(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3829165号
(P3829165)
(45)発行日　平成18年10月4日(2006.10.4)　(24)登録日　平成18年7月21日(2006.7.21)

(51)Int.Cl.　F I
H04L 12/46 (2006.01)　H04L 12/46　100B

請求項の数 36　(全 25 頁)

(21)出願番号	特願平9-500466
(86)(22)出願日	平成8年4月9日(1996.4.9)
(65)公表番号	特表平11-506288
(43)公表日	平成11年6月2日(1999.6.2)
(86)国際出願番号	PCT/US1996/004892
(87)国際公開番号	WO1996/041455
(87)国際公開日	平成8年12月19日(1996.12.19)
審査請求日	平成15年3月26日(2003.3.26)
(31)優先権主張番号	08/483,501
(32)優先日	平成7年6月7日(1995.6.7)
(33)優先権主張国	米国(US)

(73)特許権者
アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド
アメリカ合衆国、９４０８８―３４５３ カリフォルニア州、サニィベイル、ピィ・オゥ・ボックス・３４５３、ワン・エイ・エム・ディ・プレイス、メイル・ストップ・６８（番地なし）
(74)代理人
弁理士　深見　久郎
(74)代理人
弁理士　森田　俊雄
(74)代理人
弁理士　伊藤　英彦

最終頁に続く

(54)【発明の名称】全二重イーサネットのための８０２．３メディアアクセスコントロールおよび付随する信号伝達方式の強化

(57)【特許請求の範囲】

【請求項１】

ネットワークスイッチの第１のポートで受取られた入力データパケットを、スイッチの第２のポートと第３のポートとの間に確立された通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、

入力データパケットを阻止するかどうかを判定するステップと、その後

阻止することが望ましい場合、フロー制御情報を含む全二重のフロー制御パケットを第１のポートへ与えるステップと、

入力データパケットを送信している第１のポートの入力に結合される送信機によって、前記フロー制御情報に応答して入力データパケットの送信を打切るステップと、その後

第２のポートが利用可能な場合、再送信された入力データパケットを第１のポートにおいて受取るステップと、その後

第１のポートから第２のポートへ入力データパケットを経路付けするステップとを含む、方法。

【請求項２】

前記フロー制御パケットは、第１のポートに結合される全二重受信器によってフロー制御事象として認知され、第１のポートに結合される半二重受信器において衝突状態をトリガする、請求項１に記載のスイッチングの方法。

【請求項３】

ネットワークスイッチの第１のポートで受取られた入力データパケットを、スイッチの第

２のポートと第３のポートとの間に設けられた通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
入力データパケットを阻止するかどうかを判定するステップと、その後
阻止することが望ましい場合、フロー制御情報を含む全二重のフロー制御パケットをを第１のポートへ与えるステップと、
入力データパケットを送信している第１のポートの入力に結合されるトランシーバによって、前記フロー制御情報に応答して前記トランシーバが入力データパケットを無効にするようにさせるステップと、その後
第２のポートが利用可能な場合、再送信された入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ入力データパケットを経路付けするステップとを含む、方法。

【請求項４】
前記付与するステップは、第１のポートの入力における入力データパケットの受信と同時に第１のポートから前記フロー制御パケットを送信するステップを含む、請求項３に記載のスイッチングの方法。

【請求項５】
前記トランシーバは、予め定められた時間の遅延の後、再送信信号をアサートするためのタイマを含み、前記方法は
前記フロー制御情報の受信で前記タイマを起動するステップと、その後
前記タイマによる前記再送信信号のアサートの後、入力データパケットの再送信を開始するステップとをさらに含む、請求項３に記載のスイッチングの方法。

【請求項６】
全二重ネットワークスイッチの第１のポートの入力において受取られた入力データパケットを、スイッチの第２のポートの出力と第３の入力との間に設けられた通信チャネルを有するスイッチの第２のポートの出力へ、出力データパケットが第１のポートの出力から送信されている間に、スイッチングする方法であって、前記方法は
入力データパケットを阻止するかどうかを決定するステップと、その後
阻止することが望ましい場合、第１のポートからフロー制御情報を含むフロー制御パケットを送信するステップと、
前記フロー制御情報に応答して、第１のポートに結合されるトランシーバによって入力データパケットの送信を打切るステップと、その後
第２のポートの出力が利用可能である場合、第１のポートにおいて入力データパケットを受取るステップと、その後
入力データパケットを第１のポートから第２のポートへ経路付けするステップとを含む、方法。

【請求項７】
フロー制御メッセージを特定された第１の時間期間内に受取るための手段と、前記特定された第１の時間期間後前記フロー制御メッセージを受取るとＭＩＩ型インターフェイスを介して前記フロー制御メッセージ送るための手段と、
ＭＩＩ型インターフェイスを介する後前記フロー制御メッセージの転送を遅延させるための手段とを含む、トランシーバ。

【請求項８】
特定された時間は５１２ビット時間衝突ウィンドウである、請求項７に記載のトランシーバ。

【請求項９】
特定された時間はプログラムされるか、または共同で別の値にネゴシエーションされる、請求項８に記載のトランシーバ。

【請求項１０】
ネットワークスイッチの第１のポートで受取られた第２の入力データパケットを、第１の

入力データパケットの後に、スイッチの第２のポートと第３のポートとの間に設けられた通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
第２の入力データパケットを阻止するかどうかを第１の入力データパケットの送信中に決定するステップと、その後
阻止することが望ましい場合、第１の入力データパケットの送信中にフロー制御情報を含む全二重のフロー制御パケットを第１のポートへ与えるステップと、
入力データパケットを送信している第１のポートの入力に結合される送信機によって、前記フロー制御情報に応答して第２の入力データパケットの送信を無効にし中断するステップと、その後
第２のポートが利用可能である場合、再送信された第２の入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ第２の入力データパケットを経路付けするステップとを含む、方法。

【請求項１１】
ネットワークスイッチの第１のポートで受取られた第２の入力データパケットを、第１の入力データパケットの後、スイッチの第２のポートと第３のポートとの間に設けられた通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
第２の入力データパケットを阻止するかどうかを第１の入力データパケットの送信中に決定するステップと、その後
阻止することが望ましい場合、第１の入力データパケットの送信の後、第２の入力データパケットの送信の前にフロー制御情報を含む全二重のフロー制御パケットを第１のポートへ与えるステップと、
入力データパケットを送信している第１のポートの入力に結合される送信機によって、前記フロー制御情報に応答して第２の入力データパケットの送信を中断するステップと、その後
第２のポートが利用可能である場合、第２の入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ第２の入力データパケットを経路付けするステップとを含む、方法。

【請求項１２】
フロー制御情報を生じさせるための手段と、
フロー制御情報を、遠隔装置に対し、ＩＥＥＥ８０２．３規格に従ったイーサネットパケットフォーマットを用いて送信するための手段とを含む、全二重可能イーサネット装置。

【請求項１３】
パケットは、特別なプリアンブル、ＳＦＤまたはメッセージデータフォーマットのいずれかを用いた、埋込まれたフロー制御情報を含む、請求項１２に記載のイーサネット装置。

【請求項１４】
フロー制御情報を生じさせるための手段と、
フロー制御情報の、遠隔装置との間の送信または受信を、独自の符号化されたパターンを用いて行なうための手段とを含む、全二重可能イーサネット装置。

【請求項１５】
パターンは複数のコードグループであり得る、請求項１４に記載のコントローラ。

【請求項１６】
パターンはＩＰＧの間に送信され得る、請求項１４に記載のコントローラ。

【請求項１７】
パターンはＩＰＧまたはパケットデータの間に送信され得る、請求項１４に記載のコントローラ。

【請求項１８】
コードグループは様々なメッセージとして解釈される、請求項１５に記載のコントローラ。

【請求項１９】
特別なコードグループは「送信停止」インジケータとして解釈される、請求項１８に記載のコントローラ。
【請求項２０】
特別なコードグループは「送信開始」インジケータとして解釈される、請求項１８に記載のコントローラ。
【請求項２１】
特別なコードグループは「ＩＰＧ時間変更」インジケータとして解釈される、請求項１８に記載のコントローラ。
【請求項２２】
特別なコードグループは「バックオフ間隔変更」インジケータとして解釈される、請求項１８に記載のコントローラ。
【請求項２３】
特別なコードグループは「スロット時間間隔変更」インジケータとして解釈される、請求項１８に記載のコントローラ。
【請求項２４】
特別なデータフィールドをＭＩＩ型インターフェイスのＲＸＤ＜３：０＞ラインにわたって送り、特殊フロー制御関連の符号化を示す、請求項７に記載のトランシーバ。
【請求項２５】
データパケットを同時に送受信するための方法を動作させる全二重イーサネット互換コントローラであって、
パケットを受取りながら同時にパケットを送信するステップと、
フロー制御情報の受信を検出するステップと、
送信の試行を打切り、前記フロー制御情報に基づいて決定されたインターバルの後、その試行を再びスケジュールするステップとを含む、コントローラ。
【請求項２６】
ネットワークスイッチの第１のポートで受取られた入力データパケットを、スイッチの第２のポートと第３のポートとの間に確立された通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
入力データパケットを阻止するかどうかを判定するステップと、その後
阻止することが望ましい場合、フロー制御情報を含む全二重のフロー制御パケットを第１のポートへ与えるステップとを含み、前記与えるステップは、入力データパケットの受信と同時に第１のポートから送信される出力データパケットに続いてパケット間の間隔において前記フロー制御パケットを送信するステップを含み、前記方法はさらに、
入力データパケットを送信している第１のポートの入力に結合されるトランシーバによって、前記フロー制御情報に応答して前記トランシーバに入力データパケットを無効にさせるステップと、その後
第２のポートが利用可能な場合、再送信された入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ入力データパケットを経路付けするステップとを含む、方法。
【請求項２７】
ネットワークスイッチの第１のポートで受取られた入力データパケットを、スイッチの第２のポートと第３のポートとの間に確立された通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
入力データパケットを阻止するかどうかを判定するステップと、その後
阻止することが望ましい場合、フロー制御情報を含む全二重のフロー制御パケットを第１のポートへ与えるステップと、
入力データパケットを送信している第１のポートの入力に結合されるトランシーバによって、前記フロー制御情報に応答して前記トランシーバに入力データパケットを無効にさせ

るステップと、その後
第２のポートが利用可能な場合、再送信された入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ入力データパケットを経路付けするステップとを含み、
前記フロー制御情報はバックオフ持続時間を含み、前記方法はさらに、
前記バックオフ持続時間を前記フロー制御情報から抽出するステップと、その後
前記バックオフ持続時間の後、入力データパケットを再送信するステップとを含む、方法。

【請求項２８】
ネットワークスイッチの第１のポートで受取られた入力データパケットを、スイッチの第２のポートと第３のポートとの間に確立された通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
入力データパケットを阻止するかどうかを判定するステップと、その後
阻止することが望ましい場合、フロー制御情報を含む全二重のフロー制御パケットを第１のポートへ与えるステップと、
入力データパケットを送信している第１のポートの入力に結合されるトランシーバによって、前記フロー制御情報に応答して前記トランシーバに入力データパケットを無効にさせるステップと、その後
第２のポートが利用可能な場合、再送信された入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ入力データパケットを経路付けするステップとを含み、
さらに、
第１のポートから前記トランシーバへ再開信号を送信するステップと、その後
前記トランシーバによって応答して入力データパケットを第１のポートへ再送信するステップとを含む、方法。

【請求項２９】
前記再開信号が予め定められた遅延の後に受信されなかった場合、端末局は前記予め定められた時間の後に送信を再開する、請求項２８に記載のスイッチングの方法。

【請求項３０】
全二重ネットワークスイッチの第１のポートの入力において受取られた入力データパケットを、スイッチの第２のポートの出力と第３の入力との間に設けられた通信チャネルを有するスイッチの第２のポートの出力へ、出力データパケットが第１のポートの出力から送信されている間に、スイッチングする方法であって、
入力データパケットを阻止するかどうかを決定するステップと、その後
阻止することが望ましい場合、第１のポートからフロー制御情報を含むフロー制御パケットを送信するステップと、
前記フロー制御情報に応答して、第１のポートに結合されるトランシーバによって入力データパケットの送信を打切るステップと、その後
第２のポートの出力が利用可能である場合、第１のポートにおいて入力データパケットを受取るステップと、その後
入力データパケットを第１のポートから第２のポートへ経路付けするステップとを含み、
前記送信するステップは、第１のポートから前記トランシーバへ送信された出力データパケットの後に前記フロー制御情報を送信する、方法。

【請求項３１】
ネットワークスイッチの第１のポートで受取られた第２の入力データパケットを、第１の入力データパケットの後に、スイッチの第２のポートと第３のポートとの間に設けられた通信チャネルを有するスイッチの第２のポートへスイッチングする方法であって、
第２の入力データパケットを阻止するかどうかを第１の入力データパケットの送信中に決定するステップと、その後
阻止することが望ましい場合、フロー制御情報を含む全二重のフロー制御パケットを第１

の入力データパケットの送信後かつ第２の入力データパケットの送信前に第１のポートに与えるステップと、
入力データパケットを送信している第１のポートの入力に結合される送信機によって、前記フロー制御情報に応答して第２の入力データパケットの送信を中断するステップと、その後
第２のポートが利用可能である場合、第２の入力データパケットを第１のポートにおいて受取るステップと、その後
第１のポートから第２のポートへ第２の入力データパケットを経路付けするステップとを含み、
前記フロー制御情報は再送信されるべき特定のパケットを定義するＳＡ／ＤＡフィールドを含む、方法。
【請求項３２】
データパケットを同時に送受信する全二重イーサネット互換コントローラを動作させるための方法であって、
第１のパケットを送信しながら同時に第２のパケットを受信するステップと、
フロー制御表示のアサートを検出するステップと、
フロー制御表示に応答して第１のパケットの送信を打切り、規定された期間の後、第１のパケットの送信を再びスケジュールするステップとを含む、方法。
【請求項３３】
規定された期間はフロー制御表示によって規定される、請求項３２に記載の方法。
【請求項３４】
フロー制御表示は第１のデータパケットの再送信の前において待つよう特定の数のＩＰＧを示す、請求項３３に記載の方法。
【請求項３５】
フロー制御表示は後に送信されたパケット間において待つよう特定の数のＩＰＧを示す、請求項３３に記載の方法。
【請求項３６】
フロー制御表示は再送信の前において待つよう特定の数のスロット時間を示す、請求項３３に記載の方法。
【発明の詳細な説明】
発明の背景
この発明は、一般的にはコンピュータローカルエリアネットワークに用いられるデータパケットスイッチに関し、より特定的には、そのようなコンピュータネットワークのスイッチのポートに宛てられたパケットのフロー制御を実現するための方法および装置に関する。
コンピュータのネットワークは今日のビジネス環境で一般的に用いられている。一般的なネットワークシステム構造では、１つまたはそれ以上のリピータを用いる。リピータは典型的にはいくつかのポートを含む。１つのポートで受取られたあるデータパケットはリピータの他のポートから再送信される。各々のリピータは、１つのポートにおいて受信されたデータパケットのタイミングおよび振幅劣化を修復し、これらを他のすべてのポートに、したがってネットワーク全体にわたって、再送信する。イーサネットネットワークのような、ＣＳＭＡ／ＣＤ方式のネットワークを用いるネットワークでは、すべてのデータパケットはすべてのリピータを通過する。ネットワーク管理者は、したがってネットワークの稼働に関する情報を集めるためのネットワーク上の道具として、各々のリピータを便利に利用することができる。
従来のイーサネット（８０２．３　１０ＢＡＳＥ５）およびチーパネット（８０２．３　１０ＢＡＳＥ２）においては、ローカルエリアネットワークのすべてのノードが接続されるリニアバスを同軸ケーブルが提供している。ＩＥＥＥによって公表された規格（ＩＥＥＥ規格８０２．３）は、コンピュータネットワークのためのさまざまな機能性を規定する。この規格を、すべての目的のために明示的に引用により援用する。信号伝達は、電流同

期技術を用いて達成され、同軸ケーブルの中心導体は信号に用いられ、同軸ケーブルの遮蔽導体は基準電圧（典型的には接地）に用いられる。より線対イーサネット（８０２．３　１０ＢＡＳＥ－Ｔ）は、同軸ケーブルではなく標準音声声域電話ケーブルを用いる。電話ケーブルは、送信および受信のために別々の導線の対を用いる。

より線対イーサネットを用いる際、ネットワーク構成はスタートポロジーである。スタートポロジーはいくつかの端末局またはデータ端末（ＤＴＥ）装置を考慮したものであり、これらの装置はすべて、スターの中心に位置するマルチポートリピータに結合されている。リピータは信号の振幅およびタイミングの復元を行なう。リピータは、そのポートの１つにおいてビットストリームを受け、すべての適当な出力ポートへの信号振幅レベルおよびタイミング要件を修復する。リピータはその他のポートのすべてに、整形し直されタイミングが調整し直された入力ビットストリームを中継する。ある意味でリピータは論理的同軸ケーブルとして作用し、同軸ケーブルが用いられるときとちょうど同じように、より線対ネットワークに接続されたすべてのノードが、他のどのノードからでも各々の送信を受けることを可能にする。導体の対は差分信号送信を用い、１対は送信用、もう１対は受信用である。

従来のやり方で配線された同軸イーサネットネットワークにおいて、リピータはネットワークの物理的距離制限を拡張するための機構として用いられるが、ＩＥＥＥ８０２．３　１０ＢＡＳＥ－Ｔにおける規格は、２つ以上のノードが存在する場合はいつでもノード間に接続可能性を提供するというリピータの用途を指定している。従来のイーサネット方式のリピータとより線対方式のリピータとでは電線の物理的信号伝達が異なるが、リピータの機能は同じであり、同様に、ネットワークに関係しているノード間でメッセージを伝達するのに用いられるフレームまたはパケット形式も同じである。

パケットは、交互の（「１」および「０」）パターンであるプリアンブルシーケンスで始まる。プリアンブルシーケンスはネットワーク上に各フレームの始めにおいてこの場合は５メガヘルツ（ＭＨｚ）である単一の周波数を提供し、受信器が関連のビットストリームを捕捉しロックするのを可能にする。プリアンブルシーケンスの後に、フレーム開始識別子が続き、そのすぐあとに送信のデータ部分がある。フレーム開始デリミタ（８０２．３）または同期シーケンス（イーサネット）のいずれかがメッセージのデータ部分の開始を表わす。フレーム開始識別子に続くのは、行先アドレス（ＤＡ）および発信元アドレス（ＳＡ）の２つのアドレスフィールドである。これらのアドレスはともに４８ビット値であり、最下位ビット（ＬＳＢ）から先に送信される。

各ＤＴＥに対応するメディアアクセスコントローラ（ＭＡＣ）は、行先アドレスを用いて、入来パケットが対応のノードにアドレスされているかどうかを判定する。受信ノードがそのノードアドレスと行先アドレスフィールド内に送信されているアドレスとの間に一致を検出すると、そのノードはパケットを受信しようと試みる。一致するアドレスを検出しなかったＭＡＣを有するノードでは、典型的にはパケットの残り部分を無視する。

８０２．３規格で指示されている行先アドレス指定には、以下の３種類がある。

１．個別　ＤＡフィールドは、ネットワーク上の単一のノードに割当てられた個別で一意のアドレスを含む。

２．マルチキャスト　ＤＡの最初のビット（ＬＳＢ）がセットされているとき、ＤＡの残り部分はグループアドレスを含む。実際にアドレスされたノード群は上位層機能により決定される。一般的に、グループアドレスの使用は、ネットワーク上の論理的に類似したノードのサブセットへメッセージを送信するように設計されている。

３．ブロードキャスト　ブロードキャストとは、ＤＡフィールドがすべて「１」にセットされているマルチキャストアドレスの特別な形態である。このアドレスは予約されており、ネットワーク上のすべてのノードはブロードキャストメッセージを受信できなければならない。

データパケットを送信するＭＡＣはＳＡフィールドに自らのアドレスを書込む。これによって送信ＭＡＣが自ら生成するそれらのパケットを識別することが可能になる。８０２．３規格によれば、受信ＭＡＣがＳＡフィールドに基づいて動作を起こすことは要求されて

いない。管理、セキュリティ、または構成などのアプリケーションによっては、ＳＡフィールドは追跡され監視され得る。
２バイトの長さ／タイプフィールドがＳＡフィールドの後に続く。長さまたはタイプの選択は、フレームがＩＥＥＥ８０２．３またはイーサネット規格のいずれと互換性があるのかに依存している。長さ／タイプフィールドの上位バイトが最初に送信され、各バイトのＬＳＢが最初に送信される。
データフィールドは端末局間で転送され、４６から１５００バイトの長さである実際のパケットデータを含む。論理リンク制御（ＬＬＣ）機能が、ネットワーク上での送信に適したブロックサイズにデータを細分化する責任を負っている。データバイトは連続して送信され、各バイトのＬＳＢが最初に送信される。
フレームチェックシーケンス（ＦＣＳ）は、フレーム全体のための巡回冗長検査（ＣＲＣ）を含む４バイトフィールドである。送信局はＤＡ、ＳＡ、長さ／タイプフィールド、およびデータフィールドのすべてにわたってＣＲＣを計算する。送信局はＦＣＳをフレームの最後の４バイトとして末尾に追加する。受信局は同様のＣＲＣアルゴリズムを用いて、受信したフレームのためのＣＲＣを計算する。受信局はそれの計算するＣＲＣ値と送信されたＦＣＳのＣＲＣ値とを比較する。不一致は不良データフレームなどのエラーを示す。ＦＣＳのＣＲＣビットは、最上位ビット（ＭＳＢ）からＬＳＢへ順に送信される。
図１および図２はそれぞれ、ＩＥＥＥ８０２．３規格に従ったパケットおよびイーサネットパケットのフォーマットを図解する図である。パケットフォーマットを比較することにより示されるパケットタイプの主な違いは、８０２．３のフレーム開始デリミタ（ＳＦＤ）が「１０１０１０１１」のパターンを有するバイトとして規定される一方、イーサネットの開始フレーム（同期）は「１１」のシーケンスであることである。とはいうものの、いずれの場合においても、プリアンブルのビット総数にフレーム開始表示を加えたものは６４ビット長である。
８０２．３およびイーサネット規格はともに、パケットが６４から１５１８バイト（プリアンブル／ＳＦＤを除く）の範囲内でなければならないことを規定している。しかし、８０２．３システムの実際のデータフィールドは、この最小サイズを確保するのに必要な４６バイト値より小さくてもよいことになっている。小さめのサイズのデータフィールドを扱うために、送信局のＭＡＣは、ネットワーク上にデータを送る前にＬＬＣデータフィールドの末尾に埋込み文字を追加する。データをＭＡＣに送る前に最小データフィールドが４６バイトであることを上位層が確認することをイーサネット規格は想定しているため、イーサネットフォーマットを実装するＭＡＣには末尾に追加された埋込み文字の存在は知られない。
８０２．３規格はまた、データフィールドのみにあるデータバイトの数を示す長さフィールドを用いる。一方イーサネットでは、メッセージプロトコルタイプを識別するために同じ２バイトにあるタイプフィールドを用いる。有効なイーサネットのタイプフィールドは常に８０２．３パケット長サイズの有効最大値外に割当てられるため、８０２．３およびイーサネットのパケットの双方が同じネットワーク上に共存できる。そこで、さまざまな理由のため、アドレスの追跡と監視とが可能なことが重要であることがわかってきた。たとえば、ネットワークに接続されたノードの数が変わるにつれ、アドレスとネットワーク内の特定のポート等とを関連づけることが可能であることが重要になる。
さらに背景の情報を提供すると、衝突ドメインの概念を考慮することが有益である。衝突ドメインとは、ＣＳＭＡ／ＣＤプロトコルを用いる場合に利用可能な全帯域幅に集合的にアクセスするノードおよび端末局の集合である。ＣＳＭＡ／ＣＤシステムは衝突状態（ネットワーク上の２つ以上の装置が同時に送信を試みる場合）を検出する機構を提供する。ＣＳＭＡ／ＣＤ装置は衝突が存在すると、これを検出し、さまざまな装置を制御する手順を提供して、衝突ドメイン内で一度に１つのデータパケットしか送信されないようにする。たとえば、イーサネットでは、衝突事象が検出されると、衝突が検出された際に送信を試みた各端末局はバックオフアルゴリズムを実現するように仕向けられる。バックオフは、特定の端末局が再び送信を試みる前に待機する期間を選択する。いくつかの異なった値

が可能であるため、干渉する端末局はそのうち異なる値を選択することとなり、その結果その端末局のうちの１つがそのデータパケットを送信することになる。
付加的な制御機構がキャリア検出によって提供される。端末局は、別の端末局が衝突ドメイン内に送信している場合にはその衝突ドメイン内にパケットを送信することができない。周知であるように、端末局は別の端末局が送信しているかどうかをキャリア検出機構によって判定する。
個々の端末局の計算能力および帯域幅の要求が高まるにつれ、益々衝突ドメインが飽和しやすくなり、その結果、ネットワークの動作は効率が悪くなり、ネットワーク上のさまざまな資源にアクセスを試みる端末局の動作も効率が悪くなる。
さまざまなネットワークの管理者らがネットワークの輻そうを減少するために用いてきた解決策は、端末局を複数の衝突ドメインにセグメント化することである。各ドメインが関係のあるネットワークの帯域幅を完全に割当てられるので、さまざまなセグメント化された衝突ドメインの動作は改善される。特に高い帯域幅の要件を有する端末局では、これをそれ自体の衝突ドメインにセグメント化することが可能である。
しばしば、さまざまな自律的衝突ドメインは情報を交換することになる。スイッチまたはブリッジなどの特別な装置が複数の衝突ドメインに存在し、さまざまな衝突ドメイン間にデータパケットを経路付ける。これらの特別な装置を設計するにあたって、それらがドメイン間パケットを適切な行先へ効率よく転送しながらも、メッセージを間で転送する衝突ドメインの性能を大幅に劣化させないことが重要である。一般的には、ある衝突ドメイン内の衝突は他のどの衝突ドメインにも伝搬されない。別の衝突ドメイン内の行先に当てられた有効パケットのみが転送されることになる。
性能を改善し、既存のネットワークを強化するため、常に規格が開発されている。これらの規格のうちのいくつかは全二重通信の実現例に取り組むものである。これらの新しい規格の考案の結果、８０２．３／イーサネットネットワークではいくつかの半二重の選択肢が存在し、これはＵＴＰおよびファイバなどの潜在的全二重のメディアタイプにわたって動作する。その選択肢は以下のとおりである。
（ｉ）　１０ＢＡＳＥ－Ｔ．現在据付けされたベースの中では最も一般的なイーサネットのバージョンである。カテゴリ３またはよりよいＵＴＰの２対（片方の１対は送信専用、他方の対は受信専用）を必要とする。２対の束、４対の束、および２５対の束にわたっての動作に適している。
（ii）　１０ＢＡＳＥ－Ｆ．元の１０Ｍｂ／ｓイーサネットの一般的なバージョンであり、光ファイバケーブルにわたって動作し、長距離リンク、ＥＭＩ／ＲＦＩからの免疫、およびＥＭＩ／ＲＦＩの自己放出の減少を可能にする。二重の（一方は送信専用、他方は受信専用）６２．５／１２５ミクロン多モードファイバを必要とする。この規格は概括的に以前のＦＯＩＲＬ規格に代わるものであり、ここに明示的にあらゆる目的のために引用により援用する。
（iii）　１００ＢＡＳＥ－Ｔ４．カテゴリ３ケーブルプラントにわたっての動作に適した１００ＢＡＳＥ－Ｔのバージョンの１つである。カテゴリ３またはよりよいＵＴＰの４対（双方向性送信／受信が２つ、送信専用が１つ、受信専用が１つ）を必要とする。個別の４対ケーブルにわたっての動作には適するが、２５対の束には適さない。１００ＢＡＳＥ－Ｔ４は、送信または受信する際に３対で動作することが必要であるため（第４の対は衝突検出に用いられる）、１００Ｍｂ／ｓ全二重動作には適当な候補ではない。さらに詳しくは、８０２．３ｕ草稿補遺の条項２３を参照されたい。ここは、ここに明示的にあらゆる目的のために引用により援用される。
（iv）　１００ＢＡＳＥ－ＴＸ．カテゴリ５ケーブルプラントにわたっての動作に適した１００ＢＡＳＥ－Ｔのバージョンの１つである。カテゴリ５ＵＴＰの２対（一方は送信専用、他方は受信専用）を必要とする。個別の２対ケーブルにわたっての動作には適しているが、束には適していない。さらに詳しくは、８０２．３ｕ草稿補遺の条項２４および２５を参照されたい。これは、ここに明示的にあらゆる目的のために引用により援用される。

（ｖ）　１００ＢＡＳＥ－ＦＸ．光ファイバケーブルにわたっての動作に適した１００ＢＡＳＥ－ＴＸのバージョンの１つである。二重の（１つは送信専用、もう１つは受信専用）６２．５／１２５ミクロン多モードファイバを必要とする。さらに詳しくは、８０２．３ｕ草稿補遺の条項２４および２６を参照されたい。これは、ここに明示的にあらゆる目的のために引用により援用される。

全二重動作を提供することには２つの根本的な理由がある。

（ｉ）　同時の送信受信活動を提供すること。このことは実際上半二重システムの利用可能な帯域幅を２倍にし、マルチタスクの環境（マルチタスク動作システムを備えたネットワークサーバまたはワークステーション）およびインタラクティブサービス（音声またはテレビ会議）に有効である。

（ii）　ネットワークのトポロジー制約を増すこと。このことは特に、ＩＥＥＥ１００ＢＡＳＥ－Ｔ草案に規定されるような、新しい「ファストイーサネット」トポロジーに基づくネットワークについて言えることである。イーサネットの速度の高められたこの派生物においては、ネットワークは１０倍に増加したデータ速度（１００Ｍｂ／ｓ）で動作する。しかしながら、このネットワークデータ速度の増加に伴って、ネットワークのトポロジー（またはスパン）の削減が起こり、ネットワークの直径はほぼ２００ｍとなる。全二重能力を提供することで、ラウンドトリップが不適当でありエンドツーエンドの減衰のみがリンクの距離に対して重要である場合に、光ファイバなどの長距離リンクを用いることを可能にする。全二重通信のためのフロー制御を実現することは、特にデータスループット速度が急速にバッファ空間を圧倒する速度にまで増加する場合に非常に望ましい。

ＥＰ－Ａ－０　６４８　０３４では、データの送信を停止させる禁止手段に背圧信号を利用する先行技術のイーサネットネットワークを開示する。この構成は全二重動作には適していない。

ＥＰ－Ａ－０　５２９　７７４は、ターゲットのネットワークが半二重メディアを用いる通信ネットワークブリッジ装置における通常輻そうの制御のための先行技術の装置を開示する。

ＵＳ－Ａ－４　７２７　５３７では、公知のデータ送信フロー制御構成を開示するが、これは全二重動作には適さない。

Ｌ．メラッティ（L. Melatti）による“Fast Ethernet: 100M bit/s Made Easy”（Data Communications, Vol 23, No.16, November 1994, p.111-114）もまた、先行技術を記述する。

発明の概要

この発明は、全二重、および半二重のポートを有するイーサネットスイッチにおいて、簡単に、有効に、かつ経済的にフロー制御を提供するための装置および方法を提供する。好ましい実施例の１つの利点は、複数の入力または出力バッファを備えた既存のスイッチ実現例に利用できることである。実現例では、入力／出力バッファの数を減少することができ、同時に設計の元のデータ経路を保つことができる。バッファリング要件を大幅に減少するために機能を追加するには最小の論理を追加するだけでよい。好ましい実施例では、全くどのようなスイッチアーキテクチャでも対応することが可能であり、これはスイッチが専用または共有の入力／出力キュー構造を提供するのかどうかにかかわらず、またはそれがメモリベースであるか、またはバスベースであるかということにさえ関係がない。要するに、説明される好ましい実施例の機構は、スイッチ構成の利用可能なバッファリングおよび集合帯域幅に応じてスイッチの最適化を可能にする。

この発明の別の利点では、フロー制御を用いるスイッチを強化して、実質的により少ない専用バッファメモリを用いた卓越した性能を提供することができ、これはバッファリング方式を簡略化し、フロー制御を与えないスイッチと比べてコストの削減が可能である。

たとえば、出力キュー付スイッチでは、出力バッファリングを減少することができる。フロー制御インジケータは、現出力バッファが存在しないことを示すためだけに生成される。本質的に、背圧フロー制御は出力ポートのための利用可能な出力バッファがない場合に発生する。同様に、入力キュー付スイッチでは、入力バッファリングを減少することがで

き、フロー制御インジケータは現入力バッファが存在しないことを合図するのに用いられる。時分割多重（ＴＤＭ）バスベースのスイッチについては、フロー制御インジケータはＴＤＭバス上にある、現時点で利用可能なスロットのいずれも利用できないことを合図する。

この発明の１局面によれば、ネットワークスイッチの第１のポートにおいて受取られた入力データパケットをネットワークスイッチの第２のポートへスイッチングする方法が含まれ、ネットワークスイッチの第２のポートと第３のポートとの間の通信チャネルが設置されている。この発明の好ましい実施例は以下のステップを含む。すなわち、

入力データパケットを阻止するかどうかを判定するステップと、その後

阻止することが望ましい場合、全二重背圧インジケータを第１のポートへ与えるステップと、

第１のポートの入力に結合される送信機によって、背圧インジケータに応答して入力データパケットの送信を打切るステップと、その後

第２のポートが利用可能な場合、再送信された入力データパケットを第１のポートにおいて受取るステップと、その後

第１のポートから第２のポートへ入力データパケットを経路付けするステップとを含む。

結論としては、２つの潜在的な実現例が説明される。第１は、パケットを全二重局へ送信することに基づき、その局はパケットをフロー制御情報を含むものとして認知し、そのパケットの送信を中断するか、停止するか、またはパケットの以後の伝送速度を送らせるなどの措置を講じる。メッセージを含む明示的なパケット、または全二重局内で動作する時間切れのいずれかによって、後に送信の特性がその通常の、または最適の送信速度に回復される。この方式は既存の１０ＢＡＳＥ－Ｔおよび（ファイバイーサネットの最も一般的に用いられるバージョンであるＦＯＩＲＬを伴う）１０ＢＡＳＥ－ＦＬの装置に用いるのに最適である。なぜなら、全二重局の送信の特性を変えるためか、または、衝突状態（同時の送信および受信動作）の生成によって、半二重局による送信の試みを阻止するために同じ種類のフロー制御パケットを利用できるからである。この方式の不利な点は、局でのフロー制御パケットの受信を用いることによって、実質的に半二重衝突と同じ状態を作り出すには衝突が通常のスロット時間である５１２ビット時間内に起こらなければならないこと（「ウィンドウ内」衝突）を意味することである。このことはさらに、ネットワークのトポロジーを制約してスロット時間内にすべてのネットワークの局にフロー制御パケットの伝搬遅延が届くことを確実にしなければならないことを意味する。しかしながら、インストールされた機能には、これは厳しい問題ではない。既存の１０ＢＡＳＥ－Ｔおよび１０ＢＡＳＥ－ＦＬネットワークは既にトポロジーが制約されており、最大ラウンドトリップ遅延が５１２ビット時間内であることを保証しており、これはすべてのリピータを通る経路も含んでいる。リピータをスイッチによって置き換えて、たとえば帯域幅を強化する場合、以前は同じ衝突ドメイン内に存在していたリピータの各ポートはそれ自体の個別の衝突ドメインに隔離される。本質において、リピータをスイッチで置き換えることは、すなわち、結果として生じる衝突ドメイン間でのエンドツーエンド遅延を減少することである。ある端末局のアプリケーションによる要求が小さなワークグループで帯域幅を共有することすら超える場合、「専用イーサネット」アプローチを用いることができる。専用イーサネットは実質的にはネットワークのセグメント化を論理的に極端化したものであり、ただ１つの端末局しかスイッチの各ポートに接続されていない。各ポートは完全な１０Ｍｂ／ｓを達成できるため、各デスクトップは有効な「専用」１０Ｍｂ／ｓを達成する。専用イーサネットの強化は「全二重イーサネット」を用いることである。これは、性能をより高め、同時の送信および受信活動が重要である、インタラクティブの応用例でのサポートを強化するのに用いられる。しかしながら、イーサネット端末局のインストールされた機能のほとんどは半二重モードのみで動作しているため、必ずといっていいほど端末局をアップグレードする必要がある。また、スイッチされたイーサネットを使用するには全二重動作でなければならないわけではないが、全二重イーサネットを使用するにはスイッチを使用しなければならないことに注意されたい。

端末局での全二重動作は典型的には、以下に示される、当業者には周知であるＭＡＣおよびレイヤ管理機能に加えられる小さな変更によって特徴付けられる。

ＭＡＣ－据置および衝突中止

衝突ディセーブル	4.2.8	“procedure WatchForCollision ”を不能化／変更する
据置ディスエーブル	4.2.8	“process Deference ”を不能化／変更する

レイヤ管理－ＭＩＢ破壊を避け、制御／ステータスを加える

ループバックディセーブル	5.2.2.1.2	“Number of carrierSenseErrors”を不能化／変更する
ＳＱＥテストディセーブル	5.2.3.1.1	“Number of SQETestErrors ”を不能化／変更する

全二重動作（管理によりサポートされる）をイネーブル／不能化する

全二重動作（管理によりサポートされる）を報告する

フロー制御の機構を有さずに無差別の送信および受信活動を可能にする、以前の全二重端末局実現例に対するこの発明の主な違いは、（任意に、特別なプリアンブル、ＳＦＤ、または特定のデータパターンのいずれかを含む）フロー制御パケットの検出と、それに対する反応、すなわち、現送信パケットに衝突を引起こすことである。たとえば全二重モードで動作している場合に衝突スロット時間を増加させたり、局の送信性能調整に、より細かい粒度を提供すること（ＩＰＧ間隔を長くすること等）の他の強化策も可能である。しかし、これは元の半二重イーサネットＭＡＣ機能からさらに逸脱するため、実現化がより複雑になるであろう。

この第１の方式は設置された１０Ｍｂ／ｓＵＴＰおよびファイバのインストールされた機能に最適であるが、これは半二重モードにおけるフロー制御では１００ＢＡＳＥ－Ｔ４にも適用可能であり、全二重または半二重モードのいずれかにおいて１００ＢＡＳＥ－ＴＸ／ＦＸ局にも適用可能である。

次に第２の方式が説明される。これは、現在予約されている制御コードグループが局面でフロー制御情報を転送するのに用いられる、１００ＢＡＳＥ－Ｘ実現例の信号伝達の特性に合わせて最適化されている。これもまた、半二重局とともに用いることができる。なぜなら、これらのコードグループは、受取られた際、アイドルから非アイドル状態へ遷移するラインを示す「偽キャリア」表示として解釈されるが、ＳＳＤ（「Ｊ」コードグループと、その後の「Ｋ」コードグループ）は識別子として受取られないからである。このことは、半二重局が送信している場合は衝突状態につながり、そのため１０Ｍｂ／ｓシステムにおけるフロー制御パケットの発生に本質的に類似している（すなわち、局は同時の送信および受信活動を検出する）。

この方式には明らかな利点がある。第１は、ＭＡＣおよび調停サブレイヤ内での変化を最小限に抑え、変化のほとんどをＰＨＹに強いることである。たとえば、特別なコードグループの受信をＸＯＮ／ＸＯＦＦプロトコルに制限することができ、これはまた、任意のバックグランドタイマによって明らかなＸＯＮメッセージの必要をなくす。ＰＨＹはＸＯＦＦメッセージを受取り、通常の態様でＭＩＩ　ＣＯＬ表示を用いて衝突を発生させる。以後に説明されるように、他の特別なコードグループを衝突表示の間に送ることによって拡張も可能になるが、このことはまた、ＭＡＣレイヤにおける、通常のＣＳＭＡ／ＣＤ半二

重バージョンからのずれがより大きくなることを意味する。ＰＨＹは衝突表示を用いるため、唯一必要なＭＡＣの変更は受信活動が進行中である際に送信活動を可能にすることだけである。このことは、別個の送信および受信キャリア検知表示を用いることによって簡単に達成できる。受信キャリア検知はＭＩＩ　ＲＸ＿ＤＶ信号から得ることができ、送信キャリア検知信号は通常のＭＩＩ　ＣＲＳ信号から得ることができる。また、この技術は当業者には既に周知である。

第２の利点は、通常のパケット内にフロー制御を埋込むことが可能なことであり、このことでフロー制御の粒度を大幅に補助することができる。さらに、ＰＨＹが送信活動を監視することができ、送信側ＭＡＣがスロット時間内に存在する場合に、（たとえば）フロー制御ＸＯＦＦ信号のみを印加するため、トポロジーを拡張することができる。このようにして、送信側ＭＡＣは「ウィンドウ内」衝突を検出してバックオフし、適宜再試行間隔をスケジュールする。スロット時間が満了した後、ＸＯＦＦ表示が受信された場合、ＰＨＹは次の送信パケットまで衝突表示を遅延することができる。

いずれの方式においても、ＩＰＧおよび／またはスロット時間は変更および／またはネゴシエーションすることができる。たとえば、電源投入時では、自動ネゴシエーションプロトコル（８０２．３ｕ条項２８）を用いるか、またはコードグループ内容を含むパケットを動的に用いてこれらのパラメータを変更することである。

図面および請求の範囲を含むこの明細書の残りの部分を参照することによって、この発明の他の特徴および利点が明らかになるであろう。この発明のさらなる特徴および利点も、この発明のさまざまな実施例の構造および動作も、添付の図面に関連して以下に詳しく説明される。図面では同じ参照番号は同一または機能上類似した要素を示す。

【図面の簡単な説明】

図１は、ＩＥＥＥ８０２．３規格に従ったパケットのためのフォーマットを示す図である。

図２は、イーサネットパケットのためのフォーマットを示す図である。

図３は、スイッチを有し、ドメイン間パケットを適切なドメイン内に経路付けるための多衝突ドメインネットワークのブロック図である。

図４は、受信側衝突ドメインが使用中である場合の１つの衝突ドメインへの背圧の付与を図示する概略ブロック図である。

図５は、この発明の好ましい実施例を組込んだスイッチの概略ブロック図である。

図６は、背圧および論理的衝突ドメインをサポートするスイッチ構成のための詳細な論理である。

図７は、全二重部分を含むネットワークスイッチのブロック図である。

好ましい実施例の説明

図３は、スイッチ２０を有し、ドメイン間パケットを適切な衝突ドメイン内へ経路付けるための多衝突ドメインネットワーク１０のブロック図である。ネットワーク１０は、第１のドメイン３０、第２のドメイン４０、および第３のドメイン５０を含み、これらのすべてはスイッチ２０の半二重ポートに結合される。図３では、第１のドメイン３０は一連のパケット６０をスイッチ２０を通して経路付けることによって第２のドメイン４０へ送信していると想定する。スイッチ２０はパケット６０を第２のドメイン４０の中へ経路付け、ドメイン４０の中に経路付けられるパケットはパケット７０として示される。以下の説明のために、スイッチの概念はその広義において用いられ、スイッチ、ハブ、ゲートウェイ、ルータ、集信装置、またはその他の術語のいずれで称されるかにかかわらず、衝突ドメイン間パケット経路制御装置を意味する。

最初に、スイッチ２０は、入来パケット６０の行先アドレスフィールドに含まれる行先アドレス情報に応答して、第１のドメイン３０と第２のドメイン４０との間に通信チャネルを設定する。パケットを１つの衝突ドメインから別の衝突ドメインへ送るプロセスはＩＥＥＥ８０２．３規格に従うものである。パケット６０のうちの１つが第１のドメイン３０から第２のドメイン４０へスイッチ２０を通して送信される間、第３のドメイン５０はパケット８０を第１のドメイン３０または第２のドメイン４０へ送信することを試みること

があり得る。第３のドメイン５０は送信中に別のドメインの１つにパケット８０を送ることが可能である。なぜなら、第３のドメイン５０は他の衝突ドメインからネットワーク活動について情報を受取らないからである。
パケット８０を受取ると、スイッチ２０はパケット８０が有効なパケットであるかどうか、かつ第３のドメイン５０以外の衝突ドメインの端末局に宛てられているかどうかを判定する。パケット８０が別の衝突ドメインに宛てられている場合は必ず、スイッチ２０は所望のドメインが使用中であるかどうかを判定する。
説明を簡単にするために、パケット８０は第２のドメイン４０へ送信されるものと想定する。パケット８０がスイッチ２０において受信された際、所望のドメインが使用中であった場合、スイッチ２０はパケット８０を第２のドメイン４０へ送ることができるまでパケット８０をバッファするためにメモリのいくらかを利用することができるであろう。しかしながら、コストの競合のあるスイッチでは、バッファメモリの量は必然的に限られており、場合によってはスイッチ２０はパケット８０をバッファできないことがある。従来のスイッチは、フロー制御機構を通して第３のドメイン５０を十分に制御して、第３のドメイン５０がパケット８０を第２のドメイン４０へ送信するのを中断することができない。パケット８０はそのため送信され消失する。従来のシステムの物理レイヤは、高位誤り検出および訂正機構によってこの種類の誤り状態から回復しなければならない。このことは大幅な性能劣化につながることがある。
図４は、使用中である受信側ドメインへのパケット８０の送信に応答しての第３のドメイン５０への背圧の付与を図示する概略ブロック図である。スイッチ２０は、パケット８０を阻止しなければならないことを判断した後（スイッチ２０がパケット８０をバッファすることも第２のドメイン４０へ転送することもできないため）、第３のドメイン５０内へ背圧を与える。好ましい実施例ではファントムデータパケット９０である背圧は、スイッチ２０から第３のドメイン５０へ送られる。ファントムパケット９０は事実上どんなタイプのパケットでもよく、好ましい実施例においては、スイッチ２０によって作り出されるが、そうでなければネットワーク１０上に存在しないことからファントムパケットと呼ばれる。ファントムパケット９０に要求されることの１つは、パケット８０がスイッチ２０へ送信されている間に第３のドメイン５０においてファントムパケット９０が受信されると、第３のドメイン５０内の衝突状態がトリがされることである。
第３のドメイン５０はパケット８０の送信を中断するか、またはＩＥＥＥ規格８０２．３に従った周知の態様でパケット８０を再送信することで望ましい、ＩＥＥＥ８０２．３に従った、衝突状態に応答する。最も簡単な実現例では、スイッチ２０は、第２のドメイン５０がパケット８０の受信を検出し、その受信を阻止する必要がある度に（たとえば、バッファリングが使用不能であり所望のドメインが使用中である場合）第２のドメイン５０に単一のファントムパケット９０の形で背圧を与える。第３のドメイン５０はファントムパケット９０に応答して、バックオフアルゴリズムを実現し、これは、第３のドメイン５０がパケット８０の送信を再度試みることができる前にランダムに遅延を割当てる。応用例によっては、この簡単なタイプのフロー制御で十分である。
他の応用例では、上述のフロー制御手順は完全な解決をもたらさないこともある。例によっては、所望のドメインが長い連続したデータパケットを受信することもある。第３のドメイン５０がパケット８０の送信を再度試みるのがいつなのかを調整することなしには、第２のドメイン４０に送信される一連のパケット６０のパケット間の間隔（ＩＰＧ）にパケット８０がスイッチ２０において受信される確率は低い。標準のバックオフアルゴリズムでは、以後、衝突が検出されることなくパケット８０を送信するのに失敗する度に遅延が増加するという点で、パケット８０がＩＰＧの間に受取られる可能性は減少する。可能性が減少する理由は、バックオフアルゴリズムにより、ファントムパケット９０によって引起こされるパケット８０の衝突の度に、次の送信を試みる前に経過していなければならない、起こり得るバックオフ遅延の範囲が増大することである。ＩＥＥＥ８０２．３規格は、プロセスを終了し再試行エラーをアサートするまでに、１つの特定の送信の試みに対して１６回の失敗（衝突）しか許容しない。

この制約を効果的に回避するために、好ましい実施例では、単に単一のファントムパケット９０を第３のドメイン５０に送る以上のことをする。スイッチ２０は実は、関係のある３つの衝突ドメインから１つの論理的衝突ドメインを作るのである。換言すると、スイッチ２０は第１のドメイン３０、第２のドメイン４０、および第３のドメイン５０を組合せることによって単一の論理的衝突ドメインを作る。これを実行するために、スイッチ２０は第１のドメイン３０からのデータパケット６０の送信が持続する間、キャリア活動信号として一連のファントムパケット９０を発生する。一連のファントムパケット９０の第１のものはバックオフアルゴリズムを開始する。後続のファントムパケット９０の送信は第３のドメイン５０によってキャリア活動として検出され、キャリア活動が停止するまで第３のドメイン５０がパケット８０の送信を再び試みることを禁じる。キャリア活動が停止し、第１のドメイン３０と第２のドメイン４０との間の通信チャネルがアイドルとなり、第２のドメイン４０がパケット８０を受けることができるようになると、第３のドメイン５０は第２のドメイン４０へのアクセスを巡って攻撃的に争う機会を与えられる。

第３のドメイン５０が第２のドメイン４０を巡って攻撃的に争う際に、現トランザクションに係わるドメインと、関係のあるドメインのうちの１つとの通信を開始するよう試みるドメインとに対して論理的衝突ドメインを用いることは、ネットワーク１０のフロー制御およびアクセス可能性／性能を大いに改善する。バースト転送の場合など、第１のドメイン３０が第２のドメイン４０を独占している例においては、第３のドメイン５０はなお第２のドメイン４０にアクセスするのが困難であることがある。すべての衝突ドメイン内のすべてのＤＴＥの各々はＩＰＧカウンタを含み、ＩＰＧカウンタはパケットの送信の間の最小の間隔を制御する。ＩＥＥＥ８０２．３規格を実現するにあたっては、キャリア活動が完了した後までＤＴＥがそのＩＰＧカウンタを開始しないことが要求される。そのため、第１のドメイン３０による第２のドメイン４０へのバックツーバック送信では、第３のドメイン５０による第２のドメイン４０へのアクセスは厳密には公平ではない可能性がある。公平さを増し、機能性を強化するために、スイッチ２０は優先順位付け機構を実現し、いずれかの衝突ドメインによる、別の衝突ドメインへのアクセスも制御する。

好ましい実施例では、この優先順位付け機構は、各ポートに１つずつの、複数のスロットルカウンタ（図示せず）を含む。スイッチ２０がある衝突ドメインに背圧を与えると、その衝突ドメインに関連付けられたスロットルカウンタがインクリメントされる。スロットルカウンタのカウントは優先順位の問題を判定する際スイッチ２０によって利用される。パケット８０が第２のドメイン４０に首尾よく送信された後、スイッチ２０はスロットルカウンタをクリアし、論理的衝突ドメインを取除く。

優先順位付け機構を効率よく実現するために、スイッチ２０は、現在使用中の行先ドメインに宛てられたパケットを有する各衝突ドメインに、行先ドメインがアイドルになるまで、背圧をアサートする。次に、スイッチ２０は、行先ドメインに送信されるパケットを有する衝突ドメイン（この例では第３のドメイン５０）に与えられている背圧を軽減することによって行先ドメイン（第２のドメイン４０）へのアクセスを認める。

好ましい実施例では、スロットルカウントは優先順位を与える際に用いられる。スロットルカウントが１６回のパケットの再試行制限に近づくにつれ、第３のドメイン５０にアクセスを与えることは益々切迫したものとなる。スイッチ２０は、特定のポート（ポート番号）または端末局（パケットの発信元アドレス）へのアクセスを強化する他の優先順位付けアルゴリズムを含んでもよい。

好ましい実施例では、予め定められた時間内に２つの入力ポートが活性状態になる場合、各関連のあるポートにおいてＭＡＣからスロットルカウント値が読出され、最も高い（または予め定められた）値のスロットルカウントを有する受信ポートに優先順位が与えられる。実際上同時に受取られたと考えられる、到着しつつある受信パケットの期間はプログラム可能にすることもでき、これによって可能な限り遅くまで入力ポートが出力ポートを巡って争うことを可能にし、公平さを最大にする。

好ましい実施例では、ファントムパケットの発生の開始はプログラム可能にされる。ファントムパケットの発生が起こるのは、パケットの早期の段階のある時点において（スロッ

ト時間内のある時点において）であるため、このことは事実上、衝突ドメインの論理的スパンをその物理的スパンより大きく見せる。実質的には、ファントム衝突の発生の遅延は、付加的なラウンドトリップ遅延として発信元端末局に現れる。スイッチポートの下でさまざまなトポロジーに対応可能にするために、ファントムパケットの発生の開始がプログラム可能であることが必要であるが、通常は妥当なデフォルトが用いられる。拡張されたトポロジーが必要である場合には、ファントムパケットの発生が開始するまでの時間は減少されることとなり、潜在的な遅い衝突の問題を避ける。

好ましい実施例では、ファントムパケットの内容および／または持続時間を制御して、システムのロバストネスを強化することができる。半二重実現例のファントムパケットに含まれる実際のデータは任意のものである。しかしながら、ファントムパケットを受信する装置がデータを有効なパケットと解釈しないことが望ましい。このことを確実にする１つの方法は、単純に、ファントムパケットがプリアンブル（交互の「１」および「０」）のみを有するようにすることである。そのような単純なファントムパケットは（１０ＢＡＳＥ－Ｔなどの）リンクセグメントには十分であるかもしれないが、そのようなファントムパケットが（１０ＰＡＳＥ２などの）混合セグメントに対しては不適当である可能性がある。すなわち、応用例によっては、ファントムパケットは小型（ファントムパケットはプリアンブルを含んで５７６ビットより小さくなければならない）として解釈されるか、または規定の長さのファントムパケットは無効なＣＲＣ値を有するとして解釈されなければならない。

好ましい実施例では、ファントムパケットがプログラム可能な大きさを有することはオプションである。あるモードでのファントムパケットの発生は、発生したファントムパケットのすべてが確実に小型であるようにする。小型パケットは受信器によって自然に無効パケットとして拒絶される。ファントムパケットの発生はまた、行先ドメインが使用中である限りファントムパケットを続けるようにプログラム可能である。

このように、好ましい実施例は、行先ドメインが使用中である間中、通常のプリアンブルシーケンスとしてのファントムパケットを背圧として発生する妥協案を提供する選択肢を含んでいる。プログラム可能な選択肢として、ファントムパケットは複数の小さいパケット（小型ファントム）にセグメント化され、この小さいパケットは許されているより近づけて連続して送信される（つまり、２つの小型ファントムの間のＩＰＧは違法的に短い）。このタイプの背圧によって、フロー制御が適用されている装置がまず衝突を検出し、短い沈黙の後、次に第１の部分のフレーム間間隔（ＩＦＳ）に達する前に、受信器がキャリア活動を検出しバックオフすることが確実になる。小型ファントムの各々はパケットの大きさが無効であるため受信側メディアアクセスコントローラ（ＭＡＣ）によって拒絶されるので、不法ＣＲＣ値を作り出す必要はほとんどない。

図５は、この発明の好ましい実施例を組込んだスイッチ２０の概略のブロック図である。スイッチ２０は、スイッチ処理エンジン１００、データ転送ユニット１０５、ファントムパケットおよび論理的衝突ドメイン論理１１０、ならびに１対のＱＵＡＤ統合イーサネットＭＡＣ装置１１５（統合パケット１１５につき４つのＭＡＣ）を含む。スイッチ２０はすなわち、８ポートのスイッチ構成を提供する。ＱＵＡＤ　ＭＡＣ１１５は、カリフォルニア州サニィベイルのアドバンスト・マイクロ・ディバイシズ（Advanced Micro Devices,Inc.Sunnyvale,California）からの商業的に入手可能なＱＵＡＤ－ＭＡＣＥチップ（Ｐ／Ｎ　ＡＭＭ７９Ｃ９４４）を変形したものである。ＱＵＡＤ－ＭＡＣＥのための現在利用可能な仕様はここに明示的にあらゆる目的のために引用により援用する。

ＱＵＡＤ－ＭＡＣＥはポートごとに２つのピンを加えることによって変更される。出力ピンの１つ（ＣＲＳｘ）は、対応するポートｘが使用中である場合（送信中または受信中）アサートされ、もう１つの入力ピンはＢＬＯＣＫｘ信号を受ける。ＢＬＯＣＫｘ信号がアサートされると、ポートｘはＢＬＯＣＫｘ信号がアサートされている限り背圧をアサートする。

ポートｘでパケットが受取られると、スイッチ処理エンジン１００は適当なＭＡＣ１１５のポートｘからの入来パケットの開始を読む。ルックアップを実行して物理的行先ポート

であるポートｙを定めた後、ポートｙが既に使用中であることが判定された場合、スイッチ処理エンジン１００は活性状態の入力ポートを３ビットＰＯＲＴ　ＩＤコードとして、適当な出力ポートｙの論理的衝突ドメインコントローラへ書く。論理的衝突ドメインコントローラは３ビットのＰＯＲＴ　ＳＥＬＥＣＴコードを用いて選択される。各ポートは関連付けられた３－ｔｏ－８線デコーダを有する（以下の図６参照）。ポートＩＤが書き出されると、ＣＲＳ出力（この場合ポートｙの送信または受信状態を示す）はＢＬＯＣＫ信号としてポートｘへアサートされる。スイッチ処理エンジン１００は、ポートｘがポートｙへ送信するのに成功するまでこの構成を維持する。ポートのＣＲＳ出力は同じポートのＢＬＯＣＫ入力へフィードバックされないことに注意されたい。スイッチ処理エンジン１００は、ポート上の入来パケットが物理的にそれ自体に宛てられている場合、適切な措置を講じるものと想定され、ハードウェアブロッキングは実行されない。

図６は、背圧および論理的衝突ドメインをサポートするスイッチ構成のための詳細な論理である。ファントムパケットおよび論理的衝突ドメインコントローラ１１０は、８ポートのスイッチに対して、８つのポートｘ受信パケット行先デコーダ２００ｘ（ポートごとに１つ）および８つのブロックキング７入力ＯＲゲート（Ｇ０－Ｇ７）を含む。各デコーダ２００ｘはＭＡＣ１１５からＣＲＳｘを受取り、ＣＲＳｘがアサートされるとポートｙを識別するデコードされたＰＯＲＴ　ＩＤ信号に応答して、デコーダ出力信号ＤＥＣＯＤＥｙをアサートする。デコーダ２００ｘからの対応するＤＥＣＯＤＥｙ信号（すなわちポートｙを識別するもの）は、各々対応するブロッキングＯＲゲートＧｙの入力の１つに提供される。

各デコーダ２００ｘは、３－ｔｏ－８デコーダ２５０、７つの二入力ＮＯＲゲート（Ｇ１０－Ｇ１６）およびインバータＧ１７を含む。デコーダ２００ｘには、ＣＲＳｘがインバータＧ１７へ入力される。インバータＧ１７の出力は各ゲートＧ１０－Ｇ１６の第１の入力に結合される。ＮＯＲゲートＧｘの出力はＤＥＣＯＤＥｙ信号を提供する。ＣＲＳｘ入力はまた、デコーダ２５０のイネーブル入力に提供される。デコーダ２５０は、ＰＯＲＴ　ＩＤ信号とイネーブル信号のアサートとに応答して、活性状態の受信ポートを識別するビットコードを８ビットコードに変換する。変換されたＰＯＲＴ　ＩＤコードのビットの各々のうちの１つはＮＯＲゲートＧ１０－Ｇ１６の第２の入力の１つに結合される。

たとえば、ポート０が活性受信器であり、識別される行先出力ポートがポート７である場合（ＣＲＳ７はアサートされていると想定する）、値０がポート７受信パケット行先デコーダ２００_{7}へ書かれると、ＣＲＳ７はＤＥＣＯＤＥ０においてブロッキングＯＲゲートＧ０へ経路付けられる。ＤＥＣＯＤＥ０が活性状態である場合（たとえば活性ＣＲＳ７がＧ０へ経路付けられている）、ＢＬＯＣＫ０がアサートされ、ＣＲＳ７がアサートされている限り、ファントムパケットがポート０に結合される衝突ドメインの中へ送信される。スイッチ処理エンジン１００は、２つの入力ポートが活性状態である場合、適当なスロットルカウント値を読むと同時に（または時間の間隔をあまり置かずに）同じ出力資源へのアクセスを要求することができると想定する。スイッチ処理エンジン１００は次に、最大のスロットルカウントを有するポートを選び、それが出力ポートを占有できるようにする一方、低いスロットルカウント値を有するポートは、そのポートＩＤを関連のある出力ポートのための論理的衝突ドメインコントローラへ書くようにする。

以上の説明はスイッチ内で用いるための半二重フロー制御システムで、典型的に１０ＢＡＳＥ－Ｔ、１０ＢＡＳＥ－ＦＬまたはＦＯＩＲＬ、および／または１００ＢＡＳＥ－Ｔ４リンクセグメントが用いられるシステムを明らかにした。以下の説明では、そのようなスイッチを相互接続したり、または全二重動作によって利益を得られるような装置へのアクセスを提供するための全二重の使用と、これらの環境に対するフロー制御機構の拡張可能性とについて考察する。

図７は、第１のネットワークスイッチ４０５および第２のネットワークスイッチ４１０を組込んだネットワーク４００のブロック図であり、各スイッチは全二重ポートと半二重ポートとを有する。ネットワーク４００はいくつかの衝突ドメイン４１５ｉを組込んでおり、各衝突ドメインは、データパケットが衝突ドメイン間に転送される場合を除いて他のド

メインとは別個に動作する。
ネットワーク４００には多くの構成が可能であるが、そのうちの１つが好ましい実施例の説明を簡略化するために以下に説明される。ネットワーク４００の他の構成およびオリエンテーションも、この発明を実現しながら依然として可能である。ネットワーク４００は、スイッチ４０５とそれに結合される衝突ドメイン（つまり、衝突ドメイン415_1、415_2および415_3）との間の１０Ｍｂ／ｓまたは１００Ｍｂ／ｓ半二重リンクの必要に応え、スイッチ４１０はそれ自身と衝突ドメイン415_4と衝突ドメイン415_5との間の１０Ｍｂ／ｓまたは１００Ｍｂ／ｓの半二重リンクで動作する。ここに明示的にあらゆる目的のために引用により援用される、ＩＥＥＥ規格部門による、１９９３年度版ＡＮＳＩ／ＩＥＥＥ規格８０２．３、文書＃Ｐ８０２．３ｕ／Ｄ５、「１００Ｍｂ／ｓ動作のためのＭＡＣパラメータ、物理レイヤ、メディア接続ユニットおよびリピータ（第５版）」（MAC Parameters,Physical Layer,Medium Attachment Units and Repeater for 100Mb/s Operation(Version 5.0)）の草稿補遺では、１００Ｍｂ／ｓ動作のための環境の１つが規定される。この規格を実現することには端末局の物理的距離が制限されている（たとえば１００ＢＡＳＥ－Ｔ草案を用いると約２００ｍ）という欠点がある。
この物理的なケーブルの制約を改善するために、ネットワーク４００はスイッチ４０５とスイッチ４１０との間に全二重光ファイバリンク４２０を実装する。一実現例としては、信号減衰によって定められる物理的リンク（現在は約２ｋｍの限界）を有する、１００ＢＡＳＥ－ＴＸ環境の全二重バージョンを用いることであろう（ここに明示的にあらゆる目的のために引用により援用される、１９９３年度版ＡＮＳＩ／ＩＥＥＥ規格８０２．３の１００ＢＡＳＥ－Ｔ草稿補遺の条項２６および２７）。このようにして、スイッチ４０５およびスイッチ４１０は光ファイバリンクをわたって衝突ドメイン間パケットを交換し、通常の衝突ドメインリンクの物理的限界を避けることができる。
他の例においては、スイッチ４０５が全二重送信プロトコルを用いてデータベース４２５（サーバなど）と通信し、性能を高め衝突ドメイン415_iのいずれかにある端末局によるデータへのアクセスを改善することが望ましい。スイッチ４０５は１００Ｍｂ／ｓ全二重より線対リンク４３０によってデータベース４２５に結合される。リンク４３０の一実現例では、光ファイバケーブルの費用をかけずに全二重１００Ｍｂ／ｓ動作の性能を提供する、１００ＢＡＳＥ－ＴＸ環境の全二重バージョンを用いる（ここに明示的にあらゆる目的のために引用により援用される、１９９３年度版ＡＮＳＩ／ＩＥＥＥ規格８０２．３の１００ＢＡＳＥ－Ｔ草稿補遺の条項２４および２５）。ネットワーク４００の動作速度においては、衝突ドメイン間パケットの送信を制御するために効率のよい全二重フロー制御機構を提供することが重要である。また、半二重環境および全二重環境が混在していることはフロー制御機構をより複雑にする。スイッチはまた、１００Ｍｂ／ｓポートに加えて全二重１０Ｍｂ／ｓポートを組込むことが可能である。
可能な１０Ｍｂ／ｓシステムについては、１０ＢＡＳＥ－Ｔはアイドルの期間中（かなりの時間パケット活動がない状態）、「リンクテストパルス」を用いて、エンドツーエンドリンクが無傷であることを監視する。１０ＢＡＳＥ－ＦＬ（ファイバリンク）および１０ＢＡＳＥ－ＦＢ（ファイババックボーン）は「活性アイドル」信号を用い、それぞれ１ＭＨｚまたは２．５ＭＨｚ信号を送信する。これらの方式は両方とも、検出の速度に関する理由のため、フロー制御データをアイドルインジケータにコード化するのにはあまり適していない。
１００ＢＡＳＥ－Ｔ４はまた、リンクのエンドツーエンド整合性を判定するために「リンクテストパルス」を用いる。１００ＢＡＳＥ－Ｔ４は対称１００Ｍｂ／ｓ全二重動作には適していないが、全二重の非対称バージョン（公称６６Ｍｂ／ｓを送信し３３Ｍｂ／ｓを受信するなど）の候補ではないというわけではない。この場合、１０ＢＡＳＥ－Ｔシステムのように、「リンクテストパルス」はパルス間に必要な分離のためと、システムのノイズに対するロバストネスを増加するために複数のメッセージの複写を送信しなければならないかもしれないため、リアルタイムのフロー制御情報を担持するのには適していない。
そのため、１０ＢＡＳＥ－Ｔ、１０ＢＡＳＥ－Ｆ、および１００ＢＡＳＥ－Ｔ４システム

は前に概説した帯域内フロー制御機構により適しているようである。すなわち、パケットがフロー制御情報を担持しているものとして識別されることを可能にする、特別なＳＦＤまたはデータフィールドなどの埋込インジケータを含むフロー制御パケットを作ることである。パケットは良好に形成され（有効なＣＲＣを有し）、その中に、行先の送信機の後の性能予想に関係のある情報を保持していることに注意されたい。このようにして、半二重フロー制御はファントムパケットの発生によって生じる衝突機構を用いて提供することができる。１０ＢＡＳＥ－Ｔおよび１０ＢＡＳＥ－ＦＬの全二重バージョンのためのフロー制御は、ファントムパケットに対して異なる態様で反応し、ファントムパケット内に埋込まれた情報を用いてその送信動作を変更するように構成してもよい。

他方では、１００ＢＡＳＥ－ＴＸおよび１００ＢＡＳＥ－ＦＸ信号伝達方式では、送信されているのがデータであるかアイドル情報であるかにかかわらず、連続したキャラクタのストリームを送る。これはこれらの技術の独特な局面であり、両方とも前述の帯域外信号方式で動作するのに非常に適したものにする。１００ＢＡＳＥ－Ｘ（ＴＸおよびＦＸ）システムは４Ｂ／５Ｂ符号化技術を用いて、４ビットのデータパケットを取り、それを５ビットのコードグループに変換し、これが実際にメディアにわたって送信される。逆デコード技術によって、受取られた５ビットのコードグループを４ビットのデータフィールドに再マップし、これをＭＡＣに送り返す。１００ＢＡＳＥ－Ｔシステムでは、ＭＡＣとＰＨＹとの間で４ビットの情報を送るための標準化されたインターフェイスはメディア独立インターフェイス（ＭＩＩ）と称され、８０２．３ｕ草稿補遺の条項２２の中に挙げられている。

１００ＢＡＳＥ－Ｘでは、ＰＨＹは４ビットの２進データをメディア上を送信される５ビットのコードグループに変換したり、その逆をしたりする。４ビットのデータには、明らかに１６通りの組合せが可能である。５ビットコードがマッピングに利用可能であるため、コードグループには３２通りの可能な組合せがあり得る。コードグループの３２通りの可能な組合せのうち１６通りだけがデータビットに割当てられるが、その残りのコードグループは特別な機能のために用いられる。たとえば、アイドルコードグループ（１１１１１）はＩＰＧの間、ＭＡＣから実質的にデータが送られてこないときに送信される。さらに、コードグループのいくつかは制御機能のために予約されており、たとえば「Ｊ」（１０００）コードグループと「Ｋ」（１００００１）コードグループとは、ストリーム開始デリミタ（ＳＳＤ）を示すのに共に用いられ、アイドルからフレームデータへの遷移を示す。「Ｔ」（０１１０１）コードグループと「Ｒ」（００１１１）コードグループとはストリーム終了デリミタ（ＥＳＤ）を示すように共に用いられ、フレームデータからアイドルへの遷移を示す。「Ｈ」（００１００）コードグループは無効データを示すのに用いられ、特別な誤り状態の下でのみ送信されるべきである。さらなる１０のコードグループはこの時点では規定されないが、いずれも受信した場合無効のコードを示す。これらは前に引用した１００ＢＡＳＥ－Ｘのための条項２４の表２４－１に規定されている。

これによって未使用のコードグループはフロー制御情報を送信するのに利用することができる。これはパケット間の間隔（ＩＰＧ）内の信号伝達が可能なだけでなく、コードグループを慎重に選択することによって、パケットデータストリーム内にフロー制御情報を信号で送ることも可能である。このようにデータストリーム内で信号伝達をする能力があることで、ＰＨＹに付加的な遅延（すなわちバッファリング）制約を強いることとなり、これは１００ＢＡＳＥ－ＴＸの場合には、ネットワークの遅延見積りは既に限界に近く、最小のマージンしか提供していないため重大である。

いくつかの未使用のコードグループを利用することによって簡単で柔軟性のあるフロー制御を実現することが可能になる。たとえば、コードグループを「ＸＯＮ、ＸＯＦＦ」を意味するようにマッピングすることで使用するコード空間を最小にすることができる。別のより柔軟なコードグループの割当のやり方としては、１つのコードグループを「フロー制御情報ヘッダ」として割当てることであろう。このようにして、他のコードグループを「ＸＯＮ」、「ＸＯＦＦ」、「時間ｘ分送信を遅延せよ」、「ＩＰＧを期間ｘ分増加せよ」などの意味を持たせるように割当てることができる。

ＰＨＹはこれらのフロー制御コードグループを解釈し、調停サブレイヤおよびＭＩＩを介してＭＡＣへ送るものと想定される。簡単なＸＯＮ／ＸＯＦＦ制御の場合では、前に示されるように、ＰＨＹは単に衝突をアサートしパケットの送信を打切るものと想定される。この方式を拡張すると、ＰＨＹは実際に「遅い衝突」しきい値そのものを「取締まる」のに用いることもできるのに注意されたい。たとえば、フロー制御インジケータが遅い衝突しきい値の後に受取られた場合、ＰＨＹはその表示を次のパケットの送信の開始までバッファする。このようにして、次のパケット送信は明らかな「ウィンドウ内」衝突を経験し、（全二重モードである場合に起きた、調整によるバックオフ期間が何であれ、それを用いて）通常バックオフする。このことは、送信側ＭＡＣが、スロット時間後にフロー制御／衝突信号が受取られ、ＭＡＣが同じパケットを再試行しようとする場合、送信を破壊するかどうかを判定しなければならないという問題を軽減する。そのため、このことは全二重と半二重動作との間でＭＡＣが必要とする変更を最小限にする。

明らかに、フロー制御情報をパケット内に埋込んで送信する能力には、遠隔局の送信挙動の制御の粒度がより厳しく制御できるという点において利点がある。たとえば、次の２つの例では、ＩＰＧ内のみでフロー制御を用いることを試みる場合に予想され得るレーテンシの種類を考察する。

ＩＰＧベースのフロー制御の例：

スイッチは、特定のポートに接続される全二重局へ最大長パケット（１５００データバイト）の送信を開始する。

スイッチが送信を開始した直後、スイッチはフロー制御を送る必要があることを認識する（たとえば、同時に局からパケットを受取っており、バッファ資源が不足しつつある）。

局は、スイッチがスイッチ自身の送信を開始するのとほぼ同時に現送信を終了する。

この時点で局はフロー制御情報を受取っていない（最大長パケットが送信されているため）ので、局のＩＰＧタイマは満了し、以下のいずれかを送信する。

（ｉ）　最大長パケット

（ii）　複数の最小長（６４データバイト）パケット（最小のＩＰＧを備えた約１８個の６４バイトパケットが、単一の１５００バイトパケットの要した時間の間に送信され得る）

（iii）　いくつかの中間長のパケット（６４より大きくかつ１５００より小さいデータバイトのパケット）

スイッチは、一旦、最大長パケットを完了すると、ＩＰＧ内に局へフロー制御メッセージを発する。

局が局の次のパケット送信において５１２ビット時間しきい値に対する前にフロー制御情報が局に届いた場合、その送信は打切られることになる。フロー制御情報が「ウィンドウ内」衝突時間の後に届いた場合、パケットは完了し、後続のパケットが遅延または中断されることになる。

このことから、リンクのエンドツーエンド遅延が（公称）５１２ＢＴスロット時間より少ない場合には、スイッチは最大長パケットの１つ、または最大１８個の最小長パケットのいずれかをバッファする能力がなければならないことがわかる。フロー制御情報がスロット時間ウィンドウの後に届いた場合、スイッチは最大２個の最大長パケット、または１パケットが最小長または最大長のいずれかであるパケットを１９個、バッファする能力がなければならない。

パケット内のフロー制御は、このため、スイッチのバッファリングの要件を大幅に減少することを可能にし、潜在的により費用効果のある実現を可能にする。

上に概説したような、より複雑な信号伝達が用いられる場合、（特定した時間、送信を遅延するためのフロー制御コードグループなどの）この情報はＭＩＩのＲＸＤ＜３：０＞ラインにわたって供給されなければならなくなる。この際、ＣＯＬラインの両方はアサートされ、ＲＸ＿ＤＶラインはデアサートされ、ＲＸ＿ＥＲラインはアサートされる。さらに、０００１から１１０１、または１１１１のコードのうちの１つがＲＸＤ＜３：０＞に割当てられる（ＲＸ＿ＤＶラインがデアサートされ、ＲＸ＿ＥＲラインがアサートされてい

る場合、コード１１１０は偽キャリア表示のために予約される）。この機構を用いて、追加の情報をＭＡＣまたは管理エンティティへ効率よく送り、ＭＡＣの再送信挙動を変更することができる。

単にパケットが拡張された距離を伝搬するのにかかった遅延が原因で、全二重リンク上のエンドツーエンド遅延がスロット時間より長くなり得るということは、潜在的な問題につながることがあり、これは避けなければならない。たとえば、パケットが端末局によって再試行される状況では、スイッチはパケットを複製すべきではない。ここでのよい例としては最小の大きさのフレームを送信することである。これは６４バイト＋６４ビットのプリアンブルを有し、合わせて５７６ビット時間のパケットの大きさとなる。スイッチがフロー制御信号を送ると想定して、このフロー制御信号は端末局にビット時間５１１において到達するなら、端末局がフレームの送信を完了して送信を再度スケジュールできる可能性がある。スイッチには端末局がフロー制御信号を正確にどの時点で受取ったかを示す表示がなされないため、パケットが再送信されると、スイッチは以前のパケットの複製を受取ることがある。これにはいくつかの潜在的な解決策がある。

（ａ）　フロー制御信号が受取られなければならないウィンドウを拡張する。これにより、フロー制御信号が送られると、それは必ず再送信を起こし、そのため現在受取られたパケットは常にスイッチによって廃棄できるようにすることを保証できる。これには、再試行ウィンドウが大きくなり、パケットが完全に端末局コントローラの中へ再ロードされなければならないことがあり、バスの帯域幅の利用を劣化させるという不利な点があった。

（ｂ）　ＩＰＧの間のみフロー制御を有効にさせる。これにより、フロー制御フレームを受取った装置は必ず現送信パケットを完了するのに成功するが、次のパケットを送信する前に、これよりかなり大きいＩＰＧを加えることになる。この機構を用いると、再試行は必要ではなくなる。不利な点は、スイッチが現パケットをバッファするのを強いられることである。

（ｃ）　ウィンドウ内にフロー制御パケットが受信されると端末局からの送信されたパケットが無効のＦＣＳを有するようになることを保証する。このことで、スイッチまたは最後の行先は、フレームを受取った時点でこれを廃棄することになる。端末局コントローラは再試行をスケジュールする場合には送信されたパケットを破壊しなければならず、フロー制御要求を許されているウィンドウの後に受取った場合にはパケットを通常どおり届けなければならない。

（ｄ）　フロー制御情報が５１２ＢＴの制限までにターゲットの装置に到着することを保証する。これは実質的にリンクの片道遅延を５１１ビット時間に制限するので望ましくないかもしれない。

（ｅ）　ＰＨＹを用いて「遅い衝突」しきい値を「取締る」。たとえば、フロー制御インジケータが遅い衝突しきい値の後に受取られた場合、ＰＨＹはその表示を次のパケットの送信の開始までバッファすることになる。このようにして、次のパケット送信は明らかな「ウィンドウ内」衝突を経験し、（全二重モードである場合に起きた、調整によるバックオフ期間が何であれ、それを用いて）通常バックオフする。このことは、送信側ＭＡＣが、スロット時間後にフロー制御／衝突信号が受取られ、ＭＡＣが同じパケットを再試行しようとする場合、送信を破壊するかどうかを判定しなければならないという問題を軽減する。そのため、このことは全二重と半二重動作との間でＭＡＣが必要とする変更を最小限にする。

結論としては、アルゴリズムには以下に概説されるように２つの可能な派生物がある。

１．再試行ベース

端末局は送信する。

スイッチはフロー制御信号（ファントムパケットまたはコードグループ）を送信する

端末局はフロー制御信号を受取る

端末局は送信を破壊し、試行を再びスケジュールする

端末局は（時間が経過するまでか、または「再開」信号を待って）待機する

端末局は再試行する

２．遅延ベース
端末局は送信する
スイッチはフロー制御信号（ファントムパケットまたはコードグループ）を送信する
端末局はフロー制御信号を受取る
端末局は送信を完了する
端末局は（時間が経過するまで、または「再開」信号を待って）待機する
端末局は次のパケットを送る
１００ＢＡＳＥ－Ｘベースのネットワークに特有の再試行ベースのアルゴリズムの代替案は以下のとおりである。
３．再試行ベース（１００ＢＡＳＥ－Ｘ）
端末局は送信する
スイッチはフロー制御信号を送信する
端末局はフロー制御信号を受取る
ＢＨＹは「ウィンドウ内」フロー制御であるか、または「ウィンドウ外」フロー制御であるかを判定する
「ウィンドウ内」の場合、ＰＨＹはフロー制御を付与する
「ウィンドウ外」の場合、ＰＨＹは次の送信の試行までフロー制御を遅延する端末局は送信を打切り、試行を再びスケジュールする
端末局は（時間が経過するまでか、または「再開」信号を待って）待機する
端末局は再試行する
３通りのアルゴリズムに当てはまる別の考慮点は、端末局がその時点で送信していない場合に端末局がフロー制御信号を受取ることの影響についてである。２つの明らかな選択肢が考えられる。待ち時間をスケジュールすることもできれば、何の措置も講じないでいることもできる。次の送信までに待ち時間がスケジュールされることが好ましい。これは、フロー制御要求が以前のパケットの結果、発生したものであるかもしれず、伝搬の遅延のためにパケットが完了した後に届くかもしれないという事実を見込んでいる。
以上から、再試行ベースの方式が帯域内信号伝達方式（ファントムパケット）または帯域外方式（コードグループ）によく適したものであるように思われる。なぜなら、フロー制御信号を素早く（つまりスロット時間内に）検出することが必要だからである。さらに、この方式では、半二重にも全二重にも適用可能であることと、フロー制御信号が通常のデータフォーマットパケットに実質的に（たとえば特別なプリアンブル／ＳＦＤシーケンスまたはコードグループを用いることによって）「ピギーバック」できるということなどの利点がある。この方式ではバッファが利用できなくても動作可能なため、スイッチ内での最小のバッファリングが可能である。遅延ベースの方式は帯域外および帯域内信号伝達においてうまく動作するが、資源制限の検出／信号伝達がスロット時間内に実行できるのでない限り（このことは伝搬遅延だけをとっても実際的ではないが）、スイッチがある程度の最小限のバッファリング能力を提供する必要がある（つまり、現在送信されているパケットを受信する能力があり、さらに潜在的な遅延を考慮してフロー制御パケットの行先局への配達を確実にする）。
結論としては、この発明は半二重イーサネットスイッチのフロー制御の問題に簡単で有効な解決策を提供する。上記はこの発明の好ましい実施例の完全な説明であるが、さまざまな代替案、修正例、および均等物を用いることもできる。このように、上述の説明は添付の請求の範囲によって規定されるこの発明の範囲を限定するものとして受取られるべきではない。

【図１】

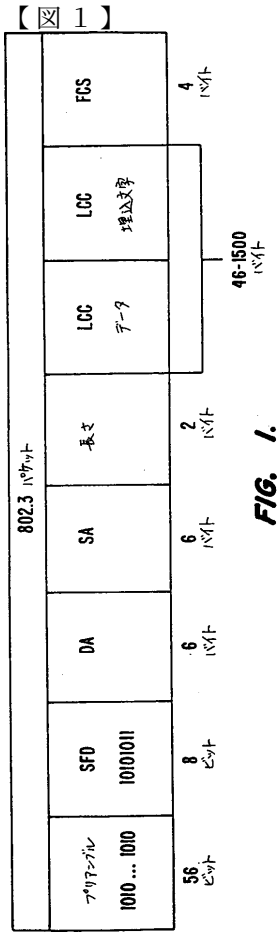

FIG. 1.

【図２】

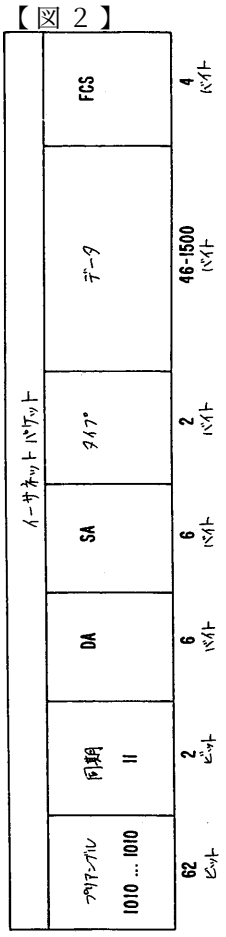

FIG. 2.

【図３】

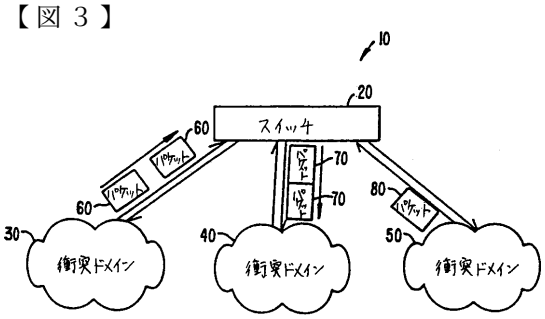

FIG. 3.

【図４】

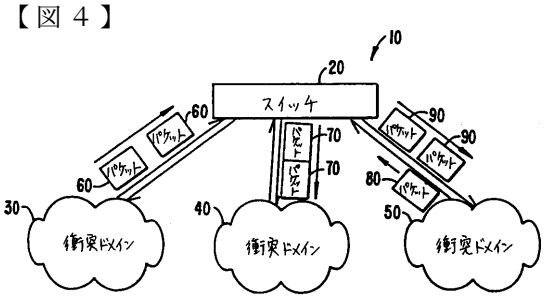

FIG. 4.

【図５】

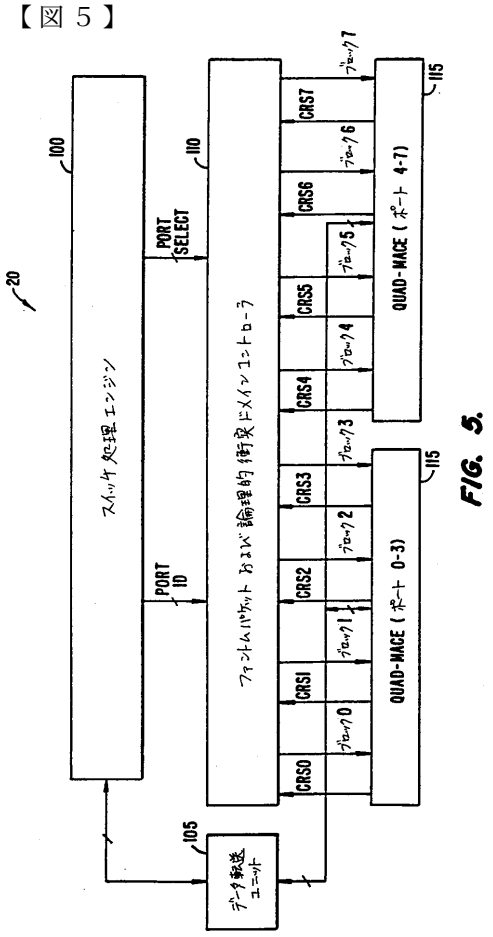

FIG. 5.

【図６】

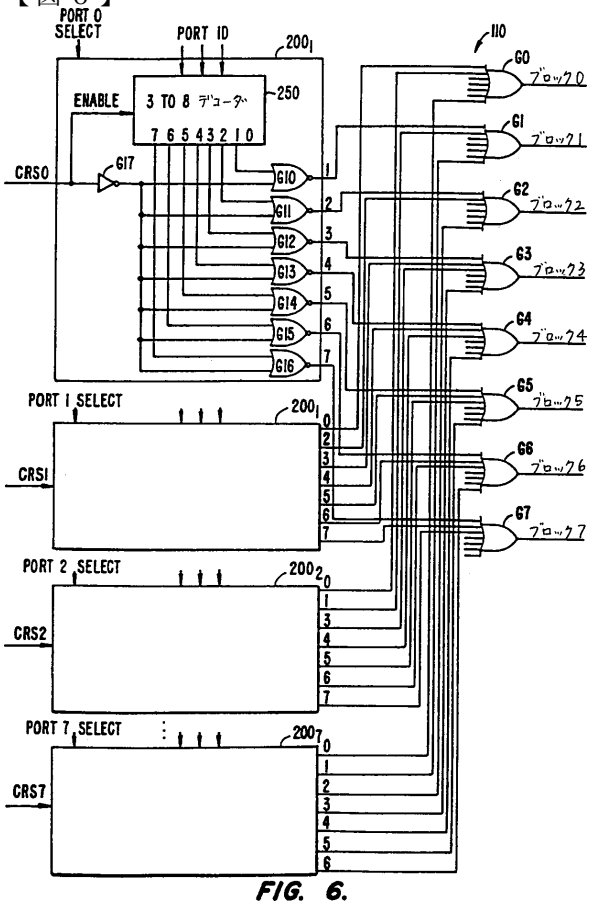

FIG. 6.

【図７】

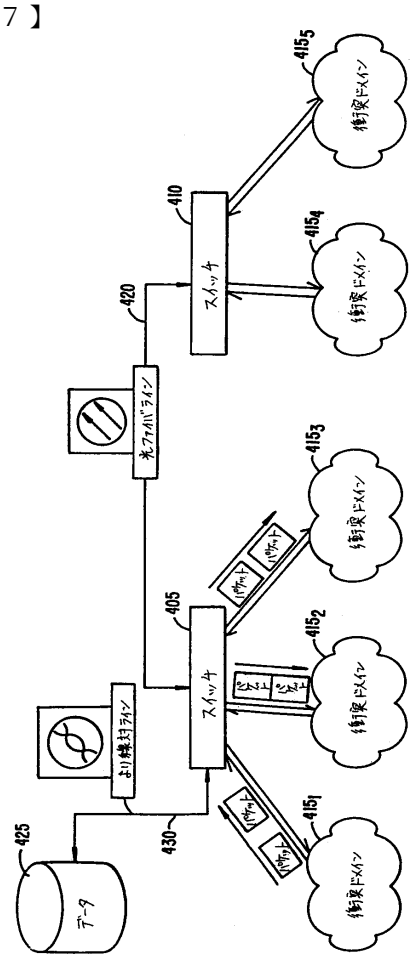

FIG. 7.

フロントページの続き

(74)代理人
弁理士　堀井　豊
(72)発明者　クレイフォード，イーアン
アメリカ合衆国、９５１２９　カリフォルニア州、サン・ノゼ、アイリーン・ドライブ、５３８０

審査官　矢頭　尚之

(56)参考文献　特開平０５－２０７０３５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
H04L 12/28