

(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
2014 年 8 月 7 日(07.08.2014)



(10) 国際公開番号
WO 2014/118969 A1

- (51) 国際特許分類:
G06F 13/10 (2006.01) **G06F 9/46** (2006.01)
G06F 3/06 (2006.01)
- (21) 国際出願番号: PCT/JP2013/052377
- (22) 国際出願日: 2013 年 2 月 1 日(01.02.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 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: 山田 陽介(YAMADA, Yosuke); 〒2591392
神奈川県秦野市堀山下 1 番地 株式会社日立製
作所 I T プラットフォーム事業本部内
Kanagawa (JP). 清田 雄策(KIYOTA, Yuusaku); 〒
2591392 神奈川県秦野市堀山下 1 番地 株式会
社日立製作所 I T プラットフォーム事業本部
内 Kanagawa (JP). 井場 徹(IBA, Tooru); 〒2591392
神奈川県秦野市堀山下 1 番地 株式会社日立製

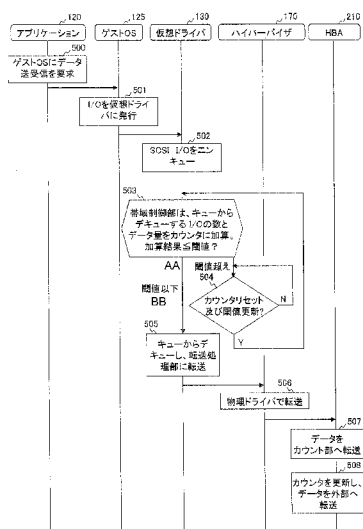
作所 I T プラットフォーム事業本部内
Kanagawa (JP).

- (74) 代理人: 藤井 正弘(FUJII, Masahiro); 〒1050001
東京都港区虎ノ門一丁目 1 番 4 号アーバン虎
ノ門ビル 藤井特許事務所 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, BN, BR, BW, BY, BZ, CA, CH, CL, 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, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR,
LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW.
- (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, LR, LS, MW,
MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシ
ア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ

[続葉有]

(54) Title: VIRTUAL COMPUTER SYSTEM AND DATA TRANSFER CONTROL METHOD FOR VIRTUAL COMPUTER SYSTEM

(54) 発明の名称: 仮想計算機システムおよび仮想計算機システムのデータ転送制御方法



120 Application
125 Guest OS
130 Virtual driver
170 Hypervisor
500 Request data transmission to and reception from guest OS
501 Issue I/O to virtual driver
502 Enqueue SCSI I/O
503 Bandwidth control unit adds number of I/O units to dequeue from queue, and quantity of data, to counter. Addition result $\leq$ threshold?
504 Counter reset and threshold update?
505 Dequeue from queue, and transfer to transfer processing unit
506 Transfer using physical driver
507 Transfer data to counting unit
508 Update counter, and transfer data to outside
AA Exceed threshold
BB Less than or equal to threshold

(57) Abstract: The present invention is a method which controls data transfer by a virtual computer system which is provided with a computer that is provided with a processor and memory and a virtualization unit which virtualizes a resource of the computer and assigns the same to a plurality of virtual computers, wherein: the computer has an adapter which is connected to storage devices; the adapter transmits data to and receives data from the storage devices, and measures the transfer amount of data that has been transmitted and received, and the number of I/O accesses, for each virtual computer; the virtualization unit, on the basis of the transfer amount of the data, and the number of I/O accesses, acquired from the adapter, computes an upper bound for the data transfer amount and an upper bound for the number of I/O accesses for each virtual computer and reports the same to the virtual computers; and the virtual computers retain the data to transfer to and to receive from the storage devices in a queue, and the virtual computers control data to output from the queue so as not to exceed the upper bound of the data transfer amount or the upper bound of the number of I/O accesses.

(57) 要約: プロセッサとメモリとを備えた計算機と、前記計算機の資源を仮想化して複数の仮想計算機に割り当てる仮想化部と、を備えた仮想計算機システムでデータ転送を制御する方法であって、前記計算機は、ストレージ装置に接続されるアダプタを有し、アダプタがストレージ装置との間でデータを送受信し、当該送受信したデータの転送量と I/O 回数とを仮想計算機毎に測定し、仮想化部は、アダプタから取得したデータの転送量と I/O 回数に基づいて、データ転送量の上限値と、I/O 回数の上限値を仮想計算機毎に演算して仮想計算機に通知し、仮想計算機が、ストレージ装置との間で送受信するデータをキューに保持し、仮想計算機が、データ転送量の上限値と、I/O 回数の上限値を超えないように前記キューから出力するデータを制御する。

WO 2014/118969 A1



(AL, 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, RS, SE, SI, SK, SM, TR), OAPI
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

仮想計算機システムおよび仮想計算機システムのデータ転送制御方法

技術分野

[0001] 本発明は、ファイバチャネルに接続されるHBA（Host Bus Adapter）を、複数の仮想計算機で共用する技術に関する。

背景技術

[0002] 計算機の仮想化技術およびクラウドコンピューティング技術の発展により、ホスト計算機における物理的な計算機資源を、複数の仮想的なゲスト計算機（仮想計算機）で共用して利用するシステムは、一般的に利用されるようになってきている。仮想計算機システムでは、物理的な計算機資源を効率的に利用しやすくなる反面、適切にゲスト計算機間の負荷を制御しなければならない。ゲスト計算機の負荷を制御しない場合、複数のゲスト計算機間で共用している計算機資源を1つのゲスト計算機が占有したり、重要なゲスト計算機のために確保しておきたい計算機資源を、重要でないゲスト計算機が使ってしまうたりする可能性がある。

[0003] 近年は、ミッションクリティカルな分野で仮想計算機システムが用いられる事例も増加傾向にあり、各々のゲスト計算機に割り当てる計算機資源を制御する機能はますます重要となってくると考えられる。特に、ディスク装置と通信するために用いられるファイバチャネルHBAは、ゲスト計算機とディスク装置またはSAN（Storage Area Network）とのインターフェースとして重要な役割を担っている。このため、多数のゲスト計算機が存在する仮想計算機システムにおいても、各々のゲスト計算機に対して、効率良くファイバチャネルHBAの帯域を割り当てられる技術の実現が求められている。

[0004] 特許文献1は、ネットワークにおけるパケット転送装置において帯域制御を実現した例である。特許文献1では、通信インターフェースにおいて送信パケットのサイズを解析することにより、送信するデータの量を検出し、パ

ケットを送信するタイミングを制御することで帯域制御を実現していた。

- [0005] 非特許文献1は、仮想化ソフトウェアを用いて複数のゲスト計算機間で共有される、ファイバチャネルHBAの帯域制御を行う例である。非特許文献2では、仮想化ソフトウェアにおけるI/O数を、SCSI I/O単位で計測し、帯域を制御するものである。

先行技術文献

特許文献

- [0006] 特許文献1：特開2006-109299号公報

非特許文献

- [0007] 非特許文献1：“FIBER CHANNEL Physical Interface-4 (FC-PI-4) Rev.7.00”、Chapter.6、Table 6、Single-mode link classes1 (OS1, OS2) 800-SML C-L: Data rate: 800MB/s、[online]、Global Engineering、2007年9月20日、[平成25年1月15日検索]

非特許文献2：「VMware vSphere 5.0 評価ガイド Vol.2：高度なストレージ機能テクニカル ホワイト ペーパー」、第48頁 IOPS 制限の効果の確認、[online]、グイェムウェア株式会社、[平成25年1月15日検索]

発明の概要

発明が解決しようとする課題

- [0008] ファイバチャネルHBAにおいて、1つの物理的なHBAポートから発行可能な1秒あたりのデータ送受信量(MB/s)は、共通規格(例えば、非特許文献1)によって定められている。例えば、8Gbpsファイバチャネルにおいては、800MB/sである。このため、このHBAポートを共有する各々のゲスト計算機が使用できる1秒あたりのデータ送受信量を保障するためには、各々のゲスト計算機の所定の周期あたりのデータ送受信量に閾値を設け、データ送受信量の合計がファイバチャネルHBAの規格(8Gbpsの場合は800MB/s)を超えないように制御することが必要である。

[0009] 上記非特許文献2の手法は、SCSI I/O数とデータ送受信量が比例するものと見なしてSCSI I/O数をカウントすることにより負荷を推定し、SCSI I/Oの数を制御する手法である。この手法を用いると、「1つのSCSI I/Oあたりのデータ送受信量」が常に一定ならば、ファイバチャネルHBAにおける各々のゲスト計算機の制御間隔あたりのデータ送受信量を制御できる。しかし、「1つのSCSI I/Oあたりのデータ送受信量」がゲスト計算機によって異なる場合は、SCSI I/O数がデータ送受信量に比例するという仮定が成り立たないため、各々のゲスト計算機の制御間隔あたりのデータ送受信量のカウント値と、実際のデータ通信量との間の誤差が大きくなるという課題があった。

[0010] たとえば、ゲスト計算機1とゲスト計算機2の間で、ゲスト計算機2のデータ送受信量をゲスト計算機1の100分の1に制限するために、ゲスト計算機2のコマンド数をゲスト計算機1の100分の1となるように制限したと仮定する。このとき、ゲスト計算機2の「1つのSCSI I/Oあたりのデータ送受信量」がゲスト計算機1の100倍であれば、ゲスト計算機2のSCSI I/O数をゲスト計算機1の100分の1に抑えることができたとしても、ゲスト計算機2のデータ送受信量はゲスト計算機1のデータ送受信量とほぼ同じとなる。その場合、ゲスト計算機2のデータ送受信量をゲスト計算機1の100分の1に抑えるという目的は達成できない、という課題があった。

[0011] データ送受信量に基づいて帯域制御を実現した例として、上記特許文献1が知られている。しかしながら、特許文献1の手法では、実際にパケットを送信する前に、パケットのデータ量を、「制御間隔あたりのデータ送信量を保持するカウント値」に加算しているため、カウント値と実際のデータ送受信量との間に誤差が生じる可能性がある。

[0012] 例として、フレームサイズ8MBのSCSI I/Oコマンドを送受信する場合を考える。この場合、データ送受信開始から1ミリ秒経過した時点では、実際には800KB分のデータしか送受信できていないにもかかわらず、

特許文献１の手法では、８ＭＢのデータを送受信したとカウントする。つまり、特許文献１の手法は、１つのＩ／Ｏコマンドで送受信するデータサイズが大きくなるほどに、カウント値と実際に使用しているデータ送受信量との誤差が大きくなる、という課題があった。

- [0013] 一方、データベースの分野ではＳＣＳＩに準じた通信を行う際に、１つのＳＣＳＩコマンドで転送するデータ量を大きく（１ＭＢ程度）設定することがある。この様な分野では、「１つのＩ／Ｏコマンドで転送するデータサイズが大きくなったとしても、カウント値と実際に使用しているデータ送受信量との誤差が変わらない」手法が求められていた。

課題を解決するための手段

- [0014] 本発明は、プロセッサとメモリとを備えた計算機と、前記計算機の資源を仮想化して１以上の仮想計算機に割り当てる仮想化部と、を備えた仮想計算機システムであって、前記計算機は、ストレージ装置に接続されるアダプタを有し、前記アダプタは、前記ストレージ装置との間でデータを送受信し、当該送受信したデータの転送量とＩ／Ｏ回数とを前記仮想計算機毎に測定する転送処理部と、前記データの転送量とＩ／Ｏ回数を前記仮想計算機毎に格納するカウンタと、を有し、前記仮想計算機は、前記ストレージ装置との間で送受信するデータを保持するキューと、前記データの転送量とＩ／Ｏ回数を制御する帯域制御部と、を有し、前記仮想化部は、前記アダプタのカウンタから取得した前記データの転送量とＩ／Ｏ回数に基づいて、前記データの転送量の上限値と、前記Ｉ／Ｏ回数の上限値を前記仮想計算機毎に演算する閾値演算部を有し、前記帯域制御部は、前記仮想化部が演算した前記データの転送量の上限値と、前記Ｉ／Ｏ回数の上限値を超えないように前記キューから出力するデータを制御する。

発明の効果

- [0015] 本発明により、ストレージ装置に接続されるアダプタ（例えば、ＨＢＡ）のＩ／Ｏについて、Ｉ／Ｏ回数に基づく帯域制御のみでなく、実際に各ゲスト計算機が送受信したデータ量に基づく帯域制御を実現することが可能とな

る。これにより、従来の I/O 回数のみによる帯域制御に対して、HBA の帯域を超えることなく各ゲスト計算機毎に正確な帯域制御が実現できる。

図面の簡単な説明

[0016] [図1]本発明の実施例を示し、仮想計算機システムの一例を示すブロック図である。

[図2]本発明の実施例を示し、ハイパバイザの一例を示すブロック図である。

[図3]本発明の実施例を示し、ゲスト計算機の一例を示すブロック図である。

[図4]本発明の実施例を示し、HBAの一例を示すブロック図である。

[図5]本発明の実施例を示し、仮想計算機システムで行われる SCSI I/O 処理の一例を示すシーケンス図である。

[図6]本発明の実施例を示し、仮想計算機システムで行われる閾値更新処理の一例を示すシーケンス図である。

[図7]本発明の実施例を示し、ハイパーバイザが管理する閾値更新ルールの一例を示す図である。

[図8]本発明の実施例を示し、コマンド数とデータ量と目標帯域の関係を示す図である。

[図9]本発明の実施例を示し、仮想WWNテーブルの一例を示す図である。

発明を実施するための形態

[0017] 以下、本発明を適用する仮想計算機システムについて、図面を参照して詳細に説明する。

[0018] 図1は、本発明の仮想計算機システムの一例を示すブロック図である。図1において、ホスト計算機100は、演算を実行する複数の物理プロセッサ109-1~109-nと、データやプログラムを格納する物理メモリ114と、LAN280と通信を行うNIC (Network Interface Card) 270と、SAN (Storage Area Network) 250を介してストレージ装置260を制御するファイバチャネルHBA (HOST BUS ADAPTER) 210と、ファイバチャネルHBA210及びNIC270と各物理プロセッサ109-1~109-nとを接続するチップセット108と、を備える。

[0019] ハイパーバイザ（仮想化部）170は、ホスト計算機100の物理プロセッサ109-1～109-nや物理メモリ114等の物理計算機資源を分割して仮想計算機資源300（図2参照）を生成し、仮想プロセッサや仮想メモリなどの仮想計算機資源（または論理計算機資源）をゲスト計算機（または仮想計算機）1～n（105-1～105-n）に割り当てることで、ゲスト計算機105-1～105-nをホスト計算機100上に構成する。

[0020] なお、ゲスト計算機n（105-n）も、ゲスト計算機1（105-1）と同様の構成であるため、本実施例ではゲスト計算機n（105-n）の説明を省略し、ゲスト計算機1（105-1）について説明する。なお、以下では、ゲスト計算機105-1～105-nの総称を符号105で表記し、物理プロセッサ109-1～109-nの総称を符号109で表記する。以下、他の符号についても同様であり、構成要素の総称を示す符号は”-”以降を削除した符号を用いる。

[0021] <概要>

本発明は、ファイバチャネルHBA（以下、HBAとする）210を複数のゲスト計算機105で共有し、各ゲスト計算機105がそれぞれ使用するHBA210の帯域とI/O回数の上限値をハイパーバイザ170が決定し、各ゲスト計算機105の仮想ドライバに含まれる帯域制御部がHBA210の帯域とI/O回数を規制する。

[0022] このため、各ゲスト計算機105が使用したHBA210の帯域（データ量）とI/O回数を、HBA210の転送処理部がゲスト計算機105毎に測定する。そして、ハイパーバイザ170は、所定の時間間隔（例えば、10msec）ごとに、各ゲスト計算機105が使用した帯域とI/O回数をHBA210から取得して、各ゲスト計算機105がそれぞれ使用するHBA210の帯域の上限値である帯域閾値と、I/O回数（以下IOPSとする）の上限値であるIOPS閾値（I/O回数閾値）とを算出して更新する。なお、IOPS閾値は、所定の時間間隔あたりにゲスト計算機105が発行可能なI/O数である。帯域閾値は、所定の時間間隔あたりにゲスト計算機1

１０５が発行可能なデータの送受信量である。なお、所定の時間間隔としては、ゲストＯＳ１２５のタイマ割り込みの周期等を用いるようにしてもよい。

[0023] ゲスト計算機１０５は、後述するように所定のタイミングで帯域閾値と、ＩＯＰＳ閾値をハイパーバイザ１７０から取得し、仮想ドライバに含まれる帯域制御部が、ＨＢＡ２１０の帯域を制御する。

[0024] <ハイパーバイザ>

図２は、ハイパーバイザ１７０の一例を示すブロック図である。

[0025] ハイパーバイザ１７０は、ゲスト計算機１０５を制御するためのプログラムであり、ホスト計算機１００の物理メモリ１１４にロードされて、物理プロセッサ１０９によって実行される。

[0026] ハイパーバイザ１７０は、ホスト計算機１００の物理プロセッサ１０９－１～１０９－ n や物理メモリ１１４等の物理資源を分割し、仮想プロセッサ３０１－１～３０１－ n や仮想メモリ３０２－１～３０２－ n 及び仮想ＨＢＡ３０３－１～３０３－ n を仮想計算機資源３００として生成し、ゲスト計算機１～ n （１０５－１～１０５－ n ）に割り当てる。なお、図示はしないが、ＮＩＣ２７０についても、ＨＢＡ２１０と同様にハイパーバイザ１７０が仮想ＮＩＣをゲスト計算機１０５に提供する。

[0027] ハイパーバイザ１７０は、ゲスト計算機１０５－１～１０５－ n に割り当てる仮想ＨＢＡ３０３－１～３０３－ n のそれぞれについて、仮想ＷＷＮ（図中ＶＷＷＮ－１～ＶＷＷＮ－ n ）を割り当てる。そして、ハイパーバイザ１７０は、仮想ＨＢＡ３０３に仮想ＷＷＮ（World Wide Name）を割り当てる度に、当該仮想ＷＷＮと当該仮想ＨＢＡ３０３を割り当てたゲスト計算機１０５の識別子を物理的なＨＢＡ２１０へ通知する。

[0028] ＨＢＡ２１０は、ＳＣＳＩ Ｉ／Ｏ（以下、単にＩ／Ｏとする）を受け付けると、上記通知された仮想ＷＷＮとゲスト計算機１０５の識別子からＩ／Ｏの発行元のゲスト計算機１０５を特定する。なお、ハイパーバイザ１７０が仮想ＨＢＡ３０３－１～３０３－ n に割り当てる仮想ＷＷＮ－１～ＶＷＷＮ－ n の値は、仮想計算機システム内でユニークであれば良い。

[0029] そして、上述のようにゲスト計算機 105 毎の HBA 210 に対する I/O の帯域と IOPS を制御するため、ハイパーバイザ 170 は、I/O の帯域閾値と IOPS 閾値を算出する閾値計算部 185 と、閾値を更新する際のルールを保持する閾値更新ルール 186 と、HBA 210 を制御するための物理ドライバ 187 を有する。なお、図示はしないが、ハイパーバイザ 170 は、NIC 270 などの他のデバイスについても物理ドライバを備えて制御する。また、図 2 において、仮想計算機資源（または論理計算機資源）の提供については、公知または周知の技術を適用すれば良いので、本実施例では説明しない。

[0030] ハイパーバイザ 170 は、物理メモリ 114 の所定の領域を使用し、データやプログラムを保持する。ハイパーバイザ 170 は、ゲスト計算機 1～n（105-1～105-n）のそれぞれについて、HBA 210 から取得した IOPS を格納する IOPS 値（I/O 回数値）1～n（190-1～190-n）と、閾値計算部 185 が算出した IOPS 閾値を格納する IOPS 閾値 1～n（200-1～200-n）と、HBA 210 から取得した I/O の帯域（データ転送量）を格納する帯域値 1～n（195-1～195-n）および閾値計算部 185 が算出した I/O の帯域閾値を格納する帯域閾値 1～n（205-1～205-n）を保持する。

[0031] 閾値計算部 185 は閾値計算プログラムで構成され、物理メモリ 114 の所定の領域にロードされて、物理プロセッサ 109 で実行される。

[0032] 閾値計算部 185 は、HBA 210 が測定した IOPS カウンタ 1～n（235-1～n：図 4 参照）の値と、帯域カウンタ 1～n（240-1～240-n：図 4 参照）の値を取得し、上記ハイパーバイザ 170 の IOPS 値 190 および帯域値 195 に格納し、閾値計算部 185 が IOPS 閾値 200 および帯域閾値 205 を更新する。そして、閾値計算部 185 からの閾値更新完了の通知に基づいて、仮想ドライバ 130 は IOPS 値 1'（150）と帯域値 1'（155）をリセットする。

[0033] なお、ハイパーバイザ 170 は、ゲスト計算機 105-1～105-n の

仮想ドライバ130からI/Oの要求（リード要求またはライト要求）を受け付けると、物理ドライバ187でI/O要求を物理的なHBA210に転送する。そして、I/O要求には仮想HBA303-1~303-nの仮想WWNを付与することで、ハイパーバイザ170及び物理的なHBA210はI/Oの発行元のゲスト計算機105-1~105-nを特定することができるのである。

[0034] また、ハイパーバイザ170は、LAN280を介して図示しない管理計算機などからゲスト計算機の生成、起動または停止、削除の指令を受けて付けて仮想計算機資源の割り当てを制御する。

[0035] 図7は、ハイパーバイザ170が管理する閾値更新ルール186の一例を示す図である。閾値更新ルール186は、ゲスト計算機105-1~105-nの識別子を格納するエントリ1861と、各ゲスト計算機105毎の閾値更新ルールを格納するルール1862のエントリを備え、ゲスト計算機105-1~105-nに対応するフィールドを有する。ルール1862は、LAN280を介して図示しない管理計算機などから受け付けた値を格納する。ここで、ルール1862が「増大」のゲスト計算機105では、IOPS値150がIOPS閾値140を超えたり、帯域値155が帯域閾値145を超えた場合、次の制御間隔で用いる帯域閾値145を増大させ、また、IOPS閾値140を増大させる。帯域閾値145の増大は、予め設定した増分値を加算することができる。ただし、帯域閾値145を増大する上限を設けても良い。また、IOPS閾値140の増大も、予め設定した増分値を加算することができ、IOPS閾値140を増大する上限を設けても良い。

[0036] <ゲスト計算機>

図3は、ゲスト計算機105-1の一例を示すブロック図である。なお、他のゲスト計算機105-nも同様の構成であるので、重複した説明は省略する。ゲスト計算機105-1は、ハイパーバイザ170が提供する仮想計算機資源上で稼動する仮想計算機である。

- [0037] ゲスト計算機 105-1 は、ハイパーバイザ 170 から仮想プロセッサ 301-1 と、仮想メモリ 302-1 及び仮想 HBA 303-1 の提供を受けて、ゲスト OS 125 を実行する。ゲスト OS 125 上では、仮想 HBA 303-1 に対してアクセスする仮想ドライバ 130 と、仮想ドライバ 130 に I/O 要求を行うアプリケーション 120 が稼働する。アプリケーション 120 は、ゲスト OS 125 上で稼働するソフトウェアであり、ゲスト OS 125 に対してデータの送受信を要求する。
- [0038] 仮想ドライバ 130 は、ゲスト OS 125 からデータ送受信の要求である I/O 要求を受け付けて、仮想 HBA 303-1 に対して当該 I/O 要求に応じたデータの送受信を実行するプログラムである。
- [0039] 仮想ドライバ 130 の帯域制御部 135 は、SCSI I/O キュー（以下、I/O キューとする）160 中の I/O に記録されているデータ量 165 を取得して、IOPS 値 1'（150）の値が IOPS 閾値 1'（140）を超えないように、また帯域値 1'（155）の値が帯域閾値 1'（145）を超えないように、仮想 HBA 303-1（HBA 210）が発行する I/O の量を所定の時間間隔（例えば、10 msec）ごとに、制限する。
- [0040] I/O キュー 160 は、複数のキュー 165-1 ~ 165-n を有し、送受信するデータを一時的に格納する。I/O キュー 160 は、仮想ドライバ 130 がゲスト OS 125 から受け取った I/O 要求を一時的に保持するための記憶領域である。I/O キュー 160 中の各々の I/O には、送受信するデータ量が記録されている。
- [0041] 仮想 HBA 303-1 の所定の時間間隔毎のデータ転送量は、仮想 HBA 303-1 に発行した IOPS を格納する IOPS 値 1'（150）と、仮想 HBA 303-1 から物理ドライバ 187 に転送したデータ量を格納する帯域値 1'（155）と、からなる。
- [0042] また、ハイパーバイザ 170 から取得した仮想 HBA 303-1 の閾値は、仮想 HBA 303-1 の仮想 WWN-1 に対応付けられた HBA 210 の IOPS を規制する IOPS 閾値 1'（140）と、帯域を規制する値を格

納する帯域閾値 1' (145) とを含む。

[0043] アプリケーション 120 は、ストレージ装置 260 へのアクセスが発生すると、仮想ドライバ 130 に I/O 要求を仮想 HBA 303-1 に発行しデータを I/O キュー 160 に格納する。

[0044] 仮想ドライバ 130 の帯域制御部 135 は、IOPS 値 1' (150) 及び帯域値 1' (155) が IOPS 閾値 1' (140) 及び帯域閾値 1' (145) 以内であれば、キュー 160 のデータをハイパーバイザ 170 の物理ドライバ 187 へ転送するように仮想 HBA 303-1 へ指令する。

[0045] 一方、仮想ドライバ 130 の帯域制御部 135 は、IOPS 値 1' (150) または帯域値 1' (155) の何れかが、IOPS 閾値 1' (140) あるいは帯域閾値 1' (145) の何れかを超過していれば、所定の時間間隔 (例えば、10msec) が経過するまで I/O 要求をキュー 160 に保留し、ハイパーバイザ 170 によって IOPS 閾値 1' (140) 及び帯域閾値 1' (145) が更新されるのを待つ。

[0046] 上記帯域制御部 135 により、ゲスト計算機 105-1 が利用する仮想 HBA 303-1 (HBA 210) の帯域 (データ転送量) 及び IOPS は、所定の時間間隔ごとに閾値以内に制御される。

[0047] 本発明の帯域制御部 135 における帯域制御では、ゲスト計算機 1 (105-1) が発行する I/O 要求の数およびデータの送受信量を一定の時間間隔 (10ms) ごとに、閾値を設けて制限する。この一定間隔を制御間隔と呼び、当該制御間隔毎にハイパーバイザ 170 が閾値を更新する。閾値としては、I/O 要求の数と、データの送受信量 (帯域) についてそれぞれの値が設定され、各々の閾値を超過しないように、ゲスト計算機 105 で仮想 HBA 303-1 を制御する仮想ドライバ 130 の帯域制御部 135 が I/O を発行するタイミングを制御することにより、帯域制御を実現する。

[0048] IOPS 閾値 1' (140) は、制御間隔あたりにゲスト計算機 1 (105-1) が発行可能な I/O 数を保持する変数である。IOPS 値 1' (150) は、制御間隔あたりにゲスト計算機 1 (105-1) が発行した I/O

０数を保持する変数である。帯域閾値１'（１４５）は、制御間隔あたりにゲスト計算機１（１０５－１）が発行可能なデータの送受信量を保持する変数である。帯域値１'（１５５）は、制御間隔あたりのゲスト計算機１（１０５－１）のデータ送受信量を保持する変数である。

[0049] ＩＯＰＳ閾値１'（１４０）および帯域閾値１'（１４５）には制御間隔が開始する時点で、該制御間隔の間にゲスト計算機１０５－１が送受信可能なＩ／Ｏ数およびデータ送受信量が、後述の閾値計算部１８５によって設定される。ＩＯＰＳ値１'（１５０）および帯域値１'（１５５）は、所定の制御間隔でハイパーバイザ１７０からの通知によって０にリセットされる。

[0050] ホスト計算機１００は以上のように構成され、ハイパーバイザ１７０の閾値計算部１８５と、ゲストＯＳ１２５、アプリケーション１２０、仮想ドライバ１３０、帯域制御部１３５は、物理メモリ１１４に格納されたプログラムとして物理プロセッサ１０９により実行される。

[0051] 物理プロセッサ１０９は、各機能部のプログラムに従って動作することによって、所定の機能を実現する機能部として動作する。例えば、物理プロセッサ１０９は、帯域制御プログラムに従って動作することで帯域制御部１３５として機能し、閾値計算プログラムに従って動作することで閾値計算部１８５として機能する。他のプログラムについても同様である。さらに、物理プロセッサ１０９は、各プログラムが実行する複数の処理のそれぞれを実現する機能部としても動作する。計算機及び計算機システムは、これらの機能部を含む装置及びシステムである。

[0052] ホスト計算機１００の各機能を実現するプログラム、テーブル等の情報は、ストレージ装置２６０や不揮発性半導体メモリ、ハードディスクドライブ、ＳＳＤ（Solid State Drive）等の記憶デバイス、または、ＩＣカード、ＳＤカード、ＤＶＤ等の計算機読み取り可能な非一時的データ記憶媒体に格納することができる。

[0053] ＜ＨＢＡ＞

図４は、ＨＢＡ２１０の一例を示すブロック図である。

- [0054] HBA 210は、ホスト計算機100とファイバチャネルで構成されたSAN 250及びストレージ装置260との間でデータの送受信を実行する装置である。
- [0055] HBA 210は、組み込みプロセッサ215と、記憶部220と、カウント回路230と、ホスト計算機100に接続されるI/F部236と、SAN 250に接続されるポート237とを備える。I/F部236は、例えば、PCI expressなどで構成される。
- [0056] カウント回路230は、データの送受信量を測定する論理回路である。記憶部220には、データの送受信処理を行う転送処理部225と、ゲスト計算機105-1～105-n毎に仮想HBA 303-1～303-nのI/O数の測定結果を格納するIOPSカウンタ1～n（235-1～23-n）と、データ転送量（帯域）の測定結果を格納する帯域カウンタ1～n（240-1～240-n）と、ハイパーバイザ170から受信した仮想WWNとゲスト計算機105の識別子を対応付けて保持する仮想WWNテーブル245が格納される。転送処理部225は、記憶部220に転送処理プログラムをロードして、組み込みプロセッサ215が実行することで機能する。
- [0057] 転送処理部225は、ハイパーバイザ170から仮想WWNとゲスト計算機105の識別子を受信すると、仮想WWNとゲスト計算機105の識別子を対応付けて仮想WWNテーブル245に格納する。そして、転送処理部225は、仮想WWN毎に、IOPSカウンタ235と帯域カウンタ240を割り当てる。図9は、仮想WWNテーブル245の一例を示す図である。仮想WWNテーブル245は、ハイパーバイザ170が仮想HBA 303に割り当てた仮想WWNを格納するカラム2451と、仮想WWNの仮想HBA 303を割り当てたゲスト計算機105の識別子を格納するカラム2452と、転送処理部225が当該仮想WWNに割り当てたIOPSカウンタ235の識別子を格納するカラム2453と、転送処理部225が当該仮想WWNに割り当てた帯域カウンタ240の識別子を格納するカラム2454と、からひとつのエントリが構成される。

[0058] なお、ハイパーバイザ170は、ひとつの物理的なHBA210から生成した複数の仮想HBA303を、各ゲスト計算機105にひとつずつ割り当てるものとする。

[0059] 転送処理部225は、ホスト計算機100からのI/O要求に応じてSAN250を介してストレージ装置260と通信を行う。このとき、転送処理部225はカウント回路230を用いてゲスト計算機105毎にデータの転送量を帯域として測定し、帯域カウンタ240に格納する。また、ゲスト計算機105毎にI/O要求の回数を測定し、IOPSカウンタ235に格納する。

[0060] ここで、転送処理部225は、I/Oに含まれる仮想WWNから仮想WWNテーブル245を参照して、ゲスト計算機105とIOPSカウンタ235及び帯域カウンタ240を特定し、当該仮想WWNに対応するIOPSカウンタ235と帯域カウンタ240に値を格納する。例えば、I/O要求に含まれる仮想WWNが「VWWN-1」であれば、転送処理部225は、ゲスト計算機105-1の仮想HBA303-1が発行したI/O要求であると判定し、ゲスト計算機105-1に対応するIOPSカウンタ235-1と帯域カウンタ240-1に値を格納する。

[0061] そして、後述するように、HBA210はハイパーバイザ170からの要求を受け付けると、IOPSカウンタ235と帯域カウンタ240の値を通知する。また、HBA210は、ハイパーバイザ170から読み出し要求に応じて値を読み出したIOPSカウンタ235及び帯域カウンタ240をリセットする。

[0062] <I/O処理>

図5は、仮想計算機システムで行われるSCSI I/O処理（以下、I/O処理とする）の一例を示すシーケンス図である。図5のシーケンス図は、ゲスト計算機105のアプリケーション120がゲストOS125にI/O要求を送信したときに実行される。

[0063] 本処理では、まずゲストOS125が、アプリケーション120から発行

されたデータ送受信要求を受信する（ステップ500）。ゲストOS125は、受信したデータ送受信要求をSCSI I/Oに変換し、仮想ドライバ130にI/Oを発行する（ステップ501）。更に仮想ドライバ130は、ゲストOS125から受信したI/Oを、SCSI I/Oキュー160にエンキューする（ステップ502）。

[0064] さらに、仮想ドライバ130の帯域制御部135は、ゲスト計算機1（105-1）が各I/OをSCSI I/Oキュー160からデキュー可能か否かの判定を行う（ステップ503）。判定方法については後述する。

[0065] ステップ503の判定で、デキュー可能と判断した場合、帯域制御部135は、SCSI I/Oキュー160にエンキューされた順番でI/Oをデキューし、ハイパーバイザ170にI/Oを発行する（ステップ505）。ハイパーバイザ170は受信したI/Oを物理ドライバ187でHBA210に転送する（ステップ506）。

[0066] 一方、ステップ503の判定で、デキューが不可能と判断した場合は、判断した時点で、該ゲスト計算機105のI/Oの発行を停止する制御を行う。

[0067] 上記ステップ503の帯域制御部135の処理は、まず、SCSI I/Oキュー160にエンキューされた順にI/Oを選択し、制御間隔内でゲスト計算機1（105-1）が該I/Oを発行可能であるか判定するため、仮想ドライバ130のIOPS値1'（150）に、1（該I/O分）を加算し、I/Oに記入されているデータ量を、帯域値1'（155）に加算する。

[0068] その後、帯域制御部135は、IOPS値1'（150）とIOPS閾値1'（140）を比較し、更に帯域値1'（155）と帯域閾値1'（145）を比較する（ステップ503）。

[0069] 上記比較の結果、IOPS値1'（150）がIOPS閾値1'（140）以下であり、かつ、帯域値1'（155）が帯域閾値1'（145）以下であるならば、該ゲスト計算機がI/Oを追加して発行可能であると帯域制御部135は判定する。そして、帯域制御部135は、SCSI I/Oキュー

ー 160 から該 I/O をデキューし、HBA 210 の転送処理プログラム 225 に対して I/O の発行処理を行う（ステップ 505）。IOPS 値 1'（150）が IOPS 閾値 1'（140）より大きい、または、帯域値 1'（155）が帯域閾値 1'（145）より大きいならば、該ゲスト計算機 105 が追加で I/O を発行不可であると帯域制御部 135 は判定する。このため、帯域制御部 135 は該 I/O の発行処理を行わず、該 I/O は SCSI I/O キュー 160 に滞留されたままとなる（ステップ 504）。すなわち、SCSI I/O キュー 160 のデータは出力が禁止される。

[0070] そして、ステップ 504 では、IOPS 値 1'（150）が IOPS 閾値 1'（140）を超えたか、または、帯域値 1'（155）が帯域閾値 1'（145）を超えたときには、仮想ドライバ 130 はハイパーバイザ 170 に対して閾値を超えたことを通知する。

[0071] I/O が SCSI I/O キュー 160 に滞留する場合は、制御間隔が終了する時点で、IOPS 値 1'（150）および帯域値 1'（155）が、後述するようにハイパーバイザ 170 の通知でリセットされることを契機として帯域制御部 135 により、ステップ 503 へ戻って再び加算結果と閾値の比較が実行される。

[0072] 帯域制御部 135 により、I/O がハイパーバイザ 170 の物理ドライバ 187 を介して転送処理部 225 に転送された後、転送処理部 225 は、受信した I/O をストレージ装置 260 へ発行する。また、転送処理部 225 は発行した I/O をカウント回路 230 へ転送する（ステップ 507、508）。

[0073] HBA 210 のカウント回路 230 は、IOPS カウンタ 235 に発行した I/O 数を加算した後、I/O の転送を開始してから終了するまでのデータ転送量を仮想 WWN に基づいて帯域カウンタ 240 に加算する（ステップ 508）。

[0074] 以上の処理により、ゲスト計算機 105 の仮想ドライバ 130 の帯域制御部 135 は、ハイパーバイザ 170 が決定した IOPS 閾値 1'（140）

と帯域閾値' (145) で所定の制御間隔 (例えば、10msec) ないで、I/O回数またはI/Oのデータ量 (帯域) が閾値を超えないように監視する。I/O回数またはI/Oの帯域がIOPS閾値1' (140) または帯域閾値1' (145) を超えると、帯域制御部135は、現在の制御間隔でのI/Oの発行を停止することで、ゲスト計算機105のI/O回数と帯域を目標とする閾値以内に制限することが可能となる。

[0075] <閾値の更新>

次に、上記ステップ504で行われるIOPS閾値200 (140) と、帯域閾値205 (145) の演算について図6を参照しながら説明する。

[0076] 図6は、仮想計算機システムで行われる閾値更新処理の一例を示すシーケンス図である。この処理はハイパーバイザ170のブートを契機に開始し、その後、所定の制御間隔 (10msec) ごとに繰り返して実行される。以下では、ゲスト計算機105-1に割り当てられた仮想HBA303-1の閾値更新の一例を示すが、他の仮想HBA303-2~303-nの閾値更新についても同様に行えば良い。なお、他の仮想HBA303-2~303-nの閾値更新は、以下の各ステップで実施してもよい。あるいは、仮想HBA303-1~303-nの閾値更新のタイミングをずらして、以下のステップ601~607を各仮想HBA303毎に実行しても良い。

[0077] この処理は、ハイパーバイザ170の閾値計算部185が、HBA210からIOPSカウンタ1 (235-1) および帯域カウンタ1 (240-1) の値を受信し、IOPS閾値1 (200-1) および帯域閾値1 (205-1) を演算し、ゲスト計算機1 (105-1) に閾値の更新完了を通知する処理である。この通知により、仮想ドライバ130の帯域制御部135は、ゲスト計算機105-1のIOPS値1' (150) および帯域値1' (155) をリセットする。

[0078] まず、ハイパーバイザ170の閾値計算部185が、HBA210に対してIOPSカウンタ1 (235-1) と帯域カウンタ1 (240-1) の読み出しを要求する (ステップ601)。

- [0079] 次に、HBA201は、要求されたIOPSカウンタ1(235-1)と帯域カウンタ1(240-1)の値を閾値計算部185へ送信する(ステップ602)。HBA201は、読み出しが完了したIOPSカウンタ1(235-1)と帯域カウンタ1(240-1)の値を0にリセットする。
- [0080] 閾値計算部185は、受信したIOPSカウンタ1(235-1)の値をIOPS値1(190-1)に格納し、受信した帯域カウンタ1(240-1)の値を帯域値1(195-1)に格納する(ステップ603)。
- [0081] 閾値計算部185は、閾値更新ルール186を参照し、更新対象の仮想HBA303-1を割り当てたゲスト計算機105-1の閾値の更新ルールを取得する。閾値の更新ルール1862が「維持」であれば、閾値計算部185は、IOPS閾値1(200-1)と帯域閾値1(205-1)を維持する。なお、「維持」の場合は、IOPS閾値1(200-1)と帯域閾値1(205-1)の増分値を、それぞれ0としてもよい。
- [0082] 一方、閾値の更新ルール1862が「増大」で、かつ、前記図5のステップ504で、仮想ドライバ130から閾値を超えた通知を受けたゲスト計算機105であれば、閾値計算部185は、IOPS閾値1(200-1)と帯域閾値1(205-1)にそれぞれ所定の増分値を加算して更新する(ステップ604)。なお、所定の増分値は、IOPS閾値1(200-1)の増分値と、帯域閾値1(205-1)の増分値がそれぞれ予め設定されている。また、これらの増分値を閾値更新ルール186に格納するようにしてもよい。
- [0083] 更に、閾値計算部185は、IOPS閾値1(200-1)および帯域閾値1(205-1)の更新が完了したことを、ゲスト計算機105-1で稼働する仮想ドライバ130の帯域制御部135に通知する。
- [0084] 帯域制御部135は、ハイパーバイザ170から閾値更新通知の受信を契機に、ハイパーバイザ170からIOPS閾値1(200-1)を読み出して、IOPS閾値1'(140)に格納する。次に、帯域制御部135は、ハイパーバイザ170から帯域閾値1(205-1)を読み出し、帯域閾値

1' (145) に格納する。

[0085] その後、ハイパーバイザ170からの閾値更新通知の受信を契機として、帯域制御部135は、仮想ドライバ130で保持していたIOPS値1' (150) および帯域値1' (155) をリセットする(ステップ605)。さらに、帯域制御部135は、SCSI I/Oキュー160に滞留しているI/Oがある場合は、該I/Oのデキューを再開(図5のステップ504、503、505)する。

[0086] 最後に、閾値計算部185は、閾値更新のタイマを起動し(ステップ606)、このタイマがステップ607でタイムアウトした際に、図中のステップ601から再度、次の制御間隔の閾値の更新処理を実行する。

[0087] 以上の処理により、ハイパーバイザ170では閾値計算部185が、HBA210のカウント回路230によって測定されたIOPSカウンタ1(235-1)の値および帯域カウンタ1(240-1)の値と、閾値更新ルール186に基づいて、IOPS閾値1(200-1)および帯域閾値1(205-1)を更新する。この結果、閾値計算部185は、ゲスト計算機105の仮想HBA303ごとのIOPS閾値1' (140-1) および帯域閾値1' (145-1) を更新する。

[0088] さらに、仮想ドライバ130の帯域制御部135がIOPS閾値1' (140) および帯域閾値1' (145) をハイパーバイザ170から取得して更新する。帯域制御部135は、更新されたIOPS閾値1' (140) および帯域閾値1' (145) に基づいて帯域制御を実施する。これにより、制御間隔あたりのデータ送受信量に基づく帯域制御を実現できる。

[0089] また、本発明では、HBA210が仮想HBA303毎に帯域とI/O回数を測定し、ハイパーバイザ170が制御間隔ごとに帯域閾値とI/O回数の閾値を更新してゲスト計算機105に通知し、ゲスト計算機105の仮想ドライバ130が制御間隔ごとにI/O回数と帯域(データ転送量)の閾値を超えないように仮想HBA303の帯域制御を行う。これにより、帯域の測定と閾値の算出及び帯域制御の実行を、HBA210とハイパーバイザ1

70及びゲスト計算機105で分散したので、一箇所に負荷が集中するのを防ぐことが可能となる。

[0090] なお、本発明において説明した計算機等の構成、処理部及び処理手段等は、それらの一部又は全部を、専用のハードウェアによって実現してもよい。

[0091] また、本実施例で例示した種々のソフトウェアは、電磁的、電子的及び光学式等の種々の記録媒体（例えば、非一時的な記憶媒体）に格納可能であり、インターネット等の通信網を通じて、コンピュータにダウンロード可能である。

[0092] また、本発明は上記した実施例に限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施例は本発明をわかりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。

請求の範囲

[請求項1]

プロセッサとメモリとを備えた計算機と、
前記計算機の資源を仮想化して1以上の仮想計算機に割り当てる仮想化部と、を備えた仮想計算機システムであって、
前記計算機は、
ストレージ装置に接続されるアダプタを有し、
前記アダプタは、
前記ストレージ装置との間でデータを送受信し、当該送受信したデータの転送量とI/O回数とを前記仮想計算機毎に測定する転送処理部と、
前記データの転送量とI/O回数を前記仮想計算機毎に格納するカウンタと、を有し、
前記仮想計算機は、
前記ストレージ装置との間で送受信するデータを保持するキューと、
前記データの転送量とI/O回数を制御する帯域制御部と、を有し、
前記仮想化部は、
前記アダプタのカウンタから取得した前記データの転送量とI/O回数に基づいて、前記データの転送量の上限値と、前記I/O回数の上限値を前記仮想計算機毎に演算する閾値演算部を有し、
前記帯域制御部は、
前記仮想化部が演算した前記データの転送量の上限値と、前記I/O回数の上限値を超えないように前記キューから出力するデータを制御することを特徴とする仮想計算機システム。

[請求項2]

請求項1に記載の仮想計算機システムであって、
前記仮想計算機は、
前記データの転送量を格納する帯域値と、前記I/O回数を格納す

る I / O 回数値と、前記データの転送量の上限值を格納する帯域閾値と、前記 I / O 回数の上限値を格納する I / O 回数閾値と、を有し、
前記帯域制御部は、

前記キューから出力する度にデータの転送量を前記帯域値に加算し、前記 I / O 回数を前記 I / O 回数値に加算して、前記帯域値が帯域閾値以下で、かつ前記 I / O 回数値が I / O 回数閾値以下であれば前記キューからデータを出力することを特徴とする仮想計算機システム。

[請求項3] 請求項2に記載の仮想計算機システムであって、

前記帯域制御部は、

前記キューから出力する度にデータの転送量を前記帯域値に加算し、前記 I / O 回数を前記 I / O 回数値に加算して、前記帯域値が帯域閾値を超え、または I / O 回数値が I / O 回数閾値を超える場合は、前記帯域値及び前記 I / O 回数閾値がリセットされるまでは前記キューからデータの出力を禁止することを特徴とする仮想計算機システム。

[請求項4] 請求項2に記載の仮想計算機システムであって、

前記閾値演算部は、

所定の周期が経過する度に、前記アダプタから前記データの転送量と I / O 回数を取得して、所定のルールに基づいて前記取得した前記データの転送量と I / O 回数から、前記データの転送量の上限值と、前記 I / O 回数の上限値を前記仮想計算機毎に演算し、当該演算した前記データの転送量の上限值と、前記 I / O 回数の上限値を前記仮想計算機に通知し、

前記帯域制御部は、

前記閾値演算部の前記データの転送量の上限值と前記 I / O 回数の上限値の通知を受信したことに基づいて、前記仮想計算機の帯域値と、前記 I / O 回数値とをリセットすることを特徴とする仮想計算機シ

ステム。

[請求項5]

請求項4に記載の仮想計算機システムであって、

前記所定のルールは、

前記仮想計算機ごとに予め設定された増分値を加算することを特徴とする仮想計算機システム。

[請求項6]

プロセッサとメモリとを備えた計算機と、前記計算機の資源を仮想化して1以上の仮想計算機に割り当てる仮想化部と、を備えた仮想計算機システムでデータ転送を制御する方法であって、

前記計算機は、ストレージ装置に接続されるアダプタを有し、

前記アダプタが、前記ストレージ装置との間でデータを送受信し、当該送受信したデータの転送量とI/O回数とを前記仮想計算機毎に測定する第1のステップと、

前記仮想化部は、前記アダプタから取得した前記データの転送量とI/O回数に基づいて、前記データの転送量の上限値と、前記I/O回数の上限値を前記仮想計算機毎に演算し、当該演算した前記データの転送量の上限値と、前記I/O回数の上限値を前記仮想計算機に通知する第2のステップと、

前記仮想計算機が、前記ストレージ装置との間で送受信するデータをキューに保持する第3のステップと、

前記仮想計算機が、前記データの転送量の上限値と、前記I/O回数の上限値を超えないように前記キューから出力するデータを制御する第4のステップと、

を含むことを特徴とする仮想計算機システムのデータ転送制御方法。

[請求項7]

請求項6に記載の仮想計算機システムのデータ転送制御方法であって、

前記第2のステップは、

前記仮想計算機が、前記データの転送量を帯域値に格納し、前記I/O回数をI/O回数値に格納し、前記データの転送量の上限値を帯

域閾値に格納し、前記 I / O 回数の上限値を I / O 回数閾値に格納するステップを含み、

前記第 4 のステップは、

前記仮想計算機が、前記キューから出力する度にデータの転送量を前記帯域値に加算し、前記 I / O 回数を前記 I / O 回数値に加算して、前記帯域値が帯域閾値以下で、かつ前記 I / O 回数値が I / O 回数閾値以下であれば前記キューからデータを出力することを特徴とする仮想計算機システムのデータ転送制御方法。

[請求項 8] 請求項 7 に記載の仮想計算機システムのデータ転送制御方法であって、

前記第 4 のステップは、

前記仮想計算機が、前記キューから出力する度にデータの転送量を前記帯域値に加算し、前記 I / O 回数を前記 I / O 回数値に加算して、前記帯域値が帯域閾値を超え、または I / O 回数値が I / O 回数閾値を超える場合は、前記帯域値及び前記 I / O 回数閾値がリセットされるまでは前記キューからデータの出力を禁止することを特徴とする仮想計算機システムのデータ転送制御方法。

[請求項 9] 請求項 7 に記載の仮想計算機システムのデータ転送制御方法であって、

前記第 2 のステップは、

前記仮想計算機は、所定の周期が経過する度に、前記アダプタから前記データの転送量と I / O 回数を取得して、所定のルールに基づいて前記取得した前記データの転送量と I / O 回数から、前記データの転送量の上限値と、前記 I / O 回数の上限値を前記仮想計算機毎に演算し、当該演算した前記データの転送量の上限値と、前記 I / O 回数の上限値を前記仮想計算機に通知するステップと、

前記仮想計算機は、前記データの転送量の上限値と前記 I / O 回数の上限値の通知を受信したことに基づいて、前記仮想計算機の帯域値

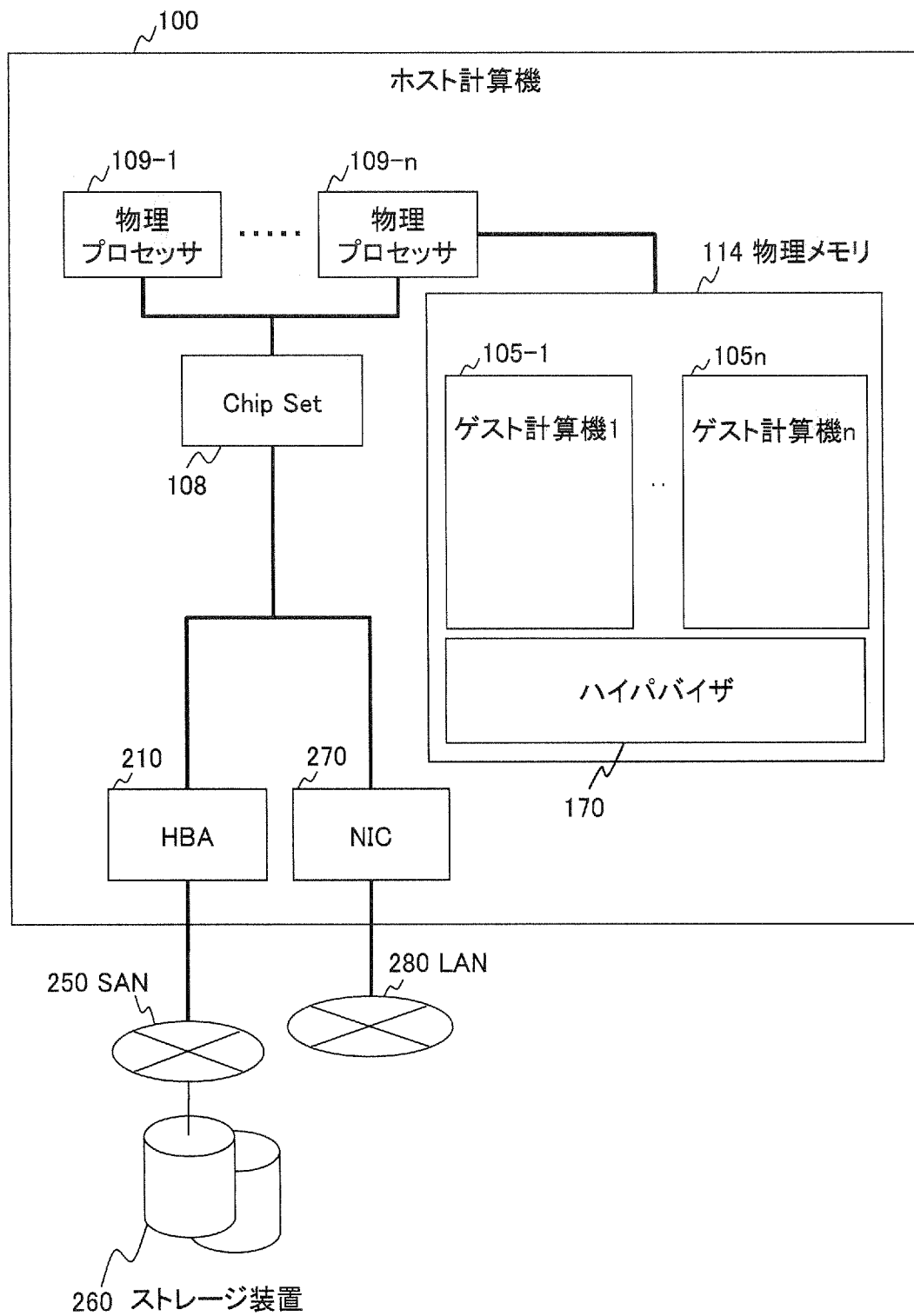
と、前記 1 / 0 回数値とをリセットするステップと、
を含むことを特徴とする仮想計算機システムのデータ転送制御方法。

[請求項10] 請求項 9 に記載の仮想計算機システムのデータ転送制御方法であって、

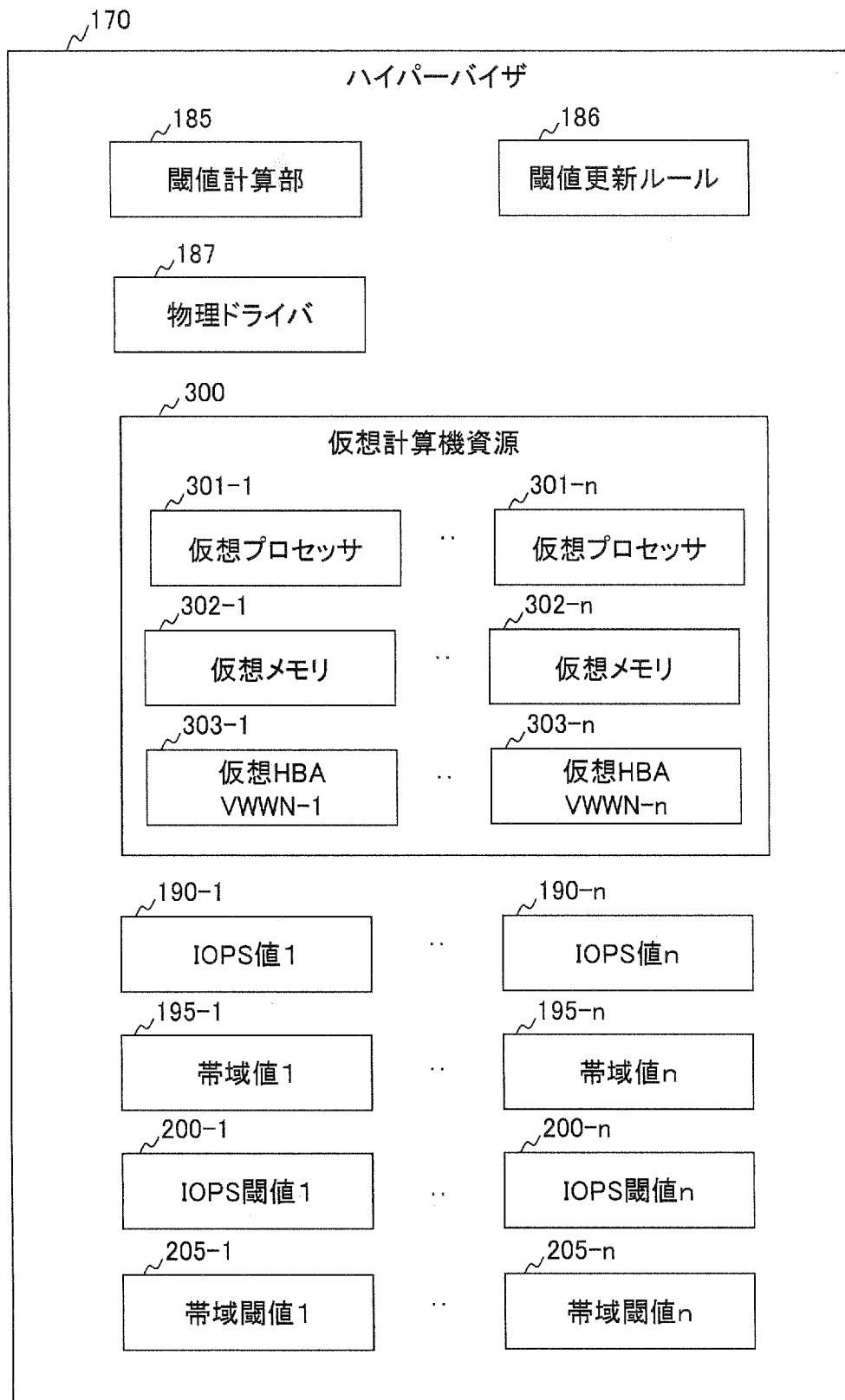
前記所定のルールは、

前記仮想計算機ごとに予め設定された増分値を加算することを特徴とする仮想計算機システムのデータ転送制御方法。

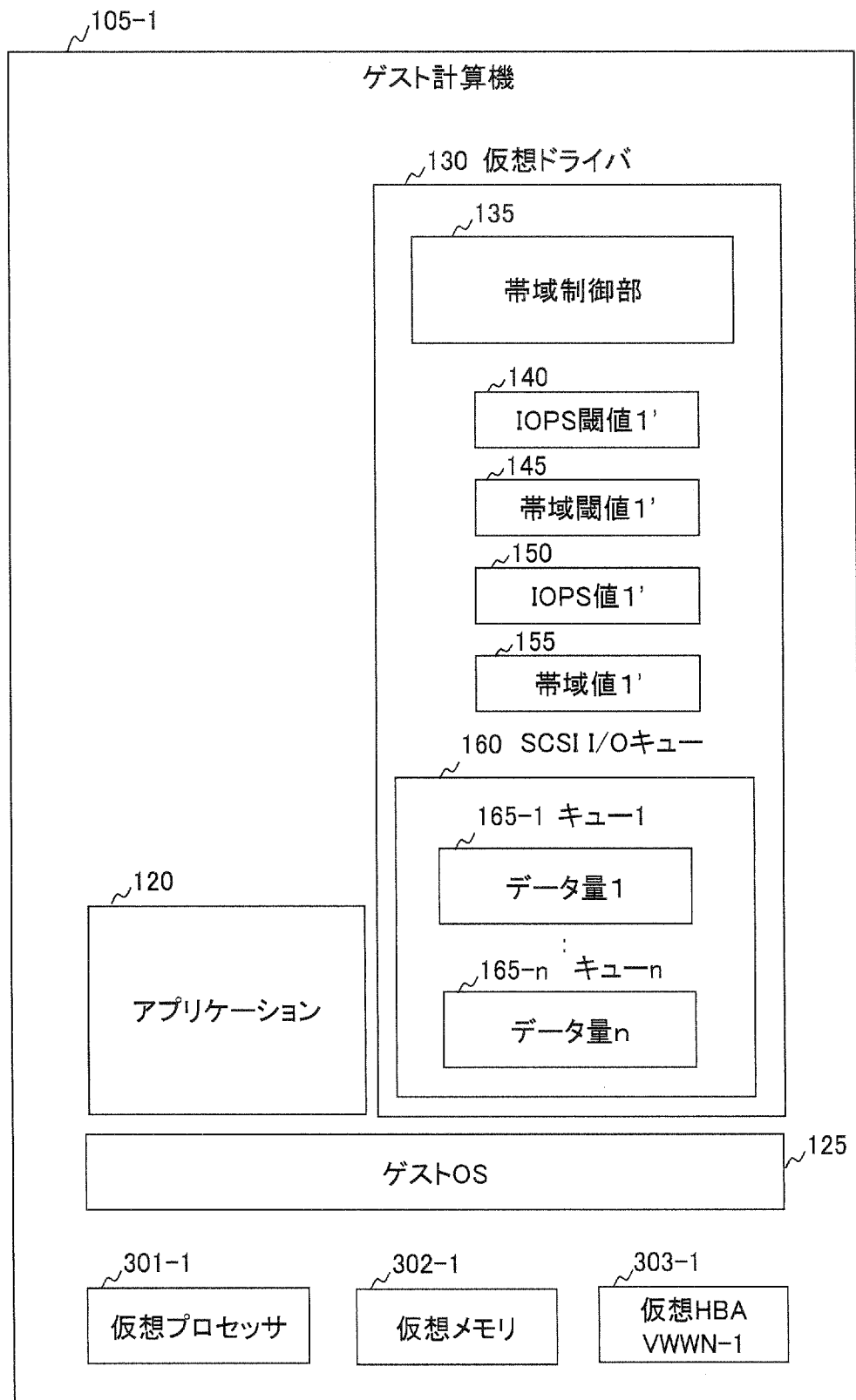
[図1]



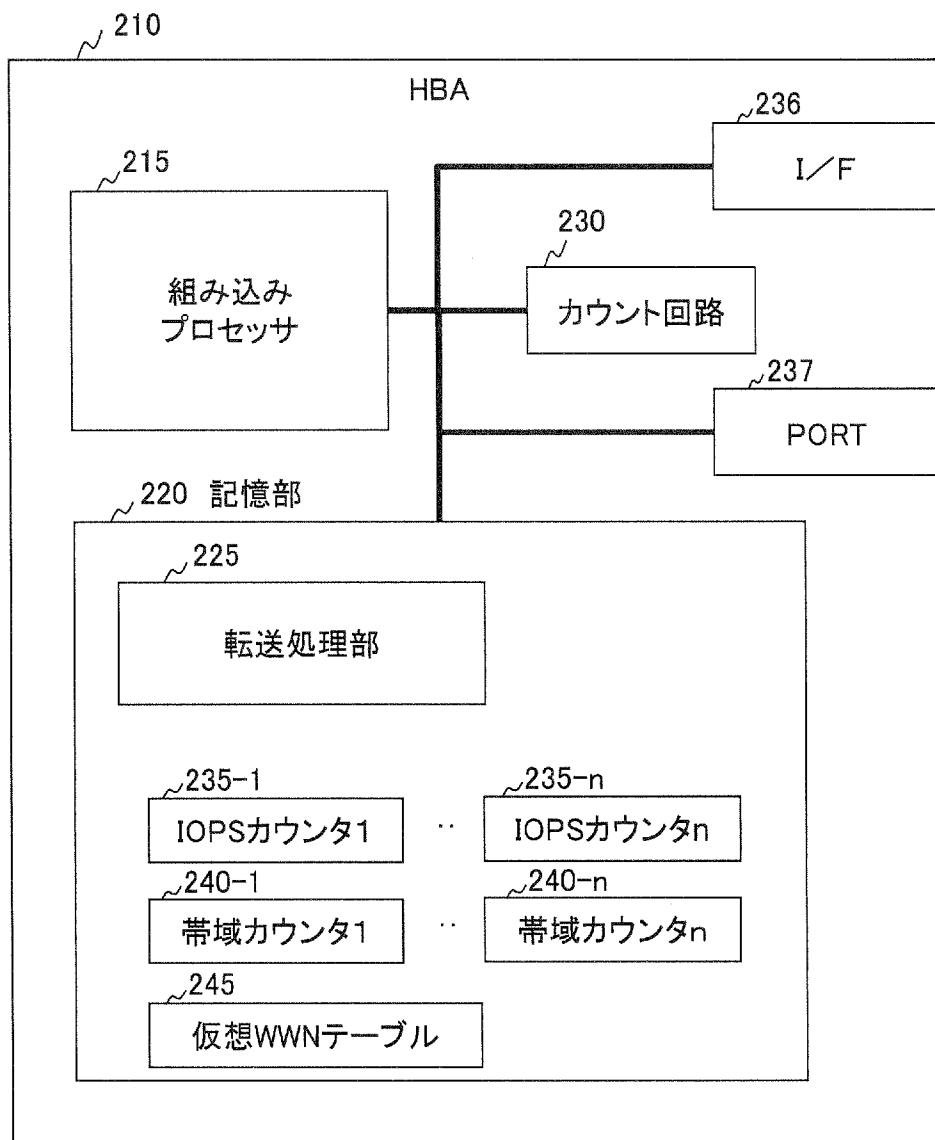
[図2]



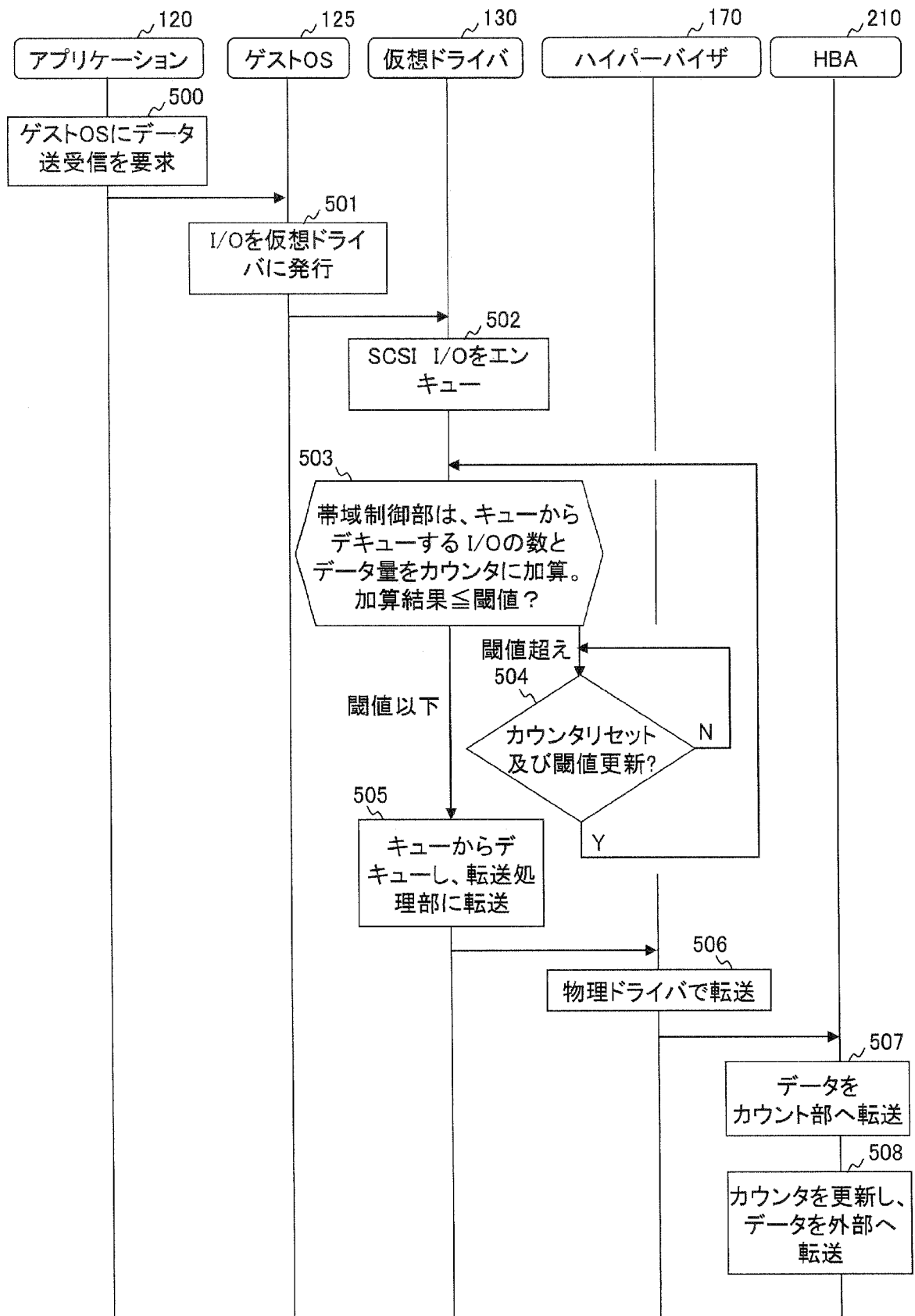
[図3]



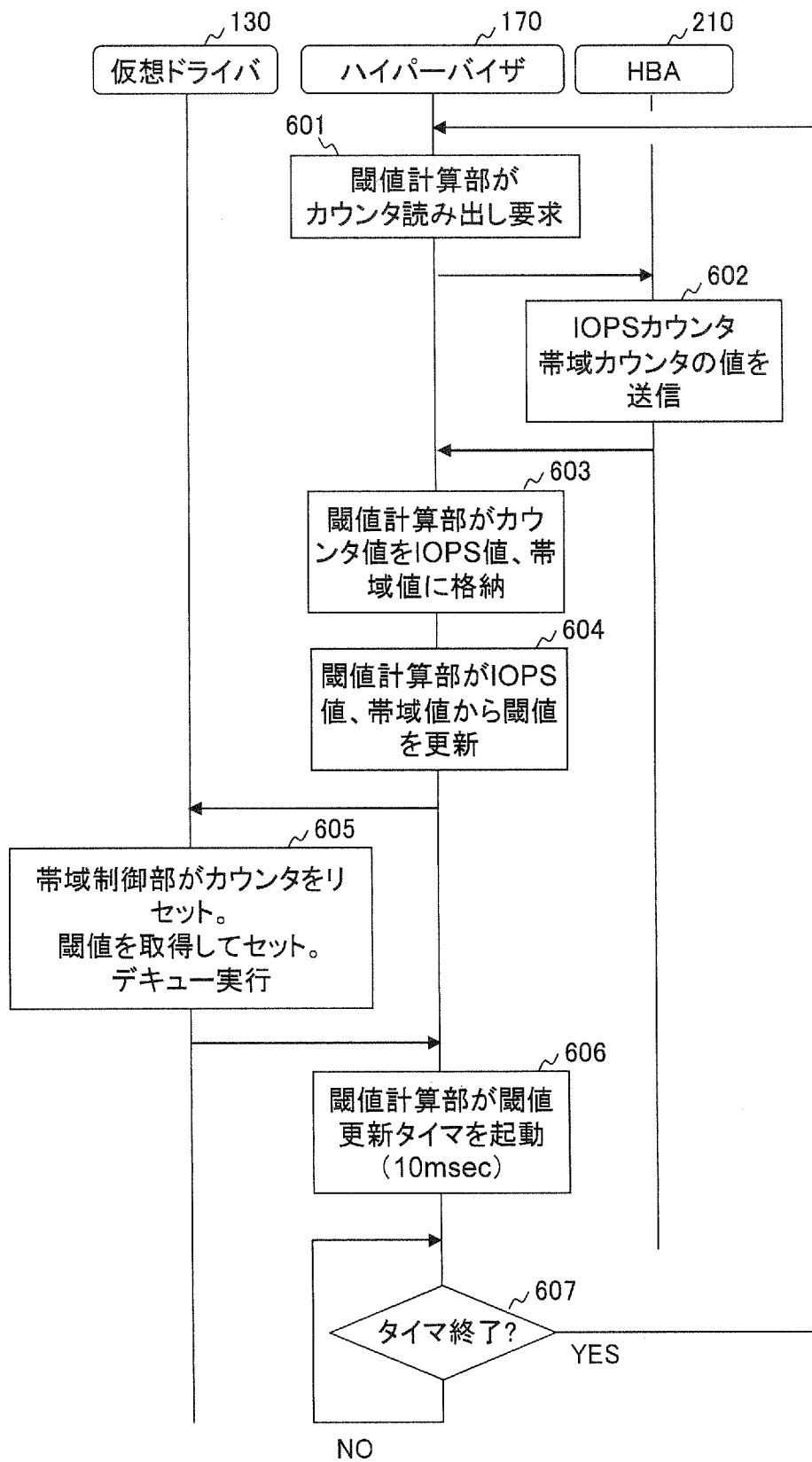
[図4]



[図5]



[図6]

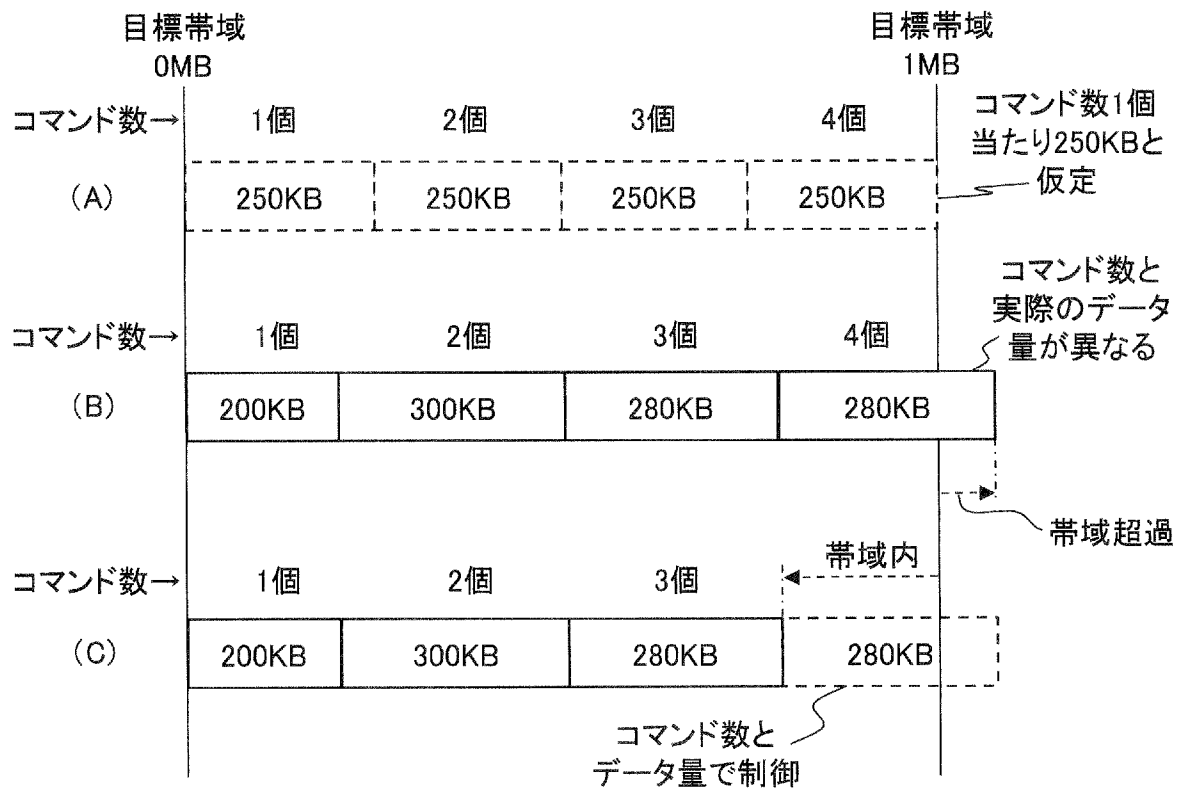


[図7]

186 閾値更新ルール

1861	ゲスト計算機	1	2	...	n
1862	ルール	増大	増大		維持

[図8]



[図9]

245 仮想WWNテーブル

仮想WWN	ゲスト計算機	IOPSカウンタ	帯域カウンタ
VWWN-1	1	1	1
VWWN-2	2	2	2
:	:	:	:

2451 2452 2453 2454

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/052377

A. CLASSIFICATION OF SUBJECT MATTER

G06F13/10(2006.01)i, G06F3/06(2006.01)i, G06F9/46(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F13/10, G06F3/06, G06F9/46

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2008-186211 A (Hitachi, Ltd.), 14 August 2008 (14.08.2008), paragraphs [0017] to [0020], [0040] to [0054]; fig. 1 (Family: none)	1-10
A	JP 2012-133630 A (Nomura Research Institute, Ltd.), 12 July 2012 (12.07.2012), paragraphs [0024] to [0061]; fig. 1 (Family: none)	1-10
A	JP 2012-123556 A (Hitachi Solutions, Ltd.), 28 June 2012 (28.06.2012), paragraphs [0017] to [0022]; fig. 2 (Family: none)	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
15 February, 2013 (15.02.13)

Date of mailing of the international search report
26 February, 2013 (26.02.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F13/10(2006.01)i, G06F3/06(2006.01)i, G06F9/46(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F13/10, G06F3/06, G06F9/46

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X	JP 2008-186211 A（株式会社日立製作所）2008.08.14, 段落【0017】－【0020】、【0040】－【0054】、図1（ファミリーなし）	1-10
A	JP 2012-133630 A（株式会社野村総合研究所）2012.07.12, 段落【0024】－【0061】、図1（ファミリーなし）	1-10
A	JP 2012-123556 A（株式会社日立ソリューションズ）2012.06.28, 段落【0017】－【0022】、図2（ファミリーなし）	1-10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

1 5 . 0 2 . 2 0 1 3

国際調査報告の発送日

2 6 . 0 2 . 2 0 1 3

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

稲葉 崇

5 T

3 8 5 9

電話番号 03-3581-1101 内線 3568