ノードのリストに R_e を追加する.

3. R_e は Child Node 以下のノードに影響を与えずに離脱を行うために, 自身からの離脱を促す **Leave Request** を自身の Child Node のリストに含まれる全 Child Node に対し送信する. この際 R_e の Child Node は R_f であるため, R_f に対して **Leave Request** が送信される.

4. **Leave Request** を受けた R_f は再度 Join Procedure を行い, 新たな Parent Node に対し参加を行う. 新たに S_a から **PPL Request** により PPL を受け取った R_f は, R_b を新たな Parent Node としてマルチキャストツリーに参加する. R_b から **Data StartReqeuest** によりデータを受け取り始める段階で, 以前の Parent Node である R_e と R_b の 2 つの Parent Node からデータを受けている.

5. R_f は新たに受け取り始めた R_b からのデータへ切り替えが完了すると同時に, 以前の Parent Node である R_e に対しデータの送信停止要求を行う **Data StopRequest** を送信

する.

6. **Data StopRequest** を受けた R_e は R_f に対するデータの送信を停止する.

7. 以前の Parent Node である R_e からのデータの送信が停止し, 単一の Parent Node からのデータを受けている状態になった R_f は, **Leave Complete** を R_e に対し送信する.

8. **Leave Complete** を受けた R_e は R_b は管理 Child Node のテーブルから R_f のノード情報を削除する. これにより, R_e は Child Node を一切管理していない状態となる.

9. R_f は Parent Node が R_e から R_b に変更され, マルチキャストツリーが更新されたことを Source Node に伝える必要がある. R_f は新たな Parent Node である R_b と自身のノード情報を含んだ **Notify ParentChanged** を S_a に送信する.

10. **Notify ParentChanged** を受けた S_a は, R_f の Parent Node を R_e から R_b に繋ぎ変え, ツリー構造を最新の状態に更新する.

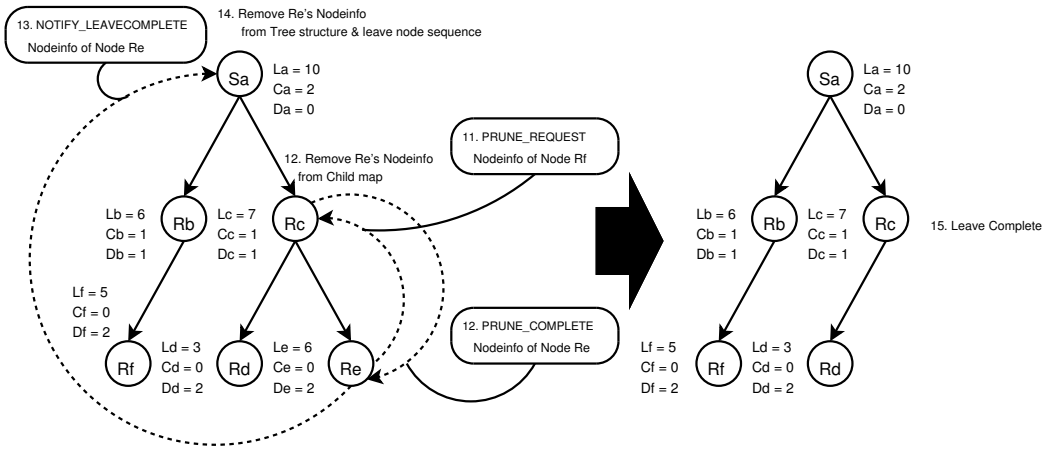

図 5.10: Leave Procedure 2/2

11. Leave Procedure の 9. において Child Node 一切管理しない状態になった R_e は Parent Node である R_c に対し自身を切り離す要求を行う **Prune Request** を送信する.

12. **Prune Request** を受けた R_c は管理子ノードのリストから R_e のノード情報を削除する. 削除を終えた, R_c は **Prune Reuqest** への返信として R_e に **Prune Compelte** を送信する.

13. **Prune Complete** を受けた R_e は, Parent Node も管理する Child Node も一切無く, マルチキャストツリーから離脱した状態となる. 最後に R_e は Source Node に対し **Notify LeaveCompelte** を送り, 自身をツリー構造から削除するように要求する.

14. **Notify LeaveComplete** を受けた, S_a はツリー構造から R_e のツリーノード情報を削除し, Leave Procedure が完了する.

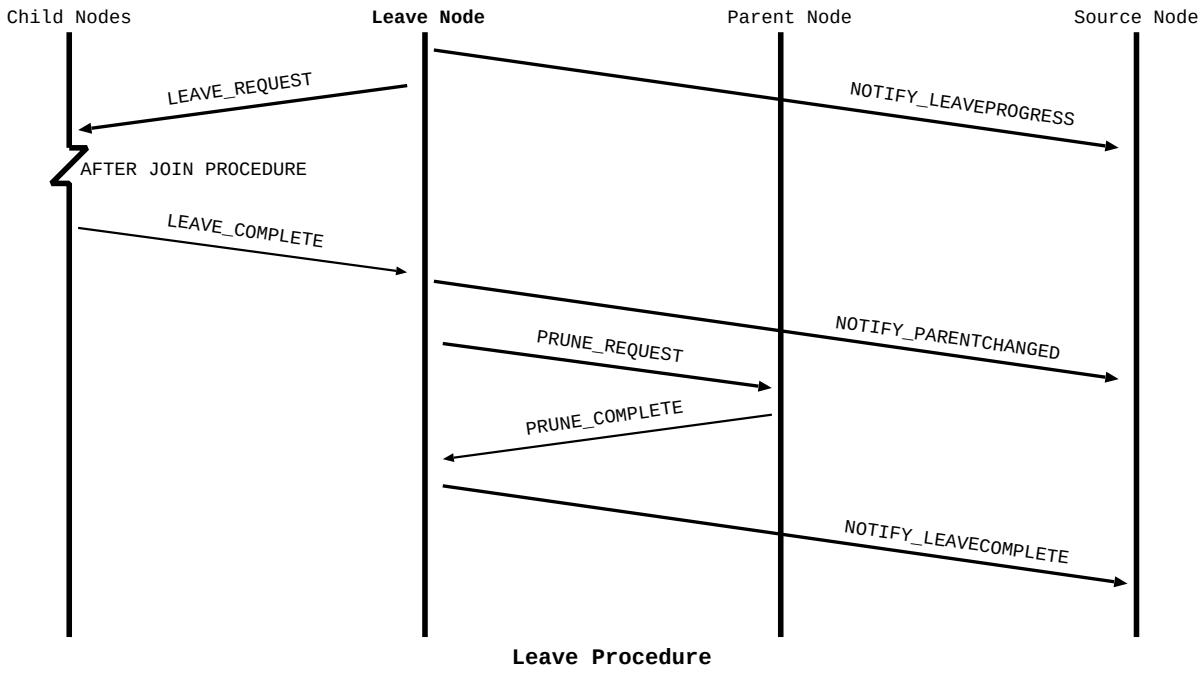

図 5.11: Leave Procedure におけるメッセージパッシング

第6章　実装

6.1　実装概要

LOLCAST プロトコルのシミュレータによる評価, およびアプリケーションによる評価を単一のプロトコル処理部分で行うために図 6.1のようなモジュールに分割されたシステムを実装した. LOLCAST は 4 つのモジュールより構成される. LOLCAST プロトコル処理モジュール, シミュレーションモジュール, アプリケーションモジュール, ネットワークモジュールの 4 つである. 各モジュールの持つ機能と関係を述べる.

- **LOLCAST プロトコル処理モジュール**
  LOLCAST プロトコル処理モジュールは本システムの核となるプロトコルに関連する処理を行うモジュールである. 本モジュールにより, 到着メッセージによるデータ構造への処理や, それに対する返信となるメッセージを下部に位置するモジュール接合部に引き渡す役割を担っている. メッセージパッシングを用いたプロトコルの処理により, マルチキャストツリーの構築を行う.

- **モジュール接合部**
  モジュール接合部ではネットワークモジュールやシミュレータモジュールから引き渡されたメッセージをプロトコル処理モジュールに対しメッセージにあった処理への振分けをおこなう. また, プロトコル処理モジュールにおいて処理を終えた返信のメッセージをネットワークモジュールやシミュレータモジュールに引き渡す役割を担っている.

- **アプリケーションモジュール**
  アプリケーションモジュールはデータをプロトコル処理モジュールより指定されたフォーマットに従いレイヤ分割結合し, ネットワークモジュールにその送信要求を行う. 受信したデータを表示する機能もここで提供される. レイヤ数, 送信先の情報のみが与えられるため, データフォーマットに依存すること無く処理ができる.

- **ネットワークモジュール**
  ネットワークモジュールは, 指定されたアドレスに対し渡されたデータの送信, 受信したデータのアプリケーションモジュールへの引き渡しの機能を担う.

- **シミュレータモジュール**
  シミュレータモジュールでは, 仮想的に複数のノードを用意しプロトコル処理モジュールからのメッセージに従った処理を指定された仮想的なノードに対して行う.

6.2　実装環境

本実装を行った環境を表 6.1に示す.

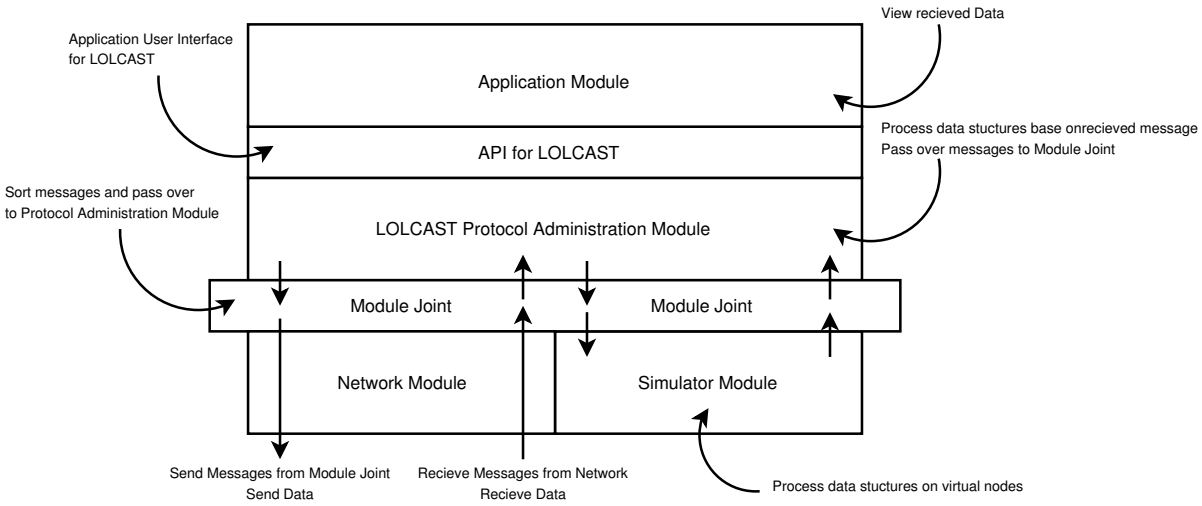

図 6.1: LOLCAST システム図

表 6.1: 実装環境

プロセッサ	PowerPC G4 1.25Ghz / Pentium III 1Ghz
メモリ	768MB / 512MB
OS	MacOSX 10.3.7 / NetBSD 1.6.2
開発言語	C 言語
利用ライブラリ	ctmpl

6.2.1 CTmpLib

C 言語においてテンプレートを扱う CTmpLib を本実装で利用した．CTmpLib は抽象化されたデータ構造の指定により複数の実装方法を選択できる．シーケンス, セット, マップ, マルチセット, マルチマップに対応しており, それぞれの構造に対し複数の実装方法が選択できる．また, すべてのデータ構造に対する操作は単一なインターフェイスにより利用できる．本実装では, リストとテーブルの作成に CTmpLib を利用した．

6.3 データ構造

本節では 5節で述べた各ノードの保持するデータ構造の実装方法を解説する．

6.3.1 Source Node

図 6.2に Source Node の保持するデータ構造とそれぞれの関係を述べる． Source Node は全ノードを含むツリー構造の保持を行う **disttree** 型の構造体と, 自身のノード情報を保持する **nodeinfo** 型の構造体をもつ． **disttree** 内には, 自身のツリーノード情報の保持を行う **disttree_node** 型の **dt_source**, Leave Node を保持する **distnodeseq** 型のリスト, 全ノード

のツリーノード情報の検索と保持を行うマップである **distnodemap** 型の **dtmap_nodes** が保持される. **distnodemap** の実装方法としてスキップリストを採用した.

6.3.2 Recipient Node

Recipient Node の保持するデータ構造とその関係を, 図 6.3に示す. Recipient Node は自身のノード情報, Parent Node のノード情報, Old Parent Node のノード情報, Source Node のノード情報を **nodeinfo** 型で保持している. さらに **nodeinfo** 型のリスト構造である, **nodeseq** 型で管理 Child Node のリストと PPL を保持している.

6.4 LOLCAST プロトコル処理モジュール

本節では LOLCAST プロトコル処理モジュールの実装において特徴的な処理を行う箇所の解説を行う.

6.4.1 ツリー操作関数群

ツリー構造関数群は Source Node の保持するツリー構造に対しノードを追加, 削除, 入れ替えを行う. 本節ではツリー構造関数群の中から重要な役割を持つ関数を取り上げ, それぞれについて解説を行う.

- **disttree_addnode()**
  disttree_addnode() はツリー構造にノードを追加する際に利用する. 引数として新規参加ノードと Parent Node の **disttree_node** を取る. Parent Node の検索を **dtmap_nodes** を用いて行い, その **dtmap_children** に新規参加ノードを追加する.

- **disttree_removenode()**
  disttree_removenode() はツリー構造からのノードの削除に利用する. 引数として削除を行うノードとその Parent Node の **disttree_node** を取る. **disttree_addnode()** と同様に, **dtmap_nodes** を用いて Parent Node の検索を行う. 見つかった Parent Node の **disttree_node** の **dtmap_children** から離脱ノードのエントリを消去する. 最後に離脱ノードの領域を解放する.

- **disttree_setdepth()**
  disttree_changeparent() はノードの深さを設定する関数である. 引数として Parent Node と Child Node の **disttree_node** を取り, Child Node 以下に存在するすべての子ノードの深さを更新を再起的に行う.

- **disttree_changeparent()**
  disttree_changeparent() は Parent Node の繋ぎ変えを行い, **Notify ParentChanged** のメッセージを受け取った際に呼び出される. 引数として, Child Node の **disttree_node**, 新たな Parent Node の **disttree_node** をとり, ノードの繋ぎ変えを行う. まず, Child Node の検索を **dtmap_nodes** で行い, Old Parent Node の **dtmap_children** から Child

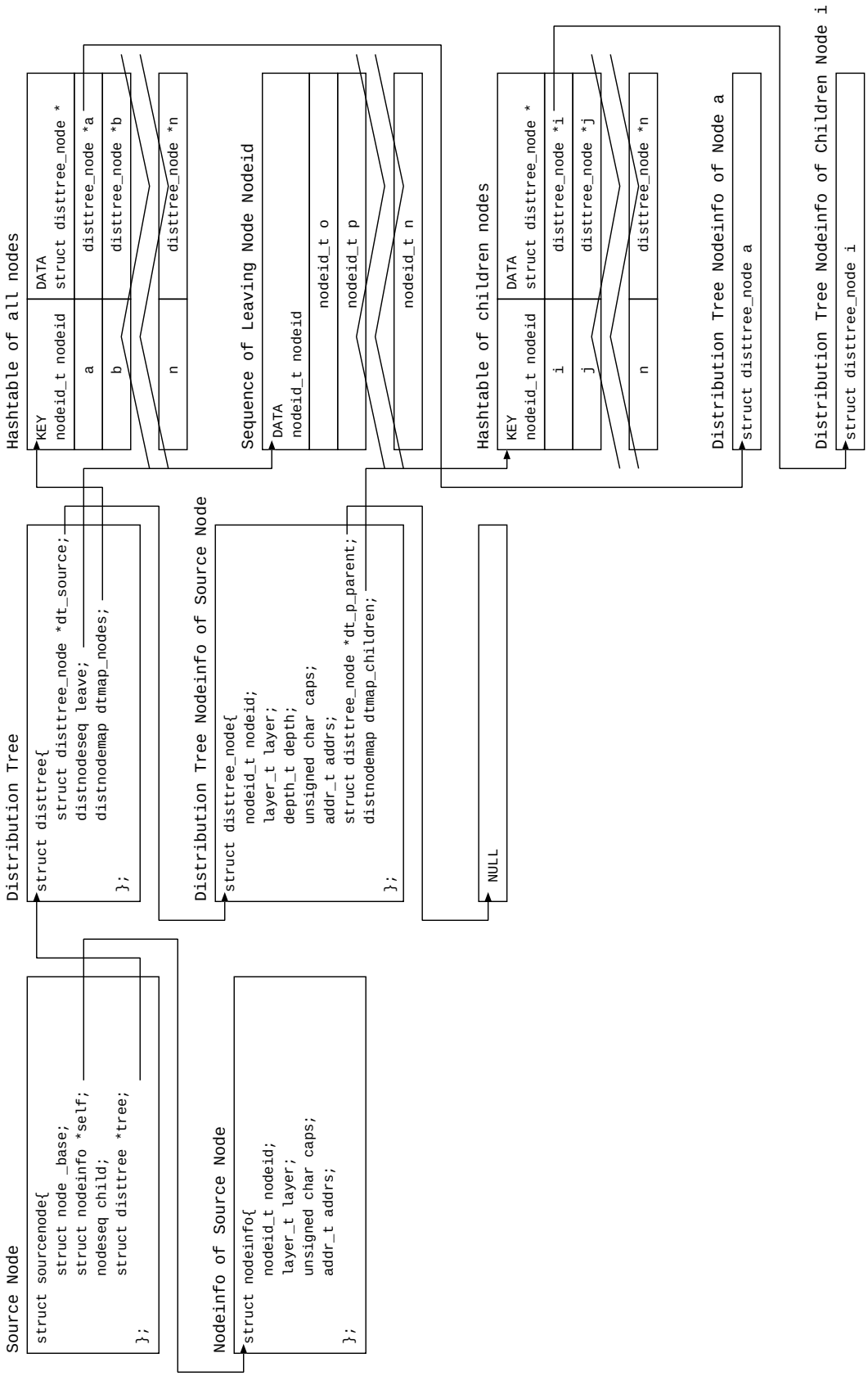

図 6.2: Source Node の保持するデータ構造

図 6.3: Recipient Node の保持するデータ構造

```
struct disttree *disttree_new(void);
void disttree_delete(struct disttree *disttree_p);
struct disttree_node *disttree_node_new(void);
void disttree_node_delete(struct disttree_node *disttree_node_p);
struct disttree_node *disttree_node_find(struct disttree *disttree_p, nodeid_t nodeid);
void disttree_switchnode(struct disttree *tp, struct disttree_node *parent,
                         struct disttree_node *child);
void disttree_addnode(struct disttree *disttree_p, struct disttree_node *parent,
                      struct disttree_node *child);
void disttree_removenode(struct disttree *disttree_p, struct disttree_node *child);
void disttree_changeparent(struct disttree *disttree_p, struct disttree_node *newparent,
                           struct disttree_node *child);
void disttree_setdepth(struct disttree_node *parent, struct disttree_node *child);
distnodeseq disttree_getparentlist(struct disttree *tp, layer_t layer);
```

図 6.4: ツリー構造関数一覧

```
void disttree_changeparent(struct disttree *tp, struct disttree_node *parent,
                           struct disttree_node *child){
    distnodemap_iterator itr;

    /* check for layer, parent > child */
    if(parent->layer < child->layer){
        abort();
    }

    /* remove child data from current parent child map */
    itr = iterator_first(child->dt_p_parent->dtmap_children);
    itr = iterator_find(child->dt_p_parent->dtmap_children, child->nodeid);
    if(!iterator_is_end(itr)){
        dprintf(("\t---node %d found\n", child->nodeid));
        itr = iterator_erase(itr);
    } else {
        dprintf(("\t---node %d not found\n", child->nodeid));
        abort();
    }

    /* add child data to new parent child map */
    iterator_insert_pair(parent->dtmap_children, child->nodeid, child);
    /* change parent entry of child */
    child->dt_p_parent = parent;
    /* set depth */
    disttree_setdepth(parent, child);
}
```

図 6.5: disttree_changeparent()

Node のエントリを削除する．新たな Parent Node の **dtmap_children** に Child Node のエントリを追加する．Child Node の **p_parent** に Parent Node の Nodeinfo を入れる．最後に **disttree_setdepth()** が呼ばれ, 深さの更新が行われる．

- **disttree_getparentlist()**
  disttree_getparentlist() は PPL Extraction に用いられる関数である．5.3節で述べたように, Temporary PPL Extraction と PPL Extraction の二つの処理により PPL が生成される．図 6.6に **disttree_getparentlist()** 関数を示す．

  ppl_layercheck() により, レイヤ比較が行われ, Temporary PPL が生成される．次に,

```
distnodeseq disttree_getparentlist(struct disttree *tp, layer_t layer){
    distnodeseq tmpppl;
    distnodeseq ppl;
    distnodemap_iterator itr;
    int pplnum;

    /* 略 */

    tmpppl = distnodeseq_new(PPL_MAX, PPL_MAX, NULL, NULL);
    ppl = distnodeseq_new(PPL_MAX, PPL_MAX, NULL, NULL);
    itr = iterator_first(tp->dtmap_nodes);

    while(1){
        /* extract temporary ppl */
        itr = ppl_layercheck(tp->dtmap_nodes, tmpppl, layer, itr);

        if(iterator_count(tmpppl) > 0){
            /* remove leave node */
            pplnum = ppl_removedeny(tp->leave, tmpppl);
            /* check number of childs  */
            pplnum = ppl_childcheck(tp->dtmap_nodes, tmpppl);
            if(iterator_count(tmpppl) > 0){
                ppl = ppl_extractbydepth(tp->dtmap_nodes, tmpppl);
                /* sort ppl with depth */
                dprintf(("ppl with out switch\n"));
                break;
            }
        }
    /* 略 */
    }
    return ppl;
}
```

図 6.6: disttree_getparentlist()

ppl_removedeny() により Leave Node が削除され, **ppl_childcheck()** により, Child Node 数の空きがチェックされる. 最後に生成された PPL を **ppl_extractbydepth()** により, 深さが浅いノードから順に並べかえられ, PPL の生成が完了する.

6.4.2 メッセージ処理関数群

メッセージの処理を行う関数は, メッセージ毎に受信, 送信を行う recv 関数, send 関数が用意される. 到着したメッセージはモジュール結合部により, **Message Type**, **Message Subtype** フィールドを利用して各 recv 関数に渡される. recv 関数はメッセージの種類に従った処理をデータ構造に対し行い, 必要であれば send 関数を呼び出す. 呼び出された send 関数はメッセージの作成を行い, **Message Data** にデータを格納する.

メッセージの送受信を行うノードが Source Node か Recipient Node で処理の変わるメッセージがある. C 言語では多様性を仕様で表現できないため, 上述した **node** 型の構造体でノードの保持するデータ構造を引き渡し, **nodetype** を利用し処理を分けた. 図 6.7にメッセージ処理で用いられる関数の一覧を示す.

```
void send_rendezvous_request(struct node *self);
void recv_rendezvous_request(struct node *self, struct msg *msgdata);
void send_rendezvous_accept(struct node *self, struct nodeinfo *peer);
void recv_rendezvous_accept(struct node *self, struct msg *msgdata);
void send_rendezvous_reject(struct node *self, struct nodeinfo *peer);
void recv_rendezvous_reject(struct node *self, struct msg *msgdata);
void send_ppl_request(struct node *self);
void recv_ppl_request(struct node *self, struct msg *msgdata);
void send_ppl_data(struct node *self, struct nodeinfo *peer);
void recv_ppl_data(struct node *self, struct msg *msgdata);
void send_join_request(struct node *self);
void recv_join_request(struct node *self, struct msg *msgdata);
void send_join_accept(struct node *self, struct nodeinfo *peer);
void recv_join_accept(struct node *self, struct msg *msgdata);
void send_join_reject(struct node *self, struct nodeinfo *peer);
void recv_join_reject(struct node *self, struct msg *msgdata);
void send_data_startrequest(struct node *self);
void recv_data_startrequest(struct node *self, struct msg *msgdata);
void send_data_stoprequest(struct node *self);
void recv_data_stoprequest(struct node *self, struct msg *msgdata);
void send_leave_request(struct node *self);
void recv_leave_request(struct node *self, struct msg *msgdata);
void send_leave_complete(struct node *self);
void recv_leave_complete(struct node *self, struct msg *msgdata);
void send_prune_request(struct node *self);
void recv_prune_request(struct node *self, struct msg *msgdata);
void send_prune_complete(struct node *self, struct nodeinfo *peer);
void recv_prune_complete(struct node *self, struct msg *msgdata);
void send_notify_joincomplete(struct node *self);
void recv_notify_joincomplete(struct node *self, struct msg *msgdata);
void send_notify_parentchanged(struct node *self);
void recv_notify_parentchanged(struct node *self, struct msg *msgdata);
void send_notify_leaveprogress(struct node *self);
void recv_notify_leaveprogress(struct node *self, struct msg *msgdata);
void send_notify_leavecomplete(struct node *self);
void recv_notify_leavecomplete(struct node *self, struct msg *msgdata);
```

図 6.7: メッセージ処理関数一覧

6.4.3 モジュール結合部

モジュール結合部は **lc_recv()**, **lc_send()** の 2 つの関数から構成される. **lc_recv()** により, 受信メッセージの振分けが行われプロトコル処理モジュールにメッセージが引き渡される. **lc_send()** はプロトコル処理モジュールより受け取ったメッセージをシミュレータモジュールやアプリケーションモジュールに引き渡す.

6.5 シミュレータモジュール

シミュレータモジュールでは, ノードが新たにマルチキャストツリーに参加する際に Join Procedure にかかる時間と, その際に発生する PPL Extraction にかかる時間の計測を行う機能を実装した. シミュレータモジュールは図 6.8に示す関数群から構成される.

simu_init_sourcenode(), **simu_init_recipientnode()** により, Source Node, Recipient Node のデータ構造が初期化される. **simu_recv()** は **lc_send()** に引き渡されたメッセージの受信に利用される. **evaluate_joinprocedure()** により, ノードが Join Procedure により参加するまでの処理にかかる時間を計測する. **evaluate_joinprocedure()** を複数回実行することに

```
nodeid_t simu_nodeid_alloc(void);
struct sourcenode *simu_init_sourcenode(void);
struct recipientnode *simu_init_recipientnode(struct sourcenode *snp, layer_t layer);
void simu_recv(struct msg *msgdata, struct nodeinfo *peer);
struct timeval evaluate_joinprocedure(struct sourcenode *snp, struct msg *msgdata, layer_t layer);
```

図 6.8: シミュレータ関数群

```
nodeid: 1 (0 depth, 10 layer, 3 children)
         nodeid: 2 (1 depth, 8 layer, 3 children)
                 nodeid: 5 (2 depth, 8 layer, 3 children)
                         nodeid: 13 (3 depth, 3 layer, 0 children)
                         nodeid: 14 (3 depth, 6 layer, 0 children)
                         nodeid: 16 (3 depth, 3 layer, 0 children)
                 nodeid: 6 (2 depth, 7 layer, 0 children)
                 nodeid: 7 (2 depth, 7 layer, 0 children)
         nodeid: 10 (1 depth, 9 layer, 1 children)
                 nodeid: 4 (2 depth, 1 layer, 0 children)
         nodeid: 11 (1 depth, 10 layer, 2 children)
                 nodeid: 3 (2 depth, 6 layer, 3 children)
                         nodeid: 8 (2 depth, 2 layer, 0 children)
                         nodeid: 9 (2 depth, 5 layer, 0 children)
                         nodeid: 12 (3 depth, 2 layer, 0 children)
                 nodeid: 15 (2 depth, 10 layer, 0 children)
```

図 6.9: マルチキャストツリー出力の例

より, ツリーに参加するノードの増加による Join Procedure にかかる時間の増加を測定できる. **evaluate_joinprocedure()** は引数として新規ノードの要求レイヤ数を取り, 0 を指定することによりランダムなレイヤ数を指定できる. 乱数の生成には random 関数を用いた. また, 時間の計測には gettimeofday 関数を用い, 返り値として経過時間を返す. **evaluate_joinprocedure()** をランダムなレイヤ数を指定し実行した際に生成されるマルチキャストツリーの一例を図 6.9に示す.

第7章　評価

7.1　評価の目的

本節では, LOLCASTが 1.2節で述べた目標環境実現のための必要要件を満たしているか考察を行う. 必要要件を満たすことにより,「発信者が一般利用者にとって現実的な資源で, 多くの受信者に対し, 個々の受信者が要求する品質でサービスを提供できる環境」が実現される. また, LOLCASTにおいて提案した階層的な構造を持つデータの配信に伴うオーバーヘッドの計測を行い, プロトコルの有効性を述べる.

7.2　定性評価

定性評価では, 1.2節で述べた目標環境実現のための必要要件をLOLCASTが満たしているか考察を行う. 図 7.1に示すように, LOLCASTは既存オーバーレイマルチキャスト技術と比べ, 単一の配信網の管理により, 複数の受信者の資源環境に適した配信が行える. 配信者の帯域的資源の負担も, 品質毎のデータを送ること無いため, 少なく, 多様な受信者に対応できる.

7.3　定量評価

定量評価として, ノードの増加に伴うSource Nodeにおける処理時間の増加を計測した. この際, 新規ノードが均質な値を要求レイヤ数として指定した場合と要求レイヤ数をランダムとした場合の二通りの計測を行った. 均質な値を要求した際はPPL Extractionにおいて, 全ノードの含まれるマップに対し線形に検索を行う. よって上位のエントリから順に管理子ノード数の空きが埋まるため常に最悪値となる. このデータを新規ノードの参加における最悪値として扱いランダムな値を要求した際と比較を行う. 計測対象の処理としてJoin Procedureとこの処理の一部であるPPL Extractionを用いた.

表 7.1: 定性評価

評価項目/配信技術	LOLCAST	既存オーバーレイマルチキャスト技術
提供可能な品質	複数の品質	単一な品質
配信網の維持コスト	単一なマルチキャストツリー	品質毎のマルチキャストツリー
予期しない離脱への対処	データパスの冗長化, 輻輳回避手法	配信パスの冗長化 (HostCast)

前述した Join Procedure と PPL Extraction の処理時間を計測するシミュレータは 6章で実装を行った LOLCAST プロトコル処理モジュールとシミュレータモジュールを用いた.

次にシミュレータの実行方法と出力されるデータについて述べる. 実装を行ったシミュレータは引数としてマルチキャストツリーに参加する総ノード数を指定する. 実行すると, 現在の参加ノード数, 参加ノードが増加した際の Join Procedure にかかる時間, 参加ノードが増加した際の PPL Extraction にかかる時間を出力する. シミュレータの実行例を図 7.1に示す. この際, **-r** オプションをつけることにより, ランダムなレイヤ数を要求するノードの参加時間を出力できる. **-r** オプションを指定しない場合は同一のレイヤ数を要求するノードの参加時間を出力する. それぞれの時間の単位はマイクロ秒で表される.

```
% ./lolcast -r 10
#nodenum joinprocedure pplextraction
1 93 20
2 63 12
3 29 12
4 23 9
5 48 9
6 25 10
7 26 9
8 24 9
9 48 10
10 35 18
```

図 7.1: シミュレータ実行例

7.3.1 実験環境

以下の環境においてシミュレータを実行した. この際, プロセススイッチによるデータのブレを防ぐため, 他プロセスの影響を受ける可能性の低い Single User Mode で計測を行った.

表 7.2: 評価環境

プロセッサ	Pentium III 1Ghz
メモリ	512MB
OS	NetBSD 1.6.2
コンパイラ	gcc 2.95.3 (20010315)
コンパイラオプション	-O3

7.3.2 実験結果

固定レイヤ数による測定

図 7.2に計 10,000 ノードをマルチキャストツリーに 1 ノードづつ追加を行った際の Join Procedure に伴う処理時間を示す. 図 7.3に計 10,000 ノードをマルチキャストツリーに 1 ノードづつ追加を行った際の PPL Extraction に伴う処理時間を示す. 共に, 新規参加ノードの要求レイ

ヤ数は固定値を取る. 図 7.3に 1000 ノード毎の Join Procedure, PPL Extraction の平均処理時間を示す.

表 7.3: 同一のレイヤ数の値を指定した際の処理時間平均

ノード数	Join Procedure (usec)	PPL Extraction (usec)
1000	75.211	39.284
2000	174.944	134.929
3000	355.41	311.035
4000	527.597	480.662
5000	692.782	643.399
6000	860.21	809.592
7000	1032.33	980.978
8000	1234.35	1180.21
9000	1468.23	1411.35
10000	1714.89	1654.12

ランダムレイヤ数

図 7.5に計 10,000 ノードをマルチキャストツリーに 1 ノードづつ追加を行った際の Join Procedure に伴う処理時間を示す. 図 7.6に計 10,000 ノードをマルチキャストツリーに 1 ノードづつ追加を行った際の PPL Extraction に伴う処理時間を示す. 共に, 新規参加ノードの要求レイヤ数はランダムに決定される. 図 7.4に 1000 ノード毎の Join Procedure, PPL Extraction の平均処理時間を示す.

表 7.4: ランダムにレイヤ数を指定した際の処理時間平均

ノード数	Join Procedure (usec)	PPL Extraction (usec)
1000	73.134	37.76
2000	146.678	108.842
3000	304.653	262.234
4000	475.288	430.582
5000	618.879	573.492
6000	770.426	722.807
7000	920.903	870.902
8000	1078.18	1026.74
9000	1254.48	1200.84
10000	1443.25	1386.22

7.4 評価結果

LOLCAST の想定規模である数千人を対象とした配信を考える．ランダムなレイヤ数を要求する 5,000 ノードが参加を行った際には, 図 7.4に示す様に 0.6msec 程度で Join Procedure が完了している．また, 10,000 ノードが参加を行っている場合でも 2msec 程度で Join Procedure が完了している点から, ランダムなレイヤ数を用いても大きなオーバーヘッドは発生しないと言える．また PPL の処理が Join Procedure において非常に大きなウェイトを占めていることがわかる．以上の結果により, LOLCAST を用いることにより, 既存技術と比較しより効率的な配信手法を提供できた．

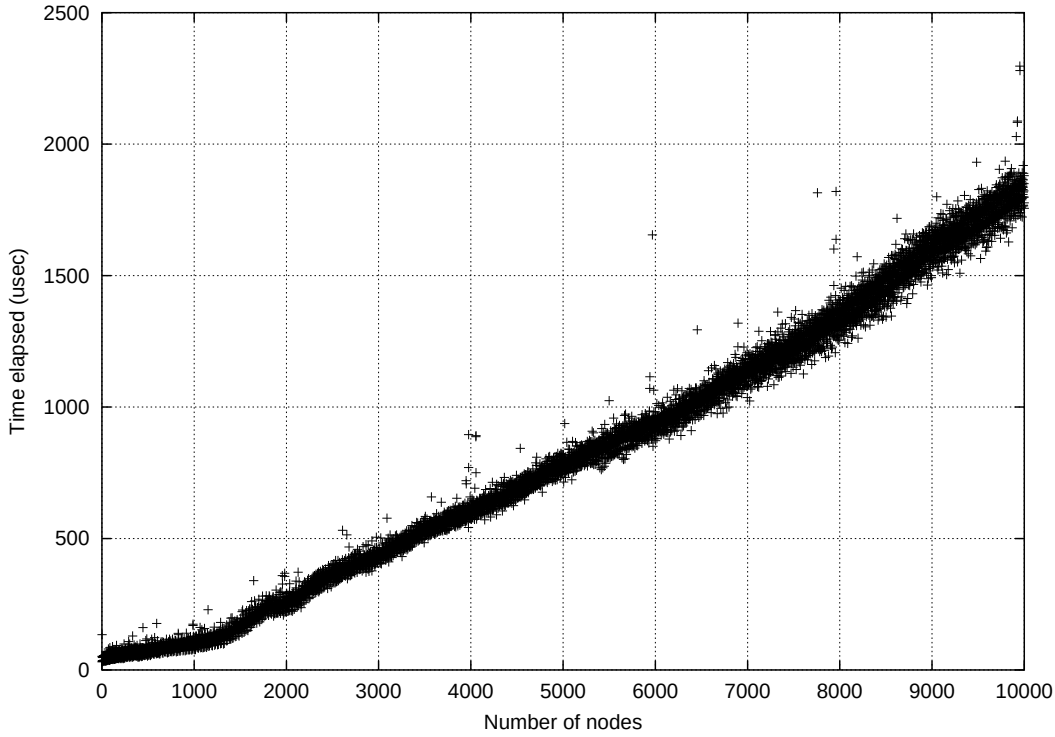

図 7.2: Join Procedure 所要時間 (10,000 nodes / Fixed layer / Skip List)

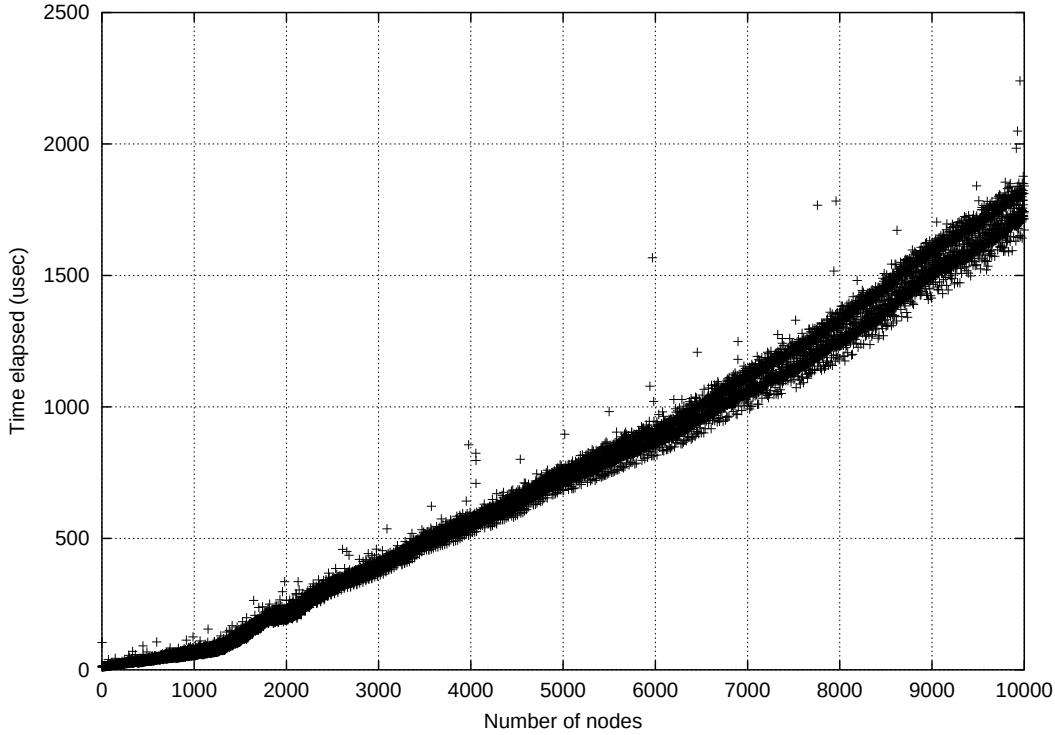

図 7.3: PPL Extraction 所要時間 (10,000 nodes / Fixed layer / Skip List)

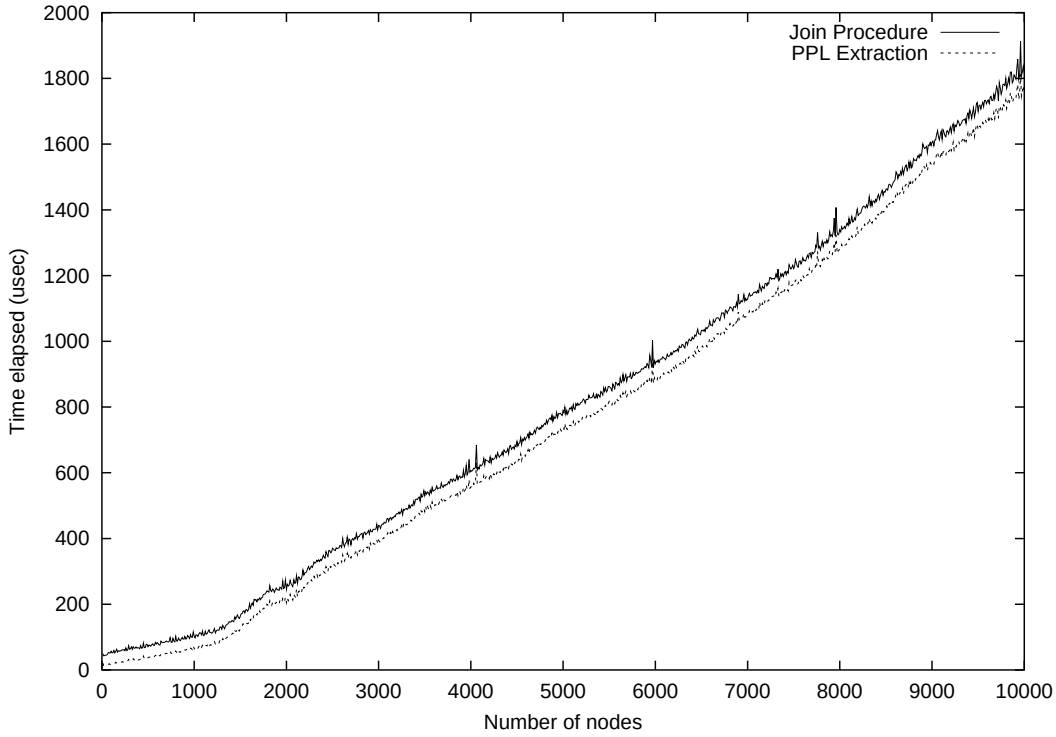

図 7.4: Join Procedure における PPL Extraction 所要時間 (10,000 nodes / Fixed layer / Skip List)

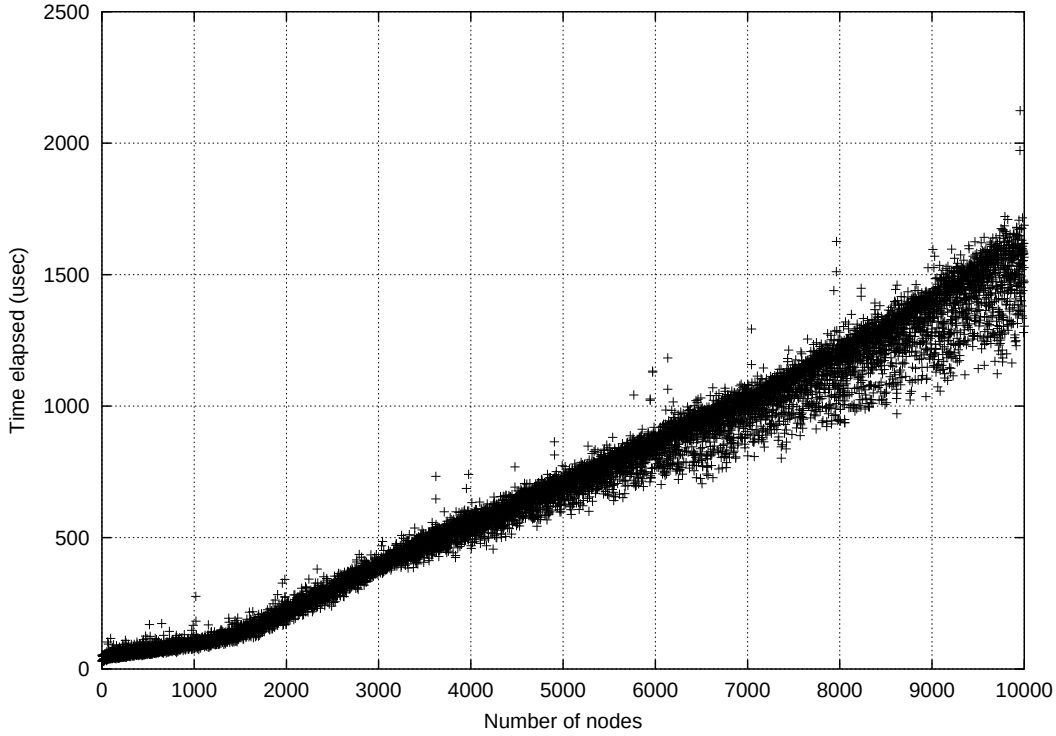

図 7.5: Join Procedure 処理時間 (10,000 nodes / Random layer / Skip List)

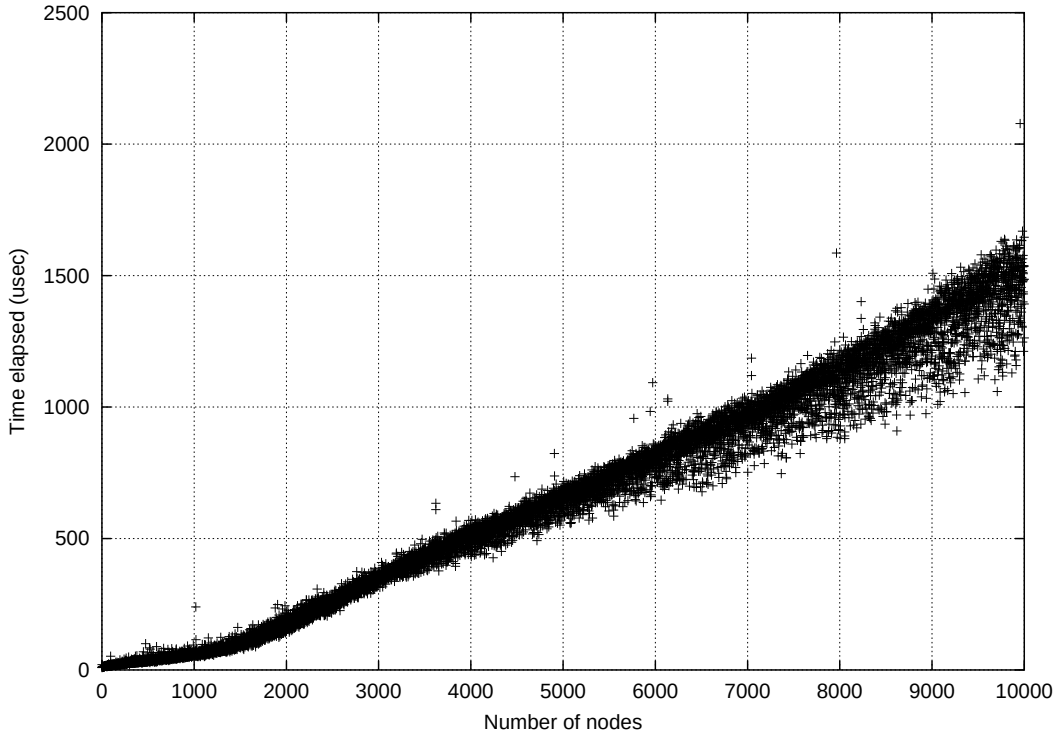

図 7.6: PPL Extraction 処理時間 (10,000 nodes / Random layer / Skip List)

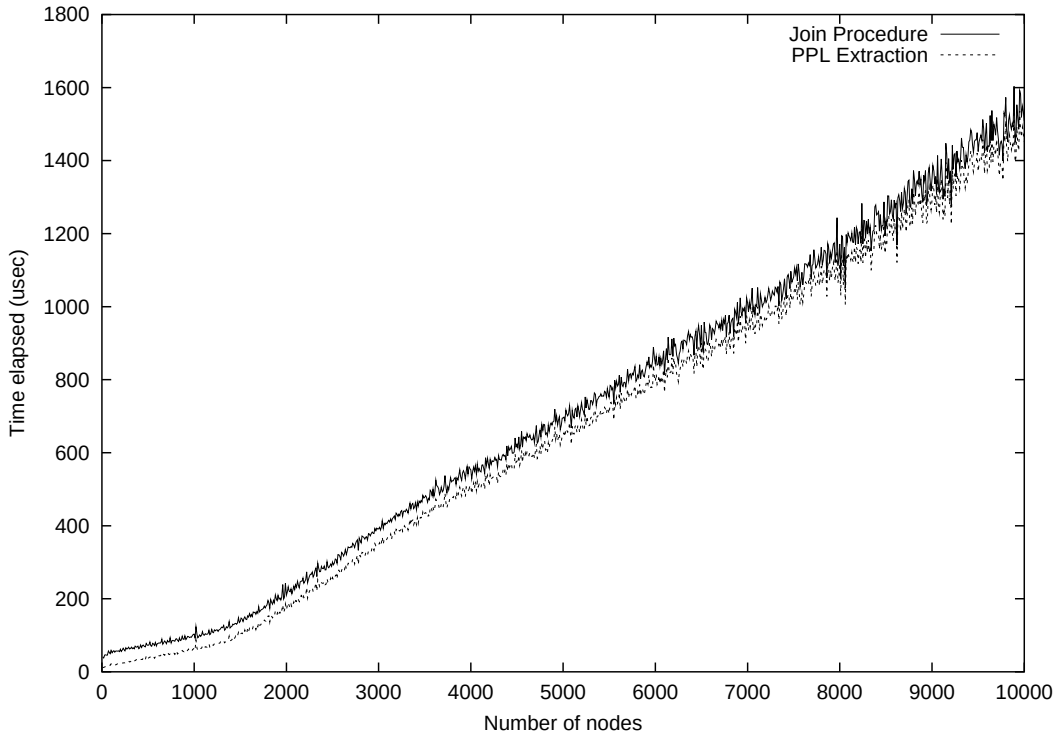

図 7.7: Join Procedure における PPL Extraction 所要時間 (10,000 nodes / Random layer / Skip List)

第8章　結論

8.1　まとめ

本研究では, 階層構造を持つデータの配信を行う新たなオーバーレイ・マルチキャストプロトコルであるLOLCASTを提案した. LOLCASTにより発信者は単一の配信網を構築・管理することにより受信者の資源環境に適応した配信が行える. 受信者は資源環境によらず配信網への参加が可能となり, 環境に合わせ品質の選択が可能となった. 評価の結果, LOLCASTは既存OLM技術と比較し, 発信者にかかる配信のためのコストを削減することができた.

8.2　今後の課題

8.2.1　PPL Extractionの性能改善

7節で述べたように, Join Procedureにおける処理時間の大部分はPPL Extractionの処理に裂かれており, この処理を改善することにより全体のパフォーマンスを向上できる. 現状ではTemporary PPL Extraction時にレイヤ数を満たすノードを線形にサーチを行っているため, 負荷が大きい. ノードの検索を検索木に変更することにより計算量で比較すると $O(n)$ から $O(logn)$ に改善するため, 処理時間自体の向上も期待できる.

8.2.2　フレームワークの提供

現状の設計では, 個々のデータフォーマット毎にデータの加工、表示を行うアプリケーションモジュールが必要となってしまう. アプリケーションモジュール内の機能を分割し, データフォーマット依存の処理を行うモジュールをアプリケーションモジュールに組み込める構造を持たせる. データフォーマットモジュールでは, そのフォーマットにおけるレイヤの定義とデータの加工部分の実装を行う. これにより, アプリケーションモジュールはデータフォーマット依存の処理を隠蔽することが可能となり, レイヤ数という統一されたインターフェイスで複数の異なるフォーマットに対応できる.

8.2.3　アプリケーションによる評価

今回は, シミュレータの実装とそれを用いた評価を行った. オーバーレイ・マルチキャスト技術は対象アプリケーションに沿った設計を行うことが重要とされているため, アプリケーションによる評価も必要となる. 本研究の目標で挙げた, リアルタイム性の高い映像配信手法としてLOLCASTが利用可能であるかを映像配信アプリケーションを利用し評価する.

謝辞

本研究を進めるにあたり，御指導を頂きました，慶應義塾大学環境情報学部教授徳田英幸博士，村井純博士，同学部助教授の楠本博之博士，中村修博士，同学部専任講師の南政樹氏，重近範行博士に感謝致します.

研究において絶えず御指導と御助言を頂きました慶応義塾大学政策メディア研究科 今泉英明氏, 石原知洋氏に深く感謝します. 最後の最後で心強いサポートをくれた後輩, 空閑洋平氏に感謝します. また同じ境遇に立ち, 良き話し相手, 相談相手となってくれた, 慶応義塾大学政策メディア研究科 入野仁志氏, 慶応義塾大学環境情報学部, 谷隆三郎氏, 千代佑氏, 松園和久氏, 同大学総合政策学部 吉田雅史氏に感謝します.

最後に論文執筆中, 良質の音楽を提供してくれた以下のアーティスト達に感謝します. Tidy Trax, Untidy Trax, ACO, Clammbon, Honda Lady, YMCK, Rainstick Orchestra, Laurent Garnier, X-Dream, Hixxy, Kurli, Four Tet, London Elektricity, Massive Attack, Parasence, Infected Mushroom, Radio Head, Bjork, Royksopp, David Lewston, Scion, Shpongle, Suitei Shyoujyo, DJ Q'Hey, DJ Shinkawa, DJ Hoekzema, DJ Uraken, Every Little Thing, Mr.Children, DJ Die, DJ Marix, Lalgudi G. Jayaraman, Ustad Amjadali Khan, Untidy DJ's, Tidy Boys, Perpetual Motion, 1200mics, Sorma, Kindzadza, Psyside, DJ Now!, Aoa, Kojima Mayumi, Eminem, Scott Brown, Juno Reactor, UA, Pinback, Zobra, Raja Ram, etc.

参考文献

[1] HITACHI. 360 度どこからでも見ることができる立体映像ディスプレイ技術. *http://www.hitachi.co.jp/New/cnews/040224a.html*, 2003.

[2] Stephen E. Deering and David R. Cheriton. Multicast routing in datagram internetworks and extended lans. *ACM Trans. Comput. Syst.*, 8(2):85–110, 1990.

[3] Christophe Diot and Brian Neil Levine and Bryan Lyles and Hassan Kassem and Doug Balensiefen. Deployment issues for the ip multicast service and architecture. In *IEEE Network Vol.14, num 1*, pages 78–88, 2000.

[4] P. Francis. Yoid : Extending the internet multicast architecture. In *Technical report, AT&T Center for Internet Research at ICSI (ACIRI)*, April 2000.

[5] S. Banerjee and B. Bhattacharjee. A comparative study of application layer multicast protocols. 2002.

[6] Yang hua Chu, Sanjay G. Rao, and Hui Zhang. A case for end system multicast (keynote address). In *Proceedings of the 2000 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 1–12. ACM Press, 2000.

[7] Yang Chu, Sanjay Rao, Srinivasan Seshan, and Hui Zhang. Enabling conferencing applications on the internet using an overlay muilticast architecture. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 55–67. ACM Press, 2001.

[8] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Schenker. A scalable content-addressable network. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 161–172. ACM Press, 2001.

[9] Suman Banerjee, Bobby Bhattacharjee, and Christopher Kommareddy. Scalable application layer multicast. In *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 205–217. ACM Press, 2002.

[10] B. Zhang, S. Jamin, and L. Zhang. Host multicast: A framework for delivering multicast to end users. In *IEEE Infocom*, 2002.

[11] Zhi Li and Prasant Mohapatra. Hostcast: A new overlay multicasting protocol. In *IEEE International Communications Conference (ICC)*, 2003.

[12] Anne-Marie Kermarrec Miguel Castro, Michal B. Jones and Antony Rowstron. An evaluation of scalable application-level multicast built using peer-to-peer overlay. In *Infocom 2003, San Francisco, CA*, 2003.

[13] 村田正幸 CaoLe Thanh Man, 長谷川 剛. サービスオーバーレイネットワークのためのインラインネットワーク計測に関する一検討. In **電子情報通信学会技術研究報告** *(IN03-176)*, pages 53–58, 2003.

[14] Dimitris Pendarakis, Sherlia Shi, Dinesh Verma, and Marcel Waldvogel. ALMI: An application level multicast infrastructure. In *Proceedings of the 3rd USNIX Symposium on Internet Technologies and Systems (USITS '01)*, pages 49–60, San Francisco, CA, USA, March 2001.

[15] Yatin Chawathe. Scattercast: an adaptable broadcast distribution framework. *Multimedia Syst.*, 9(1):104–118, 2003.

[16] John Jannotti, David K. Gifford, M. Frans Kaashoek, and James W. O'Toole Jr. Overcast: Reliable multicasting with an overlay network. In *5th Symposium on Operating System Design and Implementation (OSDI)*, December 2000.

[17] Yan Zhu, Min-You Wu, and Wei Shu. Comparison study and evaluation of overlay multicast networks. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference*, July 2003.

[18] Yogen K. Dalal and Robert M. Metcalfe. Reverse path forwarding of broadcast packets. *Commun. ACM*, 21(12):1040–1048, 1978.

[19] Bb@nifty: Bb clip. *http://bb.nifty.com/clip/movie/*.

[20] MPEG-2 Generic coding of moving pictures and associated audio information. http://www.chiariglione.org/mpeg/standards/mpeg-2/mpeg-2.htm.

[21] David Taubman. High performance scalable image compression with ebcot. In *IEEE Transactions on Image Processing Vol.9 No.7*, pages 1158–1170, 2000.

[22] David Taubman, Erik Ordentlich, Marcelo Weinberger, Gadiel Seroussi, Ikuro Ueno, and Fumitaka Ono. Embedded block coding in jpeg2000. In *IEEE International Conference on Image Processing (ICIP)*, pages 33–36, 2000.

[23] HD DIGITAL VCR CONFERENCE. Specifications of Consumer-Use Digital VCRs PART1 General Specifications of Consumer-Use Digital VCRs. *Specifications of Consumer-Use Digital VCRs using 6.3mm magnetic tape*, pages 1–61, Dec 1994.

[24] Sally Floyd, Mark Handley, Jitendra Padhye, and Joerg Widmer. Equation-based congestion control for unicast applications. In *Proceedings of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 43–56. ACM Press, 2000.