

(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2006-107506
(P2006-107506A)

(43) 公開日 平成18年4月20日(2006.4.20)

(51) Int. Cl. F I テーマコード(参考)
G06F 12/00 (2006.01) G06F 12/00 545A 5B082

審査請求 未請求 請求項の数 20 O L (全 34 頁)

(21) 出願番号 特願2005-290093 (P2005-290093)
(22) 出願日 平成17年10月3日(2005.10.3)
(31) 優先権主張番号 60/615,411
(32) 優先日 平成16年10月1日(2004.10.1)
(33) 優先権主張国 米国 (US)
(31) 優先権主張番号 11/053,386
(32) 優先日 平成17年2月7日(2005.2.7)
(33) 優先権主張国 米国 (US)

(71) 出願人 500046438
マイクロソフト コーポレーション
アメリカ合衆国 ワシントン州 9805
2-6399 レッドモンド ワン マイ
クロソフト ウェイ
(74) 代理人 100077481
弁理士 谷 義一
(74) 代理人 100088915
弁理士 阿部 和夫
(72) 発明者 ダニエル イー. ロビンガー
アメリカ合衆国 98052 ワシントン
州 レッドモンド ワン マイクロソフト
ウェイ マイクロソフト コーポレーシ
ョン内

最終頁に続く

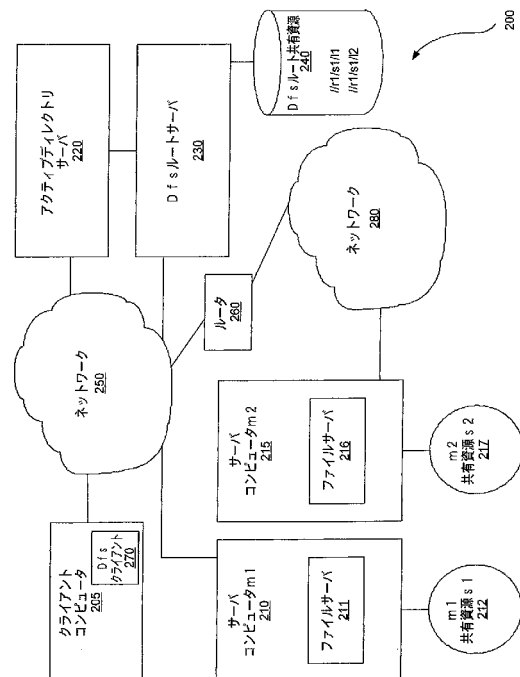
(54) 【発明の名称】 分散ファイルシステムのためのターゲットフェイルバックおよびターゲット優先度を決定するシステムおよび方法

(57) 【要約】

【課題】分散ファイルシステムのためのターゲットフェイルバックおよびターゲット優先度を決定するシステムおよび方法を提供する。

【解決手段】コンピュータシステムは、サイトコストおよび/またはターゲット優先度に従ってソートされたソート済みターゲットリストからターゲットを選択することができる。次いで、コンピュータシステムは、設定ターゲットが、ソート済みリスト中のすべての利用可能なターゲットに比べて、より好ましいターゲットに関連付けられているかどうか判定し、そうでない場合、より好ましいターゲットにスイッチバックすることができる。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

コンピュータシステムにおいて、コンピュータのネットワーク中のファイルの場所を探す方法であって、

クライアントコンピュータから、サーバコンピュータのネットワーク中の少なくとも 1 つのサーバ上に位置するファイルの場所を要求することと、

前記要求されたファイルが位置している前記サーバコンピュータのネットワーク中の複数の場所を決定することと、

前記クライアントコンピュータに、それぞれが前記サーバコンピュータのネットワーク中の要求されたファイルの場所に対応する複数の参照を含むターゲットのリストを返すことと、

各個別ファイルの場所と前記クライアントコンピュータの間で前記要求されたファイルを検索するために、ターゲット情報に基づいて前記ターゲットのリストを有界集合にソートすることと、

前記有界集合の指示を前記クライアントコンピュータに提供することと
を含むことを特徴とする方法。

【請求項 2】

ターゲット情報に基づいて前記ターゲットのリストを有界集合にソートすることは、各有界集合内の各ターゲットが前記個別有界集合中の他のターゲットと同じサイトコストに関連付けられるように前記ターゲットを有界集合に分けることを含めてサイトコストの評価に基づいて前記リストをソートすることをさらに含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記有界集合の指示を前記クライアントコンピュータに提供することは、ターゲットと前記クライアントコンピュータの間のコストを含む境界属性を提供することを含むことを特徴とする請求項 1 に記載の方法。

【請求項 4】

各有界集合の境界は、少なくとも一部は、各ターゲットに関連付けられた前記サーバコンピュータの前記場所に基づくものであることを特徴とする請求項 2 に記載の方法。

【請求項 5】

各有界集合の境界は、少なくとも一部は、各ターゲットに関連付けられた前記サーバコンピュータに関連付けられた健全性の指示に基づくものであることを特徴とする請求項 2 に記載の方法。

【請求項 6】

前記ソートすることによって確定された順序で各ターゲットに繰り返しアクセスしようと試みることと、

試みに成功すると、正常にアクセスされた第 1 のターゲットとの接続を、後続のアクセス要求に使用される設定ターゲットとして確立することと

をさらに含むことを特徴とする請求項 1 に記載の方法。

【請求項 7】

後続のアクセス要求を提供するのにより低いサイトコストに関連付けられたターゲットが利用可能であるかどうか判定することと、

より低いサイトコストのターゲットが利用可能である場合、前記設定ターゲットを前記より低いサイトコストのターゲットに変更することと

をさらに含むことを特徴とする請求項 6 に記載の方法。

【請求項 8】

前記より低いサイトコストに関連付けられたターゲットが利用可能であるかどうか判定することは、前記ターゲットに付随するターゲットフェイルバックポリシーが適用されるたびにより低いサイトコストに関連付けられたターゲットが利用可能であるかどうか判定することを含むことを特徴とする請求項 7 に記載の方法。

10

20

30

40

50

【請求項 9】

前記ターゲットフェイルバックポリシーは前記ターゲットのリストに関連付けられることを特徴とする請求項 8 に記載の方法。

【請求項 10】

前記ターゲットフェイルバックポリシーは、前記ターゲットのリスト中の各個別ターゲットに関連付けられ得るルートに関連付けられることを特徴とする請求項 8 に記載の方法。

【請求項 11】

除外ポリシーに基づいて前記ターゲットのリストから少なくとも 1 つのターゲットを取り除くことをさらに含むことを特徴とする請求項 1 に記載の方法。

10

【請求項 12】

ターゲット情報に基づいて前記ターゲットのリストを有界集合にソートすることは、各ターゲットに関連付けられた優先度設定に従って前記ターゲットをソートすることをさらに含むことを特徴とする請求項 1 に記載の方法。

【請求項 13】

前記ソートすることは、各有界集合内の各ターゲットが、前記個別有界集合中の他のターゲットと同じサイトコストに関連付けられるように前記ターゲットのリストを有界集合に分けることをさらに含み、前記優先度設定に従ってソートすることは、関連するターゲット優先度設定に従って各有界集合内の各ターゲットを順序付けすることをさらに含むことを特徴とする請求項 12 に記載の方法。

20

【請求項 14】

前記各ターゲットに関連付けられた優先度設定に従って前記ターゲットをソートすることは、大域的に高いターゲット優先度設定を持つターゲットを前記順序付けされたリスト中の前記個別有界集合の先頭に配列し、大域的に低いターゲット優先度設定を持つターゲットを前記順序付けされたリスト中の前記個別有界集合の末尾に配列することをさらに含むことを特徴とする請求項 12 に記載の方法。

【請求項 15】

ネットワークに動作可能に結合されたサーバコンピュータであって、前記アクティブディレクトリサーバコンピュータは、ターゲットを求める要求に応答して 1 つまたは複数のクライアントコンピュータにターゲットのリストを提供するように動作可能であり、各ターゲットは前記ネットワーク中のコンピュータ上に位置するファイルに対応し、前記ファイルは前記 1 つまたは複数のクライアントコンピュータによって要求され、前記アクティブディレクトリサーバコンピュータは、各ターゲットに関連付けられたサイトコストの評価に従って前記ターゲットのリストをソートするようにも動作可能であり、前記サイトコストは、前記ターゲットに対応するサーバコンピュータから前記ターゲットを要求する前記クライアントコンピュータにファイルを検索することに伴うコストを示すものであるサーバコンピュータと、

30

前記ネットワークに結合された他のコンピュータからのファイルを要求し、検索するように動作可能な、前記ネットワークに結合された少なくとも 1 つのクライアントコンピュータと、

40

前記アクティブディレクトリサーバコンピュータによって提供される前記ターゲットの少なくとも 1 つに対応する、要求されたファイルを持つ少なくとも 1 つのサーバコンピュータと

を備えることを特徴とする分散コンピューティングシステム。

【請求項 16】

前記アクティブディレクトリサーバコンピュータは、さらに、個別有界集合中の各ターゲットが前記個別有界集合中の他のターゲットと同じサイトコストに関連付けられると共に、有界集合が、前記有界集合中の各ターゲットに関連付けられた前記サイトコストに従ってソートされるように前記ターゲットのリストを有界集合にソートするように動作可能であることを特徴とする請求項 15 に記載の分散コンピュータシステム。

50

【請求項 17】

前記サーバコンピュータは、さらに、各ターゲットに関連付けられたターゲット優先度設定に従って各有界集合内の各ターゲットをソートするように動作可能であることを特徴とする請求項 16 に記載の分散コンピュータシステム。

【請求項 18】

前記クライアントコンピュータは、さらに、第 1 の利用可能なターゲットを設定ターゲットに指定し、その後、前記設定ターゲットが、前記ターゲットのリスト中の他のターゲットに比べて、最低のサイトコストに関連付けられているかどうか判定し、そうでない場合、より低いサイトコストを持つ利用可能なターゲットに切り換え、前記より低いサイトコストを持つ前記ターゲットを前記設定ターゲットに指定するように動作可能であることを特徴とする請求項 17 に記載の分散コンピュータシステム。 10

【請求項 19】

前記クライアントコンピュータは、さらに、第 1 の利用可能なターゲットを設定ターゲットに指定し、その後、前記設定ターゲットが、前記ターゲットのリスト中の他のターゲットに比べて、より好ましい優先度に関連付けられているかどうか判定し、そうでない場合、より好ましい優先度を持つ利用可能なターゲットに切り換え、前記より好ましい優先度を持つ前記ターゲットを前記設定ターゲットに指定するように動作可能であることを特徴とする請求項 17 に記載の分散コンピュータシステム。

【請求項 20】

データ構造が格納されているコンピュータ可読媒体であって、前記データ構造は、
それぞれがサーバコンピュータに関連付けられたターゲットのグループに属するターゲットを格納する第 1 のフィールドと、
クライアントコンピュータと前記サーバコンピュータの間の関係を指示する情報を格納する第 2 のフィールドと、
前記ターゲットのグループの境界を示す値を格納する第 3 のフィールドと
を備えることを特徴とするコンピュータ可読媒体。 20

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、一般に、コンピュータシステムに関し、より詳細には、分散ファイルシステムのためのターゲットフェイルバックおよびターゲット優先度を決定する改善されたシステムおよび方法に関する。 30

【背景技術】**【0002】**

分散ファイルシステム (Dfs) とは、ネットワーク上のデータを位置指定し、それを管理するネットワークサーバコンポーネントである。Dfs を使って、異なるコンピュータ上のファイルを単一の名前空間に一元化することができ、ゆえに、ユーザが、ネットワーク上の複数のファイルサーバおよびファイルサーバ共有資源 (share) の単一の、階層的ビューの構築が可能になる。1 台のサーバコンピュータ、または 1 組のサーバコンピュータの状況において、Dfs は、パーソナルコンピュータシステムにおけるハードディスクでのファイルシステムにたとえることができる。例えば、ディスク上のセクタの集合体への一様な名前によるアクセスを提供するファイルシステムの役割と同様に、Dfs は、サーバ、共有資源、およびファイルの集合体での一様な命名規則およびマッピングを提供することができる。ゆえに、Dfs は、ファイルサーバおよびそれらの共有資源を、大企業とその情報リソースをより効率よく管理し、使用することを可能にする 1 つの論理的な階層に編成することができる。 40

【0003】

さらに、Dfs は、単一のファイルプロトコルに制限されず、使用されているファイルクライアントに関係なく、そのクライアントがネイティブサーバおよび共有資源をサポートすることを条件として、サーバ、共有資源、およびファイルのマッピングをサポートす 50

ることができる。Dfsは、異種のサーバボリュームおよび共有資源への名前透過性も提供することができる。Dfsを介して、管理者は、そのコンテンツが組織の広域ネットワーク(WAN)全体に分散されている単一の階層的ファイルシステムを構築することができる。

【0004】

以前においては、国際命名規則(UNC)を用いて、ユーザまたはアプリケーションがファイル情報にアクセスするために物理サーバおよび共有資源を指定することが必要とされた。例えば、ユーザまたはアプリケーションは、\\Server\share\Path\Filenameを指定しなければならなかった。たとえUNCが直接使用され得る場合であっても、UNCは、通常、さらにそれが\\Server\shareにマッ 10
プされ得る、x:などのドライブ名にマップされる。その点から、ユーザは、ユーザがアクセスしようとするデータにマップする指定変更されたドライブを越えてナビゲートすることが必要とされた。例えば、ユーザが特定のファイルにナビゲートするためには、copy x:\Path\More_path\...\Filenameなどが必要とされた。

【0005】

ネットワークのサイズが拡大し、企業が、イントラネットなどのために、社内でも社外でも既存の記憶を使用し始めると、個々の共有資源への単一のドライブ名のマッピングは、十分に対応できなくなる。さらに、ユーザは、UNC名を直接使用することはできるが、これらのユーザは、データが格納され得る場所の数の多さに圧倒されかねない。 20

【0006】

Dfsは、サーバおよび共有資源の、より簡単で、容易にナビゲート可能な名前空間へのリンクを可能にすることによってこれらの問題を解決する。Dfsボリュームは、共有資源を他の共有資源に階層的に接続させる。Dfsは物理記憶を論理表現にマップするため、その純利益(net benefit)は、任意の数のファイルの物理位置がユーザおよびアプリケーションにとって透過的になることである。

【0007】

さらに、ネットワークサイズが世界規模のネットワークのレベルにまで拡大すると、ネットワークからファイルを検索するのに関連する(ネットワーク時間、ネットワーク負荷などの点での)コストを低減するのに役立つように、同じ1つまたは複数のファイルのいくつ 30
つかのコピーを、ネットワーク内のいくつかの異なる場所に位置させることができる。例えば、第1のサーバ位置の近くに位置する大規模ネットワークのユーザは、通常、それらのユーザの最も近くにあるサーバ上のファイルのコピーを使用する(すなわち、シアトルのユーザは、シアトルの近くに位置するレッドモンドという名前のサーバに最も近いと考えられる)。同様に、第2のサーバ位置の近くに位置する大規模ネットワークのユーザは、通常、それらのユーザの最も近くにある異なるサーバ上のファイルのコピーを使用する(すなわち、タイのユーザは、バンコクに位置するバンコクという名前のサーバに最も近いと考えられる)。ゆえに、ファイルを検索するサイトコスト(すなわち、クライアントとサーバの間の距離、サーバ分離の度合い、その他の物理的ネットワークパラメータを含むいくつかのネットワークパラメータの擬似任意(pseudo-arbitrary)の指示であるスカ 40
ラ数)は、要求される1つまたは複数のファイルを持つ最も近くのサーバにアクセスすることによって最小限にすることができる。

【0008】

ユーザがDfsからファイルを検索しようとするとき、ユーザがそこからファイルを要求しているクライアントコンピュータは、要求されたファイルをどのように検索するか決定する。クライアントコンピュータは、要求された1つまたは複数のファイルの1つまたは複数の場所を獲得することを求める参照要求を出すことができる。参照は、要求側クライアントコンピュータと、要求された1つまたは複数のファイルが見つかり得るサーバコンピュータの間の相対パスとすることができる。クライアントコンピュータは、ローカルでは利用不能であると知られている1つまたは複数のファイルを要求することができ、い 50

くつの異なる場所が要求されたファイルのコピーを提供し得るか判定することができる。通常、要求されたファイルを提供し得る場所を指示する何百または何千ものターゲット（すなわち、そのファイルへの相対パス）があり得る。したがって、参照要求に回答してクライアントコンピュータに返される参照応答は、通常、要求されたファイルを持つサーバおよび/または共有資源に対応するターゲットのリストを含む。

【0009】

しかしながら、以前においては、クライアントコンピュータに返される参照応答は、ランダムな順序でリストされた識別ターゲットを含み、または、場合によっては、サイトコスト別にリストされた識別ターゲットも含むことができた。参照応答中の各ターゲットは、必ずしも、そのターゲットの直前または直後のターゲットとの関係を持つとは限らなかった。その結果、クライアントコンピュータは、ただ単に、ランダムに配列されたターゲットリストの先頭から、ターゲットが接続して応答するまで、リスト上の連続する各ターゲットと接続を確立しようとする試みを開始し得るだけであった。

10

【0010】

しかしながら、このランダム性に関連する問題は、第1の利用可能なターゲットが、実際には、文字通り世界の反対側に位置し得るということである。ゆえに、この第1の利用可能なターゲットと通信を行うサイトコストは、長期的には、相当に高く付き、望ましくないこともある。

【0011】

しかしながら、ターゲットへの接続の連続性を保持することは、ある程度重要である。これは、「粘着性（stickinessまたはstickiness）」と呼ばれている。ゆえに、クライアントコンピュータのファイル要求を満たすことのできる第1の利用可能なターゲットが見つかり、通常は、その後のすべての参照および要求も、クライアントコンピュータのユーザが新しい参照を具体的に要求しない限り、そのターゲットに経路指定される。したがって、第1の利用可能なターゲットへの潜在的に高いサイトコストの接続が無期限に存続し、ますます多くのネットワークトラフィックおよび全体的ネットワークコストを生じることになり得る。

20

【発明の開示】

【発明が解決しようとする課題】

【0012】

連続性を保持するためにクライアントコンピュータとサーバコンピュータの間の非効率な参照を維持する問題は、結果として高サイトコストの通信セッションを生じ得る。求められているのは、参照接続の連続性を保持すると同時に、参照接続のサイトコストを低減する方法である。

30

【課題を解決するための手段】

【0013】

簡単にいうと、本発明は、参照応答として受け取られるターゲットを編成、格納し、分散ファイルシステムにおけるターゲットフェイルバックおよびフェイルバックのターゲット優先度を決定するシステムおよび方法を提供する。一実施形態では、ソート方法は、クライアントコンピュータから、コンピュータネットワーク中の1つまたは複数のコンピュータ上に位置するファイル、ディレクトリ、共有資源などの複数の場所（すなわちターゲット）を要求することを含み得る。次いで、DFSサーバなどのコンピュータは、クライアントコンピュータに、それぞれがコンピュータネットワーク中の要求されたファイルまたはディレクトリの場所に対応する複数の参照を含むターゲットリストを返すことができる。さらに、このターゲットリストは、各個別ターゲットに関連付けられたサイトコストの評価に基づいてソートされ得る。

40

【0014】

サーバコンピュータによって識別されたより低コストのターゲットが参照応答の先頭にソートされ得る。クライアントコンピュータは、より高コストのターゲットとの接続を確立しようとする前に、単に、より低コストのターゲットから始まる参照応答を論理的に

50

構文解析するだけ試みられた接続に至ることができる。

【0015】

ターゲットの優先順位を使用することを含む、そのようなソートシステムも実施され得る。より高い優先度のターゲットも参照応答の先頭にソートされ得る。さらに、参照応答は、サイトコストとターゲット優先度両方での措置(provision)を含むようにもソートされ得る。ゆえに、同等の関連サイトコストを持つターゲットのグループが、さらに、そのグループ内で、各ターゲットの個別関連ターゲット優先度に従ってソートされ得る。

【0016】

本発明の別の実施形態では、参照応答として提供されるソート済みターゲットのリストを使用し得るターゲットフェイルバックおよびターゲット優先度ポリシーが実施され得る。したがって、コンピュータシステムは、クライアントコンピュータにおいて少なくとも1つの要求されたファイルまたはディレクトリを検索するために、サイトコストに従ってソートされたターゲットのリストからターゲットを選択し、それを設定ターゲットに指定することができる。次いで、コンピュータシステムは、設定ターゲットが、ソート済みリスト中のすべての利用可能なターゲットに比べて最低のサイトコストに関連付けられたものであるかどうか判定することができる。そうでない場合、システムは、その設定ターゲットより低いサイトコストに関連付けられた別のターゲットにフェイルバックし、その新しいターゲットを設定ターゲットに指定することができる。

10

【0017】

他の利点は、以下の詳細な説明を図面と併せて読めば明らかになるであろう。

20

【発明を実施するための最良の形態】

【0018】

(例示的動作環境)

図1に、本発明が実施され得る適当なコンピューティングシステム環境100の一例を示す。コンピューティングシステム環境100は、適当なコンピューティング環境の一例にすぎず、本発明の用途または機能の範囲についてのどんな限定を示唆するためのものでもない。また、コンピューティング環境100は、例示的動作環境100に示す構成要素のいずれか1つまたはそれらの組合せに関連するどんな依存関係または要件を有するものであるとも解釈すべきではない。

【0019】

本発明は、他の多数の汎用または専用コンピューティングシステム環境または構成を用いて動作する。本発明と共に使用するのに適し得るよく知られているコンピューティングシステム、環境、および/または構成の例には、それだけに限らないが、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドまたはラップトップ機器、タブレット機器、ヘッドレスサーバ、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能家庭用電化製品、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、上記のシステムまたは機器のいずれかを含む分散コンピューティング環境などが含まれる。

30

【0020】

本発明は、コンピュータにより実行される、プログラムモジュールなどのコンピュータ実行可能命令の一般的コンテキストで説明され得る。一般に、プログラムモジュールには、個々のタスクを実行し、または個々の抽象データ型を実施するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などが含まれる。本発明は、タスクが、通信ネットワークを介してリンクされたりリモート処理装置により実行される分散コンピューティング環境でも実施され得る。分散コンピューティング環境では、プログラムモジュールは、メモリ記憶装置を含むローカルおよび/またはリモートのコンピュータ記憶媒体に位置し得る。

40

【0021】

図1を参照すると、本発明を実施する例示的システムは、コンピュータ110の形で汎用コンピューティングデバイスを含む。コンピュータ110の構成要素には、それだけに

50

限らないが、処理装置 120、システムメモリ 130、およびシステムメモリを含む様々なシステム構成要素を処理装置 120 に結合するシステムバス 121 が含まれ得る。システムバス 121 は、様々なバスアーキテクチャのいずれかを使用した、メモリバスまたはメモリコントローラ、周辺バス、およびローカルバスを含む数種類のバス構造のいずれでもよい。例をあげると、それだけに限らないが、そのようなアーキテクチャには、産業標準アーキテクチャ (ISA) バス、マイクロチャネルアーキテクチャ (MCA) バス、拡張 ISA (EISA) バス、ビデオ電子装置規格化協会 (VESA) ローカルバス、およびメザンバスとも呼ばれる周辺装置相互接続 (PCI) バスが含まれる。

【0022】

コンピュータ 110 は、通常、様々なコンピュータ可読媒体を含む。コンピュータ可読媒体は、コンピュータ 110 によってアクセスされ得る任意の使用可能な媒体とすることができ、それには揮発性媒体と不揮発性媒体の両方、着脱式媒体と固定式媒体の両方が含まれる。例をあげると、それだけに限らないが、コンピュータ可読媒体には、コンピュータ記憶媒体および通信媒体が含まれ得る。コンピュータ記憶媒体には、コンピュータ可読命令、データ構造、プログラムモジュールまたはその他のデータなどの情報を記憶するための任意の方法または技術で実施された、揮発性および不揮発性、着脱式および固定式の媒体が含まれる。コンピュータ記憶媒体には、それだけに限らないが、RAM、ROM、EEPROM、フラッシュメモリまたは他のメモリ技術、CD-ROM、デジタル多用途ディスク (DVD) または他の光ディスク記憶、磁気カセット、磁気テープ、磁気ディスク記憶または他の磁気記憶装置、あるいは所望の情報を格納するのに使用でき、コンピュータ 110 によってアクセスされ得る他の任意の媒体が含まれる。通信媒体は、通常、コンピュータ可読命令、データ構造、プログラムモジュールまたはその他のデータを、搬送波や他の搬送機構などの変調データ信号として実施し、任意の情報配信媒体を含む。「変調データ信号」という用語は、その特性の 1 つまたは複数、その信号に情報を符号化するような方式で設定または変更されている信号を意味する。例をあげると、それだけに限らないが、通信媒体には、有線ネットワークや直接配線接続などの有線媒体、および音響、RF、赤外線、その他の無線媒体などの無線媒体が含まれる。上記のいずれかの組合せも、コンピュータ可読媒体の範囲内に含めるべきである。

【0023】

システムメモリ 130 は、読取り専用メモリ (ROM) 131 やランダムアクセスメモリ (RAM) 132 などの揮発性および / または不揮発性メモリの形でコンピュータ記憶媒体を含む。基本入出力システム (BIOS) 133 は、始動時などに、コンピュータ 110 内の諸要素間での情報転送を支援する基本ルーチンを含み、通常、ROM 131 に格納される。RAM 132 は、通常、処理装置 120 へ直ちにアクセス可能であり、かつ / または処理装置 120 によって現在操作されているデータおよび / またはプログラムモジュールを含む。例として、それだけに限らないが、図 1 に、オペレーティングシステム 134、アプリケーションプログラム 135、その他のプログラムモジュール 136、およびプログラムデータ 137 を示す。

【0024】

コンピュータ 110 は、他の着脱式 / 固定式、揮発性 / 不揮発性コンピュータ記憶媒体も含み得る。例にすぎないが、図 1 に、固定式、不揮発性磁気媒体との間で読取りまたは書込みを行うハードディスクドライブ 141、着脱式、不揮発性磁気ディスク 152 との間で読取りまたは書込みを行う磁気ディスクドライブ 151、および CD-ROM や他の光媒体などの着脱式、不揮発性光ディスク 156 との間で読取りまたは書込みを行う光ディスクドライブ 155 を示す。例示的動作環境で使用され得る他の着脱式 / 固定式、揮発性 / 不揮発性のコンピュータ記憶媒体には、それだけに限らないが、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、ソリッドステート RAM、ソリッドステート ROM などが含まれる。ハードディスクドライブ 141 は、通常、インターフェース 140 などの固定式メモリインターフェースを介してシステムバス 121 に接続され、磁気ディスクドライブ 151 および光ディスクドライブ 15

5 は、通常、インターフェース 150 などの着脱式メモリアンターフェースによってシステムバス 121 に接続される。

【0025】

前述の、図 1 に示す各ドライブおよびそれに関連するコンピュータ記憶媒体は、コンピュータ 110 のためのコンピュータ可読命令、データ構造、プログラムモジュールおよびその他のデータの記憶を提供する。図 1 では、例えば、ハードディスクドライブ 141 は、オペレーティングシステム 144、アプリケーションプログラム 145、その他のプログラムモジュール 146、およびプログラムデータ 147 を格納するものとして示されている。これらのコンポーネントは、オペレーティングシステム 134、アプリケーションプログラム 135、その他のプログラムモジュール 136、およびプログラムデータ 137 と同じでも、異なってもよいことに留意されたい。オペレーティングシステム 144、アプリケーションプログラム 145、その他のプログラムモジュール 146、およびプログラムデータ 147 には、少なくともそれらが異なるコピーであることを示すために、図では異なる番号を付与してある。ユーザは、タブレットまたは電子ディジタイザ 164、マイクロホン 163、キーボード 162、あるいは一般にマウス、トラックボール、タッチパッドなどと呼ばれるポインティングデバイス 161 などの入力装置を介してコンピュータ 110 にコマンドおよび情報を入力することができる。図 1 に示さない他の入力装置には、ジョイスティック、ゲームパッド、衛星パラボラアンテナ、スキャナ、あるいは、バイオメトリックセンサ、環境センサ、位置センサまたはその他の種類のセンサを備える装置を含むその他の装置が含まれ得る。上記その他の入力装置は、しばしば、システムバスに結合されたユーザ入力インターフェース 160 を介して処理装置 120 に接続されるが、パラレルポート、ゲームポート、ユニバーサルシリアルバス (USB) といった他のインターフェースおよびバス構造によっても接続され得る。モニタ 191 または他の種類の表示装置も、ビデオインターフェース 190 などのインターフェースを介してシステムバス 121 に接続される。モニタ 191 は、タッチスクリーンパネルなどと一体化されることもある。モニタおよび/またはタッチスクリーンパネルは、タブレット型パーソナルコンピュータなどのように、コンピューティングデバイス 110 が組み込まれている筐体に物理的に結合され得ることに留意されたい。また、コンピューティングデバイス 110 などのコンピュータは、スピーカ 194 やプリンタ 195 など他の周辺出力装置も含むことができ、それらは、出力周辺インターフェース 193 などを通じて接続され得る。

【0026】

コンピュータ 110 は、リモートコンピュータ 180 など、1 つまたは複数のリモートコンピュータへの論理接続を使用したネットワークで接続された環境で動作し得る。リモートコンピュータ 180 は、パーソナルコンピュータ、サーバ、ルータ、ネットワーク PC、ピアデバイスまたは他の一般的なネットワークノードとすることができ、通常は、コンピュータ 110 に関連して前述した要素の多くまたはすべてを含むが、図 1 にはメモリ記憶装置 181 だけを示してある。図 1 に示す論理接続には、ローカルエリアネットワーク (LAN) 171 および広域ネットワーク (WAN) 173 が含まれるが、他のネットワークも含まれ得る。そのようなネットワーク環境は、オフィス、企業規模のコンピュータネットワーク、イントラネットおよびインターネットではよく見られるものである。LAN ネットワーク環境で使用されるとき、コンピュータ 110 はネットワークインターフェースまたはアダプタ 170 を介して LAN 171 に接続される。WAN ネットワーク環境で使用されるとき、コンピュータ 110 は、通常、モデム 172 またはインターネットなどの WAN 173 を介して通信を確立する他の手段を含む。モデム 172 は、内蔵でも外付けでもよく、ユーザ入力インターフェース 160 または他の適当な機構を介してシステムバス 121 に接続され得る。ネットワークで接続された環境では、コンピュータ 110 に関連して示すプログラムモジュール、またはその一部は、リモートのメモリ記憶装置にも格納され得る。例として、それだけに限らないが、図 1 に、リモートアプリケーションプログラム 185 をメモリ装置 181 上にあるものとして示す。図示のネットワーク接続は例示的であり、コンピュータ間での通信リンクを確立する他の手段も使用され得るこ

とが理解されるであろう。

【0027】

(参照応答のソート)

本発明は、一般に、分散ファイルシステムのためのターゲットフェイルバックおよびターゲット優先度を決定するシステムおよび方法を対象とする。このシステムおよび方法は、コンピュータネットワーク中の所望のファイルまたはディレクトリを含み得るコンピュータのリストを要求するクライアントコンピュータのためのソート方法を提供することができる。要求に 응답して、クライアントコンピュータに提供され得るターゲットのリストが、各個別ターゲットに関連付けられたサイトコストを含む、特定のパラメータの評価に基づいてソートされ得る。明らかのように、クライアントコンピュータが、より高コストのターゲットとの接続を確立しようと試みる前に、単に、より低コストのターゲットから始まる参照応答を論理的に構文解析するだけで接続を試みることができるように、識別されるより低コストのターゲットが参照応答の先頭にソートされ得る。

10

【0028】

そのようなソートシステムは、ターゲットの優先順位を使っても実施することができ、一実施形態では、より高い優先度のターゲットが、やはり、参照応答の先頭にソートされ得る。さらに、参照応答は、サイトコストとターゲット優先度両方での措置を含むようにもソートされ得る。ゆえに、同等の関連サイトコストを持つターゲットのグループは、各ターゲットの個別の関連ターゲット優先度に従って、そのグループ内でさらにソートされ得る。明らかのように、本明細書で説明する様々なブロック図、流れ図およびシナリオは例にすぎず、本発明が適用される他の多くのシナリオがある。

20

【0029】

図面の図2を見ると、本発明の一態様による、ターゲットフェイルバックおよびターゲット優先度を実施するための分散コンピューティング環境の例示的アーキテクチャを一般的に表すブロック図が示されている。通常、クライアントコンピュータ205は、図1のパーソナルコンピュータ110と類似のものとすることができ、企業イントラネットの一部またはインターネットの一部とすることのできる第1のネットワーク250に動作可能に結合され得る。クライアントコンピュータ205は、Dfsの一部とすることもでき、Dfs全体のどこかに位置するファイルへのシームレスなアクセスを提供し得るDfsクライアント270を含むこともできる。また、ネットワークも、コンピュータm1 210や、アクティブディレクトリサーバ220や、Dfsルートサーバ230など、他のコンピュータに動作可能に結合され得る。アクティブディレクトリサーバ220を使って、Dfsリンク参照情報を含むDfs情報が格納され得る。しかしながら、様々な実施形態において、この情報は、ネットワーク250からアクセス可能な任意のコンピュータ可読媒体によって格納され得る。

30

【0030】

そしてさらに、Dfsサーバ230は、やはり、別の企業イントラネットまたはインターネットの別の部分とすることのできる別のネットワーク280にも動作可能に結合され得る。ネットワーク280は、どんなコンピュータネットワークでも見られるようなルータ260を介してネットワーク250に動作可能に結合され得る。Dfsルートサーバ230は、1つまたは複数のリンク参照を含み得るDfsルート共有資源240に動作可能に結合され得る。例えば、リンク//r1/s1/l1は、//m1/s1に位置するファイルまたはディレクトリへの参照を提供し得る。同様に、リンク//r1/s1/l2は、//m2/s1に位置するファイルまたはディレクトリへの参照を提供し得る。さらに、1つのリンクは、複数のコンピュータ、サーバ、共有資源、および/またはディレクトリへの参照を提供し得る。ゆえに、サーバコンピュータm2 215を含む多くのコンピュータが、無数のネットワークおよびコンピュータを介してクライアントコンピュータ205に通信するように結合され得る。

40

【0031】

クライアントコンピュータ205は、事実上、ネットワーク250、またはネットワー

50

ク 2 5 0 および 2 8 0 中のどんなサーバ位置からのファイルおよび/またはディレクトリも要求し、検索することができる。しかしながら、クライアントコンピュータ 2 0 5 が、ネットワーク 2 5 0 に接続され得るすべてのコンピュータに関する情報を維持することは実現不可能である。一般に、要求は、ファイルまたはディレクトリを求めるものとして行うことができる。そのため、ネットワーク 2 5 0 中の任意のコンピュータから任意のファイル（および/または本開示の残りの部分全体を通じてファイル要求および検索によって示唆されるディレクトリなど）が検索され得るが、クライアントコンピュータ 2 0 5 には、通常、別のソースからの要求されたファイルの場所が提供される。ゆえに、D f s サーバコンピュータ 2 3 0 は、クライアントコンピュータ 2 0 5 がローカルで利用できないファイルを要求したときに、D f s サーバコンピュータ 2 3 0 が、各ターゲットが要求されたファイルへのパスに対応するターゲットのリストを提供することができるように、ネットワーク 2 5 0 および 2 8 0 に結合された多くのコンピュータに関する情報を維持するように動作することができる。

10

【 0 0 3 2 】

通常、D f s サーバコンピュータ 2 3 0 は、参照要求に回答して何百または何千ものターゲットを返すことができ、いくつかのパラメータに従ってそれらのターゲットの順序を提供することができる。参照をソートする 2 つのそのような方法として、各ターゲットに関連付けられたサイトコストによるソートおよび各ターゲットに関連付けられたターゲット優先度でのソートが考えられる。これらのソート方法のそれぞれについて以下でさらに詳細に説明する。

20

【 0 0 3 3 】

ターゲットのリストが D f s サーバコンピュータ 2 3 0（または参照要求に回答してターゲットを提供し、ソートするように動作可能な他の専用コンピュータシステム）によってソートされると、それがクライアントコンピュータ 2 0 5 に返される。次いで、クライアントコンピュータ 2 2 0 は、ターゲットと接続が確立されるまで、繰り返し、（ソートされている順序で）各ターゲットと接続を確立しようと試み始めることができる。ファイルは、共有資源 s 1 2 1 2 におけるサーバコンピュータ m 1 2 1 0、または単に / / m 1 / s 1 に対応するターゲットに位置し得る。共有資源 s 1 2 1 2 は、サーバコンピュータ m 1 2 1 0 上のファイルサーバ 2 1 1 と関連付けることができ、ゆえに、D f s サーバコンピュータ 2 3 0 によって参照され得る。

30

【 0 0 3 4 】

同様に、要求されたファイルは、共有資源 s 2 2 1 7 におけるサーバコンピュータ m 2 2 1 5 に位置し得る。共有資源 s 2 2 1 7 もまたサーバコンピュータ m 2 2 1 5 上のファイルサーバ 2 1 6 と関連付けることができ、ゆえに、D f s サーバコンピュータ 2 3 0 によって参照され得る。しかしながら、サーバコンピュータ 2 1 5 はネットワーク 2 8 0 を介して動作可能に結合され得るため、クライアントコンピュータによって参照され得るパスはより複雑になり得る。というのは、通信接続がルート D f s サーバ 2 3 0 を介して確立され得るからである。図 2 のアーキテクチャにおける D f s を使用することによって、ネットワーク中の任意の場所からのファイルが、D f s サーバコンピュータ 2 3 0 によってクライアントコンピュータ 2 0 5 に提供される参照を介して検索され得る。

40

【 0 0 3 5 】

図 3 に、本発明の一態様による、分散ファイルシステムを使ってファイルを要求し、検索するために行われるステップを一般的に表す流れ図を示す。クライアントコンピュータ 2 0 5 は、D f s の一部とすることもでき、D f s 全体のどこかに位置するファイルへのシームレスなアクセスを提供し得る D f s クライアント 2 7 0 を含むこともできる。したがって、クライアントコンピュータ 2 0 5 は、あたかも要求されるすべてのファイルがローカルで見つかるかのようにファイルを要求し、検索することができる。

【 0 0 3 6 】

クライアントコンピュータ 2 0 5 が、クライアントコンピュータ 2 0 5 においてローカルで利用できない特定のファイルを必要とするときには、D f s クライアント 2 7 0 を使

50

ってクライアントコンピュータ205においてファイル要求が開始され得る。様々な実施形態において、Dfsクライアントは、アクティブディレクトリサーバ220上にも常駐することができ、Dfsルートサーバ230の場所を探すのに使用され得る。Dfsサーバコンピュータ230は、クライアントコンピュータ205に、ステップ308で、Dfsを介して利用可能なリモートサーバコンピュータに対応するターゲットのリストを含み得る参照応答を提供することができる。ターゲットリストは、サイトコスト(その方法については以下で図4との関連で説明する)、関連優先度(その方法については以下で図5との関連で説明する)、および/またはサイト認識(以下で一般的に説明する)といった様々なパラメータに従ってソートされ得る。

【0037】

さらに、参照応答は、それぞれがターゲットのグループを含む有界集合(bounded set)の指示を含み得る。一実施形態において、有界集合の先頭の指示は、その有界集合中の第1のターゲットに関する情報に関連付けられた境界値(demarcation value)を提供することによって行われ得る。これは、参照応答としてクライアントコンピュータに、ターゲットとそのクライアントコンピュータの間でファイルまたはディレクトリを検索することに伴うコストなどの境界属性を提供することによって達成され得る。有界集合は、サイトコスト、サイト認識、ターゲット優先度、および/または参照に関与するコンピュータの健全性を含む任意の数のパラメータに基づくものとして行うことができる。本明細書で使用する場合、健全性(health)とは、参照を提供するための応答時間、トラフィック負荷、プロセッサ負荷、および、性能メトリックを含む参照に関与するコンピュータのその他の属性を意味し得る。

【0038】

クライアントコンピュータ205が参照応答を受け取ると、クライアントコンピュータ205は、ソートされた順序でリスト中の各ターゲットと通信セッションを確立しようと試み始めることができる。ゆえに、クライアントコンピュータ205は、ステップ310でターゲット1を試みるが、接続に失敗し得る。次いで、クライアントコンピュータ205は、ステップ312でターゲット2を試みるが、やはり接続に失敗し得る。クライアントコンピュータ205は、ステップ314でターゲットmとの接続が確立され得るまで、参照応答中の各ターゲットに接続しようと試み続けることができる。ターゲットへの接続が確立されると、そのターゲットは設定ターゲットに指定され、そのターゲットとの通信の連続性を保持するために、利用不能になるまで使用され得る。例えば、ターゲットmは、クライアントコンピュータ205が、例えば、新しいターゲットを発見するために新しい参照を要求するまで、あるいは別の方法で指示されるまで、その後のファイル検索のための設定ターゲットとして設定され得る。

【0039】

実装形態は、同じ効果を達成しつつ、しかも本発明の範囲を逸脱することなく、これらのステップを異なる順序で実行するよう選択することもでき、あるいは効率性や柔軟性のためにこれらのステップの一部だけを実行するよう選択することもできることを当分野の技術者は理解するであろう。さらに、この同じ概念は、そもそも、Dfsサーバ230の場所を探す際にも適用され得る。というのは、特定のDfsサーバの参照は、ほとんど同じやり方でアクティブディレクトリサーバ220によって提供され得るからである。

【0040】

図4に、本発明の一態様による、サイトコストに基づいて参照応答を組み立てるために行われるステップを一般的に表す流れ図を示す。一般に、参照応答は、サイトコストやサイト認識など、いくつかのパラメータに従ってソートされ得る。一実施形態において、クライアントコンピュータ205と同じサイト内にあるターゲットは、サイトコストに従ってソートする前にサイト認識に従ってソートされ得る。例えば、クライアントコンピュータ205と同じサイト内のターゲットは、参照応答の先頭にソートすることができ、次いで、サイトコストパラメータを使って残りのターゲットがソートされ得る。図4は、Dfsサーバコンピュータ230によって生成された参照応答中のターゲットをソートする

10

20

30

40

50

ためにサイトコストおよびサイト認識を使用する方法の一実施形態を示し得るものである。しかしながら、任意の数のパラメータまたはパラメータの組合せを使って参照応答のターゲットがソートされ得ることを当分野の技術者は理解すると考えられる。

【0041】

Dfsサーバコンピュータ230は、クライアントコンピュータ205からの参照要求を満たすことのできるターゲットのリストを編成することができる。これらのターゲットは、まず初めにステップ402で、ランダムな順序で組み立てられ得る。次いで、Dfsサーバコンピュータ230は、クライアントコンピュータ205と同じサイト（例えば、/ / m 1 / s 1 / など）内に存在し得る各ターゲットを識別することができる。ステップ404で、これらの識別されたターゲットが、次にサイト認識パラメータについてソートされる参照応答を生じるように、リストの先頭に移動され得る。様々な実施形態において、参照応答は、この時点でクライアントコンピュータ205に返送され得る。しかしながら、他の実施形態では、参照応答は、図4に示すように、別のパラメータ、サイトコストに従ってさらにソートされ得る。

10

【0042】

ステップ408で、残りのターゲットが関連サイトコストに従ってソートされ、有界集合に区分され得る。そのターゲットに関連付けられた第1のサイトコストを持つ各ターゲットが、同じ関連サイトコストを持つ他のターゲットと共に有界集合に移動され得る。同様に、第2の有界集合は、第2の関連サイトコストを持つ1つまたは複数のターゲットを持ち得る。ゆえに、この場合、ターゲットのリストを有界集合にソートすることができ、第1の有界集合は最低の関連サイトコストを持つターゲットを含むことができ、次の有界集合は次に低い関連サイトコストを持つターゲットを含むことができ、以下同様である。

20

【0043】

その場合、各有界集合は、ランダムな順序でリストされ、それに関連付けられた同じサイトコストを持つ多数のターゲットを含み得る。ゆえに、クライアントコンピュータ205が通信セッションを確立しようとする試みを次々と反復する際、クライアントコンピュータ205は、最初に、最低のサイトコストのターゲットを次々と反復することができる。

【0044】

さらに、ステップ410で、各有界集合の先頭にリストされたターゲットは、有界集合の先頭を示すように設定され得る境界ビットにも関連付けられ得る。このようにして、クライアントコンピュータは、各有界集合間の境界を容易に識別することができる。境界ビットは、そのターゲットが、リスト中の前のターゲットより高い関連サイトコストを持つという指示として働くことができる。ゆえに、サイトコスト計算(site-costing)に基づく有界集合を提供することによって、フェイルバックまたはフェイルオーバーに関するより多くの情報に基づく決定が行われ得る。次いで、ステップ412で、Dfsサーバコンピュータ230は、サイト認識およびサイトコストに基づくソート済みターゲットリストを持つ最終的な参照応答を組み立てることができ、その参照応答をクライアントコンピュータ205に返送することができる。

30

【0045】

図5に、本発明の一態様による、ターゲット優先度に基づいて参照応答を組み立てるために行われるステップを一般的に表す流れ図を示す。この場合もやはり、参照応答は、簡単に前述したようにターゲット優先度および/またはサイト認識によるなど、いくつかのパラメータに従ってソートされ得る。図5は、Dfsサーバコンピュータ230によって生成された参照応答中のターゲットをソートするためにターゲット優先度およびサイト認識を使用する方法の一実施形態を示し得るものである。しかしながら、任意の数のパラメータまたはパラメータの組合せを使って参照応答のターゲットがソートされ得ることを当分野の技術者は理解すると考えられる。前述のように、Dfsサーバコンピュータ230は、クライアントコンピュータ205からの参照要求を満たすことのできるターゲットのリストを編成することができる。これらのターゲットは、まず初めにステップ502で、

40

50

ランダムな順序で組み立てられ得る。次いで、D f sサーバコンピュータ230は、クライアントコンピュータ205と同じサイト（例えば / / m 1 / s 1 / など）内にある各ターゲットを識別することができる。ステップ504で、これらの識別されたターゲットは、次に、サイト認識パラメータについてソートされる参照応答を生じるようにリストの先頭に移動され得る。いくつかの実施形態では、参照応答は、この時点でクライアントコンピュータ205に返送され得る。しかしながら、他の実施形態では、参照応答は、図5に示すように、別のパラメータ、ターゲット優先度に従ってさらにソートされ得る。

【0046】

それに関連付けられた、大域的に高いなど第1のターゲット優先度を持つ各ターゲットが、同じ関連ターゲット優先度を持つ他のターゲットと共に有界集合に移動され得る。同様に、第2の有界集合は、大域的に低いなど第2の関連ターゲット優先度を持つ1つまたは複数のターゲットを含み得る。ゆえに、この場合、ターゲットのリストを有界集合にソートすることができ、第1の有界集合は最高の関連ターゲット優先度、すなわち大域的に高い優先度を持つターゲットを含むことができ、次の有界集合は、次に高い関連ターゲット優先度、すなわち普通の高さの優先度を持つターゲットを含み、以下同様である。

10

【0047】

その場合、各有界集合は、ランダムな順序でリストされ、それに関連付けられた同じターゲット優先度を持つ多数のターゲットを含み得る。ゆえに、クライアントコンピュータ205がターゲットを次々と反復して、通信セッションを確立しようとする際、クライアントコンピュータ205は、最初に、最高のターゲット優先度のターゲットを次々と反復することができる。

20

【0048】

さらに、ステップ510で、各有界集合の先頭にリストされたターゲットは、その有界集合の先頭を指示するように設定された境界ビットを持つこともできる。このようにして、クライアントコンピュータ205は、各有界集合間の境界を容易に識別することができる。境界ビットは、そのターゲットが、リスト中の前のターゲットより低い関連ターゲット優先度を持ち得ることの指示として働くことができる。ゆえに、ターゲット優先度に基づく有界集合を提供することによって、フェイルバックまたはフェイルオーバーに関するより多くの情報に基づく決定が行われ得る。次いで、ステップ512で、D f sサーバコンピュータ230は、サイト認識およびターゲット優先度に基づくソート済みターゲットリストを持つ最終的な参照応答を組み立てることができ、その参照応答をクライアントコンピュータ205に返送することができる。

30

【0049】

D f sサーバコンピュータ230は、ソート済み参照応答を設定するに際してソートパラメータの組合せも使用することができる。例えば、第1のソートは、同じサイトのターゲットをリストの先頭に移動することとすることができる。次いで、ターゲットは、それぞれに関連付けられた大域的優先度に基づいてソートされ得る。例えば、関連付けられた大域的に高い優先度を持つすべてのターゲットは、参照応答の先頭にソートされ得る。同様に、関連付けられた大域的に低い優先度を持つすべてのターゲットは、リストの末尾にソートされ得る。次いで、（大域的に普通と呼ばれ得る）残りのターゲットが、サイトコストに基づいてさらに有界集合にソートされ得る。

40

【0050】

別の実施形態において、第2のソートは、サイトコストに基づき、各集合の先頭にあるターゲット中の集合境界ビットによって表される有界集合を設定することができる。次いで、各有界集合は、（本来的に同じ関連サイトコストを持ち得る）各有界集合内のターゲットが、各ターゲットに関連付けられた優先度に従ってさらに順序付けされ得るように、ターゲット優先度に従ってソートされ得る。ゆえに、サイトコストに基づく有界集合内で、その有界集合の先頭にあるターゲットは、高い優先度に関連付けることができ、次のターゲットのグループは、次に高い優先度（例えば普通など）に関連付けることができ、以下同様である。

50

【 0 0 5 1 】

さらに別の実施形態では、ターゲットが、大域的な優先度に基づき、各有界集合内でソートされるように、前述の2つの実施形態と一緒に実施され得る。

【 0 0 5 2 】

別の例として、第1のソートは、やはり、同じサイトのターゲットをリストの先頭に移動することとすることができる。次いで、第2のソートは、ターゲット優先度に基づき、各集合の先頭にあるターゲット中の集合境界ビットによって表される有界集合を設定することができる。次いで、各有界集合は、(本来的に同じ関連ターゲット優先度を持ち得る)各有界集合内のターゲットが、各ターゲットに関連付けられたサイトコストに従ってさらに順序付けされるように、サイトコストに従ってソートされ得る。ゆえに、ターゲット優先度に基づく有界集合内で、その有界集合の先頭にあるターゲットは、最低のサイトコストに関連付けることができ、次のターゲットのグループは、次に低いサイトコストに関連付けることができ、以下同様である。

10

【 0 0 5 3 】

参照応答中のターゲットを設定された順序でソートし、参照応答中の各ターゲット内の一定のビットを設定することにより、次いで、クライアントコンピュータは、以下で説明するように、参照応答中のソート済みターゲットを使って効率のよいターゲットフェイルバックポリシーおよび優先フェイルバックポリシーを実施することができる。

【 0 0 5 4 】

(ターゲットフェイルバックおよび優先度)

サーバコンピュータのネットワークにおける所望のファイルを持ち得るサーバコンピュータのリストを要求するクライアントコンピュータのためのソート方法に加えて、このシステムおよび方法は、参照応答として提供されるソート済みターゲットのリストを使用し得るターゲットフェイルバックおよびターゲット優先度ポリシーも提供することができる。明らかのように、コンピュータシステムは、サイトコストに従ってソートされたターゲットのリストからターゲットを選択し、設定ターゲットに指定することができる。次いで、コンピュータシステムは、その設定ターゲットが、ソート済みリスト中のすべての利用可能なターゲットに比べて、最低のサイトコストに関連付けられているかどうか判定することができる。そうでない場合、システムは、設定ターゲットより低いサイトコストに関連付けられている別のターゲットにフェイルバックし、その新しいターゲットを設定ターゲットに指定することができる。理解されるように、本明細書で説明する様々な流れ図およびシナリオは例にすぎず、説明するターゲットフェイルバックおよびターゲット優先度ポリシーが適用される他の多くのシナリオがある。

20

30

【 0 0 5 5 】

図面の図6を見ると、本発明の一態様による、ソート済み参照応答と共に使用され得るより低いサイトコストのターゲットにフェイルバックするために行われるステップを一般的に表す流れ図が示されている。ターゲットフェイルバックポリシーの様々な実施形態において、Dfsサーバコンピュータ230は、参照応答を生成するときDfsによって使用され得るターゲットに関する情報をソートすることができる。一実施形態では、Dfsサーバコンピュータ230は、サイト認識に基づいてターゲットに関する情報をソートすることができる。このモードでは、参照応答は、本質的に、2つのターゲット集合で構成され得る。一方の集合は、クライアントコンピュータ205と同じサイトにあるターゲットを含み、他方の集合はクライアントコンピュータ205以外のサイトにあるターゲットを含む。最初、各集合中のターゲットはランダムに配列され得る。

40

【 0 0 5 6 】

別の実施形態において、Dfsサーバコンピュータ230は、サイトコストに基づいて参照をソートすることができる。この動作モードでは、参照応答は、複数の有界集合にソートされ得る。各有界集合は、(要求側クライアントコンピュータに関連するサイトコスト情報に基づいて決定され得る)Dfsサーバによって決定されるのと同じサイトコストを持つターゲットを含み得る。有界集合は、クライアントと同じサイトにあるターゲット

50

が第1の有界集合に含まれ、次に低いサイトコストを持つターゲットが第2の集合に含まれ、以下同様とすることができるよう、サイトコストの昇順で配列され得る。最初、各集合内のターゲットは、ランダムに配列され得る。

【0057】

ゆえに、ステップ602でクライアントコンピュータが参照を要求すると、Dfsサーバコンピュータ230によってソート済み参照応答が提供され得る。次いで、ステップ604で、クライアントコンピュータは、第1の利用可能なターゲットを決定するために、参照応答中の各ターゲットと接続を確立しようと試み始めることができる。ターゲットが、ネットワークエラー、ターゲットクラッシュなど、いくつもの理由でアクセス不能であるとき、Dfsクライアントは、次の利用可能なターゲットにフェイルオーバーすることができる。例えば、参照応答のターゲットリスト中の次のターゲットと接続を確立しようと試みることができる。ターゲットとの接続が確立されると、ステップ606で、その特定のターゲットが設定ターゲットに指定され得る。

10

【0058】

設定ターゲットより最適であり得る1つまたは複数のターゲットが再度利用可能になったとき、Dfsクライアントは、連続性およびシームレス性のために設定ターゲットを引き続き使用することもでき、より最適な(more optimal)利用可能なターゲットを獲得し、または再獲得することもできる。フェイルバックポリシーおよびターゲットの有界集合を持つことにより、利用可能な場合に、より低いサイトコストを持つターゲットなど、より最適なターゲットを獲得し、または再獲得することが可能になる。

20

【0059】

ゆえに、ターゲットフェイルバックポリシーが指定されている場合(例えば、ターゲットフェイルバックポリシービットが設定されているなど)、ステップ608で、クライアントコンピュータは、より低いサイトコストのターゲットなど、より好ましいターゲットが利用可能であり得ることを判定することができる。この判定は、いつでも行うことができ、クライアントコンピュータ205によっても、Dfsサーバコンピュータ230によっても、あるいは他の任意のリモートコンピュータによっても開始され得る。さらに、フェイルバック時刻は、その設定ターゲットとの接続が確立され得た時刻からの経過期間に対応し得る。例えば、クライアントコンピュータは、第1の設定ターゲットでの接続が確立されてから30分後により最適なターゲットの有無をチェックすることができる。さらに、フェイルバック時刻は、半時間ごとや、5分ごとなど、特定の時刻にも対応し得る。さらに、フェイルバック時刻は、クライアントコンピュータ205からのファイルを求めるまさしく次の要求にも対応し得る。

30

【0060】

フェイルバック時に、より低いサイトコストのターゲットなどより好ましいターゲットが利用可能である場合、ステップ612で、その利用可能なより好ましいターゲットとの接続を確立することができ、利用可能なより好ましいターゲットが設定ターゲットに指定され得る。しかしながら、より低いサイトコストを持つものなど、より好ましい利用可能なターゲットがない場合、ステップ610で、設定ターゲットが現在のターゲットとして引き続き使用され得る。フェイルバックポリシーは、参照応答中の所与のDfsルート/リンクについて実施され得る。ターゲットフェイルバックが個々のリンクレベルで使用可能とされ得ない場合、Dfs名前空間全体での設定が使用され得る。一実施形態において、ターゲットフェイルバックは、ターゲットフェイルバックがDfs名前空間レベルで使用可能とされ得るときには、リンクレベルでは使用不可とされ得る。

40

【0061】

ターゲットフェイルバック情報は、一実施形態では、Dfsメタデータなど、他のDfs情報と共に格納され得る。例えば、各ターゲットのメタデータのルート/リンク別情報における「Type」フィールドが使用され得る。このフィールド中の自由なビット位置が、既存のシステム全体にわたって機能するように識別され、検証され得る。この新しいビット位置は、PKT__ENTRY__TYPE__TARGETFAIL.BACKである

50

と定義され得る。

【 0 0 6 2 】

ゆえに、一実施形態では、以下の擬似構造によって、(「 I D B L O B 」と呼ばれる)ドメインベースの D f s のメタデータでの典型的な形式が定義され得る。

【 0 0 6 3 】

【表 1】

GUID	VolumeId;	
USHORT	PrefixLength;	10
WCHAR	Prefix [PrefixLength];	
USHORT	ShortPrefixLength;	
WCHAR	ShortPrefix[ShortPrefixLength] ;	
ULONG	Type;	
ULONG	State;	
USHORT	CommentLength;	20
WCHAR	Comment[CommentLength];	
FILETIME	PrefixTimeStamp;	
FILETIME	StateTimeStamp;	
FILETIME	CommentTimeStamp;	
ULONG	Version;	30

【 0 0 6 4 】

同様に、以下の擬似構造によって、(「 I D B L O B 」と呼ばれる)独立型 D f s のメタデータでの典型的な形式が定義され得る。

【 0 0 6 5 】

【表 2】

```

USHORT    PrefixLength;

WCHAR     Prefix[PrefixLength] ;

USHORT    ShortPrefixLength;

WCHAR     ShortPrefix [ShortPrefixLength];

GUID      VolumeId;

ULONG     State;

ULONG     Type;

USHORT    CommentLength;

WCHAR     Comment[CommentLength];

FILETIME  PrefixTimeStamp;

FILETIME  StateTimeStamp;

FILETIME  CommentTimeStamp;

```

10

20

【0066】

NetDfsGetInfo/NetDfsEnum APIは、ルート/リンク別メタデータから設定を取り出すことができる。パラメータを検証することのできるAPIフロントエンドがしかるべく変更され、「infoレベル」特有の情報を解釈し、返すことのできるルーチンもしかるべく変更され得る。

【0067】

ターゲット優先度とターゲットフェイルバック両方をサポートするために、Dfsクライアントは、より低いサイトコストおよび/またはより高い優先度の別のターゲットへのフェイルバックを試みることができる。一実施形態では、Dfsクライアントは、現在アクティブなターゲットと同じサイトコストおよび/または同じ優先度の別のターゲットにフェイルバックすることができない。

30

【0068】

ターゲット集合は、一実施形態では、ターゲットの有界集合であると定義され得る。例えば、ターゲットの有界集合は、同じサイトコストを持つランダムにソートされたターゲットの集合とすることができる。ターゲットの有界集合での境界の指示は、例えば、参照応答では、参照エントリに境界の指示を含めることによって行われ得る。

【0069】

したがって、あるルート/リンクでのDfsクライアントへの参照応答は、一実施形態では、そのルート/リンクにフェイルバックが使用可能であるか否か指示することができる。返されるターゲット間の有界集合境界も指示することができる。一実施形態では、あるルート/リンクでの既存の参照応答の形式が、以下のように、参照ヘッダと参照エントリの両方における既存のビットフィールド定義に新しいビットフィールドを付加することによって使用され得る。

40

【0070】

【表 3】

```
typedef struct {  
    USHORT PathConsumed; // Number of WCHARs consumed in  
    DfsPathName
```

【 0 0 7 1 】

【表 4】

```

USHORT NumberOfReferrals; // Number of referrals
    contained here

struct {
    ULONG ReferralServers : 1; // Elements in
        Referrals[] are referral servers
    ULONG StorageServers : 1; // Elements in Referrals[]
        are storage servers
    ULONG TargetFailback : 1; // Target Fail-back is
        enabled for this namespace root/link
};
    ULONG ReferralHeaderFlags
};

union { // The vector of referrals
    DFS_REFERRAL_V1 v1;
    DFS_REFERRAL_V2 v2;
    DFS_REFERRAL_V3 v3;
    DFS_REFERRAL_V4 v4;
    } Referrals[i]; // [ NumberOfReferrals ]
    //
    // WCHAR String Buffer[]; // Used by DFS ..
        REFERRAL_V2
    //
} RESP _GET_DFS_REFERRAL;

```

【 0 0 7 2 】

【表 5】

```

typedef RESP_GET_DFS_REFERRAL *PRESP_GET_DFS_REFERRAL;

typedef struct {

    USHORT VersionNumber; // == 4

    USHORT Size; // Size of this whole element

    USHORT ServerType; // Type of server: 0 == Don't know,
                        // 1 == SMB, 2 == Netware
};
union {
    struct {
        USHORT StripPath : 1; // Strip off PathConsumed
                                // characters from front of DfsPathName prior to
                                // submitting name to UncShareName
        USHORT NameListReferral : 1; // This referral
                                // contains an expanded name list
        USHORT TargetSetBoundary : 1 // Denotes this target
                                // is the first in a target set // All targets in
                                // have the same site-cost or priority rank;
    };
    USHORT ReferralEntryFlags
};
    ULONG TimeToLive; // In number of seconds
};
union {
    struct {
        USHORT DfsPathOffset; // Offset from beginning of
                                // this element to Path to access
    };
};

```

【 0 0 7 3 】

【表 6】

```

USHORT DfsAlternatePathOffset; // Offset from
    beginning of this element to 8.3 path
USHORT NetworkAddressOffset; // Offset from
    beginning of this element to Network path
GUID ServiceSiteGuid; // The guid for the site
};
10
struct {
    USHORT Special NameOffset; // Offset from this
        element to the special name string
    USHORT NumberOfExpandedNames; // Number of expanded
        names
        20
    USHORT ExpandedNameOffset; // Offset from this
        element to the expanded name list
};
} DFS_REFERRAL_V4;
typedef DFS_REFERRAL_V4 *PDFS_REFERRAL_V4;
30

```

【0074】

参照応答中のターゲット間の有界集合境界の指示を提供し、またはフェイルバックが使用可能にされているか否か指示するために、分離データ構造(separate data structure)など、他のデータ構造も使用され得ることを当分野の技術者は理解するであろう。

【0075】

図7に、本発明の一態様による、ソート済み参照応答と共に使用され得るより高い優先度のターゲットにフェイルバックするために行われるステップを一般的に表す流れ図を示す。ターゲットフェイルバックポリシー環境の様々な実施形態において、Dfsサーバコンピュータ230は、サイト認識に基づいてターゲットに関する情報をソートすることができる。このモードでは、参照応答は、本質的に、2つのターゲット集合で構成され得る。一方の集合はクライアントコンピュータ205と同じサイトにあるターゲットを含み、他方の集合はクライアントコンピュータ205以外サイトにあるターゲットを含む。最初、各集合中のターゲットはランダムに配列され得る。

【0076】

別の実施形態において、Dfsサーバコンピュータ230は、ターゲット優先度に基づいて参照をソートすることができる。この動作モードでは、参照応答は、複数の有界集合にソートされ得る。各有界集合は、参照を要求しているDfsサーバによって決定されるのと同じターゲット優先度を持ち得るターゲットを含むことができる。有界集合は、最高の優先度のターゲットが第1の有界集合に含まれ、次に高い優先度を持つターゲットが第

2の集合に含まれ、以下同様とすることができるよう、ターゲット優先度の降順に配列され得る。最初、各集合中のターゲットは、ランダムに配列され得る。

【0077】

Dfsサーバコンピュータ230は、ソート済み参照応答を設定するに際してソートパラメータの組合せを使用することもできる。例えば、第1のソートは、同じサイトのターゲットをリストの先頭に移動させることとすることができる。次いで、ターゲットは、それぞれに関連付けられた大域的優先度に基づいてソートされ得る。例えば、関連する大域的に高い優先度を持つすべてのターゲットは、参照応答の先頭にソートされ得る。同様に、関連する大域的に低い優先度を持つすべてのターゲットは、リストの末尾にソートされ得る。次いで、(大域的に普通と呼ばれ得る)残りのターゲットがサイトコストに基づいてさらに有界集合にソートされ得る。

10

【0078】

別の実施形態において、第2のソートは、サイトコストに基づき、各集合の先頭にあるターゲット中の集合境界ビットによって表される有界集合を設定することができる。次いで、各有界集合は、(本来的に同じ関連サイトコストを持ち得る)各有界集合内のターゲットが、各ターゲットに関連付けられた優先度に従ってさらに順序付けされ得るように、ターゲット優先度に従ってソートされ得る。ゆえに、サイトコストの基づく有界集合内では、その有界集合の先頭にあるターゲットは、高い優先度に関連付けることができ、次のターゲットのグループは次に高い優先度(例えば普通など)に関連付けることができ、以下同様である。

20

【0079】

さらに別の実施形態においては、ターゲットが大域的な優先度により、各有界集合内でソートされるように、前述の2つの実施形態が一緒に実施され得る。ゆえに、ステップ702で、クライアントコンピュータ205が参照を要求すると、Dfsサーバコンピュータ230によってソート済み参照応答が提供され得る。次いで、ステップ704で、クライアントコンピュータは、第1の利用可能なターゲットを決定するために参照応答中の各ターゲットとの接続を確立しようと試み始めることができる。例えば、参照応答のターゲットリスト中の次のターゲットとの接続を確立しようと試みることができる。ターゲットとの接続が確立されると、ステップ706で、その特定のターゲットが設定ターゲットに指定され得る。

30

【0080】

設定ターゲットより最適であり得る1つまたは複数の前のターゲットが再度利用可能になったとき、Dfsクライアントは、連続性およびシームレス性のためにその設定ターゲットを引き続き使用することもでき、より最適なターゲットを獲得し、または再獲得することもできる。フェイルバックポリシーおよびターゲットの有界集合を持つことにより、利用可能な場合に、より高いターゲット優先度を持つターゲットなど、より最適なターゲットを獲得し、または再獲得することが可能になる。

【0081】

ゆえに、ターゲットフェイルバックポリシーが指定される場合(例えば、ターゲットフェイルバックビットが設定されているなど)には、ステップ708で、クライアントコンピュータは、より好ましい有界集合内のターゲットなど、より好ましいターゲットが利用可能であることを判定することができる。この判定はいつでも行うことができ、クライアントコンピュータ205によっても、Dfsサーバコンピュータ230によっても、あるいは他の任意のリモートコンピュータによっても開始され得る。さらに、優先度チェック時刻は、設定ターゲットの設定からの経過期間に対応し得る。例えば、クライアントコンピュータ205は、第1の設定ターゲットが設定されてから30分後により最適なターゲットの有無をチェックすることができる。さらに、優先度チェック時刻は、半時間ごとや5分ごとなど、特定の時刻にも対応し得る。さらに、優先度チェック時刻は、クライアントコンピュータ205からのファイルを求めるまさしく次の要求にも対応し得る。

40

【0082】

50

優先度チェック時刻において、より高いターゲット優先度のターゲットが利用可能である場合、ステップ 712 で、その利用可能なより高いターゲット優先度のターゲットと接続が確立され、そのより高いターゲット優先度のターゲットが設定ターゲットに指定され得る。しかしながら、より高いターゲット優先度を持つどんな利用可能なターゲットもない場合、ステップ 710 で、設定ターゲットが現在のターゲットとして引き続き使用され得る。

【0083】

ゆえに、Dfs 自体の中でターゲットを順序付けする単純な方法がソートに役立つ利点を提供し得る。サーバターゲット優先度は、前述のようにサイトコスト計算と同時に行われ得る。サーバターゲット優先度は、より高い優先度のターゲットが利用可能である場合にはより低い優先度のターゲットがトラフィックを受け取れないようにターゲット間の厳然たる順序付け(hard ranking)を作成することができる。

10

【0084】

前述のように、参照応答は、各ターゲットの関連サイトコストに基づいて有界集合にソートされ得る。サーバターゲット優先度を用いる場合も、これらの有界集合は、やはり、アクセスするターゲットのコストに基づくものとして行うことができる。サーバターゲット優先度は、単に、ターゲットのコストソート基準を拡張し得るだけであり、そのため、これらの集合は、同じサイトコストおよびサーバターゲット優先度を持つターゲットとすることができる。

【0085】

サーバターゲット優先度は、一実施形態では、2つの値、すなわち優先クラスと優先順位で表すことができる。優先クラスは、等しいサイトコストのターゲットの集合内、すなわち局所的レベルと、大域的レベルの2つのレベルで定義され得る。これらの各レベル内で、高い、普通、および低い優先度のターゲットという大雑把な順序付けがなされ得る。これは、以下の5つの優先クラスを提供することができる。

20

大域的に高い

サイトコストが高い

サイトコストが普通

サイトコストが低い

大域的に低い

30

【0086】

これらは、上記の優先順位で順序付けされ得る。「大域的に普通」のクラスは、サイトコストクラスと等価であるとみなされ得るので、別個の「大域的に普通」のクラスはあり得ないことに留意されたい。優先順位は、単なる整数の順位、0、1、2、等々とすることができる。

【0087】

参照を順序付けする際の、一実施形態でのプロセスは以下の通りとすることができる。

1. 大域的に高いターゲットおよび大域的に低いターゲットの集合が、残りの「大域的に普通」のターゲットと共に識別され得る。

2. これら3つの集合は、大域的に高い、大域的に普通、大域的に低い、の優先順位で配列され得る。

40

3. 除外ポリシーが設定され得る場合、除外集合内のターゲットは除去され得る。

4. これら3つの各集合内で、ターゲットは、サイトコスト機構によって(ローカルサイトまたは完全なサイトコスト計算によって)順序付けされ、等しいサイトコストのターゲットの有界集合が生成され得る。

5. 等しいサイトコストの「大域的に普通」のターゲットの集合内で、ターゲットは、優先クラス、すなわち、サイトコストが高い、普通、低いにより順序付けされ得る。

6. 等しいサイトコストおよび優先クラスのターゲットの有界集合内で、ターゲットは、優先順位(0が最高である)によって順序付けされ得る。

7. 等しいサイトコスト、優先クラスおよび優先順位のターゲットの有界集合内で、タ

50

ーゲットは、負荷平衡化のためにランダムにシャッフルされ得る。

【0088】

実装形態は、同じ効果を達成しつつ、しかも本発明の範囲を逸脱することなく、これらのステップを異なる順序で実行するよう選択することもでき、あるいは効率性や柔軟性のためにこれらのステップの一部だけを実行するよう選択することもできることを当分野の技術者は理解するであろう。

【0089】

図式的には、これらは、クライアントがターゲットを受け取るはずの順序の集合とすることができる。

[大域的に高い優先クラス]

[サイトコスト サイトコスト = 0 のターゲットでの高い優先クラス]

[サイトコスト サイトコスト = 0 のターゲットでの普通の優先クラス]

[サイトコスト サイトコスト = 0 のターゲットでの低い優先クラス]

[サイトコスト サイトコスト = 1 のターゲットでの高い優先クラス]

[サイトコスト サイトコスト = 1 のターゲットでの普通の優先クラス]

[サイトコスト サイトコスト = 1 のターゲットでの低い優先クラス]

[大域的に低い優先クラス]

10

【0090】

ターゲット優先度情報はターゲット別とすることができるため、この情報を維持する無理のない場所は、各ルート/リンクごとのターゲットリストを含む Dfs レプリカメタデータ情報中であると考えられる。各ターゲットの情報は、以下の擬似構造によって定義され得る。

20

【0091】

【表7】

FILETIME ReplicaTimeStamp;

ULONG ReplicaState;

ULONG ReplicaType;

USHORT ServerNameLength;

WCHAR ServerName[ServerNameLength];

USHORT ShareNameLength;

WCHAR ShareName[ShareNameLength];

30

【0092】

一実施形態では、ターゲット優先度が、メタデータに格納するために以下のように U C H A R で符号化され得る。

40

- ・ ビット 0 ~ 4 : 優先クラス内の優先順位
- ・ ビット 5 ~ 7 : 優先クラス

【0093】

優先クラスは、一実施形態では、以下の値で表され得る。

サイトコスト普通 (デフォルト) 0 x 0

大域的に高い 0 x 1

大域的に低い 0 x 2

サイトコストが高い 0 x 3

サイトコストが低い 0 x 4

【0094】

50

一例として、大域的に低い優先クラス内の優先順位 1 のターゲットは以下のように符号化され得る。

【0095】

【表 8】

ビット 7 6 5 / 4 3 2 1 0

0 1 0 / 0 0 0 0 1

【0096】

この符号化は、32の優先順位、および8つの可能な優先クラスを提供することができ、8つの可能な優先クラスのうちの5つが、大域的に高い、サイトコストが高い、サイトコストが普通、サイトコストが低い、および大域的に低いという優先順位を使って定義され得る。優先順位は、0から31までの値とすることができ、0が最高の優先順位とみなされる。ゆえに、各集合内で、優先順位0を持つターゲットが最初に返され、優先順位31を持つターゲットが最後に返され得る。

10

【0097】

DfsUTILなどのユーティリティを使って、設定/表示操作のために、優先クラス定義を直接公開することができる。代替として、ユーザインターフェースが、ターゲット属性ページ上のラジオボタンによって優先クラスを公開することもできる。DfsUTILもUIも、「サイトコストが普通の優先クラス」をデフォルトとして選択し得る。

【0098】

ターゲット優先度は、Dfsターゲットの属性とみなされ得る。したがって、前述のUCHAR値は、一実施形態では、アクティブディレクトリとレジストリの両方で、Dfsオブジェクトにターゲット別属性として格納され得る。新しいDFS_REPLICA_INFORMATION構造は以下のように提供され得る。

20

【0099】

【表 9】

```
typedef struct DFS_REPLICA_INFORMATION
```

```
{
```

```
    PVOID      pData;
```

30

```
    ULONG      DataSize;
```

```
    UCHAR      TargetPriority;
```

```
    UCHAR[7]   Unused;
```

```
    ULONG      ReplicaState;
```

```
    ULONG      ReplicaType;
```

40

```
    UNICODE_STRING ServerName;
```

```
    UNICODE_STRING ShareName;
```

```
    DFS_REPLICA_INFORMATION, *PDFS_REPLICA_INFORMATION;
```

```
}
```

【0100】

DFS_TARGET_PRIORITY_CLASSは、一実施形態では、以下のよ

50

うに5つの可能なターゲット優先クラス設定を定義し得る新しいenum型を提供することができる。

【0101】

【表10】

```
typedef enum {
    DfsInvalidPriorityClass = -1,
    DfsSiteCostNormalPriorityClass 0,
    DfsGlobalHighPriorityClass,
    DfsSiteCostHighPriorityClass,
    DfsSiteCostLowPriorityClass,
    DfsGlobalLowPriorityClass
} DFS_TARGET_PRIORITY_CLASS;
```

10

【0102】

デフォルトのターゲット優先度設定は、<PriorityClass=SiteCostNormalPriorityClass, PriorityRank=0>とすることができる。これは意図的なものとするすることができる。というのは、ダウンレベルメタデータのデフォルト値が0になるからである。DfsSiteCostNormalPriorityClassの値は、それがDfsSiteCostHighPriorityClassに比べて優先度が低い場合でさえも、0とすることができる。

20

【0103】

一実施形態では、タプル<優先クラス, 優先順位>とすることができるターゲット優先度の定義をカプセル化するために、以下の構造が提供され得る。

【0104】

【表11】

```
typedef struct _DFS_TARGET_PRIORITY {
    DFS_TARGET_PRIORITY_CLASS    TargetPriorityClass;
    USHORT                        TargetPriorityRank
    USHORT                        Reserved;
} DFS_TARGET_PRIORITY;
```

30

40

【0105】

Dfsサービスでのルートまたはリンクターゲットを表し得るDfsReplicaクラスは、以下の新しいクラスDfsTargetPriorityのインスタンスを含み得る。

【0106】

【表 1 2】

```
class DfsTargetPriority (
    private:
        DFS_TARGET_PRIORITY _TargetPriority;
```

【0107】

DfsTargetPriorityクラスは、DFSメタデータから取り込まれたターゲット優先度情報を含むパックUCHARを、より使用に適したDFS_TARGET_PRIORITYの形に変換することができる。これはまた、DFS_TARGET_PRIORITYが与えられると、パックUCHAR値を作成することもできる。さらに、DFS_TARGET_PRIORITY_CLASSの優先クラス定義と、メタデータに格納するのに使用される優先クラス定義の間のマッピングを実行することもできる。

【0108】

既存のstruct REPLICATION_COST_INFORMATIONは、以下のDfsTargetCostInfoという新しいクラスに変換され得る。

【0109】

【表 1 3】

```
class DfsTargetCostInfo
    private:
        ULONG TargetCost; // Site-cost
        DfsReplica *pReplica; // Server priority is an
        attribute of DfsReplica BOOLEAN IsBoundary;
    public:
        BOOLEAN operator<=(DfsTargetCostInfo &rhs);
        BOOLEAN operator!=(DfsTargetCostInfo &rhs);
        DfsTargetCostInfo& operator=(const DfsTargetCostInfo
        &rhs);
```

【0110】

既存のstruct REFERRAL_INFORMATIONは、DfsTargetCostInfoオブジェクトの配列を含むことになるクラスに変換され得る。参照のシャッフル、ソートおよび生成に関連する論理は、この新しいクラスでカプセル化され得る。このクラスは、生成されたDfsTargetCostInfoオブジェクトの配列を、DFSプロトコルが期待するREFERRAL_HEADER形式に変換することもできる。

【0111】

ターゲット優先度による参照のソートは以下のように実施され得る。2つのターゲット、r1およびr2が与えられたとすると、r1は、r1がr2の前に配列される必要がある場合には、r2であるとみなされ得る。そのため、r1 r2である場合、ソート済みの参照は{r1, r2}のように見えることになる。その比較の実施は、以下の所見

に依存するものである。

【0112】

大域的優先クラスのターゲットでは、まず、優先クラスが比較され得る。また、非大域的優先クラスのターゲットでは、まず、サイトコストが比較され得る。

【0113】

1. 一方または両方のターゲットが大域的優先クラスにある場合：

a. 優先クラスが同じでない場合には、優先クラスに基づいて順序を決定する。例えば、 r_1 が `DfsGlobalHighPriorityClass` にある場合、 r_1 は、 r_2 の優先クラスに関係なく r_2 の前に配置され得る。

b. そうではなく、優先クラスが同じである場合には、 r_1 と r_2 のサイトコストを比較する。 r_1 のサイトコスト $<$ r_2 のサイトコストである場合、 $r_1 < r_2$ である。

c. 優先クラスおよびサイトコストが同じである場合には、ステップ3に進む。

【0114】

2. どちらのターゲットも大域的優先クラスにない場合：

a. ターゲットサイトコストが同じでない場合で、 r_1 のサイトコスト $<$ r_2 のサイトコストである場合には、 $r_1 < r_2$ である。

b. そうではなく、サイトコストが同じである場合には、優先クラスをチェックする。 r_1 が r_2 に比べて高い優先クラスにある場合には、 $r_1 < r_2$ である。

c. サイトコストおよび優先クラスが同じである場合には、ステップ3に進む。

【0115】

3. 優先クラスおよびサイトコストが同じである場合。 r_1 が、 r_2 に比べて高い(数値的に低い)優先順位を持つ場合には、 $r_1 < r_2$ である。

【0116】

大域的優先クラスでは、参照は、まず、サイトコストにより配列され、次いで、優先順位により配列され得る。これが行われるのは、それを他の場合における比較でのモデルとすることができるからである。例えば、大域的でない優先クラスでは、サイトコストが等しいときには、優先順位だけが考慮され得る。

【0117】

図8に、本発明の一態様による、参照応答をソートするために一実施形態で行われるステップを一般的に表す流れ図を示す。参照応答がソートされるとき、前述の方法の様々な組合せが実施され得る。図8の流れ図に示す実施形態では、サイト認識、サイトコスト、およびターゲット優先度が、それぞれ、参照応答中のターゲットをソートする際に使用される。

【0118】

ステップ802で、クライアントコンピュータ要求を満たし得るターゲットのリストが、ソートのために参照応答として組み立てられ得る。次いで、ステップ804で、要求側コンピュータと同じサイト内にあるターゲットが、サイト認識として前述したように、リストの先頭にソートされ得る。ステップ806で、残りのターゲットが大域的優先度に従ってソートされ得る。より具体的には、大域的に高い優先度に関連付けられたターゲットはターゲットリストの先頭の方へソートされ、大域的に低い優先度に関連付けられたターゲットはターゲットリストの末尾の方へソートされ得る。

【0119】

次に、大域的に高いにも大域的に低いにも指定され得ないターゲットが、関連サイトコストに従ってソートされ得る。一実施形態では、そのような大域的に高いにも大域的に低いにも指定され得ないターゲットは、大域的に普通と指定され得る。ゆえに、これらのターゲットは、任意の所与の有界集合中の各ターゲットが、その特定の有界集合中の他のすべてのターゲットと類似の、または同等のサイトコストに関連付けられるように、ターゲットの有界集合にグループ化され得る。さらに、各有界集合中の各先頭のターゲットは、一実施形態では、有界集合の先頭を指示し得る集合境界ビットを含み得る。

【0120】

10

20

30

40

50

各有界集合は、ステップ 810 で、その有界集合中の各ターゲットに関連付けられた優先度に基づいてさらにソートされ得る。ゆえに、高い優先度に関連付けられた有界集合内のターゲットはその有界集合の先頭にソートされ、低い優先度に関連付けられた有界集合内のターゲットはその有界集合の末尾にソートされ得る。このようにして、各有界集合を優先度によってソートし、サイトコストソート内にネストさせ、それを、さらに、大域的優先度ソート内にネストさせることができる。

【0121】

この場合もやはり、実装形態は、同じ効果を達成しつつ、しかも本発明の範囲を逸脱することなく、これらのステップを異なる順序で実行するよう選択することもでき、あるいは効率性や柔軟性のためにこれらのステップの一部だけを実行するよう選択することもできることを当分野の技術者は理解するであろう。

10

【0122】

このように、このシステムおよび方法は、ターゲット優先度とターゲットフェイルバックの両方をサポートすることができる。クライアントコンピュータシステムは、より低いサイトコストおよび/またはより高い優先度の別のターゲットにフェイルバックしようと試みることができる。本発明は、サイト認識および/またはターゲット優先度に基づく参照のソートをサポートすることができる。サイト認識に基づいて参照をソートすることにより、クライアントコンピュータシステムに参照を提供する Dfs サーバと同じサイトにある 1 組のターゲット、および他のすべてのターゲット含む別の組のターゲットが提供され得る。ターゲット優先度に基づいて参照をソートすることにより、各有界集合が、参照を要求しているクライアントコンピュータシステムによって決定されるのと同じターゲット優先度を持つターゲットを含み得る有界集合が、優先度の順序で提供され得る。一実施形態において、参照応答は、同じサイトコストおよびサーバターゲット優先度を持つ有界集合にソートされ得る。

20

【0123】

以上の説明から明らかなように、本発明は、分散ファイルシステムのためのターゲットフェイルバックおよびターゲット優先度を決定する改善されたシステムおよび方法を提供するものである。クライアントコンピュータからの要求に回答して、クライアントコンピュータに、各個別ターゲットに関連付けられたサイトコストの評価に基づいてソートされ得るターゲットのリストが提供され得る。クライアントコンピュータが、より高コストのターゲットとの接続を確立しようと試みる前に、単に、より低コストのターゲットから始まる参照応答を論理的に構文解析するだけで接続を確立しようと試みることができるように、より低コストのターゲットが参照応答の先頭にソートされ得る。さらに、ターゲットの優先順位を使ったターゲットのソートを実施することもでき、そのため、一実施形態では、より高い優先度のターゲットも参照応答の先頭にソートされ得る。さらに、参照応答は、サイトコストとターゲット優先度両方での措置を含むようにもソートされ得る。ゆえに、同等の関連サイトコストを持つターゲットのグループは、さらに、そのグループ内で、各ターゲットの個別関連ターゲット優先度に従ってソートされ得る。任意のコンピュータシステムが、さらに、本発明のソート済み参照応答を使用し、サイトコストおよび/またはターゲット優先度に従ってソートされたターゲットのリストから新しい設定ターゲットを選択し、指定することによって、より高い優先度のターゲットにフェイルバックすることができる。以上から明らかなように、ゆえに、このシステムおよび方法は、現代のコンピューティングで必要とされる大きな利点および利益を提供するものである。

30

40

【0124】

本発明では、様々な変更および代替構成が可能であるが、図面には本発明の特定の例示的实施形態を示し、それらを本明細書で詳細に説明している。しかしながら、本発明を開示の特定の形だけに限定する意図なく、それとは逆に、本発明はすべての変更形態、代替構成、および本発明の精神および範囲内に含まれる均等物を包含するものであることを理解すべきである。

【図面の簡単な説明】

50

【 0 1 2 5 】

【図 1】本発明が組み込まれ得るコンピュータシステムを一般的に表すブロック図である。

【図 2】本発明の一態様による、一実施形態においてターゲットフェイルバックおよびターゲット優先度を実施するための分散コンピューティング環境の例示的アーキテクチャを一般的に表すブロック図である。

【図 3】本発明の一態様による、分散ファイルシステムを使ってファイルを要求し、検索するために行われるステップを一般的に表す流れ図である。

【図 4】本発明の一態様による、サイトコストに基づいて参照応答を組み立てるために一実施形態で行われるステップを一般的に表す流れ図である。

【図 5】本発明の一態様による、ターゲット優先度に基づいて参照応答を組み立てるために一実施形態で行われるステップを一般的に表す流れ図である。

【図 6】本発明の一態様による、ソート済み参照応答と共に使用され得る、より低いサイトコストのターゲットにフェイルバックするために一実施形態で行われるステップを一般的に表す流れ図である。

【図 7】本発明の一態様による、ソート済み参照応答と共に使用され得る、より高い優先度のターゲットにフェイルバックするために一実施形態で行われるステップを一般的に表す流れ図である。

【図 8】本発明の一態様による、参照応答をソートするために一実施形態で行われるステップを一般的に表す流れ図である。

【符号の説明】

【 0 1 2 6 】

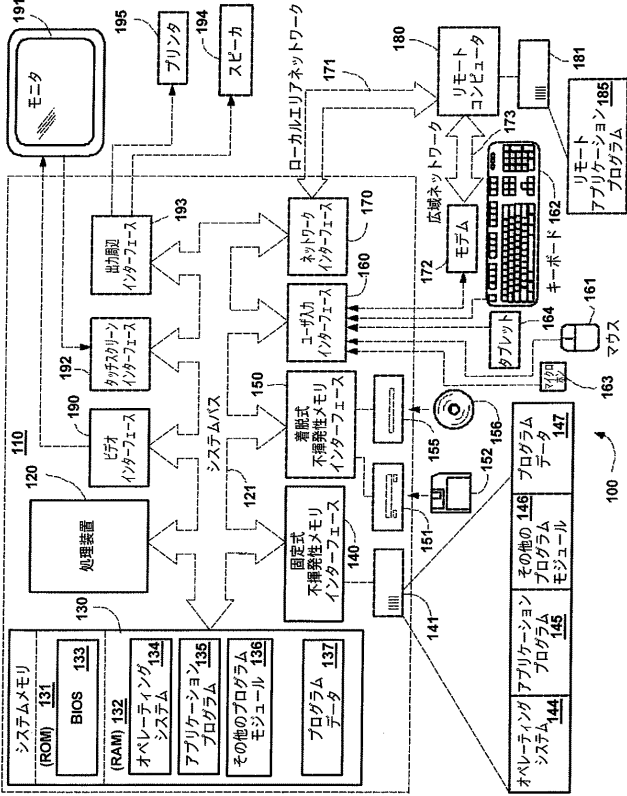
2 0 5 クライアントコンピュータ
 2 1 0 サーバコンピュータ m 1
 2 1 1 ファイルサーバ
 2 1 2 m 1 共有資源 s 1
 2 1 5 サーバコンピュータ m 2
 2 1 6 ファイルサーバ
 2 1 7 m 2 共有資源 s 2
 2 2 0 アクティブディレクトリサーバ
 2 3 0 D f s ルートサーバ
 2 4 0 D f s ルート共有資源
 2 5 0 ネットワーク
 2 6 0 ルータ
 2 7 0 D f s クライアント
 2 8 0 ネットワーク

10

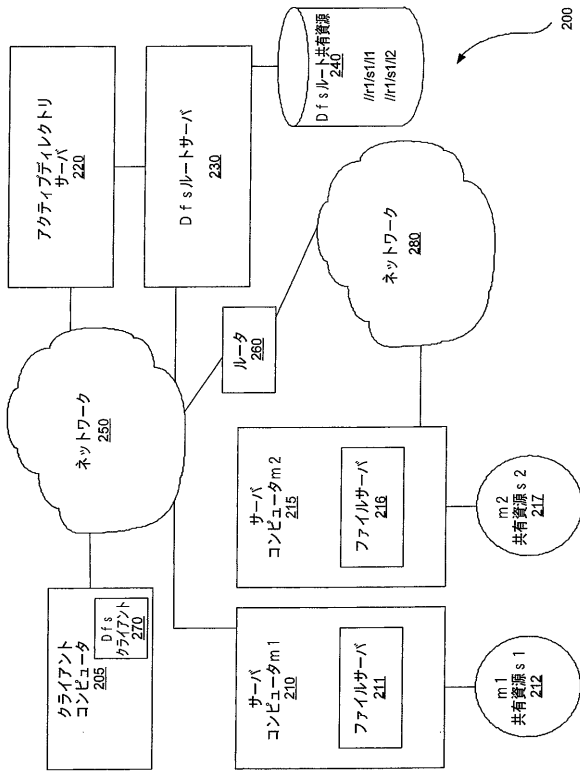
20

30

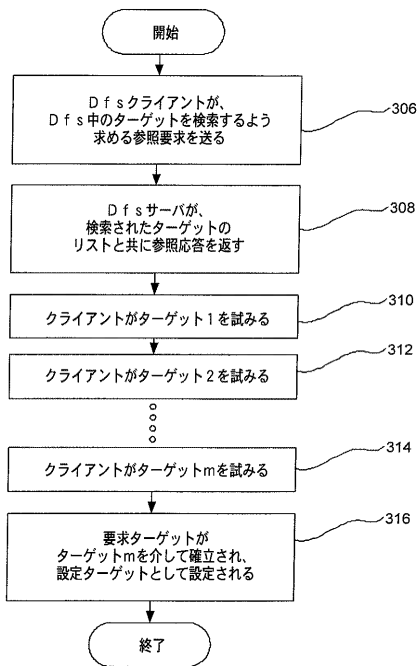
【図1】



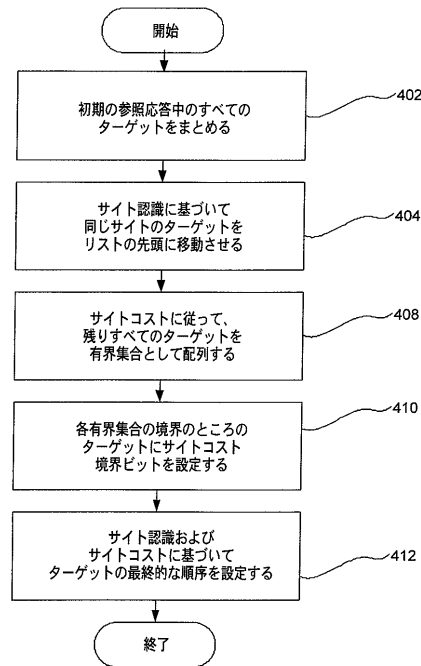
【図2】



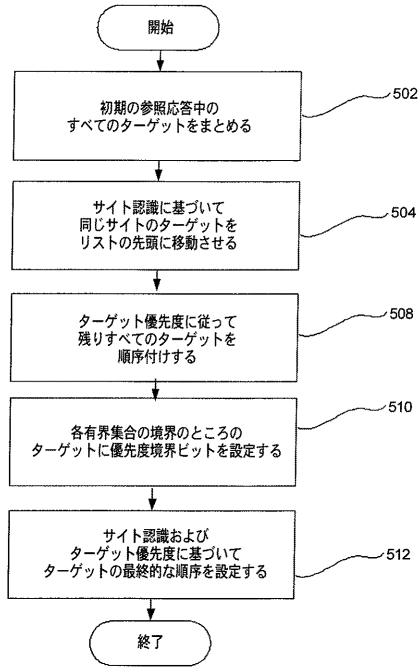
【図3】



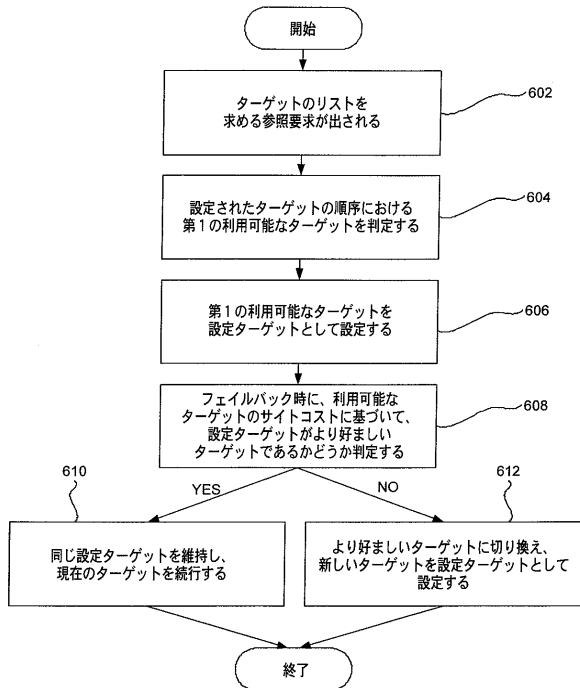
【図4】



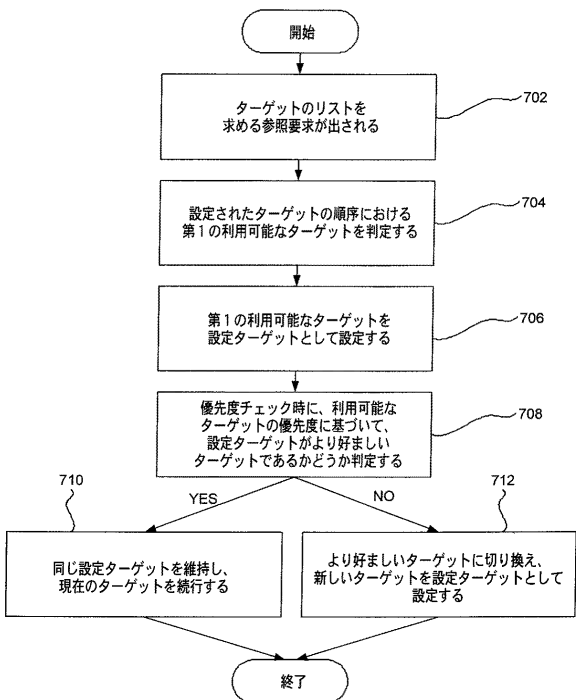
【 図 5 】



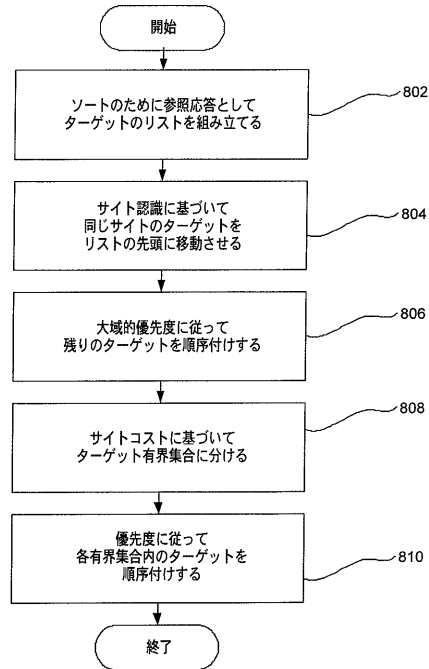
【 図 6 】



【 図 7 】



【 図 8 】



フロントページの続き

- (72)発明者 ムクール グプタ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 パトリック イー. ボーズマン
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 ラメッシュ シャンカー
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 ラビスンカー ブディペディー
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 スコット コールビル
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 スープリヤ ウィクラマチラケ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内

Fターム(参考) 5B082 HA08