

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号
特開2010-9472
(P2010-9472A)

(43) 公開日 平成22年1月14日(2010.1.14)

(51) Int.Cl.	F I	テーマコード (参考)
G 0 6 F 13/00 (2006.01)	G O 6 F 13/00 3 5 3 A	5 B 0 1 4
G 0 6 F 13/10 (2006.01)	G O 6 F 13/10 3 4 0 A	5 B 0 6 5
G 0 6 F 3/06 (2006.01)	G O 6 F 3/06 3 0 1 M	5 B 0 8 9
H 0 4 L 12/56 (2006.01)	H O 4 L 12/56 2 0 0 E	5 K 0 3 0
H 0 4 L 29/06 (2006.01)	H O 4 L 13/00 3 0 5 C	5 K 0 3 4

審査請求 未請求 請求項の数 20 O L (全 28 頁)

(21) 出願番号	特願2008-170579 (P2008-170579)	(71) 出願人	305013910 国立大学法人お茶の水女子大学 東京都文京区大塚2丁目1番1号
(22) 出願日	平成20年6月30日 (2008. 6. 30)	(71) 出願人	800000080 タマティーエルオー株式会社 東京都八王子市旭町9番1号 八王子スク エアビル11階
		(71) 出願人	307022424 ソフトバンクテレコム株式会社 東京都港区東新橋一丁目9番1号
		(74) 代理人	100118821 弁理士 祖父江 栄一
		(74) 代理人	100102635 弁理士 浅見 保男

最終頁に続く

(54) 【発明の名称】 ネットワークシステム、ネットワークシステムの制御方法および制御プログラム

(57) 【要約】

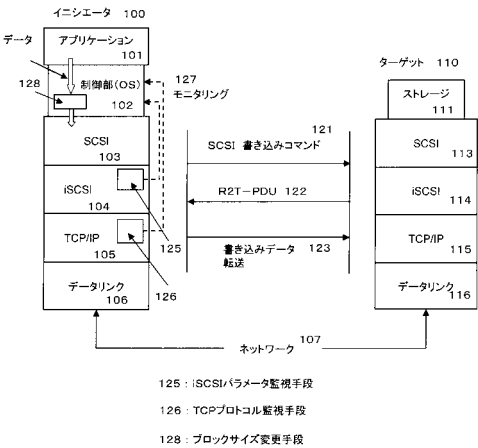
【課題】

iSCSIにより接続されている第1の装置から第2の装置にデータを送信するTCP/IPネットワークシステムにおいて、iSCSIコネクションのR2Tのデータ送信サイズとTCPコネクションの輻輳ウィンドウサイズの制御に適合したブロックサイズによりSCSIコネクションのデータの送信制御を行う。

【解決手段】

iSCSI層に設けられたiSCSIパラメータ監視手段125によりR2Tのデータ送信サイズと、TCP層に設けられたTCPプロトコル監視手段126によりTCP層のバッファを監視し、両情報に基づいてSCSI層103が処理するデータブロックサイズを、TCP層のバッファがオーバーフローしない大きさに制御する。

【選択図】 図1 - 1



【特許請求の範囲】**【請求項 1】**

SCSI コネクションと iSCSI コネクションと TCP コネクションを介して第 1 の装置から第 2 の装置にデータを送信するネットワークシステムであって、

前記 iSCSI コネクションを監視し iSCSI コネクションのデータ送信パラメータを取得する iSCSI パラメータ監視手段と、

前記 TCP コネクションを監視し TCP コネクションのデータ送信状態情報を取得する TCP プロトコル監視手段と、

前記 iSCSI コネクションのデータ送信パラメータと前記 TCP コネクションのデータ送信状態情報に基づいて SCSI コネクションのデータブロックサイズを決定するブロックサイズ決定手段とからなることを特徴とするネットワークシステム。

10

【請求項 2】

SCSI コネクションと iSCSI コネクションとデータリンクを介して第 1 の装置から第 2 の装置にデータを送信するネットワークシステムであって、

前記 iSCSI コネクションを監視し iSCSI コネクションのデータ送信パラメータを取得する iSCSI パラメータ監視手段と、

前記データリンクを監視し、データリンク層のデータ送信状態情報を取得するデータリンク層監視手段と、

前記 iSCSI コネクションのデータ送信パラメータと前記データリンク層のデータ送信状態情報に基づいて SCSI コネクションのパラメータを決定する SCSI パラメータ決定手段とからなることを特徴とするネットワークシステム。

20

【請求項 3】

請求項 1 又は請求項 2 記載のネットワークシステムであって、

前記 iSCSI コネクションのデータ送信パラメータは前記第 2 の装置の iSCSI 層から前記第 1 の装置の iSCSI 層に送信される R2T-PDU に含まれるデータ送信サイズであることを特徴とするネットワークシステム。

【請求項 4】

請求項 1 又は請求項 3 記載のネットワークシステムであって、

前記 TCP コネクションのデータ送信状態情報は前記 TCP コネクションの輻輳ウィンドウサイズであることを特徴とするネットワークシステム。

30

【請求項 5】

請求項 1 又は請求項 3 記載のネットワークシステムであって、

前記 TCP コネクションのデータ送信状態情報は前記第 1 の装置の TCP 層のバッファの状態情報であることを特徴とするネットワークシステム。

【請求項 6】

請求項 2 又は請求項 3 記載のネットワークシステムであって、

前記データリンク層のデータ送信状態情報は前記第 1 の装置のデータリンク層のバッファの状態情報であることを特徴とするネットワークシステム。

【請求項 7】

請求項 1 乃至請求項 6 記載のネットワークシステムであって、

前記 iSCSI コネクションのデータ送信パラメータと前記 TCP コネクションのデータ送信状態情報に基づいて SCSI 層において並列に実行する SCSI シーケンスの数を指定する並列実行シーケンス数指示手段を有することを特徴とするネットワークシステム。

40

【請求項 8】

請求項 1 乃至請求項 6 記載のネットワークシステムであって、

前記 iSCSI コネクションのデータ送信パラメータと前記 TCP コネクションのデータ送信状態情報に基づいて iSCSI コネクション再構成指示手段を有することを特徴とするネットワークシステム。

【請求項 9】

SCSI コネクションと iSCSI コネクションと TCP コネクションを介して第 1 の装置から第 2

50

の装置にデータを送信するネットワークシステムの制御方法であって、

前記iSCSIコネクションを監視し、iSCSIコネクションのデータ送信パラメータの情報を取得するiSCSIパラメータ監視ステップと、

前記TCPコネクションを監視し、TCPコネクションのデータ送信状態情報を取得するTCPプロトコル監視ステップと、

前記iSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報と前記TCPコネクションのデータ送信状態情報に基づいてSCSIコネクションのデータブロックサイズを決定するブロックサイズ決定ステップとからなることを特徴とするネットワークシステムの制御方法。

【請求項 10】

10

SCSIコネクションとiSCSIコネクションとデータリンクを介して第1の装置から第2の装置にデータを送信するネットワークシステムの制御方法であって、

前記iSCSIコネクションを監視しiSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報を取得するiSCSIパラメータ監視ステップと、

前記データリンクを監視し、データリンク層のデータ送信状態情報を取得するデータリンク層監視ステップと、

前記iSCSIコネクションのデータ送信パラメータの情報と前記データリンク層のデータ送信状態情報に基づいてSCSIコネクションのデータブロックサイズを決定するブロックサイズ決定ステップとからなることを特徴とするネットワークシステムの制御方法。

【請求項 11】

20

請求項9又は請求項10記載のネットワークシステムの制御方法であって、

前記iSCSIコネクションのデータ送信パラメータは前記第2の装置のiSCSI層から前記第1の装置のiSCSI層に送信されるR2T-PDUに含まれるデータ送信サイズであることを特徴とするネットワークシステムの制御方法。

【請求項 12】

請求項9又は請求項11記載のネットワークシステムの制御方法であって、

前記TCPコネクションのデータ送信状態情報は前記TCPコネクションの輻輳ウィンドウサイズであることを特徴とするネットワークシステムの制御方法。

【請求項 13】

請求項9又は請求項11記載のネットワークシステムの制御方法であって、

前記TCPコネクションのデータ送信状態情報は前記第1の装置のTCP層のバッファの状態情報であることを特徴とするネットワークシステムの制御方法。

【請求項 14】

請求項10又は請求項11記載のネットワークシステムの制御方法であって、

前記データリンク層のデータ送信状態情報は前記第1の装置のデータリンク層のバッファの状態情報であることを特徴とするネットワークシステムの制御方法。

【請求項 15】

請求項9乃至請求項14記載のネットワークシステムの制御方法であって、

前記iSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報と前記TCPコネクションの輻輳ウィンドウサイズの情報に基づいてSCSI層において並列に実行するシーケンス数を指定する並列実行シーケンス数指示ステップを有することを特徴とするネットワークシステムの制御方法。

40

【請求項 16】

請求項9乃至請求項14記載のネットワークシステムの制御方法であって、

前記iSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報と前記TCPコネクションの輻輳ウィンドウサイズの情報に基づいてiSCSIコネクション再構成ステップを有することを特徴とするネットワークシステムの制御方法。

【請求項 17】

SCSIコネクションとiSCSIコネクションとTCPコネクションを介してターゲット装置にデータを送信するネットワークシステムのイニシエータ装置であって、

50

前記iSCSIコネクションを監視しiSCSIコネクションのデータ送信パラメータを取得するiSCSIパラメータ監視手段と、

前記TCPコネクションを監視しTCPコネクションのデータ送信状態情報を取得するTCPプロトコル監視手段と、

前記iSCSIコネクションのデータ送信パラメータと前記TCPコネクションのデータ送信状態情報に基づいてSCSIコネクションのデータブロックサイズを決定するブロックサイズ決定手段とからなることを特徴とするイニシエータ装置。

【請求項 18】

SCSIコネクションとiSCSIコネクションとデータリンクを介してターゲット装置にデータを送信するネットワークシステムのイニシエータ装置であって、

前記iSCSIコネクションを監視し、iSCSIコネクションのデータ送信パラメータを取得するiSCSIパラメータ監視手段と、

前記データリンクを監視し、データリンク層のデータ送信状態情報を取得するデータリンク層監視手段と、

前記iSCSIコネクションのデータ送信パラメータと前記データリンク層のデータ送信状態情報に基づいてSCSIコネクションのデータブロックサイズを決定するブロックサイズ決定手段とからなることを特徴とするイニシエータ装置。

【請求項 19】

SCSIコネクションとiSCSIコネクションとTCPコネクションを介して第1の装置から第2の装置にデータを送信するネットワークシステムの制御プログラムであって、

前記iSCSIコネクションを監視しiSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報を取得するiSCSIパラメータ監視ステップと、

前記TCPコネクションを監視しTCPコネクションのデータ送信状態情報を取得するTCPプロトコル監視ステップと、

前記iSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報と前記TCPコネクションのデータ送信状態情報に基づいてSCSIコネクションのデータブロックサイズを決定するブロックサイズ決定ステップ

の各ステップを制御部に実行させることを特徴とするネットワークシステムの制御プログラム。

【請求項 20】

SCSIコネクションとiSCSIコネクションとデータリンクを介して第1の装置から第2の装置にデータを送信するネットワークシステムの制御プログラムであって、

前記iSCSIコネクションを監視し、iSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報を取得するiSCSIパラメータ監視ステップと、

前記データリンクを監視しデータリンク層のデータ送信状態情報を取得するデータリンク層監視ステップと、

前記iSCSIコネクションのR2T-PDUに含まれるデータ送信サイズの情報と前記データリンク層のデータ送信状態情報に基づいてSCSIコネクションのデータブロックサイズを決定するブロックサイズ決定ステップ

の各ステップを制御部に実行させることを特徴とするネットワークシステムの制御プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、第1の装置から第2の装置にデータを送信するネットワークシステムに関するものであり、詳しくは、ホストコンピュータからネットワーク内に設置されたハードディスク等のストレージ装置に書き込むデータを効率よく送信するネットワークシステムとネットワークシステムの制御方法に関するものであり、iSCSI(Internet Small Computer

10

20

30

40

50

System Interface)を介してイニシエータであるホストコンピュータから、ターゲットのストレージ装置に書き込むデータを効率よく送信するネットワークシステムとネットワークシステムの制御方法を提供するものである。

【背景技術】

【0002】

iSCSIは、ホストコンピュータとストレージ装置などの周辺機器を接続する規格であるSCSI(Small Computer System Interface)コマンドをTCP/IP(Transmission Control Protocol/Internet Protocol)でカプセル化する規格である。TCP/IPネットワーク上のホストコンピュータと周辺機器をiSCSIに対応させることにより、ホストコンピュータはSCSIコマンドを利用してTCP/IPネットワーク上に設置されたストレージ装置などの周辺機器へアクセスすることが可能になる。 10

図13(a)は、iSCSI層を有するネットワークの階層モデルであり、イニシエータ501のアプリケーション502が、TCP/IPネットワーク内のターゲット511のストレージ装置512の間でデータの送受信521を行うものである。また、図13(b)は、アプリケーション502がデータ531をストレージ装置512に書き込む際、各階層で実施されるデータ変換の概要を示す図である。

アプリケーション502がストレージ装置512にデータの書き込みをイニシエータ501の制御部(例えば、オペレーティングシステム(Operating System, OS))に指示する。イニシエータ501の制御部は、ネットワーク507を介してSCSI層503、iSCSI層504、TCP/IP層(TCP層とIP層からなる)505の各層と、ターゲット511のSCSI層513、iSCSI層514、TCP/IP層515と論理的な接続路(コネクション)を形成し、下記の処理を実行し指示する。 20

【0003】

(1): イニシエータ501の制御部は、アプリケーション502により依頼された書き込みデータ531をSCSI層503に渡す。

(2): SCSI層503は、書き込みデータ531をSCSIコネクション522に適合する大きさに分割し、SCSIヘッダーを付加して1つ以上のSCSIデータブロック532を生成する。生成されたSCSIデータブロック532は順次iSCSI層504に渡される。

(3): iSCSI層504は、SCSI層503から渡されたSCSIデータブロック532を所定の大きさに分割し、iSCSIヘッダーを付加する等の処理を行い、iSCSIデータカプセル(iSCSI-PDU: iSCSI Protocol Data Unit)533を生成する。生成されたiSCSI-PDU533は順次TCP/IP層505に渡される。 30

(4): TCP/IP層505のTCP層は、iSCSI層504から渡されたiSCSI-PDU533を所定の大きさに分割し、TCPヘッダーを付加しTCPパケットを生成する。また次のIP層はTCPパケットにIPヘッダーを付加しIPパケットを生成し、データリンク層制御部506に渡す。ここでデータリンク層制御部とは具体的にはイーサネット(登録商標)、インフィニバンド、SONET/SDHなどのデータリンク層通信プロトコルの制御部を意味する。

データリンク層制御部506は所定の手順に従ってターゲット511のデータリンク層制御部516と通信し、TCP/IPパケット534を順次送信する。ターゲット511のデータリンク層制御部516、TCP/IP層515、iSCSI層514、SCSI層513はイニシエータ501の各層と逆の処理を行い、受信したデータをストレージ装置512に書き込む。 40

【0004】

イニシエータとターゲットが地理的に離れた別の2地点に設置されていた場合には、パケットが往復に要する時間(ラウンドトリップタイム R T T)が大きくなるため、後述するように通信性能をあげるために様々な工夫が必要である。

SCSI層における標準的な読み取りプロセスは図15のとおりである。はじめにSCSI読み出しコマンドをターゲットに送信する。SCSI読み出しコマンドには一度に読み出すデータのサイズ(ブロックサイズ)が指定されており、ターゲットは次のSCSI読み出しコマンドを受信するまで待たないと、次のデータを送信することはできない。従ってブロックサイズを大きくすればより効率的にデータの送信が可能になる。 50

広域ネットワークシステム(TCP)にも、長距離通信におけるデータの送信速度を向上させるための多くの施策がなされている。その1つに、複数のデータセグメントを送信する時、送信したデータセグメントに対する応答を待つことなく次のデータセグメントを連続して送信する方法がある。連続した送信が可能なデータセグメントの数は、送信先が一度に受け付け可能なデータセグメントの数(広告ウィンドウサイズ)と、ネットワークが処理可能なデータセグメントの数(輻輳ウィンドウサイズ)により決まる。長距離通信においてデータ送信の効率を高めるには、広告ウィンドウサイズの範囲内で輻輳ウィンドウサイズを大きな値とする必要がある。しかし、あまりにも大量のデータを送信すると、ネットワークにデータがあふれてしまい輻輳が生じて送信効率が低下する。このため、利用可能な輻輳ウィンドウサイズを、ネットワークの性能、ネットワーク内に滞留しているパケットの量等により動的に変化させて最適化する手法(輻輳制御アルゴリズム)が用いられる。

10

輻輳ウィンドウサイズを制御する技術、あるいはTCPコネクションやSCSIプロトコルを最適化する技術文献には下記の文献がある。

【非特許文献1】豊田真知子, 山口実靖, 小口正人"iSCSIアクセス時のTCP輻輳ウィンドウ制御を用いたシステム性能向上手法の一検討" 社団法人電子情報通信学会 Vol.104, No.537(20041213) pp. 1-6

【非特許文献2】山口実靖, 小口正人, 喜連川優"iSCSI解析システムの構築と高遅延環境におけるシーケンシャルアクセスの性能向上に関する考察" 社団法人電子情報通信学会 Vol.J87-D-I, No.2(20040201) pp. 216-231

20

【非特許文献3】藤原啓成, 若宮直紀, 志賀賢太"広域 IP 網を介した iSCSI 通信におけるプロトコルチューニングの一検討" 社団法人情報処理学会 1-55 5A-5

【非特許文献4】岡廻 隆生, 阿部 隆裕, 芦萱 吉喜, 岡本 隆行"ファイバチャネルを用いた遠隔大容量ファイル共有システム" 社団法人電子情報通信学会 信学技報 Vol.107, No.175 (20080802) pp71-76

【発明の開示】

【発明が解決しようとする課題】

【0005】

従来のiSCSIを用いた通信プロセスでは上記のように各レイヤ(上記の例では、SCSI層、TCP/IP層)ごとに最適な送信サイズ(SCSI層におけるブロックサイズ、TCP/IP層における輻輳ウィンドウサイズ)が決めることができるが、SCSI層とTCP/IP層の間の整合性がないために、かえって送信効率が低下することがある。例えば、SCSI層であまりに大きなブロックサイズを指定すると、ネットワーク上にあまりに多くのデータが流入し輻輳を引き起こす。これを避けるためにTCP/IP層で輻輳制御アルゴリズムが動的に輻輳ウィンドウサイズを小さくすると伝送効率がかえって低下する。

30

【0006】

上記のSCSI層とTCP/IP層の不整合による伝送効率劣化の具体例を示す。図14において、送信元からのTCPパケットを受信した送信先は、当該TCPパケットに対するAck(以下Ack)を返信する。送信元が所定の時間内にAckを受信すると、当該データ送信は正常に行われたと判断し、次のTCPパケットを送信する。送信TCPパケットの喪失等の理由によりデータ送信が正常に行われず、Ackが所定の時間内に返信されなかった場合、送信元は当該TCPパケットを再送する等のエラー処理を行う。

40

送信元においてAckが正常に受信された場合、ネットワークの送信許容量に余裕があり、輻輳ウィンドウサイズの増加が許容される可能性がある。送信元においてAckが正常に受信された場合、当該Ackの受信に応答して次のTCPパケットを送信する時、Ackを待つことなく複数のTCPパケットを連続して送信する、図14(a)に示す方式が採用されている。

【0007】

図14(a)に示される例では、送信元のTCP551は送信先のTCP552に1つのTCPパケット554を送信する。送信先552はパケット554を正常に受信したことを示すAck555を送信元551

50

に返信する。送信元551はAck555を所定の応答時間内の時間 [T1] 553に受信したことに応答して、2つのTCPパケット556を連続して送信する。送信先552は2つのパケット556を正常に受信したことを示すAck558を送信元551に返信する。送信元551はAck558に応答して、3つのTCPパケット560を連続して送信する。以降、送信元551から送信先552へのTCPパケットの送信が正常に実行される毎に、送信元551は送信先552に連続して送信するTCPパケットの数、即ち輻輳ウィンドウサイズを1ずつ増加させる。

【 0 0 0 8 】

図 1 4 (b) は送信元551が実行する輻輳ウィンドウサイズの制御を示す図である。同図のネットワークはn個を越えるTCPパケットが連続して送信されると輻輳状態となり、TCPパケットの破棄等の障害が発生するものとする。

10

時刻「t0」で、送信元551は送信先552にTCPパケットの送信を開始する。送信期間562の確認応答期間が [T1] 553、 [T2] 557、 [T3] 559と進むに従い、輻輳ウィンドウサイズは1ずつ増加する。

時刻「t1」で輻輳ウィンドウサイズが「n」となり、送信元551はn個のTCPパケットを連続して送信する。この送信はネットワークの許容量561を越え、ネットワークは輻輳状態となり、TCPパケットの破棄等の障害が発生する。この結果、送信元551は所定の時間内にAckを受信することができず、ネットワークは輻輳状態にあると判断する。当該判断に基づいて、送信元551のTCP制御部は輻輳ウィンドウサイズを「1」にリセットする(564)。当該リセット処理により、時刻「t1」の次の輻輳ウィンドウサイズは「1」となり、1つのTCPパケットが送信され、送信に対する応答を待つ。送信期間562と同様、送信期間565、及び送信期間566においても、輻輳ウィンドウサイズは「1」から始まり、1ずつ増加し、「n」となった時リセットされる動作が繰り返される。

20

【 0 0 0 9 】

図 1 4 (b) の例では、TCPパケットの送信に利用可能な輻輳ウィンドウサイズは「n-1」である。しかし、SCSI層でのブロックサイズが大きすぎるとTCPパケット送信に利用可能なサイズ以上のデータが流入するため、輻輳状態が発生すると輻輳ウィンドウが小さくなり伝送効率が低下する。従って、図 1 4 (b) に示すようなSCSI層とTCP層の整合性がない制御方法では、輻輳状態が頻繁に発生し、その度に輻輳ウィンドウサイズが「1」にリセットされるため、ネットワークの送信能力を充分に利用しておらず、大きな送信能力を得ることができない。上記問題に対応するため、輻輳ウィンドウサイズの増加量を大きくする、あるいは、輻輳ウィンドウサイズをリセットする際、リセット値を「1」とせず、例えば「n/2」とし、輻輳ウィンドウサイズを大きく維持する等の構成が提案されている。しかし、いずれの構成であっても、SCSI層におけるブロックサイズが大きいままであれば、ネットワークに許容量以上のデータが流入して輻輳状態となる、あるいはネットワークに流入する量を調整する送信側バッファ (TCP層の送信バッファあるいはデータリンク層における送信バッファ) がオーバーフローを起こすため、伝送効率をあげることができない。

30

【 0 0 1 0 】

上記文献 1 には、上記の問題を解決するためイニシエータがiSCSIを介してターゲットのストレージ装置からデータを読み取る構成において、SCSI層のブロックサイズをTCP層に適合した大きさに動的に制御する技術が記載されている。

40

図 1 5 は、イニシエータのアプリケーションがiSCSIにより接続されているターゲットのストレージ装置からデータを読み取る時のSCSI層とiSCSI層の制御コマンドとデータの流れの概要を示す図である。

イニシエータ501のアプリケーション502(図 1 3) が、ターゲット511のストレージ装置512(図 1 3) からデータの読み取りを制御部(図 1 6 の制御部581)に依頼すると、制御部はSCSI503にターゲット511のSCSI513を介してストレージ装置512からデータを読み取ることを指示する。当該指示に従って、SCSI及びiSCSIは下記の処理を実行する。

【 0 0 1 1 】

ステップ571 : SCSI503はストレージ装置512からブロックデータを読み取るコマンドをi

50

SCSI504に渡す。

ステップ572：iSCSI504はSCSI503から渡されたSCSIコマンドをカプセル化し、TCP/IPパケットとしてiSCSI514に送信する。

ステップ573：iSCSI514はTCP/IP層を介して受信したiSCSIカプセルからSCSIコマンドを取り出し、SCSI513に渡す。

ステップ574：SCSI513は当該読み取りコマンドを実行し、ストレージ装置512からブロックデータを読み取りiSCSI514に渡す。

ステップ575：iSCSI514は上記のブロックデータをカプセル化してTCP/IP層515に渡す。TCP/IP層515は図14に示される輻輳ウィンドウサイズの制御を実行しつつ、カプセル化されたブロックデータをイニシエータ501のTCP/IP層505を介してiSCSI504に送信する。

10

ステップ576, 577：ブロックデータの送信が全て終了すると、SCSI513は終了レスポンスを上記と同様にiSCSI層、TCP/IP層を介してSCSI503に送信する。

【0012】

図16(a)は上記文献1に示されるネットワークの階層構成の概要である。同文献に記載される技術は、イニシエータ501の制御部581が、SCSI503とSCSI513の間で授受されるデータのサイズ(SCSI層のブロックサイズ)を輻輳ウィンドウサイズのリセットが発生しない大きさとするものである。図16(b)は、SCSI503が処理するデータブロックサイズを変更する機能を示す図である。また、図17(a)は制御部581により実行されるデータブロックサイズを変更する動作フローの図であり、図17(b)は輻輳ウィンドウサイズの変化を示す図である。

20

【0013】

ターゲット511のTCP/IP層515に設置されている輻輳ウィンドウモニタ部585は、例えば文献2に示されるTCPの監視機能に相当するものであり、TCP/IP層515のTCP層が実行中の輻輳ウィンドウサイズを監視し、イニシエータ501の制御部581に通知する機能を有している。

イニシエータ501の制御部581は、アプリケーション502がターゲット511のストレージ装置512からデータの読み込みを要求したことに応答して、SCSI503に所定のブロックサイズによるデータの読み込みを指示する。

制御部581の輻輳ウィンドウサイズ変化検出部586は、ターゲット511の輻輳ウィンドウモニタ部585から送信される輻輳ウィンドウサイズの情報を監視し、輻輳ウィンドウサイズが減少したことを検出すると、ネットワークへのデータ流入量を小さくして輻輳状態になるのを防ぐためブロックサイズ指定/変更部588を介してSCSI503にブロックサイズを小さくすることを指示する。

30

【0014】

図17(a)はSCSIのブロックサイズを小さくし、輻輳ウィンドウサイズのリセットの発生を防止する動作フローである。また、図17(b)は当該動作フローによる輻輳ウィンドウサイズの変化図である。

ステップ601：イニシエータ501のアプリケーション502が、OS等の制御部581にターゲット511のストレージ装置512からデータの読み取りを要求すると、制御部581はSCSI503にターゲット511のSCSI513を介してストレージ装置512からデータを読み取ることを指示する。SCSI503、iSCSI504、TCP/IP505は各々SCSI513、iSCSI514、TCP/IP515と論理的な接続(コネクション)を確立しているものとする。

40

ステップ602：制御部581は、SCSI503に所定のブロックサイズ、例えば1024キロバイトのブロックサイズによるデータの読み込みを指示する。データの読み込みを指示するSCSIコマンドはiSCSI504によりカプセル化され、TCP/IPプロトコルに従ってTCP/IP515に送信され、iSCSI514を介してSCSI513に渡される。SCSI513はストレージ装置512から1024キロバイトのデータを読み出し、iSCSI514を介してTCP/IP515に渡す。TCP/IP515は、図14に示される輻輳ウィンドウサイズの制御を実行しつつ、TCP/IP505にTCPパケットに変換されたデータを送信する。

ステップ603：輻輳ウィンドウサイズ変化検出部586はターゲット511の輻輳ウィンドウ

50

モニタ部585から送信される輻輳ウィンドウサイズの情報監視する。図17(b)に示されるように、輻輳ウィンドウサイズは順次増加し、時刻「t1」で、ネットワークの許容量607を超える。輻輳ウィンドウサイズがネットワークの許容量607を超えたことにより発生するパケットの破棄等の障害により、ネットワークは輻輳状態にあると判断したTCP/IP515の制御部は輻輳ウィンドウサイズをリセットする。当該リセットにより、輻輳ウィンドウサイズが減少したことを検知した輻輳ウィンドウサイズ変化検出部586はステップ603の判定を「Yes」とし、処理をステップ604に進める。

ステップ604：輻輳ウィンドウサイズの低下を検知した輻輳ウィンドウサイズ変化検出部586からの情報により、ブロックサイズ指定/変更部588はSCSI層のデータブロックサイズを所定量、例えば64キロバイト小さくすることを決定し、960(1024-64)キロバイトのブロックサイズによるデータの読み込みをSCSI503に指示し、ステップ603に戻る。また、当該指示に基づいて、SCSI503はターゲット511のSCSI513に960キロバイトのブロックサイズによるデータの送信処理を指示する。

【0015】

時刻「t1」でリセットされた輻輳ウィンドウサイズは、TCP/IP515が実行する輻輳ウィンドウサイズの制御により順次増加する。図17(b)に示される例では、時刻「t2」で輻輳ウィンドウサイズがネットワークの許容量607を超え、輻輳ウィンドウサイズは再度リセットされる。輻輳ウィンドウサイズ変化検出部586が輻輳ウィンドウサイズの減少を検知することにより、ステップ604においてブロックサイズは64キロバイト小さくされ、896キロバイトに設定される。

時刻「t2」でリセットされた輻輳ウィンドウサイズは、TCP/IP515による輻輳ウィンドウサイズの制御により順次増加する。しかし、図17(b)に示される例では、SCSI層でのブロックサイズを小さくすることによってネットワークへのデータ流入量が小さくなったため、輻輳ウィンドウサイズ608はネットワークの許容量607に達していない。従って、制御部581は引き続き896キロバイトのデータの送信を指示し、TCP/IP515は輻輳ウィンドウサイズ608(図17(b))によるデータの送信を実行する。

以上のように、輻輳ウィンドウサイズの減少が検知される毎に、SCSIが処理するデータのブロックサイズを小さくする処理を繰り返すことにより、ブロックサイズはネットワークの許容量を越えない最適な輻輳ウィンドウサイズによる送信が可能な大きさとなり、安定した送信が可能となる。

【0016】

文献1に記載される構成は、イニシエータ501のアプリケーション502がSCSI及びiSCSIを介してターゲット511のストレージ装置512からデータを読み込む処理に対応する構成である。図15に示されるように、iSCSIを介してターゲットからデータを読み込む動作では、ターゲット511のiSCSI514は読み出しコマンド572により要求されたデータを連続してイニシエータ501のiSCSI504に送信する。従ってiSCSI層は下位の層であるTCP層や上位の層であるSCSI層の動作を乱すことはない。

しかし、イニシエータ501からターゲット511のストレージ装置512にデータを書き込む場合には、iSCSI層においても、TCP層やSCSI層と同様に決められたデータ単位ごとにR2T-PDU(Ready to Transfer-PDU)と呼ばれるパケットの受信を待って次のデータを送出する必要がある。R2T-PDUとはiSCSI層においてターゲット側でデータの受信準備ができたことを示すパケットであり、TCP層のAckや、SCSI層のSCSIレスポンスのような確認パケットに類似した役割をiSCSI層にて果たすものである。iSCSI層におけるデータ送出手の単位は、iSCSIコネクション確立時に設定されるパラメータMaxBurstLengthを超えることはできない。つまり、アプリケーション502が書き込みを要求したデータは、SCSI層のブロックサイズの大きさに関わらず上記MaxBurstLengthにより指定された量より小さいデータに分割されてターゲット511に送信される。図18は、イニシエータ501からターゲット511のストレージ装置へデータを書き込むシーケンスの概要を示す図である。

【0017】

イニシエータ501のアプリケーション502がストレージ装置512にデータの書き込みを要

10

20

30

40

50

求すると、制御部581はSCSI503にターゲット511のSCSI513を介してストレージ装置512にデータを書き込むことを指示する。

図18のステップ621で、SCSI503はiSCSI504に書き込みコマンドを渡す。iSCSI504は書き込みコマンドをカプセル化しターゲット511のiSCSI514に送信する(ステップ622)。

ストレージ装置512によるデータの書き込みの準備が完了すると、ターゲット511の制御部は一回の受信動作により受け取り可能なデータ長の情報を含むR2T-PDUをiSCSI504に送信する(以降、当該データ長を「R2T」とする。「R2T」は特にストレージ装置に問題がない限り前述のMaxBurstLengthと同じ値である(ステップ623))。

R2T-PDUを受信したiSCSI504は、R2T-PDUの「R2T」により指定された長さのデータをターゲット511に送信する(ステップ624)。送信データの処理を終了したターゲット511は次のデータを要求するR2T-PDUをiSCSI504に送信する(ステップ625)。その後、全てのデータの送信が終了するステップ628まで、ターゲット511がイニシエータ501にR2T-PDUを送信するステップと、R2T-PDUに応答してイニシエータ501が指定された「R2T」長のデータをターゲット511に送信するステップを繰り返す。

最後にステップ631とステップ632によりターゲット511からイニシエータ501にSCSIレスポンスが送信される。つまりiSCSIを介してTCP/IPネットワークに接続されているターゲットのストレージ装置にデータを書き込む場合、書き込むデータは一回の送信動作によりに授受可能なデータ量「R2T」に分割し、複数回に分けて送信される。

ターゲットからイニシエータにデータを読み込む場合には、SCSI層とTCP/IP層のパラメータを動的に変化させることにより長距離伝送時のネットワークの送信効率を上げることが可能であったが、イニシエータからターゲットにデータを書き込む場合には、上記のようにSCSI層とTCP/IP層のパラメータの制御だけでは不十分であり、SCSI層のパラメータ、TCP/IP層のパラメータに加えて、iSCSI層のパラメータを考慮して制御を行う必要がある。

【課題を解決するための手段】

【0018】

本発明は、iSCSIにより接続されたターゲットのストレージ装置等へ書き込みデータを効率よく送信することを可能とするものであり、SCSI層のパラメータ(ブロックサイズや並列に実行するシーケンス数)、TCP/IP層のパラメータ(輻輳ウィンドウサイズ)に加えて、iSCSI層のパラメータ(MaxBurstLength)を考慮して制御を行うことを特徴とする。

本発明の第1の実施例では、イニシエータのiSCSI部にターゲットのiSCSIから送信されるR2T-PDUを監視し、iSCSIの動作パラメータを取得する「iSCSIパラメータ監視手段」と、イニシエータのTCP制御部に輻輳ウィンドウサイズ、あるいはTCP層のバッファの状態を監視する「TCPプロトコル監視手段」を有している。また、制御部に「TCPプロトコル監視手段」の情報に基づいてTCPプロトコルの障害を検知する「TCP障害検知手段」を有し、「iSCSIパラメータ監視手段」と「TCP障害検知手段」の出力に基づいてSCSI層に指示する送信データのブロックサイズを決定する「ブロックサイズ決定部」および、決定されたブロックサイズの変更をSCSI層に指示する「ブロックサイズ変更部」を有している。

本発明の第2の実施例では、上記「TCPプロトコル監視手段」、「TCP障害検知手段」の代わりに、データリンク層のバッファの状態を監視する「データリンク層監視手段」、「データリンク層障害検知手段」を有し、上記「iSCSIパラメータ監視手段」と「データリンク層障害検知手段」の出力に基づいてSCSI層に指示する送信データのブロックサイズを決定する「ブロックサイズ決定部」および、決定されたブロックサイズの変更をSCSI層に指示する「ブロックサイズ変更部」を有している。

本発明の第3の実施例では、第1の実施例における「ブロックサイズ決定部」の代わりに、SCSI層のブロックサイズだけでなくSCSIレスポンスを待たずに並列に実行するSCSIシーケンスの数(以下「並列実行シーケンス数」と呼ぶ)を決定し変更を指示する「ブロックサイズ・並列実行シーケンス数決定部」を有している。

本発明の第4の実施例では、iSCSI層の制御部に「TCP障害検知手段」の出力に基づいてiSCSIコネクションの動作パラメータを変更する機能を備えている。当該機能は、イニシ

10

20

30

40

50

エータとターゲット間のMaxBurstLengthの設定値を変更するために、iSCSIコネクションを再構成する機能を有しているため、R2T-PDUが指定する送信データのサイズを大きくすることを可能とする。

【実施例 1】

【0019】

図1-1は、本発明の第1の実施例に係るネットワークの構成を示す図であり、イニシエータ100の制御部102はブロックサイズ変更手段128を有している。アプリケーション101がターゲット110のストレージ装置111にデータを書き込む際、ブロックサイズ変更手段128はiSCSIプロトコルのパラメータとTCPプロトコルの状態に基づいて、SCSI層103に送信を指示するデータのブロックサイズを変更する。また、図1-2は上記機能の詳細を示す図である。同図のブロックサイズ決定部134とブロックサイズ変更部133が図1-1のブロックサイズ変更手段128に相当する。TCP/IP層に設けられたTCPプロトコル監視手段126はTCP層の状態、例えばTCP層のバッファの状態、あるいは輻輳ウィンドウサイズに関する情報をTCPプロトコル状態情報127.2として制御部102のTCP障害検知手段131に渡す。TCP障害検知手段131はTCPプロトコル監視手段が送信した情報を解析し、TCP層105、データリンク層106、およびネットワーク107の状態を判別する。

【0020】

TCPプロトコル監視手段の具体的な例としては以下のようなものがある。ネットワーク内に設定されたTCPコネクションによる通信が正常に実行されている時、TCPの制御部は輻輳ウィンドウサイズを順次増加させる。また、TCPの制御部がTCPパケットに対する応答がタイムアウトしたことによりTCPコネクションが輻輳状態となったことを検知すると、輻輳ウィンドウサイズをリセットする。従って、輻輳ウィンドウサイズを監視することによりネットワーク107にて輻輳が発生しているか否かの判別が可能となる。

また、ネットワークのデータ送信が遅滞するとTCP層からデータリンク層へのデータの送信やデータリンク層からネットワークへの送信も遅滞する。このような状態の時、上位の層であるiSCSIから送信すべきデータが送られるとTCP層やデータリンク層の送信バッファはオーバーフローし輻輳ウィンドウがリセットされる。従って、TCP層の輻輳ウィンドウサイズをモニターすることにより、間接的にTCP層やデータリンク層の送信バッファがオーバーフローしているかどうかの状況を監視することが可能となる。

以下では、TCPプロトコル監視手段126によりTCP層105のバッファを監視する構成を例に本発明を説明する。

【0021】

iSCSIパラメータ監視手段125は、ターゲット110のiSCSI114から送信されるR2T-PDUのパラメータを監視する。図1-8に示される様に、イニシエータ100のiSCSI104からターゲット110のiSCSI114にカプセル化されたSCSI書き込みコマンド121が送信されると、ターゲット110の制御部は、データの受信が可能となったことを示すR2T-PDU122をイニシエータ100のiSCSI104に送信する。R2T-PDU122には次のデータ送信において、1回の送信動作により送信するデータ量「R2T」に関する情報が格納されている。iSCSIパラメータ監視手段125はiSCSI114から送信されたR2T-PDU122を監視し、「R2T」の情報をiSCSIパラメータ127.1としてブロックサイズ決定部134に渡す。

【0022】

アプリケーション101がターゲット110のストレージ装置111に所定量のデータの書き込みを制御部102に指示すると、データ書き込み処理部132はSCSI103にデータの書き込みを指示する。SCSI103はブロックサイズ変更手段128に指示されたブロックサイズにて前記データを処理しiSCSI104に渡す。iSCSI層104はターゲット110のiSCSI層114からR2T-PDU122が送信されたことに応答して、「R2T」により指定された量のデータをカプセル化してTCP/IP層105に渡す。

【0023】

ブロックサイズ決定部134は、TCP障害検知手段131の出力とiSCSIパラメータ監視手段125がR2T-PDUから抽出した「R2T」に基づいてSCSI103のデータブロックサイズを決定する。

ブロックサイズ変更部133はブロックサイズ決定部134が決定したデータブロックサイズをSCSI層に指示する。

図2は、TCP層のバッファの状態を監視するTCPプロトコル監視手段126の出力と、iSCSIパラメータ監視手段125が取得した「R2T」に基づいて、ブロックサイズ決定部134が実行するデータブロックサイズの決定フローであり、図3はブロックサイズ決定部134がブロックサイズ変更部133を介してSCSI103に指示するデータのブロックサイズの変化を示す図である。

【0024】

ステップ141：アプリケーション101からデータ送信要求が出されると、SCSI層、iSCSI層、TCP/IP層の接続を通じて、上記データ送信が開始される。

ステップ142：iSCSI層104のiSCSIパラメータ監視手段125は、ターゲット110のiSCSI層114から送信されたR2T-PDUを解析し、1回のデータ送信によりiSCSI層104からiSCSI層114に送信される書き込みデータの大きさ「R2T」を取得する。

ステップ143：制御部102はSCSI層103のデータブロックサイズを初期化する。

ステップ144：ブロックサイズ決定部134は現在のデータブロックサイズがステップ142で取得した「R2T」より大きいのか否かを判断する。1回の送信によりiSCSI層104からiSCSI層114に送信される書き込みデータの最大値は「R2T」である。SCSI層103におけるデータブロックサイズが「R2T」より大きいと、SCSI層から受取ったデータを「R2T」で指定されたサイズに分割されてしまうため、これ以上ブロックサイズを大きくしても意味がないのでステップ147に進む。一方、データブロックサイズが「R2T」より小さい場合は、SCSI層のデータブロックサイズを大きくすることにより、iSCSI層でのデータの分割は行われないので、データ転送効率を増加させる可能性があるため、ステップ145に進む。

【0025】

ステップ145：制御部102のブロックサイズ決定部134はSCSI層103に指示するデータブロックサイズを所定量増加する。SCSI層103は指示されたデータブロックサイズにてデータを処理し、iSCSI層を介してTCP/IP層105に渡す。

ステップ146：TCP/IP層105はネットワークを介してTCP/IP層115へデータを送信する。ネットワークが輻輳状態にあり、TCP/IP層105からネットワークへのパケットの送信が滞ると、TCP/IP層105が有するバッファがオーバーフローする。TCP/IP層105のTCPプロトコル監視手段126は、上記バッファを監視し、制御部102のTCP障害検知手段131にバッファの状態を通知する。

TCP障害検知手段131がバッファはオーバーフローすることなくデータ送信が正常に行われていると判断すると、ステップ144に戻り、データブロックサイズの更なる増加が可能か否かの判別処理を行う。

TCP障害検知手段131がバッファのオーバーフローを検知すると、ステップ147に進み、オーバーフローを起こさないデータブロックサイズによるデータの送信を行う。

ステップ147：データブロックサイズの最適化処理は終了し、決定されたデータブロックサイズによるデータ送信を行う。

【0026】

図3(a)は、R2T-PDUで指定されるデータサイズ「R2T」152が、バッファがオーバーフローする値151より大きい場合のデータブロックサイズの変化の一例を示す図である。

t0：図2のステップ143で、制御部102はSCSI層103に指示するデータブロックサイズを初期値150に設定する。

t1～t3：データブロックサイズは「R2T」152より小さく、且つTCP/IP層でバッファオーバーフローは発生しないため、ステップ144乃至ステップ146を順次実行することによりデータブロックサイズは所定量ずつ増加する。

t4：TCP障害検知手段131により、TCP/IP層105のバッファオーバーフローが検知されると、ブロックサイズ決定部134はデータブロックサイズをバッファがオーバーフローしない値、即ち一段階前の値153に設定し、ブロックサイズ変更部133を介してSCSI層103に指示を出す。

10

20

30

40

50

t5～：ブロックサイズ決定部134が決定したデータブロックサイズ153に従ってSCSI層103はデータを送出する。

【0027】

図3(b)は、R2T-PDUにより指定されるデータサイズ「R2T」162が、バッファがオーバーフローする値161より小さい場合のデータブロックサイズの変化の例を示す図である。

t0：図2のステップ143で、制御部102はSCSI層103に指示するデータブロックサイズを初期値160に設定する。

t1～t3：データブロックサイズは「R2T」162より小さく、且つTCP/IP層でバッファオーバーフローは発生しないため、ステップ144乃至ステップ146を順次実行することにより、データブロックサイズを所定量ずつ増加する。

t4：データブロックサイズが「R2T」162より大きくなったことを検知したブロックサイズ決定部134は、データブロックサイズを所定量小さくし、「R2T」162を越えない値、即ち一段階前の値163とする。

t5～：ブロックサイズ決定部134が決定したデータブロックサイズ163に従ってSCSI層103はデータを送出する。

【0028】

図1-1乃至図3に示されるTCPプロトコル監視手段126は、TCP制御部のバッファの状態を監視し、TCP障害検知手段131はバッファがオーバーフローしたことによりTCPコネクションが輻輳状態となったと判断する構成である。TCPプロトコルの状態を監視する手段として図16に示される輻輳ウィンドウサイズの変化を監視する構成を用いることが可能である。当該構成を用いた場合、図2のステップ146で、輻輳ウィンドウサイズが減少した時にTCPコネクションが輻輳状態となったと判断する。また、この時の輻輳ウィンドウサイズからTCPコネクションが送信可能なデータ量を算定し、ブロックサイズ決定部134によるデータブロックサイズの決定に用いることが可能である。

【0029】

本発明は、iSCSIにより接続されているターゲット110のストレージ装置111にイニシエータ100からデータを書き込む場合、高いデータ送信効率を可能とするストレージネットワークシステムとストレージネットワークシステムの制御方法、及び制御プログラムを提供するものである。

本発明の第1の実施例は、TCPコネクションの状態をTCPプロトコル監視手段により監視すると共に、iSCSIコネクションのパラメータをiSCSIパラメータ監視手段により監視する構成を有している。イニシエータ100の制御部102は両監視手段からの情報に基づいてSCSI層におけるデータブロックサイズを決定する。当該構成により、iSCSI間のデータ送信において、TCP層でのバッファオーバーフロー、データリンク層でのバッファオーバーフロー、ネットワークの輻輳などの問題を回避しつつ高い送信効率を確保することが可能となる。

【実施例2】

【0030】

図4-1は、本発明の第2の実施例に係るネットワークの構成を示す図である。実施例1においては、図1-1に示すとおりTCP/IP層に設けられたTCPプロトコル監視手段126によりTCP層のバッファを監視する構成を示したが、実施例2においては、図4-1のように、データリンク層にデータリンク層監視手段135を設け、データリンク層のバッファを監視する構成とする点で実施例1と異なる。

iSCSIを用いたデータ送信で考えられる障害の代表的なものとして、前記のとおりネットワークの輻輳、TCP層でのバッファオーバーフロー、データリンク層でのバッファオーバーフローなどが考えられるが、広帯域でパケット損失の少ない高品質のネットワークにおいては、データリンク層でのバッファオーバーフローが伝送劣化の主因となることが多い。このような場合、図4-1に示すように、データリンク層に監視手段を設けることで、データリンク層でのバッファオーバーフローを直接監視することが可能であり、実施例2はこのような構成を実現したものである。

図4-2は上記機能の詳細を示す図である。データリンク層に設けられたデータリンク層監視手段135はデータリンク層の状態、例えばデータリンク層のバッファの状態に関する情報をデータリンク層状態情報127.3として制御部102のデータリンク層障害検知手段137に渡す。データリンク層障害検知手段137はデータリンク層監視手段が送信した情報を解析し、データリンク層106の状態を判別する。

図4-1と図4-2において、図1-1と図1-2に示される構成と同等の構成要素には同一の参照番号が付されている。

【0031】

図5は、データリンク層のバッファの状態を監視するデータリンク層監視手段135の出力と、iSCSIパラメータ監視手段125が取得した「R2T」に基づいて、ブロックサイズ決定部138が実行するデータブロックサイズの決定フローであり、図6はブロックサイズ決定部138がブロックサイズ変更部139を介してSCSI103に指示するデータのブロックサイズの変化を示す図である。

10

【0032】

ステップ171：アプリケーション101からデータ送信要求が出されると、SCSI層、iSCSI層、TCP/IP層の接続を通じて、上記データ送信が開始される。

ステップ172：iSCSI層104のiSCSIパラメータ監視手段125は、ターゲット110のiSCSI層114から送信されたR2T-PDUを解析し、1回のデータ送信によりiSCSI層104からiSCSI層114に送信される書き込みデータの大きさ「R2T」を取得する。

ステップ173：制御部102はSCSI層103のデータブロックサイズを初期化する。

20

ステップ174：ブロックサイズ決定部138は現在のデータブロックサイズがステップ172で取得した「R2T」より大きいかなかを判断する。1回の送信によりiSCSI層104からiSCSI層114に送信される書き込みデータの最大値は「R2T」である。SCSI層103におけるデータブロックサイズが「R2T」より大きいと、SCSI層から受取ったデータを「R2T」で指定されたサイズに分割されてしまうため、これ以上ブロックサイズを大きくしても意味がないのでステップ177に進む。一方、データブロックサイズが「R2T」より小さい場合は、SCSI層のデータブロックサイズを大きくすることにより、iSCSI層でのデータの分割は行われないので、データ転送効率を増加させる可能性があるため、ステップ175に進む。

【0033】

ステップ175：制御部102のブロックサイズ決定部138はSCSI層103に指示するデータブロックサイズを所定量増加する。SCSI層103は指示されたデータブロックサイズにてデータを処理し、iSCSI層を介してTCP/IP層105に渡す。

30

ステップ176：TCP/IP層105は、データリンク層106、ネットワーク107、データリンク層116を介してTCP/IP層115へデータを送信する。ネットワークが輻輳状態にあり、データリンク層106からネットワーク107へのパケットの送信が滞ると、データリンク層106が有する送信バッファがオーバーフローする。データリンク層106のデータリンク層監視手段135は、上記バッファを監視し、制御部102のデータリンク層障害検知手段137にバッファの状態を通知する。

データリンク層障害検知手段137がバッファはオーバーフローすることなくデータ送信が正常に行われていると判断すると、ステップ174に戻り、データブロックサイズの更なる増加が可能かなかの判別処理を行う。

40

データリンク層障害検知手段137がバッファのオーバーフローを検知すると、ステップ177に進み、オーバーフローを起こさないデータブロックサイズによるデータの送信を行う。

ステップ177：データブロックサイズの最適化処理は終了し、決定されたデータブロックサイズによるデータ送信を行う。

【0034】

図6(a)は、R2T-PDUで指定されるデータサイズ「R2T」182が、バッファがオーバーフローする値181より大きい場合のデータブロックサイズの変化の一例を示す図である。

t0：図5のステップ173で、制御部102はSCSI層103に指示するデータブロックサイズを

50

初期値180に設定する。

t1～t3：データブロックサイズは「R2T」182より小さく、且つデータリンク層でバッファオーバーフローは発生しないため、ステップ174乃至ステップ176を順次実行することによりデータブロックサイズは所定量ずつ増加する。

t4：データリンク層障害検知手段137により、データリンク層106のバッファオーバーフローが検知されると、ブロックサイズ決定部138はデータブロックサイズをバッファがオーバーフローしない値、即ち一段階前の値183に設定し、ブロックサイズ変更部139を介してSCSI層103に指示を出す。

t5～：ブロックサイズ決定部138が決定したデータブロックサイズ183に従ってSCSI層103はデータを送出する。

10

【0035】

図6(b)は、R2T-PDUにより指定されるデータサイズ「R2T」192が、バッファがオーバーフローする値191より小さい場合のデータブロックサイズの変化の例を示す図である。

t0：図5のステップ173で、制御部102はSCSI層103に指示するデータブロックサイズを初期値190に設定する。

t1～t3：データブロックサイズは「R2T」192より小さく、且つデータリンク層でバッファオーバーフローは発生しないため、ステップ174乃至ステップ176を順次実行することにより、データブロックサイズを所定量ずつ増加する。

t4：データブロックサイズが「R2T」192より大きくなったことを検知したブロックサイズ決定部138は、データブロックサイズを所定量小さくし、「R2T」192を越えない値、即ち一段階前の値193とする。

20

t5～：ブロックサイズ決定部138が決定したデータブロックサイズ193に従ってSCSI層103はデータを送出する。

【0036】

本発明は、iSCSIにより接続されているターゲット110のストレージ装置111にイニシエータ100からデータを書き込む場合、高いデータ送信効率を可能とするストレージネットワークシステムとストレージネットワークシステムの制御方法、及び制御プログラムを提供するものである。

本発明の第2の実施例は、データリンク層の状態をデータリンク層監視手段により監視すると共に、iSCSIコネクションのパラメータをiSCSIパラメータ監視手段により監視する構成を有している。イニシエータ100の制御部102は両監視手段からの情報に基づいてSCSI層におけるデータブロックサイズを決定する。当該構成により、iSCSIのデータ送信において、データリンク層のバッファオーバーフローなどの問題を回避しつつ高い送信効率を確保することが可能となる。

30

【実施例3】

【0037】

図7-1は、本発明の第3の実施例に係るネットワークの構成を示す図であり、図7-2は第3の実施例を詳細に示す図である。同実施例は、イニシエータ100のアプリケーション101がiSCSI104を介してターゲット110のiSCSI114に接続されているストレージ装置111にデータを書き込む構成において、iSCSIのパラメータとTCPプロトコルの状態を監視し、それら監視結果に基づいてSCSI層のデータブロックサイズとSCSI層が並列で実行するシーケンスの数を決定するものであり、第3の実施例の制御部102はブロックサイズ・並列シーケンス数決定部202を有し、SCSI層103に並列シーケンス数設定指示204が出力される点で、第1の実施例の構成と相違する。

40

図7-1と図7-2において、図1-1と図1-2に示される構成と同等の構成要素には同一の参照番号が付されている。

【0038】

第1の実施例では、「R2T」で指定されるデータ量以上にブロックサイズを大きくしても、分割されてしまうために、「R2T」を超える量をネットワークに供給することができなかった。

50

参考文献 4 に示されるように、SCSI 層において同時に実行される送信シーケンスの数を増やすことにより、通信の効率を高めることが可能である。

図 7 - 3 は制御部 102 が SCSI 層 103 に同時に実行する送信コマンドを増やす指示を出すことにより、SCSI コネクションが同時に実行する送信シーケンスの数を増やす例である。

図 7 - 3 (a) は、SCSI 層が送信コマンドを 1 つずつ実行する時のシーケンスであり、図 1 8 に示される従来のシーケンスと同じである。SCSI 103 が書き込みコマンド 211 を実行することにより、iSCSI 104 は書き込みコマンドをカプセル化 (211) しターゲット 110 に送信する。その後 iSCSI 104 は、R2T-PDU の受信に回答して「R2T」分の書き込みデータを送信する。すべてのデータの送信が終了し、ターゲット 110 から正常終了を示す SCSI レスポンスを受信すると、SCSI 103 は次の書き込みコマンドを実行する。

10

図 7 - 3 (b) は、SCSI 層が 2 つの送信コマンドを同時に実行する時のシーケンスと同じである。SCSI 103 は第 1 の書き込みコマンド 211.1 の実行に続いて、第 2 の書き込みコマンド 211.2 を連続して実行する。その後イニシエータ 100 とターゲット 110 の間で、2 つの書き込みコマンドに対応する R2T-PDU と書き込みデータの送受信が行われる。従って、図 7 - 3 (b) のシーケンスによる通信量は、図 7 - 3 (a) のシーケンスによる通信量のほぼ 2 倍となる。

第 3 の実施例は上記原理に基づくものであり、SCSI 層においてブロックサイズを増加させるだけでなく、必要であれば同時に実行される書き込みコマンドの数も増やし、「R2T」を超えるデータの量をネットワークに供給することにより、より効率のよい通信を行うことを可能とする。

20

【 0 0 3 9 】

本発明の第 3 の実施例は、動作中の TCP コネクションによる送信データ量が「R2T」を越えた時、制御部は SCSI の制御部に複数の書き込みコマンドを与える等により、SCSI 層が並列に実行するシーケンスの数を増加させる。

SCSI 層は並列実行するシーケンスの数を増加させることにより、R2T-PDU を待たずに送信できるデータ量を多くすることを可能とする。図 7 - 2 に示される構成において、ブロックサイズ・並列実行シーケンス数決定部 202 が iSCSI パラメータ監視手段 125 と TCP 障害検知手段 131 の情報から、動作中の TCP コネクションによる送信データ量が「R2T」により指定される許容量に達したことを検知すると、SCSI 層 103 の制御部に同時に実行する送信コマンドを増加させ、並列に実行するシーケンス数を増加させる指示を出力する。

30

図 8 は、本発明の第 3 の実施例において、ブロックサイズを増やすとともに SCSI コネクションが同時に実行する送信シーケンスの数を増やす構成の動作フローである。

【 0 0 4 0 】

ステップ 231 : アプリケーション 101 からデータ送信要求が出されると、SCSI 層、iSCSI 層、TCP/IP 層のコネクションを通じて、データの送信が開始される。

ステップ 232 : iSCSI 層 104 の iSCSI パラメータ監視手段 125 は、ターゲット 110 の iSCSI 層 114 から送信される R2T-PDU を監視し、1 回のデータ送信により iSCSI 層 104 から iSCSI 層 114 に送信するデータの大きさを示す情報「R2T」を取得する。

ステップ 233 : 制御部 102 は SCSI 層 103 におけるデータブロックサイズを初期化する。また、並列に実行するシーケンス数を、例えば「1」に初期化する。

40

ステップ 234 : ブロックサイズ・並列実行シーケンス数決定部 202 は実行中のデータブロックサイズがステップ 232 で取得した「R2T」より大きいかなかを判断する。データブロックサイズが「R2T」を越える場合はステップ 236 に進み、「R2T」を越えない場合はステップ 235 に進む。

ステップ 235 : データブロックサイズが「R2T」を越えていない場合、データブロックサイズを大きくすることにより、送信効率が上がる可能性がある。従って、ブロックサイズ・並列実行シーケンス数決定部 202 はブロックサイズ変更部 203 に指示し、SCSI 層 103 におけるデータブロックサイズを所定量増加させることを指示する。

【 0 0 4 1 】

ステップ 236 : データブロックサイズが「R2T」を越えた場合、ブロックサイズ・並列実

50

行シーケンス数決定部202はSCSI層103の制御部に並列に実行するシーケンスの数を1つ増加することを指示する。

ステップ237: TCP制御部に設けられているTCPプロトコル監視手段126はTCPのバッファの状態をTCPプロトコル状態情報127.2としてTCP障害検知手段131に通知する。

TCP障害検知手段131がバッファのオーバーフローを検知しなかった場合はステップ234に戻り、データブロックサイズを増加あるいは並列実行シーケンス数を増加する処理を繰り返す。

TCP障害検知手段131がバッファのオーバーフローを検知した場合、ネットワークは輻輳状態となったことを意味し、更にデータの送信量を増加させることは不可能である。従って、ネットワークが輻輳状態とならないデータブロックサイズと並列実行シーケンス数による通信を維持する。

上記ステップ237においては、実施例1にならってTCPプロトコル監視手段、TCPプロトコル状態情報、TCP障害検知手段を用いてTCP層におけるバッファオーバーフローを検知する方法について説明したが、この代わりに実施例2に示したデータリンク層監視手段、データリンク層状態情報、データリンク層障害検知手段を用いてデータリンク層におけるバッファオーバーフローを検知する手法を用いても良い。

【0042】

図9は第3の実施例におけるデータブロックサイズの変化の例を示す図である。

t0: 図8のステップ233で制御部102はSCSI層103におけるデータブロックサイズを初期値241に設定する。また、同時実行するシーケンス数を「1」とする。

t1~t3: データブロックサイズは「R2T」242より小さく、且つTCP/IP層のバッファはオーバーフローを起こさない。従って、制御部102は図8のステップ234、ステップ235、ステップ237を順次実行し、データブロックサイズを所定量ずつ増加させる。

t4: ブロックサイズ・並列実行シーケンス数決定部202が、並列実行シーケンス数1による実行中のデータブロックサイズが「R2T」242を越えたことを検知したことにより、並列実行シーケンス数1のデータブロックサイズを「R2T」242を越えない値243に設定する。

t5: ブロックサイズ・並列実行シーケンス数決定部202はSCSI層103の同時実行シーケンス数を2にするよう指示する。

t6: ブロックサイズ・並列実行シーケンス数決定部202はSCSI層103の同時実行シーケンス数を3にするよう指示する。

t7: ブロックサイズ・並列実行シーケンス数決定部202はSCSI層103の同時実行シーケンス数を4にするよう指示する。これによりデータの供給量がTCP/IP層105のバッファの容量を超え、バッファがオーバーフローする。バッファのオーバーフローを検知したTCP障害検知手段131からの情報により、ブロックサイズ・並列実行シーケンス数決定部202は並列実行シーケンス数を3に戻すように指示する。

【0043】

本発明の第3の実施例は、TCPプロトコル監視手段とiSCSIパラメータ監視手段の情報に基づいてSCSI層が並列で実行するシーケンスの数を決定する手段を有している。当該構成により、SCSIコネクションはより多くのデータを送信することが可能となり、より効率の良いデータ伝送を実現する。

【実施例4】

【0044】

図10-1は、本発明の第4の実施例に係るネットワークの構成を示す図であり、図10-2は第4の実施例に係る構成の詳細図である。同実施例は、イニシエータ100のアプリケーション101がiSCSI104を介してターゲット110のiSCSI114に接続されているストレージ装置111にデータを書き込む構成において、iSCSIのパラメータとTCPプロトコルの状態を監視し、それら監視結果に基づいてSCSI層のデータブロックサイズを決定する手段と、iSCSIコネクションを再設定してMaxBurstLengthと呼ばれるiSCSI層のパラメータを変更し、R2T-PDUにより指定されるデータサイズ「R2T」の大きさを制御する手段を有している。

【 0 0 4 5 】

第 4 の実施例に係る構成は、ブロックサイズ・R2Tサイズ決定部302を有し、iSCSI層104にiSCSIコネクション再構成指示304が出力される点で、第 1 の実施例の構成と相違する。

図 1 0 - 1 と図 1 0 - 2 において、図 1 - 1 と図 1 - 2 に示される構成と同等の構成には同一の参照番号が付されている。

文献 3 に記載されるように、iSCSIコネクションの設定を変更し、R2T-PDUに応答してイニシエータからターゲットに送信されるデータ長を大きくすることにより、単位時間あたりの送信データ量を増大することが可能である。図 1 0 - 3 は、R2T-PDUとR2T-PDUに応答してイニシエータからターゲットに送信される書き込みデータの送信シーケンスの概要を示す図である。

10

図 1 0 - 3 (a) は、イニシエータのアプリケーションがターゲットのストレージ装置に対してデータの書き込みを要求した時、イニシエータとターゲットのiSCSI間で設定される初期状態のiSCSIコネクションにおけるMaxBurstLengthによる送信を示すものである。同図のターゲット110は、書き込みデータの長さを初期状態のMaxBurstLengthと同じサイズの「R2T-初期値」と指示するR2T-PDU313をイニシエータ100に送信することにより、イニシエータ100は「R2T-初期値」で指定された長さの書き込みデータ314をターゲット110に送信する。

図 1 0 - 3 (b) は、ブロックサイズ・R2Tサイズ決定部302の指示に基づいて、MaxBurstLengthを初期状態の2倍として再構成されたiSCSIコネクションによる送信を示すものである。同図のターゲット110は、書き込みデータの長さが初期状態のMaxBurstLengthの2倍である「R2T-2」とするR2T-PDU313.1をイニシエータ100に送信する。イニシエータ100はR2T-PDU313.1に응答して、「R2T-2」で指定された長さの書き込みデータ314.1をターゲット110に送信する。

20

本発明の第 4 の実施例は、上記原理に基づくものであり、SCSI層103によるデータ送信のブロックサイズが「R2T」を越えた時、iSCSIコネクションの設定を変更して、同コネクションを再構成することによりMaxBurstLengthを再設定し、より大きな「R2T」によるデータの送信を可能とするものである。図 1 1 は当該実施例の動作フローである。

【 0 0 4 6 】

ステップ331：アプリケーション101からデータ送信要求が出されると、SCSI層、iSCSI層、TCP/IP層のコネクションを通じて、上記データ送信が開始される。iSCSIコネクションにおいて一度のR2T受信により可能な送信量には、所定の初期値（「R2T」）が設定される。

30

ステップ332：iSCSI層104のiSCSIパラメータ監視手段125は、ターゲット110のiSCSI層114から送信されたR2T-PDUを監視し、1回のデータ送信により一回の送信処理によりiSCSI層104からiSCSI層114に送信するデータの大きさ「R2T」を取得する。

ステップ333：制御部102はSCSI層103におけるデータブロックサイズを初期化する。

ステップ334：ブロックサイズ・R2Tサイズ決定部302は、SCSI層が実行中のデータブロックサイズがステップ332で取得した「R2T」より大きいかなかを判断する。データブロックサイズが「R2T」を越える場合はステップ336に進み、「R2T」を越えない場合はステップ335に進む。

40

ステップ335：データブロックサイズが「R2T」を越えていない場合、データブロックサイズを大きくすることにより、送信効率が上がる可能性がある。従って、ブロックサイズ・R2Tサイズ決定部302はブロックサイズ変更部303に指示し、SCSI層103におけるデータブロックサイズを所定量増加させることを指示する。

【 0 0 4 7 】

ステップ336：データブロックサイズが「R2T」を越えた場合、動作中のiSCSIコネクションは送信容量の全てを使用していることを意味する。この場合、ブロックサイズ・R2Tサイズ決定部302は、iSCSI層104に「R2T」を増加できるようにiSCSIコネクションを再構成しMaxBurstLengthをより大きい値に変更することを指示する。

ステップ337：TCP制御部に設けられているTCPプロトコル監視手段126はTCPのバッファ

50

の状態をTCPプロトコル状態情報127.2としてTCP障害検知手段131に通知する。

TCP障害検知手段131がバッファのオーバーフローを検知しなかった場合はステップ334に戻り、データブロックサイズを増加する、あるいはiSCSIコネクションを再設定しMaxBurstLengthをより大きい値に変更する処理を繰り返す。

TCP障害検知手段131がバッファのオーバーフローを検知した場合、ネットワークは輻輳状態となったことを意味し、更にデータの送信量を増加させることは不可能である。従って、ネットワークが輻輳状態とならない「R2T」の値とSCSIのデータブロックサイズによる通信を維持する。

上記ステップ337においては、実施例1にならってTCPプロトコル監視手段、TCPプロトコル状態情報、TCP障害検知手段を用いてTCP層におけるバッファオーバーフローを検知する方法について説明したが、この代わりに実施例2に示したデータリンク層監視手段、データリンク層状態情報、データリンク層障害検知手段を用いてデータリンク層におけるバッファオーバーフローを検知する手法を用いても良い。

【0048】

図12は第4の実施例におけるデータブロックサイズの変化の例を示す図である。

t0: ステップ333で、制御部102はデータブロックサイズを初期値341に設定する。

t1: データブロックサイズが「R2T」の初期値342より小さく、且つTCP/IP層でのバッファオーバーフローとならない間、図11のステップ334、ステップ335、ステップ337が順次実行され、データブロックサイズは所定量ずつ増加する。

t2: データブロックサイズが「R2T」の初期値342を越えたことを検知したブロックサイズ・R2Tサイズ決定部302は、iSCSI層104の制御部にiSCSIコネクション再構成指示304を出し、iSCSIコネクションを再構成して「R2T」を増加させることを指示する。当該指示を受けたiSCSI層104の制御部は、新たなMaxBurstLength(図12の「R2T-2」343)により、iSCSI104とiSCSI114の間のiSCSIコネクションを再構成する。

再構成に成功しなかった場合は、現在のデータブロックサイズによる送信を維持する。

再構成に成功した場合は、t2~t3の間で、図11のステップ334、ステップ335、ステップ337が実行され、データブロックサイズが増加する。

t3: t2と同様、「R2T-3」344によりiSCSIコネクションを再構成する。t3~t4の間、図11のステップ334、ステップ335、ステップ337が実行され、データブロックサイズが増加する。

t4: t3と同様、「R2T-4」345によりiSCSIコネクションを再構成する。

t5: t5において、TCPコネクションに要求されるデータ送信量はTCP層のバッファの容量を超える。TCP障害検知手段131がTCP層のバッファがオーバーフローしたことを検知すると、ブロックサイズ・R2Tサイズ決定部302はデータブロックサイズの大きさをバッファがオーバーフローする以前の値347に決定する。

t6~: バッファがオーバーフローを起こすことのないデータブロックサイズ347によるデータ送信が維持される。

【0049】

本発明の第4の実施例は、TCPプロトコル監視手段とiSCSIパラメータ監視手段の情報に基づいてiSCSIコネクションを再構成してMaxBurstLengthの値を再設定し、R2T-PDUが指定するデータサイズ「R2T」の値を大きくする手段を有している。当該手段により、iSCSIプロトコルによりイニシエータからターゲットのストレージ装置にデータを書き込むストレージネットワークシステムにおいて、ネットワークの送信能力に即した大きなスループットを得ることが可能となる。

上記の実施例1、2、3、4におけるブロックサイズなどのパラメータの最適値発見方法として、一定値ずつ単純に増減させる線形探索方法を記載したが、本発明はこれに限定されるものではない。

例えば、線形的な増減の代わりに指数関数的な増減を用いる最適値探索方法を適用することも可能であるし、探索範囲を二分しながら最適の値に近づけていく二分探索方法を用いることができる。さらに、各実施例におけるフローチャートにおいてパラメータを初期

10

20

30

40

50

化する際（図2のステップ143、図5のステップ173、図8のステップ233、図11のステップ333）、iSCSIパラメータ127.1、TCPプロトコル状態情報127.2、データリンク層状態情報127.3などから、最適に近いと考えられるブロックサイズの値を求め、その値を初期値として与えて探索を開始することにより、より速く最適なブロックサイズの値を発見することも可能である。

本発明は最適値発見がより高速になる上記の方法などを含むものであり、実施例に記載されたものに限定されるものではない。

【産業上の利用可能性】

【0050】

イニシエータからiSCSIにより接続されたターゲットのストレージ装置にデータを格納する場合、SCSI層では、イニシエータはターゲットに送信するデータをブロックサイズと呼ばれるデータ量に分割して送信し、このデータ量を送信した後は確認パケットを受取るまで次のデータ送信ができない。よってSCSI層においては、RTTが大きい長距離通信においては、ブロックサイズを大きくすることが望ましい。

同様にiSCSI層ではイニシエータはターゲットに送信するデータを、R2T-PDUにおける「R2T」で指定したデータ量に分割して送信される。よってiSCSI層でスループットを増加させるためには「R2T」を大きくすることが望ましい。

一方、TCP/IPネットワークのTCP層では輻輳制御が行われており、輻輳ウィンドウサイズはネットワークの状態やバッファの状態に応じて増減を繰り返す。この結果、上記ブロックサイズや上記「R2T」を大きくしすぎると、ネットワークの輻輳やバッファオーバーフローの原因となり、かえって書き込み処理のスループットが低下する問題がある。

本発明の第1の実施例は、TCPコネクションの状態をTCPプロトコル監視手段により監視すると共に、iSCSIコネクションのパラメータをiSCSIパラメータ監視手段により監視する構成を有している。イニシエータ100の制御部102は両監視手段からの情報に基づいてSCSI層におけるデータブロックサイズを決定する。当該構成により、TCPコネクションによるiSCSI間のデータ送信において、TCP層でのバッファオーバーフロー、データリンク層でのバッファオーバーフロー、ネットワークの輻輳などの問題を回避しつつ高い送信効率を確保することが可能となる。

本発明の第2の実施例は、データリンク層の状態をデータリンク層監視手段により監視すると共に、iSCSIコネクションのパラメータをiSCSIパラメータ監視手段により監視する構成を有している。イニシエータ100の制御部102は両監視手段からの情報に基づいてSCSI層におけるデータブロックサイズを決定する。当該構成により、iSCSI間のデータ送信において、データリンク層でのバッファオーバーフローなどの問題を回避しつつ高い送信効率を確保することが可能となる。

【0051】

本発明の第3の実施例は、TCPプロトコル監視手段とiSCSIパラメータ監視手段の情報に基づいて、SCSI層のブロックサイズおよび同時に実行するシーケンス数を決定する構成を有している。当該構成により、「R2T」による制限をこえたデータをTCP層やネットワークに供給することが可能となるため、ネットワークの送信能力に即した大きなスループットを得ることが可能となる。

本発明の第4の実施例は、TCPプロトコル監視手段とiSCSIパラメータ監視手段の情報に基づいて、SCSI層のブロックサイズを決定する、またはiSCSIコネクションを再構成しR2T-PDUが指定するデータサイズ「R2T」の値を大きくする構成を有している。当該構成により、ネットワークの送信能力に即した大きなスループットを得ることが可能となる。

【図面の簡単な説明】

【0052】

【図1-1】本発明に係るネットワークシステムの第1の実施形態の図

【図1-2】第1の実施形態の詳細図

【図2】第1の実施形態の動作フロー図

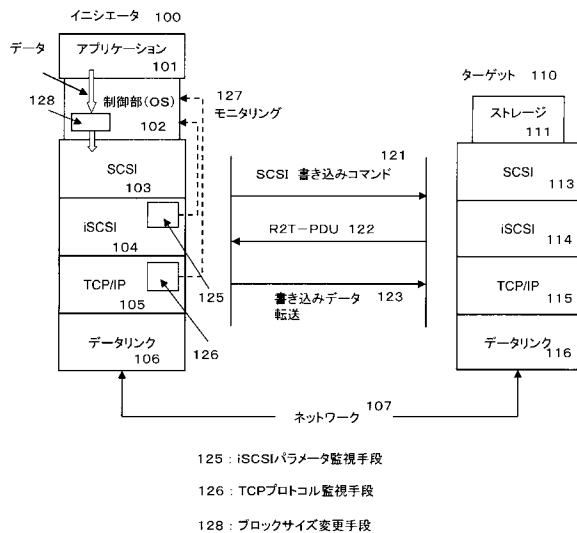
【図3】第1の実施形態におけるSCSIのデータブロックサイズの変化を示す

- 【図 4 - 1】本発明に係るネットワークシステムの第 2 の実施形態の図
 【図 4 - 2】第 2 の実施形態の詳細図
 【図 5】第 2 の実施形態の動作フロー図
 【図 6】第 2 の実施形態における SCSI のデータブロックサイズの変化を示す
 【図 7 - 1】本発明に係るネットワークシステムの第 3 の実施形態の図
 【図 7 - 2】第 3 の実施形態の詳細図
 【図 7 - 3】SCSI コマンドの数と並列実行シーケンス数の関係を示す概念図
 【図 8】第 3 の実施形態の動作フロー図
 【図 9】第 3 の実施形態における SCSI のデータブロックサイズの変化を示す図
 【図 10 - 1】本発明に係るネットワークシステムの第 4 の実施形態の図
 【図 10 - 2】第 4 の実施形態の詳細図
 【図 10 - 3】R2T-PDU による書き込みデータの大きさを指定する概念図
 【図 11】第 4 の実施形態の動作フロー図
 【図 12】第 4 の実施形態における SCSI のデータブロックサイズの変化を示す図
 【図 13】iSCSI 層を有する従来のネットワークを示す図
 【図 14】TCP 層の輻輳ウィンドウサイズの変化を示す図
 【図 15】iSCSI コネクションによる読み出しシーケンスを示す図
 【図 16】輻輳ウィンドウサイズを最適化する従来方法の構成を示す図
 【図 17】図 16 に示される従来方法による処理フローと輻輳ウィンドウサイズの変化を示す図
 【図 18】iSCSI コネクションによる書き込みシーケンスを示す図

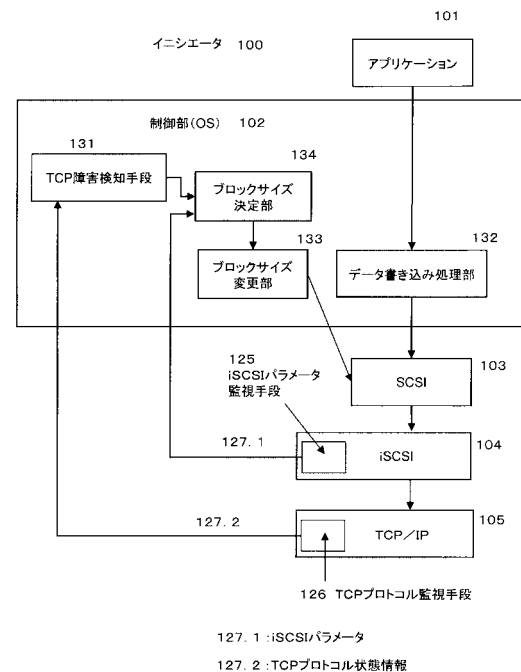
10

20

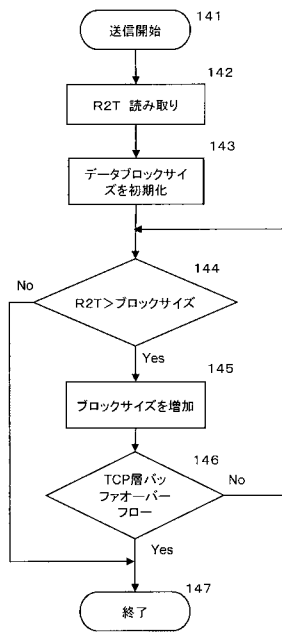
【図 1 - 1】



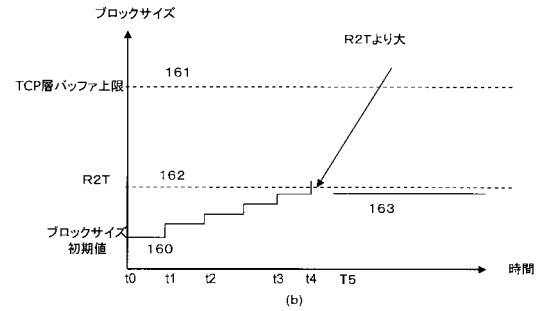
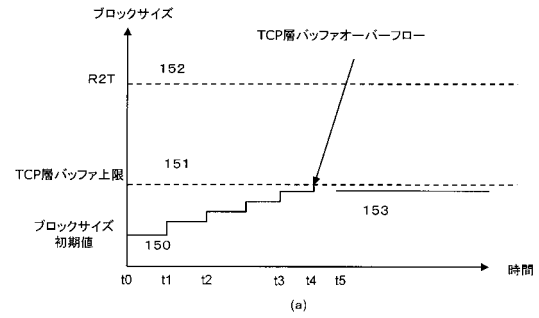
【図 1 - 2】



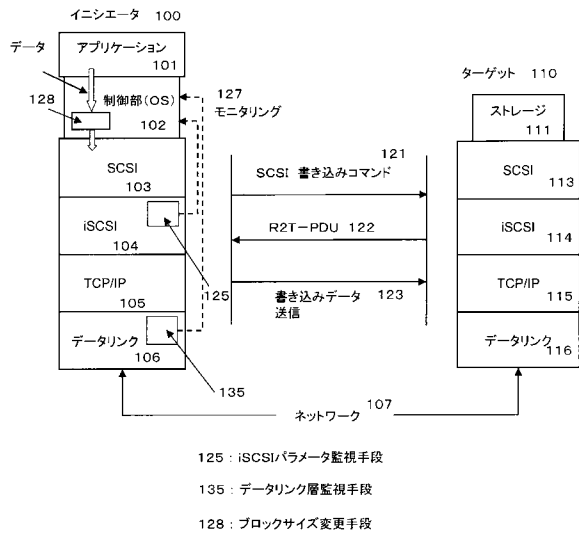
【図 2】



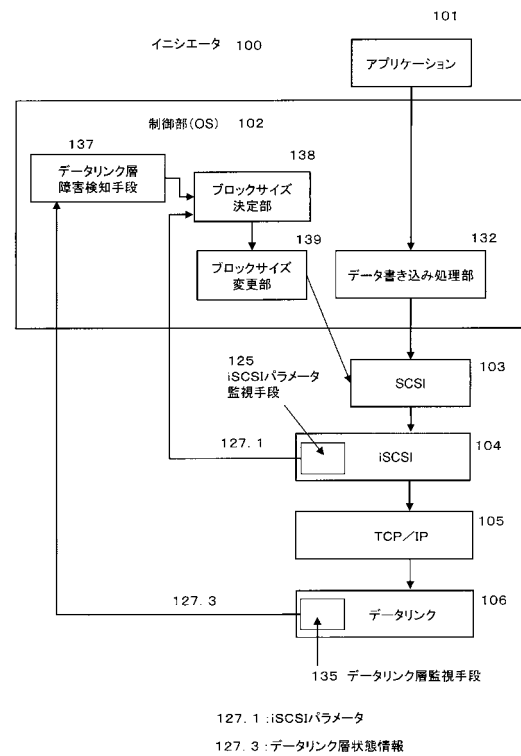
【図 3】



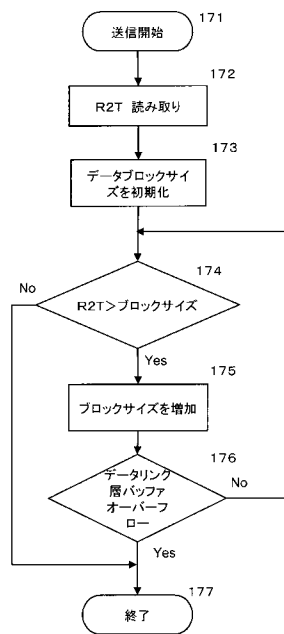
【図 4 - 1】



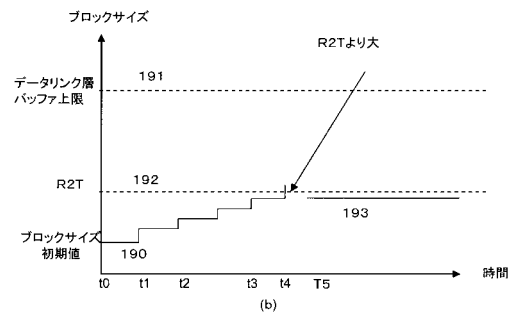
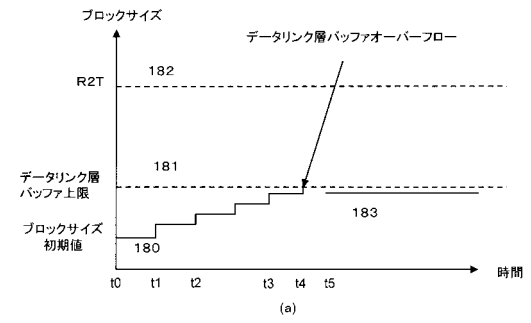
【図 4 - 2】



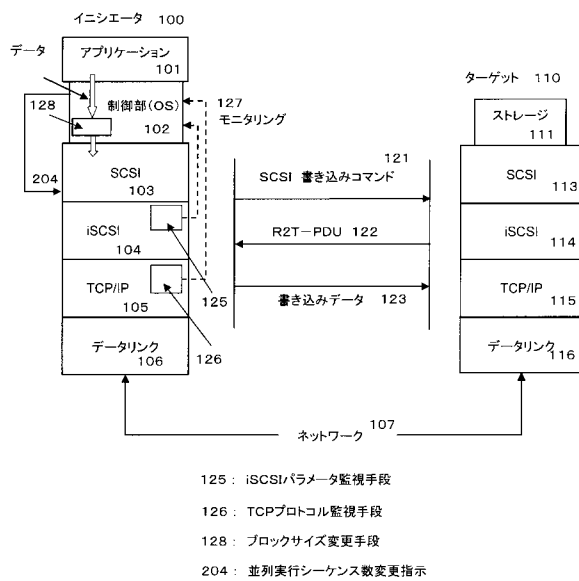
【図 5】



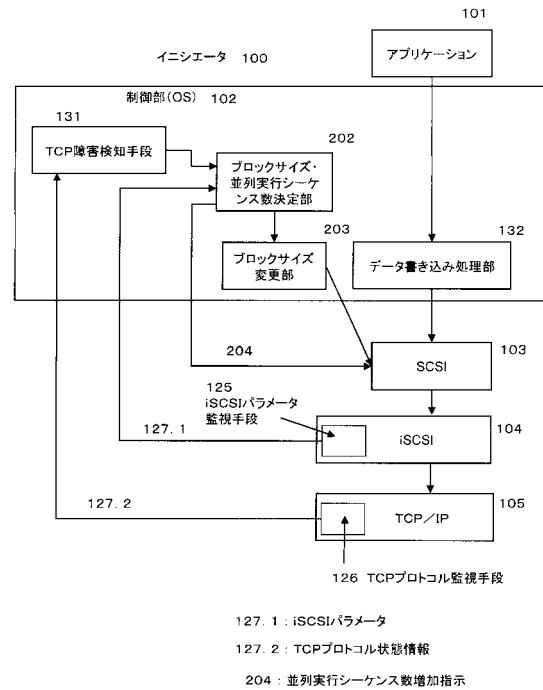
【図 6】



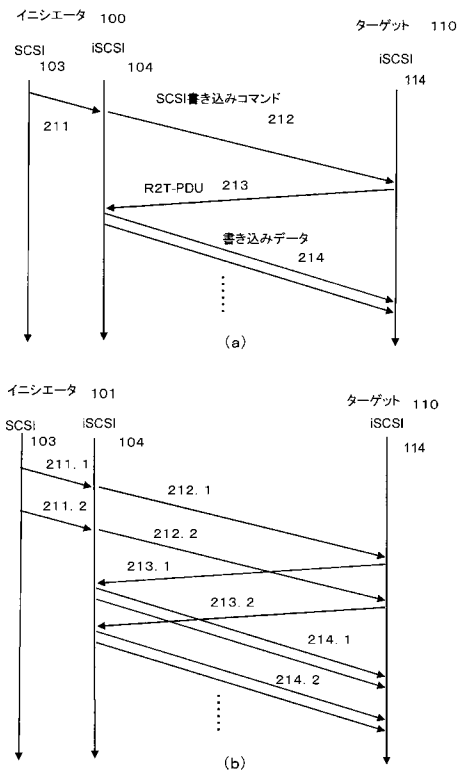
【図 7 - 1】



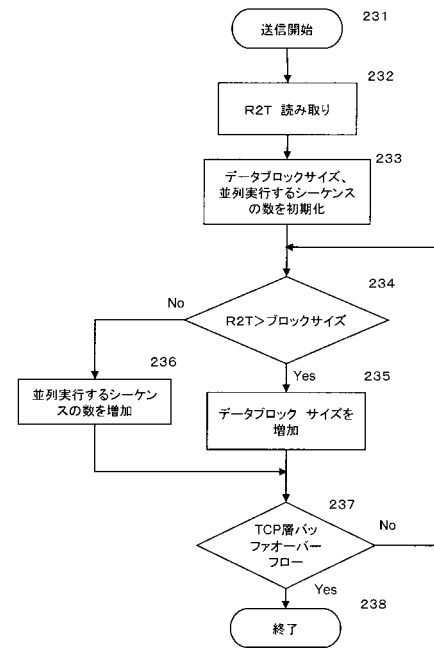
【図 7 - 2】



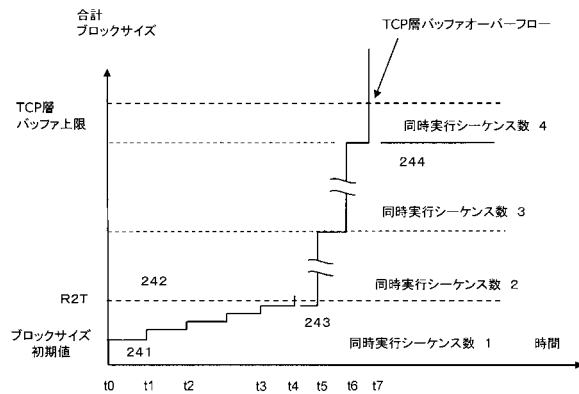
【図 7 - 3】



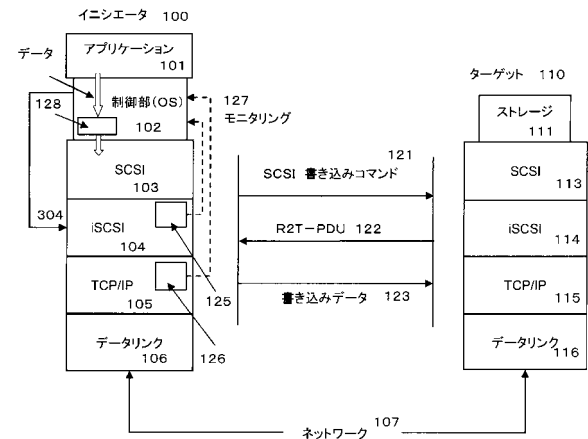
【図 8】



【図 9】



【図 10 - 1】



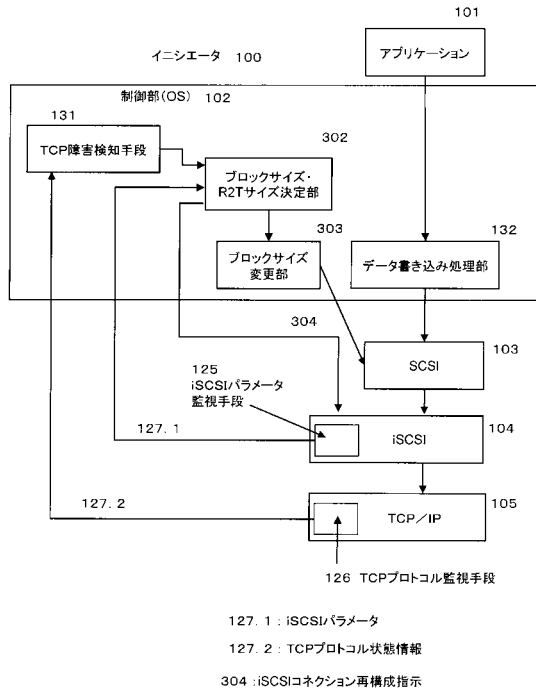
125 : iSCSIパラメータ監視手段

126 : TCPプロトコル監視手段

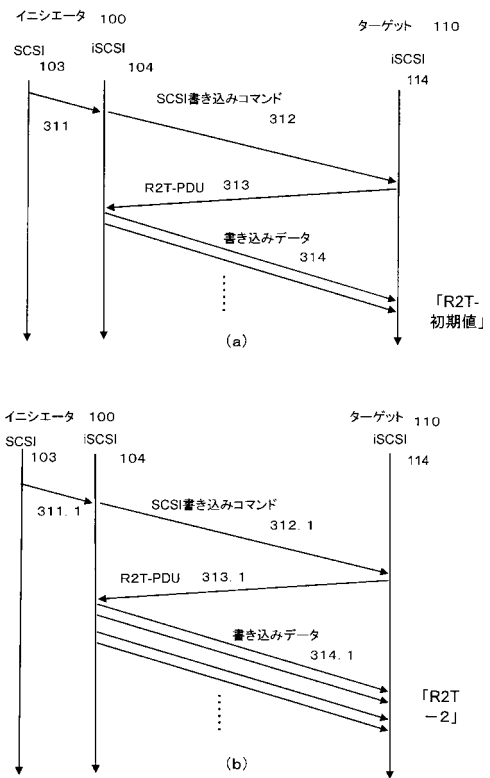
128 : ブロックサイズ変更手段

304 : iSCSIコネクション再構成指示

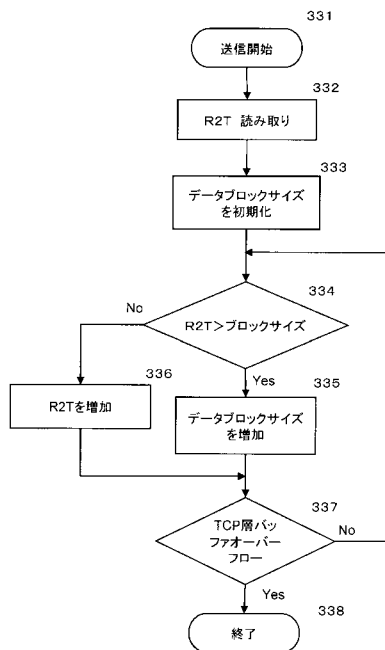
【図 10 - 2】



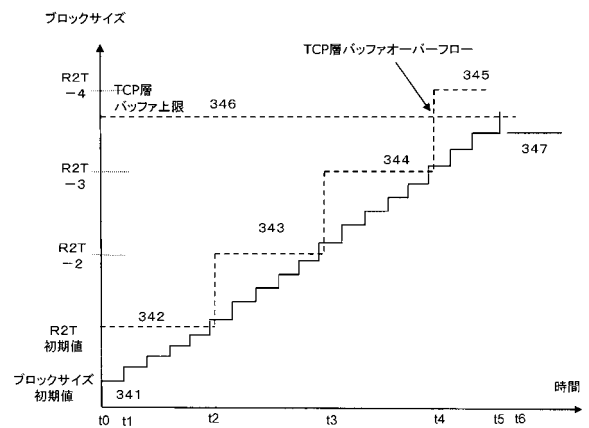
【図 10 - 3】



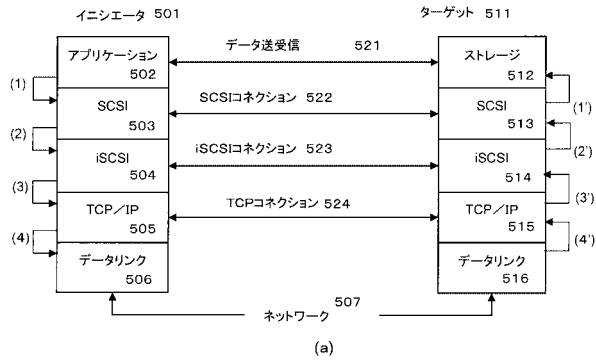
【図 11】



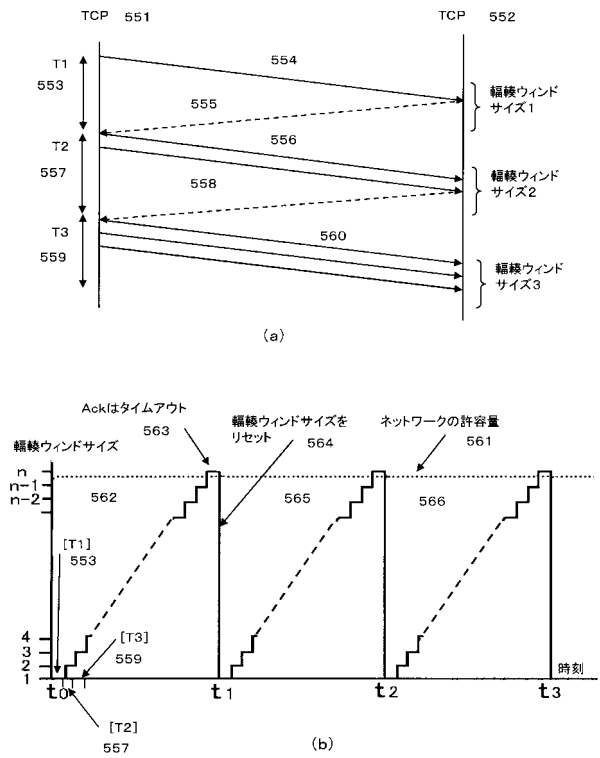
【図 12】



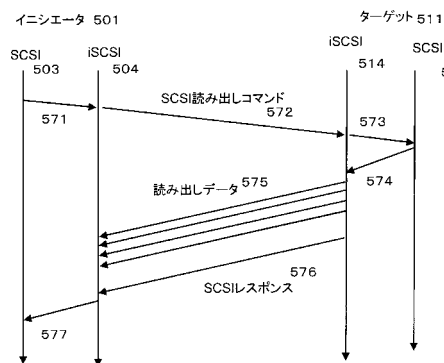
【図 13】



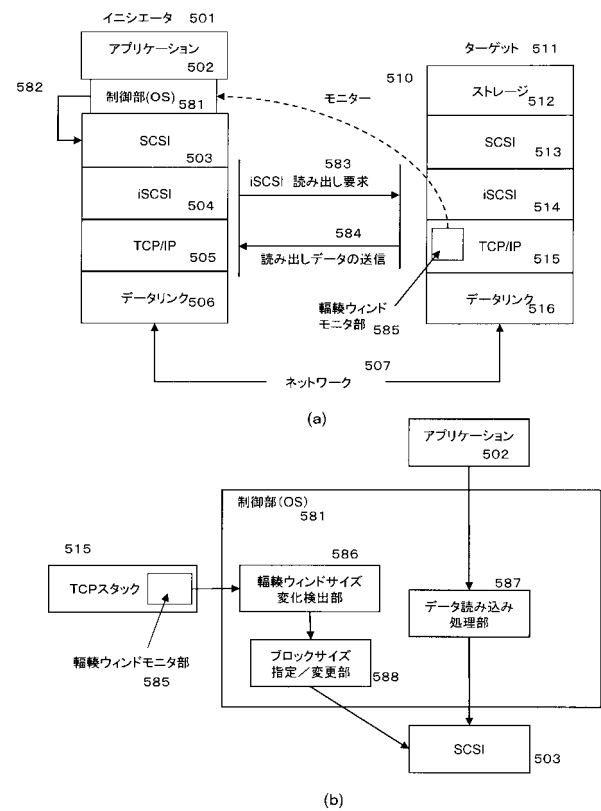
【図 14】



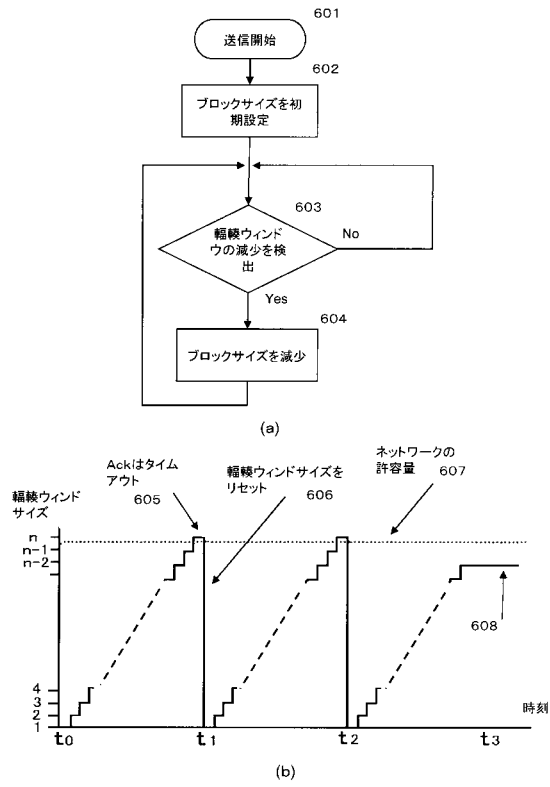
【図 15】



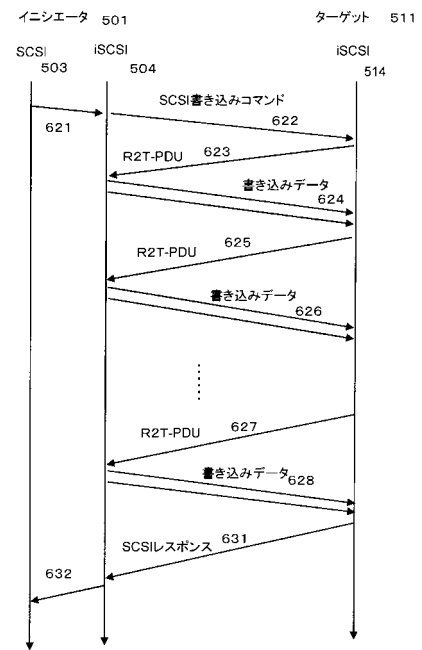
【図 16】



【図 17】



【図 18】



フロントページの続き

- (74)代理人 100106459
弁理士 高橋 英生
- (74)代理人 100105500
弁理士 武山 吉孝
- (74)代理人 100103735
弁理士 鈴木 隆盛
- (72)発明者 小口 正人
東京都文京区大塚 2 - 1 - 1 お茶の水女子大学内
- (72)発明者 山口 実靖
東京都新宿区市谷本村町 4 2 - 8 工学院大学内
- (72)発明者 岡廻 隆生
東京都港区東新橋一丁目 9 番 1 号 ソフトバンクテレコム株式会社内
- (72)発明者 松原 幸助
東京都港区東新橋一丁目 9 番 1 号 ソフトバンクテレコム株式会社内
- F ターム(参考) 5B014 EB05
5B065 CE04
5B089 GB01 HB02
5K030 GA01 HA08 JA05 KA03 LC03 MA04 MA13 MB11
5K034 HH12 HH50 HH63 HH64 MM24 MM25