

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5835043号  
(P5835043)

(45) 発行日 平成27年12月24日(2015.12.24)

(24) 登録日 平成27年11月13日(2015.11.13)

(51) Int.Cl. F I  
H O 4 L 12/723 (2013.01) H O 4 L 12/723

請求項の数 8 (全 17 頁)

(21) 出願番号	特願2012-62891 (P2012-62891)	(73) 特許権者	000005223 富士通株式会社
(22) 出願日	平成24年3月19日 (2012. 3. 19)		神奈川県川崎市中原区上小田中4丁目1番 1号
(65) 公開番号	特開2013-197896 (P2013-197896A)	(74) 代理人	100070150 弁理士 伊東 忠彦
(43) 公開日	平成25年9月30日 (2013. 9. 30)	(74) 代理人	100146776 弁理士 山口 昭則
審査請求日	平成26年12月4日 (2014. 12. 4)	(72) 発明者	藤井 雅章 福岡県福岡市早良区百道浜2丁目2番1号 富士通九州ネットワークテクノロジーズ 株式会社内

最終頁に続く

(54) 【発明の名称】 リスタート方法及びノード装置

(57) 【特許請求の範囲】

【請求項1】

ラベル転送ネットワークでリスタートを行うリスタート方法であって、  
リスタートを行うリスタートノードは、前記リスタートノードの処理能力に基づいたノード数の隣接ノードを選択してリスタートグループとし、前記リスタートグループの隣接ノードに対してリスタートを通知し、

前記リスタートノードは、前記リスタートグループの隣接ノードから通知されたパスのラベル値を基に、パスとラベルを対応付けるラベルテーブル内で前記通知されたパスに対応するエントリを復元し、

前記リスタートの通知と前記通知されたパスに対応するエントリの復元とを繰り返して前記ラベルテーブルの復元を行う  
ことを特徴とするリスタート方法。

【請求項2】

請求項1記載のリスタート方法において、  
前記リスタートノードに、優先度毎に、隣接ノードと、前記隣接ノードと自ノード間のパス数を登録したノード情報テーブルを設け、

前記ノード情報テーブルを優先度が高い順に検索して前記リスタートノードの処理能力に基づいたノード数の隣接ノードを選択する  
ことを特徴とするリスタート方法。

【請求項3】

10

20

請求項 2 記載のリスタート方法において、

前記選択する隣接ノードは、前記ノード情報テーブルを優先度が高い順に検索して、前記隣接ノード毎に前記隣接ノードと自ノード間のパス数を加算し、合計値が前記リスタートを行うノード装置の処理能力に応じた所定値を超過する直前までに検索された隣接ノードである

ことを特徴とするリスタート方法。

【請求項 4】

請求項 3 記載のリスタート方法において、

前記ノード情報テーブルは、隣接ノードから新たなパスが設定されたとき、前記新たなパスの優先度と、前記隣接ノードが既に登録されている前記ノード情報テーブルの優先度とを比較して前記新たなパスの優先度が高い場合には、前記隣接ノードを登録する前記ノード情報テーブルの優先度を前記新たなパスの優先度に変更する

ことを特徴とするリスタート方法。

【請求項 5】

ラベル転送ネットワークでリスタートを行うノード装置であって、

自ノード装置の処理能力に基づいたノード数の隣接ノードを選択してリスタートグループとし、前記リスタートグループの隣接ノードに対してリスタートを通知する通知手段と

、  
前記リスタートグループの隣接ノードから通知されたパスのラベル値を基に、パスとラベルを対応付けるラベルテーブル内で前記通知されたパスに対応するエントリを復元する復元手段と、を有し、

前記通知手段による通知と前記復元手段による復元とを繰り返して前記ラベルテーブルの復元を行う

ことを特徴とするノード装置。

【請求項 6】

請求項 5 記載のノード装置において、

優先度毎に、隣接ノードと、前記隣接ノードと自ノード間のパス数を登録したノード情報テーブルを有し、

前記通知手段は、前記ノード情報テーブルを優先度が高い順に検索して前記自ノード装置の処理能力に基づいたノード数の隣接ノードを選択する

ことを特徴とするノード装置。

【請求項 7】

請求項 6 記載のノード装置において、

前記通知手段は、前記ノード情報テーブルを優先度が高い順に検索して、前記隣接ノード毎に前記隣接ノードと自ノード間のパス数を加算し、合計値が前記リスタートを行うノード装置の処理能力に応じた所定値を超過する直前までに検索された隣接ノードを選択する

ことを特徴とするノード装置。

【請求項 8】

請求項 7 記載のノード装置において、

隣接ノードから新たなパスが設定されたとき、前記新たなパスの優先度と、前記隣接ノードが既に登録されている前記ノード情報テーブルの優先度とを比較して前記新たなパスの優先度が高い場合には、前記隣接ノードを登録する前記ノード情報テーブルの優先度を前記新たなパスの優先度に変更する変更手段を

有することを特徴とするノード装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ラベル転送ネットワークでリスタートを行うリスタート方法及びノード装置に関する。

10

20

30

40

50

## 【背景技術】

## 【0002】

近年、パケットにラベルを付与してデータを転送するMPLS (Multi-Protocol Label Switching) を用いたラベル転送ネットワークが一般化している。

## 【0003】

一方、RSVP (Resource reSerVation Protocol) はネットワーク上で送信先までの帯域を予約し、通信品質を確保するプロトコルである。現在のインターネットではどのデータも平等に扱われているため、テレビ会議やリアルタイムの動画配信など、即時性、連続性が求められるトラフィックを優先させる目的で開発されたものである。

10

## 【0004】

技術標準文書RFC (Request For Comments) 3209などで規定されたRSVP-TE (RSVP-Traffic Engineering) はRSVPに対してMPLSで利用するラベルを予約できるように拡張したものであり、隣接装置故障やリンク故障の検出のためにハローメッセージが定義されている。

## 【0005】

グレースフルリスタート (Graceful-Restart) はRSVP-TEシグナリングを用いるMPLS通信装置で用いられる。すなわち、コントロールプレーンでハローメッセージやPathメッセージやResvメッセージが途切れた場合であっても、データプレーンの通信を継続させ、コントロールプレーンの障害が復旧した後で状態情報 (Path状態情報、Resv状態情報) を障害前の状態に復旧させるために用いられる。なお、ハローメッセージは障害検出などに使用され、Pathメッセージはパスの設定などに使用され、Resvメッセージは受信応答などに使用される。

20

## 【0006】

グレースフルリスタートは、リスタートしたときに隣接ノードからのメッセージを基にリスタート前の状態を復元する。このため、多数の隣接ノードとの間で多数のLSP (Label Switched Path) が設定されているときにリスタートすると、リスタートノードで大量の処理が発生する。

## 【0007】

グレースフルリスタートではリスタートしたことを隣接ノードへ示すためにリスタートノードからハローメッセージを送信するが、このハローメッセージが予め設定されているリスタートタイム内で隣接ノードに受信されないと、新たなラベルを取得する処理が実施され、新たなラベルを取得する処理の間はデータプレーンで通信断が発生する。

30

## 【0008】

図1にグレースフルリスタートの一例のシーケンス図を示す。ここでは、隣接ノードをリスタートノードの上流、下流に各1ノードずつ接続した状態で記載しているが、実際にはリスタートノードの上流、下流それぞれに複数のノードが接続されている。

## 【0009】

シーケンスSQ1: リスタートノードはハローメッセージ内で「Restart\_Capオブジェクト」を送信し、隣接ノードに対しグレースフルリスタートであることを通知する。「Restart\_Capオブジェクト」で予めリスタートタイムを通知し、再起動にかかる最大時間を通知する。上流及び下流の隣接ノードは「Restart\_Capオブジェクト」を受信すると、送信元ノードがグレースフルリスタート可能であることを認識する。

40

## 【0010】

シーケンスSQ2: 通常通りLSPを作成する。そして、PathメッセージやResvメッセージでラベル交換を行う。

## 【0011】

シーケンスSQ3: リスタートノードは再起動した後にラベルテーブルの全エントリの

50

「stale」にマーキングを行う。リスタートノードはハローメッセージ中の「Restart\_Capオブジェクト」でリカバリタイムを指定し、このリカバリタイム内でラベルテーブルをリフレッシュして復元するリカバリ処理を行うことを上流及び下流の隣接ノードに通知する。

【0012】

シーケンスSQ4：上流及び下流の隣接ノードはハローメッセージの「Restart\_Capオブジェクト」を受信すると、ラベルテーブルのうちリスタートノードに対応するエントリの「stale」にマーキングを行う。

【0013】

シーケンスSQ5：上流の隣接ノードはPathメッセージの「Recovery\_Label\_Object」を用い、リフレッシュを行うLSPでリスタートノードが配布していたラベルをリスタートノードに通知する。

【0014】

シーケンスSQ6：リスタートノードは通知されたPathメッセージの「Recovery\_Label\_Object」のラベル値を基にラベルテーブルを参照し、下流の隣接ノードから配布されていたラベル値を取得する。それを基にLSPと上流のノードに配布したラベル値、下流のノードから配布されたラベル値を取得し、RSVP-TEの状態を復旧し、かつ、ラベルテーブルの「stale」を削除する。

【0015】

図2にラベルテーブルの一例の構成図を示す。ラベルテーブルは、LSP番号毎に、入力ラベル値(in)と、出力ラベル値(out)と、ネクストホップ(NH)と、staleとを有している。シーケンスSQ6では例えば上流隣接ノードから入力ラベル値=20のPathメッセージがリスタートノードに供給される。リスタートノードでは入力ラベル値=20でラベルテーブルを参照して出力ラベル値=10とネクストホップ=IPaを得る。リスタートノードはラベルテーブルの当該エントリの「stale」のマーキングを削除し、Pathメッセージのラベル値を出力ラベル値=10に変更し、ネクストホップ=IPaに対応する下流隣接ノードに向けて送信する。

【0016】

シーケンスSQ7：リスタートノードはPathメッセージの「Suggested\_Label\_Object」を用い、当該LSPで下流の隣接ノードが配布していたラベルを下流の隣接ノードに通知する。これを受信した下流の隣接ノードはラベルテーブルの「stale」を削除する。この後、下流の隣接ノードからのResvメッセージがリスタートノードを経て上流の隣接ノードに通知されると、上流の隣接ノードはラベルテーブルの「stale」のマーキングを削除する。

【0017】

シーケンスSQ8：全LSPについてシーケンスSQ4～SQ7を繰り返すことにより、リスタートノードはラベルテーブルの全状態を復元する。リスタートノードはリカバリタイムが経過すると、リフレッシュされていないLSPをラベルテーブルから削除するために、「stale」にマーキングが残っているエントリを削除する。

【0018】

シーケンスSQ9：隣接ノードはリスタートノードから受信したリカバリタイムが経過すると、リフレッシュされていないLSPをラベルテーブルから削除するために、「stale」にマーキングされたエントリを削除する。

【0019】

ところで、通信ネットワーク内の回線上における障害の発生を検出した場合に、障害が発生した回線を通過する複数のトンネルの切断メッセージを集約して、障害の発生を回線単位で隣接ノードに通知するトンネル障害通知装置が提案されている(例えば特許文献1参照)。

【先行技術文献】

【特許文献】

10

20

30

40

50

【 0 0 2 0 】

【特許文献 1】特開 2 0 0 6 - 3 3 1 2 4 号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 2 1 】

図 3 にグレースフルリスタートのリカバリ処理の輻輳例のシーケンス図を示す。図 1 ではリスタートノードの隣接ノードは上流、下流に各 1 ノードずつで記載しているが、実際には複数のノードが接続されている。

【 0 0 2 2 】

この場合、再起動時にリスタートノードはシーケンス S Q 4 で全ての隣接ノードにハローメッセージを送信し、全ての隣接ノードよりハローメッセージを受信する。

10

【 0 0 2 3 】

シーケンス S Q 5 でリスタートノードは「R e c o v e r y L a b e l O b j e c t」付きの P a t h メッセージによりラベルを受信する。

【 0 0 2 4 】

シーケンス S Q 6 でリスタートノードは L S P のリカバリ処理を行ってラベルテーブルを復元する。

【 0 0 2 5 】

シーケンス S Q 7 でリスタートノードは下流隣接ノードがあれば P a t h メッセージを送信する。

20

【 0 0 2 6 】

上記のシーケンス S Q 5 において上流の全ての隣接ノードから「R e c o v e r y L a b e l O b j e c t」付き P a t h メッセージを受信してシーケンス S Q 6 のリカバリ処理を行うので、リカバリ処理が輻輳する場合がある。

【 0 0 2 7 】

なお、技術標準文書 R F C 3 4 7 3 では、上流の隣接ノードでの「R e c o v e r y L a b e l O b j e c t」付きのメッセージを分散して送信することなどが規定されているが、リスタートノードでの対処については言及されておらず、リスタートノードでは複数の上流隣接ノードから同時に「R e c o v e r y L a b e l O b j e c t」付きの P a t h メッセージを受信する。

30

【 0 0 2 8 】

その結果、リカバリタイム内でラベルテーブルの全エントリの復元ができない場合があり、復元できなかったリスタートノードは再度ラベル取得処理を実行するため、データプレーンにおける通信断が発生するという問題があった。

【 0 0 2 9 】

近年のネットワークでは、1つのノードに収容する L S P 数が増加しているため、シーケンス S Q 4 のハローメッセージ送受後に、1つの隣接ノードからリスタートノードに返ってくるシーケンス S Q 5 の「R e c o v e r y L a b e l O b j e c t」付き P a t h メッセージ数が増加している。また、各ノードの制御処理速度が上がったため、シーケンス S Q 4 のハローメッセージ送受後に、各隣接ノードからリスタートノードに返ってくるシーケンス S Q 5 の「R e c o v e r y L a b e l O b j e c t」付き P a t h メッセージ受信までの遅延差が小さくなっている。また、L S P 別の P a t h メッセージ送信間隔が短くなっている。

40

【 0 0 3 0 】

以上の状況から、各ノードが返すシーケンス S Q 5 の「R e c o v e r y L a b e l O b j e c t」付き P a t h メッセージ受信によるシーケンス S Q 6 の L S P 処理が短時間に集中し、リカバリ処理の輻輳が生じるという問題があった。

【 0 0 3 1 】

開示のリスタート方法は、リカバリ処理の輻輳を抑制することを目的とする。

【課題を解決するための手段】

50

## 【0032】

開示の一実施形態によるリスタート方法は、ラベル転送ネットワークでリスタートを行うリスタート方法であって、

リスタートを行うリスタートノードは、前記リスタートノードの処理能力に基づいたノード数の隣接ノードを選択してリスタートグループとし、前記リスタートグループの隣接ノードに対してリスタートを通知し、

前記リスタートノードは、前記リスタートグループの隣接ノードから通知されたパスのラベル値を基に、パスとラベルを対応付けるラベルテーブル内で前記通知されたパスに対応するエントリを復元し、

前記リスタートの通知と前記通知されたパスに対応するエントリの復元とを繰り返して前記ラベルテーブルの復元を行う。

10

## 【発明の効果】

## 【0033】

本実施形態によれば、リカバリ処理の輻輳を抑制することができる。

## 【図面の簡単な説明】

## 【0034】

【図1】グレースフルリスタートの一例のシーケンス図である。

【図2】ラベルテーブルの一例の構成図である。

【図3】グレースフルリスタートのリカバリ処理の輻輳例のシーケンス図である。

【図4】ノード装置の一実施形態の構成図である。

20

【図5】グレースフルリスタートの一実施形態のシーケンス図である。

【図6】ラベルテーブルの一実施形態の構成図である。

【図7】ノード情報テーブルの一実施形態の構成図である。

【図8】テーブル更新処理の一実施形態のフローチャートである。

【図9】リスタート順決定処理の一実施形態のフローチャートである。

【図10】グレースフルリスタートのリカバリ処理の一実施形態のシーケンス図である。

【図11】Session\_Attributeオブジェクトのフォーマットを示す図である。

## 【発明を実施するための形態】

## 【0035】

30

以下、図面に基づいて実施形態を説明する。

## 【0036】

## &lt;ノード装置の構成&gt;

図4にノード装置の一実施形態の構成図を示す。図4において、ノード装置は複数の入出力ユニット11-1~11-Nと、スイッチ部12と、制御部13を有している。入出力ユニット11-1~11-Nそれぞれはデータ受信部15、テーブルメモリ16、ルート&フォワード制御部17、データ送信部18を有している。データ受信部15は複数のポートから入力される入力データの入力データチェック、ラベルチェック、制御パケットの抽出などを行う。テーブルメモリ16にはラベルテーブルが格納される。

## 【0037】

40

データ受信部15は入力パケットのラベルでラベルテーブルを参照して得たネクストホップ(NH)をルート&フォワード制御部17に供給し、当該パケットをスイッチ部12に供給する。ルート&フォワード制御部17からの制御は制御部13内のCPU21に通知され、CPU21によってスイッチ部17のスイッチングが制御されて、当該パケットはネクストホップに応じた入出力ユニット11-1~11-Nのデータ送信部18に供給される。また、データ受信部15は抽出したハローメッセージやPathメッセージ又はResvメッセージなどの制御パケットをルート&フォワード制御部17に供給する。

## 【0038】

データ送信部18はルート&フォワード制御部17の制御に応じて出力するパケットのラベルを変更し、また、ルート&フォワード制御部17で生成された制御パケットを挿入

50

されて複数のポートから出力する。

【0039】

制御部13のCPU21はメモリ22に格納されている各種プログラムを実行することで、スイッチ部17のスイッチング制御処理、グレースフルリスタート処理、テーブル更新処理、リスタート順決定処理などを実施する。なお、メモリ22にはテーブル更新処理、リスタート順決定処理で使用されるノード情報テーブルが格納される。

【0040】

<グレースフルリスタートの一実施形態のシーケンス>

図5にグレースフルリスタートの一実施形態のシーケンス図を示す。ここでは、隣接ノードを上流、下流に各1ノードずつで記載しているが、実際には複数のノードが接続されている。

10

【0041】

シーケンスSQ11：リスタートノードはハローメッセージ内で「Restart\_Capオブジェクト」を送信し、隣接ノードに対し、グレースフルリスタートであることを通知する。「Restart\_Capオブジェクト」で予めリスタートタイムを通知し、再起動にかかる最大時間を通知する。上流及び下流の隣接ノードは「Restart\_Capオブジェクト」を受信すると、送信元ノードがグレースフルリスタート可能であることを認識する。

【0042】

シーケンスSQ12：通常通りLSPを作成する。そして、PathメッセージやResvメッセージでラベル交換を行う。なお、上流の隣接ノードは当該Pathメッセージに優先度(Priority)を含めてリスタートノードに送信する。

20

【0043】

ここで、優先度の情報はMPLSのラベル配布プロトコルであるLDP(Label Distribution Protocol)メッセージ内のSession\_Attributeのセットアッププライオリティ値や、ホールドプライオリティ値を使用して送信する。図11にSession\_Attributeオブジェクトのフォーマットを示す。セットアッププライオリティ値はLSPトンネルを確立する際の優先度を示す。ホールドプライオリティ値は他のLSPトンネルによる経路使用の競合が発生した際に原稿のパスを保持する優先度を示す。この他に、PathメッセージやResvメッセージ等のLDPメッセージにおける未使用ビットを使用して優先度の送信を行っても良い。

30

【0044】

なお、LSPを設定する場合にはPathメッセージのTIME\_VALUEでリフレッシュ間隔が設定され、その後、上記リフレッシュ間隔内にLSPを確認してリフレッシュする必要がある。このため、上記Pathメッセージで明示されたリフレッシュ間隔(TIME\_VALUE)が短いLSPを優先度の高いパスと判断することも可能である。即ちリフレッシュ間隔から優先度を設定しても良い。また、上記の方法を組み合わせると優先度を設定しても良い。

【0045】

リスタートノードは従来通り上流、下流の隣接ノードとラベル交換を行い、ラベル情報(入力ラベル値(in),出力ラベル値(out))に加えて、上流ノード情報と、受信した当LSPの優先度をラベルテーブルに書き込む。そして、後述するテーブル更新処理を実行する。

40

【0046】

図6にラベルテーブルの一実施形態の構成図を示す。ラベルテーブルは、LSP番号をエントリとして、送信元のノード名と、入力ラベル値(in)と、出力ラベル値(out)と、ネクストホップ(NH)と、優先度(Priority)と、staleとを有している。

【0047】

シーケンスSQ13：リスタートノードは再起動した後にラベルテーブルの全エントリ

50

の「stale」にマーキングを行う。

【0048】

シーケンスSQ14：リスタートノードはリスタート順決定処理を実行してリスタートを行う一又は複数の隣接ノードを選択する。リスタートノードはハローメッセージ中の「Restart\_Capオブジェクト」でリカバリタイムを指定し、このリカバリタイム内に、リスタート順決定処理で選択された一又は複数の隣接ノードである上流の隣接ノードと下流の隣接ノードに対し、ラベルテーブルをリフレッシュして復元するリカバリ処理を行うことを通知する。上流及び下流の隣接ノードはハローメッセージの「Restart\_Capオブジェクト」を受信すると、ラベルテーブルのうちリスタートノードに対応するエントリの「stale」にマーキングを行う。

10

【0049】

シーケンスSQ15：上流の隣接ノードはPathメッセージの「Recovery Label Object」を用い、リフレッシュを行うLSPでリスタートノードが配布していたラベルを通知する。

【0050】

シーケンスSQ16：リスタートノードは通知されたPathメッセージの「Recovery Label Object」のラベル値を基にラベルテーブルを参照し、下流の隣接ノードから配布されていたラベル値を取得する。それを基にLSPと上流のノードに配布したラベル値、下流のノードから配布されたラベル値を取得し、RSVP-TEの状態を復旧し、かつ、ラベルテーブルの「stale」を削除する。

20

【0051】

例えば上流の隣接ノードAから入力ラベル値=20のPathメッセージがリスタートノードに供給されると、図6の第1行に示すように、リスタートノードでは入力ラベル値=20でラベルテーブルを参照して出力ラベル値=10とネクストホップ=IPaを得る。リスタートノードはラベルテーブルの当該エントリの「stale」のマーキングを削除し、Pathメッセージのラベル値を出力ラベル値=10に変更し、ネクストホップ=IPaに対応する下流の隣接ノードに向けて送信する。

【0052】

シーケンスSQ17：リスタートノードはPathメッセージの「Suggested Label Object」を用い、当該LSPで下流の隣接ノードが配布していたラベルを下流の隣接ノードに通知する。これを受信した下流の隣接ノードはラベルテーブルの「stale」を削除する。この後、下流の隣接ノードからのResvメッセージがリスタートノードを経て上流の隣接ノードに通知されると、上流の隣接ノードはラベルテーブルの「stale」のマーキングを削除する。

30

【0053】

シーケンスSQ18：全LSPについてシーケンスSQ14~SQ17を繰り返すことにより、リスタートノードは全状態を復旧する。リスタートノードはリカバリタイムが経過すると、リフレッシュされていないLSPをラベルテーブルから削除するために、「stale」にマーキングされたエントリを削除する。

【0054】

シーケンスSQ19：隣接ノードはリスタートノードから受信したリカバリタイムが経過すると、リフレッシュされていないLSPをラベルテーブルから削除するために、「stale」にマーキングされたエントリを削除する。

40

【0055】

<ノード情報テーブルの構成>

図7にノード情報テーブルの一実施形態の構成図を示す。ノード情報テーブルはメモリ22内に優先度毎に作成される。図7(A)は優先度0のノード情報テーブルを示し、図7(B)は優先度1のノード情報テーブルを示し、図7(C)は優先度2のノード情報テーブルを示す。なお、優先度の値が小さいほど優先度が高い(優先される)ものとする。各優先度のノード情報テーブルは、上流の隣接ノード名をエントリとして、当該上流の隣

50



接ノードと自ノード（つまりリスタートノード）間に作成された L S P 数を登録している。

【 0 0 5 6 】

例えば図 7 ( A ) では、第 1 のエントリが上流の隣接ノード A との間に  $N_a$  個の L S P が作成され、上記  $N_a$  個の L S P のうち最も高い優先度が優先度 0 であることを表している。また、図 7 ( B ) では、第 1 のエントリが上流の隣接ノード E との間に  $N_e$  個の L S P が作成され、上記  $N_a$  個の L S P のうち最も高い優先度が優先度 1 であることを表している。

【 0 0 5 7 】

< テーブル更新処理 >

図 8 にテーブル更新処理の一実施形態のフローチャートを示す。このテーブル更新処理はシーケンス S Q 1 2 で実行される。このテーブル更新処理は上流の隣接ノードから L S P 作成要求が通知されると実行を開始する。

【 0 0 5 8 】

図 8 において、ステップ S 1 でデータ受信部 1 5 は上流の隣接ノードから受信した L S P 作成要求に従ってテーブルメモリ 1 6 内のラベルテーブルに L S P を追加する。ここでは、例えば上流の隣接ノード x から優先度 y の L S P を追加することにする。

【 0 0 5 9 】

次に、ステップ S 2 で C P U 2 1 は L S P 作成要求を送信元である上流の隣接ノード x を用いてメモリ 2 2 内のノード情報テーブルを検索する。これにより、例えば優先度 z のノード情報テーブルで上流の隣接ノード x を検索できたものとする。なお、隣接ノード x を検索できなかった場合には優先度 z として最大値（最も低い優先度）を設定する。

【 0 0 6 0 】

次に、ステップ S 3 で C P U 2 1 は検索で得たノード情報テーブルの優先度 z と L S P 作成要求の優先度 y を比較して、 $z < y$  であればステップ S 4 に進み、 $z \geq y$  であればステップ S 6 に進む。ステップ S 4 で C P U 2 1 は優先度 y のノード情報テーブルに隣接ノード x のエントリを追加し、ステップ S 5 で C P U 2 1 は優先度 z のノード情報テーブルから隣接ノード x のエントリを削除する。

【 0 0 6 1 】

なお、上記ステップ S 4 では優先度 z のノード情報テーブルから隣接ノード x のエントリをコピーして優先度 y のノード情報テーブルに貼り付ける。つまり、隣接ノード x からの新たに作成する L S P の優先度が、それまでの隣接ノード x からの L S P の優先度より高い場合には隣接ノード x の優先度を新たに作成する L S P の優先度に更新する。

【 0 0 6 2 】

ステップ S 5 の実行後、又は、ステップ S 3 で  $z \geq y$  であればステップ S 6 に進む。ステップ S 6 で C P U 2 1 は優先度 y 又は z のノード情報テーブルで隣接ノード x のエントリの L S P 数  $N_x$  に 1 を加算する。この結果、上流の隣接ノードは所持する L S P のうち最も高い優先度に該当するテーブルに属することとなる。

【 0 0 6 3 】

< リスタート順決定処理 >

図 9 にリスタート順決定処理の一実施形態のフローチャートを示す。リスタート順とはリスタート時にリスタートノードからハローメッセージを送信する隣接ノードの順番である。このリスタート順決定処理では隣接ノードの優先度に従い、自ノードのノード処理能力に合わせてリスタート順を決定して行く。ここで、ノード処理能力とは、リスタートノードがリカバリタイム内にリカバリ処理が可能な L S P 数を表し、ノードの演算速度等によって決定される。自ノードのノード処理能力（L S P 数）についてはハードウェアによって決定される上限値以下であれば、対向装置などの処理能力に合わせて自由に設定してよい。

【 0 0 6 4 】

図 9 において、装置リスタート後、ステップ S 1 1 で C P U 2 1 はノード情報テーブル

10

20

30

40

50

を参照し、高優先度の優先度 0 のテーブルに属する隣接ノードから順々に L S P 数を合計 L S P 数に加算し、ステップ S 1 2 で合計 L S P 数がノード処理能力の L S P 数を超過したか否かを判別する。合計 L S P 数が自ノード処理能力の L S P 数未満であると、ステップ S 1 3 で C P U 2 1 はノード情報テーブルにおける次の隣接ノードの L S P 数を加算してステップ S 1 2 の判別を繰り返す。

【 0 0 6 5 】

ステップ S 1 2 で合計 L S P 数がノード処理能力の L S P 数を超過した場合は加算を中止し、ステップ S 1 4 でノード処理能力の L S P 数を超過する直前の隣接ノードまでの複数の隣接ノードをリスタートグループとし、このリスタートグループの複数の隣接ノードに対しハローメッセージを送信する。そして、このリスタートグループの隣接ノードから

10

【 0 0 6 6 】

次に、ステップ S 1 5 で C P U 2 1 は合計 L S P 数をクリアつまりゼロリセットし、前回の L S P 数の加算において、最後に加算した隣接ノードを C P U 2 1 はノード情報テーブルにおける次の隣接ノードに設定してステップ S 1 3 に進む。そして、ステップ S 1 3 で C P U 2 1 はノード情報テーブルにおける次の隣接ノードの L S P 数を合計 L S P 数に加算してステップ S 1 2 の判別を繰り返す。

【 0 0 6 7 】

例えば図 7 に示す例では、まず図 7 ( A ) の優先度 0 のノード情報テーブルに属する L S P 数  $N_a$  ,  $N_b$  ,  $N_c$  を加算して行く。合計 L S P 数が自ノード処理能力を超過した場合は加算を中止し、処理能力範囲内までの例えば隣接ノード A , B を第 1 リスタートグループとし、第 1 リスタートグループの隣接ノード A , B に対してハローメッセージを送信し、リカバリ処理を実行する。その後、L S P 数をクリアし、前回計算時の最後に加算したノード C から、再び同様の処理を実施する。

20

【 0 0 6 8 】

再計算により求められた第 2 リスタートグループについては、前回ハローメッセージを送信した第 1 リスタートグループについてのリカバリ処理が全て完了した後にハローメッセージを送信し、リカバリ処理を実行して行く。計算において優先度 0 のノード情報テーブルのノードを加算し終えたら、次は優先度 1 のノード情報テーブルに属するノードの L S P 数  $N_e$  ,  $N_f$  , ... を加算して行く。これらの処理を全てのノード情報テーブルに記録された全てのノードに対して実行する。

30

【 0 0 6 9 】

図 1 0 にグレースフルリスタートのリカバリ処理の一実施形態のシーケンス図を示す。図 1 0 において、リスタートノードであるノード # 1 は障害発生から再起動し ( シーケンス S Q 2 3 a ) 、ノード # 1 から上流の隣接ノード # 2 に対しハローメッセージを送信する ( シーケンス S Q 2 4 a ) 。

【 0 0 7 0 】

ノード # 1 はノード # 2 のリカバリ処理完了後 ( シーケンス S Q 2 5 a ~ S Q 2 7 a ) 、ノード # 3 に対してハローメッセージを送信し ( シーケンス S Q 2 4 b ) 、ノード # 2 と同様にリカバリ処理を実施する ( シーケンス S Q 2 5 b ~ S Q 2 7 b ) 。以降も同様にノード # 4 , ノード # 5 とリカバリ処理を繰り返す。

40

【 0 0 7 1 】

図 1 0 ではノード # 2 ~ 5 は個別のノードとして表現しているが、ノード # 2 ~ 5 をリスタートグループとしても良く、ノード # 1 からリスタートグループ # 2 に対してリカバリ処理する。この時、リスタートグループ # 2 に含まれる L S P の数はノード # 1 がリカバリタイム内にリカバリ処理を完了できる範囲の数である。その後、同様にリスタートグループ # 3 , リスタートグループ # 4 と処理を繰り返す。

【 0 0 7 2 】

また、ハローメッセージの送信順に関しては、L S P の優先度に基づいてノード情報テーブルから決定されたリスタート順に従う。ところで予め設定されるリスタートタイム内

50

に復旧できないノード（又はリスタートグループ）# 3, # 4 に対しては、図 10 に示すように、新たにリスタートタイムの延長を行っても良い（シーケンス S Q 2 8 b, S Q 2 9 b）。

【 0 0 7 3 】

本実施形態では、リスタートノードから隣接ノードに対するハローメッセージをリスタートノード自身が処理可能な数の上流の隣接ノードに対してのみ送信する。そのリカバリ処理完了後、同様のプロセスにて処理可能な数の別の上流の隣接ノードに対してハローメッセージを送信し、リカバリ処理を行う。以降、上記の操作を順次繰り返してリカバリ処理を実施することで、リカバリ処理の輻輳を回避することができ、円滑にリスタートすることができる。

10

【 0 0 7 4 】

ここで、M P L S ラベルには 2 0 ビットのフィールドが割り当てられており、ラベル値は 0 ~ 1 0 4 8 5 7 5 の間で付与される。仮に、リカバリ処理輻輳により 1 0 4 8 5 7 5 個全てのラベルについて新ラベルの取得処理が行われたとする。1 ラベルの取得処理に要する時間を 1 m s とすると、この場合の通信断時間は、 $1 0 4 8 5 7 5 \times 1 m s = 1 0 4 8 . 6 s \quad 1 7 m i n$  となるが、本実施形態により、上記の通信断時間を 0 にすることが可能となる。

（付記 1）

ラベル転送ネットワークでリスタートを行うリスタート方法であって、

リスタートを行うリスタートノードは、前記リスタートノードの処理能力に基づいたノード数の隣接ノードを選択してリスタートグループとし、前記リスタートグループの隣接ノードに対してリスタートを通知し、

20

前記リスタートノードは、前記リスタートグループの隣接ノードから通知されたパスのラベル値を基に、パスとラベルを対応付けるラベルテーブル内で前記通知されたパスに対応するエントリを復元し、

前記リスタートの通知と前記通知されたパスに対応するエントリの復元とを繰り返して前記ラベルテーブルの復元を行う

ことを特徴とするリスタート方法。

（付記 2）

付記 1 記載のリスタート方法において、

30

前記リスタートノードに、優先度毎に、隣接ノードと、前記隣接ノードと自ノード間のパス数を登録したノード情報テーブルを設け、

前記ノード情報テーブルを優先度が高い順に検索して前記リスタートを行うノード装置の処理能力に基づいたノード数の隣接ノードを選択する

ことを特徴とするリスタート方法。

（付記 3）

付記 2 記載のリスタート方法において、

前記選択する隣接ノードは、前記ノード情報テーブルを優先度が高い順に検索して、前記隣接ノード毎に前記隣接ノードと自ノード間のパス数を加算し、合計値が前記リスタートを行うノード装置の処理能力に応じた所定値を超過する直前までに検索された隣接ノードである

40

ことを特徴とするリスタート方法。

（付記 4）

付記 3 記載のリスタート方法において、

前記ノード情報テーブルは、隣接ノードから新たなパスが設定されたとき、前記新たなパスの優先度と、前記隣接ノードが既に登録されている前記ノード情報テーブルの優先度とを比較して前記新たなパスの優先度が高い場合には、前記隣接ノードを登録する前記ノード情報テーブルの優先度を前記新たなパスの優先度に変更する

ことを特徴とするリスタート方法。

（付記 5）

50

付記 2 乃至 4 のいずれか 1 項記載のリスタート方法において、  
前記優先度は、前記隣接ノードと自ノードの間で送受信されるメッセージに含まれる  
ことを特徴とするリスタート方法。

(付記 6)

付記 2 乃至 4 のいずれか 1 項記載のリスタート方法において、  
前記優先度は、前記隣接ノードと自ノードの間で送受信されるメッセージに含まれるリ  
フレッシュ間隔が短いパスほど高い値とする  
ことを特徴とするリスタート方法。

(付記 7)

ラベル転送ネットワークでリスタートを行うノード装置であって、  
自ノード装置の処理能力に基づいたノード数の隣接ノードを選択してリスタートグルー  
プとし、前記リスタートグループの隣接ノードに対してリスタートを通知する通知手段と

10

、  
前記リスタートグループの隣接ノードから通知されたパスのラベル値を基に、パスとラ  
ベルを対応付けるラベルテーブル内で前記通知されたパスに対応するエントリを復元する  
復元手段と、を有し、

前記通知手段による通知と前記復元手段による復元とを繰り返して前記ラベルテーブル  
の復元を行う  
ことを特徴とするノード装置。

(付記 8)

20

付記 7 記載のノード装置において、  
優先度毎に、隣接ノードと、前記隣接ノードと自ノード間のパス数を登録したノード情  
報テーブルを有し、

前記通知手段は、前記ノード情報テーブルを優先度が高い順に検索して前記自ノード装  
置の処理能力に基づいたノード数の隣接ノードを選択する  
ことを特徴とするノード装置。

(付記 9)

付記 8 記載のノード装置において、  
前記通知手段は、前記ノード情報テーブルを優先度が高い順に検索して、前記隣接ノ  
ード毎に前記隣接ノードと自ノード間のパス数を加算し、合計値が前記リスタートを行うノ  
ード装置の処理能力に応じた所定値を超過する直前までに検索された隣接ノードを選択す  
る  
ことを特徴とするノード装置。

30

(付記 10)

付記 9 記載のノード装置において、  
隣接ノードから新たなパスが設定されたとき、前記新たなパスの優先度と、前記隣接ノ  
ードが既に登録されている前記ノード情報テーブルの優先度とを比較して前記新たなパス  
の優先度が高い場合には、前記隣接ノードを登録する前記ノード情報テーブルの優先度を  
前記新たなパスの優先度に変更する変更手段を  
有することを特徴とするノード装置。

40

(付記 11)

付記 8 乃至 10 のいずれか 1 項記載のノード装置において、  
前記優先度は、前記隣接ノードと自ノードの間で送受信されるメッセージに含まれる  
ことを特徴とするノード装置。

(付記 12)

付記 8 乃至 10 のいずれか 1 項記載のノード装置において、  
前記優先度は、前記隣接ノードと自ノードの間で送受信されるメッセージに含まれるリ  
フレッシュ間隔が短いパスほど高い値とする  
ことを特徴とするノード装置。

【符号の説明】

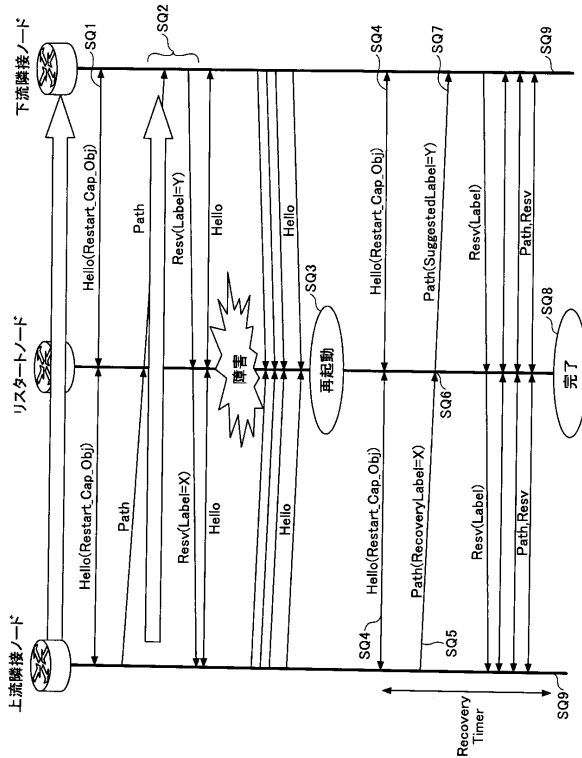
50

【 0 0 7 5 】

- 1 1 - 1 ~ 1 1 - N 入出力ユニット
- 1 2 スイッチ部
- 1 3 制御部
- 1 5 データ受信部
- 1 6 テーブルメモリ
- 1 7 ルート&フォワード制御部
- 1 8 データ送信部
- 2 1 C P U
- 2 2 メモリ

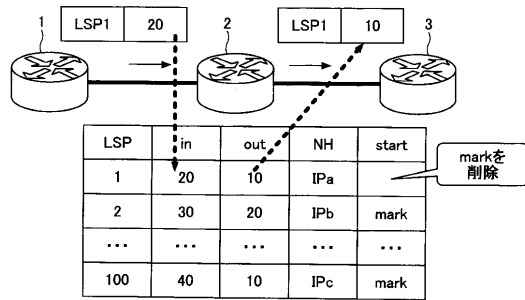
【 図 1 】

グレースフルリスタートの一例のシーケンス図



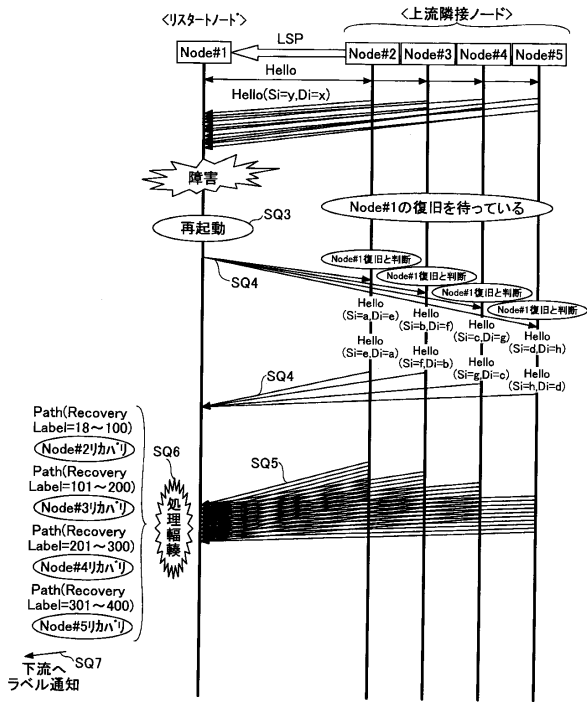
【 図 2 】

ラベルテーブルの一例の構成図



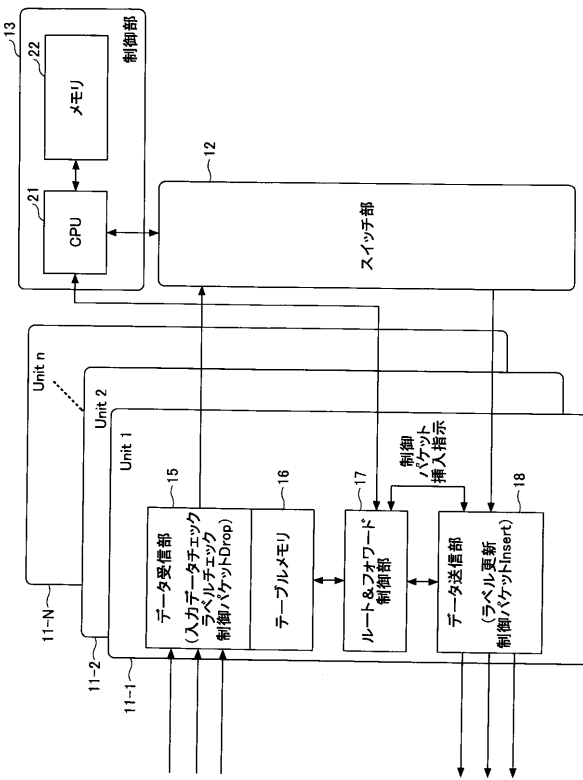
【図3】

グレースフルリスタートのリカバリ処理の輻輳例のシーケンス図



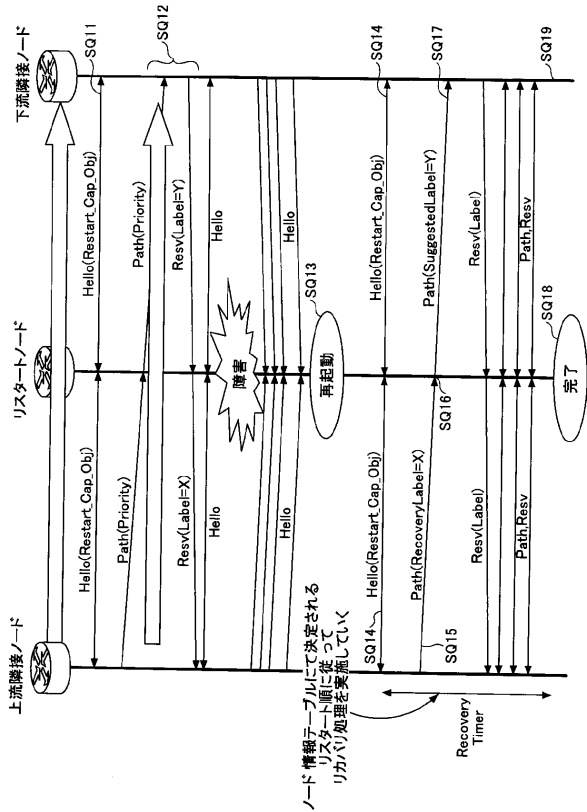
【図4】

ノード装置の一実施形態の構成図



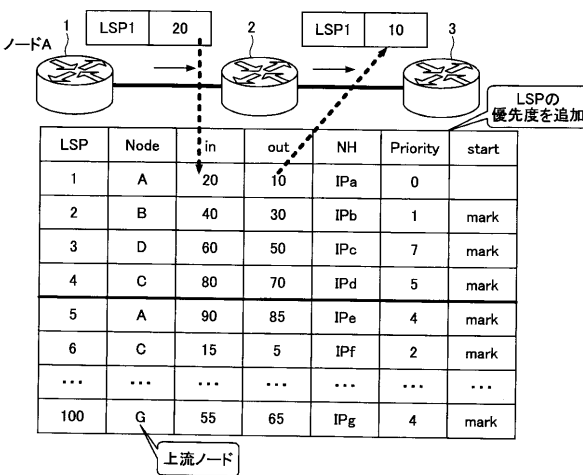
【図5】

グレースフルリスタートの一実施形態のシーケンス図



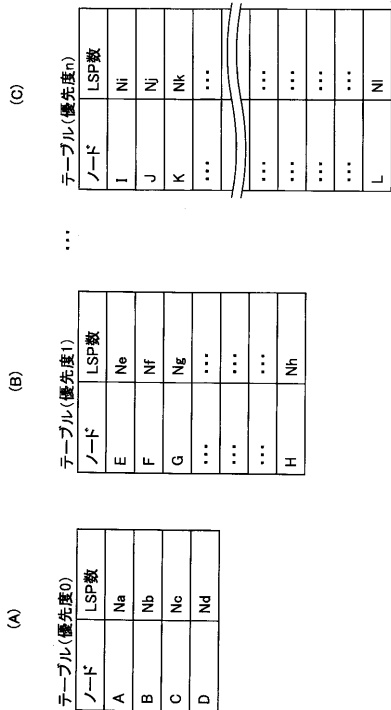
【図6】

ラベルテーブルの一実施形態の構成図



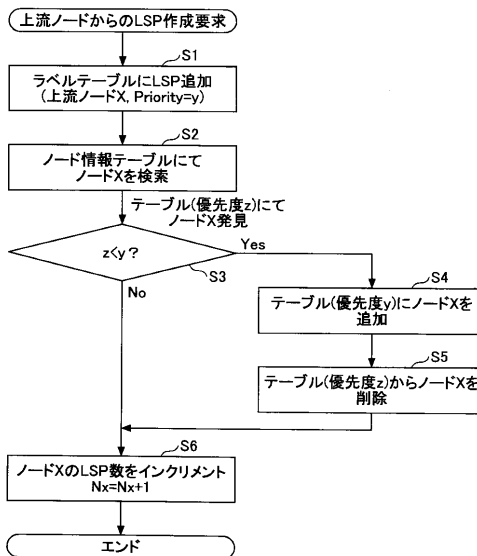
【図7】

ノード情報テーブルの一実施形態の構成図



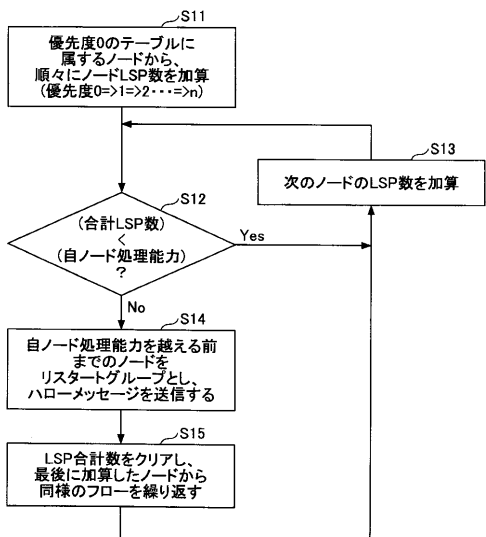
【図8】

テーブル更新処理の一実施形態のフローチャート



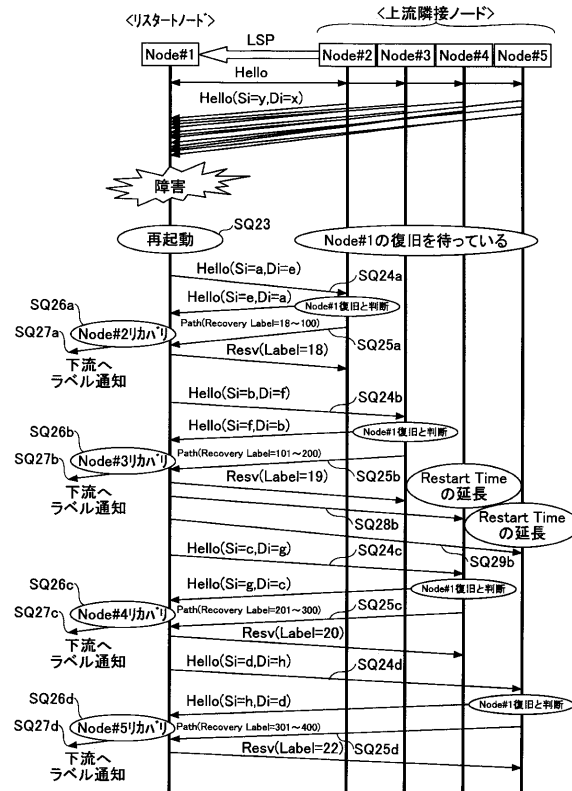
【図9】

リスタート順決定処理の一実施形態のフローチャート



【図10】

グレースフルリスタートのリカバリ処理の一実施形態のシーケンス図



【 図 1 1 】

Session-Attributeオブジェクトのフォーマットを示す図

Setup Priority	Holding Priority	Flags	Name Length
Session Name			



---

フロントページの続き

(72)発明者 山本 洋照

福岡県福岡市早良区百道浜2丁目2番1号 富士通九州ネットワークテクノロジーズ株式会社内

審査官 宮島 郁美

(56)参考文献 特開2009-303087(JP,A)

特開2009-124566(JP,A)

特開2008-042807(JP,A)

米国特許出願公開第2008/0037426(US,A1)

米国特許出願公開第2010/0260199(US,A1)

国際公開第2011/045733(WO,A1)

(58)調査した分野(Int.Cl.,DB名)

H04L12/00-12/955,13/00