

(11) 特許出願公開番号

特開2012-124700

(P2012-124700A)

(43) 公開日 平成24年6月28日(2012.6.28)

(51) Int. Cl.
H04L 12/56

F 1
H O 4 L 12/56 1 0 0 Z

テーマコード (参考)
5K030

審査請求 未請求 請求項の数 9 O L (全 14 頁)

(21) 出願番号 特願2010-273553 (P2010-273553)
(22) 出願日 平成22年12月8日 (2010. 12. 8)

(71) 出願人 000004226
日本電信電話株式会社
東京都千代田区大手町二丁目3番1号

(74) 代理人 100123788
弁理士 宮崎 昭夫

(74) 代理人 100148297
弁理士 ▲高▼森 俊夫

(72) 発明者 平松 淳
東京都千代田区大手町二丁目3番1号 日
本電信電話株式会社内

(72) 発明者 栗本 崇
東京都千代田区大手町二丁目3番1号 日
本電信電話株式会社内

Fターム(参考) 5K030 GA11 KA05 KX12 KX13 LB05

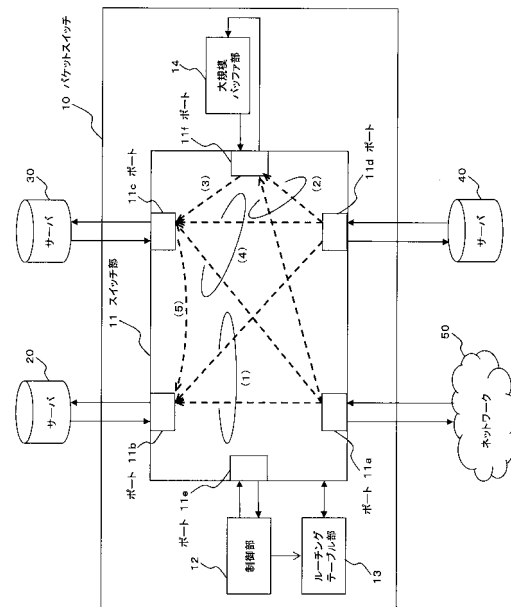
(54) 【発明の名称】 パケットスイッチ

(57) 【要約】

【課題】サーバによる機能の提供の高信頼性を確保する

【解決手段】パケットを入出力する複数のポートを備えたスイッチ部と、ルーチングテーブルとを有するパケットスイッチであって、スイッチ部は、複数のポートのうち少なくとも1つを示すポート情報およびアドレスを含む第1のメッセージを複数のポートのいずれかを介して受信し、複数のポートのうちの第1のポートを介してスイッチ部と接続され、第1のポートから出力されたパケットを受け付けて記憶するバッファ部と、第1のメッセージが受信されると、複数のポートのうち当該第1のメッセージに含まれるポート情報が示すポート以外のポートにて受け付けられ、当該第1のメッセージに含まれるアドレスを宛先アドレスとするパケットを第1のポートから出力するようにルーチングテーブルを設定する制御部とを有する。

【選択図】 図 3



【特許請求の範囲】

【請求項 1】

パケットを入出力する複数のポートを備えたスイッチ部と、前記複数のポートのうちパケットを受け付けたポートと当該パケットを出力するポートとをパケットの宛先アドレスに応じて対応付けて記憶するルーチングテーブルとを有するパケットスイッチであって、

前記スイッチ部は、前記複数のポートのうちの少なくとも1つを示すポート情報、および、アドレスを含む第1のメッセージを前記複数のポートのいずれかを介して受信し、

前記複数のポートのうちの第1のポートを介して前記スイッチ部と接続され、前記第1のポートから出力されたパケットを受け付けて記憶するバッファ部と、

前記第1のメッセージが受信されると、前記複数のポートのうち当該第1のメッセージに含まれるポート情報が示すポート以外のポートにて受け付けられ、当該第1のメッセージに含まれるアドレスを宛先アドレスとするパケットを、前記第1のポートから出力するように前記ルーチングテーブルを設定する制御部と、を有するパケットスイッチ。

10

【請求項 2】

請求項 1 に記載のパケットスイッチにおいて、

前記スイッチ部は、前記複数のポートのいずれかを介して第2のメッセージを受信し、

前記制御部は、前記第2のメッセージが受信されると、該受信された第2のメッセージに基づいて前記複数のポートの中から第2のポートを決定し、前記第1のポートにて受け付けられたパケットを前記第2のポートから出力するように前記ルーチングテーブルを設定した後、前記バッファ部に記憶されたパケットを出力させるパケットスイッチ。

20

【請求項 3】

請求項 2 に記載のパケットスイッチにおいて、

前記第2のメッセージは、前記複数のポートのうちのいずれかを示すポート情報を含み、

、

前記制御部は、前記受信された第2のメッセージに含まれるポート情報が示すポートを前記第2のポートとするパケットスイッチ。

【請求項 4】

請求項 2 に記載のパケットスイッチにおいて、

前記制御部は、前記複数のポートのうち前記第2のメッセージが受信されたポートを前記第2のポートとするパケットスイッチ。

30

【請求項 5】

請求項 2 乃至 4 のいずれか 1 項に記載のパケットスイッチにおいて、

前記制御部は、前記バッファ部に記憶されたパケットがなくなると、前記複数のポートのうち前記第1のメッセージに含まれるポート情報が示すポート以外のポートにて受け付けられ、前記第1のメッセージに含まれるアドレスを宛先アドレスとするパケットを前記第2のポートから出力するように前記ルーチングテーブルを設定するパケットスイッチ。

【請求項 6】

請求項 2 乃至 5 のいずれか 1 項に記載のパケットスイッチにおいて、

前記バッファ部は、前記記憶したパケットを所定のレートで出力するパケットスイッチ。

40

【請求項 7】

請求項 2 乃至 6 のいずれか 1 項に記載のパケットスイッチにおいて、

前記バッファ部は、前記記憶したパケットを出力する際、当該パケットが記憶されていた時間が所定の時間よりも短いかどうかを判定し、当該パケットが記憶されていた時間が前記所定の時間よりも短い場合に当該パケットを出力するパケットスイッチ。

【請求項 8】

請求項 1 乃至 7 のいずれか 1 項に記載のパケットスイッチにおいて、

前記バッファ部は、前記受け付けたパケットのヘッダ部分が所定のパターンと一致する場合に当該受け付けたパケットを記憶するパケットスイッチ。

【請求項 9】

50

請求項 1 乃至 8 のいずれか 1 項に記載のパケットスイッチにおいて、

前記バッファ部は、前記受け付けたパケットを記憶させるための空きエリアがない場合、当該バッファ部に記憶された複数のパケットのうち最も早い時期に記憶されたパケットを破棄し、当該受け付けたパケットを記憶するパケットスイッチ。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、パケットを転送するパケットスイッチに関する。

【背景技術】

【0002】

現在、多様なサーバが多数の利用者に対してネットワークを介して様々な機能を提供するようになってきている。利用者数が増加し、また、提供される機能の重要性が高くなってきていることから、機能の提供の安定性や継続性に対する関心が高まっており、サーバの高信頼性が求められるようになってきている。

【0003】

そこで、複数のサーバをパケットスイッチで接続しておき、通常運用中に機能を提供する運用サーバのソフトウェアやハードウェアを交換する場合や、運用サーバに障害等が発生した場合に、機能を提供するサーバを運用サーバからバックアップサーバへ切り替える。これにより、機能の提供が中断される時間を短くすることができる。

【0004】

また、運用サーバに障害等が発生した場合に、複数のバックアップサーバのうちのどのサーバに対してどのソフトウェアの実行を委ねるか、というような機能の提供の継続に向けたサーバ群の管理を行うための技術も考えられている。このような技術が例えば、特許文献 1 に開示されている。

【0005】

特許文献 1 に開示されている技術では、複数のサーバによって機能の提供を行っているシステムにおいて、障害が検出された場合に障害の原因等に応じてどのサーバにどのプロセスの処理を実行させるかを管理している。これにより、機能の提供が中断される時間を可能な限り短くすることができる。

【0006】

ここで、機能を提供するサーバを切り替えるときの動作の一例を説明する。なお、ここでは、運用サーバのソフトウェアやハードウェアを計画的に交換する場合について説明する。

【0007】

図 5 は、機能を提供するサーバを切り替えるときの動作の一例を説明するための図である。

【0008】

図 5 に示すパケットスイッチ 100 は、パケットを入出力するポート 11a～11c を有するスイッチ部 101 と、制御部 102 と、ルーチングテーブル部 103 とを備えている。サーバ 20, 30 は、パケットスイッチ 100 を介してネットワーク 50 に接続されている。

【0009】

ルーチングテーブル部 103 は、ルーチングテーブルを記憶している。ルーチングテーブルは、ポート 11a～11c のうち、パケットを受け付けたポートとそのパケットを出力するポートとを、パケットの宛先アドレスに応じて対応付けて記憶している。

【0010】

スイッチ部 101 は、ポート 11a～11c のそれぞれにてパケットが受け付けられると、受け付けられたポートとそのパケットの宛先アドレスとから、ルーチングテーブルを参照することによってそのパケットを出力するポートを決定する。そして、スイッチ部 101 は、決定したポートからそのパケットを出力する。

10

20

30

40

50

【0011】

制御部102は、パケットスイッチ100全体の動作を制御する。具体的には例えば、ルーティングテーブルの設定および更新や、スイッチ部11の状態管理を行う。

【0012】

図5に示した構成においてサーバを切り替える動作は以下に説明するとおりである。なお、ここでは、サーバ20を運用サーバとし、サーバ30をバックアップサーバとする。

(手順a)サーバ20は、ネットワーク50に接続された端末等に対して機能を提供しながらサーバ30を起動させる。

(手順b)サーバ20は、機能の提供を継続しながら様々な環境データをサーバ30へコピーする。

(手順c)サーバ20は、ネットワーク50に接続された端末等に対する機能の提供を停止する。

(手順d)サーバ20は、処理中のサービスに関する状態データをサーバ30へコピーする。

(手順e)サーバ20は、サーバ30を運転状態に移行させる。

(手順f)パケットスイッチ100の制御部102は、ネットワーク50からサーバ20へ送信されたパケットが、サーバ30へ転送されるようにルーティングテーブルを設定する(図5における(A)から(B)への切り替え)。

【0013】

以降、サーバ30が、サーバ20の処理を引き継ぎ、ネットワーク50に接続された端末に対する機能の提供が再開される。

【0014】

上述した(手順c)において、サーバ20は機能の提供を停止している。これは、サーバ20からサーバ30へ状態データのコピーを行っている間にサーバ20が何らかの処理を行うことにより、サーバ20の内部状態が変化してしまうことを回避するためである。例えば、サーバ20からサーバ30へ状態データのコピーを行っている間にサーバ20が何らかの処理を行うと、サーバ30にコピー済みの状態データとの不整合が生じること可能性がある。

【先行技術文献】

【特許文献】

【0015】

【特許文献1】特許第4112191号

【発明の概要】

【発明が解決しようとする課題】

【0016】

機能を提供するサーバを切り替える動作において、上述した手順cから手順fまでの間は、運用サーバもバックアップサーバも機能を提供していない。そのため、この間に運用サーバへ送信されたパケットは処理されず、サーバによる機能の提供の高信頼性を確保できないという問題点がある。

【0017】

なお、サーバ上のソフトウェアの処理を工夫することにより、運用サーバからバックアップサーバへコピーする状態データの量を少なくしたり、コピー処理の性能を改善したりするなどして、手順cから手順fまでにかかる時間を短縮する取り組みが行われている。しかし、この時間を0にすることは困難である。

【0018】

また、運用サーバに障害等が発生した場合には、運用サーバのソフトウェアの再起動や、バックアップサーバのソフトウェアの起動などの処理が行われるが、機能の提供が再開されるまでの間に運用サーバへ送信されたパケットは処理されない。

【0019】

本発明は、サーバによる機能の提供の高信頼性を確保することを可能にするパケットス

10

20

30

40

50

イッチを提供することを目的とする。

【課題を解決するための手段】

【0020】

上記目的を達成するために本発明のパケットスイッチは、パケットを入出力する複数のポートを備えたスイッチ部と、前記複数のポートのうちパケットを受け付けたポートと当該パケットを出力するポートとをパケットの宛先アドレスに応じて対応付けて記憶するルーチングテーブルとを有するパケットスイッチであって、

前記スイッチ部は、前記複数のポートのうちの少なくとも1つを示すポート情報、および、アドレスを含む第1のメッセージを前記複数のポートのいずれかを介して受信し、

前記複数のポートのうちの第1のポートを介して前記スイッチ部と接続され、前記第1のポートから出力されたパケットを受け付けて記憶するバッファ部と、

前記第1のメッセージが受信されると、前記複数のポートのうち当該第1のメッセージに含まれるポート情報が示すポート以外のポートにて受け付けられ、当該第1のメッセージに含まれるアドレスを宛先アドレスとするパケットを、前記第1のポートから出力するように前記ルーチングテーブルを設定する制御部と、を有する。

【発明の効果】

【0021】

本発明は以上説明したように構成されているので、サーバによる機能の提供がされていない間にそのサーバへ送信されたパケットを一時的にバッファ部に記憶させることができる。

【0022】

バッファ部に一時的に記憶されたパケットは、機能の提供が再開された後に、機能を提供するサーバに転送されて処理される。そのため、サーバによる機能の提供の高信頼性を確保することが可能となる。

【図面の簡単な説明】

【0023】

【図1】本発明のパケットスイッチの実施の一形態の構成を示すブロック図である。

【図2】図1に示した大規模バッファ部の構成の一例を示すブロック図である。

【図3】図1および図2に示したパケットスイッチの動作の一例を説明するための図である。

【図4】図1および図2に示したパケットスイッチの動作の他の例を説明するための図である。

【図5】機能を提供するサーバを切り替えるときの動作の一例を説明するための図である。

【発明を実施するための形態】

【0024】

以下に、本発明の実施の形態について図面を参照して説明する。

【0025】

図1は、本発明のパケットスイッチの実施の一形態の構成を示すブロック図である。

【0026】

本実施形態のパケットスイッチ10は図1に示すように、パケットを入出力するポート11a～11fを有するスイッチ部11と、制御部12と、ルーチングテーブル部13と、大規模バッファ部14とを備えている。なお、図1においては、スイッチ部11が有するポートの数を6個としているが、6個に限定されるものではない。

【0027】

ルーチングテーブル部13は、ルーチングテーブルを記憶している。ルーチングテーブルは、ポート11a～11fのうち、パケットが受け付けられたポートとそのパケットを出力するポートとをパケットの宛先アドレスに応じて対応付けて記憶している。なお、アドレスとは例えば、IP(Internet Protocol)アドレスである。

【0028】

スイッチ部 11 は、ルーチングテーブルに記憶された内容に従ってパケットを転送する。具体的には、スイッチ部 11 は、ポート 11 a～11 f のそれぞれにてパケットが受け付けられると、受け付けられたポートとそのパケットの宛先アドレスとから、ルーチングテーブルを参照することによってそのパケットを出力するポートを決定する。そして、スイッチ部 11 は、決定したポートからそのパケットを出力する。また、スイッチ部 11 は、ポート 11 a～11 d を介して接続された端末やサーバなどから送信され、制御部 12 を宛先とするメッセージを受信する。このメッセージの詳細については後述する。

【0029】

制御部 12 は、パケットスイッチ 10 全体の動作を制御する。具体的には例えば、ルーチングテーブルの設定および更新や、スイッチ部 11 の状態管理、大規模バッファ部 14 の状態管理や制御を行う。また、制御部 12 は、ポート 11 e を介してスイッチ部 11 と接続されており、スイッチ部 11 にて受信されたメッセージをポート 11 e を介して受け付ける。メッセージには例えば、パケットの転送の停止を要求する第 1 のメッセージである停止要求メッセージや、停止したパケットの転送の再開を要求する第 2 のメッセージである停止解除要求メッセージがある。停止要求メッセージには、ポート 11 a～11 f のうちの少なくとも 1 つを示すポート情報およびアドレスが含まれている。制御部 12 は、停止要求メッセージを受け付けると、受け付けた停止要求メッセージに含まれるポート情報およびアドレスに応じてルーチングテーブルを設定する。また、制御部 12 は、停止解除要求メッセージを受け付けると、受け付けた停止解除要求メッセージに基づき、ポート 11 a～11 f の中から第 2 のポートを決定する。そして、制御部 12 は、ポート 11 f にて受け付けられたパケットを第 2 のポートから出力するようにルーチングテーブルを設定する。なお、制御部 12 がルーチングテーブルを設定する動作の詳細については、後述する動作フローで説明する。

【0030】

大規模バッファ部 14 は、第 1 のポートであるポート 11 f を介してスイッチ部と接続されており、ポート 11 f から出力されたパケットを受け付けて記憶する。

【0031】

図 2 は、図 1 に示した大規模バッファ部 14 の構成の一例を示すブロック図である。

【0032】

図 1 に示した大規模バッファ部 14 は図 2 に示すように、バッファメモリ 141 と、読み出し部 142 と、書き込み部 143 と、管理部 144 と、レートレジスタ 145 と、タイマ 146 と、最大記憶時間レジスタ 147 と、パターンテーブル 148 とを備えている。

【0033】

管理部 144 は、大規模バッファ部 14 の状態を管理する。なお、制御部 12 は、後述するレートレジスタ 145 や、最大記憶時間レジスタ 147、パターンテーブル 148 の設定を管理部 144 を介して行う。また、制御部 12 は、管理部 144 を介してバッファメモリ 141 の空き状態を監視する。

【0034】

書き込み部 143 は、ポート 11 f から出力されたパケットを受け付け、受け付けたパケットをバッファメモリ 141 の空きエリアに書き込む。

【0035】

バッファメモリ 141 は、パケットを記憶するメモリである。バッファメモリ 141 のサイズは、ポート 11 a～11 f の転送速度とパケットを記憶する最長時間との積によって決定される。例えば、ポート 11 a～11 f の転送速度が 1 Gbps で、最長で 80 秒間パケットを記憶する場合には、バッファメモリ 141 のサイズは 10 G バイトとなる。なお、この場合、バッファメモリ 141 は、パケットの書き込みとパケットの読み出しとを 1 Gbps の速度で並行して行える性能を有していなければならない。

【0036】

読み出し部 142 は、管理部 144 を介して制御部 12 から読み出し開始の指示を受け

10

20

30

40

50

付けると、バッファメモリ 141 に記憶されたパケットを読み出して出力する。

【0037】

ここで、大規模バッファ部 14 は図 2 に示すように、レートレジスタ 145 と、タイマ 146 と、最大記憶時間レジスタ 147 と、パターンテーブル 148 とを備えていてもよい。

【0038】

レートレジスタ 145 には、バッファメモリ 141 に記憶されたパケットを出力する際の所定のレートが制御部 12 によって予め設定されている。レートレジスタ 145 が備えられている場合、読み出し部 142 は、レートレジスタ 145 に設定された所定のレートでバッファメモリ 141 に記憶されたパケットを出力する。

10

【0039】

パターンテーブル 148 には、パケットのヘッダ部分の所定のパターンが制御部 12 によって予め設定されている。パターンテーブル 148 が備えられている場合、書き込み部 143 は、ポート 11f から出力されたパケットを受け付けると、受け付けたパケットのヘッダ部分がパターンテーブル 148 に設定されている所定のパターンと一致するかどうかを判定する。この判定の結果、受け付けたパケットのヘッダ部分がパターンテーブル 148 に設定されている所定のパターンと一致する場合、書き込み部 143 は、受け付けたパケットをバッファメモリ 141 に書き込む。一方、受け付けたパケットのヘッダ部分がパターンテーブル 148 に設定されている所定のパターンと一致しない場合、書き込み部 143 は、受け付けたパケットを破棄する。なお、パターンテーブル 148 には、複数の

20

【0040】

タイマ 146 は、時刻を示すカウンタである。また、最大記憶時間レジスタ 147 には、パケットを記憶する最大の時間である最大記憶時間が制御部 12 によって予め設定されている。

【0041】

タイマ 146 と最大記憶時間レジスタ 147 が備えられている場合、書き込み部 143 は、ポート 11f から出力されたパケットを受け付けると、受け付けたときのタイマ 146 が示す時刻を、受け付けたパケットとともにバッファメモリ 141 の空きエリアに書き込む。この場合、受け付けたパケットを記憶させるのに必要なメモリのサイズは、時刻を記憶させる分だけ大きくなる。また、読み出し部 142 は、バッファメモリ 141 に記憶されたパケットを読み出すときに、時刻も合わせて読み出す。そして、読み出し部 142 は、パケットを読み出したときのタイマ 146 が示す時刻と、バッファメモリ 141 から読み出した時刻との差を算出する。算出した差が最大記憶時間レジスタ 147 に設定された最大記憶時間よりも小さな場合、読み出し部 142 は、読み出したパケットを出力する。一方、算出した差が最大記憶時間レジスタ 147 に設定された最大記憶時間以上である場合、読み出し部 142 は、読み出したパケットを破棄する。

30

【0042】

以下に、上記のように構成されたパケットスイッチ 10 の動作について説明する。ここでは、運用サーバのソフトウェアやハードウェアを計画的に交換する場合について説明する。

40

【0043】

図 3 は、図 1 および図 2 に示したパケットスイッチ 10 の動作の一例を説明するための図である。

【0044】

図 3 に示すように、ネットワーク 50、サーバ 20、サーバ 30 およびサーバ 40 のそれぞれは、ポート 11a, 11b, 11c, 11d を介してスイッチ部 11 に接続されている。なお、ここでは、サーバ 20 を運用サーバとし、サーバ 30 をバックアップサーバとする。

【0045】

50

通常運用中は、宛先アドレスがサーバ 20 を示すパケットがポート 11 a、ポート 11 c、ポート 11 d にて受け付けられると、受け付けられたパケットをポート 11 b から出力するようにルーティングテーブルが設定されている。これは、図 3 において (1) および (5) で示されている。

【0046】

また、通常運用中は、バッファメモリ 141 にはパケットが記憶されておらず、読み出し部 142 は動作を停止している。

【0047】

(手順 1) サーバ 20 は、ネットワーク 50 に接続された端末等やサーバ 40 に対して機能を提供しながらサーバ 30 を起動させる。

【0048】

(手順 2) サーバ 20 は、機能の提供を継続しながら様々な環境データをサーバ 30 へコピーする。

【0049】

ここまでの動作は、本実施形態のパケットスイッチ 10 を用いない場合でも同様である。

【0050】

(手順 3) サーバ 20 は、ネットワーク 50 に接続された端末等やサーバ 40 に対する機能の提供を停止する。このとき、サーバ 20 は、ポート 11 c を示すポート情報とサーバ 20 のアドレスとを含む停止要求メッセージを制御部 12 へ送信する。制御部 12 は、サーバ 20 から送信された停止要求メッセージをポート 11 e を介して受け付けると、サーバ 30、40 およびネットワーク 50 と接続されたポート 11 a、11 c、11 d のうちポート 11 c 以外のポート 11 a、11 d にて受け付けられ、宛先アドレスがサーバ 20 を示すパケットをポート 11 f から出力するようにルーティングテーブルを設定する。これにより、図 3 の (2) に示すように、ネットワーク 50 およびサーバ 40 から送信され、宛先アドレスがサーバ 20 を示すパケットは、大規模バッファ部 14 へ転送されることになる。なお、受け付けた停止要求メッセージに含まれるポート情報が示すポート 11 c にて受け付けられ、宛先アドレスがサーバ 20 を示すパケットは引き続き、図 3 の (5) に従って転送される。つまり、サーバ 30 から送信されたパケットは、大規模バッファ部 14 へ転送されることはない。従って、サーバ 20 とサーバ 30 との間では、通信を行うことが可能である。また、図 3 には明示的に示していないが、ポート 11 a ~ 11 f と制御部 12 との間の通信については、制御部 12 が管理しており、サーバ 20 から送信された停止要求メッセージの有無によらず、転送の停止を行わないものとする。

【0051】

(手順 4) サーバ 20 は、処理中のサービスに関する状態データをサーバ 30 へコピーする。

【0052】

(手順 5) サーバ 20 は、サーバ 30 を運転状態に移行させる。

【0053】

手順 4 および手順 5 の動作は、本実施形態のパケットスイッチ 10 を用いない場合でも同様である。

【0054】

(手順 6) 停止解除要求メッセージが制御部 12 へ送信される。このとき、サーバ 20 がポート 11 c を示すポート情報を含む停止解除要求メッセージを送信する方法と、サーバ 30 が停止解除要求メッセージを送信する方法とがある。前者の方法では、第 2 のポート、すなわちパケットの新たな転送先が明示的に指定されている。前者の方法では、運用サーバだけではなく、例えばサーバ群の運転状態やサーバの切り替えを管理するサーバなどから、パケットの転送の停止の解除を指示する場合に適している。一方、後者の方法において制御部 12 は、ポート 11 a ~ 11 f のうち停止解除要求メッセージが受け付けられたポートを識別することによって第 2 のポート、すなわちパケットの新たな転送先を決

定することになる。後者の方法では、バックアップサーバが接続されているポートを示すポート情報等をバックアップサーバに設定しておかなくても、パケットの転送を再開することができる。制御部 12 は、サーバ 20 またはサーバ 30 から送信された停止解除要求メッセージをポート 11 e を介して受け付けると、ポート 11 f にて受け付けられたパケットをポート 11 c から出力するようにルーティングテーブルを設定する。これは図 3 に示す (3) を設定することとなる。そして、制御部 12 は、読み出し開始の指示を大規模バッファ部 14 へ出力する。これにより、バッファメモリ 141 に記憶されているパケットが出力されてポート 11 f にて受け付けられ、受け付けられたパケットがポート 11 c から出力されることになる。つまり、バッファメモリ 141 に記憶されているパケットがサーバ 30 へ転送される。なお、この時点では、ポート 11 a, 11 d にて受け付けられ、宛先アドレスがサーバ 20 を示すパケットは引き続き、ポート 11 f から出力される。これは、新たに受け付けられたパケットが、大規模バッファ部 14 に記憶されているパケットに引き続いて、サーバ 30 へ転送されるようにするためである。これにより、パケットの順序の逆転などを回避することができる。なお、ポート 11 b にて受け付けられ、宛先アドレスがサーバ 20 を示すパケットについては、そのまま折り返してポート 11 b から出力する方法と、破棄する方法が考えられるが、いずれの処理を行うかはパケットスイッチ 10 の利用者が設定可能とする。

10

【0055】

上記の手順 6 により、サーバ 20 が運用サーバとして行っていた処理を引き継いだサーバ 30 は機能を提供できることとなる。

20

【0056】

次に、大規模バッファ部 14 を経由しているパケット転送を、大規模バッファ部 14 を経由しないパケット転送の状態に戻すために、以下に示す手順 7 が実行される。

【0057】

(手順 7) 制御部 12 は、バッファメモリ 141 の空き状況を監視する。そして、制御部 12 は、バッファメモリ 141 に記憶されたパケットがなくなると、ポート 11 a およびポート 11 d にて受け付けられ、宛先アドレスがサーバ 20 を示すパケットをポート 11 c から出力するようにルーティングテーブルを設定する。これにより、図 3 の (4) に示すように、ネットワーク 50 およびサーバ 40 から送信され、宛先アドレスがサーバ 20 を示すパケットは、サーバ 30 へ転送されることになる。そして、制御部 12 は、読み出し部 142 の動作を停止させる。これにより、大規模バッファ部 14 は、別のサーバの切り替えなどに対応することが可能な状態となる。

30

【0058】

このように本実施形態においてパケットスイッチ 10 は、停止要求メッセージが受信されると、複数のポートのうち当該停止要求メッセージに含まれるポート情報が示すポート以外のポートにて受け付けられ、当該停止要求メッセージに含まれるアドレスを宛先アドレスとするパケットを、ポート 11 f から出力するようにルーティングテーブルを設定する。

【0059】

これにより、運用サーバおよびバックアップサーバがともに機能を提供していない間に運用サーバへ送信されたパケットを一時的に大規模バッファ部 14 に記憶させることができる。

40

【0060】

大規模バッファ部 14 に一時的に記憶されたパケットは、バックアップサーバによる機能の提供が開始された後に、バックアップサーバに転送されて処理される。そのため、サーバによる機能の提供の高信頼性を確保することが可能となる。

【0061】

また、バッファメモリ 141 を備える大規模バッファ部 14 は、スイッチ部 11 と接続されており、ポート 11 a ~ 11 e のいずれにおいて受け付けられたパケットも一時的に記憶することができる。そのため、複数のサーバでバッファメモリ 141 を共有してサー

50

バの切り替えを行うことが可能である。従って、例えば複数のポート毎に設けられているバッファの容量を拡張して一時的にパケットを記憶させる場合と比べ、サーバによる機能の提供の高信頼性を経済的に確保することが可能となる。

【0062】

また、バックアップサーバが機能を提供するための準備を完了するまでの間には、非常に多くのパケットがバッファメモリ141に記憶される可能性がある。バックアップサーバが機能の提供を開始した直後に、バッファメモリ141に記憶された非常に多くのパケットが一斉に出力されると、バックアップサーバが輻輳状態に陥ってしまう。

【0063】

これは、レートレジスタ145に設定される所定のレートを、バックアップサーバが輻輳状態に陥らないようなレートにすることによって回避することができる。

【0064】

また、バックアップサーバが機能を提供するための準備を完了するまでの間には、様々な種類のパケットが様々な端末から送信されている可能性がある。バックアップサーバが機能の提供を開始したときに、これらのパケットの全てを処理するのではなく、これらのパケットのうち機能の提供の継続に不可欠なパケットのみを処理するような対応を行うことも考えられる。

【0065】

パターンテーブル148に設定された所定のパターンと一致するヘッダ部分のパケットだけをバッファメモリ141に記憶させることにより、機能の提供の継続に不可欠なパケットだけをバックアップサーバに転送することが可能となる。

【0066】

また、バックアップサーバが機能を提供するための準備を完了するまでの間に、端末等が運用サーバに向けてパケットを送信した場合、そのパケットを送信した端末等は、一定時間を経過してもそのパケットに対する応答がないと、そのパケットの再送することがある。

【0067】

端末等がこのような動作を実行する場合、バックアップサーバが機能の提供を開始したときに、端末が再送を行うほど古いパケットについては処理をしないという対応方法も考えられる。

【0068】

タイマ146および最大記憶時間レジスタ147を備えることにより、バックアップサーバが処理をする必要のないパケットをバックアップサーバへ転送することを回避することができる。

【0069】

また、上述したように、停止要求メッセージにおいては、大規模バッファ部14へのパケットの転送を明示する必要がない。従って、大規模バッファ部14が接続されているポートを識別するポート情報を運用サーバ等に通知したり、運用サーバ等がポート情報を管理したりする必要がない。

【0070】

また、停止要求メッセージは、大規模バッファ部14へのパケットの転送を指示するメッセージではなく、パケットの転送の停止を指示するメッセージである。これにより、大規模バッファ部14の有無に応じて停止要求メッセージの形式を変更しなくても済む。すなわち、パケットスイッチが停止要求メッセージを受信したとき、そのパケットスイッチが大規模バッファ部14を備えていれば大規模バッファ部14にパケットを転送し、大規模バッファ部14を備えていなければパケットを破棄するということが可能となる。

【0071】

なお、パケットが大規模バッファ部14に転送されているときに、バッファメモリ141に空きエリアがなくなったときには、バッファメモリ141に記憶されている複数のパケットうち最も早い時期に記憶されたパケットを破棄し、大規模バッファ部14に新たに

10

20

30

40

50

転送されてきたパケットを記憶するようにしてもよい。これは、バッファメモリ 141 に記憶された時期が早いパケットについては、これを送信した端末が再送の手順などを開始している可能性があるためである。

【0072】

これにより、バックアップサーバが機能を提供する際に、機能の提供に必要な可能性が高いパケットを大規模バッファ部 14に残しておくことができる。

【0073】

また、上述した実施形態において停止要求メッセージは、バックアップサーバが接続されたポートのポート情報を含むものとして説明した。ただし、機能を提供するサーバを切り替えるために運用サーバは、切り替えの工程を管理する運用管理サーバ等と通信をする可能性もある。そのため、停止要求メッセージは、バックアップサーバが接続されたポートを示すポート情報だけではなく、このような運用管理サーバ等が接続されたポートのポート情報等も含むようにしてもよい。

【0074】

ここで、図3を参照しながら説明したのは、運用サーバのソフトウェアやハードウェアを計画的に交換する場合に、機能を提供するサーバを運用サーバからバックアップサーバへ切り替えるときのパケットスイッチ10の動作であった。

【0075】

次に、運用サーバに障害等が発生したときのパケットスイッチ10の動作について説明する。

【0076】

図4は、図1および図2に示したパケットスイッチ10の動作の他の例を説明するための図である。ここでは、運用サーバがソフトウェア障害を検出し、運用サーバ内のソフトウェアが再起動されるときのパケットスイッチ10の動作について説明する。

【0077】

図4に示すように、ネットワーク50、サーバ20およびサーバ40のそれぞれは、ポート11a, 11b, 11dを介してスイッチ部11に接続されている。また、ここでは、サーバ20を運用サーバとする。

【0078】

通常運用中は、宛先アドレスがサーバ20を示すパケットがポート11a、ポート11dにて受け付けられると、受け付けられたパケットをポート11bから出力するようにルーティングテーブルが設定されている。この状態は、図4において(6)で示されている。

【0079】

また、通常運用中は、バッファメモリ141にはパケットが記憶されておらず、読み出し部142は動作を停止している。

【0080】

サーバ20がソフトウェア障害を検出し、ソフトウェアの再起動を開始すると、再起動の開始時に停止要求メッセージを制御部12へ送信する。これにより、制御部12は、ポート11a, 11dにて受け付けられ、宛先アドレスがサーバ20を示すパケットをポート11fから出力するようにルーティングテーブルを設定する。これにより、図4の(7)に示すように、ネットワーク50およびサーバ40から送信され、宛先アドレスがサーバ20を示すパケットは、大規模バッファ部14へ転送されることになる。

【0081】

次に、サーバ20がソフトウェアの再起動を完了すると、サーバ20は、停止解除要求メッセージを制御部12へ送信する。これにより、制御部12は、図4の(8)に示すように、ポート11fにて受け付けられたパケットをポート11bから出力するようにルーティングテーブルを設定する。そして、制御部12は、読み出し開始の指示を大規模バッファ部14へ出力する。これにより、サーバ20のソフトウェアを再起動している間にサーバ20へ送信されたパケットが、ソフトウェアの再起動後に処理される。

【0082】

10

20

30

40

50

そして、制御部 1 2 は、バッファメモリ 1 4 1 の空き状態を監視する。制御部 1 2 は、バッファメモリ 1 4 1 に記憶されたパケットがなくなると、図 4 の (6) に示すように、ポート 1 1 a , 1 1 d にて受け付けられ、宛先アドレスがサーバ 2 0 を示すパケットをポート 1 1 b から出力するようにルーチングテーブルを設定する。これにより、通常運用中の状態に戻ることになる。

【 0 0 8 3 】

図 4 を参照しながら説明した動作の場合、運用サーバにおいて障害が発生してから障害が検出されるまでの間にサーバ 2 0 が受信したパケットについては正常に処理されないが、障害が検出された後のパケットについては正常に処理される。

【 0 0 8 4 】

従って、ネットワーク 5 0 に接続された端末等から見ると、見掛け上の中断時間は、パケットスイッチ 1 0 を用いない場合に比べて大きく短縮されることになる。これは、運用サーバにおいてソフトウェアの再起動を行う場合だけでなく、運用サーバ内のデータの再初期化を行ったり、運用サーバ内の 2 重化した装置を切り替えたりするような場合でも同様である。

【 0 0 8 5 】

つまり、運用サーバに障害等が発生した場合にも、サーバによる機能の提供の高信頼性を確保することが可能となる。

【 0 0 8 6 】

なお、本発明は、従来のパケットスイッチに大規模バッファ部 1 4 を追加し、制御部 1 2 のソフトウェアを変更するなどによって簡単に実現が可能であり、多様なパケットスイッチへの適用が可能である。

【 0 0 8 7 】

特に、サーバ間での状態データの引き継ぎを行いながら、機能提供の継続性を実現するようなサーバを多数収容する可能性が高く、かつ、サーバの機能提供の中断により非常に多数の利用者に影響をあたえてしまう可能性が高い大規模データセンタを構成するパケットスイッチなどに適用されることが想定される。

【 符号の説明 】

【 0 0 8 8 】

- 1 0 パケットスイッチ
- 1 1 スイッチ部
- 1 1 a ~ 1 1 f ポート
- 1 2 制御部
- 1 3 ルーチングテーブル部
- 1 4 大規模バッファ部
- 2 0 , 3 0 , 4 0 サーバ
- 5 0 ネットワーク
- 1 4 1 バッファメモリ
- 1 4 2 読み出し部
- 1 4 3 書き込み部
- 1 4 4 管理部
- 1 4 5 レートレジスタ
- 1 4 6 タイマ
- 1 4 7 最大記憶時間レジスタ
- 1 4 8 パターンテーブル

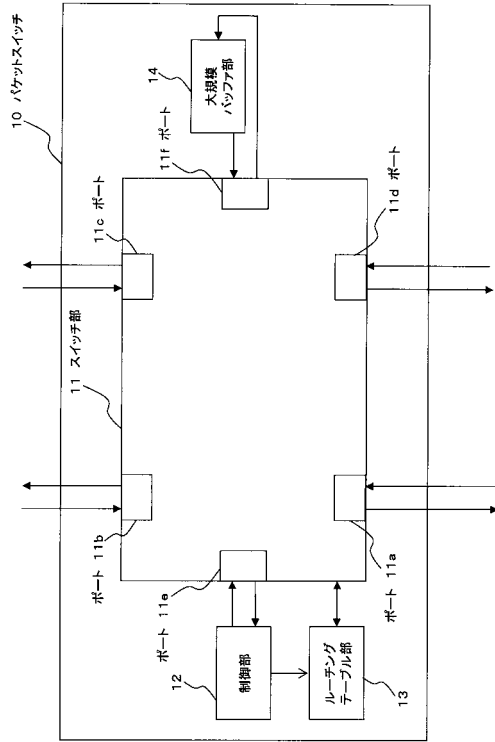
10

20

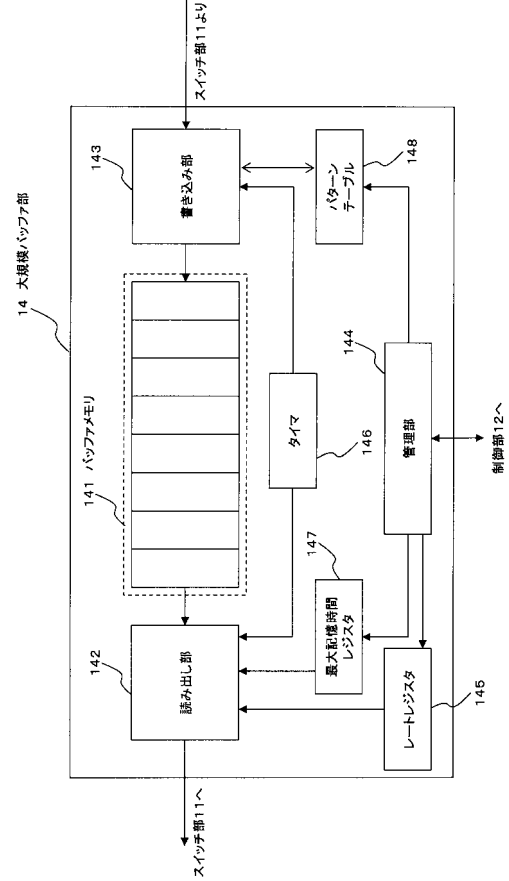
30

40

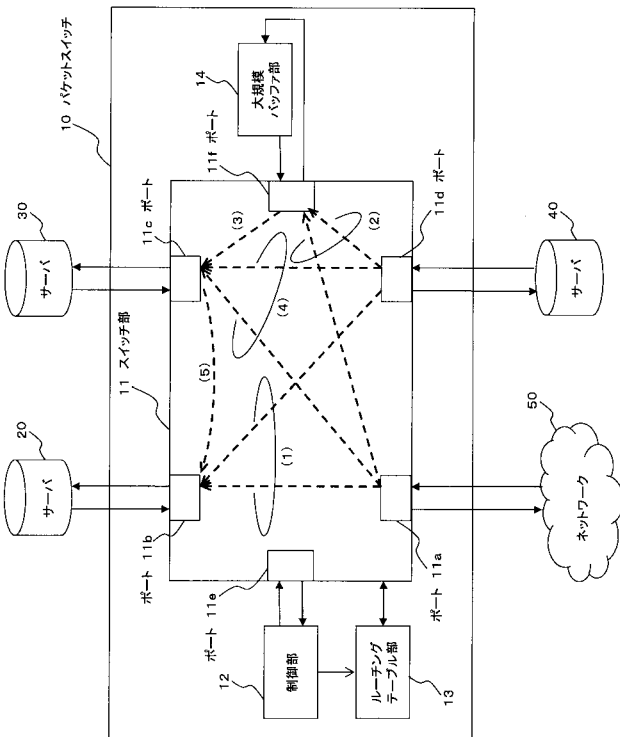
【図 1】



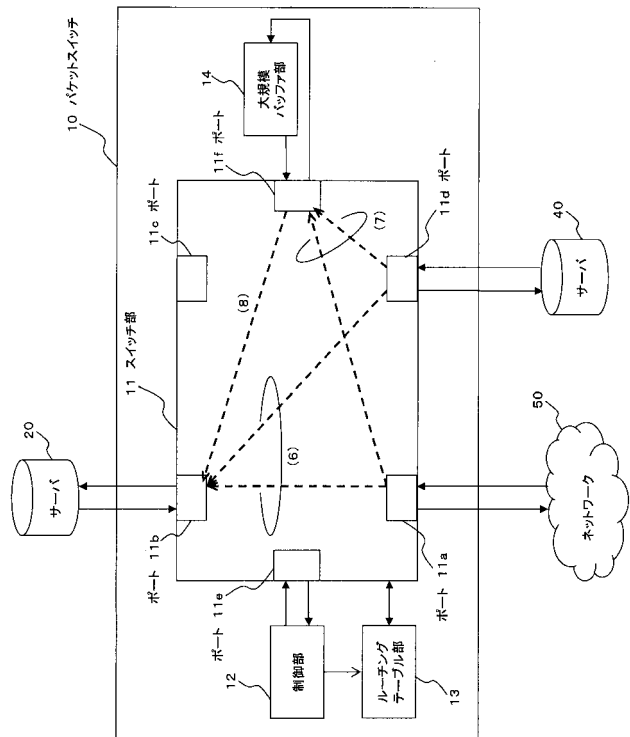
【図 2】



【図 3】



【図 4】



【図 5】

