

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4474286号  
(P4474286)

(45) 発行日 平成22年6月2日(2010.6.2)

(24) 登録日 平成22年3月12日(2010.3.12)

(51) Int. Cl. F I  
**GO6F 13/10 (2006.01)** GO6F 13/10 340A  
**HO4L 12/56 (2006.01)** HO4L 12/56 200E

請求項の数 10 (全 20 頁)

(21) 出願番号	特願2004-563465 (P2004-563465)	(73) 特許権者	390009531
(86) (22) 出願日	平成15年12月12日(2003.12.12)		インターナショナル・ビジネス・マシーンズ・コーポレーション
(65) 公表番号	特表2006-512662 (P2006-512662A)		INTERNATIONAL BUSINESS MACHINES CORPORATION
(43) 公表日	平成18年4月13日(2006.4.13)		アメリカ合衆国10504 ニューヨーク州 アーモンク ニュー オーチャードロード
(86) 国際出願番号	PCT/IB2003/006061	(74) 代理人	100108501
(87) 国際公開番号	W02004/059943		弁理士 上野 剛史
(87) 国際公開日	平成16年7月15日(2004.7.15)	(74) 代理人	100112690
審査請求日	平成18年12月8日(2006.12.8)		弁理士 太佐 種一
(31) 優先権主張番号	10/335, 323	(74) 代理人	100091568
(32) 優先日	平成14年12月31日(2002.12.31)		弁理士 市位 嘉宏
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 i S C S I のためのサービスの質

(57) 【特許請求の範囲】

【請求項1】

i S C S I 環境のためのサービスの質 ( Q o S ) を与えるように適合したシステムであって、

1 つ以上のイニシエータであって、各前記イニシエータが、  
フロー仕様パラメータを用いて1 つ以上の P D U をシェーピングするイニシエータ Q o S シェーパを含む i S C S I イニシエータと、

前記 i S C S I イニシエータから前記 シェーピングされた P D U を受け取り、 Q o S ポリシーを用いて前記 P D U をマーキングする Q o S マーカを含む T C P レイヤと、  
 を含むイニシエータと、

I P ネットワークを介して前記イニシエータから前記 P D U を受信する 1 つ以上のターゲットであって、各前記ターゲットが、

I / O リソースを処理するための i S C S I ターゲットにおける第 1 の Q o S コンポーネントであって、前記 P D U から抽出した 1 つ以上の読み取りコマンドおよび 1 つ以上の書き込みコマンドを、フロー仕様パラメータを用いてシェーピングする第 1 の Q o S コンポーネントと、

ネットワーク・リソースを処理するための T C P レイヤにおける第 2 の Q o S コンポーネントと、

を含む、ターゲットとを含み、

前記 P D U はコマンド P D U であり、

前記イニシエータ Q o S シェーパは、コマンド・シーケンス番号が割り当てられる前に前記コマンド P D U をシェーピングし、

前記フロー仕様パラメータは、I / O レート、バイト・レート、および I / O フロー過剰の 1 つ以上を含み、

前記ターゲットの前記第 1 の Q o s コンポーネントは、前記フロー仕様パラメータを用いて前記コマンド P D U をシェーピングする、

システム。

【請求項 2】

前記イニシエータおよびターゲットと通信を行うサービス・レベル契約 ( S L A ) エージェントであって、前記イニシエータおよび前記ターゲットに前記 Q o S ポリシーを提供する S L A エージェントを更に含む、請求項 1 に記載のシステム。

10

【請求項 3】

前記イニシエータおよび前記ターゲットが、i S C S I セッションごとに多数のサービス・クラスのための多数の別個の接続をサポートする、請求項 1 に記載のシステム。

【請求項 4】

複数のフィルタ仕様属性を含み、前記フィルタ仕様属性の 1 つ以上が i S C S I トラヒック・タイプを反映するように構成されている、請求項 1 に記載のシステム。

【請求項 5】

前記フィルタ仕様属性が、T C P 4 タプル、i S C S I セッション i d、L U N ( 論理ユニット番号)、イニシエータ・ユーザ i d、およびアプリケーション識別属性のうち 1 つ以上を含む、請求項 4 に記載のシステム。

20

【請求項 6】

前記フィルタ仕様属性が、ネットワーク・リソースおよび I / O リソースに Q o S を与えるように適合している、請求項 4 に記載のシステム。

【請求項 7】

前記フィルタ仕様属性が、1 つ以上のイニシエータおよび 1 つ以上のターゲットによって用いられ、前記イニシエータによって用いられる前記フィルタ仕様属性が、前記ターゲットによって用いられる前記フィルタ仕様属性とは異なる、請求項 4 に記載のシステム。

【請求項 8】

前記 フロー仕様パラメータが、ネットワーク・リソースおよび I / O リソースに Q o S を与えるように適合している、請求項 1 に記載のシステム。

30

【請求項 9】

前記イニシエータによって用いられる前記 フロー仕様パラメータが、前記ターゲットによって用いられる前記 フロー仕様パラメータとは異なる、請求項 1 に記載のシステム。

【請求項 10】

i S C S I 環境のために Q o S を与えるための方法であって、

i S C S I イニシエータにおいて、コマンド・シーケンス番号が割り当てられる前に、フロー仕様パラメータを用いて 1 つ以上のコマンド P D U をシェーピングするステップと

、

前記 i S C S I イニシエータを含むイニシエータにおける T C P レイヤにおいて、Q o S ポリシーを用いて、前記シェーピングしたコマンド P D U をマーキングするステップと

40

、

I P ネットワークを介して、i S C S I ターゲットにおいて前記コマンド P D U を受信するステップと、

前記 i S C S I ターゲットにおいて、フロー仕様パラメータを用いて、前記コマンド P D U から抽出した 1 つ以上の読み取りコマンドおよび 1 つ以上の書き込みコマンドをシェーピングするステップと

を含み、

前記フロー仕様パラメータは、I / O レート、バイト・レート、および I / O フロー過剰の 1 つ以上を含む、

50

方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般にサービスの質に関し、より具体的には、ネットワーク・ストレージ・システムにおけるサービスの質に関する。

【背景技術】

【0002】

サービスの質(QoS:Quality ofService)は、広義には、選択されたネットワーク・トラフィックに、より良いサービスを提供することができるネットワークの能力を指す。QoSは、可用性、誤り性能、応答時間、ネットワーク輻輳によるトラフィック喪失または伝送不良等を含む、いくつかの基本的な性能規準によって特徴付けることができる。

10

【発明の開示】

【発明が解決しようとする課題】

【0003】

ネットワークのためのQoSは、かなり開発された分野であるが、ネットワーク・ストレージ・システムでは、従来技術のQoSプロトコルおよび技法によって対処されない特別な問題が生じる。1例として、インターネット・プロトコル(IP)ネットワークにおいて、重いトラフィックのほとんどは、サーバからクライアントへのものである。通常、トラフィックは、サーバからクライアントへのダウンロードの形態とすることができる。従って、従来技術のQoS技法は、サーバからクライアントへの一方向トラフィックに対処するように設計されている。

20

【0004】

しかしながら、ネットワーク・ストレージでは、重いトラフィックは、読み取り動作または書き込み動作のいずれかが等、双方向である場合がある。読み取り動作は、ストレージ・サーバからクライアントへの重いトラフィックを生成することがある。これに対して、書き込み動作は、クライアントからストレージ・サーバへの重いトラフィックを生成する場合がある。従来技術のQoS技法では、双方向トラフィックに対処するような設備が整っていない。

【0005】

30

別のネットワーク・ストレージの問題は、不従順(non-compliant)クライアントである。IPネットワーク・トラフィックでは、クライアントからのトラフィックのほとんどは、サーバからのトラフィックに比べて、ごくわずかである。従って、サービス・レベル契約(SLA:service level agreements)に従わないクライアントによって引き起こされる損害は、通常、取るに足りないものである。このため、従来技術のQoS技法は、不従順クライアントの問題に取り組まない。これに対して、ネットワーク・ストレージ・システムでは、不従順クライアントは、データの読み取りまたは書き込み等、多数の重いトラフィック動作を開始する可能性がある。従って、ネットワーク・ストレージのためのQoSでは、不従順クライアントのトラフィックを追跡し、これを含むことが必要である。現在、この必要性は、従来技術のQoS技法では満たされていない。

40

【0006】

上述のことは、単に、従来技術のQoS技法がネットワーク・ストレージ・システムのQoS問題に対処できないことの2つの例である。従って、ネットワーク・ストレージ・システムの独特かつ特定の問題に対処することができるQoS技法を提供することが要望されている。

【課題を解決するための手段】

【0007】

iSCSI(Internet SmallComputer System Interface、アイスカジー)プロトコルによって、伝送制御手順(TCP:Transmission Control Protocol)上にSCSI(Small Computer System Interface、スカジー)プロトコルをトランスポートすることが可能

50

となる。i S C S Iトラヒックは、通常、別個のネットワーク・ノード上に常駐するイニシエータとターゲットとの間の双方向性トラヒックである。i S C S Iトラヒックは、共有ネットワークおよび入力/出力(I/O)リソースを利用し、従って、輻輳の影響を受ける。

**【0008】**

本発明の1つの目的は、i S C S Iトラヒックを他のネットワーク・トラヒックと区別するためのQoS技法およびプロトコルを提供することである。同様に、更に別の目的は、例えば、異なるイニシエータからのトラヒックおよび異なる論理ユニット(LU: logical unit)へのトラヒックのように、i S C S Iトラヒックの様々なインスタンスを区別し優先順位を付けるQoS技法およびプロトコルを提供することである。

10

**【0009】**

本発明の好適な実施形態は、i S C S Iのためのエンド・ツー・エンド(end-to-end)QoSを提供することができる。本発明のQoS技法およびプロトコルは、イニシエータからターゲットへ、およびその逆のフロー経路に沿って、多数のリソースの制御および管理を行うことができる。

**【0010】**

本発明のいくつかの実施形態では、i S C S Iトラヒックを、ネットワーク帯域のある部分に限定し、これによって、他のトラヒックの十分なサービスの質を可能とすることができる。

**【0011】**

20

本発明の1つの態様によれば、i S C S I環境のためのサービスの質を得るように構成されたイニシエータが提供される。イニシエータは、i S C S IイニシエータおよびTCPレイヤ(TCP layer)を含む。i S C S Iイニシエータは、QoSポリシーに従って1つ以上のPDUをシェーピングする(shaping)ためのQoSシェーパ(shaper)を含む。TCPレイヤは、i S C S IイニシエータからPDUを受け取り、QoSポリシーに従ってPDUにマーキングする(marking)ためのQoSマーカ(marker)を含む。

**【0012】**

好ましくは、QoSポリシーは、ネットワーク・リソースのためのポリシーおよびI/Oリソースのためのポリシーを含む。

**【0013】**

30

代替的な実施形態では、PDUはコマンドPDUである。かかる場合、コマンドPDUがコマンド・シーケンス番号を受け取る前に、シェーパがそれらをシェーピングする。

**【0014】**

他の代替的な実施形態では、PDUはデータPDUおよびコマンドPDUである。データPDUおよびコマンドPDUは、TCPレイヤに受け取られる前にシェーピングされる。

**【0015】**

また、本発明は、TCPレイヤにおいて第2のシェーパも含むことができる。第2のシェーパは、i S C S I以外のアプリケーションのネットワーク・トラヒックをシェーピングする。

40

**【0016】**

本発明の別の態様によれば、i S C S I環境のためのサービスの質を与えるように構成したターゲットが提供される。ターゲットは、第1および第2のQoSコンポーネントを含む。

**【0017】**

第1のQoSコンポーネントは、i S C S Iターゲット内にあり、I/Oリソースを処理する。第1のQoSコンポーネントは、QoSポリシーに従って読み取りコマンドおよび書き込みコマンドをシェーピングする。このため、第1のQoSコンポーネントは、不従順イニシエータからのトラヒックにQoSを与える。

**【0018】**

50

第2のQoSコンポーネントは、TCPレイヤ内にあり、ネットワーク・リソースを処理する。iSCSI PDUは、iSCSIのQoSポリシーに従ってシェーピングおよびマーキングされる。

【0019】

あるいは、ターゲットは、複数の論理ユニット(LU)に対するアクセスを提供する。LUの各々は多数のタスク・セットを有し、タスク・セットの各々は関連するサービス・クラスを有する。いくつかの実施形態では、関連するサービス・クラスの2つ以上は同一である。代替的な実施形態では、関連するサービス・クラスの2つ以上は互いに異なっている。

【0020】

本発明の別の態様によれば、iSCSI環境のためのサービスの質(QoS)を与えるように構成したシステムが提供される。このシステムは、1つ以上のイニシエータおよび1つ以上のターゲットをサービス。いくつかの実施形態では、システムは、サービス・レベル契約(SLA)エージェントも含む。エージェントは、イニシエータおよびターゲットと通信し、イニシエータおよびターゲットにQoSポリシーを提供する。

【0021】

代替的な実施形態では、イニシエータが用いるQoSポリシーは、ターゲットにおいて用いられるQoSポリシーとは異なる。更に、イニシエータおよびターゲットは、iSCSIセッションごとに多数のサービス・クラスのための多数の別個の接続をサポートすることができる。

【0022】

本発明の更に別の態様によれば、iSCSI環境のために構成したQoSシステムが提供される。QoSシステムは、複数のフィルタ仕様属性を含む。フィルタ仕様属性の1つ以上は、iSCSIトラフィック・タイプを反映するように構成されている。

【0023】

いくつかの実施形態では、フィルタ仕様属性は、TCP 4タプル(four-tuple)、iSCSIセッションid、LUN、イニシエータ・ユーザid、もしくはアプリケーション識別属性またはそれら全てを含む。あるいは、フィルタ仕様属性は、ネットワーク・リソースおよびI/OリソースにQoSを与えるように構成されている。

【0024】

通常、フィルタ仕様属性は、1つ以上のイニシエータおよび1つ以上のターゲットによって用いられる。場合によっては、イニシエータによって用いられるフィルタ仕様属性は、ターゲットによって用いられるフィルタ仕様属性とは異なる。

【0025】

本発明の更に別の態様によれば、iSCSI環境のために構成した別のQoSシステムが提供される。QoSシステムは、複数のフィルタ仕様パラメータを含む。フィルタ仕様パラメータの1つ以上は、iSCSIトラフィック・タイプのために構成されている。

【0026】

いくつかの実施形態では、フロー仕様パラメータは、I/Oレート、バイト・レート、もしくはI/Oフロー過剰またはそれら全てを含む。あるいは、フロー仕様パラメータは、ネットワーク・リソースおよびI/OリソースにQoSを与えるように構成されている。

【0027】

通常、フロー仕様パラメータは、1つ以上のイニシエータおよび1つ以上のターゲットによって用いられる。場合によっては、イニシエータによって用いられるフロー仕様パラメータは、ターゲットによって用いられるフィルタ仕様パラメータとは異なる。

【0028】

本発明の別の態様によれば、iSCSI環境のために構成した別のQoSシステムが提供される。QoSシステムは、シェーパ、マーカ、および第1のターゲットQoSコンポーネントを含む。

10

20

30

40

50

## 【0029】

シェーパは、iSCSIイニシエータ内にあり、QoSポリシーに従って1つ以上のPDUをシェーピングする。マーカは、イニシエータにおけるTCPLレイヤ内にあり、iSCSIイニシエータからPDUを受け取る。また、マーカは、QoSポリシーに従ってPDUをマーキングする。第1のターゲットQoSコンポーネントは、iSCSIターゲット内にある。第1のターゲットQoSコンポーネントは、PDUから抽出した1つ以上のコマンドを受信し、QoSポリシーに従ってコマンドをシェーピングする。

## 【0030】

いくつかの実施形態では、第1のターゲットQoSコンポーネントは、I/Oリソースを処理する。あるいは、システムは、TCPLレイヤにおいて第2のターゲットQoSコンポーネントを含む。第2のターゲットQoSコンポーネントは、ネットワーク・リソースを処理する。iSCSI PDUは、iSCSIのためのQoSポリシーに従ってシェーピングおよびマーキングされる。

10

## 【0031】

本発明の別の態様によれば、iSCSI環境のためにQoSを与えるための方法が提供される。この方法は、QoSポリシーに従ってiSCSIイニシエータにおいて1つ以上のPDUをシェーピングすることを含む。また、この方法は、QoSポリシーに従ってイニシエータにおけるTCPLレイヤにおいてPDUをマーキングすることを含む。PDUがコマンドPDUである場合、この方法は、コマンドPDUにコマンド・シーケンス番号を割り当てる前に、それらをシェーピングすることを含む。コマンドPDUのシェーピングは、通常、I/OリソースのためのQoSポリシーに従って行われる。

20

## 【0032】

いくつかの実施形態では、コマンドPDUは、即値データを含む書き込みコマンドPDUである。書き込みコマンドPDUは、ネットワーク・リソースのためのQoSポリシーに従って、更に、I/OリソースのためのQoSポリシーに従ってシェーピングされる。

## 【0033】

代替的な実施形態では、PDUはデータPDUおよびコマンドPDUである。データPDUおよびコマンドPDUをTCPLレイヤに送る前に、iSCSIイニシエータにおいてデータPDUおよびコマンドPDUのシェーピングが行われる。PDUがデータPDUである場合、データPDUは、ネットワーク・リソースのためのQoSポリシーに従ってシェーピングされる。

30

## 【0034】

あるいは、この方法は、IPネットワークを介して、ターゲットにPDUを送信することを含む。iSCSIターゲットにおいて、I/OリソースのためのQoSポリシーに従って、PDUから読み取りコマンドおよび書き込みコマンドを抽出し、シェーピングし、これによって不従順トラフィックにQoSを与える。

## 【0035】

本発明の別の態様によれば、iSCSI環境のためにQoSを与えるための更に別の方法が提供される。この方法は、iSCSIターゲットにおいて1つ以上のLUからデータを受信することと、iSCSIのためのQoSポリシーに従って、QoSコンポーネントにおいてデータをシェーピングおよびマーキングすることとを含む。QoSポリシーは、ネットワーク・リソースのためのポリシーである。

40

## 【0036】

本発明の別の態様によれば、iSCSI環境においてサービスの質を与えるためのシステムが提供される。このシステムは、QoSポリシーに従ってiSCSIイニシエータにおいて1つ以上のPDUをシェーピングするための手段を含む。また、システムは、QoSポリシーに従ってイニシエータにおけるTCPLレイヤにおいてシェーピングしたPDUをマーキングするための手段を含む。

## 【0037】

本発明の別の態様によれば、コンピュータ読み取り可能媒体ソフトウェア上で具現化さ

50

れるコンピュータ・プログラムが提供される。このコンピュータ・プログラムは、QoSポリシーに従ってiSCSIイニシエータにおいて1つ以上のPDUをシェーピングするように動作する第1のセグメントを含む。第2のセグメントは、QoSポリシーに従ってシェーピングしたPDUをイニシエータにおけるTCPレイヤにおいてマーキングするように動作する。

【0038】

代替的な実施形態では、プログラムは、IPネットワークを介して、ターゲットにおいてPDUを受信するように動作する第3のセグメントを含む。第4のセグメントは、iSCSIターゲットにおいて、QoSポリシーに従って、1つ以上の読み取りコマンドおよび1つ以上の書き込みコマンドをシェーピングし、これによって不従順トラヒックにQoSを与えるように動作する。

10

【0039】

これより、本発明の実施形態について、添付図面を参照して、一例として説明する。

【発明を実施するための最良の形態】

【0040】

用語

・サービスのクラス(CoS:Class ofService)

所与のレベルのサービス優先順位を有するクラスに同様のタイプのトラヒックをグループ化することによって、ネットワーク内で、トラヒックを管理する方法。クラスは、トラヒック・ドメインに関連する1組の属性によって決定する(フィルタ仕様を参照のこと)。

20

【0041】

・コマンド記述子ブロック(CDB:CommandDescriptor Block)

SCSIコマンドおよびその属性を含むデータ構造。CDBは、iSCSIイニシエータからiSCSIターゲットに、コマンドPDU内で送信される。CDBは、SCSIアーキテクチャにおいて定義される。

【0042】

・コマンドPDU

SCSIコマンドの詳細、具体的にはCDBを含むiSCSIプロトコルでのPDU。また、コマンドPDUは、即値データ・アウトを含む場合がある。

30

【0043】

・コマンド・シーケンス番号

iSCSIプロトコルによってSCSIコマンドに関連付けられた数値。シーケンス番号は、iSCSIイニシエータによって割り当てられ、コマンドPDUのフィールド内で搬送される。コマンド・シーケンス番号は、セッション全体を通して用いられ、セッションの全接続を通して、ターゲットにおけるコマンドの順次供給のために用いられる。iSCSIターゲットは、コマンド・シーケンス番号の順序で、コマンドをSCSI実行エンジンに供給しなければならない。

【0044】

・接続

イニシエータ・ノードとターゲット・ノードとの間のTCP接続。各接続は、1つのセッションに属する。1セッション内に多数の接続がある場合がある。各接続は、iSCSIプロトコルにおいて一意のidを有する。また、接続は、ローカルTCPポート番号によって各端末で識別することができる。各接続は、単一のサービス・クラスに関連付けられている。多数の接続を用いて、1セッション内で多数のサービス・クラスをサポートすることができる。

40

【0045】

・接続忠誠(ConnectionAllegiance)

双方向で、同一の接続を介して、個々のSCSIコマンドの全てのPDUを送信するためのiSCSIの要件。これは、コマンドPDU、データPDU、ならびに応答および制

50

御 P D U を含む。

【 0 0 4 6 】

・データ・イン (Data In)

S C S I 読み取りコマンドのプロトコルの一部として、ターゲットからイニシエータに送信されるデータ。

【 0 0 4 7 】

・データ・アウト (Data-Out)

S C S I 書き込みコマンドのプロトコルの一部として、イニシエータからターゲットに送信されるデータ。データ・アウトは、ターゲットによって請求されるか、または請求がなくても与えられるものとする事ができる。

10

【 0 0 4 8 】

・データ P D U

データ・インまたはデータ・アウトを含む i S C S I プロトコルにおける P D U。

【 0 0 4 9 】

・区分化サービス (D i f f S e r v)

トラフィック分類に基づいて、インターネットにおいてスケラブル・サービス区分化を実施するための標準規格 (R F C 2 4 7 4)。D i f f S e r v は、インターネット・エンジニアリング・タスク・フォース (I E T F) のネットワーキング・ワーキング・グループによって定義される。

20

【 0 0 5 0 】

・フィルタ仕様

1 組の属性であり、トラフィックのサービス・クラスを決定するために用いる。フィルタ仕様属性の例には、T C P 4 タプル、L U N、セッション i d 等が含まれる場合がある。

【 0 0 5 1 】

・フロー

D i f f S e r v e アーキテクチャによって区別可能なトラフィック・ストリームの最も小さい粒度 (granularity)。

【 0 0 5 2 】

・フロー仕様

フローを記述する際に用いる 1 組のパラメータ。フロー仕様パラメータの例は、帯域幅要件、許容可能な最大遅延、損失率等を含む場合がある。各サービス・クラスは、それ自身のフロー仕様を有する。

30

【 0 0 5 3 】

・ジェネリック S C S I 実行エンジン

確定した I / O リソースに関連付けられていない S C S I 実行エンジン。実行エンジンは、ローカル S C S I L U にアクセスするか、または、ストレージ・エリア・ネットワーク (S A N : storage area network) 等の別のネットワークへのゲートウェイとして機能することができる。

【 0 0 5 4 】

・即値データ (Immediate Data)

i S C S I 書き込みコマンド P D U の一部として送信されるデータ・アウト。これは、請求がなくても与えられるデータの第 1 の (または唯一の) 部分を構成する。即値データは、イニシエータとターゲットとの間で交渉可能な最大サイズを有する。

40

【 0 0 5 5 】

・イニシエータ \*

i S C S I プロトコルにおけるクライアント。これは、ターゲットとして知られるサーバに要求を送信して、論理ユニットに保存されたデータにアクセスする。また、この言葉は、イニシエータの機能を実施するシステム・コンポーネント (ドライバ) を識別するために用いられる。また、この言葉は、S C S I プロトコルにおけるクライアント、および、この機能を実施するシステム・コンポーネント (ドライバ) を指す。

50

## 【 0 0 5 6 】

## ・イニシエータ・ノード\*

SCSIまたはiSCSIイニシエータあるいはその両方の機能を具現化するネットワーク・ノード。

\*本明細書中では、「イニシエータ」という言葉は、文脈によって、総称的に、iSCSIイニシエータ、SCSIイニシエータ、またはイニシエータ・ノードを指すことを注記しておく。

## 【 0 0 5 7 】

## ・iSCSI

インターネットSCSI。SCSIプロトコルのTCPへのトランスポートのマッピング。iSCSIでは、イニシエータおよびターゲットは、別個のネットワーク・ノード上に常駐し、TCPを介して送信されるPDUによって通信を行う。iSCSIは、インターネット・エンジニアリング・タスク・フォース(IETF)のインターネット・プロトコル・ストレージ(IPS)ワーキング・グループによって定義された標準規格である。

10

## 【 0 0 5 8 】

## ・I-T-L連結

特定のイニシエータ、特定のターゲット、および特定の論理ユニット間の関係であり、SCSIコマンドが供給される文脈を規定する。この言葉は、大まかにタスク・セットに関連している。

20

## 【 0 0 5 9 】

## ・論理ユニット(LU)

ターゲットにおいてアクセス可能なデータ・デバイスであり、これにおいてSCSIコマンドを実行する。各ターゲットは、多数の論理ユニットにアクセスすることができる。各論理ユニットは一意に識別可能である。

## 【 0 0 6 0 】

## ・論理ユニット番号(LUN)

ターゲットにおいて各論理ユニットを識別する一意の番号。また、この言葉は、特定の論理ユニットを指すために用いられる。

## 【 0 0 6 1 】

## ・マーキング

規定されたポリシーに基づいて、パケットのIPヘッダ内のTOSバイトを設定するプロセス。

30

## 【 0 0 6 2 】

## ・不従順イニシエータ

不従順トラヒックを発生させるiSCSIイニシエータ。

## 【 0 0 6 3 】

## ・不従順トラヒック

特定のサービス・クラスのフロー仕様に従わないトラヒック。

## 【 0 0 6 4 】

## ・ポリシー

ネットワーク内等で、どのようにリソースをカスタマ間で共有するかを規定する、公式の1組のステートメント。

40

## 【 0 0 6 5 】

## ・プロトコル・データ・ユニット(PDU)

iSCSIプロトコルの一部として、イニシエータからターゲットに、またはその逆に送信されるメッセージ。PDUは、コマンド、応答、制御情報、およびデータを含む。

## 【 0 0 6 6 】

## ・SCSI

スモール・コンピュータ・システム・インタフェース。I/Oデバイス、特にストレージ・デバイスと通信を行うプロトコル・ファミリー。SCSIは、情報技術規格委員会(N

50

C I T S : National Committee for Information Technology Standards) の T 1 0 技術委員会によって定義された標準規格である。これによって提供されるクライアント / サーバ・アーキテクチャでは、クライアント (イニシエータと呼ぶ) がサーバ (ターゲットと呼ぶ) にコマンドを送信して、データの読み取りおよび書き込み等の I / O サービスを実行する。

【 0 0 6 7 】

・ S C S I 実行エンジン

ターゲット・ノードにおけるコンポーネント。S C S I 実行エンジンは、i S C S I ターゲットから、コマンド P D U から抽出された S C S I コマンドを受信し、コマンドを実行のため渡す責任がある。

10

【 0 0 6 8 】

・サービス・レベル契約 ( S L A )

サービス・プロバイダとカスタマとの間で、通常は測定可能な条件で、サービス・プロバイダがカスタマにどのサービスを与えるかを指定する取り決め。

【 0 0 6 9 】

・サービス・レベル契約 ( S L A ) サーバ

S L A を維持し、それらをポリシーに翻訳することができるネットワーク上のノード。S L A サーバは、ポリシーを他のネットワーク・ノードに広める責任がある。

【 0 0 7 0 】

・セッション

i S C S I プロトコルにおける特定のイニシエータと特定のターゲットとの間の関連。1 つのセッションは、1 つ以上の接続から成る。セッションは、イニシエータに、ターゲットにおいてアクセス可能な L U 集合の不変のビューを提供する。S C S I コマンドは、ターゲットにおいて、セッション全体に渡って順次供給され、コマンド・シーケンス番号によって実施される。各セッションは 1 組の属性を有し、これはプロトコルを支配し、イニシエータとターゲットとの間で交渉可能である。各セッションは、セッションの各側で一意的識別子によって識別される。各イニシエータとターゲットとの間で、多数のセッションがある場合がある。

20

【 0 0 7 1 】

・シェーピング

サービス・クラス内でトラヒックを遅延させて、定義したフロー仕様に従わせるプロセス。

30

【 0 0 7 2 】

・請求データ

ターゲットからの請求に回答して送信されるデータ・アウト。請求は、ターゲットからの ( R 2 T ) P D U を送信するための要求の形態である。所与の書き込みコマンドのために多数の R 2 T が存在し、その各々が異なるデータ部分を請求する場合がある。R 2 T の数および各データ部分のサイズは、ターゲットによって決定される。

【 0 0 7 3 】

・ターゲット \* \*

i S C S I プロトコルにおけるサーバ。これは、L U 集合に対するアクセスを与える。ターゲットは、クライアント (イニシエータと呼ぶ) から要求を受信する。また、この言葉は、ターゲットの機能を実施するシステム・コンポーネント (ドライバ) を識別するために用いられる。また、この言葉は、S C S I プロトコルにおけるサーバおよびこの機能を実施するシステム・コンポーネント (ドライバ) を指す。

40

【 0 0 7 4 】

・ターゲット・ノード \* \*

S C S I または i S C S I ターゲットあるいはその両方の機能を具現化するネットワーク・ノード。

\* \* 本明細書中では、「ターゲット」という言葉は、文脈によって、総称的に、i S C

50

S Iターゲット、S C S Iターゲット、またはターゲット・ノードを指すことを注記しておく。

【0075】

・タスク

インシエータから発するi S C S I要求であり、ターゲットからの応答を待つ。また、この言葉は、S C S Iアーキテクチャにおいて定義されたように、進行中のS C S IコマンドまたはS C S Iコマンドのリンク・リストを指す。

【0076】

・タスク・セット

同一の論理ユニットにアクセスするタスク集合。ターゲットにおける構成に応じて、各L Uごとに、全インシエータからのタスクを含む単一タスク・セット、または、インシエータごとに別個のタスク・セットが存在する。後者は、I - T - L連結に対応する。

10

【0077】

・サービス・タイプ(T O S)バイト

D I f f S e r vに一致して解釈された場合、I Pバージョンの4ヘッダ「サービス・タイプ」オクテット・フィールド、またはI Pバージョンの6「トラヒック・クラス」オクテット・フィールド。

【0078】

・請求なしで与えられるデータ

ターゲットからの請求を待たずに送信されるデータ・アウト。請求なしで与えられるデータは、交渉可能な最大サイズを有し、1つ以上のデータP D U内で送信される。請求なしで与えられるデータの第1の部分は即値データである。

20

【0079】

本発明の説明

ここで図1を参照すると、i S C S I環境等のネットワーク・ストレージ環境のためのサービスの質(Q o S)アーキテクチャおよびシステム10が示されている。Q o S 10は、i S C S Iトラヒック用のエンド・ツー・エンドのサービスの質のための技法およびプロトコルを提供することができる。

【0080】

説明の明確さのため、i S C S Iによって、T C P上へのS C S Iプロトコルのトランスポートを可能とすることを注記しておく。本発明は、i S C S I環境内で説明するが、本発明は、ネットワーク・ストレージ機能を提供する他の技法またはプロトコルに適用可能であることは、当業者には明らかである。

30

【0081】

本発明の好適な実施形態は、i S C S I環境において用いるために、既存のネットワークQ o S技法を適合させるが、大きな方法上の相違がある。従来技術のQ o Sシステムによって対処されていないがQ o Sシステム10によって対処される大きな問題のいくつかは、以下のものである。

【0082】

1) i S C S Iのためのエンド・ツー・エンドQ o Sには、ネットワーク・リソースおよびI / Oリソースを含む多数のタイプのリソースが関与する。I / Oリソースは、論理ユニットとすることができる。従来技術のQ o S技法は、典型的にネットワーク・リソースである1つのみのタイプのリソースに対処した。

40

【0083】

2) I Pネットワークを介したS C S Iトラヒックのトランスポートでは、ネットワークのためのQ o Sに、以下のような固有の課題が生じる。

a) 厳密な時間感度。i S C S Iトラヒックは通常、S C S I性能要件のため、他のI Pネットワーク・トラヒックよりも応答時間に敏感である。ネットワークを介したストレージ・トラヒックのトランスポートでは、余分な遅延および潜在的なデータ損失が加わる場合があり、これらは双方とも応答時間に影響を与える。

50

b) 長時間続く通信セッション。典型的なネットワーク接続の使用は短時間である。これに対して、i S C S Iセッションの接続は、長時間持続することを意図している。

c) i S C S Iトラヒックは双方向である。IPネットワークにおいて、通常、重いトラヒックのほとんどは、サーバからクライアントへの一方向である。1例として、サーバからクライアントへのダウンロードの形態である。

【0084】

従来技術のQoS技法は、2a)、b)、およびc)の項に記載した固有の問題のいずれにも対処しない。

【0085】

3) i S C S Iプロトコルの要件に対する依存性。従来技術では、ネットワーキングQoS技法およびプロトコルは、TCPLレイヤにおいて実施される。下に詳述する理由により、i S C S I要件のため、TCPLレイヤにおいてi S C S Iのための全てのQoS技法およびプロトコルを実施することはできない。

【0086】

4) i S C S Iイニシエータは、サービス・レベル契約に従わない場合がある。i S C S Iの分散型環境において、いくつかのイニシエータはQoSポリシーに従わないことがあり、または、QoSが全く実施されていない場合さえある。かかるイニシエータは、ターゲットをコマンドPDUでいっばいにし、ターゲットにおいてI/O輻輳を引き起こすことがある。ターゲットは、この状況を防ぐQoS方法を有しなければならない。従来技術では、宛先は不従順ソースを追跡しない。

【0087】

本発明の好適な実施形態は、フィルタおよびフロー仕様のためにDiffServ技法およびプロトコルを適合させる。本発明は、新たなフィルタ仕様属性および新たなフロー仕様パラメータを追加的に識別している。

【0088】

好適な実施形態において、多数のQoSコンポーネントは、各々、i S C S Iトラヒックの一部を処理することができる。このため、各QoSコンポーネントは、i S C S Iトラヒックのエンド・ツー・エンド経路に沿って異なるリソースに位置付けることができ、更に、コマンド、データ・イン、およびデータ・アウト等、特定のタイプのトラヒックを処理することができる。

【0089】

以下に詳しく述べるように、QoSコンポーネントのレイアウトは、i S C S Iに対する依存性によって影響を受ける場合がある。QoS動作のあるものはTCPLレイヤで行うことができ、QoS動作の他のものはi S C S Iコンポーネントで行うことができる。

【0090】

当業者には明らかであるように、本発明のQoS技法は、全てのトラヒックをTCPLレイヤで処理する従来技術のネットワークQoS技法とは異なる。

【0091】

再び図1を参照すると、QoSシステム10は、イニシエータ・ノード20、ターゲット・ノード30、およびSLAサーバ40を含む。QoS10における要素は、IPネットワーク46を介して通信を行う。明確さのため、図1は1つのみのイニシエータ・ノード20および1つのターゲット・ノード30を示すが、QoSは、多くのイニシエータ・ノード20または多くのターゲット・ノード30あるいはその両方を含む場合があり、その場合も本発明の原理内であることは、当業者には明らかである。SLAサーバ40は、1つ以上のイニシエータ20および1つ以上のターゲット30と通信を行う別個のサーバとすることができる。SLAサーバは、当技術分野の要素において公知であり、本明細書中においてこれ以上は論じないことを注記しておく。

【0092】

イニシエータ・ノード20は、SCSIイニシエータ・コンポーネント22、i S C S Iイニシエータ・コンポーネント24、QoSシェーピング・コンポーネント26、Qo

10

20

30

40

50

S マーキング・コンポーネント 28、および TCP レイヤ 32 を含むことができる。ターゲット・ノード 30 は、SCSI 実行エンジン・コンポーネント 42、iSCSI ターゲット・コンポーネント 44、TCP レイヤ 34、ならびに QoS コンポーネント 36 および 38 を含むことができる。

【0093】

イニシエータ・ノード 20 とターゲット・ノード 30 との間で、セッションを確立することができる。図 1 には 1 つのみのセッションを示すが、各イニシエータ・ノード 20 と各ターゲット・ノード 30 との間には多数のセッションが存在する場合があります、その場合も本発明の真の精神および原理内であることは、当業者には明らかである。

【0094】

QoS 10 は、2 つのタイプのリソースすなわちネットワークおよび I/O のための QoS ポリシーを用いる。イニシエータ 20 におけるシェーピング・コンポーネント 26 は、双方のタイプのリソースを処理することができることを注記しておく。更に、QoS コンポーネント 36 および 38 は、各々、1 つのタイプのリソースを処理することができることに留意するべきである。QoS 36 はネットワーク・リソースを処理し、QoS 38 は I/O リソースを処理する。また、イニシエータにおいて用いる QoS ポリシーは、ターゲットにおいて用いる QoS と同一のものかまたは異なるものであり得ることも注記しておく。

【0095】

いくつかの好適な実施形態では、SCSI 実行エンジン 42 は、ジェネリックなものとしてすることができる。SCSI 実行エンジン 42 が、ストレージ・エリア・ネットワーク (SAN) 等の別のネットワークへのゲートウェイである場合、SCSI 実行エンジン 42 は、他のネットワークにおける QoS の実施を継続させる責任を有することができる。

【0096】

従来技術の実施形態では、QoS シェーピングおよびマーキングが双方とも TCP レイヤにおいて行われることを注記しておく。本発明の好適な実施形態では、イニシエータ・ノードにおけるシェーピングは、TCP から iSCSI イニシエータ内へ移ることができる。シェーピングは、QoS ポリシーと iSCSI プロトコル要件との間の潜在的な競合状況を解決するため、TCP の前に実行される。これについては、以下で十分に説明する。次いで、TCP において、マーキングが実行される。

【0097】

また、ターゲット・ノードでは、かかる潜在的な競合状況は存在しないことを注記しておく。このため、シェーピングおよびマーキングは双方とも TCP において実行することができる。

【0098】

以下に列挙した図は、iSCSI トラヒック・フローを表す。

1) 図 2 は、イニシエータ 20 とターゲット 30 との間のセッションを示す。図 2 は、更に、イニシエータ 20 からターゲット 30 への SCSI 読み取り / 書き込みコマンドのフローを示す。

2) 図 3 は、イニシエータ 20 とターゲット 30 との間のセッションを示す。図 3 は、更に、イニシエータ 20 からターゲット 30 へのデータ・アウト PDU のフローを示す。

3) 図 4 は、イニシエータ 20 とターゲット 30 との間のセッションを示す。図 2 は、更に、ターゲット 30 からイニシエータ 20 へのデータ・イン PDU のフローを示す。

【0099】

イニシエータ 20 からターゲット 30 へのコマンド PDU のトラヒック・フロー

ここで図 2 を参照すると、イニシエータ 20 とターゲット 30 との間のセッションが示され、SCSI 読み取りおよび書き込みコマンドの例示的なトラヒック・フローが重ねて図示されている。

【0100】

SCSI コマンドは、SCSI イニシエータ 22 で発生する。次いで、コマンドを、i

10

20

30

40

50

SCSIイニシエータ24に渡すことができる(矢印50)。iSCSIイニシエータ24は、各コマンドをPDUに圧縮し、次いで、コマンドPDUを、シェーピングのため、QoSシェーピング・コンポーネント26に渡すことができる。

【0101】

当技術分野において公知のように、シェーピング・コンポーネント26はTCPでなくiSCSIにあることに留意すべきである。TCPの前にiSCSIにおいてシェーピングを実行するここでの目的は、QoSポリシーとiSCSIプロトコル要件との間の潜在的な競合状況を解決することである。これについて以下で説明する。

【0102】

iSCSIイニシエータ24によってTCPに送信されたコマンドPDUは、コマンド・シーケンス番号を有する。コマンド・シーケンス番号は、同一セッションの多数の接続に渡って、ターゲット30においてコマンドを順次供給するために用いられる。もしもイニシエータにおけるシェーピングがTCPにおいて行われると、コマンド・シーケンス番号の小さいコマンドPDUは、TCPにおいて、遅いサービス・クラスに関連した接続上で引き止められる場合がある。一方で、シーケンス番号の大きい他のコマンドPDUが、速いサービス・クラスを有する接続を介してiSCSIターゲット44に到着することがある。シーケンス番号の順序のため、後者の、速いネットワーク・サービス・クラスを有するコマンドを、iSCSIターゲット44において遅らせて、遅いネットワーク・サービス・クラスを有する前者のコマンドを待つ。この挙動により、性能の劣化が生じ、QoSの目的が果たされないことがある。

【0103】

この潜在的な競合状況を解決するため、コマンド・シーケンス番号を割り当てる前に、イニシエータ20においてシェーピングを実行する。本発明の好適な実施形態では、シェーピング・コンポーネント26は、iSCSIイニシエータ24内に配置する。

【0104】

本発明において、コマンドPDUは、I/OリソースのためのQoSポリシーに従ってシェーピングされる。コマンドPDUが即値データも含む場合、PDUは、ネットワーク・リソースのためのQoSポリシーに従って、更にシェーピングされる。

【0105】

次いで、iSCSIイニシエータ24は、コマンドPDUをTCPレイヤ32に送信することができる(矢印52)。次いで、TCPレイヤ32は、トラヒックを、QoSマーキング・コンポーネント28に送信することができる(矢印54)。

【0106】

マーキングされたトラヒックは、TCPレイヤ32に戻ることができる(矢印56)。当技術分野において公知のプロセスとして、TCPレイヤ32は、IPネットワーク46を介して、コマンドPDUをターゲット30に送信することができる(矢印58)。

【0107】

次いで、ターゲット30のTCPレイヤ34は、コマンドPDUを受信することができる(矢印60)。TCPレイヤ34は、コマンドPDUをiSCSIターゲット44に送ることができる(矢印62)。iSCSIターゲット44は、SCSIコマンドを抽出して、それらをシェーピングのためQoSコンポーネント38に渡すことができる。QoSコンポーネント38は、コマンドをSCSI実行エンジン42に送る前に、コマンドをシェーピングすることができる。コマンドは、すでにQoSシェーピング・コンポーネント26によってシェーピングされている場合があるが、その場合も、それらを、シェーピングのためQoSコンポーネント38に渡すことができることを注記しておく。これは、本発明によって対処される問題の第4項に上述したように、不従順イニシエータが潜在的に存在するためである。iSCSIターゲット44は、適格および不従順トラヒックを区別することができないので、到着した全コマンドPDUをシェーピングのために渡さなければならない。

【0108】

QoSシェーピング・コンポーネント26によってすでにシェーピングされた適格イニシエータ20からのコマンドは、QoSコンポーネント38における追加のシェーピングによって影響されないで済むことを注記しておく。

【0109】

次いで、iSCSIターゲット44は、適切な順序で、コマンドをSCSI実行エンジン42に供給することができる(矢印64)。

【0110】

本例示は、TCPレイヤ32とは別個のものとしてQoSマーキング・コンポーネント28を示すことを注記しておく。しかしながら、代替的な実施形態では、QoSマーキング・コンポーネント28は、TCPレイヤ32内に存在することができ、この場合も本実施形態の原理内とすることができる。

10

【0111】

更に、マーキング・コンポーネント28は、iSCSI以外のアプリケーションからのトラフィックを含めて、イニシエータ・ノード20から出立する全ネットワーク・トラフィックをマーキングする責任があることを注記しておく。更に、イニシエータ・ノード20は、TCPレイヤにおいて第2のシェーピング・コンポーネント(図示せず)を含むことができる。第2のシェーピング・コンポーネントは、iSCSI以外のアプリケーションのネットワーク・トラフィックをシェーピングする責任を有する場合がある。上述の代替案が本発明の真の精神および範囲内に含まれることは認められよう。

【0112】

イニシエータ20からターゲット30へのデータ・アウトPDUのトラフィック・フロー

ここで図3を参照すると、イニシエータ20とターゲット30との間のセッションが例示され、データ・アウトPDUの例示的なトラフィック・フローが重ねて図示されている。

20

【0113】

データ・アウトは、SCSIイニシエータ22で発生する。イニシエータ22は、これをiSCSIイニシエータ24に渡すことができる(矢印70)。iSCSIイニシエータ24は、データ・アウトPDUを生成し、データ・アウトPDUは、次いで、シェーピングのためQoSシェーピング・コンポーネント26に渡すことができる。

【0114】

QoSシェーピング・コンポーネント26は、データ・アウトPDUをTCPレイヤ32に送ることができる(矢印72)。TCPレイヤ32は、トラフィックをマーキングのためQoSマーキング・コンポーネント28に渡すことができる(矢印74)。次いで、QoSマーキング・コンポーネント28は、マーキングしたトラフィックをTCP32に戻すことができ(矢印76)、TCP32は、データ・アウトPDUを、IPネットワーク46を介してターゲット30に送信することができる(矢印78)。

30

【0115】

TCPレイヤ34は、データ・アウトPDUを受信し(矢印80)、それらをiSCSIターゲット44に供給することができる(矢印82)。iSCSIターゲット44は、PDUからデータを抽出し、データは次いでSCSI実行エンジン42に渡すことができる(矢印84)。

40

【0116】

ターゲット30からイニシエータ20へのデータ・インPDUのトラフィック・フロー

ここで図4を参照すると、イニシエータ20とターゲット30との間のセッションが例示され、データ・インPDUの例示的なトラフィック・フローを重ねて図示されている。

【0117】

データ・インは、SCSI LU(図示せず)で発生する。本実施形態では、SCSI実行エンジン42は、データ・インをiSCSIターゲット44に渡すことができる(矢印90)。iSCSIターゲット44は、データ・インPDUを生成し、データ・インPDUをTCPレイヤ34に送ることができる(矢印92)。

【0118】

50

従来技術のネットワークQoS環境において公知のように、TCPレイヤ34は、次いで、ターゲットにおいてネットワーク・トラフィックをシェーピングしマーキングすることができる。従って、TCPレイヤ34は、シェーピングおよびマーキングのため、データ・インPDUをQoSコンポーネント36に渡すことができる(矢印94)。

【0119】

ターゲット・ノード30におけるネットワーク・フィルタ仕様は、イニシエータ・ノード20におけるよりも単純であり得ることを注記しておく。これは、iSCSI接続忠誠の要件のため、および、各接続が各方向において1つのみのサービス・クラスを有することができるからである。イニシエータは、各コマンドごとに接続を選択する責任があり、ターゲットは、コマンドの全トラフィックのためにこの接続を用いなければならない。従って、ターゲットにおけるサービス・クラスは、単に、接続の識別証またはイニシエータの識別証から決定することができる。

10

【0120】

QoSコンポーネント36は、次いで、シェーピング/マーキングしたトラフィックをTCPレイヤ34に戻すことができる(矢印96)。TCPレイヤ34は、データ・インPDUを、IPネットワーク46を介してイニシエータ20に送ることができる(矢印98)。

【0121】

TCPレイヤ32は、データ・インPDUを受信して(矢印100)、それらをiSCSIイニシエータ24に供給することができる(矢印102)。iSCSIイニシエータ24は、PDUからデータを抽出して、データ・インをSCSIイニシエータ22に供給することができる(矢印104)。

20

【0122】

本例示は、QoSコンポーネント36を、TCPレイヤ34とは別個のものとして示していることを注記しておく。しかしながら、代替的な実施形態では、QoSマーキング・コンポーネント36は、TCPレイヤ34内に存在することができ、その場合も本実施形態の原理内である。

【0123】

更に、いくつかの実施形態では、QoSコンポーネント36は、iSCSI以外のアプリケーションからのトラフィックを含めて、ターゲット・ノード30から出立する全ネットワーク・トラフィックのシェーピングおよびマーキングに責任を有する場合がある。

30

【0124】

多数のネットワーク・サービス・クラス

各TCP接続は、各方向において1つのみのサービス・クラスに関連付けることができる。1セッション内で多数のサービス・クラスを提供するためには、多数の接続を用いれば良い。

【0125】

いくつかの実施形態では、iSCSIイニシエータ24またはiSCSIターゲット44は、1セッションにつき多数の接続をサポートすることができない場合がある。代替案として、各サービス・クラスごとに別個のセッションを使用可能である。

40

【0126】

イニシエータ20において用いられるオペレーティング・システム環境またはファイル・システムによって、多数のセッションがばらばらのLUにアクセスすることを必要とする制限が加わる場合があることを注記しておく。この場合、同一のイニシエータから同一のLUにアクセスするために多数のネットワーク・サービス・クラスを用いることは不可能なことがある。

【0127】

同一論理ユニットのための多数のI/Oサービス・クラス

SCSIアーキテクチャは、各I-T-L連結ごとに別個のタスク・セットを用いるように論理ユニットを構成する能力を与える。従って、本発明の好適な実施形態は、同一L

50

Uのために多数のI/Oサービス・クラスを規定することができる。各I-T-L連結のタスク・セットには、その読み取りコマンドおよび書き込みコマンドによって、異なるI/Oサービス・クラスを割り当てることができる。あるいは、同一LUの多数タスク・セット間でI/Oサービス・クラスを共有することが可能な場合があり、これによって、いくつかの異なるイニシエータからのトラヒックに、ターゲットにおける同一のI/Oクラスを与える。

#### 【0128】

##### I/Oフロー過剰

システム10の集中的な性質のため、イニシエータにおいてI/Oフロー仕様に従うコマンドPDUが、ターゲットに到着した際に不従順であるという状況があり得る。集中的な挙動を円滑にするため、本発明のいくつかの実施形態では、選択したI/Oサービス・クラスにおいてI/Oフロー過剰を許容する場合がある。ターゲット30においてQoSコンポーネント38が用いるI/Oフロー仕様は、過剰バイト・レートおよび過剰I/Oレートのためのパラメータを含むことができる。

10

#### 【0129】

ターゲットにおける不従順コマンドは、「タスク・セット・フル(task-set full)」または「ビジー(busy)」エラーによって、QoSコマンド38によって拒絶される場合があることを注記しておく。これらのエラーは、できるだけ阻止しなければならない。I/Oフロー過剰パラメータは、システムおよびネットワーク変動によって生じる一時的な不適合を補償することができる。I/Oフロー過剰は、QoSを全く用いないイニシエータを補償しない。

20

#### 【0130】

##### フィルタ仕様およびフロー仕様

フィルタ仕様属性は、QoSが存在する場合にiSCSIフローの区分化に影響を与えるように選択することができる。かかるフィルタ属性の例には、TCP4タプル(イニシエータおよびターゲットIPアドレスおよびポート)、LUN、イニシエータ・ユーザID、セッションID、アプリケーション識別属性、動作時間等が含まれる。上に列挙したものは、単に例示であり、限定することを意図しない。記載していないがiSCSIフローの性能に影響を与え得る他のフィルタ属性も、本発明の範囲内に含まれることは認められよう。

30

#### 【0131】

いくつかの実施形態では、フィルタ属性に冗長性があり得ることを注記しておく。1例として、4タプルを指定する場合、セッションidは冗長である場合がある。それにもかかわらず、冗長性を用いて、サービス・クラスを割り当てる場合にパターンを支持することができる。1例として、1セッション内の全接続に同一クラスを割り当てることができる。分類方法において、属性のサブセットまたは各属性ごとの値の範囲を用いることができる。

#### 【0132】

QoS10は、ネットワークキング・リソースおよびI/Oリソースのために別個のフロー仕様を用いることができる。

40

#### 【0133】

ネットワークキングでは、DiffServのために規定したもの等、標準的なフロー仕様パラメータを用いることができる。ネットワーク・フロー・パラメータの例は、平均バイト・レート、ピーク・バイト・レート、バースト・サイズを含むことができる。

#### 【0134】

I/Oでは、既存の規格はない。QoS10の好適な実施形態は、SCSI実行エンジン42のI/Oリソースの共有に影響を与えるフロー仕様パラメータを選択することができる。フロー・パラメータの例は、バイト・レート、I/Oレート、I/Oフロー過剰等を含むことができる。上に列挙したものは単に例示であり、限定することを意図しない。記載していないがSCSI実行エンジン42の共有に影響を与え得る他のフロー仕様パラ

50

メータも、本発明の範囲内に含まれることは認められよう。

【0135】

イニシエータにおいて用いられるフィルタ仕様属性もしくはフロー仕様パラメータまたはその双方のいくつかは、ターゲットで用いられるものとは異なる場合があることを注記しておく。

【0136】

I/Oレートおよびバイト・レートは、ランダム・アクセスと順次アクセスとの間のトレードオフを与えることがあることを注記しておく。ランダム・アクセスは、通常、小さいデータ長によって特徴付けられ、主にI/Oレートによって制約される。一方、順次アクセスは、通常、大きいデータ長によって特徴付けられ、主にバイト・レートによって制約される。

10

【0137】

本明細書中に記載した実施形態は、iSCSIもしくはTCP/IPまたはその双方のソフトウェアおよびハードウェア実施、ならびに、ソフトウェアおよびハードウェア実施のいずれかの組み合わせに適用される。更に、1つのノードは、2つ以上のインタフェース・カードまたは2つ以上のネットワーク・リンクを有する場合がある。また、同一セッションにおけるイニシエータとターゲットとの間の多数の接続は、異なるIPアドレスを用いる場合がある。これらの全ての変形は、本発明の真の精神および範囲内に含まれる。

【0138】

本明細書中に開示した方法、装置、およびシステムは、特定のコンピュータ・ハードウェアまたはソフトウェアを参照して記載した場合もあり、そうでない場合もあるが、本明細書中に開示した方法、装置、およびシステムは、従来技法を用いたコンピュータ・ハードウェアまたはソフトウェアにおいて容易に実施可能であることは認められよう。

20

【0139】

本発明について、1つ以上の具体的な実施形態を参照して説明したが、この説明は、全体として本発明を例示することを意図し、図示した実施形態に本発明を限定するものと解釈されるべきではない。当業者は、本明細書中に具体的に示さないが本発明の真の精神および範囲内である様々な変形を認識していることは認められよう。

【図面の簡単な説明】

【0140】

30

【図1】本発明の1実施形態に従って構成および動作させる、ネットワーク・ストレージ環境のためのサービスの質(QoS)アーキテクチャおよびシステムのブロック図である。

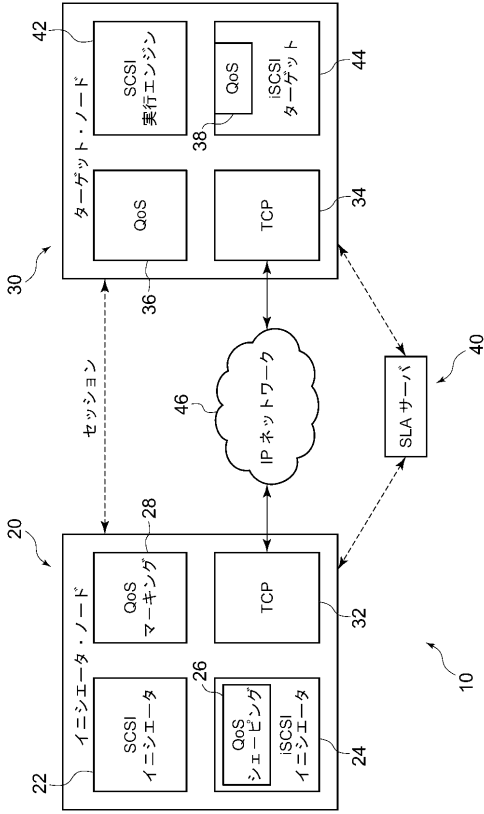
【図2】本発明の1実施形態に従って構成および動作させる、イニシエータとターゲットとの間のセッションのブロック図であり、イニシエータからターゲットへのコマンド・プロトコル・データ・ユニット(PDU)のトラフィック・フローを例示する。

【図3】本発明の1実施形態に従って構成および動作させる、イニシエータとターゲットとの間のセッションのブロック図であり、イニシエータからターゲットへのPDUのトラフィック・フローを例示する。

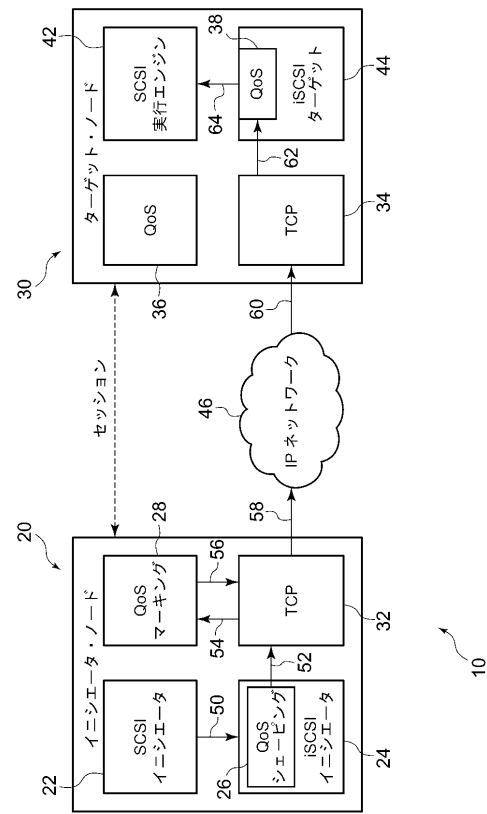
【図4】本発明の1実施形態に従って構成および動作させる、イニシエータとターゲットとの間のセッションのブロック図であり、ターゲットからイニシエータへのPDUのトラフィック・フローを例示する。

40

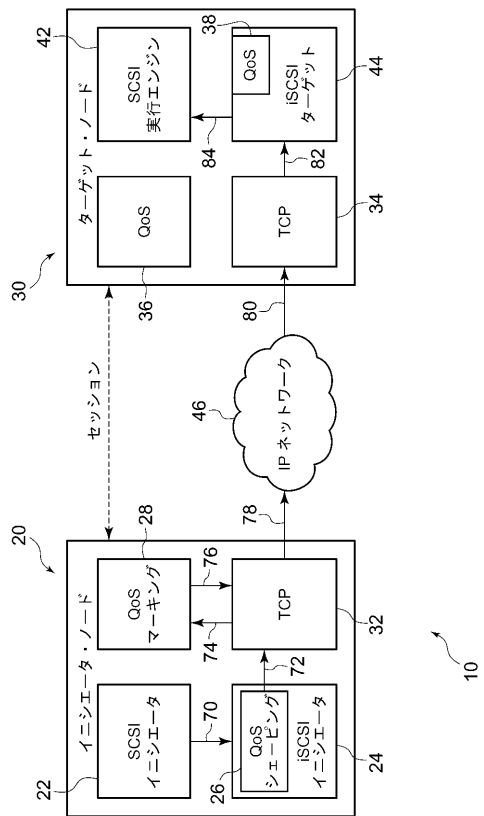
【図 1】



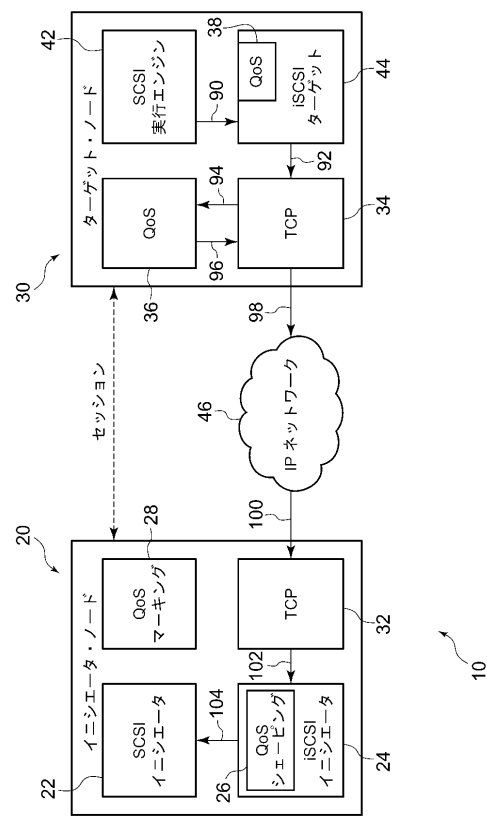
【図 2】



【図 3】



【図 4】



## フロントページの続き

(74)代理人 100086243

弁理士 坂口 博

(72)発明者 バジライ、ツイー、ポーラ

イスラエル国69499 テル アヴィヴ カシャニイ ストリート 3ビー

(72)発明者 マーベルグ、ジョン

イスラエル国34335 ハイファ ギボン ストリート 12

審査官 坂東 博司

(56)参考文献 特開平09-181740(JP,A)

特開平11-341015(JP,A)

特開2001-344228(JP,A)

特開2001-197111(JP,A)

特開2000-252997(JP,A)

特表2004-510252(JP,A)

特表2005-502225(JP,A)

特表2004-526218(JP,A)

国際公開第02/027519(WO,A1)

国際公開第02/086674(WO,A1)

国際公開第02/017034(WO,A1)

Jamal Hadi Salim, Kernel Korner IPバンド幅の管理, Linux Japan 第2巻

第7号, 日本, レーザー5 LASER 5, 1999年10月 1日, 第2巻 第7号, 115~1

20

(58)調査した分野(Int.Cl., DB名)

G06F 13/10

H04L 12/56