

DECLARATION OF HERMAN KAHN

I, HERMAN KAHN, pursuant to 28 U.S.C. § 1746, hereby declare as follows:

1. I am a freelance translator for TransPerfect, Inc.
2. I submit this declaration to certify the accuracy of the English translation of WO2010050080A1.
3. My statements are based on personal knowledge and my review of WO2010050080A1 and its Japanese to English translation. If called as a witness about the facts contained in these statements, I could testify competently based on such personal knowledge and the investigation I have conducted.
4. Attached as Exhibit A is a true and accurate copy of WO2010050080A1.
5. Attached as Exhibit B is a true and accurate copy of an English translation WO2010050080A1.
6. WO2010050080A1 Translation is a true and accurate translation of WO2010050080A1 from Japanese to English.
7. All statements made herein of my own knowledge are true, and all statements made on information and belief are believed to be true. Further, I am aware that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001. I declare under penalty of perjury that the foregoing is true and correct.
8. I also understand that by submitting this declaration I may be asked to appear for a deposition asking me questions limited to the material in my declaration. With my signature below, I agree to make reasonable efforts to make myself available for such a deposition at a reasonable place and time of my choosing.

* * *

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge. Executed on December 18, 2024 at Woodland, California.

A handwritten signature in black ink, appearing to read 'Herman Kahn', with a long horizontal flourish extending to the right.

HERMAN KAHN

Exhibit A

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2010年5月6日(06.05.2010)

PCT

(10) 国際公開番号
WO 2010/050080 A1

- (51) 国際特許分類:
G06F 1/28 (2006.01) G06F 1/26 (2006.01)
G06F 1/20 (2006.01)
- (21) 国際出願番号: PCT/JP2009/000806
- (22) 国際出願日: 2009年2月24日(24.02.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2008-280698 2008年10月31日(31.10.2008) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社日立製作所 (HITACHI, LTD.) [JP/JP]; 〒1008280 東京都千代田区丸の内一丁目6番6号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 志賀 陽子 (SHIGA, Yoko) [JP/JP]; 〒2150013 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所 システム開発研究所内 Kanagawa (JP). 加藤

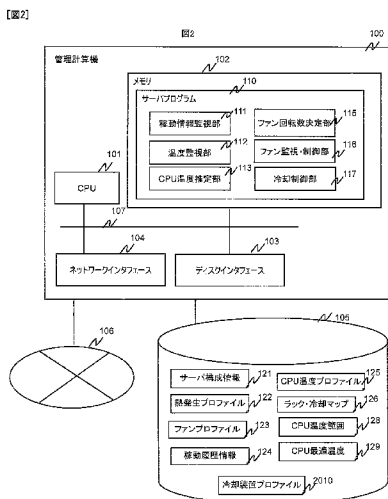
猛 (KATO, Takeshi) [JP/JP]; 〒1858601 東京都国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所 中央研究所内 Tokyo (JP). 高本 良史 (TAKAMOTO, Yoshifumi) [JP/JP]; 〒2448555 神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア事業部内 Kanagawa (JP). 林 真一 (HAYASHI, Sinichi) [JP/JP]; 〒2150013 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所 システム開発研究所内 Kanagawa (JP).

- (74) 代理人: 特許業務法人ウィルフォート国際特許事務所 (WILLFORT INTERNATIONAL); 〒1010044 東京都千代田区鍛冶町2丁目3番2号 神田センタービルディング5階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT,

[続葉有]

(54) Title: PHYSICAL COMPUTER, METHOD FOR CONTROLLING COOLING DEVICE, AND SERVER SYSTEM

(54) 発明の名称: 物理計算機及び冷却装置の制御方法及びサーバシステム



- 100 MANAGEMENT COMPUTER
- 102 MEMORY
- 110 SERVER PROGRAM
- 111 OPERATION INFORMATION MONITORING BLOCK
- 112 TEMPERATURE MONITORING BLOCK
- 113 CPU TEMPERATURE ESTIMATING BLOCK
- 114 FAN ROTATIONAL SPEED DETERMINING BLOCK
- 116 FAN MONITORING/CONTROLLING BLOCK
- 117 COOLING CONTROL BLOCK
- 104 NETWORK INTERFACE
- 103 DISK INTERFACE
- 121 SERVER CONFIGURATION INFORMATION
- 122 HEAT PRODUCTION PROFILE
- 123 FAN PROFILE
- 124 OPERATION HISTORY INFORMATION
- 125 CPU TEMPERATURE PROFILE
- 126 RACK/COOLING MAP
- 128 CPU TEMPERATURE RANGE
- 129 CPU OPTIMUM TEMPERATURE
- 2010 COOLING DEVICE PROFILE

(57) Abstract: Cooling control for reducing the overall power consumption of a system is determined. A method for controlling a management computer connected to a server device having a processor and a fan and a cooling device is characterized in that the temperature and operating rate of the processor, the rotational speed of the fan, and the temperature of the air sent into the server are acquired from the server, the temperature of the processor after a predetermined time elapse is estimated by calculation on the basis of the temperature and operating rate of the processor, the rotational speed of the fan, and the temperature of the sent air, the target rotational speed of the fan at which the estimated temperature after the time elapse is below a first predetermined value is determined if the estimated temperature is above the first predetermined value, and the server device is instructed to change the rotational speed to the target one.

(57) 要約: システム全体の消費電力をより小さくするための冷却制御を決定する。プロセッサとファンを有するサーバ装置と冷却装置とに接続する管理計算機の制御方法であって、前記サーバから前記プロセッサの温度及び稼働率と前記ファンの回転数と前記サーバへの入気温度を取得し、前記プロセッサの前記温度及び前記稼働率と、前記ファンの前記回転数と、前記入気温度とから、予め定められた期間を経過した後の前記プロセッサの推定温度を算出し、前記推定温度が第1の所定値以上である場合に、前記期間を経過した後の前記推定温度が前記所定値以下となる前記ファンの目標回転数を決定し、前記目標回転数とするように前記サーバ装置に指示する、ことを特徴とする制御方法である。

WO 2010/050080 A1



LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

物理計算機及び冷却装置の制御方法及びサーバシステム

技術分野

[0001] 本発明は、物理計算機及び冷却装置の制御方法に関し、特にCPUの稼働率に応じた物理計算機のファン及び冷却装置の出力の制御方法に関する。

背景技術

[0002] プロセッサの高性能化、ブレードサーバのような高密度なIT(Information Technology)機器の出現により、ITシステムの消費電力の増加が大きな問題となっている。この問題に対して、低消費電力プロセッサ、高効率冷却方式など消費電力を削減する技術の開発が進められているが、装置単体の省電力化には限界があり、より大きな省電力効果を得るためにはITシステムと冷却装置とを含むコンピュータ室全体での取り組みが必要である。

[0003] 現在のITシステムの冷却について説明する。サーバの稼働率上昇に伴いCPUやメモリなどの部品は熱を発生する。サーバは冷却ファンを備えており、サーバ筐体内の温度を検出し、温度が一定のしきい値を超えると、気流を発生させて発熱部品を冷却する。熱はこの気流により筐体の外に排出される。この冷却ファンの制御に関して、発熱部品の温度と温度変化に基づいてファンを制御し冷却を効率化する発明が特許文献1に開示されている。また、発熱部品の温度と温度変化に基づいてファンを制御し、冷却を効率化する発明が特許文献2で開示されている。

[0004] 一方、サーバを設置するコンピュータ室には機器の発生する熱を冷却するための冷却装置が設置されており、固定のセンサの温度をもとに出力を決定し、室内の温度を一定に保つ。しかし、IT機器の発する熱の偏りや機器の配置によって、コンピュータ室の温度を均一することは難しい。この冷却装置の制御に関して、マシン室内の空気流を監視し、空気流に応じて換気を行う発明が特許文献3に開示されている。

- [0005] コンピュータ室の熱の偏りが大きい場合は、熱だまりにある機器がCPU熱暴走により動作が不安定になる、また、CPU温度制御回路が作動し、強制的に処理性能を落としたり、シャットダウンしたりすることがある。強制的なシャットダウンによる不具合の発生を回避するため、温度が一定値を超えた場合ハイバネーションにより放熱制御して障害、シャットダウンを回避する発明が特許文献4で開示されている。
- [0006] また、GPUの現在の処理内容、GPUの温度特性、外部から入力されたユーザの指示に基づいて一定時間経過後の温度を予測し、予測された温度が基準値を上回った場合には、GPUの処理を他のGPUに移動する方法が、特許文献5に開示されている。
- [0007] 特許文献1：特開2002-268775号公報
特許文献2：特開2008-84173号公報
特許文献3：特開2006-208000号公報
特許文献4：特開2008-158787号公報
特許文献5：特開2007-241376号公報

発明の開示

発明が解決しようとする課題

- [0008] 今後は、熱の偏りを無くすためコンピュータ室全体を大規模な冷却装置で均一に冷やすのではなく、局所的に温度を制御できる冷却装置が重要になる。例えば、指向性のある冷却装置、ラック背面に取り付けるラック単位の冷却装置などが既に出荷されている。また、床下からの冷風によって冷却を行う設備の場合には、床のグレーティング板（穴あきタイル）の開閉を制御して、床下からの送風箇所を変更し、温度の高い箇所を集中的に冷やすようになる。
- [0009] また、発熱部品の稼働率が上昇してから熱が発生し、その熱が温度センサに検出されるまでには遅延がある。また、温度上昇が検出されてファンや空調の出力が上がり、実際に冷却効果が出るまでにはさらに遅延がある。このため、バースト的な負荷が発生した場合は、ファンの出力が上がっても一時

的には温度が上昇し、その後安定稼動可能な温度に落ち着く。しかし、CPUは自身の温度が上昇すると、リーク電流が指数関数的に増加して冷却の電力もより多く必要になる。

[0010] この結果、冷却能力が不足してCPU熱暴走により動作が不安定になる、または、CPU温度制御回路が作動し、強制的に処理性能を落としたり、シャットダウンすることがある。ハイバネーションにより、強制シャットダウンを回避しても、処理は中断せざるを得ない。

[0011] また、サーバの冷却ファンが回転数を上げて、サーバ背面の空気流が不足していると、熱が排出穴に運ばれずに周囲の機器に拡散し、他の機器の温度も上昇し障害を招くおそれがある。現在の冷却装置は一点で計測した温度情報を元に出力を調整しているため、あるサーバで発生した熱がその計測点に回り込むまで空調は連動しないので、このような状態になることがある。

課題を解決するための手段

[0012] 前記目的を達成するために、本発明の制御方法は、プロセッサとファンを有するサーバ装置と冷却装置とに接続する管理計算機の制御方法であって、前記サーバから前記プロセッサの温度及び稼働率と前記ファンの回転数と前記サーバへの入気温度を取得し、前記プロセッサの前記温度及び前記稼働率と、前記ファンの前記回転数と、前記入気温度とから、予め定められた期間を経過した後の前記プロセッサの推定温度を算出し、前記推定温度が第1の所定値以上である場合に、前記期間を経過した後の前記推定温度が前記所定値以下となる前記ファンの目標回転数を決定し、前記目標回転数とするように前記サーバ装置に指示する、ことを特徴とする制御方法である。

[0013] また、本発明のサーバシステムは、プロセッサとファンを有し、前記プロセッサの温度及び稼働率と前記ファンの回転数と入気温度とを計測する前記サーバ装置と、前記サーバ装置及び前記冷却装置に接続し、前記プロセッサの前記温度及び前記稼働率と、前記ファンの前記回転数と、前記入気温度とから、予め定められた期間を経過した後の前記プロセッサの推定温度を算出し、前記推定温度が第1の所定値を以上である場合に、前記期間を経過した

後の前記推定温度が前記所定値以下となる前記ファンの目標回転数を決定する管理計算機とを有する、ことを特徴とするサーバシステムである。

発明の効果

[0014] 本発明によれば、GPUを事前に冷却して最適温度に保つことにより、リーク電流による電力の消費を最小にし、冷却効率を上げることができる。

図面の簡単な説明

[0015] [図1]本発明の第一の実施例のシステム構成を示すシステム構成図の一例である。

[図2]本発明の第一の実施例の省電力制御サーバのハードウェア構成を示す構成図の一例である。

[図3]本発明の第一の実施例の管理対象である物理計算機ハードウェア構成を示す構成図の一例である。

[図4] (a) は、本発明の第一の実施例のコンピュータ室の内部構成図、(b) は、コンピュータ室の床に設置されたモータと吹出口との関係を示す要部断面図の一例である。

[図5]本発明の第一の実施例のサーバ構成情報を示す構成図の一例である。

[図6]本発明の第一の実施例のラック・冷却マップを示す構成図の一例である。

[図7]本発明の第一の実施例における稼動情報および消費電力情報を示す構成図の一例である。

[図8]本発明の第一の実施例における熱発生プロファイルを示す構成図の一例である。

[図9]本発明の第一の実施例におけるGPU温度プロファイルを示す構成図の一例である。

[図10]本発明の第一の実施例におけるファンプロファイルを示す構成図の一例である。

[図11]本発明の第一の実施例における省電力制御フローを示すフローチャートの一例である。

[図12]本発明の第一の実施例におけるCPU温度推定フローを示すフローチャートの一例である。

[図13]本発明の第一の実施例におけるファン回転数決定フローを示すフローチャートの一例である。

[図14]本発明の第二の実施例のシステム構成を示すシステム構成図の一例である。

[図15]本発明の第二の実施例のジョブ実行スケジュール、サーバ・ジョブマップを示す構成図の一例である。

[図16]本発明の第二の実施例の省電力制御フローを示すフローチャートの一例である。

[図17]本発明の第三の実施例のシステム構成を示すシステム構成図の一例である。

[図18]本発明の第三の実施例の冷却制御ルールを示す構成図の一例である。

[図19]本発明の第三の実施例の省電力制御フローを示すフローチャートの一例である。

[図20]本発明の第一の実施例の冷却装置プロファイルの一例である。

[図21]本発明の第一の実施例の冷却装置部のフローチャートの一例である。

符号の説明

- [0016] 110…省電力制御サーバ、111…稼動情報監視部、112…温度監視部、113…CPU温度推定部、114…冷却制御決定部、115…ファン回転数決定部、116…ファン監視・制御部、117…冷却制御部、121…サーバ構成情報、122…熱発生プロファイル、123…ファンプロファイル、124…サーバ稼動履歴、125…CPU温度プロファイル、126…ラック・冷却マップ、127…冷却プロファイル、200…物理計算機、223…計測エージェント

発明を実施するための最良の形態

- [0017] 以下、図面を用いて、本発明の幾つかの一実施形態を説明する。

実施例 1

[0018] 図1は、本発明の一実施例のシステム構成を示す図である。本実施例のシステム構成は、情報処理システムあるいはストレージシステム等である。例えば、1つの管理計算機100と、1つ以上の物理計算機200と、1つ以上のストレージ装置230と、該管理計算機100、該物理計算機200、該ストレージ装置230とを設置したコンピュータ室を冷却する冷却装置151と、冷却装置151を制御する冷却装置制御部150を備えて構成されている。そして、管理計算機100、物理計算機200及び冷却装置151は、管理ネットワーク225経由で接続している。また、物理計算機200とストレージ装置230は、例えばファイバチャネルネットワーク226で接続している。ここで、冷却装置制御部150は、冷却装置151を一括に制御すべく、物理計算機のメモリにプログラムとして格納されてもよい。また、冷却装置制御部151は、管理計算機内のメモリにプログラムとして格納されていてもよい。

[0019] 図2は、本発明の一実施例における管理計算機100を示す図である。

[0020] 管理計算機100の動作の概要を説明する。詳細については、以下に図面に沿って説明を加える。管理計算機100は、物理計算機200とストレージ装置230および冷却装置制御部150を管理する。そして、複数の物理計算機200と情報の授受を行って、複数の物理計算機200の稼働状況を検出する。検出した複数の物理計算機200の稼働状況に応じて、冷却装置制御機能151を介して冷却装置151を個別に制御する。また、検出した複数の物理計算機200の稼働状況に応じて複数の物理計算機200のファン回転数や冷却装置を制御する。

[0021] 管理計算機100は、中央演算装置CPU(Central Processing Unit)101、ハードディスク装置やフラッシュメモリ等の記憶装置105、メモリ102、バス107、ネットワークインタフェース104、ディスクインタフェース103から構成される。

[0022] メモリ102には、サーバプログラム110が格納されている。そして、サーバプログラムには、稼働情報監視部111、温度監視部112、CPU

温度推定部 113、ファン回転数決定部 115、ファン監視・制御部 116、冷却制御部 117が含まれる。これらのプログラムは、当初、磁気ディスク 105に格納され、必要に応じてメモリ 102に転送された後、CPU 101で実行される。稼動情報・電力監視部 111は、物理計算機 200の稼動情報および消費電力情報を収集する。温度監視部 112は、物理計算機の入気温度、CPU温度、排気温度を取得する。ファン監視・制御部 116は、ファン回転数の情報を取得し、ファンの回転数を変更する指示を出す。CPU温度推定部 113は、物理計算機に内蔵されるCPUの一定時間後の温度を推定する。ファン回転数決定部 115は、CPUの一定時間後の温度を目標値まで下げるファン回転数を決定する。冷却制御決定部 114は、ラック・冷却マップ 126を取得し、冷却装置 151の出力を決定する。冷却制御部 117は、冷却装置 151の出力を制御する指示を出す。

[0023] そして、記憶装置 105に、稼動情報履歴 124と、サーバ構成情報 121と、熱発生プロファイル 122と、ファンプロファイル 123と、CPU温度プロファイル 125と、冷却装置プロファイル 127と、CPU温度範囲 128と、CPU最適温度 129と、ラック・冷却マップ 126と、冷却装置プロファイル 2010と、が格納される。

[0024] 図 3は、本発明の一実施例における物理計算機 200のハードウェア構成を示す図である。

[0025] 物理計算機 200は、中央演算装置 CPU 201、メモリ 202、ハードディスク装置やフラッシュメモリ等の記憶装置 205、バス 207、ネットワークインタフェース 204、ディスクインタフェース 203、ファン 208、BMC (Baseboard Management Controller) 207とから構成される。

[0026] BMC 207は、サーバ入気温度、排気温度、CPU温度の監視や、ファン回転数の監視・制御、電源制御を行う。

[0027] メモリ 202には、OS 222、この物理計算機の稼動情報を収集する計測エージェントプログラム 223、業務サービスプログラム 224が格納さ

れる。これらのプログラムは、まず、磁気ディスク 205 に格納され、必要に応じてメモリ 202 に転送された後、CPU 201 で実行される。なお、これらのプログラムは、可搬型記録媒体から読み出されることにより、または、各々の装置に接続されたネットワーク経由で、他の計算機または記憶装置からダウンロードされることにより、磁気ディスク 205 に格納されるものであってもよい。

[0028] また、管理計算機のサーバプログラム 110 の各処理は、各プログラムを CPU で実行することにより実現するが、これらは計測エージェント決定部、計測部など、各処理を行う処理部として集積回路化するなどしてハードウェアで実現することもできる。

[0029] 計測エージェント 223 は、計算機 200 上で稼動し、自身が稼動する装置の CPU 使用率やメモリ使用率、ネットワークインタフェースの使用率などの稼動情報を収集し、計測カウンタとして記録するソフトウェアプログラムである。管理計算機のサーバプログラム 110 の稼動情報・電力監視部 111 は、計測エージェント 223 に対して、SNMP (Simple Network Management Protocol) による稼動情報収集リクエストを送信する。計測エージェント 223 は、この稼動情報収集リクエストを受け取り、リクエスト中のオブジェクト ID (Identification) で指定された計測カウンタの値を稼動情報・電力監視部 111 に送信する。サーバプログラム 110 は、この計測カウンタの値を受け取り、稼動情報として記録することで、複数の管理対象の稼動情報を一元的に管理することができる。

[0030] 図 4 は、本発明の一実施例における、物理計算機 200、ストレージ装置 230、冷却装置 151 などが設置されるコンピュータ室 400 の機器配置を示す図である。

[0031] コンピュータ室 400 には、4 つのラック 401 a、401 b、401 c、401 d と、冷却装置 151 a、151 b が配置されている。各ラック 401 と冷却装置 151 a、151 b は、床上に固定されている。床には、複

