

(11)特許出願公表番号

特表2004-504734

(P2004-504734A)

(43) 公表日 平成16年2月12日(2004.2.12)

(51) Int.Cl.⁷
H04L 12/56

F I
H O 4 L 12/56

E テーマコード (参考)
5K030

審查請求 有 予備審查請求 有 (全 74 頁)

| | |
|---------------|------------------------------|
| (21) 出願番号 | 特願2001-548974 (P2001-548974) |
| (86) (22) 出願日 | 平成12年12月28日 (2000.12.28) |
| (85) 翻訳文提出日 | 平成14年6月28日 (2002.6.28) |
| (86) 国際出願番号 | PCT/US2000/035490 |
| (87) 国際公開番号 | W02001/048983 |
| (87) 国際公開日 | 平成13年7月5日 (2001.7.5) |
| (31) 優先権主張番号 | 09/474, 039 |
| (32) 優先日 | 平成11年12月28日 (1999.12.28) |
| (33) 優先権主張国 | 米国 (US) |

(71) 出願人 502279445
エーディーシー ブロードバンド アクセ
ス システムズ, インコーポレイティド
アメリカ合衆国, マサチューセッツ 01
752, ウェストボロー, テクノロジー
ドライブ 8

(74) 代理人 100077517
弁理士 石田 敬

(74) 代理人 100092624
弁理士 鶴田 準一

(74) 代理人 100119987
弁理士 伊坪 公一

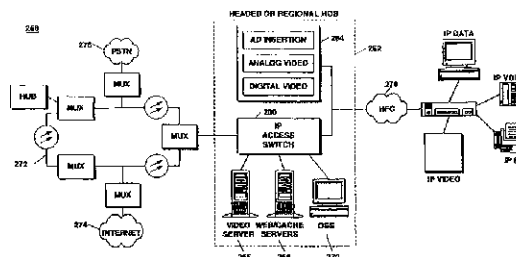
(74) 代理人 100082898
弁理士 西山 雅也

[最終頁に続く](#)

(54) 【発明の名称】 ブロードバンド網内でデータパケットを高利用度で、直接的に、フレキシブルにかつ拡張可能に交換するシステムおよび方法

(57) 【要約】

ブロードバンド・データ処理システムとネットワークのノード間で、データパケット・フローの接続を直接的かつフレキシブルに交換するためのシステムおよび方法。このシステムは、単一のIPスイッチとして動作する。



【特許請求の範囲】**【請求項 1】**

データ配信サブネットワークを接続するための装置であって、

(a) 該装置が、メッシュ線を有するメッシュに対する複数の同一のインタフェイスを有する少なくとも 1 つのシャーシを有しており、前記メッシュ線が、前記インタフェイスの各対相互間でデータを通信するための連続的なダイレクト・チャネルを提供しており、さらに、

(b) 前記装置が、データ配信サブネットワークに接続するように適合された少なくとも 1 つのデータ・ポートを備えた少なくとも 1 つのデータ処理モジュールを有しており、該モジュールが、前記シャーシの前記インタフェイスのうちの 1 つと関与していて、前記ポートのうちの 1 つを介して受信されたデータの情報に対して応答して、前記モジュールとインタフェイスされている前記メッシュ線のうちの選択されたメッシュ線に前記データをルーティングする

10

ことを特徴とする、データ配信サブネットワークを接続するための装置。

【請求項 2】

前記データの前記ルーティングが、前記データに付けられたタグによって決定され、前記タグが、前記同一のインタフェイスの中から選択された 1 つまたは複数の宛先インタフェイスを示している、請求項 1 に記載の装置。

【請求項 3】

前記装置がさらに、前記データ処理モジュールに論理的に接続されたプロセス・コントローラ・モジュールを有している、請求項 1 に記載の装置。

20

【請求項 4】

前記装置がさらに、

(c) 少なくとも 1 つのシャーシ間データリンクと、

(d) 請求項 1 に記載のシャーシと同一の少なくとも 1 つの二次シャーシと、

(e) データ配信サブネットワークに接続するように適合された少なくとも 1 つのデータ・ポートを備えた少なくとも 1 つのデータ処理モジュールとを有しており、該モジュールが、前記二次シャーシの前記インタフェイスのうちの 1 つと関与していて、前記シャーシ間データリンクから受信されたデータの情報に対して応答して、選択されたデータ配信ポートに前記データをルーティングする、

30

請求項 1 に記載の装置。

【請求項 5】

(c1) 前記シャーシ間データリンクが、シャーシ間データリンクメディアに接続するように適合された少なくとも 1 つのデータ・ポートを備えた少なくとも 1 つのデータ処理モジュールとを有しており、該モジュールが、前記一次シャーシの前記インタフェイスのうちの 1 つと関与していて、前記インタフェイスを介して受信されたデータの情報に対して応答して、前記シャーシ間データ・ポートに前記データをルーティングし、

(c2) 前記シャーシ間データリンクが、前記シャーシ間データポートに一方の端部で接続されたシャーシ間データリンクメディアを有しており、さらに、

(c3) 前記シャーシ間データリンクが、少なくとも 1 つのデータ・ポートを備えたデータ処理モジュールを有しており、該データ・ポートが、シャーシ間データリンクメディアに接続するように適合されていて、該シャーシ間データリンクメディアの他方の端部に接続されており、前記モジュールが、前記二次シャーシの前記インタフェイスのうちの 1 つと関与していて、前記シャーシ間データ・ポートを介して受信されたデータの情報に対して応答して、前記モジュールとインタフェイス状態にある前記メッシュ・チャネルのうちの 1 つに前記データをルーティングする、

40

請求項 4 に記載の装置。

【請求項 6】

前記装置がさらに、各データ処理モジュールに論理的に接続されたプロセス・コントローラ・モジュールを有している、請求項 4 に記載の装置。

50

【請求項 7】

前記プロセス・コントローラ・モジュールが、前記シャーシ間データリンクを介して、前記二次シャーシモジュールと関与するデータ処理モジュールに論理的に接続されている、請求項 6 に記載の装置。

【請求項 8】

データ配信サブネットワークを接続するための装置であって、該装置が：

(a) ルート・サーバを提供する少なくとも 1 つのプロセッサをそれぞれ含む少なくとも 2 つのシャーシと、

(b) シャーシ間リンクと、

から成っており、

前記シャーシが、前記データ配信サブネットワークに対するポートを提供し、1 つの前記ルート・サーバがマスター・ルート・サーバとして作用して、前記シャーシが単層 3 スイッチとして作用するように、それぞれ他のルート・サーバと通信する、ことを特徴とする、データ配信サブネットワークを接続するための装置。

【請求項 9】

前記シャーシのうちのいずれか 1 つの前記ポートのうちの 1 つを介して受信されたデータに、スイッチ・インターナル・タグが割り当てられて、前記データ配信サブネットワーク内の宛先に転送するための、前記ポートのうちの適切な 1 つを特定する、請求項 8 に記載の装置。

【請求項 10】

前記マスター・ルート・サーバが、それぞれ他のルート・サーバにルーティング・テーブルを提供する、請求項 8 に記載の装置。

【請求項 11】

前記シャーシのうちの少なくとも 1 つが複数のプロセッサから成っており、該プロセッサがそれぞれ、切換え可能に互いに接続されていて、それぞれが前記ポートのうちの 1 つまたは複数のポートを提供しており、それぞれが前記シャーシ間リンクに論理的に接続されている、請求項 8 に記載の装置。

【請求項 12】

前記複数のプロセッサを有するシャーシの前記ポートのうちの 1 つを介して、または、前記シャーシに対する前記シャーシ間リンクを介して受信されたデータに、シャーシ・インターナル・タグが割り当てられて、転送のための、前記複数のプロセッサを有するシャーシの前記ポートのうちの適切な 1 つのポート、または前記シャーシに対する前記シャーシ間リンクを特定する、請求項 11 に記載の装置。

【請求項 13】

データ処理アプリケーション・モジュールを保持するための複数の同一のスロットを有するシャーシであって、各スロットが、メッシュ線を有するバックプレーン・メッシュに対する同一のインタフェースを有しており、前記メッシュ線がそれぞれ、前記インタフェースの各対相互間でデータを通信するための連続的なダイレクト・チャネルを提供しており、前記メッシュ線が、毎秒 1.5 ギガビット以上の速度でのデジタル・データパケット通信に相応する処理能力を有していることを特徴とする、データ処理アプリケーション・モジュールを保持するための複数の同一のスロットを有するシャーシ。

【請求項 14】

前記シャーシがさらに、管理制御信号バスを有しており、前記同一のインタフェースが前記バスとインタフェースしている、請求項 13 に記載のシャーシ。

【請求項 15】

前記シャーシがさらに、電気通信バスを有している、請求項 13 に記載のシャーシ。

【請求項 16】

前記シャーシが、前記シャーシのバックプレーン側で前記モジュールに対応する外部データ・ポートに接続可能である、請求項 13 に記載のシャーシ。

【請求項 17】

10

20

30

40

50

前記シャーシが、前記モジュールにののための、完全に分配されたDC電源装置を有している、請求項13に記載のシャーシ。

【請求項18】

前記同一のスロットの数が4つよりも大きい、請求項13に記載のシャーシ。

【請求項19】

データ処理アプリケーション・モジュールであって、該モジュールが、データ配信サブネットワークに接続するように適合された少なくとも1つのデータ・ポートを備えており、前記モジュールがさらに、シャーシのスロット内に挿入するように適合されて、これにより前記スロットと関連するインタフェースに関与しており、前記インタフェースが複数のチャンネルを有しており、該チャンネルがそれぞれ、毎秒1.5ギガビット以上の速度でのデジタル・データパケット通信能力をそれぞれ有しており、前記モジュールが、前記データ・ポートを介して受信されたデータの情報に対して応答して、前記複数のチャンネルのうちの選択されたチャンネルに前記データをルーティングすることを特徴とする、データ処理アプリケーション・モジュール。

10

【請求項20】

前記データの前記ルーティングが、前記データに前記モジュールによって付けられたタグによって決定されるようになっており、前記タグが、前記データが最終的にルーティングされるべき、前記シャーシにおける1つまたは複数のスロット・インタフェースを示しており、前記モジュールが、第1のルートが機能しない場合には代替ルートを選択可能であり、さらに、前記モジュールのスロット・インタフェースとは異なる最終的なスロット・インタフェースを示す添付タグを有する受信データを再ルーティング可能である、請求項19に記載のモジュール。

20

【請求項21】

前記モジュールが、並列ブロックで検査された前記受信データに対して応答し、前記チャンネル上で前記データを直列に伝送するようになっている、請求項19に記載のモジュール。

【請求項22】

前記モジュールが各前記チャンネル専用の別個のトランシーバを有している、請求項19に記載のモジュール。

【請求項23】

データ処理アプリケーション・モジュールであって、該モジュールが、データ配信サブネットワークに接続するように適合された少なくとも1つのデータ・ポートを有しており、前記モジュールが、同一のモジュールと組み合わせられた前記モジュールが単層3スイッチとして作用できるように、マスター・ルート・サーバと同期するように適合されたルート・サーバを有していることを特徴とする、データ処理アプリケーション・モジュール。

30

【請求項24】

前記データ・ポートのうちの1つを介して受信されたデータに、スイッチ・インターナル・タグが割り当てられて、宛先データ配信サブネットワークに転送するための、前記ポートのうちの適切な1つを特定する、請求項23に記載の装置。

【請求項25】

前記モジュールが、実質的に同一のデータ処理モジュールを有する単一のシャーシに取り付けるように適合されており、前記データ・ポートのうちの1つを介して受信されたデータに、シャーシ・インターナル・タグが割り当てられて、宛先データ配信サブネットワークに転送するための、前記ポートのうちの適切な1つを特定する、請求項23に記載の装置。

40

【請求項26】

1つまたは複数のシャーシ内に組み込まれたデータ処理アプリケーション・モジュールにそれぞれ接続されたデータ配信サブネットワークにユーザを接続するための方法であって、該方法が以下のステップ、すなわち：

A) シャーシ内に組み込まれて1つのデータ配信サブネットワークに接続されたモジュ

50

ール上でデータを受信するステップと、

- B) i) 前記シャーシに組み込まれたモジュールに接続された、ii) 前記シャーシに対するシャーシ間データリンクを有する別のシャーシに組み込まれたモジュールに接続された、または、iii) 前記シャーシ、または前記シャーシに対するシャーシ間データリンクを有する別のシャーシのモジュールを介してアクセス可能なインターネットまたは同様のネットワークに接続された、データ配信サブネットワーク上に、前記受信データの宛先があるか否かを、前記受信データのヘッダ情報から決定するステップと、
- C) 前記シャーシ内で前記データをルーティングするために、前記受信データにヘッダ・タグを付けるステップと、
- D) 前記シャーシ内に組み込まれたモジュールから前記ヘッダ・タグなしで前記データを伝送するステップと
- から成ることを特徴とする、データ配信サブネットワークにユーザを接続するための方法

【請求項 27】

シャーシ内で相互接続されたデータ通信モジュール間でデータ・フローを交換するための方法であって、該方法が以下のステップ、すなわち：

- A) 前記データを 1 つの前記モジュールで受信するステップと、
- B) 前記モジュール内の m ビット幅の平行ル・バスに一度に m ビットのデータを提供するステップと、
- C) 前記データを m ビットの平行ル待ち行列に蓄積するステップと、
- D) どのモジュールが適切な宛先モジュールであることを決定するステップと、
- E) どの宛先モジュールに前記データがルーティングされるべきかを示すために、前記データにタグを付けるステップと、
- F) 前記 m ビットのデータを前記 m ビット幅の平行ル・バスに一度に供給するステップと、
- G) 前記宛先モジュールへの経路と関連するシャーシ内チャネル内に、前記 m ビットのデータをゲーティングするステップと、
- H) 前記 m ビットのデータをシリアル化するステップと、
- I) 前記経路と関連する前記チャネル専用の高速シリアル・リンクに、前記シリアル化された m ビットのデータを供給するステップと、
- J) 前記宛先モジュール内の m ビット幅の平行ル・バスに前記 m ビットのデータを一度に供給するステップと、
- K) 前記データを m ビットの平行ル待ち行列に蓄積するステップと、
- L) 前記宛先モジュール内の前記 m ビット幅の平行ル・バスに前記 m ビットのデータを一度に供給するステップと、
- M) 前記宛先モジュールから前記データを伝送するステップと
- から成ることを特徴とする、データ・フローを交換するための方法。

【請求項 28】

複数のシャーシに対するポートを有するデータ配信サブネットワークを接続するための方法であって、該方法が以下のステップ、すなわち：

- a) 前記複数のシャーシのそれぞれに対応するルート・サーバを同期するステップと、
- b) 前記シャーシのうちの 1 つのポートで受信されたデータにタグを付け、前記データの情報と、前記シャーシに対応する同期されたルート・サーバとにしたがって転送するためのポートを指定するステップと、
- c) 前記タグによって指定された前記ポートに前記データを転送するステップと
- から成る、データ配信サブネットワークを接続するための方法。

【請求項 29】

データ配信サブネットワークにユーザを接続するための装置であって、該装置が、シャーシ内の組込みに適合されて、同一または他のシャーシ内に組み込まれた実質的に同様の 1 つまたは複数のモジュールに論理的に接続されるべきデータ処理アプリケーション・モジ

10

20

30

40

50

ジュールから成っており、該モジュールが、

A) データを受信するための手段と、

B) i) 前記シャーシに組み込まれたモジュールに接続された、ii) 前記シャーシに対するシャーシ間データリンクを有する別のシャーシに組み込まれたモジュールに接続された、または、iii) 前記シャーシ、または前記シャーシに対するシャーシ間データリンクを有する別のシャーシのモジュールを介してアクセス可能なインターネットまたは同様のネットワークに接続された、データ配信サブネットワーク上に、前記受信データの宛先があるか否かを、前記受信データのヘッダ情報から決定するための手段と、

C) 前記シャーシ内で前記データをルーティングするために、前記受信データにヘッダ・タグを付けるための手段と、

D) 前記シャーシ内に組み込まれたモジュールから前記ヘッダ・タグなしで前記データを伝送するための手段と

から成っていることを特徴とする、データ配信サブネットワークにユーザを接続するための装置。

【請求項 30】

シャーシ内で相互接続された実質的に同様のモジュール間でデータ・フローを交換するためのデータ通信モジュールから成る装置であって、前記モジュールが：

A) 前記データを受信するための手段と、

B) mビット幅の平行・バスに一度に前記mビットのデータを供給するための手段と、

C) 前記データをmビットの平行待ち行列に蓄積するための手段と、

D) どのモジュールが適切な宛先モジュールであるかを決定するための手段と、

E) どの宛先モジュールに前記データがルーティングされるべきかを示すために、前記データにタグを付けるための手段と、

F) 前記mビットのデータを前記mビット幅の平行・バスに一度に供給するための手段と、

G) 前記宛先モジュールへの経路と関連するシャーシ内チャネル内に、前記mビットのデータをゲーティングするための手段と、

H) 前記mビットのデータをシリアル化するための手段と、

I) 前記経路と関連する前記チャネル専用の高速シリアル・リンクに、前記シリアル化されたmビットのデータを供給するための手段と

から成っていることを特徴とする、データ・フローを交換するためのデータ通信モジュールから成る装置。

【請求項 31】

複数のシャーシに対するポートを有するデータ配信サブネットワークを接続するための装置であって、該装置が：

a) 前記シャーシのそれぞれに組み込まれた少なくとも1つのデータ・プロセッサと、

b) 前記シャーシのそれぞれに設けられた少なくとも1つのルート・サーバと、

c) 前記ルート・サーバのそれぞれを同期するための手段と、

d) 前記シャーシのうちの1つのポートで受信されたデータにタグを付け、前記データの情報と、前記シャーシに対応する同期されたルート・サーバとにしたがって転送するためのポートを指定するための手段と、

e) 前記タグによって指定された前記ポートに前記データを転送するための手段と

から成ることを特徴とする、データ配信サブネットワークを接続するための装置。

【発明の詳細な説明】

発明の背景

本発明の分野は、データ処理システムのブロードバンド網を介する通信分野、例えば、SONETバックボーンと、ユーザをバックボーンに接続するためのハイブリッド・ファイバー・同軸(ケーブル)(「HFC」)アクセス網とを使用するネットワークを介する通信分野である。HFC環境に対応するパケット通信基準はケーブル・ラボ(Cable L

10

20

30

40

50

abs)のDOCSIS(ケーブル経由データ・サービス・インタフェース仕様(Data Over Cable Service Interface Specification)、資料番号SP-RFI-I05-991105)である。

図1は、(それぞれ200~2000件の適合世帯(HHP)に接続される)HFCシステムへの広域データ供給の典型的な形態を示す。ヘッドエンド・ターミネーション・システム2は、リング6においてマルチプレクサ・ドロップを介してSONETリング4に接続されている。ヘッドエンド2は、ファイバー下り側接続部7とファイバー上り側接続部8とによって、ファイバー・ノード9に接続されている。ファイバーノード9は、ケーブルモデム加入者(すなわち家庭11)に出る同軸コネクタ10を有している。これらのマルチプレクサには目下のところ、ヘッドエンド・ターミネーション・システムに加えて、およそ\$50,000の費用がかかり、コミュニティへのサービス拡大は、新しいマルチプレクサおよびサーバを必要とする。

ヘッドエンド・ターミネーションシステム上のコンポーネントが機能しないと、「下り(ヘッドエンドからエンドユーザへの)サブネットワーク(図7)全体が接続されるネットワークとの通信機能を奪われてしまうおそれがある。

コストを節減し、システム管理を容易にするために、システムを統合する試みがなされてきた。現行の「統合」データ供給システム20を図2の機能ブロックとして示す。図2は、リバース・パス・モニタリング・システム22と、IPスイッチ24と、ルータ26と、変調器およびアップコンバータ28と、プロビジョニング・システム30と、テレフォニー部分32と、複数のCMTS34(ケーブルモデム・ターミネーション・システム)とを有するシステムを示す。このタイプのシステムは、複数のベンダーからの複数のシステムを有するように構成され、種々異なる管理システムと、大きな占有面積と、高い所要電力と、高い動作コストを有しているのが普通である。

音声およびデータを供給するための代表的な現行のブロードバンド・ケーブル網を図3に示す。SONET(POS)リンク40,42を介した2つのOC-12パケットが、それぞれ2つのバックボーン・ルータ44,46のうちの一方に接続されている。これらのバックボーン・ルータは2つのスイッチ48,50にネットワークされている。スイッチ48,50はCMTSヘッド末端ルータ52,54にネットワークされている。CMTSヘッドエンド・ルータ52,54は、CMTS61を介して複数の光学ノード56に接続されている。スイッチ48,50はまた、複数の電話幹線ゲートウェイ58に接続されており、これらのゲートウェイは公衆交換電話網(PSTN)60に接続されている。図2に示した「統合」システムと同様に、このタイプのネットワークも、複数のシステムに対応する複数のベンダーを有しており、種々異なる管理システムと、大きな占有面積と、高い所要電力と、高い動作コストを有している。システムのサイズ、必要な電力およびコストを低減し、データ供給における整合性の増大および信頼性の向上を保証することが依然として望まれる。

本発明の目的は、高品質の音声、データおよびビデオサービスをブロードバンド供給するための統合システムおよびプロセスを提供することである。

本発明の別の目的は、高いネットワーク信頼性を有すると共に、ライフライン電話サービスを信頼性高く支援する能力、および、階層化された音声・データサービスを供給する能力をも備えたケーブル・アクセス・プラットフォームのためのシステムおよびプロセスを提供することである。

本発明のさらに別の目的は、安全かつ拡張可能なネットワーク・スイッチのためのシステムおよびプロセスを提供することである。

発明の概要

ブロードバンド網環境においてデータ処理を行うための統合ネットワーク・ソリューションを提供する上での問題点は、本発明による、高品質の音声、データおよびビデオサービスをブロードバンド供給するための、高度に統合された搬送クラス・ブロードバンド・アクセス・システムにより解決される。

本発明において、HFC経路への切換えがIP(インターネット・プロトコル)またはネ

10

20

30

40

50

ットワーク層（OSI層3）において、広域網（WAN）、インタフェース・カード、HFCインタフェース（ケーブルモデム・ターミネーション）カード、および、クラスター・マネージャによって制御された、完全にメッシュ化された共通シャーシ内のシャーシ間リンク（ICL）カードの組み合わせを用いて行われるようなプロセスおよびシステムが提供される。ICLは数マイルにわたって延びてよいが、しかし、クラスター・マネージャによって制御される、シャーシのクラスターは、単一のIPスイッチとして働く。シャーシは、同様に他のブロードバンド・アクセス・メディア、例えばデジタル加入者ライン（DSL）、固定無線ローカル・マルチポイント配信サービス（LMDS）およびマルチチャンネル・マルチポイント配信サービス（MMDS）、「縁石までの光ファイバ（fiber to the curb）」（FTTC）とインタフェースすることができる。

本発明は、種々様々な局地的および広域的形態を配備する上でのフレキシビリティを提供し、規模拡張のために準備されている経路を必要とする小規模なオペレータに、このような形態に適したものにする。電話モデム・リターンを有する一方向ケーブルのように、種々異なるブロードバンド通信様式には、種々異なるメディア・カードが使用されてよい。本発明はまた、種々異なるネットワーク・エレメントに複数レベルの冗長性を許し、特にストレスを与えられるエレメントの冗長性を有することにより、より高い利用可能性をオペレータに提供するのを可能にする。本発明は、信頼性の高い拡張可能なシステムを、僅かな設置面積および低所要電力と共に提供する。

詳細な説明

図4は、本発明の原理に従って動作するシャーシ200を示す。このシャーシ200は、複数のネットワーク・インタフェースとアプリケーションとを単一のスイッチシステム内に統合している。本発明のシャーシは、シャーシ・コントローラ内のクラスター・マネージャによって制御された低レベルの組込みソフトウェアを使用した、高性能のパケット転送、フィルタリングおよびQoS/CoS（サービス品質/サービス・クラス）能力を有する、完全にメッシュ化されたIPスイッチである。パケット転送、フィルタリングおよびQoS/CoSは、シャーシ内部のアプリケーション・モジュール（「カード」とも呼ばれる）にわたって分布されている。動作は、高速IPプロセッサによってそれぞれのモジュールにおいて実行される。シャーシ・コントローラおよびクラスター・マネージャは動作を制御し、それぞれのモジュールを構成する。クラスター・マネージャ内には、ルータ・サーバ機能（RIPv1、RIPv2、OSPFなど）、ネットワーク管理（SNMP V1/V2）、セキュリティ、DHCP、LDAPおよびリモート・アクセス・ソフトウェア（VPNs、PPTP、L2TPおよびPPP）を含む、より高レベルのソフトウェアが位置しており、このソフトウェアは容易に変更またはアップグレードすることができる。

本発明のこの実施例の場合、シャーシ200は14個のモジュール用スロットを有している。これら14個のスロットのうち12個はアプリケーション・モジュール205を保持しており、2つのスロットはシャーシ・コントローラ・モジュール210を保持している。各アプリケーション・モジュールはDC-DCコンバータ202を搭載しており、シャーシ内に「ホットプラグ可能」である。シャーシ・コントローラ210は、冗長システムのクロック/バス・アービトレーションのために働く。シャーシはまた、冗長電源モジュールを有している。電源モジュールおよびDC-DCコンバータは、シャーシ内で完全分配された電源装置から成っている。シャーシ内に統合可能なアプリケーションの例はCMTSモジュール215、イーサネット・モジュール220、SONETモジュール225およびテレフォニー・アプリケーション230である。他のアプリケーションとしては、シャーシ間リンク（ICL）ポート235があつてよく、このポート235を介して、シャーシを別のシャーシにリンクすることができる。ICLポートはただの出口ポートである。ICLポートは、あらゆる出口モジュール、例えば10/100イーサネット、IGイーサネットおよびパケット・オーバー・SONET（PoS）を使用して実行可能である。

図5は、図4のシャーシ200を有する統合ケーブル・インフラストラクチャ260を示

す。シャーシ 200 は音声およびデータの供給のための地域ハブ 262 の一部である。ハブ 262 は、ビデオコントローラ・アプリケーション 264 と、ビデオサーバ 266 と、ウェブ/キャッシュ・サーバ 268 と、動作支援システム (OSS) 270 と、IP アクセススイッチとして働くシャーシ 200 にネットワークされた全てのものを含んでいる。シャーシ 200 は S O N E T リング 272 に、ハブ 262 の外部で接続されている。S O N E T リング 272 はインターネット接続部 274 と、公衆交換電話網 (PSTN) 接続部 276 とを有している。シャーシ 200 はまた、H F C リング 278 によってケーブル顧客に接続されており、音声、データ、ビデオおよびファックスサービスを含む、IP に基づくサービスを提供する。各 H F C アプリケーション・モジュールは、最大 2000 件までのケーブルモデム・サービス加入者を取り扱う。加入者数の論理的限界は 8192 10 件である。シャーシは複数の H F C リングをサポートすることができ、また、100 万件を超えるケーブルモデム加入者をサポートするために、複数のシャーシと一緒にネットワークされてよい (後述のとおり)。

今日の慣例により、ユーザ (例えばデスクトップ型コンピュータ、ファクシミリ機または電話セット) への送信 (ダウンロード) のために 1 つの広帯域チャネル (27 ~ 40 Mbps) があり、アップロードのためには、より狭帯域の複数のチャネル (320 Kbps ~ 10 Mbps) がある。これは、O / E ノードで双方向通信することにより H F C カードによって処理される。ローカルな H F C ケーブル・システムまたはループは、ケーブルモデムへのドロップを有する同軸ケーブル分布網である。

図 6 は、図 4 の複数のシャーシを使用するリング・ネットワーク 300 を示す。複数のシャーシ 302, 304, 306, 308 は、全二重シャーシ間リンク 310, 312, 314, 316 にわたって延びるリングとして接続されている。このリング形態は、シャーシ 304 または I C L 310, 312 がシャーシ 302 および 306 の間で機能しないとしても、シャーシ 302 およびその関連ケーブル網がシャーシ 306 およびその関連ケーブル網と通信するのを可能にする。 20

図 7 は、本発明の原理に従って動作する一次シャーシ 330 と二次シャーシ 332 とを有する第 1 の階層ネットワーク 328 を示す。一次シャーシはファイバー光通信ライン (O C - n) 334, 336 に接続されている。O C - 12 ラインが図示されているが、O C - 3 または O C - 48 も支持されている。一次シャーシ 330 は P S T N 338 に接続されてもよい。一次シャーシ 330 と二次シャーシ 332 とは全二重リンク 340 によって 30 リンクされている。全二重リンクは高速イーサネット (F E)、ギガビット・イーサネットまたはパケット・オーバー・S O N E T (P o S) タイプの接続部であってよい。二次シャーシ 332 は複数の光学ノード 342 に接続されている。一次シャーシと 1 つまたは複数の二次シャーシとがあるネットワークにおいては、一次シャーシがクラスターをそのクラスター・マネージャで管理する。

他の形態、例えば一次シャーシ 352 と二次シャーシ 354, 356, 358, 360, 362 とを備えた、図 8 に示したような「星」形態 350、または、二次階層に一次シャーシ 382 と二次シャーシ 384 および 386 を備え、三次階層に二次シャーシ 388, 390, 392, 394 とを備えた階層状配列 380 が可能である。シャーシ 402, 404, 406, 408, 410 を備えたメッシュ形態 400 が図 10 に示されている。この場合、各シャーシが他の全てのシャーシにシャーシ間リンク内で接続されている。一次 および二次シャーシはシステム形態によって決定される。各シャーシは他のどのシャーシとも実質的に同じである。 40

上述の形態は、長距離ファイバー光リンクを使用して数 10 マイルにわたって延びることができ、あるいは、同一のマスタ・ヘッドエンド・サイトで単一ラックに配列することもできる。

図 11 は、図 4 のシャーシ 200 のバックプレーン 420 に接続されたアプリケーション・モジュールを示す。各アプリケーション・モジュール 422 は、メッシュ通信チップ (M C C) 424 を介してバックプレーン 420 とインタフェイスする。この M C C についてはより詳細に後述する。各 M C C 424 は 12 個のシリアル・リンク・インタフェイス 50

4 2 6 を有しており、1 1 個がバックプレーン 4 2 0 に通じている。各アプリケーション・モジュールにおいてバックプレーンに通じる 1 1 個のシリアル・リンクは、アプリケーション・モジュールをシャーシ内の他の全てのアプリケーション・モジュールに接続するために設けられている。1 つのリンクはモジュールをそれ自体に接続するために、つまりループ・バックのために設けられている。アプリケーション・モジュール 4 2 2 はシャーシ管理バス 4 3 2 を介してシャーシ・コントローラ 4 2 8 , 4 3 0 に接続されている。第 2 のシャーシ・コントローラ 4 3 0 は、オプションとして、システムをより信頼性高くするため、冗長性のために使用される。好ましい実施例の場合、やはり冗長性 / 信頼性を目的として、第 2 のシャーシ管理バス (図示せず) が設けられている。

バックプレーンは完全にメッシュ化されている。このことはどのアプリケーション・モジュールも、シリアル・リンクを介して、シャーシ内の他の全てのアプリケーション・モジュールに対して直接的な二地点間リンクを有することを意味する。メッシュ・バックプレーンにおけるメッシュ線はそれぞれ、毎秒 1 . 5 ギガビット以上の速度で、データ通信のための連続的なダイレクト・チャンネルを提供する。接続部 2 0 0 の一部だけを例として図 1 1 に示す。バックプレーン・メッシュは図 1 2 に示す。

M C C のシリアル・リンクを備えた 1 2 個のチャンネルは 0 ~ 1 1 の番号を有している。この番号を、チャンネル I D または C I D と呼ぶ。チャンネルについてはより詳しく後述する。バックプレーンのスロットもまた 0 ~ 1 1 の番号を有している (スロット I D または S I D) 。しかしシャーシシステムは、チャンネル 0 がバックプレーンのスロット 0 に結線されることを必要とはしない。シリアル・リンクはどのスロットにも接続できる。スロット I D はシステム・トポロジーに応じて動的に構成される。このことはバックプレーン結線の自由を可能にし、このような結線の自由はルーティングの複雑さを低減する。

図 1 1 に戻ると、各アプリケーション・モジュールはまた、シャーシ管理バス 4 3 2 に接続されている。シャーシ管理バスはモジュールとシャーシ・コントローラとの接続を可能にする。

シャーシ間リンク (I C L) を介してシャーシ相互間でパケットを交換するため、また、M C C リンクを介してシャーシ内部でパケットを交換するために、シャーシはシャーシ間交換層を有している。このシャーシ間交換層は I P 層 3 (L 3 、ネットワーク層) の下方に位置している。シャーシ内のパケット処理は、2 つのタイプのトラヒック、すなわちユニキャストと、ブロードキャスト / マルチキャストとに分類される。シャーシ間交換層を介した交換は、図 1 3 に示したシャーシ間ヘッダ 5 0 0 (シャーシ間タグとも呼ばれる) を使用して達成される。

B a s _ T y p e フィールド 5 0 2 は x ビットを有しており、シャーシ間ヘッダタイプを示す。このフィールドは、ファブリック・インタフェース・アドレス (F I A) が論理ポートを有することを示し、パケットがブロードキャスト・パケットであるかまたはユニキャスト・パケットであることを示し、また、B A S ヘッダ内の該当フィールドを示すのに使用されてよい。B a s _ T y p e フィールド 5 0 2 は、B A S ヘッダをデコードし、パケット転送を助けるのに使用される。

保持 (K e e p) フィールド 5 0 4 は x ビットを有しており、パケットが混雑のために落下するおそれがあるかどうかを決定する。

優先順位フィールドは x ビットを有しており (図 1 3 には示さず) 、パケット優先順位を決定し、混雑のためにパケットをどこで待ち行列させるかを定める。

断片フィールド 5 0 6 は x ビットを有しており、パケットの断片化およびパケットが 2 つのフレームから成るか否かを示す。

n e x t _ H o p フィールド 5 0 8 は複数のビットを有しており、シャーシ内パケット伝送に関する次のホップ情報のために使用される。このフィールドは B a s _ T y p e フィールドが特に指示しない場合には非アクティブである。

E n c a p _ T y p e フィールド 5 1 0 は、1 ビット・フィールドであり、このフィールドは、パケットがさらに層 2 (シャーシ間交換層の下方のデータ層) の処理を必要とするかどうか、または、パケットをさらなる処理なしで転送することができるかどうかを示す

10

20

30

40

50

。

Mcast__Typeフィールド512は1ビット・フィールドであり、このフィールドは、パケットがブロードキャストかまたはマルチキャストかを示す。

Dest__FIA__Typeフィールド514は1ビット・フィールドであり、このフィールドは、宛先FIAが短い形(chassis/slot/port)で提供されるか、長い形(chassis/slot/port/logical port)で提供されるかを示す。

Src__FIA__Typeフィールド516は1ビット・フィールドであり、このフィールドは、ソースFIAフィールドが短い形で提供されるか長い形で提供されるかを示す。

Date__Typeフィールド518は複数のビットを有しており、ペイロードで運搬されているトラフィック・タイプを示すのに使用される。 10

TTLフィールド(図示せず)は複数のビットを有しており、パケットがシステム内に不定に留まることを防止するためのフェイルセーフ・メカニズムである。TTLはそれがICLポートで受信されるたびに減小される。このフィールドが減小後にゼロになると、パケットは廃棄される。

転送Infoフィールド520は複数のビットを有しており、転送テーブル・リビジョン、forward__table__entry key/id、および次のホップ情報を含んでいる。

宛先FIAフィールド522は複数のビットを有しており、パケットの最終宛先を示す。このフィールドは、シャーシ/スロット/ポートおよび論理ポートの情報を含んでいる。 20

ゼロのシャーシ値は特別な意味を有し、マスター・エージェント(後述する)を有するシャーシを意味する。ゼロのポート値もまた特別な意味を有し、パケットの受信側がアプリケーション・モジュールであることを意味する。論理ポートは、モジュール内のどのスタック/エンティティがそのパケットを受信すべきかを示すのに使用することができる。したがって全てのエッジ・ポートおよびICLポートは1をベースとしている。すなわち、ポート数はゼロよりも多い。

Src FIAフィールド524は複数のビットを有しており、パケット・ソースを示す。このフィールドは、ルート・サーバによって使用されることにより、入来パケットのソースを識別する。 30

システム・アーキテクチャにしたがって、シャーシ内にプラグインする全てのアプリケーション・カードは共通の構成要件を共有する。アプリケーション・カードはデータ・プロセッサ、データ処理アプリケーション・モジュールおよびデータ通信モジュールとも呼ばれる。図14はシャーシ内のイーサネット・アプリケーション・モジュール550のブロック図である。このアプリケーション・モジュールは、メッシュ・バックプレーン552にも、管理バスA554およびB556にも接続されている。イーサネット・アプリケーション・モジュール550においては、高速IPプロセッサ(FIPP)558がPCIバス560およびF-バス562の双方に接続されている。FIPP558は、アドバンストRISCマシン(ARM)プロセッサ・コア564と6つのマイクロ・エンジン566とを有している。FIPP558はパケットを記憶するためのDRAM568と、ルーティング・テーブルを記憶するためのSRAM570とに接続されている。 40

8つのポートを有するイーサネット・デバイス572は、F-バス562に接続されている。MCC574もまたF-バスに接続されている。MCC574はメッシュ・バックプレーン552との接続部を提供する。イーサネット・デバイス572は、シャーシ外部との接続部を提供する。2つのMACデバイスは、PCIバス560と管理バスとの間に接続されている。MAC A576はPCIバス560と管理バスA554との間に接続されている。MAC B578は、PCIバス560と管理バスB556との間に接続されている。

データパケットがイーサネット・デバイス572内に入来すると、このデータパケットはFIPP558を通り、DRAM568内に記憶される。マイクロ・エンジン566はデータパケットを並列の状態で検査する。マイクロエンジン566はパケットのIPアドレスを見、次いで、SRAM570に記憶された転送テーブル内の宛先アドレスを調べる。 50

転送テーブルはパケットが出る際に通ることになるシャーシ、スロットおよびポートを提供する。パケットがイーサネット・デバイスを介して入来するときには、パケットはイーサネット・ヘッダを有している。パケットが入来したのと同じスロットから出る場合には、シャーシ間ヘッダは付けられない。F I P Pは送信ポートを見極め、パケットをそのポートから送出する。パケットが異なるスロットによって出るようになる場合、イーサネット・ヘッダは除去され、B A Sヘッダがデータパケットに加えられる。B A Sヘッダ内の最小情報は宛先データ、シャーシ、スロットおよびポートの情報である。パケットはさらに、I Pヘッダとペイロードとを含んでいる。F I P Pは送信待ち行列を有している。この待ち行列に、パケットは送信前に並べられる。パケットはF - バス5 6 2を介してM C C 5 7 4に送られる。M C C 5 7 4はデータを6 4バイト・チャンクで送出する。

10

パケットがイーサネット・モジュール5 5 0のM C C 5 7 4を通して入来すると、アプリケーション・モジュール5 5 0はパケットまたはパケットに対するポインタを、D R A M 5 7 0の待ち行列内に置く。パケットがシリアル・ポートのうちの1つから出るようになる場合には、F I P P 5 5 8はS R A M 5 7 0のA R Pテーブルにおいて宛先を調べる。A R Pテーブルを図1 5に示す。F I P Pはシャーシ、スロットおよびポートを見出す。シャーシアドレスは、宛先がI C Lでない限り変わらない。同一シャーシの場合には、アプリケーション・モジュールはI Pパケットに対応するM C Cアドレスを見出し、パケットを送出する。パケットが他のシャーシに出ることになる場合、アプリケーション・モジュールはI C Lであるポートがあるかどうかを問う。もしあるならば、アプリケーション・モジュールはパケットをそのポートから送出する。

20

図2 0は、H F C D O C S I Sインタフェース・カードとも呼ばれるC M T Sアプリケーション・カードを示す。広域通信技術が現在利用可能であることに基づき、H F C - オン・P C I技術が利用される。しかし、将来的には他のタイプのインタフェース技術を使用することが可能である。

図1 6は、シャーシ内の2つのアプリケーション・モジュール6 0 0、6 0 2を示している。アプリケーション・モジュール6 0 0、6 0 2は、管理バス6 0 4、6 0 6と、メッシュ・バックプレーン6 0 8とを介して接続されている。第2のアプリケーション・モジュール6 0 2はP C Iバス6 1 2に取り付けられたルート・サーバ6 1 0を有している。シャーシに対応するルート・サーバはいずれのアプリケーション・モジュールにも位置することができる。本発明のこの実施例の場合、ルート・サーバはP e n t i u m（登録商標）プロセッサである。シャーシネットワーク内の各シャーシは、ルートサーバを有している。しかしどの時点においても、ネットワーク内のルート・サーバはただ1つしかアクティブになることはできない。ルート・サーバは管理バス、およびI C Lを介して、それがルート・サーバであることをブロードキャストする。

30

複数のシャーシから成るネットワークの場合、一次ルート・サーバであるように、1つのルート・サーバが指定される。他の全てのルート・サーバは二次サーバである。ネットワーク内のルート・サーバはルーティング情報を3 0秒毎に送出する。一次ルート・サーバはそのシャーシ、スロットおよびポートを他の全てのルート・サーバにブロードキャストする。

図1 7は、「コントロール」または一次ルート・サーバによるルート・サーバの更新を概念的に示す。

40

全ての転送テーブルは、そのシャーシのルート・サーバがどこに位置しているかを認識する。一次ルート・サーバはイーサネット・ポートと同等の論理ポートを有している。一緒にネットワークされたシャーシのうちの1つのシャーシのルート・サーバにパケットが送られると、一次ルータは、あたかも一次ルータがそのパケット自体を受信しているかのようにパケット情報を受信する。一次ルータはルーティング・テーブルと転送テーブルとを形成し、これらを全てのアプリケーション・モジュールおよびネットワーク上の他の全てのシャーシにブロードキャストする。転送テーブルは、ルーティング・テーブル内の情報から形成される。転送テーブルは、シャーシ、スロットおよびパケットのポートのような情報、Q o Sデータ、C M T S情報、マルチキャスト・ドメイン情報およびネクスト・ホ

50

ップ I C L 情報を含む。

図 18 は、シャーシ内で管理バスに接続されたシャーシ・コントローラを示す。冗長性を得るために、各シャーシは 2 つのシャーシ・コントローラを有している。シャーシ・コントローラは、メッシュ・バックプレーンには接続されていない。シャーシ・コントローラはネットワーク管理タスクと提供タスクとを有している。シャーシ・コントローラは、シャーシクラスター全体が、1 つの管理されたエレメントに見えることを可能にする。シャーシ・コントローラはプロセッサと、メモリと、クラフト・インタフェイスとを有している。本発明のこの実施例では、プロセッサは P e n t i u m (登録商標) プロセッサである。クラフト・インタフェイスは、10 / 100 ベースのイーサネットを使用したネットワーク管理インタフェイスである。

10

論理的には、図 8 ~ 10 に示したクラスターは、多数の外部ポートを備えた単一ルータ、すなわち I P 層 3 スイッチとして共同で機能する。物理的には、これらの外部ポートは、複数のシャーシにわたって分布されていることになる。これらのシャーシは隔たった場所に分配されていてよい。I C L は、外部ポートの物理的分布を仮想化する。

論理的には、クラスターは単一エンティティとして管理される (図 19)。I P から見ると、それぞれのシャーシ内の各カードは管理サブエージェントを有することになる。各カードに局限されたサブエージェントは管理のためのカードを意味する。クラスター・マネージャの管理サーバ (マスターエージェント) は、サブエージェントと通信し、またサブエージェントの物理的分布を仮想化することになる。

一次および二次シャーシの双方は、種々の伝送メディアをサポートすることができ、最初は H F C インタフェイスに照準を合わせることで、ケーブルモデムの接続を支援して、インターネット・アクセスおよび I P 電話を可能にする。シャーシはケーブルモデム・インフラストラクチャのための、電話会社をベースとするリターンパス、ならびに、通常の (P S T N) モデムをベースとするインターネット・アクセスをサポートすることができる。

20

クラスター・マネージャ 100 は、W i n d o w s N T のような汎用オペレーティング・システムを走行させる。より具体的なエンティティの集合は、例えば：(1) マスター・エージェント (M A) またはネットワーク・マネージャ、(2) アプリケーション・モジュール、(3) カード・リソースのセンサスを可能にするためのリソース・マネージャ (R M)、(4) クラスターによって機能させられる、I P 末端ユーザの動的ビューを維持するユーザ・マネージャ (U M)、(5) 転送テーブルを作成 / 配信するための転送テーブル・マネージャ (F T)、(6) 一般的な I P フィルタリングのためのフィルタ・マネージャ (F M)、(7) サービス・マネージャの帯域幅 (サービス品質) / クラス、および (8) 負荷分散マネージャ (L S) である。

30

クラスター・マネージャのシステム管理サーバはまた、クラスター化されたシャーシを単一の仮想エンティティとして表す。クラスター・マネージャはまた、ダイヤルイン・モデムのための R A S サーバとして機能することもできる。クラスター・マネージャ上の R A D I U S クライアントは、ネットワーク管理者によって提供された R A D I U S サーバと通信する。

ウェブサーバおよびメールサーバのような多目的機能は、クラスター・マネージャ上で実行させることができる。F T P および T e l n e t は、管理者がクラスター・マネージャにリモート・アクセスすることを可能にする。

40

本発明はリソース共有を可能にする。例えば二次シャーシの H F C インタフェイス内に入来した I P トラヒックは、潜在的に音声をそのペイロードとして有している。このような I P パケットは、V O I P / P S T N カードに交換されることになる。このカードは異なる場所の異なるシャーシに位置してよい。音声ペイロードを備えた I P パケットは、分布された I P 交換ファブリックによって、適切な V O I P / P S T N カードにルーティングされる。原シャーシは、最高「サービスクラス」の優先順位をパケットにタグ付けし、ルート / スイッチ決定が最小限の待ち時間で行われることを可能にする。クラスター・マネージャは、全てのリソース (利用可能な D S の場所を含む) の集中リソース・マップをク

50

ラスタ化されたシステム内に維持する。

シャーシ・コントローラは、管理バスを介してシャーシ内のアプリケーション・モジュールと通信する。シャーシ・コントローラおよびアプリケーション・モジュールはそれぞれ、通信を目的としたエージェントを有している。シャーシ・コントローラはマスターエージェントを有し、アプリケーション・モジュールはサブエージェントを有している（図19）。サブエージェントはシャーシ、アプリケーション・モジュール、およびスロットの情報をそのシャーシのマスターエージェントに通信する。

図19はネットワーク内の複数のシャーシを示す。複数のシャーシを有する環境では、形成された第1のシャーシは一次シャーシであるように指定される。一次シャーシは付加的なマスターエージェントを有している。このマスターエージェントは「マスター・マスター」と呼ばれる。ネットワーク内の全てのシャーシは、マスター・マスターエージェントと通信する。図19に示すように、各シャーシ内のそれぞれのアプリケーション・モジュールは、そのIPアドレス（「10.chassis.slot.port」）をそのシャーシに対応するマスターエージェントおよびマスター・マスターに通信する。IPアドレス中の「10」という数値は、プライベート・ネットワーク・アドレスを表す。

マスター・マスターエージェントはそのシャーシIDを管理バスを介して分配する。一次シャーシのシャーシIDは、一次シャーシが一次シャーシとして指定された後で1にシフトする。ネットワーク上に第2のシャーシが形成されると、その存在を検出するのにリンク検出プロトコル（LDP）が使用される。LDPメッセージが目下のシャーシのリンク毎に送出される。第2のシャーシがリターン信号を受け取ると、シャーシ・コントローラがそれをICLとして識別する。LDPは各シャーシがそのICLリンクを識別するのを可能にする。一次シャーシはそのICLリンクを介して、このシャーシがマスター・マスターエージェントを有することをブロードキャストする。

図21は、メッシュ通信チップ（MCC）のブロック図である。MCCASICは、全二重シリアル・リンク215として示した高速差分対を介して、シャーシ内の他の全てのカードとの接続を可能にする。各差分対は、毎秒1ギガビットよりも高い速度のデータ処理量で動作する。F-バス・インタフェース805がMCC300をFIFOバス（F-バス）に接続する。12個の送信側FIFO810および12個の受信側FIFO815が、F-バス・インタフェース805に接続されている。各送信側FIFOは並直列方式のデータ・コンプレッサ（全部で12個のデータ・コンプレッサ820）を有しており、各受信側FIFOはデータ・エキスパンダ（全部で12個のデータ・エキスパンダ825）を有している。データ・コンプレッサ820およびデータ・エキスパンダ825のために、12個の並直列変換器830が役立つ。各並直列変換器に対して1つのコンプレッサおよび1つのエキスパンダが使用される。MCCのチャンネルは、そのエンコード/デコード論理、送信待ち行列および受信待ち行列と共にシリアル・リンクとして規定される。チャンネルから延びるシリアル・ラインはバックプレーン・メッシュに接続する。全てのチャンネルは同時にデータを送信する。

本発明のこの実施例は、メッシュ通信チップを使用することにより、シリアル・リンク技術を用いて完全メッシュ状に最大13個のF-バスを相互接続する。各MCCは2つのF-バス・インタフェースと、12個のシリアル・リンク・インタフェースとを有している。MCCはF-バス上でパケットを、64バイトからパケットいっぱいまでプログラム可能にサイズを増分して送受する。MCCは12個の仮想送信プロセッサ（VTP）を含んでいる。これらのプロセッサは、F-バスからパケットを取り出し、シリアル・リンクからこれらを送出して、同時に12個のパケットが出発するのを許す。VTPはパケット先頭のMCCタグを読み取り、ヘッダに示された宛先スロットと動的にバインドする。

カード/スロット-特定プロセッサ、カード/スロット-特定MAC/PHYの対（イーサネット、SONET、HFCなど）およびMCCは双方向F-バス（または複数の一方方向F-バス）上で通信する。パケット送信経路は、PHY/MACからプロセッサへ、さらにプロセッサからMCCへ延び、メッシュで終わる。プロセッサは層3および層4の探索をFIPPで行うことにより、パケットの宛先およびサービス品質（QoS）を見極め

10

20

30

40

50

、必要に応じてヘッダを変え、パケットをMCCに送る前にパケットにMCCタグを予め取り付ける。

パケット受信経路は、メッシュからMCCおよびプロセッサへ延び、さらにプロセッサからMAC/Phyへ延び、チャンネルで終わる。プロセッサはパケットをMACに送る前にMCCタグを剥ぎ取る。

図22は、MCCタグとも呼ばれるパケット・タグを示す。MCCタグは32ビット・タグであり、バックプレーン・メッシュを介してパケットをルーティングするのに用いられる。このタグは、パケットをMCCに送る前にスロット・プロセッサによってパケット先頭に加えられる。タグは4つのフィールド、すなわち、宛先マスクフィールド、優先順位フィールド、保持フィールドおよび保留フィールドを有している。宛先マスクフィールドは、それがシステム内の最終宛先であるか否かを問わず、パケットが指定されている目下のシャーシにおけるスロットのマスクを保持するフィールドである。送信パケットのために、MCCは宛先マスクを使用することにより、パケットがどの送信待ち行列に並ぶことになるかを決定する。受信パケットのためには、MCCは優先順位フィールドおよび保持フィールドを使用することにより、過剰割り当てを受けたスロット内でどのパケットを廃棄すべきかを決定する。保留フィールドは本発明のこの実施例では使用されない。

MCCは2つの個別の送信モードセクタ、すなわちスロット・チャンネル間マッピングおよび仮想送信モードとを有している。スロット・チャンネル間マッピングにおいて、MCCはSIDをCIDにトランスペアレントにマッピングし、ソフトウェアはマッピング軌跡を保存する必要はない。仮想送信モードにおいて、MCCはマルチキャスト・パケットを半トランスペアレントに取り扱う。MCCは単一のF-バス流を取り出し、これを複数のチャンネルに案内する。MCCにおける送信ポートは、スロットをアドレスするのではなく、仮想送信プロセッサ(VTP)をアドレスする。F-バス・インタフェースはパケットを、選択された仮想送信プロセッサに案内する。VTPはMCCタグから宛先マスクフィールドを保存し、パケット・データ(MCCタグを含む)を、宛先マスクに示された送信待ち行列のセットに転送する。これに続く全ての64バイトのパケット「チャンク」が、同じポートIDを使用してスロット・プロセッサによって送られ、こうして同じVTPに案内される。VTPは、MCCタグから保存された宛先マスクフィールドに示された送信待ち行列集合に、パケット・チャンクを転送する。チャンクがEOPビット集合と一緒に到達したときには、VTPはその宛先マスクを消去する。このポートにアドレスされた次のチャンクが新しい(すなわちSOPビット集合を有する)パケットのスタートではない場合、VTPはこのチャンクをいかなる待ち行列にも転送しない。

MCCは「チャンネル・ビジー」ビットのセットを維持する。このセットを使用することにより、MCCは、複数のVTPが同じCIDにパケットを同時に送ることを防止する。このような競合防止メカニズムは、ビジー状態のチャンネルを管理してスロット・プロセッサを助けようとするものではなく、むしろ、スロット・プロセッサが偶然同じスロットに同時に2つのパケットを送ってしまった場合に、パケットの完全な崩壊を防ごうとするものである。VTPが新しいパケットを得ると、VTPは宛先CIDマスクとチャンネル・ビジー・ビットとを比較する。ビジー状態のチャンネルがある場合には、そのチャンネルは宛先マスクから除去され、そのCIDに対してエラーが記録される。ついでVTPは残りの全ての宛先チャンネルにビジー・ビット全てをセットし、パケットを送信する。VTPがそのパケットに関してF-バス上でEOPを見た場合には、VTPはその宛先CIDに対するチャンネル・ビジー・ビットを消去する。

F-バス・インタフェースは、MCCと、アプリケーション・モジュールの残りの部分との間でI/O機能を発揮する。アプリケーション・モジュールは、図22に示す32ビット・パケット・タグ(MCCタグ)を、メッシュを介してルーティングされるべき各データパケットに加える。

F-バス上で送受されるデータは最大64ビット幅である。データ送信時には、F-バス・インタフェースは送信データに4状態ビットを加え、68ビット・データ・セグメントにする。F-バス・インタフェースは68ビット・データ・セグメントを、パケット・タ

10

20

30

40

50

グから決定される適切な送信側 F I F O にドロップする。送信側 F I F O からのデータは、関連するデータ・コンプレッサに伝送される。ここで 68 ビット・データ・セグメントは 10 ビット・セグメントに圧縮される。データは次いで、関連する並直列変換器に通され、ここでデータはさらに直列ストリームに圧縮される。直列ストリームはシリアル・リンクからバックプレーンに送出される。

バックプレーンから到達したデータは、シリアル・リンクを通して、関連チャネルに来る。そのチャネルに対応する並直列変換器は、データを 10 ビット・データ・セグメントに伸張し、関連するデータ・エキスパンドはデータを 68 ビットデータ・セグメントに伸張する。このデータ・セグメントは、関連 F I F O を通過させられ、次いで F I F O から F - バス・インタフェイスに送られる。

10

高速 IP プロセッサ (F I P P) は高速シンクロナス S D R A M の 32 / 64 M バイト、高速シンクロナス S R A M の 8 M バイト、およびブート・フラッシュを備えている。F I P P は 32 ビット P C I バスと、64 ビット F I F O バス (F - バス) とを有している。F I P P はパケットデータを全ての F - バス接続デバイスへ、またこのデバースから伝送する。F I P P はユニキャストおよびマルチキャストの両モードでの IP 転送を可能にする。シャーシルート・サーバから管理バスを介して、ルーティング・テーブルが受信される。F I P P はまた、フィルタリングのようなより高い層機能および C o S / Q o S を提供する。

各ライン・カードは、各カードに必要な全てのクロックを生成するクロック・サブシステムを有している。このシステムは、システム・クロックおよび管理バス・アービトレーシ

20

ョン・カードによって提供された基準クロックに関連する。各カードはホットプラグ、パワーオン・リセット回路およびサニティ・タイマ機能を有している。全てのカードは D C - D C コンバータを搭載しており、これによりバックプレーンの - 48 V レールから、その用途に必要ないかなる電圧にも変換することができる。(C M T S カードのような)いくつかのカードは多くの場合、別個の分離した 2 つの電源回路を有し、これによりそのカードのアナログ部分の性能を最大限に活かすことができる。

以上、出願時における本発明の好ましい実施例を説明してきた。本発明の要旨から逸脱することなしに、同等の構成成分および機能を置き換えることができることは明らかである。本発明の利点を維持しながらハードウェアおよびソフトウェアの種々の組み合わせを実施することが可能である。本発明は極めてフレキシブルかつ拡張可能であろうとするものなので、重要なのは、モジュールおよびポートの数よりも、ここに開示したモジュールの協働関係である。

30

【図面の簡単な説明】

【図 1】

マルチプレクサ・ドロップを介して S O N E T リングに接続された H F C ヘッドエンド・サーバを有する従来技術のネットワークを示す図である。

【図 2】

従来技術のケーブル経由の H F C データ供給システムを示す図である。

【図 3】

従来技術のデータ供給網を示す図である。

40

【図 4】

本発明の原理による、シャーシのブロック図である。

【図 5】

図 4 のシャーシを有する統合ケーブル・インフラストラクチャを示す図である。

【図 6】

図 4 のような複数のシャーシを使用する、環状トポロジーのネットワークを示す図である。

【図 7】

本発明の原理に従って動作する一次シャーシと二次シャーシとを有する第 1 の階層状ネッ

50

トワークを示す図である。

【図 8】

図 4 のような複数のシャーシを有する星形ネットワークを示す図である。

【図 9】

図 4 のような複数のシャーシを有する第 2 の階層状ネットワークを示す図である。

【図 10】

図 4 のような複数のシャーシを有するメッシュ・ネットワークを示す図である。

【図 11】

図 4 のシャーシのアプリケーション・カードおよびバックプレーン部分のブロック図である。

10

【図 12】

交換メッシュを含む、バックプレーン相互接続の概略的な図である。

【図 13】

本発明の原理による B A S ヘッダのブロック図である。

【図 14】

ギガビット・イーサネットのシャーシ間リンクまたは出口アプリケーション・モジュールのブロック図である。

【図 15】

A R P テーブルのブロック図である。

【図 16】

シャーシ内の 2 つの出口モジュールのブロック図である。

20

【図 17】

ルート・サーバのブロック図である。

【図 18】

シャーシ・コントローラ・モジュールのブロック図である。

【図 19】

ネットワーク管理構造のブロック図である。

【図 20】

C M T S アプリケーション・モジュールのブロック図である。

【図 21】

バックプレーン・メッシュ・インタフェースのブロック図である。

30

【図 22】

M C C タグのブロック図である。

【図 1】

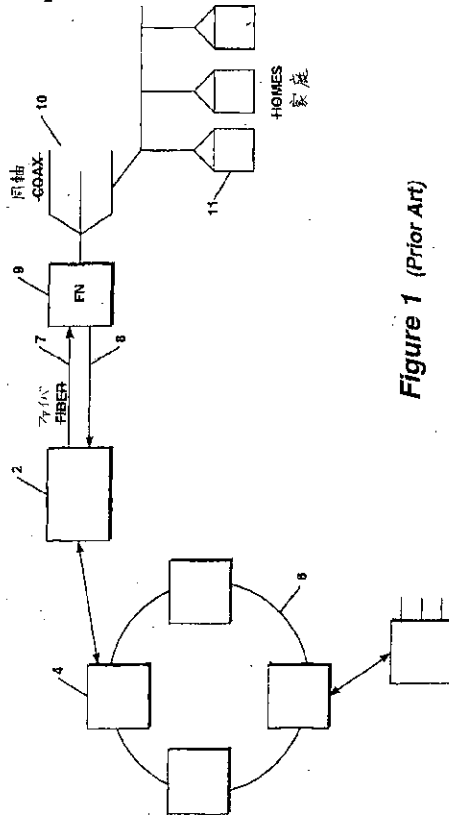


Figure 1 (Prior Art)

【図 2】

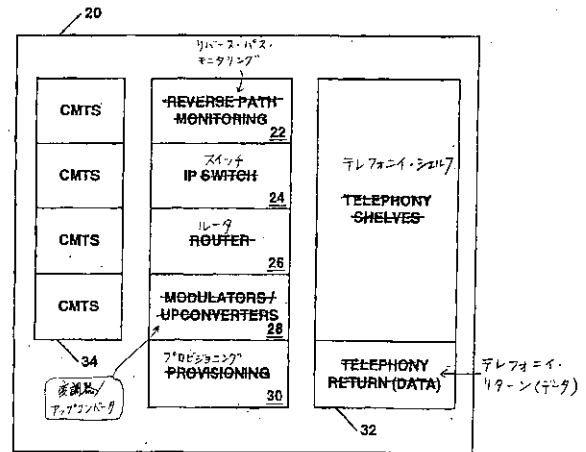


Figure 2 (Prior Art)

【図 3】

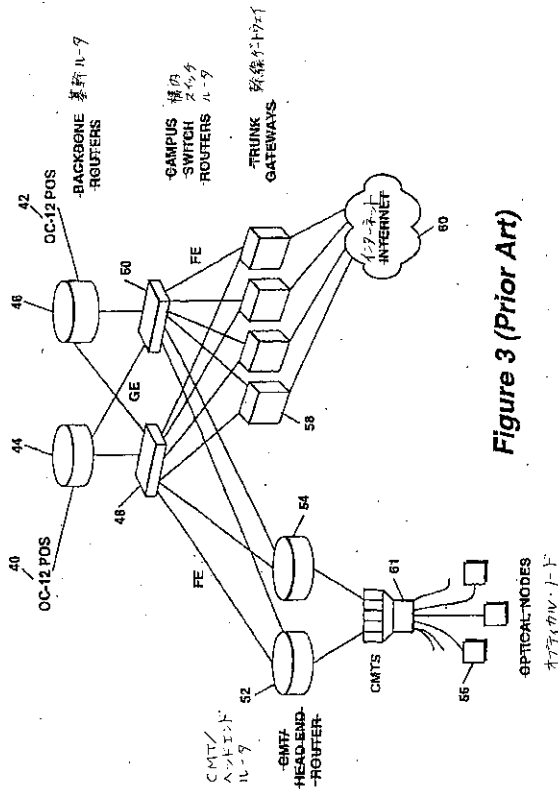


Figure 3 (Prior Art)

【図 4】

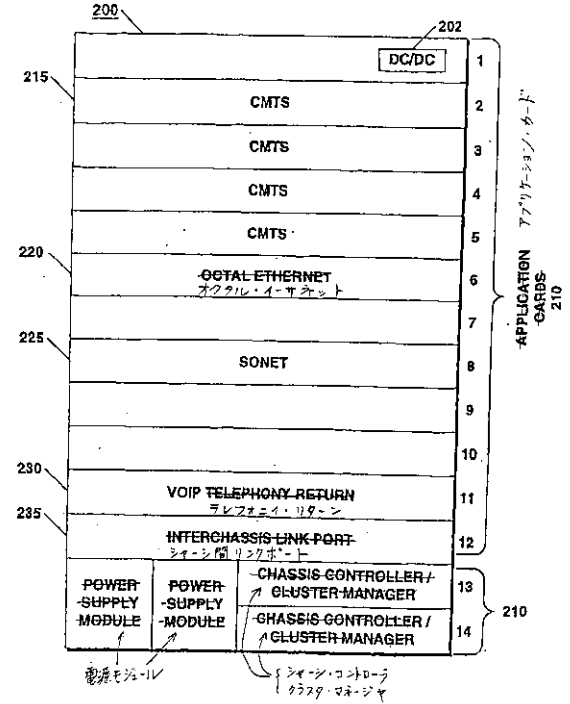


Figure 4

【図 5】

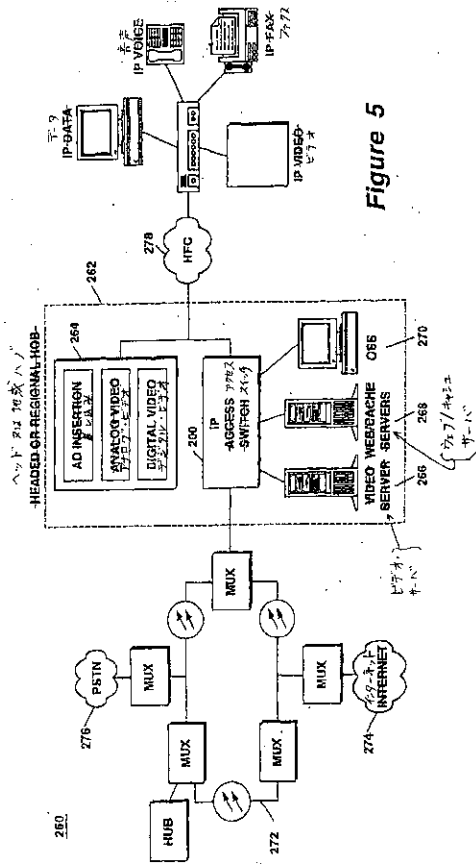


Figure 5

【図 7】

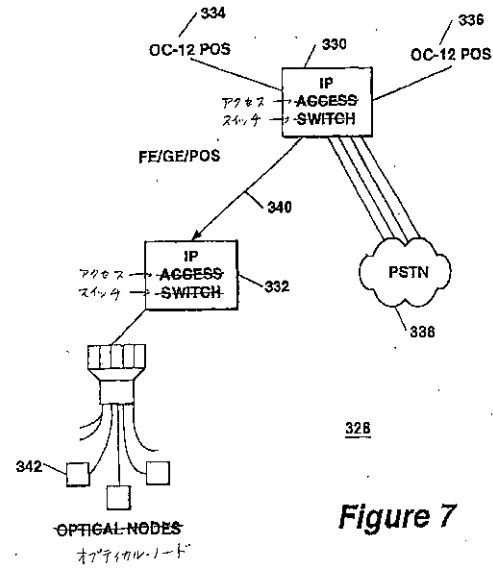


Figure 7

【図 11】

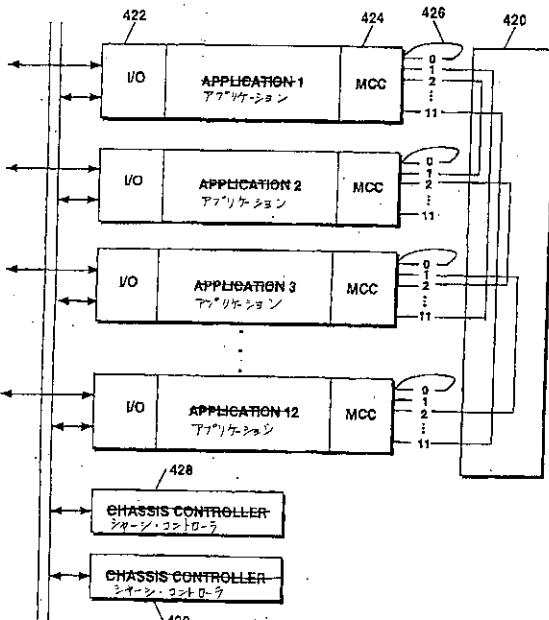


Figure 11

【図 13】

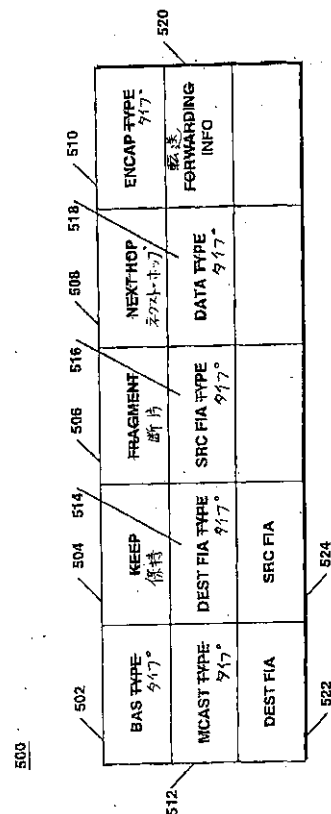


Figure 13

【図 14】

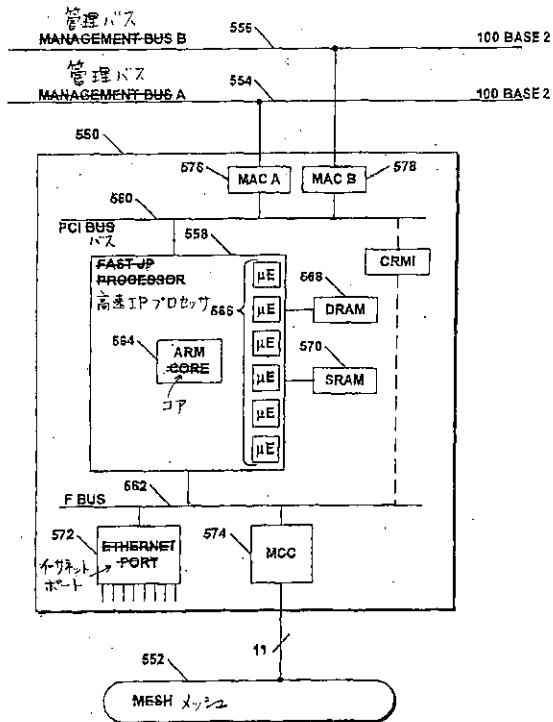


Figure 14

【図 15】

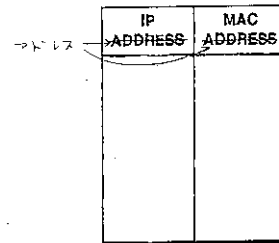


Figure 15

【図 17】

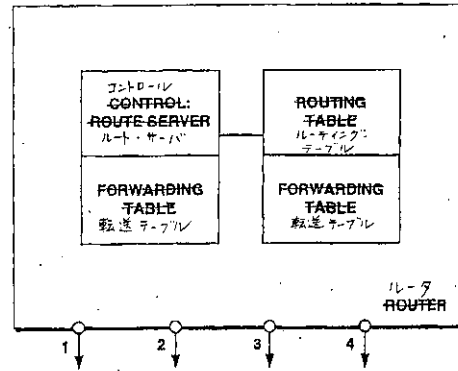


Figure 17

【図 16】

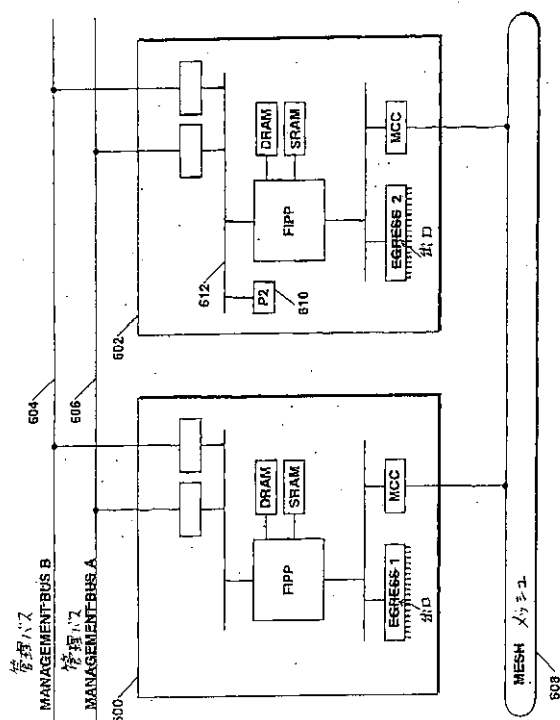


Figure 16

【図 18】

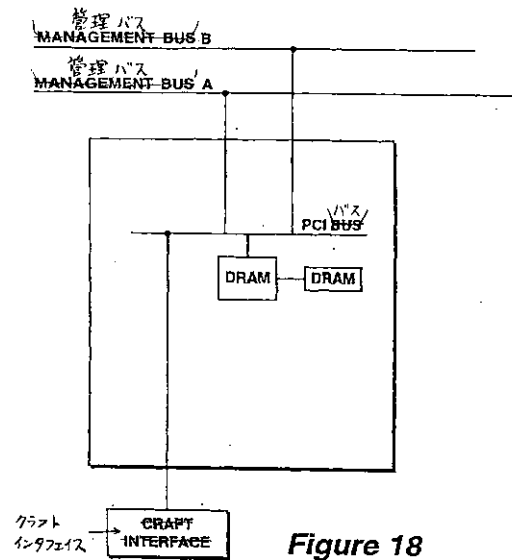


Figure 18

【図 19】

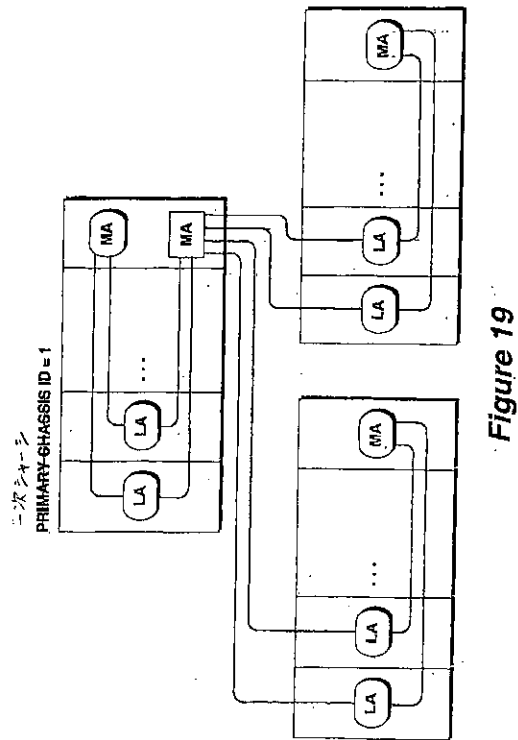


Figure 19

【図 20】

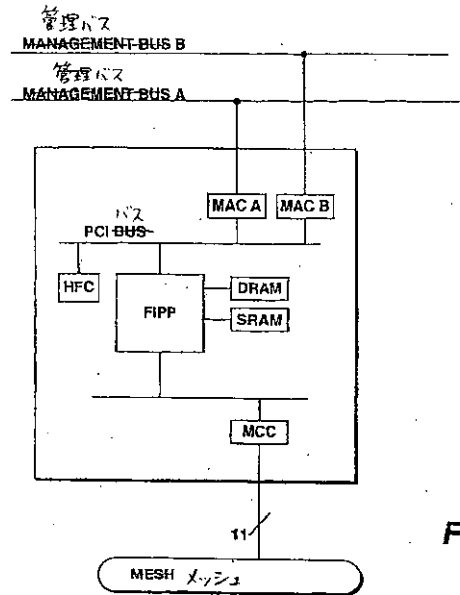


Figure 20

【図 21】

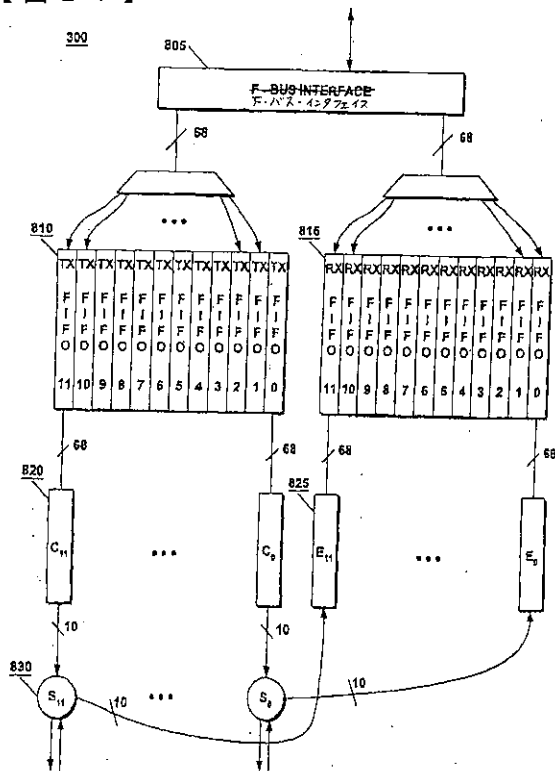


Figure 21

【図 22】

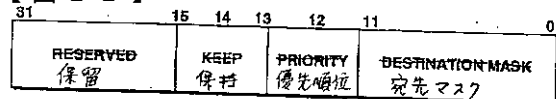


Figure 22

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 July 2001 (05.07.2001)

PCT

(10) International Publication Number
WO 01/48983 A2(51) International Patent Classification: **H04L 12/56**,
12/28

(31) International Application Number: PCT/US00/35490

(32) International Filing Date:
28 December 2000 (28.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/474,030 28 December 1999 (28.12.1999) US(71) Applicant: **BROADBAND ACCESS SYSTEMS, INC.**
(US/US), 8 Technology Drive, Westborough, MA 01752 (US).01752 (US), NGAI, Howard, Broadband Access Systems, Inc., 8 Technology Drive, Westborough, MA 01752 (US);
PAOLINO, David, R., Broadband Access Systems, Inc., 8 Technology Drive, Westborough, MA 01752 (US);
PEARCE, Kirk, R., Broadband Access Systems, Inc., 8 Technology Drive, Westborough, MA 01752 (US);
SARKAR, Jyotirmoy, B., Broadband Access Systems, Inc., 8 Technology Drive, Westborough, MA 01752 (US).

(74) Agents: COHEN, Jerry et al.; Perkins, Smith & Cohen, L.L.P., One Beacon Street, Boston, MA 02108 (US).

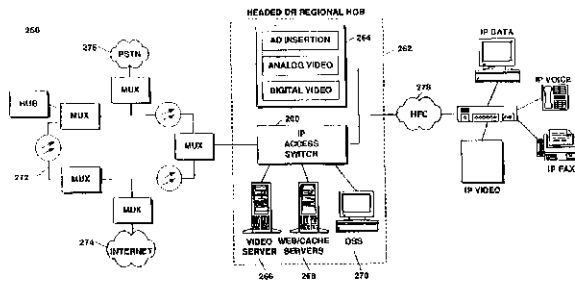
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GR, GM, HK, HU, ID, IL, IN, IS, JP, KR, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW); Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM); European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, MC, NL, PT, SE, TR); GAMP patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, NG, SN, TD, TG).

Published:
— Without international search report and to be republished upon receipt of that report.

[Continued on next page]

(54) Title: SYSTEM AND PROCESS FOR HIGH-AVAILABILITY, DIRECT, FLEXIBLE AND SCALABLE SWITCHING OF DATA PACKETS IN BROADBAND NETWORKS



(57) Abstract: A system and process for directly and flexibly switching connections of data packet flows between nodes of a broadband data processing system and network. The system acts as a single IP switch.

WO 01/48983 A2

WO 01/48983 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/48983

PCT/US00/35490

1

SYSTEM AND PROCESS FOR HIGH-AVAILABILITY, DIRECT, FLEXIBLE AND
SCALABLE SWITCHING OF DATA PACKETS IN BROADBAND NETWORKS

BACKGROUND OF THE INVENTION

The field of the invention is that of communication over a broadband network of data processing systems, for example, a network using a SONET backbone and Hybrid Fiber-Coax(ial cable) ("HFC") access network to connect users to the backbone. A packet communications standard for the HFC environment is DOCSIS (Data Over Cable Service Interface Specification, Document No. SF-RFI-I05-991105) of CableLabs.

Fig. 1 shows a typical configuration of wide-area delivery of data to HFC systems (each connected to 200-2000 households passed (HHP)). The head-end termination system 2 is connected to a SONET ring 4 via a multiplexer drop on the ring 6. The head-end 2 is connected to a fiber node 9 by a fiber downstream connection 7 and a fiber upstream connection 8. The fiber node 9 has coax connectors 10 out to cable modem subscribers (i.e. homes 11). These multiplexers currently cost some \$50,000 in addition to the head-end termination system, and scaling up of service to a community may require new multiplexers and servers.

The failure of a component on the head-end termination system can take an entire "downstream" (from the head-end to the end-user) sub-network (Figure 7) out of communication with the attached networks.

Attempts have been made to integrate systems in order to reduce costs and to ease system management. A current "integrated" data delivery system 20 is shown as functional blocks in Fig. 2. Fig. 2 shows a system 20 having a reverse path monitoring system 22, an IP switch 24, a router 26, modulators and up-converters 28, a provisioning system 30, telephony parts 32, and a plurality of CMTS's 34 (cable modem termination systems). This type of system typically is

WO 01/48983

PCT/US00/35490

2

constructed with multiple systems from multiple vendors, has different management systems, a large footprint, high power requirements and high operating costs.

A typical current network broadband cable network for delivery of voice and data is shown in Fig. 3. Two OC-12 packet over SONET (POS) links 40, 42 are each connected to one of two backbone routers 44, 46 which are in turn networked to two switches 48, 50. The switches 48, 50 are networked to CMTS head-end routers 52, 54. The CMTS head-end routers 52, 54 are connected to a plurality of optical nodes 56 through a CMTS 61. The switches 48, 50 are also connected to a plurality of telephone trunk gateways 58 which are in turn connected to the public switched telephone network (PSTN) 60. As with the "integrated" system shown in Fig. 2, this type of network also typically has multiple vendors for its multiple systems, has different management systems, a large footprint, high power requirements and high operating costs. It remains desirable to have a truly integrated solution to reduce the size of the system, its power needs, and its costs, as well as to ensure greater consistency and increased reliability in data delivery.

It is an object of the present invention to provide an integrated system and a process for broadband delivery of high quality voice, data, and video services.

It is another object of the present invention to provide a system and process for a cable-access platform having high network reliability with the ability to reliably support life-line telephony services and the ability to supply tiered voice and data services.

It is another object of the present invention to provide a system and process for a secure and scalable network switch.

SUMMARY OF THE INVENTION

The problems of providing an integrated network solution for data processing in a broadband network environment are solved by the present invention of a highly integrated

WO 01/48983

PCT/US00/35490

3

carrier-class broadband access system for delivering high quality voice, data and video services.

In the present invention, a process and system are provided in which switching to the HFC conduits is performed in the IP (Internet Protocol) or Network Layer (OSI Layer 3) using combinations of Wide Area Network (WAN) interface cards, HFC interface (Cable Modem Termination) cards, and inter-chassis link (ICL) cards in a fully meshed common chassis, controlled by a cluster manager. The ICLs may stretch over several miles, but the cluster of chassis, controlled by the cluster manager, acts as a single IP switch. The chassis can interface with other broadband access media as well, for example Digital Subscriber Line (DSL), fixed wireless Local Multi-point Distribution Service (LMDS) and Multi-channel Multi-point Distribution Service (MMDS), as well as "fiber to the curb" (FTTC).

The invention offers the flexibility of deploying a variety of local and wide-area configurations, making it suitable for smaller operators requiring a ready path for scaling up. Different media cards may be used for different modes of broadband communication, such as one-way cable with telephone modem return. The invention also allows several levels of redundancy for different network elements, allowing an operator to provide higher availability by having redundancy of particularly stressed elements. The invention supports tiered voice and data systems. The invention provides a reliable, scalable system with a small footprint and low power requirements.

The present invention together with the above and other advantages may best be understood from the following detailed description of the embodiments of the invention illustrated in the drawings, wherein:

DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a prior art network having an HFC head-end server connected to a SONET ring through a multiplexer drop;

WO 01/48983

PCT/US00/35490

4

Fig. 2 shows a prior art HFC data over cable delivery system;

Fig. 3 shows a prior art data delivery network;

Fig. 4 is a block diagram of a chassis according to principles of the invention;

Fig. 5 shows an integrated cable infrastructure having the chassis of Fig. 4;

Fig. 6 shows a network using a plurality of chassis, such as the one of Fig. 4, in a ring topology;

Fig. 7 shows a first hierarchical network having a primary chassis and a secondary chassis operating according to principles of the present invention;

Fig. 8 shows a star network having a plurality of chassis such as the one of Fig. 4;

Fig. 9 shows a second hierarchical network having a plurality of chassis such as the one of Fig. 4;

Fig. 10 shows a meshed network having a plurality of chassis such as the one Fig. 4;

Fig. 11 is a block diagram of the application cards and backplane portions of the chassis of Fig. 4;

Fig. 12 is a schematic diagram of the backplane interconnections, including the switching mesh;

Fig. 13 is a block diagram of a EAS header according to principles of the present invention;

Fig. 14 is a block diagram of a Gigabit Ethernet inter-chassis link or egress application module;

Fig. 15 is a block diagram of an ARP table;

Fig. 16 is a block diagram of a two egress modules in a chassis;

Fig. 17 is a block diagram of a route server;

Fig. 18 is a block diagram of a chassis controller module;

Fig. 19 is a block diagram of the network management architecture;

Fig. 20 is a block diagram of the CMTS application module;

WO 01/48983

PCT/US00/35490

5

Figure 21 is a block diagram of the backplane mesh interface; and

Figure 22 is a block diagram of the MCC tag.

5

DETAILED DESCRIPTION

Fig. 4 shows a chassis 200 operating according to principles of the present invention. The chassis 200 integrates a plurality of network interfaces and applications into a single switch system. The chassis of the invention is a fully-meshed IP-switch with high performance packet forwarding, filtering and QoS/CoS (Quality of Service/Class of Service) capabilities using low-level embedded software controlled by a cluster manager in a chassis controller. The packet forwarding, filtering and QoS/CoS is distributed across application modules (also called "cards") inside the chassis. The operations are performed in each of the modules by a fast IP processor. The chassis controller and cluster manager control the operation and configure each of the modules. Higher-level software resides in the cluster manager, including router server functions (RIPv1, RIPv2, OSPF, etc.), network management (SNMP V1/V2), security, DHCP, LDAP, and remote access software (VPNs, PPTP, L2TP, and PPP), and can be readily modified or upgraded.

In the present embodiment of the invention, the chassis 200 has fourteen (14) slots for modules. Twelve of those fourteen slots hold application modules 205, and two slots hold chassis controller modules 210. Each application module has an on-board DC-DC converter 202 and is "hot-pluggable" into the chassis. The chassis controller modules 210 are for redundant system clock/bus arbitration. The chassis also has redundant power supply modules. The power supply modules and the DC-DC converters comprise a fully distributed power supply in the chassis. Examples of applications that may be integrated in the chassis are a CMTS module 215, an Ethernet module 220, a SONET module 225, and a telephony application 230. Another application may be an inter-chassis link (ICL) port 235 through which the chassis may be linked to another

WO 01/48983

PCT/US00/35490

6

chassis. The ICL port is the only egress port. The ICL port may be implemented using any egress module, e.g., 10/100 Ethernet, IG Ethernet, and Packet-over-SONET (PoS).

Fig. 5 shows an integrated cable infrastructure 260 having the chassis 200 of Fig. 4. The chassis 200 is part of a regional hub 262 for voice and data delivery. The hub 262 includes a video controller application 264, a video server 266, Web/cache servers 268, and an operation support system (OSS) 270, all networked to the chassis 200 acting as an IP access switch. The chassis 200 is connected to a SONET ring 272, outside the hub 262, having an Internet connection 274 and a Public Switched Telephone Network (PSTN) connection 276. The chassis 200 is also connected by an HFC link 278 to cable customers and provides IP-based services including voice, data, video and fax services. Each HFC application module can handle up to 2000 cable modem service subscribers. The logical limit to the number of subscribers is 8192. The chassis can support a plurality of HFC links and also a plurality of chassis may be networked together (as described below) to support over one million cable modem subscribers.

By convention today, there is one wide-band channel (27 - 40 Mbps) for transmission (downloading) to users (which may be desktop computers, facsimile machines or telephone sets) and a plurality of narrower channels (320 Kbps - 10 Mbps) for uploading. This is processed by the HFC cards with duplexing at an O/E node. The local HFC cable system or loop is a coaxial cable distribution network with a drop to a cable modem.

Fig. 6 shows a ring network 300 using a plurality of chassis of Figure 4. The plurality of chassis 302, 304, 306, 308 are connected as a ring over full-duplex inter-chassis links 310, 312, 314, 316. The ring configuration allows chassis 302 and its associated cable networks to communicate with chassis 306 and its associated cable networks through chassis 308 should chassis 304 or the ICLs 310, 312 between them fail.

WO 01/48983

PCT/US00/35490

7

Fig. 7 shows a first hierarchical network 328 having a primary chassis 330 and a secondary chassis 332 operating according to principles of the present invention. The primary chassis is connected to fiber optic communication lines (OC-n) 334, 336. OC-12 lines are shown in the drawing, however OC-3 and OC-48 are also supported. The primary chassis 330 may also be connected to the PSTN 338. The primary 330 and secondary 332 chassis are linked by a full-duplex link 340 that may be a Fast Ethernet (FE), Gigabit Ethernet, or a Packet-over-SONET (PoS) type connection. The secondary chassis 332 is connected to a plurality of optical nodes 342. In networks where there is a primary chassis and one or more secondary chassis, the primary chassis manages the cluster with its cluster manager.

Other configurations are possible, including a "star" configuration 350, as shown in Fig. 8 with a primary chassis 352 and secondary chassis 354, 356, 358, 360, 362, or a tiered arrangement 380, as shown in Fig. 9, with primary chassis 382 and secondary chassis 384 and 386 at the secondary layer and secondary chassis 388, 390, 392, 394 at the tertiary layer. A mesh configuration 400 is shown in Figure 10 with chassis 402, 404, 406, 408, 410 where each chassis has an inter-chassis link to every other chassis. Primary and secondary chassis are determined by system configuration. Each chassis is substantially the same as every other chassis.

The above-described configurations can extend over tens of miles using long-haul fiber optic links, or may be collocated in a single rack at the same master head end site.

Fig. 11 shows application modules connected to a backplane 420 of the chassis 200 of Fig. 4. Each application module 422 interfaces with the backplane 420 through a Mesh Communication Chip (MCC) 424 that will be described more fully below. Each MCC 424 has twelve (12) serial link interfaces 426, eleven that run to the backplane 420. The eleven serial links that run to the backplane on each application module are for connecting the application module to every other

WO 01/48983

PCT/US00/35490

8

application module in the chassis. One link is for connecting the module with itself, i.e., a loop-back. The application modules 422 are connected to chassis controllers 428, 430 over a chassis management bus 432. The second chassis controller 5 430 is optionally used for redundancy in order to make the system more reliable. A second chassis management bus (not shown) is provided in a preferred embodiment also for redundancy/reliability purposes.

The backplane is fully meshed meaning that every 10 application module has a direct point-to-point link to every other application module in the chassis through the serial links. The mesh threads in the mesh backplane each provide a continuous direct channel for communication of data at a rate of 1.5 gigabits per second or greater. Only a portion of the 15 connections 200 are shown in Fig. 11 as an example. The backplane mesh is shown in Fig. 12.

The 12 channels with serial links of the MCC are numbered 0 to 11. The number is referred to as the channel ID or CID. Channels will be described more fully below. The 20 slots on the backplane are also numbered from 0 to 11 (slot ID, or SID). The chassis system does not require, however, that a channel 0 be wired to a slot 0 on the backplane. A serial link may be connected to any slot. The slot IDs are dynamically configured depending on system topology. This 25 allows for freedom in backplane wiring which reduces routing complexity.

Returning to Fig. 11, each application module is also connected to a chassis management bus 432 that provides the modules a connection to the chassis controllers.

30 For switching packets between chassis over the inter-chassis links (ICLs) and for switching packets inside the chassis over the MCC links, the chassis has a inter-chassis switching layer. The inter-chassis switching layer lies below the IP layer 3 (L3, the network layer). Packet processing in 35 the chassis is broken down into two types of traffic: unicast and broadcast/multicast. Switching through the inter-chassis

WO 01/48983

PCT/US00/35490

9

switching layer is accomplished using the inter-chassis header 500 (also called the inter-chassis tag) shown in Fig. 13.

The Bas_Type field 502 has x bits and indicates the inter-chassis header type. This field may be used to indicate that the fabric interface address (FIA) has logical ports, to indicate whether the packet is a broadcast or unicast packet and to indicate relevant fields in the BAS header. The Bas_Type field 502 is for decoding the BAS header and to assist in packet forwarding.

The Keep field 504 has x bits and determines whether the packet may be dropped due to congestion.

The priority field has x bits (not shown in Figure 13) and determines the packet priority and where to queue the packet due to congestion.

The fragment field 506 has x bits and indicates the packet fragmentation and whether the packet consists of two frames.

A next_hop field 508 has a plurality of bits and is used for next hop information for intra-chassis packet transfer. This field is inactive unless the Bas_Type field indicates otherwise.

The Encap_Type field 510 is a one-bit field that indicates whether the packet needs further Layer 2 (data layer, below the inter-chassis switching layer) processing or whether the packet can be forwarded without further processing.

The Mcast_Type field 512 is a one-bit field that indicates whether the packet is broadcast or multicast.

The Dest_FIA_Type field 514 is a one-bit field that indicates whether the destination FIA is provided in short form (<chassis/slot/port>) or long form (<chassis/slot/port/logical port>).

The Src_FIA_Type field 516 is a one-bit field that indicates whether the source FIA field is provided in short form or long form.

WO 01/48983

PCT/US00/35490

10

The Data_Type field 518 has a plurality of bits and is used to indicate the type of traffic being carried in the payload.

5 The TTL field (not shown) has a plurality of bits and is a fail-safe mechanism to prevent packets from staying in the system indeterminately. The TTL is decremented each time it is received at an ICL port. If this field is zero after decrementing, the packet is discarded.

10 The Forwarding Info field 520 has a plurality of bits and contains the forwarding table revision, the forward_table_entry key/id, and the next hop information.

The Destination FIA field 522 has a plurality of bits and indicates the final destination of the packet. This field contains the chassis/slot/port and logical port information.
15 A chassis value of zero has a special meaning, denoting the chassis with a Master Agent (to be described below). A port value of zero also has a special meaning, denoting that the receiver of the packet is an application module. The logical port may be used to indicate which stack/entity in the module is to receive the packet. All edge ports and ICL ports are therefore 1-based, i.e. the port numbers are greater than
20 zero.

The Src FIA field 524 has a plurality of bits and indicates the source of the packet. This field is used by a
25 route server to identify the source of the incoming packets.

According to the system architecture, all application cards that plug into the chassis share common features. The application cards are also called data processors, data processing application module, and data communication module.
30 Fig. 14 is a block diagram of an Ethernet application module 550 in the chassis. The application module is connected to both the mesh backplane 552 and to the management busses A 554 and B 556. In the Ethernet application module 550, a fast IP processor (FIPP) 558 is connected to both a PCI bus 560 and an F-bus 562. The FIPP 558 has an Advanced RISC Machines (ARM) processor core 564 and six micro-engines 566. The FIPP 558 is
35 connected to a DRAM 568 for storing packets and an SRAM 570

WO 01/48983

PCT/US00/35490

11

for storing a routing table. An Ethernet device 572 having 8 ports is connected to the F-bus 562. An MCC 574 is also connected to the F-bus. The MCC 574 provides the connections to the mesh backplane 552. The Ethernet device 572 provides connections to the outside of the chassis. Two MAC devices are connected between the PCI bus 560 and the management buses. MAC A 576 is connected between the PCI bus 560 and management bus A 554. MAC B 578 is connected between the PCI bus 560 and management bus B 556.

When a data packet comes into the Ethernet device 572, it goes through the FIPP 558 and is stored in the DRAM 568. The micro-engines 566 examine the data packets in parallel. The micro-engines 566 look at the IP address in the packet and then look up the destination address in the forwarding table stored in the SRAM 570. The forwarding table provides the chassis, slot and port that the packet will go out on. When the packet comes in over the Ethernet device, the packet has an Ethernet header. If the packet goes out the same slot through which it came, a inter-chassis header is not applied. The FIPP determines the sending port and sends the packet out that port. If the packet is to exit by a different slot, the Ethernet header is removed and a BAS header is added to the data packet. The minimum information in the BAS header is destination data, chassis, slot, and port information. The packet further includes an IP header and a payload. The FIPP has transmit queues where the packets are queued before transmission. The packets are sent over the F-bus 562 to the MCC 574 which sends data out in 64-byte chunks.

When a packet comes in through the MCC 574 of the Ethernet module 550, the application module 550 puts the packet, or pointers to the packet, into a queue in DRAM 570. If the packet is to go out one of the serial ports, the FIPP 558 looks up the destination in an ARP table in the SRAM 570. The ARP table is shown in Fig. 15. The FIPP finds the chassis, slot and port. The chassis address does not change unless the destination is an ICL. In the case of the same chassis, the application module finds the MCC address for the

WO 01/48983

PCT/US00/35490

12

IP packet and sends it out. If the packet is to go to some other chassis, the application module asks if any port is an ICL. If it is, then the application module sends the packet out that port.

5 Fig. 20 shows the CMTS application card, also referred to as a HFC DOCSIS Interface card. HFC-on-PCI technology is currently used because of currently available Broadcom technology, however, in the future other types of interface technology may be used.

10 Fig. 16 shows two application modules 600, 602 in a chassis. The application modules 600, 602 are connected over management busses 604, 606 and the mesh backplane 608. The second application module 602 has a route server 610 attached to the PCI bus 612. The route server for a chassis can reside on any application module. In the present embodiment of the invention, the route server is a Pentium® processor. Each chassis in a network of chassis has a route server. Only one route server in the network, however, may be active at any one time. The route server broadcasts over the management bus and also over the ICLs that it is a route server.

20 In a network of a plurality of chassis, one route server is designated to be the primary route server. All other route servers are secondary. The router servers in the network send out routing information every 30 seconds. The primary route server broadcasts its chassis, slot and port to all other route servers.

Fig. 17 shows conceptually the updating of the route servers by the "control" or primary route server.

30 All forwarding tables know where the route server on their chassis is. The primary route server has logical ports equivalent to Ethernet ports.

When a packet is sent to a route server in one of the chassis networked together, the primary router receives the packet information as though the primary router was receiving the packet itself. The primary router builds a routing table and a forwarding table and broadcasts them out to all application modules and all other chassis on the network.

WO 01/48983

PCT/US00/35490

13

The forwarding table is built from information in the routing table. The forwarding table contains such information as chassis, slot and port of a packet, QoS data, CMTS information, multicast domain information and next-hop ICL information.

Fig. 18 shows a chassis controller connected to the management busses in a chassis. For redundancy, each chassis has two chassis controllers. The chassis controller is not connected to the mesh backplane. The chassis controller has network management tasks and provisioning tasks. The chassis controller enables an entire chassis cluster to appear to be one managed element. The chassis controller has a processor and a memory and a craft interface. In the present embodiment of the invention, the processor is a Pentium® processor. The craft interface is a network management interface using 10/100 based Ethernet.

Logically, the clusters shown in Figs. 8-10 collectively function as a single router, that is, an IP-Layer-3 switch, with a number of external ports. Physically these external ports would be distributed across multiple chassis, which may be distributed across remote locations. The ICLs "virtualize" the physical distribution of the external points.

Logically the cluster is managed as a single entity (Figure 19). From the view of IP, each card in every chassis would have a management subagent. The subagent local to each card would represent the card for management purposes. The cluster manager's management server (master agent) would communicate with the sub agents, again "virtualizing" the physical distribution of the subagents.

Both the primary and secondary chassis are capable of supporting various transmission media, focusing initially on HFC interfaces to support the connection of cable modems to provide Internet access and IP telephony. The chassis can support Telco-based return paths for some cable modem infrastructures, as well as regular (PSTN) modem-based Internet access.

WO 01/48983

PCT/US00/35490

14

The cluster manager 100 runs a general purpose operating system such as Windows NT. It is a set of more specific entities, including: (1) a Master Agent (MA) or network manager; (2) an application module; (3) a Resource Manager (RM) for providing a census of card resources; (4) a User Manager (UM) maintaining a dynamic view of IP end users served by the cluster; (5) a Forwarding Table Manager (FT) for creating/distributing the forwarding table; (6) a Filter Manager (FM) for generic IP filtering, (7) a Bandwidth (Quality of Service)/Class of Service Manager; and (8) a Load Sharing Manager (LS).

The System Management Server on the cluster manager also represents the clustered chassis as a single virtual entity. The cluster manager can also function as an RAS Server for dial-in modems. A RADIUS client on the cluster manager communicates with a RADIUS server provided by the network administrator.

General purpose functions like Web Server and Mail Server can be run on the cluster manager. FTP and Telnet allows the administrator to access the cluster manager remotely.

The invention allows resource sharing. For example, IP traffic coming into an HFC interface on a secondary chassis potentially has voice as its payload. This IP packet would be switched to a VOIP/PSTN card, which may be on a different chassis in a different location. The IP packet with the voice payload is routed to the appropriate VOIP/PSTN card by the distributed IP switch fabric. The originating chassis tags the packet with the highest "class of service" priority, allowing the route/switch decision to be performed with minimal latency. The cluster manager keeps a centralized resource map of all the resources (including the location of DSOs available) in the clustered system.

The chassis controller communicates with the application modules in the chassis over the management busses. The chassis controller and the application modules each have an agent for communication purposes. The chassis controller has a master agent and the application modules have subagents

WO 01/48983

PCT/US00/35490

15

(Figure 19). The subagents communicate their chassis, application module, and slot information to the master agent in that chassis.

Fig. 19 shows a plurality of chassis in a network. In a multi-chassis environment, the first chassis brought up is designated to be the primary chassis. The primary chassis has an additional master agent, referred to as the "master-master". All the chassis in the network communicate with the master-master agent. Each application module in every chassis communicates its IP address ("10.chassis.slot.port") to both the master agent for its chassis and the master-master as shown in Fig. 19. The number "10" in the IP address signifies a private network address.

The master-master agent distributes its chassis ID over the management bus. The chassis ID of the primary chassis shifts to one after the primary chassis is designated as the primary chassis. When a second chassis is brought up on the network, link detection protocol (LDP) is used to determine its presence. An LDP message is sent out every link of the current chassis. If the second chassis receives a return signal, the chassis controller identifies that as an ICL. The LDP enables each chassis to identify its ICL links. The primary chassis broadcasts over its ICL links that it has the master master agent.

Fig. 21 is a block diagram of the Mesh Communication Chip (MCC). The MCC ASIC provides connectivity to all other cards in the chassis via high speed differential pairs shown as fully duplexed serial links 215. Each differential pair operates at greater than one gigabit per second data throughput. An F-bus interface 805 connects the MCC 300 to the FIFO bus (F-bus). Twelve transmit FIFOs 810 and 12 receive FIFOs 815 are connected to the F-bus interface 805. Each transmit FIFO has a parallel to serial data compressor (12 data compressors in all, 820), and each receive FIFO has a data expander (12 data expanders in all, 825). Twelve serializers 830 serve the data compressors 820 and data expanders 825, one compressor and one expander for each

WO 01/48983

PCT/US00/35490

16

serializer. A channel in the MCC is defined as a serial link together with its encoding/decoding logic, transmit queue and receive queue. The serial lines running from the channels connect to the backplane mesh. All the channels can transmit data at the same time.

A current implementation of the invention uses a Mesh Communication Chip to interconnect up to thirteen F-buses in a full mesh using serial link technology. Each MCC has two F-bus interfaces and twelve serial link interfaces. The MCC transmits and receives packets on the F-buses in programmable size increments from 64 bytes to entire packets. It contains twelve virtual transmit processors (VTPs) which take packets from the F-bus and send them out the serial links, allowing twelve outgoing packets simultaneously. The VTPs read the MCC tag on the front of the packet and dynamically bind themselves to the destination slot(s) indicated in the header.

The card/slot-specific processor, card/slot-specific MAC/PHY pair (Ethernet, SONET, HFC, etc.) and an MCC communicate on a bi-directional F-bus (or multiple unidirectional F-buses). The packet transmit path is from the PHY/MAC to the processor, then from the processor to the MCC and out the mesh. The processor does Layer 3 and Layer 4 look-ups in the FIFP to determine the packet's destination and Quality of Service (QoS), modifies the header as necessary, and prepends the MCC tag to the packet before sending it to the MCC.

The packet receive path is from the mesh to the MCC and on to the processor, then from the processor to the MAC/PHY and out the channel. The processor strips off the MCC tag before sending the packet on to the MAC.

Fig. 22 shows a packet tag, also called the MCC tag. The MCC tag is a 32-bit tag used to route a packet through the backplane mesh. The tag is added to the front of the packet by the slot processor before sending it to the MCC. The tag has four fields: a destination mask field, a priority field, a keep field, and a reserved field. The destination mask field is the field holding the mask of slots in the current chassis

WO 01/48983

PCT/US00/35490

17

to which the packet is destined, which may or may not be the final destination in the system. For a transmit packet, the MCC uses the destination mask to determine which transmit queue(s) the packet is destined for. For a receive packet, the MCC uses the priority and keep fields to determine which packets to discard in an over-committed slot. The reserved field is unused in the current embodiment of the invention.

The MCC has two independent transmit mode selectors, slot-to-channel mapping and virtual transmit mode. In slot-to-channel mapping, the MCC transparently maps SIDs to CIDs and software does not have to keep track of the mapping. In virtual transmit mode, the MCC handles multicast packets semi-transparently. The MCC takes a single F-bus stream and directs it to multiple channels. The transmit ports in the MCC address virtual transmit processors (VTPs) rather than slots. The F-bus interface directs the packet to the selected virtual transmit processor. The VTP saves the Destination Mask field from the MCC tag and forwards the packet data (including the MCC tag) to the set of transmit queues indicated in the Destination Mask. All subsequent 64-byte "chunks" of the packet are sent by the slot processor using the same port ID, and so are directed to the same VTP. The VTP forwards chunks of the packet to the set of transmit queues indicated in the Destination Mask field saved from the MCC tag. When a chunk arrives with the SOP bit set, the VTP clears its destination mask. If the next chunk addressed to that port is not the start of a new packet (i.e., with the SOP bit set), the VTP does not forward the chunk to any queue.

The MCC maintains a set of "channel busy" bits which it uses to prevent multiple VTPs from sending packets to the same CID simultaneously. This conflict prevention mechanism is not intended to assist the slot processor in management of busy channels, but rather to prevent complete corruption of packets in the event that the slot processor accidentally sends two packets to the same slot simultaneously. When the VTPs get a new packet, they compare the destination CID mask with the channel busy bits. If any channel is busy, it is removed from

WO 01/48983

PCT/US00/35490

18

the destination mask and an error is recorded for that CID. The VTP then sets all the busy bits for all remaining destination channels and transmits the packet. When the VTP sees EOP on the F-bus for the packet, it clears the channel busy bits for its destination CIDs.

5 The F-bus interface performs the I/O functions between the MCC and the remaining portion of the application module. The application module adds a 32-bit packet tag (MCC tag), shown in Figure 22, to each data packet to be routed through the mesh.

10 The data received or transmitted on the F-bus is up to 64 bits wide. In data transmission, the F-bus interface adds 4 status bits to the transmit data to make a 68-bit data segment. The F-bus interface drops the 68-bit data segment into the appropriate transmit FIFO as determined from the packet tag. The data from a transmit FIFO is transferred to the associated data compressor where the 68-bit data segment is reduced to 10-bit segments. The data is then passed to the associated serializer where the data is further reduced to a serial stream. The serial stream is sent out the serial link to the backplane.

15 Data arriving from the backplane comes through a serial link to the associated channel. The serializer for that channel expands the data to a 10-bit data segment and the associated data expander expands the data to a 68-bit data segment which is passed on to the related FIFO and then from the FIFO to the F-bus interface.

20 A Fast IP Processor (FIPP) is provided with 32/64 Mbytes of high-speed synchronous SDRAM, 8 Mbytes of high-speed synchronous SRAM, and boot flash. The FIPP has a 32-bit PCI bus and a 64-bit FIFO bus (F-bus). The FIPP transfers packet data to and from all F-bus-connected devices. It provides IP forwarding in both unicast and multicast mode. Routing tables are received over the management bus from the chassis route server. The FIPP also provides higher layer functions such as filtering, and CoS/QoS.

WO 01/48983

PCT/US00/35490

19

Each line card has a clock subsystem that produces all the clocks necessary for each card. This will lock to the reference clock provided by the System Clock and Management Bus Arbitration Card.

5 Each card has hot-plug, power-on reset circuitry, and Sanity Timer functions. All cards have on-board DC-to-DC converters to go from the -48V rail in the backplane to whatever voltages are required for the application. Some cards (such as the CMFS card) likely will have two separate
10 and isolated supplies to maximize the performance of the analog portions of the card.

The above discussion describes the preferred embodiment of the invention(s) at the time of filing. It should be clear that equivalent components and functions may be substituted
15 without departing from the substance of the invention(s). Various mixes of hardware and software implementation are possible while retaining the benefits of the invention(s). Because the invention is intended to be highly flexible and scalable, it is the cooperation of the modules here disclosed
20 that is important, rather than the number of modules and ports.

WO 01/48983

PCT/US00/35490

20

CLAIMS

What is claimed is:

- 5
1. Apparatus for connecting data distribution sub-networks, said apparatus comprising:
- (a) at least one chassis with multiple identical interfaces to a mesh having mesh threads providing a continuous direct channel for communication of data between each pair of said interfaces; and
- 10
- (b) at least one data processing module with at least one data port adapted for connecting to data distribution sub-networks, said module engaging one of said interfaces of said chassis and responsive to information in data received through one of said ports to route said data to selected ones of said mesh threads to which said module is interfaced.
- 15
2. The apparatus of Claim 1 wherein said routing of said data is determined by a tag appended to said data indicating one or more destination interfaces selected among said identical interfaces.
- 20
3. The apparatus of Claim 1 further comprising a process controller module logically connected to said data processor module.
- 25
4. The apparatus of Claim 1 further comprising:
- (c) at least one inter-chassis data link;
- (d) at least one secondary chassis identical to the chassis recited in Claim 1; and
- 30
- (e) at least one data processing module with at least one data port adapted for connecting to data distribution sub-networks, said module engaging one of said interfaces of said secondary chassis and responsive to information in data received from said inter-chassis data link
- 35
- to route said data to a selected data distribution port.

WO 01/48983

PCT/US00/35490

21

5. The apparatus of Claim 4 wherein said inter-chassis data link comprises:

(c1) a data processing module with at least one data port adapted for connecting to an inter-chassis data link medium, said module engaging one of said interfaces of said primary chassis and responsive to information in data received through said interface to route said data to said inter-chassis data port;

10 (c2) an inter-chassis data link medium connected at one end to said inter-chassis data port; and

(c3) a data processing module with at least one data port adapted for connecting to an inter-chassis data link medium and connected to the other end of said inter-chassis data link medium, said module engaging one of said interfaces of said secondary chassis and responsive to information in data received through said inter-chassis data port to route said data to one of said mesh channels to which said module is interfaced.

6. The apparatus of Claim 4 further comprising a process controller module logically connected to each data processing module.

7. The apparatus of Claim 6 wherein said process controller module is logically connected through said inter-chassis data link to a data processing module engaging said secondary chassis module.

8. Apparatus for connecting data distribution sub-networks, said apparatus comprising:

(a) at least two chassis each containing at least one processor providing a route server; and

30 (b) an inter-chassis link; wherein said chassis provide ports to said data distribution sub-networks and one said route server acts as a master route server and communicates with each other route server such that said chassis act as a single Layer 3 switch.

35 9. The apparatus of Claim 8 wherein data received through one of said ports in any one of said chassis is assigned a switch-internal tag to specify an appropriate one of said

WO 01/48983

PCT/US00/35490

22

ports for forwarding to a destination in said data distribution sub-network.

10. The apparatus of Claim 8 wherein said master route server provides routing tables for each other route server.

5 11. The apparatus of Claim 8 wherein at least one of said chassis comprises multiple processors each switchably connected to one another and each providing one or more of said ports and each logically connected to said inter-chassis link.

10 12. The apparatus of Claim 11 wherein data received through one of said ports of, or said inter-chassis link to, said multiple processor chassis is assigned an chassis-internal tag to specify an appropriate one of said ports of, or said inter-chassis link to, said multiple processor chassis for
15 forwarding.

13. A chassis having multiple identical slots for holding data processing application modules, each slot having identical interfaces to a back-plane mesh having mesh threads each providing a continuous direct channel for communication
20 of data between each pair of said interfaces, said mesh threads capable of throughput commensurate with communication of digital data packets at a rate of 1.5 gigabits per second or greater.

14. The chassis of Claim 13 further comprising a management control signal bus and wherein said identical interfaces
25 interface with said bus.

15. The chassis of Claim 13 further comprising a telecommunication bus.

16. The chassis of Claim 13 providing connection to external
30 data ports for said modules at the back-plane side of said chassis.

17. The chassis of Claim 13 further comprising a fully distributed DC power supply for said modules.

18. The chassis of Claim 13 wherein the number of said
35 identical slots is greater than four.

WO 01/48983

PCT/US00/35490

23

19. A data processing application module with at least one data port adapted for connecting to data distribution sub-networks, said module further adapted for insertion into a slot in a chassis and thereby engaging an interface associated with said slot, wherein said interface includes multiple channels each capable of communication of digital data packets at a rate of 1.5 gigabits per second or greater, and wherein said module is responsive to information in data received through said data port to route said data to selected ones of said multiple channels.
20. The module of Claim 19 wherein said routing of said data is determined by a tag appended by said module to said data indicating one or more slot-interfaces in said chassis to which said data ultimately is to be routed, said module capable of selecting an alternate route if a first route fails and of rerouting received data having an appended tag indicating an ultimate slot-interface other than that of said module.
21. The module of Claim 19 wherein said module responds to said received data examined in parallel blocks and transmits said data serially on said channel.
22. The module of Claim 19 wherein said module comprises a separate transceiver dedicated to each said channel.
23. A data processing application module with at least one data port adapted for connecting to data distribution sub-networks, said module having a route server adapted for synchronization with a master route server such that said module combined with an identical module can act as a single Layer 3 switch.
24. The module of Claim 23 wherein data received through one of said data ports is assigned a switch-internal tag to specify an appropriate port for forwarding to a destination data distribution sub-network.
25. The module of Claim 23 adapted for mounting in a single chassis with substantially identical data processing modules wherein data received through one of said data ports is assigned an chassis-internal tag to specify an appropriate

WO 01/48983

PCT/US00/35490

24

ports for forwarding to a destination data distribution sub-network.

26. A process for connecting users on data distribution sub-networks, each connected to a data processing application module installed in one or more chassis, said process comprising the steps of:

A) receiving data on a module installed in a chassis and connected to one data distribution sub-network;

10 B) determining from header information in said received data whether the destination of said received data is on a data distribution sub-network i) connected to a module installed in said chassis, ii) connected to a module installed in another chassis with an inter-chassis data link to said chassis, or iii) connected on the Internet or similar network accessible through a module on said chassis or another chassis with an inter-chassis data link to said chassis;

15 C) appending a header-tag to said received data to route said data within said chassis; and

20 D) transmitting said data without said header-tag from a module installed in said chassis.

27. A process for switching data flow among data communication modules interconnected within a chassis, said process comprising the steps of:

A) receiving said data at one said module;

25 B) applying said data m bits at a time to an m-bit-wide parallel bus in said module;

C) storing said data in m-bit parallel queues;

D) determining which module is the appropriate destination module;

30 E) appending a tag to said data to indicate to which destination module said data is to be routed;

F) applying said data m bits at a time to said m-bit-wide parallel bus;

35 G) gating said m bits of data into an intra-chassis channel associated with a path to said destination module;

H) serializing said m bits of data;

WO 01/48983

PCT/US00/35490

25

- I) applying said serialized m bits of data to a high-speed serial link dedicated to said channel associated with said path;
- J) applying said data m bits at a time to an m-bit-wide parallel bus in said destination module;
- 5 K) storing said data in m-bit parallel queues;
- L) applying said data m bits at a time to said m-bit-wide parallel bus in said destination module; and
- M) transmitting said data from said destination module.
- 10 28. A process for connecting data distribution sub-networks ported to multiple chassis, said process comprising the steps of:
- a) synchronizing route servers for each of said multiple chassis;
- 15 b) appending a tag to data received at a port of one of said chassis designating a port for forwarding according to information in said data and the synchronized route server for said chassis;
- c) forwarding said data to said port designated by said tag.
- 20 29. Apparatus for connecting users on data distribution sub-networks comprising a data processing application module adapted for installation in a chassis and to be logically connected to one or more substantially similar modules
- 25 installed in the same or other chassis, said module comprising:
- A) means for receiving data;
- B) means for determining from header information in said received data whether the destination of said received data is
- 30 on a data distribution sub-network i) connected to a module installed in said chassis, ii) connected to a module installed in another chassis with an inter-chassis data link to said chassis, or iii) connected on the Internet or similar network accessible through a module on said chassis or another chassis
- 35 with an inter-chassis data link to said chassis;
- C) means for appending a header-tag to said received data to route said data within said chassis; and

WO 01/48983

PCT/US00/35490

26

- D) means for transmitting said data without said header-tag from a module installed in said chassis.
30. Apparatus comprising a data communication module for switching data flow among substantially similar modules interconnected within a chassis, said module comprising:
- 5 A) means for receiving said data;
- B) means for applying said data m bits at a time to an m -bit-wide parallel bus;
- C) means for storing said data in m -bit parallel queues;
- 10 D) means for determining which module is the appropriate destination module;
- E) means for appending a tag to said data to indicate to which destination module said data is to be routed;
- F) means for applying said data m bits at a time to said
- 15 m -bit-wide parallel bus;
- G) means for gating said m bits of data into an intra-chassis channel associated with a path to said destination module;
- H) means for serializing said m bits of data; and
- 20 I) means for applying said serialized m bits of data to a high-speed serial link dedicated to said channel associated with said path.
31. Apparatus for connecting data distribution sub-networks ported to multiple chassis, said apparatus comprising:
- 25 a) at least one data processor installed in each of said chassis;
- b) at least one route server at each of said chassis;
- c) means for synchronizing each of said route servers;
- d) means for appending a tag to data received at a port
- 30 of one of said chassis designating a port for forwarding according to information in said data and the synchronized route server for said chassis; and
- e) means for forwarding said data to said port designated by said tag.

WO 01/48983

PCT/US00/35490

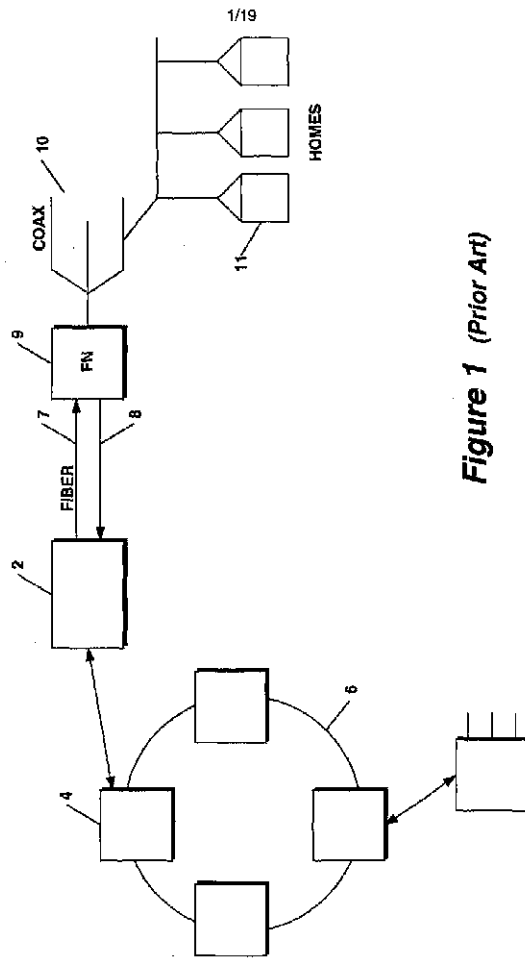


Figure 1 (Prior Art)

SUBSTITUTE SHEET (RULE 26)

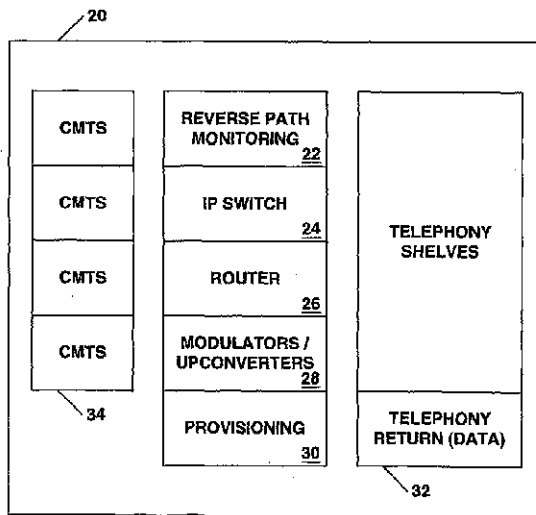


Figure 2 (Prior Art)

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

3/19

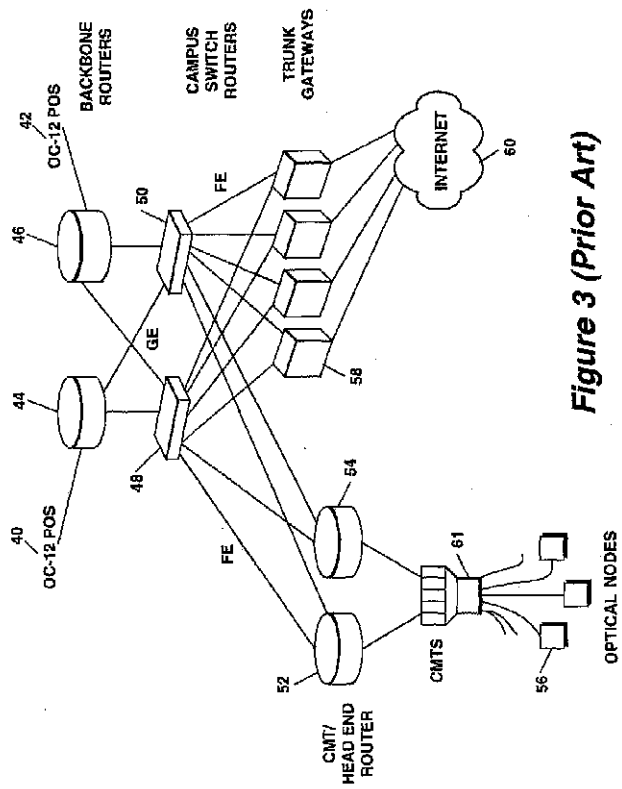


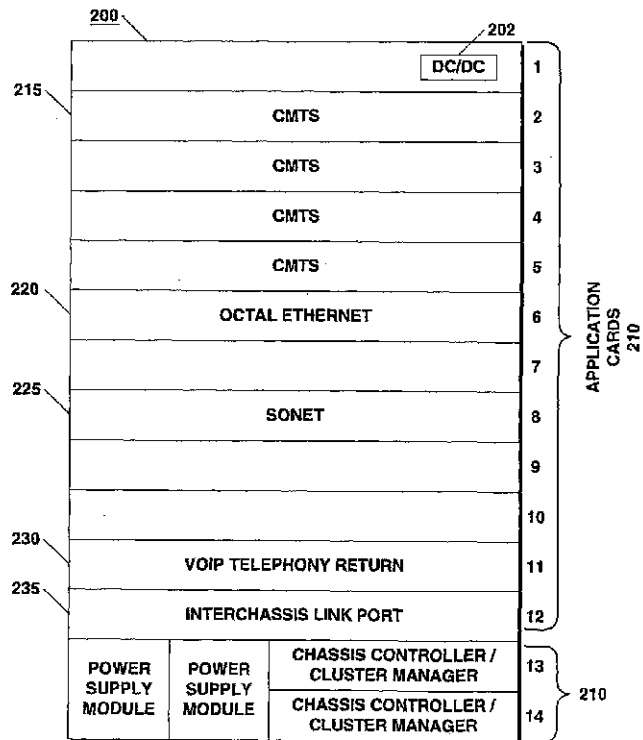
Figure 3 (Prior Art)

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

4/19

**Figure 4**

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

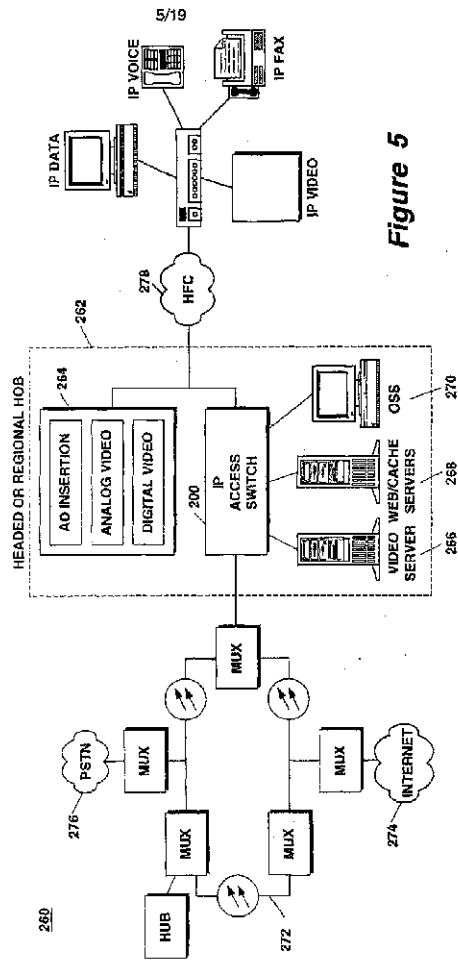
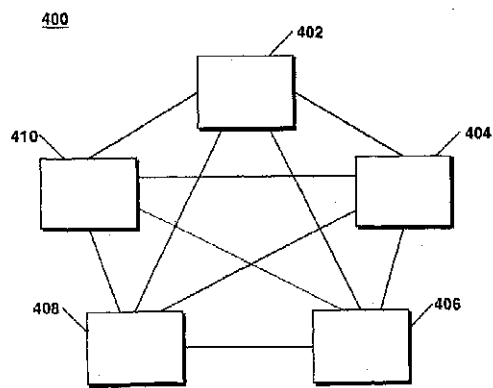
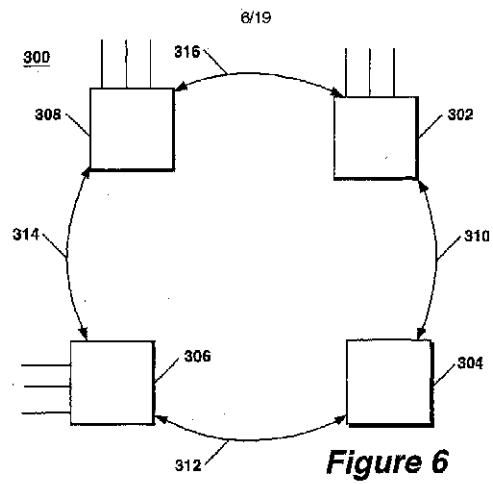


Figure 5

WO 01/48983

PCT/US00/35490

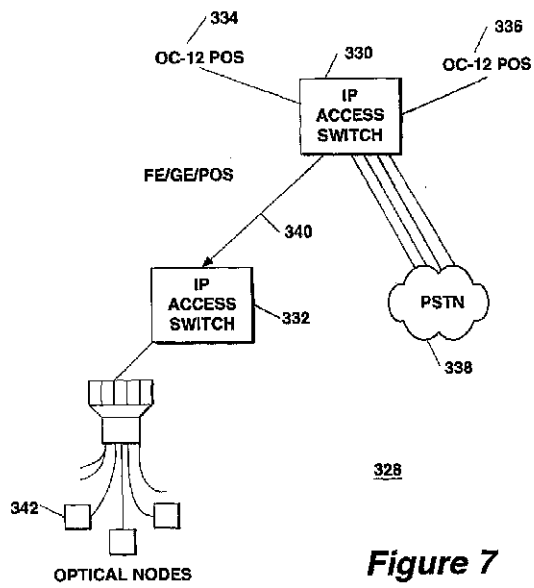
**Figure 10**

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

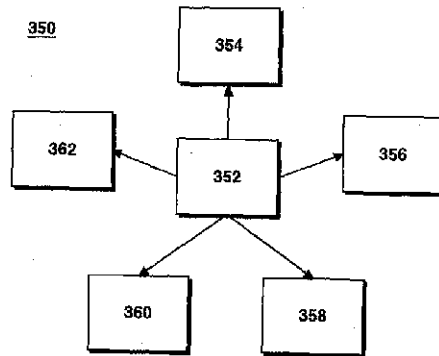
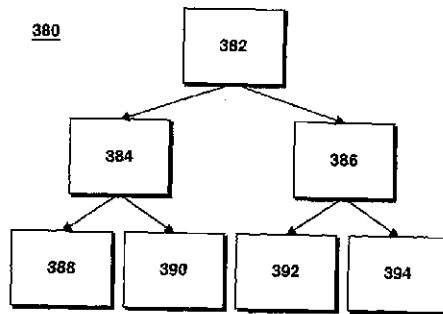
7/19

**Figure 7**

WO 01/48983

PCT/US00/35490

8/19

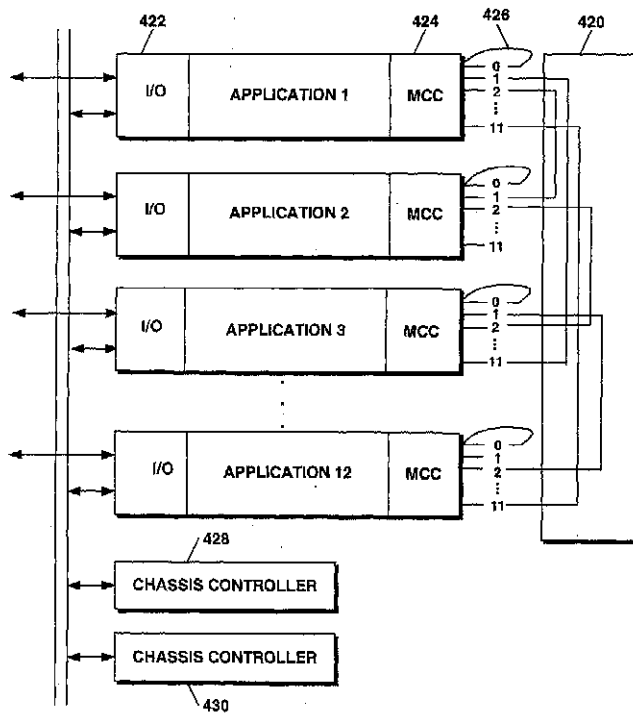
**Figure 8****Figure 9**

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

9/19

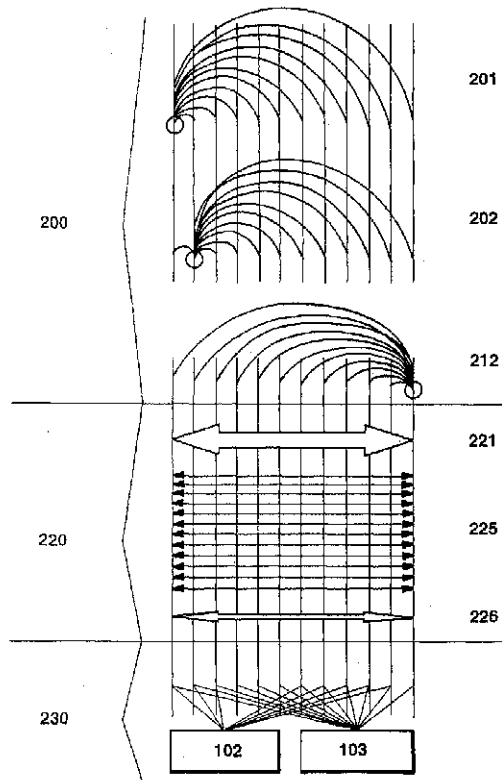
**Figure 11**

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

10/19

**Figure 12**

SUBSTITUTE SHEET (RULE 26)

11/19

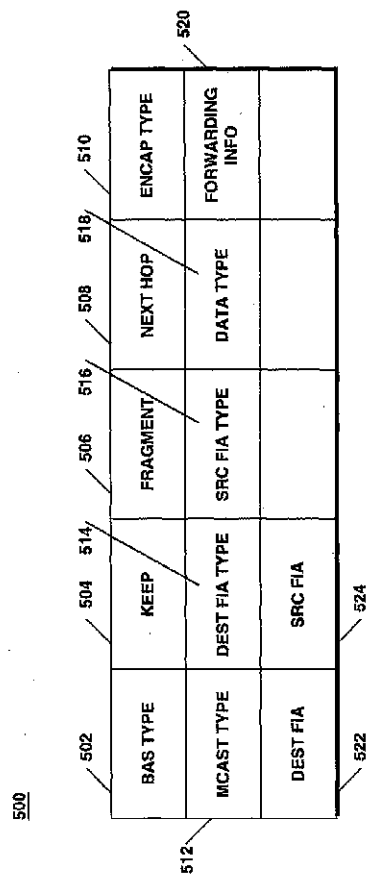
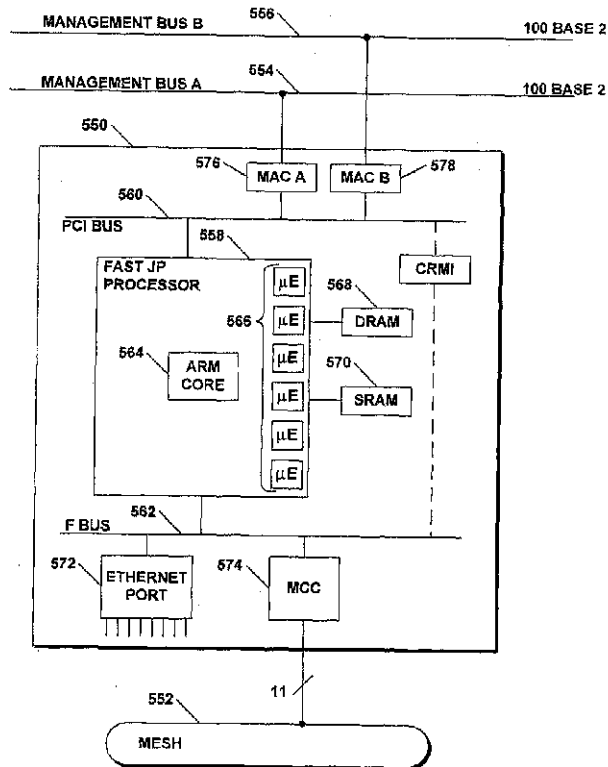


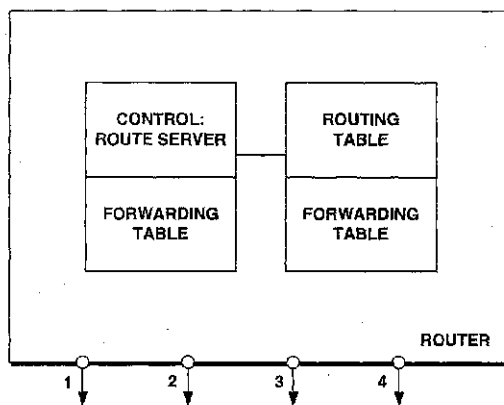
Figure 13

12/19

**Figure 14**

SUBSTITUTE SHEET (RULE 26)

| IP ADDRESS | MAC ADDRESS |
|---------------|----------------|
| | |

Figure 15**Figure 17**

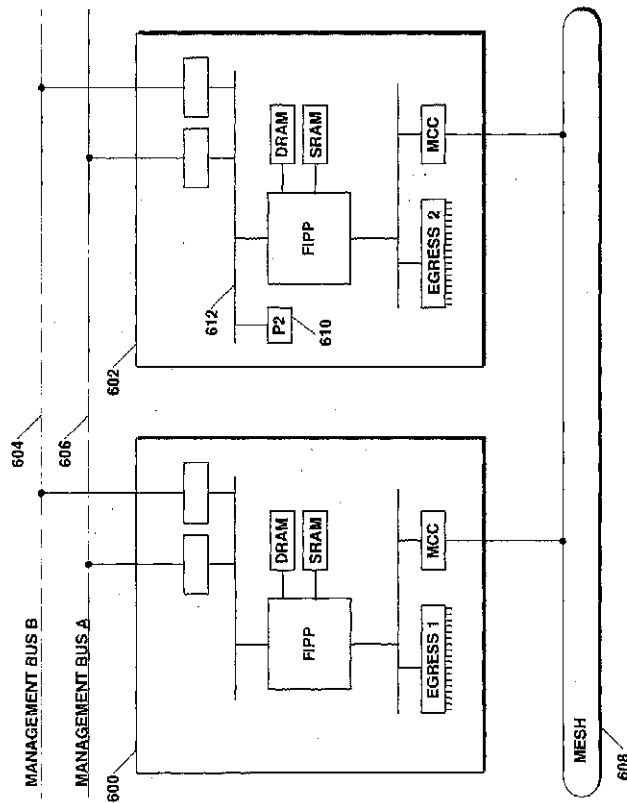


Figure 16

WO 01/48983

PCT/US00/35490

15/19

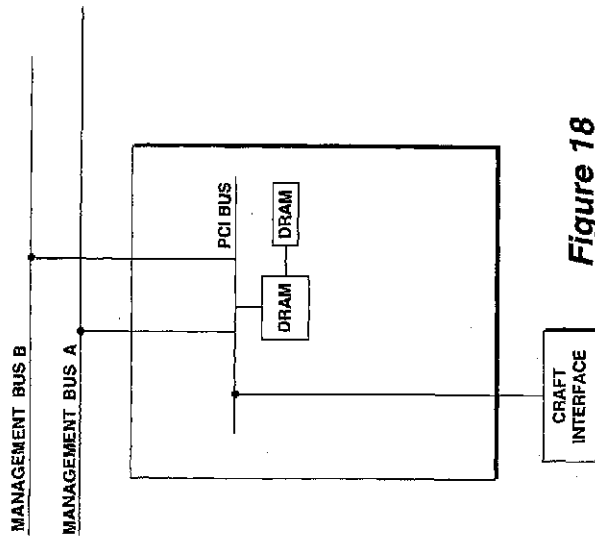


Figure 18

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

16/19

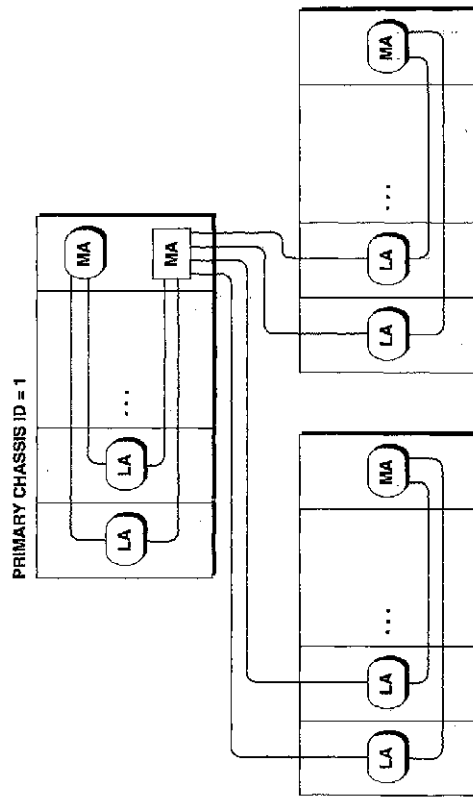


Figure 19

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

17/19

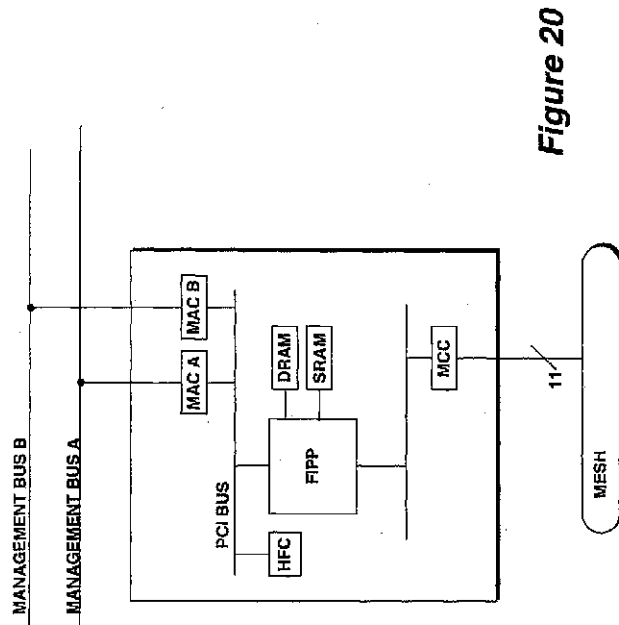


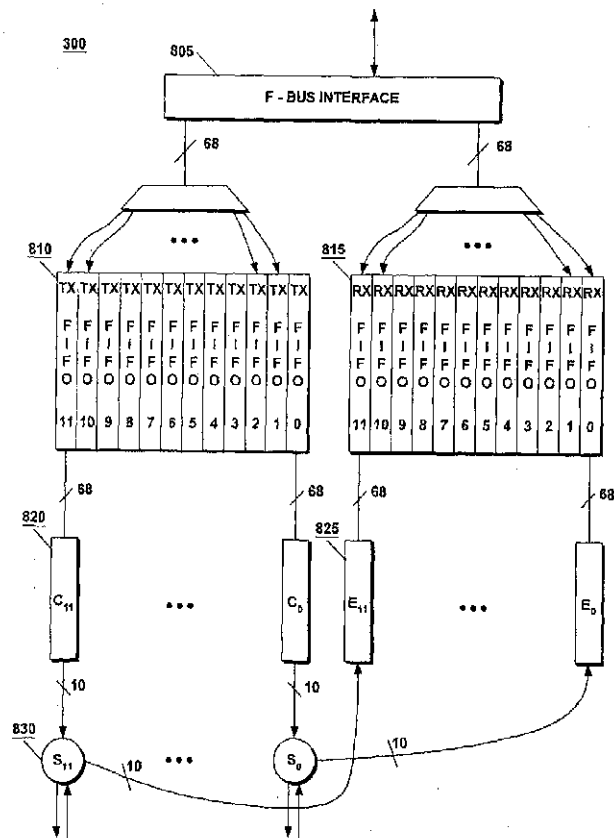
Figure 20

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

18/19

**Figure 21**

SUBSTITUTE SHEET (RULE 26)

WO 01/48983

PCT/US00/35490

19/19

| | | | | | | |
|----------|----|----|----|------|----------|------------------|
| 31 | 15 | 14 | 13 | 12 | 11 | 0 |
| RESERVED | | | | KEEP | PRIORITY | DESTINATION MASK |

Figure 22

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)



PCT

Inc., 8 Technology Drive, Westborough, MA 01752 (US).
PAOLINO, David, R.: Broadband Access Systems, Inc.,
 8 Technology Drive, Westborough, MA 01752 (US).
PEARCE, Kirk, B.: Broadband Access Systems, Inc.,
 8 Technology Drive, Westborough, MA 01752 (US).
SARKAR, Jyotirmoy, R.: Broadband Access Systems,
 Inc., 8 Technology Drive, Westborough, MA 01752 (US).

(21) International Application Number: PCT/US00/35490

(22) International Filing Date: 28 December 2000 (28.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/474,039 28 December 1994 (28.12.1994) US

(71) Applicant: BROADBAND ACCESS SYSTEMS, INC.
[US/US]: 8 Technology Drive, Westborough, MA 01752
(US).

(74) Agents: COHEN, Jerry et al.; Perkins, Smith & Cohen, LLP, One Beacon Street, Boston, MA 02108 (US).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, FE, ES, FI, GB, GR, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JE, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

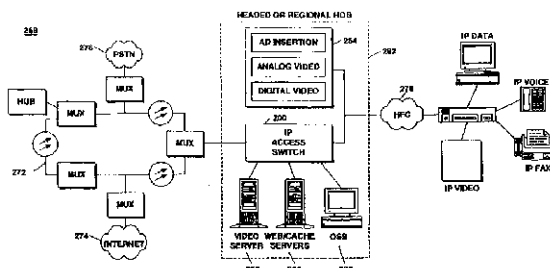
(84) Designated States (regional): ARIPO patent (GH, GM, KH, JS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report

(88) Date of publication of the international search report:
7 March 2002

(Continued on next page)

(54) Title: SYSTEM AND PROCESS FOR HIGH-AVAILABILITY, DIRECT, FLEXIBLE AND SCALABLE SWITCHING OF DATA PACKETS IN BROADBAND NETWORKS



(57) Abstract: A system and process for directly and flexibly switching connections of data packet flows between nodes of a broadband data processing system and network. The system acts as a single IP switch.

WO 01/48983 A3

WO 01/48983 A3

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

【国際調査報告】

| INTERNATIONAL SEARCH REPORT | | International Application No. PCT/US 00/35490 |
|--|---|--|
| A. CLASSIFICATION OF SUBJECT MATTER IPC 7 H04L12/56 H04L12/28 According to International Patent Classification (IPC) or to both national classification and IPC | | |
| B. FIELDS SEARCHED Manner of documentation searched (classification system followed by classification symbols) IPC 7 H04L | | |
| Documentation searched other than minimum documentation in the extent that such documents are located in the fields searched | | |
| Databases consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, PAJ | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category ¹ | Criteria of document, with indication, where appropriate, of the relevant passages | Relevance to claim(s) |
| A | US 5 485 455 A (ANDLAUER PHIL ET AL) 16 January 1996 (1996-01-16) abstract column 2, line 65 - column 4, line 20 column 13, line 40 - column 14, line 23 column 14, line 35 - line 58 claims 1,2 figures 4-6,8 | 1-18, 26-31 |
| A | WO 99 53652 A (3COM CORP) 21 October 1999 (1999-10-21) abstract page 3, line 1 - page 4, line 27 page 5, line 21 - line 26 page 9, line 2 - page 12, line 20 figures 4-6 | 1-18, 26-31 |
| -/- | | |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex | | |
| ¹ Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another claim or other special provision (see, for example) "O" document (relating to an oral disclosure, use, exhibition or other means) "P" document published prior to the international filing date but after the priority date claim(s) | | |
| Date of the actual completion of the international search: 4 October 2001 | | |
| Date of mailing of the international search report: 11/10/2001 | | |
| Name and mailing address of the ISA European Patent Office, P.O. Box 2018 Patentish 2 Tel. (+31-70) 340 2040, Tx (+31-651) 601 601 Fax (+31-70) 340 2016 | | Authorized officer Lai, C |

Form PCT/ISA/210 (September 2000)

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 00/35490

| C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|---|-----------------------|
| Category | Character of document, with indicators, where appropriate, of the relevant passages | Relevant to claim No. |
| A | US 5 951 649 A (CORMIER LEN ET AL) 14 September 1999 (1999-09-14) abstract column 4, line 18 - line 45 claims 1,2 | 1-31 |

1

Form PCT/ISA/210 (continuation of Form PCT/ISA/200) (July 1999)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. J. Appl. No.
PCT/US 00/35490

| Patent document cited in search report | | Publication date | | Patent family member(s) | | Publication date |
|--|---|------------------|----|-------------------------|--|------------------|
| US 5485455 | A | 16-01-1996 | AT | 205652 T | | 15-09-2001 |
| | | | AU | 678687 B2 | | 05-06-1997 |
| | | | AU | 1607895 A | | 15-08-1995 |
| | | | AU | 681062 B2 | | 14-08-1997 |
| | | | AU | 6190296 A | | 03-10-1996 |
| | | | AU | 681063 B2 | | 14-08-1997 |
| | | | AU | 6190396 A | | 03-10-1996 |
| | | | EP | 0741937 A1 | | 13-11-1996 |
| | | | EP | 0746175 A2 | | 04-12-1996 |
| | | | EP | 0746176 A2 | | 04-12-1996 |
| | | | JP | 9508509 T | | 26-08-1997 |
| | | | WO | 9520850 A1 | | 03-08-1995 |
| | | | US | 5491694 A | | 13-02-1996 |
| | | | US | 5521910 A | | 28-05-1996 |
| | | | US | 5790546 A | | 04-08-1998 |
| WO 9953652 | A | 21-10-1999 | US | 6058116 A | | 02-05-2000 |
| | | | WO | 9953652 A1 | | 21-10-1999 |
| US 5951649 | A | 14-09-1999 | US | 5509123 A | | 16-04-1996 |
| | | | AU | 678109 B2 | | 15-05-1997 |
| | | | AU | 2191495 A | | 09-10-1995 |
| | | | EP | 0752180 A1 | | 08-01-1997 |
| | | | JP | 9506752 T | | 30-06-1997 |
| | | | WO | 9526090 A1 | | 28-09-1995 |

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CR,CU,CZ,DE,DK,DM,DZ,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZW

(特許庁注：以下のものは登録商標)

イーサネット

Windows NT

(74)代理人 100081330

弁理士 樋口 外治

(72)発明者 ニコリッチ, ポール イー.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 チンナスワミー, クマー

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 ドールミツァー, ポール エイチ.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 エンゲルセ, ウィレム エー. エイチ.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 マーラ, ウォルター ジー.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 ガイ, ハワード

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 パオリノ, デイビッド アール.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 パース, キルク ビー.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

(72)発明者 サーカー, ジョティルモイ ビー.

アメリカ合衆国, マサチューセッツ 01752, ウェストボロー, テクノロジー ドライブ 8
 , ブロードバンド アクセス システムズ, インコーポレイティド

F ターム(参考) 5K030 JL08 JL10 KA21 LA15 LB05