

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3975703号

(P3975703)

(45) 発行日 平成19年9月12日(2007.9.12)

(24) 登録日 平成19年6月29日(2007.6.29)

(51) Int. Cl.

G06F 9/48 (2006.01)

F I

G06F 9/46 452C

請求項の数 36 (全 26 頁)

(21) 出願番号 特願2001-247157 (P2001-247157)
 (22) 出願日 平成13年8月16日(2001.8.16)
 (65) 公開番号 特開2003-58382 (P2003-58382A)
 (43) 公開日 平成15年2月28日(2003.2.28)
 審査請求日 平成16年6月18日(2004.6.18)

(73) 特許権者 000004237
 日本電気株式会社
 東京都港区芝五丁目7番1号
 (74) 代理人 100088959
 弁理士 境 廣巳
 (72) 発明者 田淵 仁浩
 東京都港区芝五丁目7番1号 日本電気株
 式会社内

審査官 鈴木 修治

最終頁に続く

(54) 【発明の名称】 情報処理システムにおける優先実行制御方法及びその装置並びにプログラム

(57) 【特許請求の範囲】

【請求項1】

データ記憶装置に記憶されたデータを参照更新する処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納するステップと、前記優先キューに対応して設けられ、前記優先キューに格納された処理要求を専門に処理する優先処理手段が、前記優先キューに処理要求が格納されるのを待ち、前記優先キューに処理要求が格納されると、直ちにその処理要求を前記優先キューから取り出して処理するステップと、前記通常キューに対応して設けられ、前記優先処理手段と同じ機能を有し、前記通常キューに格納された処理要求を専門に処理する通常処理手段が、前記通常キューに処理要求が格納されるのを待ち、前記通常キューに処理要求が格納されると、直ちにその処理要求を前記通常キューから取り出して処理するステップとを含む情報処理システムにおける優先実行制御方法。

【請求項2】

前記優先処理手段および前記通常処理手段のそれぞれは、コンピュータ上で互いに並行して動作するスレッドである請求項1記載の情報処理システムにおける優先実行制御方法。

【請求項3】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度より低く設定されている請求項2記載の情報処理システムにおける優先実行制御方法。

10

20

【請求項 4】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度と等しく設定されている請求項 2 記載の情報処理システムにおける優先実行制御方法。

【請求項 5】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度より高く設定されている請求項 2 記載の情報処理システムにおける優先実行制御方法。

【請求項 6】

前記優先処理手段を実現するスレッドまたは前記通常処理手段を実現するスレッドのスレッド優先度を動的に変更するステップを含む請求項 2 記載の情報処理システムにおける優先実行制御方法。

10

【請求項 7】

処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納するステップは、予め定められた優先ルールを参照して優先度の判定を行う請求項 1 記載の情報処理システムにおける優先実行制御方法。

【請求項 8】

前記優先ルールを動的に変更するステップを含む請求項 7 記載の情報処理システムにおける優先実行制御方法。

【請求項 9】

処理要求の処理結果を処理要求元に対して通知するステップを含む請求項 1 記載の情報処理システムにおける優先実行制御方法。

20

【請求項 10】

前記優先処理手段を実現するスレッドが処理要求を処理した結果及び前記通常処理手段を実現するスレッドが処理要求を処理した結果の少なくとも一方を処理要求元に通知するステップを含む請求項 2 記載の情報処理システムにおける優先実行制御方法。

【請求項 11】

前記優先キュー及び前記優先処理手段を実現するスレッドの組を動的に生成するステップと、生成した前記優先キュー及び前記優先処理手段を実現するスレッドの組を動的に消滅させるステップとを含む請求項 2 記載の情報処理システムにおける優先実行制御方法。

30

【請求項 12】

ネットワークを介してクライアントに接続され、クライアントからの処理要求を処理した結果を要求元のクライアントに返却するサーバにおける、処理要求を処理して応答結果を生成する処理部および該処理部で生成された応答結果を要求元のクライアントに返却する処理を行う応答部の少なくとも一方に、請求項 1 乃至請求項 11 の何れか 1 項に記載された優先実行制御にかかる手順を実行させるようにした情報処理システムにおける優先実行制御方法。

【請求項 13】

優先キュー及び通常キューと、データ記憶装置に記憶されたデータを参照更新する処理要求をその処理の優先度に応じて前記優先キュー、前記通常キューの何れか一方に格納する分配手段と、前記優先キューに対応して設けられ、前記優先キューに格納された処理要求を専門に処理する手段であって、前記優先キューに処理要求が格納されるのを待ち、前記優先キューに処理要求が格納されると、直ちにその処理要求を前記優先キューから取り出して処理する優先処理手段と、前記通常キューに対応して設けられ、前記優先処理手段と同じ機能を有し、前記通常キューに格納された処理要求を専門に処理する手段であって、前記通常キューに処理要求が格納されるのを待ち、前記通常キューに処理要求が格納されると、直ちにその処理要求を前記通常キューから取り出して処理する通常処理手段とを備えた情報処理システムにおける優先実行制御装置。

40

【請求項 14】

前記優先処理手段および前記通常処理手段のそれぞれは、コンピュータ上で互いに並行

50

して動作するスレッドである請求項 1 3 記載の情報処理システムにおける優先実行制御装置。

【請求項 1 5】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度より低く設定されている請求項 1 4 記載の情報処理システムにおける優先実行制御装置。

【請求項 1 6】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度と等しく設定されている請求項 1 4 記載の情報処理システムにおける優先実行制御装置。

10

【請求項 1 7】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度より高く設定されている請求項 1 4 記載の情報処理システムにおける優先実行制御装置。

【請求項 1 8】

前記優先処理手段を実現するスレッドまたは前記通常処理手段を実現するスレッドのスレッド優先度を動的に変更するスレッド優先度変更手段を含む請求項 1 4 記載の情報処理システムにおける優先実行制御装置。

【請求項 1 9】

処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納する分配手段によって参照される優先ルールを記憶する優先ルール記憶手段を含む請求項 1 3 記載の情報処理システムにおける優先実行制御装置。

20

【請求項 2 0】

前記優先ルール記憶手段中の優先ルールを動的に変更する優先ルール変更手段を含む請求項 1 9 記載の情報処理システムにおける優先実行制御装置。

【請求項 2 1】

処理要求の処理結果を処理要求元に対して通知する状態通知手段を含む請求項 1 3 記載の情報処理システムにおける優先実行制御装置。

【請求項 2 2】

前記優先処理手段を実現するスレッドが処理要求を処理した結果及び前記通常処理手段を実現するスレッドが処理要求を処理した結果の少なくとも一方を処理要求元に通知する状態通知手段を含む請求項 1 4 記載の情報処理システムにおける優先実行制御装置。

30

【請求項 2 3】

前記優先キュー及び前記優先処理手段を実現するスレッドの組を動的に生成し、且つ、生成した前記優先キュー及び前記優先処理手段を実現するスレッドの組を動的に消滅させるスレッド割付手段を含む請求項 1 4 記載の情報処理システムにおける優先実行制御装置。

【請求項 2 4】

ネットワークを介してクライアントに接続され、クライアントからの処理要求を処理した結果を要求元のクライアントに返却するサーバにおける、処理要求を処理して応答結果を生成する処理部および該処理部で生成された応答結果を要求元のクライアントに返却する処理を行う応答部の少なくとも一方に、請求項 1 3 乃至請求項 2 3 の何れか 1 項に記載された優先実行制御にかかる手段を備える情報処理システムにおける優先実行制御装置。

40

【請求項 2 5】

コンピュータを、データ記憶装置に記憶されたデータを参照更新する処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納する分配手段、前記優先キューに対応して設けられ、前記優先キューに格納された処理要求を専門に処理する手段であって、前記優先キューに処理要求が格納されるのを待ち、前記優先キューに処理要求が格納されると、直ちにその処理要求を前記優先キューから取り出して処理する優先処理手段、前記通常キューに対応して設けられ、前記優先処理手段と同じ機能を有し、前記

50

通常キューに格納された処理要求を専門に処理する手段であって、前記通常キューに処理要求が格納されるのを待ち、前記通常キューに処理要求が格納されると、直ちにその処理要求を前記通常キューから取り出して処理する通常処理手段、として機能させるプログラム。

【請求項 26】

前記優先処理手段および前記通常処理手段のそれぞれは、コンピュータ上で互いに並行して動作するスレッドである請求項 25 記載のプログラム。

【請求項 27】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度より低く設定されている請求項 26 記載のプログラム。

10

【請求項 28】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度と等しく設定されている請求項 26 記載のプログラム。

【請求項 29】

前記優先処理手段を実現するスレッドのスレッド優先度は前記通常処理手段を実現するスレッドのスレッド優先度より高く設定されている請求項 26 記載のプログラム。

【請求項 30】

前記コンピュータを、更に、前記優先処理手段を実現するスレッドまたは前記通常処理手段を実現するスレッドのスレッド優先度を動的に変更するスレッド優先度変更手段、として機能させる請求項 26 記載のプログラム。

20

【請求項 31】

前記コンピュータを、更に、処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納する分配手段によって参照される優先ルールを記憶する優先ルール記憶手段中の優先ルールを動的に変更する優先ルール変更手段、として機能させる請求項 25 記載のプログラム。

【請求項 32】

前記コンピュータを、更に、処理要求の処理結果を処理要求元に対して通知する状態通知手段、として機能させる請求項 25 記載のプログラム。

【請求項 33】

前記コンピュータを、更に、前記優先処理手段を実現するスレッドが処理要求を処理した結果及び前記通常処理手段を実現するスレッドが処理要求を処理した結果の少なくとも一方を処理要求元に通知する状態通知手段、として機能させる請求項 26 記載のプログラム。

30

【請求項 34】

前記コンピュータを、更に、前記優先キュー及び前記優先処理手段を実現するスレッドの組を動的に生成し、且つ、生成した前記優先キュー及び前記優先処理手段を実現するスレッドの組を動的に消滅させるスレッド割付手段、として機能させる請求項 26 記載のプログラム。

【請求項 35】

ネットワークを介してクライアントに接続され、クライアントからの処理要求を処理した結果を要求元のクライアントに返却するサーバコンピュータを、クライアントからの処理要求をその処理の優先度に応じて第 1 の優先キュー及び第 1 の通常キューの何れか一方に格納する第 1 の分配手段、前記第 1 の優先キューに対応して設けられ前記第 1 の優先キューに格納された処理要求を処理する第 1 の優先処理手段、前記第 1 の通常キューに対応して設けられ前記第 1 の通常キューに格納された処理要求を処理する第 1 の通常処理手段を含む処理部として機能させるプログラム。

40

【請求項 36】

前記サーバコンピュータを、更に、前記処理部の前記第 1 の優先処理手段及び前記第 1 の通常処理手段からクライアントに返却すべき応答結果を含む処理要求を受け取ってその処理の優先度に応じて第 2 の優先キュー及び第 2 の通常キューの何れか一方に格納する第

50

2の分配手段、前記第2の優先キューに対応して設けられ前記第2の優先キューに格納された処理要求を処理する第2の優先処理手段、前記第2の通常キューに対応して設けられ前記第2の通常キューに格納された処理要求を処理する第2の通常処理手段を含む応答部として機能させる請求項35記載のプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は情報処理システムにおける処理要求の実行制御に関し、特にアプリケーションプログラムから出力される複数の処理要求のうち、優先的に処理すべき処理要求をできるだけ速やかに処理することができる優先実行制御方法及びその装置に関する。

10

【0002】

【従来の技術】

情報処理システムにおける単位時間当たりの処理要求発生頻度と単位時間当たりの処理量との差を吸収することなどを目的として、キューが一般に用いられている。図17(a)にキューを用いた典型的な実行制御方法の概略を示す。図17(a)において、アプリケーションプログラム等の処理要求元1101から発生した個々の処理要求1102は、キュー1103に一旦格納される。処理部1104は、キュー1103に処理要求1102が格納されると、キュー1103から先入れ先出し方式(FIFO)に従って1つの処理要求1102を読み出し、処理要求1102の内容に従って当該処理要求1102を処理し、処理結果1105を出力する。処理部1104は、1つの処理要求1102を処理し

20

【0003】

図17(a)の構成では処理要求1102はその発生順に逐次に処理される為、特定の処理要求を他の処理要求より優先的に処理することはできない。特定の処理要求を優先的に実行する必要がある場合、単一のキューでなく、優先キューと通常キューの2つのキューが一般に用いられる。図17(b)に2つのキューを用いた典型的な実行制御方法の概略を示す。図17(b)において、アプリケーションプログラム等の処理要求元1101から発生した個々の処理要求1102は、その処理の優先度に応じて分配部1106により、優先度の高いものは優先キュー1107へ、そうでないものは通常キュー1108へそれぞれ一旦格納される。処理部1104は、所定の規則に従って優先キュー1107及び通常キュー1108から処理要求1102を読み出して処理し、処理結果1105を出力する。所定の規則が、優先キュー1107に処理要求がある場合にはそこから処理要求を読み出し、ない場合に限って通常キュー1108から処理要求を読み出すという規則である場合、優先キュー1107に格納される優先度の高い処理要求1102が通常キュー1108に格納される処理要求1102より優先的に処理されることになる。

30

【0004】

このような優先キュー及び通常キューの2つのキューを用いた優先実行制御方法を記載した文献として、特開平4-176236号公報がある。同公報では、通常送信と優先送信の何れか一方が指定された送信データをその指定に従って優先キュー又は通常キューに格納し、予め設定された優先通信と通常通信の比率に従って優先キュー又は通常キューから送信データを読み出して送信を行うファクシミリ装置が開示されている。他に、パケット通信ネットワークにおける経路制御装置に見られるように、QoS(Quality of Service)を目的として、パケットのヘッド情報に応じて優先キューもしくはは通常キューへ振り分けることで、特定のパケット(例えば音声パケット)の送信を優先的に行う場合にも同様の優先実行制御方法が採用されている。

40

【0005】

他方、情報処理システムにおける処理能力を改善する手法として処理の並列化がある。しかし、優先キュー及び通常キューの2つのキューを用いた優先実行制御方法に対して処理の並列化技術を適用した例は見当たらない。若し、図17(b)における処理部1104

50

の並列化を考えた場合、図17(c)に示されるような構成となるであろう。図17(c)において、処理部1104A及び処理部1104Bは、図17(b)における処理部1104と同機能を有するもので、各々は、所定の規則に従って優先キュー1107及び通常キュー1108から処理要求1102を読み出して処理し、処理結果1105を出力する。処理要求1102を処理する処理部が2つ並行して動作するため、当該情報処理システムにおける単位時間当たりの処理量が増大する利点がある。

【0006】

【発明が解決しようとする課題】

図17(b)に示した優先実行制御方法では、通常キュー1108から処理要求1102を取り出す確率に比べて優先キュー1107から処理要求1102を取り出す確率を高めることで、優先キュー1107に格納された優先度の高い処理要求1102を通常キュー1108に格納された処理要求1102より優先的に処理することができる。

10

【0007】

しかしながら、優先キュー1107が空の状態であり、通常キュー1108に1つ以上の処理要求1102が格納されており、処理部1104が通常キュー1108から1つの処理要求1102を取り出し、その処理を開始した時点で、優先度の高い処理要求が優先キュー1107に格納されたとき、当該優先度の高い処理要求が速やかに処理されないという課題がある。これは、優先キュー1107に処理要求がある場合にはそこから処理要求を読み出し、ない場合に限って通常キュー1108から処理要求を読み出すという最優先の規則を適用しても改善されない。その理由は、優先度の高い処理要求が優先キュー1107に投入された時点では処理部1104は通常キュー1108から取り出した処理要求を実行しており、その処理が終わらなければ、処理部1104は優先キュー1107から処理要求を取り出さないからである。

20

【0008】

同様の問題は図17(c)に示したように処理部を並列化した構成でも起きる。つまり、処理部1104Aが通常キュー1108中の或る1つの処理要求1102を処理しており、処理部1104Bが通常キュー1108中の別の1つの処理要求1102を処理しているときに、空き状態であった優先キュー1107に優先度の高い処理要求が投入されると、この処理要求は処理部1104A、1104Bの何れか一方が現在の処理要求の処理を終了するまで待たされることになる。

30

【0009】

本発明はこのような事情に鑑みて提案されたものであり、その目的は、優先的に処理すべき処理要求をできるだけ速やかに処理することができる優先実行制御方法及びその装置を提供することにある。

【0010】

【課題を解決するための手段】

本発明は優先キュー及び通常キューのそれぞれに対して専用の処理手段を設けることを基本とする。具体的には、本発明の情報処理システムにおける優先実行制御方法においては、データ記憶装置に記憶されたデータを参照更新する処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納するステップと、前記優先キューに対応して設けられ、前記優先キューに格納された処理要求を専門に処理する優先処理手段が、前記優先キューに処理要求が格納されるのを待ち、前記優先キューに処理要求が格納されると、直ちにその処理要求を前記優先キューから取り出して処理するステップと、前記通常キューに対応して設けられ、前記優先処理手段と同じ機能を有し、前記通常キューに格納された処理要求を専門に処理する通常処理手段が、前記通常キューに処理要求が格納されるのを待ち、前記通常キューに処理要求が格納されると、直ちにその処理要求を前記通常キューから取り出して処理するステップとを含んでいる。また、本発明の情報処理システムにおける優先実行制御装置は、優先キュー及び通常キューと、データ記憶装置に記憶されたデータを参照更新する処理要求をその処理の優先度に応じて前記優先キュー、前記通常キューの何れか一方に格納する分配手段と、前記優先キューに対応して設けられ、

40

50

前記優先キューに格納された処理要求を専門に処理する手段であって、前記優先キューに処理要求が格納されるのを待ち、前記優先キューに処理要求が格納されると、直ちにその処理要求を前記優先キューから取り出して処理する優先処理手段と、前記通常キューに対応して設けられ、前記優先処理手段と同じ機能を有し、前記通常キューに格納された処理要求を専門に処理する手段であって、前記通常キューに処理要求が格納されるのを待ち、前記通常キューに処理要求が格納されると、直ちにその処理要求を前記通常キューから取り出して処理する通常処理手段とを備えている。

【0011】

【作用】

本発明の情報処理システムにおける優先実行制御方法及びその装置にあつては、優先キューに1つも処理要求が格納されていない状態では、たとえ通常キューに複数の処理要求が格納されていたとしても、通常キューから処理要求を取り出して処理するのは通常処理手段だけで、優先処理手段は優先キューに処理要求が格納されるのを待っており、優先キューに処理要求が格納されると、この処理要求が優先処理手段によって直に取り出されて処理される。これによって、優先的に処理すべき処理要求が速やかに処理されることになる。

【0012】

【発明の実施の形態】

次に本発明の実施の形態の例について図面を参照して詳細に説明する。

【0013】

図1を参照すると、本発明の一実施の形態を適用した情報処理システムにおける優先実行制御装置は、優先キュー1及び通常キュー2の2つのキューと、処理要求3を発生するアプリケーションプログラム等の処理要求元4と、処理要求元4から発生した処理要求3をその処理の優先度に応じて優先キュー1、通常キュー2の何れか一方に分配する分配部5と、優先キュー1に対応して設けられ、優先キュー1に格納された処理要求3を専門に処理する優先処理部6と、通常キュー2に対応して設けられ、通常キュー2に格納された処理要求を専門に処理する通常処理部7とを含んで構成されている。

【0014】

処理要求3は1つの完結した処理を要求する単位であり、処理すべきデータや場合によっては処理の種類などを含んで構成される。例えば、或る計算を行うための被演算データであったり、データベースを更新するためのデータであったり、`java`や`javascript`等のインタプリタ型の言語で記述されたプログラムであったりする。優先処理部6及び通常処理部7は、共に同じ機能を有しており、全ての種類の処理要求3を処理することができる。ここでの処理は、例えば、処理要求3に含まれる被演算データを用いて所定の計算を行って結果を出力したり、処理要求3で指定された通りにデータベースなどを更新したり、インタプリタ型の言語で記述されたプログラムを解釈して実行したりする処理である。

【0015】

処理要求元4で発生した処理要求3をその処理の優先度に基づいて分配部5が優先キュー1または通常キュー2に分配する方法は、任意の方法を採用することができる。例えば、処理要求元4が処理要求3を発生する際に、その優先度を付随させ、分配部5は処理要求3に付随する優先度を解析し、高い優先度の処理要求3は優先キュー1に格納し、そうでない処理要求3は通常キュー2に格納する方法を採用することができる。他の例として、処理要求元4が複数存在する場合、個々の処理要求元4毎に予め優先度を付与しておき、分配部5は、優先度の高い処理要求元4から発生した処理要求3は優先キュー1に格納し、それ以外の処理要求元4から発生した処理要求3は通常キュー2に格納する方法を採用することができる。また、分配部5が処理要求の優先度を判定するルールは分配部5に固定的に設定しておいても良いし、別の部分に優先ルールを記憶しておき、分配部5がその記憶された優先ルールを参照するようにしても良い。

【0016】

10

20

30

40

50

このように構成された本実施の形態にあっては、処理要求元4から発生した個々の処理要求3は、その処理の優先度に応じて分配部5により、優先度の高いものは優先キュー1へ、そうでないものは通常キュー2へそれぞれ分配される。優先処理部6は、優先キュー1に処理要求3が格納されるのを待っており、優先キュー1に処理要求3が格納されると、それを取り出して処理し、処理結果8を出力する。1つの処理要求3を処理し終わると、優先キュー1に他の処理要求3が存在するか否かを調べ、存在すれば処理要求3を取り出し、同様に処理する。存在しなければ、優先キュー1に処理要求3が格納されるのを待つ。

【0017】

他方、通常処理部7は、通常キュー2に処理要求3が格納されるのを待っており、通常キュー2に処理要求3が格納されると、それを取り出して処理し、処理結果8を出力する。1つの処理要求3を処理し終わると、通常キュー2に他の処理要求3が存在するか否かを調べ、存在すれば処理要求3を取り出し、同様に処理する。存在しなければ、通常キュー2に処理要求3が格納されるのを待つ。

【0018】

このように本実施の形態によれば、優先キュー1に1つも処理要求3が格納されていない状態では、たとえ通常キュー2に複数の処理要求が格納されていたとしても、通常キュー2から処理要求を取り出して処理するのは通常処理部7だけで、優先処理部6は優先キュー1に処理要求3が格納されるのを待っている。この点が処理部を並列化した図17(c)と大きく相違する点である。このため、優先キュー1に処理要求3が格納されると、この処理要求3は優先処理部6によって直ちに取出され、速やかに処理されることになる。

【0019】

【実施例】

次に本実施の形態の実施例について図面を参照して詳細に説明する。

【0020】

【第1の実施例】

図2を参照すると、本実施例にかかる情報処理システムにおける優先実行制御方法を実施するコンピュータ100は、アプリケーションプログラム(クライアント)200から出力される複数の処理要求のうち優先的に処理すべき処理要求を他の処理要求より優先的に実行する機能を有しており、優先ルール記憶手段101、処理要求分配手段102、優先キュー103、通常キュー104、優先要求処理スレッド105、通常処理スレッド106及びデータ記憶装置107を含んで構成される。

【0021】

優先キュー103は、アプリケーションプログラム200から出力される処理要求のうち優先的に処理すべき処理要求を先入れ先出し方式で格納するキューであり、例えば主記憶上に設けられている。他方、通常キュー104は、アプリケーションプログラム200から出力される処理要求のうち優先的に処理すべき処理要求以外の処理要求を先入れ先出し方式で格納するキューであり、同様に例えば主記憶上に設けられている。優先キュー103及び通常キュー104のキュー長は固定長でも、可変長でも良い。可変長の場合、キューの為に使用するメモリの効率的な利用を考慮したキュー長制御を実施するのが望ましい。例えば、処理要求をキューに投入し、またキューから取り出すためにPUTメソッドとGETメソッドをキューに実装し、処理要求分配手段102が処理要求を優先キュー103または通常キュー104に投入する際に該当するキューのPUTメソッドを呼び出し、また優先要求処理スレッド105、通常要求処理スレッド106が優先キュー103、通常キュー104から処理要求を取り出す際に該当するキューのGETメソッドを呼び出す構成を採用している場合、キュー自身がキュー長を管理し、現在のキュー長と溜まっている処理要求の数が一致している状態でPUTメソッドが呼び出されると、キュー自身がキュー長を伸ばして処理要求の取り込みを可能にし、逆にGETメソッドが呼び出された結果、溜まっている処理要求が0になり、かつ、キュー長を伸ばしていたら、キュー自身が

10

20

30

40

50

所定のキュー長に戻すようなキュー長制御を採用することができる。

【0022】

優先要求処理スレッド105は、図1の優先処理部6に相当し、優先キュー103に対応して設けられている。他方、通常要求処理スレッド106は、図1の通常処理部7に相当し、通常キュー104に対応して設けられている。つまり、本実施の形態では、マルチスレッドの技術を使って処理要求を処理する処理部を実現している。ここで、スレッドとは、狭義の意味ではプロセスと区別されるが、本明細書ではプロセスを含めた広義の意味で使用している。つまり、スレッド105、106は狭義の意味でのスレッドであっても良いし、プロセスであっても良い。

【0023】

マルチスレッド(マルチプロセス)を実現する情報処理システムにおいて、スレッドのスケジューリング方式としてスレッドの優先度に基づいて、優先度の高いスレッドほどCPU時間をより多く割り当てる方式がある。例えば、高、低の2つの優先度があり、高い優先度を持つスレッドに対して1回あたりに割り当てるCPU時間の最大値を2T、低い優先度を持つスレッドに対して1回あたりに割り当てるCPU時間をその半分のTとし、実行可能状態のスレッド群からラウンドロビン方式でスレッドを順次を選択してCPU時間を割り当てていくスケジューリング方式では、高い優先度を持つスレッドは低い優先度を持つスレッドの2倍のCPU時間が割り当てられる。本実施例ではこのように優先度の高いスレッドほどCPU時間をより多く割り当てるスケジューリング方式を使って優先要求処理スレッド105及び通常要求処理スレッド106へのCPU時間の割り当てを行う。

【0024】

各スレッド105及び106のスレッド自体の優先度は事前に設定されており、望ましくは、優先要求処理スレッド105のスレッドの優先度は、通常要求処理スレッド106のスレッドの優先度より高いか、同じに設定されている。勿論、優先要求処理スレッド105のスレッドの優先度を、通常要求処理スレッド106のスレッドの優先度より低く設定しても、一定の効果は得られる。

【0025】

優先要求処理スレッド105及び通常要求処理スレッド106は、共に同じ機能を有する。両スレッド105、106の処理の概略を図3に示す。

【0026】

図3を参照すると、各スレッド105、106は、自スレッドに対応するキューに処理要求が格納されるのを待っており(ステップS1)、自スレッドに対応するキューに処理要求が格納されると(ステップS1でYES)、直ちにその処理要求をキューから取り出し(ステップS2)、その処理要求を処理する(ステップS3)。そして、1つの処理要求の処理を終えると、ステップS1に戻り、他の処理要求がキューに格納されていれば直ちにそれを取り出して処理し、格納されていなければ、格納されるのを待つ。但し、各スレッド105、106は1回あたりのCPU時間の持ち時間が決まっているので、自分の持ち時間を使い果たすと一旦CPUを明け渡す。また、自スレッドに対応するキューに処理要求が1つも格納されていなかった場合には、持ち時間を使い果たしていてもスリープ状態になってCPUを明け渡す。スリープ状態のスレッドは、キューに処理要求が格納されると実行可能状態となる。

【0027】

本実施例は、データ記憶装置107に記憶されたデータを参照更新する処理を必要とする処理要求を想定している。従って、双方のスレッド105及び106は同じデータ記憶装置107を共有している。各スレッド105、106がデータ記憶装置107に記憶されたデータを更新する場合、更新箇所にロックをかけ排他的に変更を実施する。ロックをかける単位、つまり排他の単位はスレッド105、106の並列性を十分に確保するためにできるだけ小さく設定される。

【0028】

処理要求分配手段102は、アプリケーション200から出力される処理要求を優先キュー

10

20

30

40

50

ー 1 0 3 及び通常キュー 1 0 4 の何れか一方に分配する手段であり、優先的に処理すべき処理要求は優先キュー 1 0 3 に、そうでない処理要求は通常キュー 1 0 4 にそれぞれ格納する。本実施例では、処理要求分配手段 1 0 2 は、処理要求に付随する優先度データと優先ルール記憶手段 1 0 1 に記憶された優先ルールとに基づいて、各処理要求の優先度を判定し、その判定結果に基づいてキュー 1 0 3、1 0 4 の何れか一方に処理要求を分配する。

【 0 0 2 9 】

図 4 に、アプリケーションプログラム 2 0 0 が出力する処理要求のフォーマット例を示す。図 4 を参照すると、個々の処理要求 3 0 0 は、処理要求本体 3 0 1 とヘッダ 3 0 2 とから構成される。処理要求本体 3 0 1 には、処理すべきデータが含まれる。例えば、或る計算を行うための被演算データや、データ記憶装置 1 0 7 の更新アドレス及び更新データや、j a v a または j a v a スクリプト等のインタプリタ型の言語で記述されたプログラムなどである。また、アプリケーションプログラム 2 0 0 から複数の種類の処理要求が出される場合、処理要求の種類が処理要求本体 3 0 1 に設定される。ヘッダ 3 0 2 中の 3 0 3 は優先度データであり、当該処理要求 3 0 0 の処理の優先度を数値で表現している。

10

【 0 0 3 0 】

アプリケーションプログラム 2 0 0 は処理要求を発行する際、処理要求本体 3 0 1 に優先度データ 3 0 3 を設定したヘッダ 3 0 2 を付加して処理要求 3 0 0 を生成し、処理要求分配手段 1 0 2 へ出力する。アプリケーションプログラム 2 0 0 から図 4 の処理要求本体 3 0 1 と優先度データ 3 0 3 の値とが渡されると、それらを使ってヘッダ 3 0 2 を持つ処理要求 3 0 0 を生成する機能を別途に設け、アプリケーションプログラム 2 0 0 がこの機能

20

【 0 0 3 1 】

優先ルール記憶手段 1 0 1 には、優先度データ 3 0 3 の値と実際の優先度との関係などを定める優先ルールが事前に記憶されている。一例を示せば、優先ルールは、優先度データ 3 0 3 の値が正の整数であれば高優先、正の整数以外（例えば 0）であれば低優先とするルールを規定する。この場合、処理要求分配手段 1 0 2 は、処理要求 3 0 0 中の優先度データ 3 0 3 の値が正の整数であれば、この処理要求 3 0 0 を優先キュー 1 0 3 に格納し、正の整数以外（例えば 0）であれば、当該処理要求 3 0 0 はこれを優先的に処理する必要のない処理要求と判断して通常キュー 1 0 4 へ格納する。他の例を示せば、優先ルールは、優先度データ 3 0 3 が閾値 以上であれば高優先、そうでなければ低優先とするルールを規定する。この場合、処理要求分配手段 1 0 2 は、処理要求 3 0 0 中の優先度データ 3 0 3 の値を閾値 と比較し、 以上の優先度データ 3 0 3 を持つ処理要求 3 0 0 は優先キュー 1 0 3 に格納し、 未満の優先度データ 3 0 3 を持つ処理要求 3 0 0 は通常キュー 1 0 4 へ格納する。

30

【 0 0 3 2 】

次に本実施例の動作を説明する。図 2 において、アプリケーションプログラム 2 0 0 が優先度データ 3 0 3 をヘッダ 3 0 2 に付与した処理要求 3 0 0 を投入すると、処理要求分配手段 1 0 2 は、優先度データ 3 0 3 の値と記憶手段 1 0 1 に記憶された優先ルールとに基づいて、優先的に処理すべき処理要求か否かを判断し、優先的に処理すべき処理要求については優先キュー 1 0 3 に格納し、優先的に処理すべき処理要求以外の処理要求については通常キュー 1 0 4 に格納する。キュー 1 0 3、1 0 4 へは処理要求 3 0 0 全体を格納しても良いし、処理要求本体 3 0 1 だけを格納するようにしても良い。

40

【 0 0 3 3 】

通常キュー 1 0 4 に処理要求 3 0 0 が格納されると、通常要求処理スレッド 1 0 6 はそのことを検出し（ステップ S 1）、通常キュー 1 0 4 から処理要求 3 0 0 を取り出し（ステップ S 2）、処理要求本体 3 0 1 の内容に従って予め自スレッドに定義された所定の処理を実行することにより当該処理要求 3 0 0 を処理する（ステップ S 3）。この処理の一環としてデータ記憶装置 1 0 7 が更新される。通常要求処理スレッド 1 0 6 は、1 つの処理要求 3 0 0 を処理し終わると、通常キュー 1 0 4 に他の処理要求 3 0 0 が既に格納されて

50

いれば、それを取り出して処理する動作を継続する。

【0034】

通常キュー104への処理要求300の投入量が通常要求処理スレッド106の単位時間当たりの処理量を上回ると、処理要求300が通常キュー104に徐々に溜まっていく。しかし、たとえこのような状況になっても、優先要求処理スレッド105は通常キュー104中の処理要求を処理することはなく、優先キュー103に処理要求300が格納されるのを待っている。そして、優先キュー103に処理要求300が格納されると、優先要求処理スレッド105はそのことを検出し(ステップS1)、優先キュー103から処理要求300を取り出し(ステップS2)、処理要求本体301の内容に従って予め自スレッドに定義された所定の処理を実行することにより、当該処理要求300を処理する(ステップS3)。この処理の一環としてデータ記憶装置107が更新される。優先要求処理スレッド105は、1つの処理要求300を処理し終わると、優先キュー103に他の処理要求300が既に格納されていれば、それを取り出して処理する動作を継続する。

10

【0035】

従って、各処理要求の処理量を同じと仮定した場合、優先要求処理スレッド105の単位時間当たりの処理要求300の処理量をW個とすると、優先キュー103への単位時間当たりの処理要求300の投入量がW個以下であれば、常に投入後直ちに、各処理要求300が処理されることになる。

【0036】

他方、優先キュー103への単位時間当たりの処理要求300の投入量がW個を超えると、優先キュー103に処理要求300が徐々に溜まっていくことになる。この場合の優先キュー103中の処理要求の処理の状況を、通常キュー104中の処理要求の処理の状況と対比して、以下説明する。

20

【0037】

先ず、優先要求処理スレッド105と通常要求処理スレッド106のスレッド優先度が同じ場合を考える。この場合、優先要求処理スレッド105と通常要求処理スレッド106にはCPU時間が等しく分配されるため、優先要求処理スレッド105と通常要求処理スレッド106の単位時間当たりの処理量は等しくなる。従って、優先キュー103中の処理要求が通常キュー104中の処理要求よりも優先的に処理されるには、優先的に処理すべき処理要求の単位時間当たりの発生量が、優先的に処理する必要のない処理要求の単位時間当たりの発生量の半分より少ないことが条件となる。つまり、アプリケーションプログラム200は、このような条件を満足するように、限定された場合にのみ優先的に処理すべき処理要求を発生することで、優先的に処理すべき処理要求をそうでない処理要求よりも優先的に処理してもらうことが可能である。

30

【0038】

図5に、優先的に処理すべき処理要求とそうでない処理要求とが1対3の割合で発生する場合の処理状況を示す。図5(a)は、アプリケーションプログラム200からの処理要求の発生順を示し、番号1、5、9を付したものが優先的に処理すべき処理要求、番号2~4、6~8を付したものがそれ以外の処理要求を示す。図5(b)は優先キュー103及び通常キュー104の状態を示し、優先キュー103には番号1、5、9等の優先的に処理すべき処理要求が発生順に格納され、通常キュー104には番号2~4、6~8等のそれ以外の処理要求が発生順に格納される。各処理要求の処理量が全て同じとすると、優先要求処理スレッド105及び通常要求処理スレッド106の単位時間当たりの処理量は前述したように等しくなるので、最終的な処理要求の処理順は図5(c)のようになる。番号5の処理要求は先に発生した番号3、4の処理要求よりも先に処理され、番号9の処理要求は先に発生した番号4等の処理要求よりも先に処理されており、優先的に処理すべき処理要求がそうでない処理要求よりも優先的に処理されているのがわかる。

40

【0039】

次に、優先要求処理スレッド105のスレッド優先度が通常要求処理スレッド106のスレッド優先度よりも高い場合を考える。この場合、優先要求処理スレッド105は通常要

50

求処理スレッド106に比べてCPU時間をより多く分配されるため、優先要求処理スレッド105の単位時間当たりの処理量は通常要求処理スレッド106のそれより多くなる。この為、スレッド優先度が同じ場合と異なり、優先的に処理すべき処理要求の単位時間当たりの発生量が、優先的に処理する必要のない処理要求の単位時間当たりの発生量の半分より多い場合であっても、優先キュー103中の処理要求が通常キュー104中の処理要求よりも優先的に処理される。優先的に処理すべき処理要求の単位時間当たりの発生量と、優先的に処理する必要のない処理要求の単位時間当たりの発生量との比がどの程度までなら、優先キュー103中の処理要求が通常キュー104中の処理要求よりも優先的に処理されるかは、優先要求処理スレッド105と通常要求処理スレッド106のスレッド優先度の差に応じて決まるCPUの分配時間の差で左右されることになる。

10

【0040】

次に、優先要求処理スレッド105のスレッド優先度が通常要求処理スレッド106のスレッド優先度よりも低い場合を考える。この場合、優先要求処理スレッド105は通常要求処理スレッド106に比べて分配されるCPU時間が少なくなるため、優先要求処理スレッド105の単位時間当たりの処理量は通常要求処理スレッド106のそれより少なくなる。この為、優先的に処理すべき処理要求の単位時間当たりの発生量が、優先的に処理する必要のない処理要求の単位時間当たりの発生量の半分より十分に少なくなければ、優先キュー103中の処理要求が通常キュー104中の処理要求よりも優先的に処理されなくなる。優先的に処理すべき処理要求の単位時間当たりの発生量と、優先的に処理する必要のない処理要求の単位時間当たりの発生量との比がどの程度までなら、優先キュー103中の処理要求が通常キュー104中の処理要求よりも優先的に処理されるかは、優先要求処理スレッド105と通常要求処理スレッド106のスレッド優先度の差に応じて決まるCPUの分配時間の差で左右されることになる。

20

【0041】

【適用例】

次に本実施例の適用例について図面を参照して詳細に説明する。

【0042】

ここでは、テレビ・ラジオにおける映像音声放送番組配信システムへの適用例を取りあげる。

【0043】

テレビ・ラジオにおける映像音声放送番組配信システムは、1つのチャンネルに流される各番組の番組枠及び番組枠の中でのCMの時間枠を管理する編成装置と、送出コンテンツを作成する制作装置と、指示された送出スケジュール情報に沿って送出データの送出を行う送出装置とで構成される。この映像音声放送番組配信システムは、プレイリストと呼ばれる送出スケジュール情報マップによって制御されている。ここで、送出スケジュール情報マップとは、編成装置や制作装置が作成したコンテンツを送出装置が正しく送出するための情報が記述されている表である。具体的には、或る番組枠を細かな時間枠に分割し、その時間枠毎の送出開始時刻、送出継続時間およびその時間枠に送出される送出コンテンツの属性値が設定される。プレイリストは、各番組内時間枠の送出開始時刻、送出継続時間、その時間枠が番組かCMかの種別、さらにその時間枠に送出される映像コンテンツおよび音声コンテンツのアドレス情報などが記述されているもので、編成装置が作成する。制作装置は、プレイリストに指定されたアドレスに送出する映像コンテンツおよび音声コンテンツを設定する。送出装置は、プレイリストを元に指定されたアドレスにあるコンテンツを制作装置から取得し、送出できる形に送出データに加工するといった送出準備処理を行い、指定された送出開始時刻になったならば準備した送出データを送出する。

30

40

【0044】

放送番組によっては、事前に設定された送出スケジュールを変更して放送する場合がある。この変更は、放送中の番組を対象とすることも、次以降に放送予定の番組を対象とすることもある。放送中の番組における送出スケジュールの変更を要求する処理要求は、次以降に放送される番組における送出スケジュールの変更を要求する処理要求よりも優先的に

50

処理する必要がある。この番組単位での送出スケジュールの変更の処理を番組データ変更処理と呼ぶ。一方、特定の1つの放送番組の送出スケジュールの変更にとどまらず、後の放送番組の送出スケジュールも変更しなければならない処理がある。これを編成変更処理と呼ぶ。この編成変更処理は、例えばスポーツ中継などの試合の延長や早終わりといった場合に必要となり、放送中の番組に対する番組データ変更処理と同様に優先的に処理する必要がある。何れも、プレイリストを変更することによって送出スケジュールの変更を送出装置に伝達する。

【0045】

図6にプレイリストの概略を示す。図6において、番組データ(1)~(n)のそれぞれが個々の1つの放送番組に関する送出スケジュール情報を記述している。本実施例では、図2のデータ記憶装置107がこのような送出スケジュール情報を記憶している。ここで、個々の番組データ(1)~(n)が1つの排他単位になっている。

10

【0046】

アプリケーションプログラム200は、送出スケジュール情報の変更要求を処理要求として発行する放送業務アプリケーションプログラムに相当する。アプリケーションプログラム200は、処理要求を送出する際、図4の処理要求本体301に、番組データ変更処理、編成変更処理等の処理種別を設定し、また、番組データ変更処理または編成変更処理における変更後の送出スケジュール情報等を設定する。そして、優先度データ303をヘッダ302に付与した処理要求300を投入する。優先度データ303としては、例えば、放送中の番組に対する番組データ変更処理および編成変更処理の場合には正の整数、それ

20

【0047】

処理要求分配手段102は、優先ルール記憶手段101に記憶された優先ルールに従って優先度データ303の値が正の整数か否かによって優先的に処理すべき処理要求か否かを判断し、優先的に処理すべき処理要求については優先キュー103に格納し、優先的に処理すべき処理要求以外の処理要求については通常キュー104に格納する。従って、放送中の番組に対する番組データ変更処理にかかる処理要求および編成変更処理にかかる処理要求は優先キュー103に格納され、次以降の番組に対する番組データ変更処理などの処理要求は通常キュー104に格納される。これにより、放送中の番組に対する番組データ変更処理および編成変更処理がそれ以外の処理より優先して処理される。

30

【0048】

また、現在放送中の番組のデータを図6の例えば番組データ(2)とすると、通常要求処理スレッド106が実施する番組データ変更処理は番組データ(3)以降の処理に限られる。従って、通常要求処理スレッド106が番組データ変更処理のために番組データ(3)~(n)の何れかを排他的にロックしている最中に、放送中の番組の番組データ(2)の変更にかかる処理要求が発生しても、排他単位が異なるため、優先要求処理スレッド105は通常要求処理スレッド106に影響されずに当該処理要求の処理を進めることが可能である。

【0049】

【第2の実施例】

図7を参照すると、本実施例にかかる情報処理システムにおける優先実行制御方法を実施するコンピュータ100は、処理要求分配手段102からデータ記憶装置107を参照できる点が、図2の第1の実施例と相違する。データ記憶装置107にはスレッド105、106の処理結果が格納されるため、処理要求分配手段102からデータ記憶装置107が参照可能になると、これまでの処理結果に応じて処理要求分配手段102の処理要求の分配の仕方を動的に変更するといったことが可能となる。

40

【0050】

例えば、優先ルール記憶手段101に記憶された優先ルールが、優先度データ303が閾値以上であれば高優先、そうでなければ低優先とするルールである場合、第1の実施例では、当該閾値は固定値となるが、本実施例では、当該閾値をデータ記憶装置107

50

上で処理の進行に応じて変化する変数とすることにより、現在までの処理結果に応じて優先度の高低を切りわけると動的に設定することが可能となる。

【0051】

【適用例】

次に本実施例の適用例について図面を参照して詳細に説明する。

【0052】

ここでは、1万人分のユーザの得点データを処理して成績上位100名のユーザ名とその得点の一覧を生成するクイズアプリケーションへの適用例を取りあげる。この場合、アプリケーションプログラム200はクイズアプリケーションプログラムであり、テレビ番組で放送されているクイズに対して視聴者(ユーザ)が携帯電話等の通信装置を通じて送信してきた回答に対して採点されたデータ(ユーザ名とその得点データ)を含むパケットを受け取る毎に、図4の処理要求本体301にユーザ名とその得点データとを設定し、ヘッダ302の優先度データ303に当該得点データそのものを設定した処理要求300を処理要求分配手段102へ投入する処理を、1万人分のユーザ毎に繰り返す。

10

【0053】

また、データ記憶装置107には、図8に示すように上位データ400及び通知データ410というデータがそれぞれ独立したデータ領域に設定される。上位データ400は、成績上位100名のユーザ名とその得点を記録する上位成績者一覧表401と、上位成績者一覧表401に記録された得点のうちの最低点402と、上位成績者一覧表401に記録されたユーザの数であるユーザ数403とを含む。最低点402の初期値は満点以上の或る値、ユーザ数の初期値は0であり、上位成績者一覧表401の更新に応じて適宜変更される。また、通知データ410は、足きりの得点である足きり点(S)411を記録する。足きり点411の初期値は0であり、上位成績者一覧表401の更新に応じて適宜変更される。データ記憶装置107を更新する場合、上位データ400単位、通知データ410単位で排他制御を行う。ある者が更新のために排他されていても他の者は参照は自由に行える。

20

【0054】

さらに、優先要求処理スレッド105および通常要求処理スレッド106の各々には、図8に示すように各スレッド105、106独自のメモリ領域内に下位成績者表420、430がそれぞれ設定される。下位成績者表420、430は、成績上位100名に入らなかったユーザのユーザ名とその得点を記録する表である。

30

【0055】

優先要求処理スレッド105及び通常要求処理スレッド106は、自スレッドに対応するキュー103、104から取り出した処理要求に対する図3のステップS3における処理では、図9に示す処理を行う。

【0056】

まず、上位データ400のユーザ数403が100以上か否かを調べる(ステップS11)。100未満であれば、処理要求中のユーザ名とその得点を上位データ400の上位成績者一覧表401に追加し(ステップS12)、ユーザ数403を+1すると共に、今回追加したユーザの得点が最低点402より低ければ、今回追加したユーザの得点で最低点402を更新する(ステップS13)。そして、ユーザ数403が100以上になったか否かを調べ(ステップS14)、100以上でなければ今回の処理を終了し、100以上であれば、通知データ410の足きり点411を上位データ400の最低点402で更新する(ステップS15)。

40

【0057】

他方、上位データ400のユーザ数403が100以上であれば(ステップS11でYES)、今回の処理要求中の得点と最低点402とを比較し(ステップS16)、最低点402以上の得点であれば、今回の処理要求中のユーザ名と得点を上位成績者一覧表401に追加する(ステップS17)。そして、ユーザ数403から上位成績者一覧表401における最低点のユーザ数の合計を差し引いた値が99であるか否かを調べ(ステップS1

50

8)、99であれば、最低点の全てのユーザの情報(ユーザ名と得点)を上位成績者一覧表401から自スレッドの下位成績者表420または430に移動させ(ステップS19)、最低点402を更新後の上位成績者一覧表401における新たな最低点で更新すると共に、ユーザ数403を100に更新する(ステップS20)。また、足きり点411を更新後の最低点402で更新する(ステップS21)。ステップS18の判定結果がNOであれば、ユーザ数403を+1する(ステップS24)。

【0058】

今回の処理要求中の得点が最低点402以上でなければ、最低点402と同得点か否かを判別する(ステップS22)。同得点であれば、今回の処理要求中のユーザ名と得点を上位成績者一覧表401に追加し(ステップS23)、ユーザ数403を+1する(ステップS24)。

10

【0059】

今回の処理要求中の得点が最低点402未満であれば(ステップS22でNO)、今回の処理要求中のユーザ名と得点を自スレッドの下位成績者表420または430に追加する(ステップS25)。

【0060】

処理要求分配手段102は、処理要求300に付随する優先度データ303で与えられる得点をデータ記憶装置107に記憶されている足きり点411と比較し、足きり点411に満たない得点を持つ処理要求300は通常キュー104に格納し、足きり点411以上の得点を持つ処理要求300は優先キュー103に格納する。

20

【0061】

次に本適用例の動作を説明する。アプリケーションプログラム200は、ユーザ名とその得点データを含むパケットを受け取ると、ヘッダ302の優先度データ303に当該得点データそのものを設定し、処理要求本体301にユーザ名と得点とを設定した処理要求300を処理要求分配手段102へ投入する。処理要求分配手段102は、処理要求300の優先度データ303で示される得点とデータ記憶装置107の足きり点411とを比較して当該処理要求の優先度を判定する。足きり点411は当初は初期値の0になっているので、処理要求分配手段102は、当初は全ての処理要求300を優先的に処理すべき処理要求と判断して、優先キュー103へ格納することになる。

【0062】

30

優先要求処理スレッド105は、優先キュー103から処理要求を順次に取り出して図9に示した処理を実行する。当初はユーザ数403が100未満なので、処理要求中のユーザ名と得点を上位成績者一覧表401に追加し、最低点402を必要に応じて更新し、ユーザ数403を+1する処理が行われる(ステップS12、S13)。そして、100個目の処理要求を処理し終えた時点で、それまでの最低点を足きり点411に設定する(ステップS15)。

【0063】

足きり点411が0点以外の或る得点Xに更新されると、以後、処理要求分配手段102は、得点Xに満たない得点を持つ処理要求300は通常キュー104へ格納し、得点Xを超える得点を持つ処理要求300だけを優先キュー103へ格納するようになる。但し、優先キュー103による遅延が介在するため、得点の低い古い足きり点411に基づいて処理要求が優先キュー103に分配される可能性がある。若し、このような処理要求が優先キュー103に格納されると、その処理要求の得点は最低点402未満であるため、優先要求処理スレッド105のステップS25において、当該処理要求中のユーザ名と得点が下位成績者表420に記録されることになる。

40

【0064】

最低点402以上の得点を持つ処理要求の場合、優先要求処理スレッド105は次のように処理する。まず、最低点402を超える得点を持つ処理要求の場合(ステップS16でYES)、今回の処理要求のユーザ名と得点を上位成績表一覧401に追加する(ステップS18)。最低点を持つユーザが上位100名に入らなくなった場合には(ステップS

50

18でYES)、最低点を持つユーザのユーザ名と得点を上位成績表一覧401から下位成績者表420に移し(ステップS20)、最低点402を更新後の上位成績表一覧401における最低点で更新すると共にユーザ数403を100に更新し(ステップS20)、足り点411を更新後の最低点402で更新する(ステップS21)。他方、最低点402と同じ得点を持つ処理要求の場合(ステップS22でYES)、今回の処理要求のユーザ名と得点を上位成績表一覧401に追加し(ステップS23)、ユーザ数403を+1する(ステップS24)。

【0065】

一方、通常要求処理スレッド106は、通常キュー104から処理要求を取り出し、図9の処理を実行する。但し、通常キュー104に処理要求が格納される時点では、ユーザ数403は100以上になっており、また、最低点402以上の得点を持つ処理要求は通常キュー104には格納されない。このため、通常要求処理スレッド106が実行する処理は、処理要求中のユーザ名と得点を自スレッドの下位成績者表430に記録する処理(ステップS25)だけとなり、データ記憶装置107の上位データ400及び通知データ410への更新、そのための排他ロックを行う場面はない。従って、優先要求処理スレッド105は通常要求処理スレッド106の処理の影響を受けずに自らの処理を進めることができる。このため、成績上位100名のユーザ名と得点をディスクなどのデータ記憶装置107に記録する処理を効率良く実施することが可能となる。

【0066】

【第3の実施例】

図10を参照すると、本実施例にかかる情報処理システムにおける優先実行制御方法を実施するコンピュータ100は、処理要求分配手段102からアプリケーションプログラム200に対して任意のデータを状態通知として通知する状態通知手段108を備えている点で、図7の第2の実施例と相違する。

【0067】

このような状態通知手段108を処理要求分配手段102とアプリケーションプログラム200との間に設けることによって、例えば、アプリケーションプログラム200が投入した処理要求の結果などを迅速にアプリケーションプログラム200に通知することができる。第2の実施例の適用例を例にすれば、データ記憶装置107に作成される成績上位100名の上位成績者一覧表401をアプリケーションプログラム200に通知することができる。そして、以下のようにすれば、上位成績者一覧表401をその完成後に直ちにアプリケーションプログラム200に通知することができる。

【0068】

先ず、処理要求分配手段102は、優先キュー103及び通常キュー104に格納した処理要求300の数をカウントし、カウント値が1万に達すると、優先キュー103の状態を監視する。そして、優先キュー103に1つも処理要求300が存在しなくなった時点より、優先要求処理スレッド105が1つの処理要求300の処理にかかる時間だけ経過した時点で、データ記憶装置107から上位成績者一覧表401を読み出し、状態通知手段108を通じてアプリケーションプログラム200に通知する。この時点で通常キュー104に処理要求が幾ら溜まっても関係はない。それらが成績上位100名に入らないことが既に確定しているからである。

【0069】

【第4の実施例】

図11を参照すると、本実施例にかかる情報処理システムにおける優先実行制御方法を実施するコンピュータ100は、優先要求処理スレッド105からアプリケーションプログラム200に対してデータ記憶装置107に記録された任意のデータを状態通知として通知する状態通知手段109を備え、通常要求処理スレッド106からアプリケーションプログラム200に対してデータ記憶装置107に記録された任意のデータを状態通知として通知する状態通知手段110を備えている点で、図10の第3の実施例と相違する。

【0070】

10

20

30

40

50

このような状態通知手段109、110を設けることによって、優先キュー103中の処理要求を優先要求処理スレッド105が処理した結果の途中経過と、通常キュー104中の処理要求を通常要求処理スレッド106が処理した結果の途中経過とを、それぞれ独立にアプリケーションプログラム200に通知することができる。第2の実施例の適用例を例にすれば、優先要求処理スレッド105は、データ記憶装置107上で時々刻々と変化する上位成績者一覧表401の内容をリアルタイムにアプリケーションプログラム200に通知することができる。また、必要によっては下位成績者表420の途中結果もアプリケーションプログラム200に通知することができる。更に、通常要求処理スレッド106は、下位成績表430の途中結果を状態通知手段110を通じてアプリケーションプログラム200に通知することができる。

10

【0071】**【第5の実施例】**

図12を参照すると、本実施例にかかる情報処理システムにおける優先実行制御方法を実施するコンピュータ100は、キューとそれに対応する要求処理スレッドとをアプリケーションプログラム200からの指示に従って生成し、また、既に生成されているキューとそれに対応する要求処理スレッドとをアプリケーションプログラム200からの指示に従って消滅させるスレッド割付手段111を備えている点で、図2の第1の実施例と相違する。

【0072】

スレッド割付手段111は、キューとスレッドの組の生成時には、アプリケーションプログラム200から、優先、通常の区別と、スレッドの優先度の指定を受け取り、優先の区別が指定された場合には、優先キュー103と、指定されたスレッド優先度を持つ優先要求処理スレッド105を生成し、優先キュー103の存在場所（アドレス）を優先ルール記憶手段101に設定する。また、通常の区別が指定された場合には、通常キュー104と、指定されたスレッド優先度を持つ通常要求処理スレッド106を生成し、通常キュー104の存在場所（アドレス）を優先ルール記憶手段101に設定する。

20

【0073】

また、スレッド割付手段111は、キューとスレッドの組の消滅時には、アプリケーションプログラム200から、優先、通常の区別の指定を受け取り、優先の区別が指定された場合には、優先キュー103と優先要求処理スレッド105を消滅させ、優先ルール記憶手段101中の優先キュー103の存在場所（アドレス）を消去する。また、通常の区別が指定された場合には、通常キュー104と通常要求処理スレッド106を消滅させ、優先ルール記憶手段101中の通常キュー104の存在場所（アドレス）を消去する。

30

【0074】

処理要求分配手段102は、優先キュー103及び通常キュー104のアドレスを優先ルール記憶手段101から参照して、アプリケーションプログラム200から入力される処理要求を格納するキューの位置を認識する。優先キュー103がなく、通常キュー104だけ存在する場合には、全ての処理要求は通常キュー104へ格納する。同様に、通常キュー104がなく、優先キュー103だけ存在する場合には、全ての処理要求は優先キュー103へ格納する。優先キュー103及び通常キュー104の双方が存在しない状態で処理要求が入力されるとエラーとなる。

40

【0075】

このようなスレッド割付手段111を備えることにより、アプリケーションプログラム200は、通常キュー104と通常要求処理スレッド106を生成すると同時に優先キュー103と優先要求処理スレッド105を生成すれば、優先実行制御を利用することが可能となり、その後、優先実行制御が不要になれば、優先キュー103と優先要求処理スレッド105とを消滅させることにより、資源の無駄を無くすることができる。

【0076】

第1の実施例の放送業務アプリケーションの適用例では、放送中に番組に関する変更が生じないことが自明である場合には、アプリケーションプログラム200はスレッド割付手

50

段111を使って優先キュー103と優先要求処理スレッド105とを消滅させ、或る番組で放送中に何らかの変更が発生することが分かっている場合には、同様にスレッド割付手段111を使って優先キュー103と優先要求処理スレッド105を生成すれば、優先実行を必要なときにのみ利用することができる。また、アプリケーションプログラム200が終了する際に、スレッド割付手段111を使って、優先キュー103と優先要求処理スレッド105及び通常キュー104と通常要求処理スレッド106を消滅させることも可能である。

【0077】

【第6の実施例】

図13を参照すると、本発明の第6の実施例は、アプリケーションプログラム200からの指示に従って優先ルール記憶手段101に記憶された優先ルールを変更する優先ルール変更手段112と、アプリケーションプログラム200からの指示に従って優先要求処理スレッド105及び通常要求処理スレッド106のスレッド優先度を変更するスレッド優先度変更手段113と、処理要求分配手段102からアプリケーションプログラム200に対して任意のデータを状態通知として通知する状態通知手段108とを備えている点で、図7の第2の実施例と相違する。

【0078】

優先ルール変更手段112は、優先ルールの変更時、アプリケーションプログラム200から変更後の優先ルールを受け取り、受け取った優先ルールで優先ルール記憶手段101に記憶されている優先ルールを上書きする機能を有する。

【0079】

スレッド優先度変更手段113は、スレッド優先度の変更時、アプリケーションプログラム200から、優先、通常の区分と、変更後のスレッド優先度とを受け取り、優先要求処理スレッド105及び通常要求処理スレッド106のうち区分で指定された側のスレッドの優先度を、指定された優先度に変更する機能を有する。

【0080】

状態通知手段108は、優先キュー103及び通常キュー104の状態や、データ記憶装置107に記憶されているデータなど任意のデータを処理要求分配手段102からアプリケーションプログラム200に通知する機能を有する。

【0081】

本実施例はこのような構成を有するため、アプリケーションプログラム200は、以下のような各種の優先実行制御方法をシステムの運用中に動的に切り替えることが可能である。

(1) 優先要求処理スレッド105のスレッド優先度を通常要求処理スレッド106のそれより低く設定した優先実行制御。

(2) 優先要求処理スレッド105のスレッド優先度を通常要求処理スレッド106のそれと同じに設定した優先実行制御。

(3) 優先要求処理スレッド105のスレッド優先度を通常要求処理スレッド106のそれより高く設定した優先実行制御。

【0082】

また、前記(1)~(3)のそれぞれの優先実行制御に際して使用する優先ルールをシステム運用中に動的に切り替えることができる。

【0083】

【適用例】

次に本実施例の適用例について説明する。ここでは、第1の実施例における適用例と同様に放送業務アプリケーションを取りあげる。

【0084】

今、図6における番組データ(1)~(n)のうち、例えば番組データ(3)の放送番組だけが、特にその放送中に当該番組データ(3)を変更する頻度が高く、それ以外の番組データ(1)、(2)、(4)~(n)についてはその放送中に当該番組データ(1)、

10

20

30

40

50

(2)、(4)～(n)を変更することがないか、あっても頻度が極めて小さいことが事前に分かっているものとする。このため、番組データ(3)以外の番組の放送中は、現放送中の番組にかかる番組データ変更処理及び編成変更処理をその他の処理より優先し、番組データ(3)の番組の放送中は、当該番組データ(3)の変更処理を編成変更処理を含めてその他の処理より優先したいものとする。

【0085】

このような場合、アプリケーションプログラム200は、番組データ(3)の番組が放送されるまでは、スレッド優先度変更手段113により優先要求処理スレッド105のスレッド優先度を通常処理スレッド106のスレッド優先度と同じか、低く設定する。また、アプリケーションプログラム200は、優先度データ303として放送中の番組に対する番組データ変更処理および編成変更処理の場合には正の整数、それ以外は0を設定するものとする。優先ルール変更手段112により、例えば、優先度データ303の値が正の整数なら高優先、それ以外なら低優先とする優先ルールを優先ルール記憶手段101に設定する。

10

【0086】

こうすると、アプリケーションプログラム200から処理要求分配手段102に出力される処理要求のうち、現放送中の番組に対する番組データ変更処理と編成変更処理の要求が優先キュー103を通じて優先要求処理スレッド105へ与えられ、それ以外の処理要求は通常キュー104を通じて通常要求処理スレッド106へ与えられる。第1の実施例で説明したように、優先的に処理すべき処理要求の発生量がそれ以外の処理要求の発生量に比べて十分に少ない場合には、優先要求処理スレッド105のスレッド優先度が通常要求処理スレッド106のスレッド優先度と同じか低くても、優先的に処理すべき処理要求を優先的に処理することが可能である。つまり、今の例では、現放送中の番組に対する番組データ変更処理と編成変更処理の要求をその他の処理要求より優先的に処理できる。他方、優先的に処理すべき処理要求の発生量が少ないときに優先要求処理スレッド105のスレッド優先度を通常要求処理スレッド106より高くしておく、その分だけ通常要求処理スレッド106の処理を無駄に圧迫することになるが、この例のように優先要求処理スレッド105のスレッド優先度を通常要求処理スレッド106と同じか低く設定しておくことで、通常要求処理スレッド106の処理能力を相対的に高めることができ、優先的に処理する必要のない処理要求の発生量が多い場合であっても迅速に処理することが可能となる。

20

30

【0087】

次にアプリケーションプログラム200は、番組データ(3)の番組の開始直前に、スレッド優先度変更手段113により優先要求処理スレッド105のスレッド優先度を通常処理スレッド106のスレッド優先度より高く設定する。また、アプリケーションプログラム200は、優先ルール変更手段112により、例えば、優先度データ303の値が正の整数かつ処理種別が番組データ変更処理なら高優先、それ以外なら低優先とする優先ルールを優先ルール記憶手段101に設定する。

【0088】

こうすると、アプリケーションプログラム200から処理要求分配手段102に出力される処理要求のうち、現放送中の番組に対する番組データ変更処理にかかる処理要求だけが優先キュー103を通じて優先要求処理スレッド105へ与えられ、番組編成変更要求を含むそれ以外の処理要求は通常キュー104を通じて通常要求処理スレッド106へ与えられる。このため、現放送中の番組に対する番組データ変更処理にかかる処理要求だけを優先的に実行させることが可能となる。

40

【0089】

次にアプリケーションプログラム200は、番組データ(3)の番組が終了すると、スレッド優先度および優先ルールを元の状態に戻す。つまり、スレッド優先度変更手段113により優先要求処理スレッド105のスレッド優先度を通常処理スレッド106のスレッド優先度と同じか、低く設定し、優先ルール変更手段112により、優先度データ303

50

の値が正の整数なら高優先、それ以外なら低優先とする優先ルールを優先ルール記憶手段101に設定する。

【0090】

なお、アプリケーションプログラム200は、処理要求分配手段102から状態通知手段108を通じて優先キュー103及び通常キュー104の状態を定期的に収集し、若し、現状のスレッド優先度、優先ルールでは支障が生じる場合にはスレッド優先度変更手段113によってスレッド優先度を適宜変更し、また優先ルール変更手段112によって優先ルールを適宜変更する。例えば、番組データ(1)、(2)の番組の放送中に、事前の予想に反して現放送中の番組データ(1)、(2)に対する変更要求の発生量が増え、優先キュー103に処理要求が滞っていることが判明すると、優先要求処理スレッド105の

10

【0091】

【第7の実施例】

図14を参照すると、本発明の第7の実施例は、クライアント500とサーバ502とがネットワーク501を介して接続されたクライアント・サーバ型の情報処理システムにおいて、クライアント500からネットワーク501経由でサーバ502に送られてくる処理要求を処理する処理部503及び処理部503で得られた応答結果をネットワーク501経由で要求元のクライアント500に返却する応答部504に対して、本発明を適用したものである。即ち、処理部503は、クライアント500からの処理要求の処理に

20

【0092】

処理部503及び応答部504に対して第1の実施例の優先実行制御を適用した例を図15に示す。

【0093】

図15を参照すると、処理部503の受信手段505は、ネットワーク501を介してクライアント500から処理要求を受信すると、処理要求分配手段102に伝達する。処理要求分配手段102は、優先ルール記憶手段101の優先ルールに沿ってその処理要求の

30

【0094】

処理部503内の優先要求処理スレッド105及び通常要求処理スレッド106は、優先キュー103及び通常キュー104から処理要求を取り出し、データベース等のデータ記憶装置107を参照、更新して当該処理要求を処理する。そして、処理結果を要求元のクライアント500に返却する応答処理の要求を応答部504の処理要求分配手段102へ出力する。

40

【0095】

応答部504の処理要求分配手段102は、処理部503から渡された応答にかかる処理要求の優先度を、応答部504内の優先ルール記憶手段101に設定された優先ルールに基づき判別し、優先して処理すべき処理要求は優先キュー103へ、そうでない処理要求

50

は通常キュー 104 へ格納する。優先ルールとしては、例えば、個々のクライアント 500 毎のクライアント名に対応した優先度と優先度の高低を切り付ける閾値とを含むルールが考えられる。この場合、処理要求分配手段 102 は、処理要求元のクライアント名に対応する優先度を優先ルールから取得し、それを閾値と比較することで優先キュー 103、通常キュー 104 への振り分けを行う。他の優先ルールとして、応答結果の種別に応じた優先度と高低の優先度を切り付ける閾値とを含むルールが考えられる。この場合、処理部 504 の各スレッド 105、106 は、応答結果の処理要求にその応答結果の種別に応じた優先度を設定したヘッダを付随させ、応答部 504 内の処理要求分配手段 102 はヘッダに設定された優先度と優先ルールに基づいて処理要求を優先キュー 103、通常キュー 104 に振り分ける。応答結果の種別に応じて優先度を設定する例としては、何らかのエラーが発生して処理要求が正しく処理されなかったことをクライアント 500 に伝えるエラー応答を正常応答より優先度を高く設定することが考えられる。

10

【0096】

【第 8 の実施例】

図 16 を参照すると、第 8 の実施例は、CPU 601、それに接続された通信インタフェース 602、主メモリ 603 及び磁気ディスク装置 604 で構成されるコンピュータ 600 と、記録媒体 605 とで構成される。記録媒体 605 は、半導体メモリ、CD-ROM、磁気ディスク等の機械読み取り可能な記録媒体であり、優先実行制御プログラムが記録されている。ここに記録された優先実行制御プログラムは、コンピュータ 600 に読み取られ、コンピュータ 600 の動作を制御することにより、コンピュータ 600 上に前述した第 1 乃至第 7 の実施例の優先実行制御に必要な機能手段を実現する。

20

【0097】

【発明の効果】

以上説明したように本発明によれば、処理要求をその処理の優先度に応じて優先キュー及び通常キューの何れか一方に格納し、優先キューに格納された処理要求はそれ専門に処理する優先処理手段で取り出して処理する構成を有するため、優先的に処理すべき処理要求をできるだけ速やかに処理することが可能になる効果がある。

【0098】

優先キュー及びそれに対応するスレッドを動的に生成でき、また消滅させることができるため、資源の無駄がなくなり、また必要な場合にのみ資源を確保した優先実行制御を実施することができる。

30

【0099】

各キューに対応するスレッドのスレッド優先度を動的に変更できるため、優先的に処理すべき処理要求とそれ以外の処理要求との発生量の差に応じた柔軟な制御が可能となる。

【0100】

優先キュー及び通常キューに処理要求を分配する際の優先度判定のために参照される優先ルールを動的に変更できるため、優先的に処理すべき処理要求とそれ以外の処理要求との切り分けを動的に変更でき、柔軟な制御が可能となる。

【図面の簡単な説明】

【図 1】本発明の一実施の形態を適用した情報処理システムにおける優先実行制御装置のブロック図である。

40

【図 2】本発明の第 1 の実施例のブロック図である。

【図 3】優先要求処理スレッド及び通常要求処理スレッドの処理の概略を示すフローチャートである。

【図 4】処理要求のフォーマット例を示す図である。

【図 5】優先的に処理すべき処理要求とそうでない処理要求とが 1 対 3 の割合で発生する場合の本発明による処理状況を示す図である。

【図 6】映像音声放送番組配信システムにおけるプレイリストの概略を示す図である。

【図 7】本発明の第 2 の実施例のブロック図である。

【図 8】本発明の第 2 の実施例の適用例においてデータ記憶装置に記憶されるデータ構造

50

の例を示す図である。

【図9】本発明の第2の実施例の適用例において優先要求処理スレッド及び通常要求処理スレッドが実施する処理例を示すフローチャートである。

【図10】本発明の第3の実施例のブロック図である。

【図11】本発明の第4の実施例のブロック図である。

【図12】本発明の第5の実施例のブロック図である。

【図13】本発明の第6の実施例のブロック図である。

【図14】本発明の第7の実施例のブロック図である。

【図15】本発明の第7の実施例のより具体的なブロック図である。

【図16】本発明の第8の実施例のブロック図である。

【図17】従来技術の説明図である。

【符号の説明】

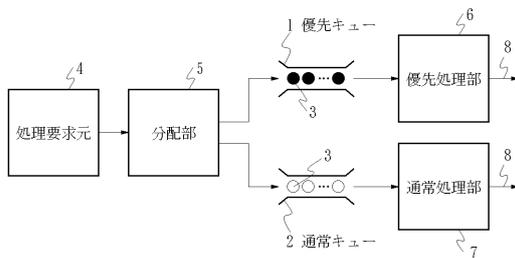
- 1 ... 優先キュー
- 2 ... 通常キュー
- 3 ... 処理要求
- 4 ... 処理要求元
- 5 ... 分配部
- 6 ... 優先処理部
- 7 ... 通常処理部
- 8 ... 処理結果

10

20

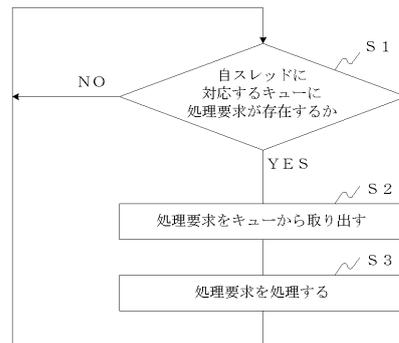
【図1】

【図1】



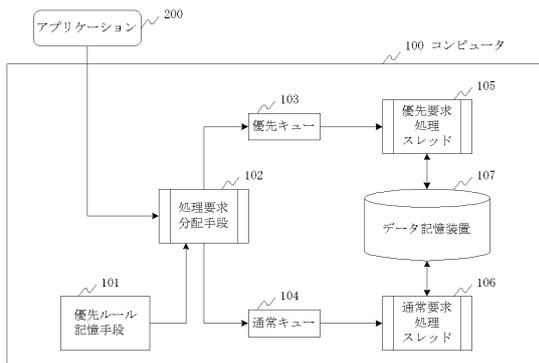
【図3】

【図3】



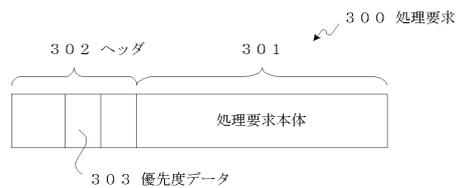
【図2】

【図2】



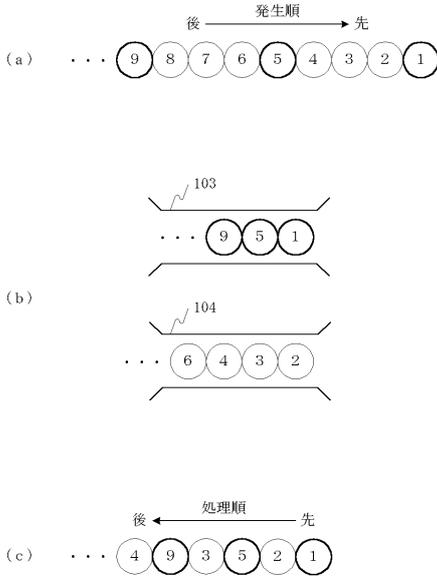
【図4】

【図4】



【図5】

【図5】



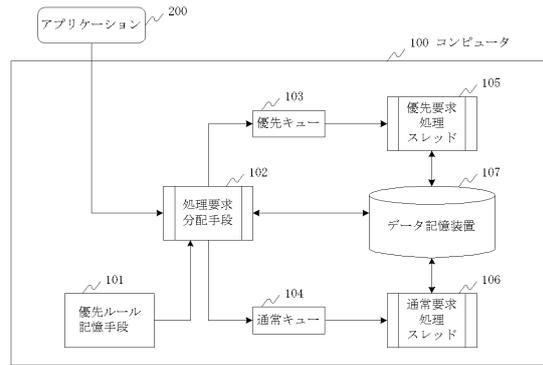
【図6】

【図6】



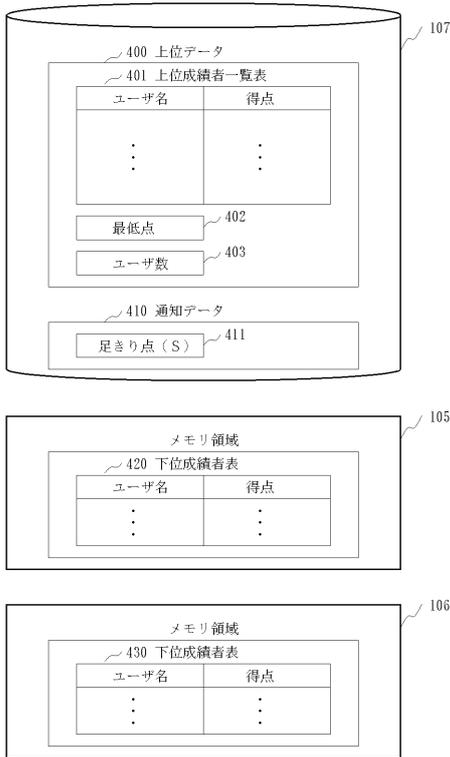
【図7】

【図7】



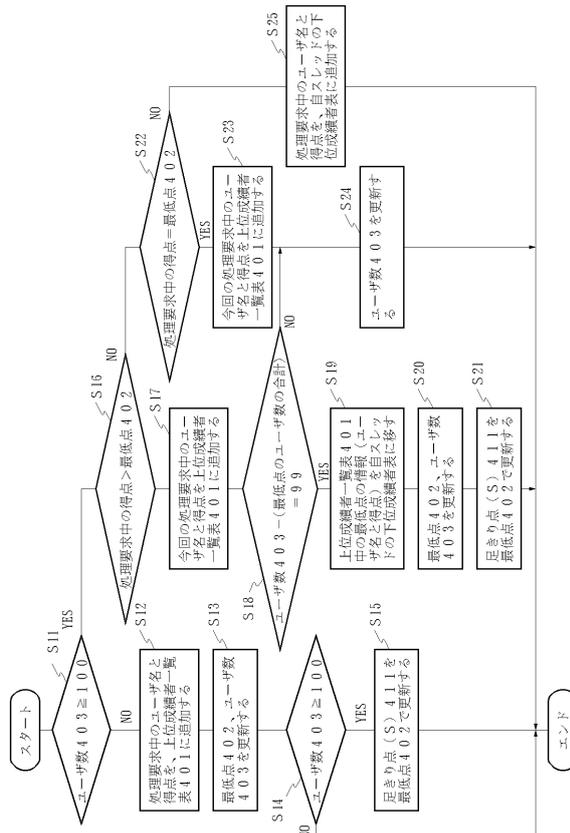
【図8】

【図8】



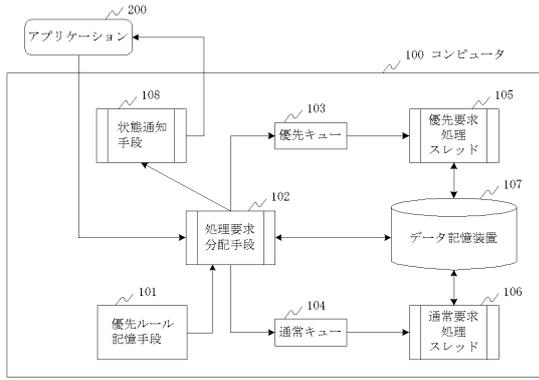
【図9】

【図9】



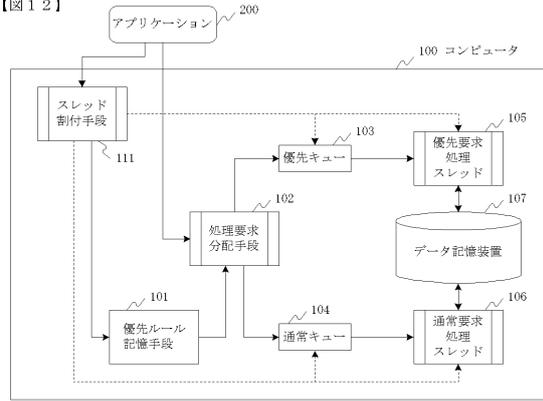
【図10】

【図10】



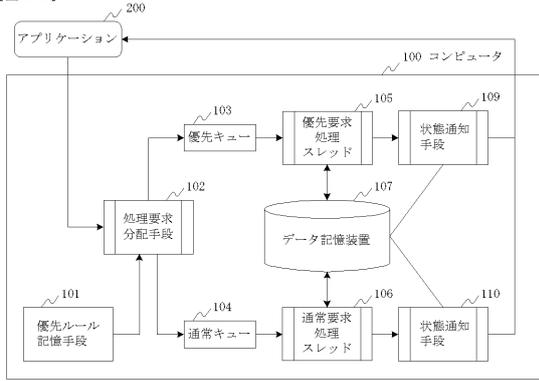
【図12】

【図12】



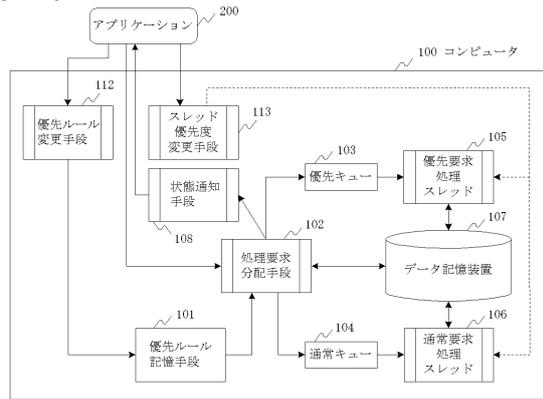
【図11】

【図11】



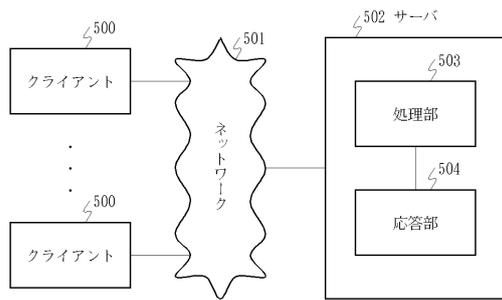
【図13】

【図13】



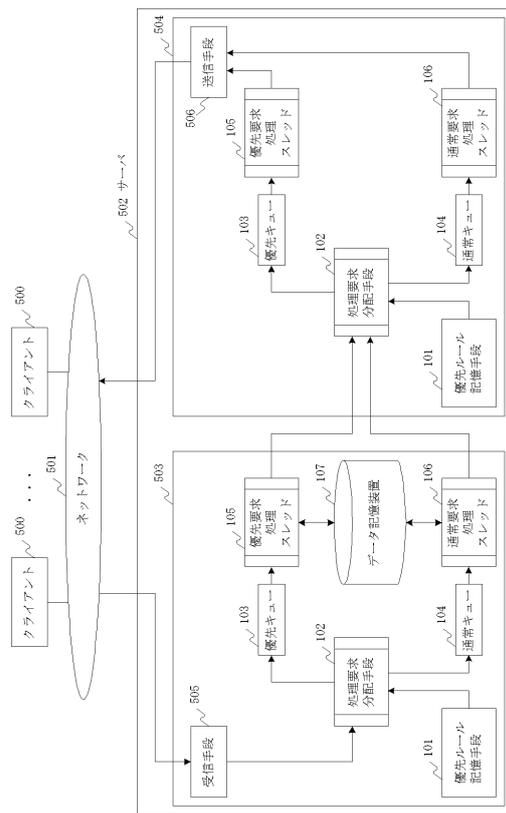
【図14】

【図14】



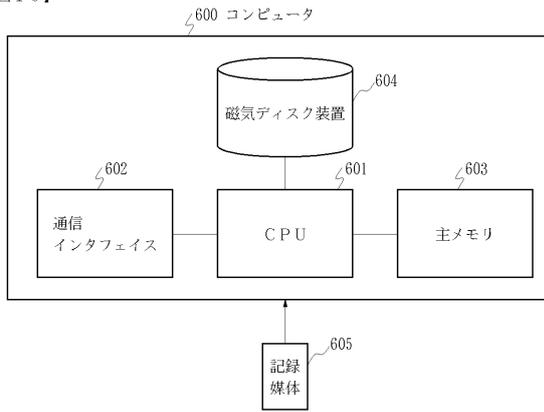
【図15】

【図15】



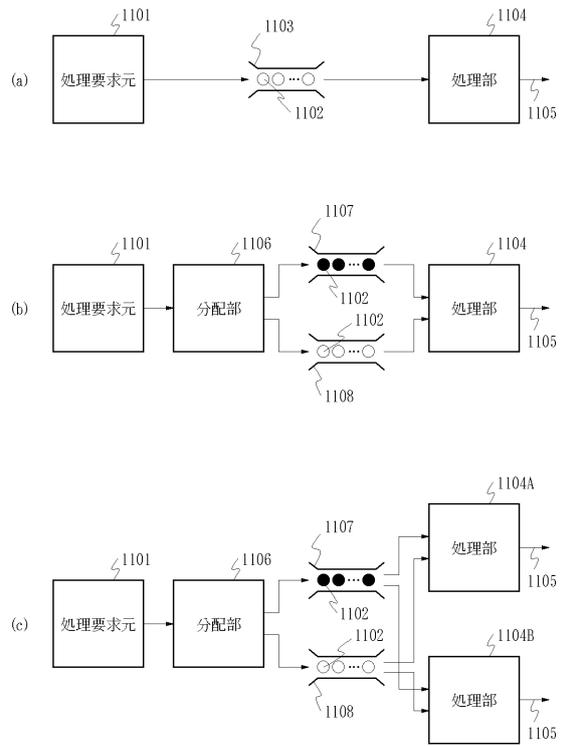
【図16】

【図16】



【図17】

【図17】



フロントページの続き

(56)参考文献 特開平09-034730(JP,A)

特開平11-237993(JP,A)

OpenTP1 Version5 分散トランザクション処理機能 OpenTP1 解説
解説書 共通マニュアル 第4版,日本,株式会社日立製作所,2001年 3月31日,第4
版,頁:102

(58)調査した分野(Int.Cl.,DB名)

G06F 9/46-9/54