

10g-wg 2004年度活動報告

久松剛 (ringo@sfc.wide.ad.jp)

千代佑 (tsk@sfc.wide.ad.jp)

松園和久 (kazuhisa@sfc.wide.ad.jp)

2005年2月7日

Copyright (C) WIDE Project (2004). All Rights Reserved.

第1章 10G WGの概要

10G WGは超広帯域ネットワークにおけるキラーアプリケーションの構築を目指し、2002年12月に発足した。本章ではまず超広帯域ネットワークにおけるキラーアプリケーションの方向性を解説し、本年度の活動について述べる。

1.1 超広帯域ネットワーク環境の有効利用

IEEE802.3ae[1]、OC-192、OC-768といった超広帯域ネットワークの登場に伴い、その有効利用方法が求められている。

10Gbps、40Gbpsというネットワークにおいて二つの観点からキラーアプリケーションのアプローチを行う。一つは帯域を潤沢に使い、スケーラビリティの高いサービスを低コストでユーザに提供するタイプのものである。VoD(Video on Demand)などが例として挙げられる。もう一つはネットワークを潤沢に消費しない代わりに高品質なサービスを提供するタイプのものである。低遅延、低ジッタな映像配信システムなどが好例である。

これらのアプリケーションを実現するにあたり、ネットワークのパフォーマンスを活かすためにPCの設定、ネットワークの構築方法などが重要となる。特に超広帯域ネットワークでは転送機器のパフォーマンスがネットワークのパフォーマンスを下回ることが多い。具体的にはハードディスク、メモリ、バスの選定、OSの設定などが挙げられる。限られた資源の中で最大のパフォーマンスを発揮する必要がある。本WGではキラーアプリケーションの構築に加え、このような最適な広帯域ネットワーク環境の模索を行う。

1.2 前年度の活動

本WGでは2003年度の活動として主に以下の6つの項目に対して取組みを行った。

- 1Gbps 帯域消費テスト
10Gbpsという超広帯域ネットワークに対する事前の取組みとして、1Gbpsのネットワークを有用なトラフィックで消費するテストを行った。

- ネットワークパフォーマンスの計測
カーネルオプションの変更によるネットワークパフォーマンステストを行った。
- 遠隔TV 試聴システムの開発
超広帯域ネットワークの利用方法として、1本の大きな帯域を消費する方法と、複数本の少量のストリームを複数のユーザーが利用する形式が考えられる。後者の有効なアプリケーション例として、遠隔地に設置したテレビチューナーを遠隔制御・試聴するシステムの開発を行った。
- オーケストラ演奏機構の開発
超高速、低遅延ネットワークにより結ばれた分散環境におけるマルチメディアコンテンツ共同製作環境例としてオーケストラ演奏機構の構築を行った。
- 分散環境における効率的な高品質映像制御に関する研究
超広帯域ネットワークにおけるアプリケーションとして、SDDVを用いたTrue VoDシステムの構築を行った。
- 適応型映像・音声配信機構の設計と構築
ネットワーク状態の変化がエンドノード間で取得可能な伝送特性として表れることに着目した服装制御手法を提案、構築を行った。

1.3 本年度の活動

超広帯域ネットワークの有効利用方法として1) 帯域を大量消費することによる有効利用、2) 大量消費はしないが品質の高い通信を保証、という2点が考えられる。

前年度の活動を踏まえ、本年度は主に以下の活動を行った。

- 計算機資源に応じた映像配信機構の研究
帯域を大量消費することによる有効利用を狙ったシステムとして開発に取り組んでいる。計算機資源に応じたストリームを受信ノードに転送することにより、効率よく多くの受信ノードが情報取得できることを目指した機構である。
- パケットフロー特性に着目した映像音声配信モデルの研究
品質の高い通信を保証するためのシステムとして開発に取り組んでいる。ネットワーク伝送遅延の揺らぎは長時間発生するものではなく、一時的なものが殆どであることに着目し、パケット伝送最適化モデルを提案し、開発を行っている。

第2章 計算機資源に応じた映像配信機構に関する研究

本研究では、計算機資源に基づく ALM(Application Layer Multicast) 映像配信機構の設計を行い、映像配信におけるエンドユーザの取得機会の拡大と配信の効率化を実現するシステムを構築している。

本研究により、配信元はネットワーク構成に依存する配信の効率化技術を利用せずに、広域へのコンテンツ配信を行うことが可能になった。また、配信元が要求するユーザ環境に満たないエンドユーザによるコンテンツの閲覧が可能になった。

2.1 目的：ユーザの情報取得機会の拡大

本研究は、コンテンツ配信におけるユーザのコンテンツ取得機会の拡大と、コンテンツの品質とユーザ要求のマッチングの実現を目的とする。

本研究の目的を達成するためには以下の項目が必要となる。

- オーバーレイネットワーク上での配信網の構成の実現

IP マルチキャストのようなネットワーク層における配信の効率化技術は、現状の IP ネットワークではインターネットサービスプロバイダ (ISP) などのサービスや運用ポリシーに依存する。このため、広域への配信を行う際にはユーザの利用している接続サービスを考慮しなくてはならない。この依存問題を解決するためには、接続サービスや運用ポリシーといった実ネットワーク構成に依存しない上位レイヤにおける配信手法が必要である。

- ユーザ要求に基づく配信元への誘導の実現

本研究で提案するコンテンツ配信システムはユーザが要求し、閲覧できる最大限の品質のコンテンツを各ユーザに提供しなければならない。このためには各ユーザの所有する計算機資源量を適切に把握する必要がある。本システムではコンテンツの品質とユーザ要求のマッチングを実現するために、エンドノードの計算機資源量の情報を統一的に管理する機能が必要である。

2.2 提案：情報取得機会の拡大

既存の技術はコンテンツ配信を行う際、ネットワーク環境への依存やコンテンツ発信元となるサーバへの負荷の増大、エンドノード環境の格差による配信モデルのボトルネック発生が起こる可能性がある。本研究ではこれらの問題点を解決する機構を組み込み、コンテンツ配信モデルの最適化を実現する。前述したように、コンテンツ配信の効率化を実現するには、計算機資源を過不足なく適切に再配分するための指標となるメトリックを設定する。

本研究の提案するコンテンツ配信システムのモデルは図 2.1 のようになる。

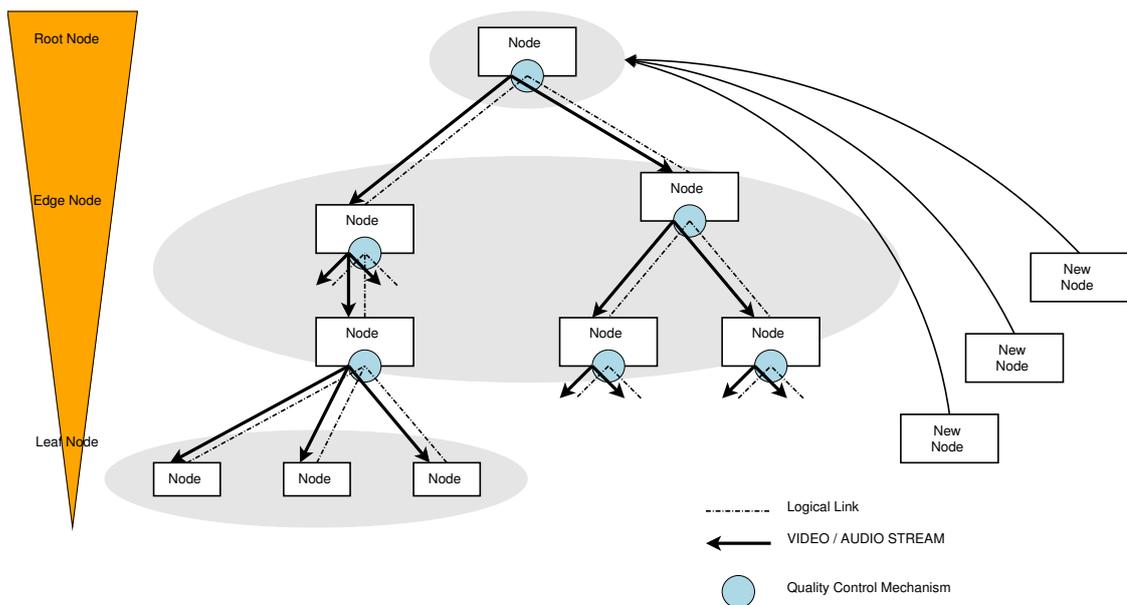


図 2.1: システムモデル

コンテンツ配信を行うノードは、ツリー構造の頂点であるルートノードとなる。コンテンツ閲覧者のノードで構成されるルート以外のノード内で、計算機資源が豊富なノードは論理リンクの上位に配置する。下位には上位ほど資源を持たないノードを配置し更に下位ノードへのコンテンツ配信を行うエッジノードとする。資源が少なく他のノードにコンテンツを提供することができない受信専用ノードは最下位のリーフノードとする。

ツリー構造の論理リンクの上位に配置されたノードはルートノードから品質の高いコンテンツを閲覧することが可能とする。その対価として下位ノードの要求に応じたコンテンツの再配信を行うことが要求される。一方、下位ノードは自身に適したコンテンツ配信を上位ノードに要求する、ただし上位ノードよりも品質の高い

コンテンツを閲覧することはできない。

2.3 実装：計算機資源に基づく ALM 映像配信機構の実装

映像配信機構としてインターネット上で DV データの送受信を行う。DVTS(Digital Video Transport System)[2] を対象に、計算機資源に基づく ALM 映像配信機構である RBMcast を設計し、その設計に基づいて実装を行っている。

本研究で実装した機構は以下の 4 つである。加えて、映像配信機構例として DVTS を用いた実装を行った。

- ノード情報取得モジュール
受信ノードの情報取得を行う。具体的な対象としては演算能力、データリンク帯域が挙げられる。今回の実装では Linux を用いたため、proc file system, MII(Media Independent Interface) からの取得を行った。
- ノード間リンクモジュール
ノード情報の送受信、生存確認、制御命令の伝達を行うことによりツリーの構築に必要な情報の収集、維持を行う。
- ノード情報管理モジュール
配信網に参加する全てのノード情報の管理を行う。また、受信ノードからの接続ノード問い合わせに対し、最適なノードの検索と映像配信機構の制御命令送信を行う。
- 映像配信機構制御モジュール
構築された ALM ツリーと、実際の映像配信機構との連携を行う。

2.4 まとめ・今後

本研究では計算機資源に基づく Application Layer Multicast(ALM) 映像配信機構の実装を行っている。本機構を用いることにより、コンテンツ配信におけるユーザのコンテンツ取得機会の拡大と、コンテンツの品質とユーザ要求のマッチングを実現することができる。

今後の発展としては、動的な配信網の再構成によるノード関係の最適化、メディア変換機構との連携が挙げられる。

第3章 パケットフロー特性に着目した映像音声配信モデルの研究

本研究では、ネットワークを用いたリアルタイム映像ストリーミングシステムに特化したパケット伝送最適化機構の構築に取り組んでいる。

本研究により、インターネットを介したリアルタイム映像転送システムの安定した送受信が品質を損なうことなく可能となる

3.1 目的: 計算機資源に応じた信頼性の高い高品質映像・音声配信システムの構築

本研究では、リアルタイム映像・音声配信システムを対象としており、受信側の計算機資源に応じた信頼性の高い高品質映像・音声配信システムの構築を目的とする。

送信側は受信側のフィードバック情報を基にパケットのフロータイプを変化させる。End to End モデルを前提としたネットワーク状況に対して効率的なパケット転送スケジューリングの構築を目標とする。

本研究では、ネットワークの帯域変化がパケット到達ジッタに現れることに着目した。送信側は受信側と協調し、受信側のパケット到達ジッタ、バッファリング状況に応じてパケットの送出間隔を変化させる。フロータイプを変化させることで、受信側のパケット到達ジッタをできるだけ押え、バッファリング量を制御する。受信側の計算機資源の差異に関係することなく、リアルタイム映像配信の信頼性を向上させることが本研究の意義である。

3.2 提案：パケットフロー特性に基づく映像・音声配信機構

受信側は送信側から転送されたストリームデータパケットの間隔と受信時の間隔を比較することにより、ネットワークの状況把握を行なう。図 3.1 にその様子を示す。

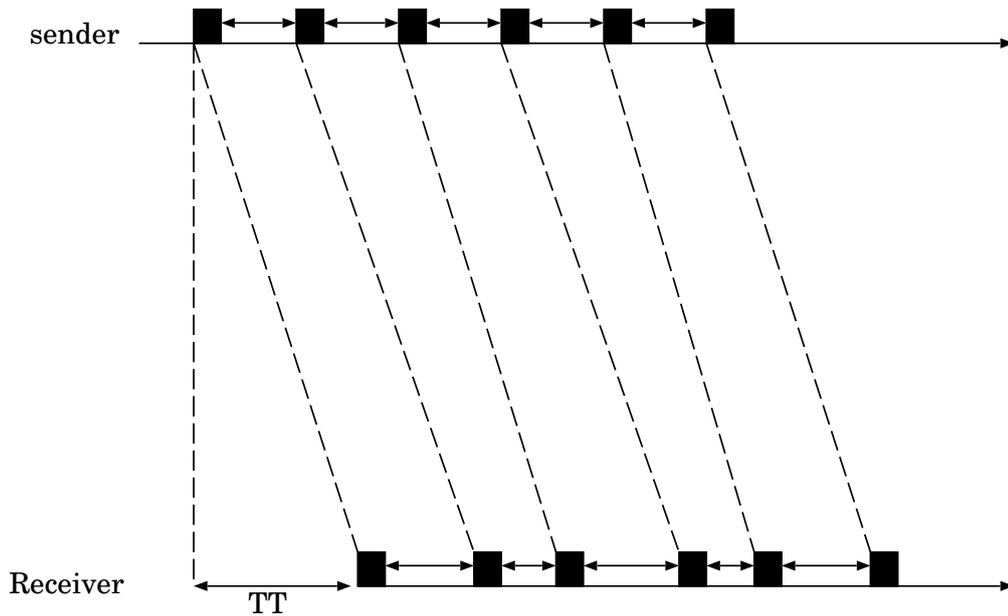


図 3.1: パケットフローの観察

ストリームデータを観察するため、送信側は計測パケットを送信する必要がない。そのため、時間間隔を細かくしてジッタ計測を行なえる。ジッタ計測によるネットワーク状況把握には、パケットペア理論 [3] を用いる。

パケットペアの概要

パケットペアは中継ノードを利用したボトルネックリンクの帯域推測の手法である。パケットペアでは、データの送信側はペアとなる2つのパケットを連続的に転送する。公平待ち行列法では、各コネクションごとにキューが割り当てられるので、ペアとなっている2つのパケットは一緒にキューイングされる。キューイングされたパケットはボトルネックリンクを通過する時にパケットの間隔が拡大される。受信側はパケットを受信すると同時に送信側にackを返す。確認応答パケッ

トの大きさは、データパケットに対して十分小さいためにボトルネックリンクでの拡大は生じない。送信側は受信側から受けとった ack の RTT (Round Trip Time) の差を計測し、ボトルネックリンクの帯域を推測することができる。図 3.2 にパケットペアによる帯域推測概念図を示す。

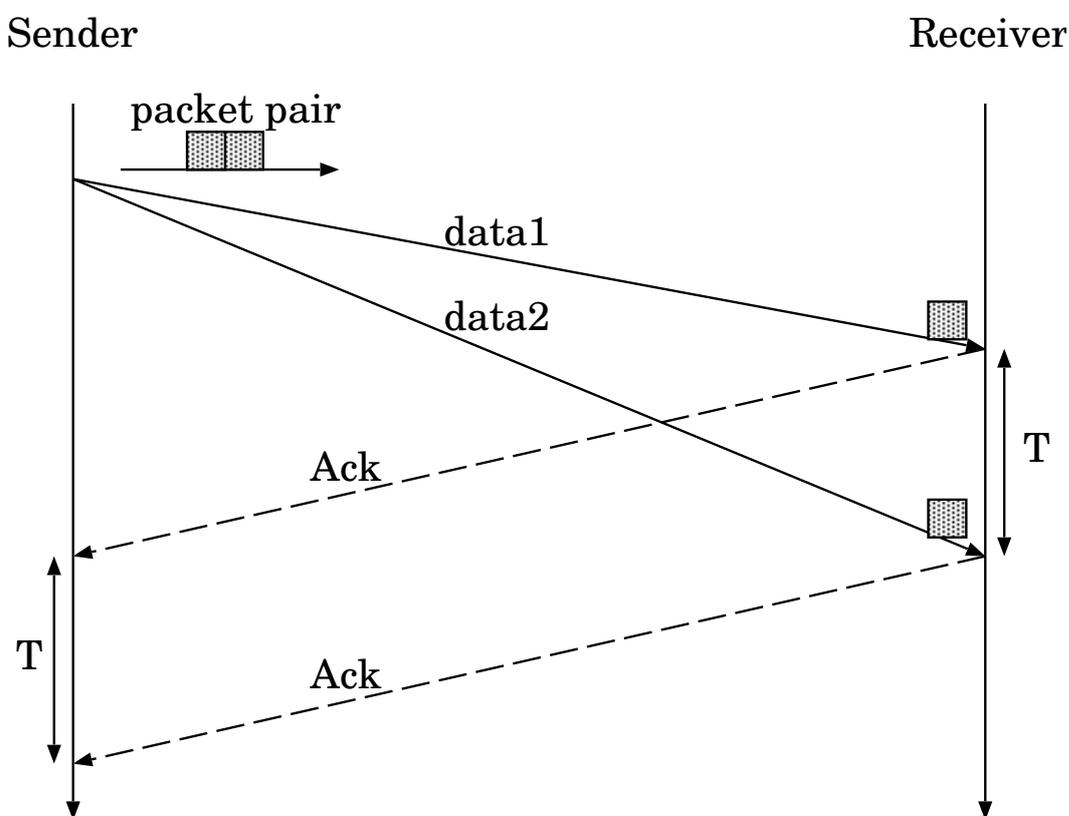


図 3.2: パケットペア方式の概念図

送信側が送ったパケットの大きさ、RTTの時間差 T から経路ネットワークにおけるボトルネックリンクの帯域を次の式で求めることができる。

$$\text{ボトルネックリンクの帯域 (byte/sec)} = \frac{\text{データパケットサイズ (byte)}}{T(\text{sec})}$$

パケットペア理論を用いる正当性の検証

パケットペアは中継ノードが公平待ち行列法を用いる場合に機能することを述べた。しかし、インターネット中の中継ノードのキューイング機構には、公平待ち行列法を採用している実装は少ない。多くの場合はFIFO(First In First Out)型のキューイング機構である。従ってペアとなる2つのパケットが連続的にキューイングされず、他のトラフィックのパケットが二つのパケット間にキューイングされてしまう可能性がある。この場合、計測結果には誤差が生じてしまう。しかし、中継ノードが公平待ち行列を採用していなくても、パケットが連続的にキューイングされる場合がある。このような考えから、パケットペアを用いて帯域推測を行なうアプリケーションとして、ボトルネックの許容帯域の推測を行なう bprobe[4] というアプリケーションがある。

上記にあげたアプリケーションはいくつかの対策を施すことで、パケットペアを用いる問題点を解決している。以下にそれらの問題点をまとめる。

- データパケットのサイズが小さい場合、ボトルネックリンクでパケットの間隔が拡大されない
- 他のトラフィックによるパケット間隔の拡大
- 計測パケットの喪失、破損
- 受信側からの応答パケット側の通信経路の輻輳

bprobe では上記にあげた問題点に対処するため、1) 転送するパケットサイズを変化させる、2) 連続的に転送するパケット数を増やす、3) 複数回計測を行なう、という実装を行なった。その解析結果によると、広域でボトルネックリンクの帯域が大きな場合に若干精度が低下したが、主に $\pm 20\%$ の範囲に80%から90%程度の計測値が存在することが報告されている。以上のことから、中継ノードが公平待ち行列を採用してない場合でも、パケットペアによる帯域推測が有効であることが分かる。

前節で述べたフロー観察手法の場合、1) 計測回数の多さ、2) パケットサイズの大きさ、3) 受信側での計測により、いくつかの問題点は解消している。しかし、送信側は連続した2つのパケットを転送するとは限らないため、他のトラフィックによるパケット間隔の拡大は、bprobe よりも大きくなる。そのため、受信側では2つのパケット間隔の開きによってネットワークの状況を推測するのではなく、統計的に推測を行なう。送信側のパケット送出タイミングの理論値を基に、ネットワークの帯域を割出し、その情報を基に送信側はパケットの送出タイミングを変

化させる。送出タイミングをボトルネックリンクの許容帯域に合わせることで、パケット到達ジッタを抑える。

3.2.1 フロー制御の要素

以下に本手法が提案するフロー制御の要素を述べる。

- パケットの送出タイミングの変更
送信側はトラフィックをシェーピングして送出する。これにより、1) バーストトラフィックによる経路ネットワーク上でのパケットロスを防ぐ、2) 受信側でパケットフローを正確に観察することが可能となる。また、受信側のパケットフロー観察の情報を基に、パケットの送出タイミングを変更する。これは、1) ネットワークの帯域幅に適した転送レートで送る、2) 瞬間的な輻輳状態によるパケットロスを防ぎ、映像品質を下げることで転送レートを下げ過ぎるのを防ぐためである。
- MTU(Maximum Transmission Unit) の変更
MTU を変更して送信側はデータパケットを転送できるようにする。これは、ジッタは安定しているが、開きが出ているため、受信側でバッファアンダーランを予測した場合である。Path MTU を変更してデータパケットを転送することで、バッファアンダーランを未然に防ぐ。

3.3 実装：パケットフロー特性に基づく映像・音声配信機構の構築

FreeBSD 5.2RELEASE を用い、カーネルランドでの実装を行っている。また利用対象ソフトウェアとして DVTS を想定した。

本研究では以下のモジュールを作成した。バッファリング監視モジュールを除き、全てカーネルランドでの実装を行っている。

- 出力制御モジュール
パケットの送出間隔の制御を行う
- スケジューリング決定モジュール
受信側から送られてきたレポート結果よりパケット送出間隔の計算を行う。
- 受信タイミング監視モジュール
パケットの受信間隔を監視する。ネットワーク伝送遅延の揺らぎ計測も行う。

- レポート送出部パケット受信状況，バッファ状況などをレポートとして送出側に送信する．
- バッファリング監視モジュール
アプリケーション層に存在する．アプリケーションのバッファ管理を行い，バッファオーバーフロー，アンダーフローの検知を行う．

3.4 まとめ・今後

本研究ではパケットフロー特性に基づく映像・音声配信機構の構築に取り組んでいる．本機構を用いることにより，ネットワーク伝送遅延の揺らぎへの対応が可能となり，映像・受信端末におけるバッファオーバーラン，アンダーランに対する信頼性が向上する．

今後は実ネットワークでの評価などを行う．

関連図書

- [1] Institute of Electrical and Electronics Engineers, Inc. Carrier Sense Multiple Access with Collision Detection(CSMA/CD) Access Method and Physical Layer Specifications -Media Access Control(MAC) Parameters, Physical Layer and Management Parameters for 10Gb/s Operation. <http://standards.ieee.org/reading/ieee/std/lanman/restricted/802.3ae-2002.pdf>, 2002.
- [2] Akimichi Ogawa. *DVTS (Digital Video Transport System) WWW page*, November 2001. <http://www.sfc.wide.ad.jp/DVTS/>.
- [3] Analysis Of Single. Packet-pair bandwidth estimation: Stochastic.
- [4] Bob Carter. Bprobe and cprobe – network probe tools home page.