

BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (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, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

(57) 要約: 光ネットワークに接続される複数の計算機 1、2 を備える通信システムであって、前記光ネットワークは、主信号通信路 4 と、制御通信路 3 とを含み、各計算機 1、2 は、RDMA (Remote Direct Memory Access) 通信路を設定する制御部 1 4 と、前記制御部 1 4 により設定された前記 RDMA 通信路の設定情報と、前記主信号通信路 4 を設定するための設定情報とを含む設定要求を、前記制御通信路 3 を介して接続先の計算機に送信する通信部 1 1 と、前記主信号通信路 4 の設定情報に基づいて、前記主信号通信路 4 を確立する主信号送受信部 1 6 と、を備える。

明 細 書

発明の名称：

通信システム、計算機、通信方法、および、プログラム

技術分野

[0001] 本発明は、通信システム、計算機、通信方法、および、プログラムに関する。

背景技術

[0002] データセンタ内などの比較的通信距離の短い計算機システム同士のデータ通信において、通信処理速度を高速化するため計算機間でCPUを介さずにメモリにアクセスしてデータ転送を行う、RDMA (Remote Direct Memory Access) 技術が用いられている（非特許文献1）。計算機が光導波路により直接、光ネットワークに接続されることで中継ノードにおける信号損失の影響がなくなるため、RDMAの長距離データ通信への適用が可能になる。

[0003] 光通信においては、5Gの普及に伴い、地理的に分散した計算機間で発生する様々な通信要求に応じて、動的に光通信路を設定もしくは削除するための分散制御方法および集中制御方法が研究されている（非特許文献2、3）。

先行技術文献

非特許文献

[0004] 非特許文献1：IETF, “A Remote Direct Memory Access Protocol Specification”, RFC5040, October 2007

非特許文献2：竹房 あつ子, 「グリッドにおける計算資源と光パスネットワーク資源のコアロケーション実験」, 情報処理学会研究報告 2006-ARC-167 2006-HPC-105

非特許文献3：神野 正彦, 「フォトニックトランスポートネットワーク アーキテクチャと制御管理技術」, NTT技術ジャーナル 2007年10月

発明の概要

発明が解決しようとする課題

- [0005] 光ネットワークを介して接続された計算機システムにおいて、RDMAによるデータ転送の設定を行う場合、一般的な通信方式では下位レイヤから上位レイヤに向けて順番に、レイヤ毎に通信路が確立される。
- [0006] RDMA通信を開始する場合、まず、最下位層のレイヤ1の主信号通信路を確立し、その後、上位レイヤのRDMA通信路を確立する。レイヤ1の主信号通信路を設定するには数分程度の時間を要するが、レイヤ1の設定が完了するのを待ってRDMA通信路が設定される。
- [0007] このため、ユーザからの要求に応じて動的に主信号通信路およびRDMA通信路を確立する場合、RDMA通信が可能になるまでに時間がかかってしまうという課題がある。
- [0008] 本発明は、上記事情に鑑みてなされたものであり、本発明の目的は、RDMA通信の開始までに要する時間を短縮することにある。

課題を解決するための手段

- [0009] 上記目的を達成するため、本発明の一態様は、光ネットワークに接続される複数の計算機を備える通信システムであって、前記光ネットワークは、主信号通信路と、制御通信路とを含み、各計算機は、RDMA (Remote Direct Memory Access) 通信路を設定する制御部と、前記制御部により設定された前記RDMA通信路の設定情報と、前記主信号通信路を設定するための設定情報とを含む設定要求を、前記制御通信路を介して接続先の計算機に送信する通信部と、前記主信号通信路の設定情報に基づいて、前記主信号通信路を確立する主信号送受信部と、を備える。
- [0010] 本発明の一態様は、接続先の計算機と光ネットワークで接続される計算機であって、前記光ネットワークは、主信号通信路と、制御通信路とを含み、前記計算機は、RDMA (Remote Direct Memory Access) 通信路を設定する制御部と、前記制御部により設定された前記RDMA通信路の設定情報と、前記主信号通信路を設定するための設定情報とを含む設定要求を、前記制御通信路を介して接続先の計算機に送信する通信部と、前記主信号通信路の設定情報に基づいて、前記主信号通信路を確立する主信号送受信部と、を備える。

[0011] 本発明の一態様は、光ネットワークに接続される複数の計算機を備える通信システムが行う通信方法であって、前記光ネットワークは、主信号通信路と、制御通信路とを含み、各計算機は、RDMA (Remote Direct Memory Access) 通信路を設定するステップと、前記設定するステップで設定された設定情報と、前記主信号通信路を設定するための設定情報とを含む設定要求を、前記制御通信路を介して接続先の計算機に送信するステップと、前記主信号通信路の設定情報に基づいて、前記主信号通信路を確立するステップと、を行う。

[0012] 本発明の一態様は、上記計算機としてコンピュータを機能させるプログラムである。

発明の効果

[0013] 本発明によれば、RDMA通信の開始までに要する時間を短縮することができる。

図面の簡単な説明

[0014] [図1]本発明の実施形態の通信システムの構成図である。

[図2]比較例の通信路の設定処理を説明するための説明図である。

[図3]本実施形態の通信路の設定処理を説明するための説明図である。

[図4]本実施形態の通信路の設定処理を示すシーケンス図である。

[図5]図4のS 1 2～S 1 7のRDMA通信路の設定を詳細に示すシーケンス図である。

[図6]図4の図4のS 1 8、S 2 1のメッセージを説明する説明図である。

[図7]図2の比較例の通信路の設定処理を示すシーケンス図である。

[図8]ハードウェア構成例である。

発明を実施するための形態

[0015] 以下、本発明の実施の形態について、図面を参照して説明する。

[0016] <通信システムの構成>

図1は、本発明の実施形態の通信システムのシステム構成図である。本実施形態の通信システムは、光ネットワーク（光伝送路）に接続される複数の

計算機 1、2 を備える。

[0017] 光ネットワークは、制御通信路 3 と、主信号通信路 4 とを含む。制御通信路 3 は、制御信号を伝送する通信路である。主信号通信路 4 は、レイヤ 1（下位レイヤ）の大容量かつ低遅延の通信路である。制御通信路 3 は常時、設定され、主信号通信路 4 は、ユーザからの要求に応じて動的に設定されるものとする。

[0018] 計算機 1、2 は、RDMA（Remote Direct Memory Access）通信によるデータ転送を行う。RDMA 通信は、一方の計算機 1 のメモリから、他方の計算機のメモリへデータを DAM（Direct Memory Access）転送する通信である。RDMA 通信では、計算機 1、2 のオペレーティングシステムを経由せずにデータが転送されるため、高スループットおよび低遅延のデータ転送を実現できる。

[0019] 本実施形態では、レイヤ 1 の主信号通信路 4（OTN プロトコル）の上位レイヤとして、RDMA 通信路（RDMA プロトコル）が、計算機 1、2 にエンド・ツー・エンドで主信号通信路 4 上に設定される。すなわち、本実施形態の通信システムは、RDMA over レイヤ 1 伝送を実現する。

[0020] 各計算機 1、2 は、RDMA 通信部 1 1（通信部）と、通信切替部 1 2 と、データ記憶部 1 3 と、RDMA 制御部 1 4（制御部）と、制御信号送受信部 1 5 と、主信号送受信部 1 6 とを備える。

[0021] RDMA 通信部 1 1 は、RDMA 通信を行うアプリケーションである。RDMA 通信部 1 1 は、CPU およびメモリを用いて実装される。RDMA 通信部 1 1 は、主信号通信路 4 の設定および削除を制御する。例えば、計算機 1 の RDMA 通信部 1 1 は、RDMA 制御部 1 4 により設定された RDMA 通信路の設定情報と、主信号通信路 4 を設定するための設定情報とを含む設定要求（制御メッセージ）を、制御通信路 3 を介して接続先の計算機 2 に送信する。本実施形態の設定要求は、RDMA 通信路および主信号通信路 4 の設定を要求する。RDMA 通信部 1 1 は、RDMA 通信路の設定を RDMA 制御部 1 4 に指示する。

[0022] 通信切替部 1 2 は、例えば Root Complex、PCI スイッチなどであって、RDMA 通信部 1 1 の通信先のデバイスを切り替える。

- [0023] データ記憶部 1 3 は、メモリまたはストレージなどであって、各種データを記憶する。データ記憶部 1 3 がメモリの場合、RDMA通信部 1 1 はデータ記憶部 1 3 上で動作する。
- [0024] RDMA制御部 1 4 は、RDMA通信によるデータ転送を制御する。RDMA制御部 1 4 には、DMA (Direct Memory Access) コントローラを用いることができる。DMAコントローラは、DMA転送を制御する専用のアプリケーション (ICチップ) であって、例えばFPGA NIC (Field Programmable Gate Array Network Interface Card) 上に実装される。本実施形態のRDMA制御部 1 4 は、RDMA通信路を設定する。具体的には、RDMA制御部 1 4 は、RDMA通信部 1 1 からの指示に基づいてRDMA通信用のメモリ領域を確保し、確保した前記メモリ領域を含む設定情報を、RDMA通信部 1 1 に通知し、RDMA通信路を設定する。
- [0025] 制御信号送受信部 1 5 は、制御信号路 3 を流れる制御信号を送受信する。制御信号送受信部 1 5 には、NIC (Network Interface Card) を用いることができる。
- [0026] 主信号送受信部 1 6 は、主信号通信路 4 を流れる主信号を送受信する。主信号送受信部 1 6 には、レーザなどを搭載したNICを用いることができる。主信号送受信部 1 6 は、主信号通信路 4 の設定情報に基づいて、主信号通信路 4 を確立する。
- [0027] <通信路の設定処理>
- 次に、主信号通信路 4 およびRDMA通信路の設定処理について説明する。なお、制御通信路 3 は、あらかじめ設定されているものとする。主信号通信路 4 およびRDMA通信路の設定処理には、以下の処理 A、B、Cが含まれる。処理 A および B と、処理 C とは、並行して実施可能である。
- [0028] 処理 A : RDMA通信を行う計算機 1、2 は、制御通信路 3 を介して、主信号通信路 4 を設定するための制御メッセージ (設定要求、設定応答) を送受信する。
- [0029] 処理 B : 計算機 1、2 は、制御メッセージに基づいて主信号通信路 4 を設定する。

[0030] 処理C：計算機1、2は、RDMA通信路を設定するための制御メッセージ（設定要求、設定応答）を送受信し、RDMA通信路を設定する。

[0031] 図2は、本実施形態の比較例における、通信路の設定処理を説明するための説明図である。図2では、処理A、処理B、処理Cの順番で、シーケンシャルに主信号通信路4およびRDMA通信路を設定する。すなわち、レイヤ毎に下位レイヤから上位レイヤに向けて順番に通信路が確立され、主信号通信路4が確立するまで処理CのRDMA通信路の設定は開始されない。比較例では、処理Cは、主信号通信路4を介して行われる。

[0032] このため、RDMA通信路が確立されるまでの時間は、処理A、処理Bおよび処理Cの各処理時間の合計となり、設定処理の開始からRDMA通信が開始されるまでに長い時間を要する。処理Bの処理時間は、数分程度と長い時間がかかる。

[0033] 図3は、本実施形態の通信路の設定処理を説明するための説明図である。

[0034] 本実施形態では、処理CのRDMA通信路の設定要求と、処理Aの主信号通信路4の設定要求とを、1つの設定要求として一括して送信する。制御信号通信路3は、主信号通信路4の設定だけでなくRDMA通信路を設定するためにも使用される。制御通信路3は、主信号通信路4を設定（もしくは削除）するために、あらかじめ確立されている。これにより、本実施形態では、設定処理の開始からRDMA通信が開始されるまでの時間を短くすることができる。

[0035] 本実施形態では、まず、計算機1のRDMA通信部11およびRDMA制御部14が、処理C1のRDMA通信路の設定を実行する。計算機1のRDMA通信部11は、RDMA通信路の設定指示をRDMA制御部14に送出する。RDMA制御部14は、前記指示に基づいて、RDMA通信用のメモリ領域を確保し、メモリ領域を含む設定情報をRDMA通信部11に通知する。

[0036] 次に、処理A1で、計算機1のRDMA通信部11は、RDMA制御部14から通知されたRDMA通信路の設定情報と、主信号通信路4を設定するための設定情報とを含む設定要求を、制御信号送受信部15および制御通信路3を介して接続先の計算機2に送信する。計算機2は、設定要求を受信し、RDMA通信路

の設定応答と、主信号通信路 4 の設定応答とを、計算機 1 に送信する。

[0037] そして、処理 B 1 で、計算機 1 の RDMA 通信部 1 1 は、主信号送受信部 1 6 を制御して、主信号通信路 4 の設定応答に基づいて接続先の計算機 2 と主信号通信路 4 を確立する。これにより、RDMA 通信路が、主信号通信路 4 上で確立される。

[0038] 本実施形態では、RDMA 通信路の設定要求と主信号通信路 4 の設定要求を、1 つの制御メッセージとして一括して制御通信路 3 を介して送信する。これにより、RDMA 通信路が確立して RDMA 通信が開始されるまでの時間を、図 2 の比較例より短縮することができる。

[0039] 図 4 は、図 3 に示す本実施形態の通信路の設定処理を示すシーケンス図である。

[0040] 制御通信路 3 は、計算機 1、2 との間で事前に確立されている（S 1 1）。

[0041] ユーザの要求を受け付けて、計算機 1 の RDMA 通信部 1 1 は、送信元アドレスおよび送信先アドレスを取得して、RDMA 通信路を設定するための設定指示を、RDMA 制御部 1 4 に送出する（S 1 2）。ユーザの要求には、接続先の計算機 2 のノード識別子が指定されている。

[0042] RDMA 制御部 1 4 は、設定指示にしたがって RDMA 用のメモリを確保するなどの設定を実行し、設定情報（設定結果）を RDMA 通信部 1 1 に送出する（S 1 3）。

[0043] 計算機 1 の RDMA 通信部 1 1 は、RDMA 通信路の設定情報と、主信号通信路 4 を設定するための設定情報とを含む設定要求を、制御信号送受信部 1 5 および制御通信路 3 を介して計算機 2 に送信する（S 1 4）。

[0044] 主信号通信路 4（光パス）の設定または削除は、GMPLS（Generalized Multi Protocol Label Switching）などの分散制御（非特許文献 2）、または SDN（Software Defined Network）もしくは NMS（Network Management SYSTEM）/ EMS（Element Management SYSTEM）などの集中制御（非特許文献 3）などを用いることができる。

- [0045] GMPLSの分散制御を用いた場合、主信号通信路4の設定情報（設定要求）には、メッセージ識別子(Session ID)、主信号通信路4のローカル端点のノード識別子、主信号通信路4のリモート端点のノード識別子、隣接ホップ/端点識別子、波長ラベル識別子（候補群）、伝送容量などが含まれる。
- [0046] SDNの集中制御を用いた場合、計算機1が集中制御サーバ（不図示）を介して計算機2に送信する設定情報（設定要求）には、メッセージ識別子(Session ID)、主信号通信路4のローカル端点のノード識別子、主信号通信路4のローカル端点としてのインターフェース/端点識別子、主信号通信路のリモート端点のノード識別子、波長ラベル識別子、伝送容量などが含まれる。
- [0047] 計算機2のRDMA通信部11は、制御信号送受信部15を介してS14の設定要求を受信する。RDMA通信部11は、RDMA通信を設定するための設定指示を、RDMA制御部14に送出する（S15）。この設定指示には、設定要求に含まれるRDMA通信路の設定情報が含まれる。RDMA制御部14は、設定指示にしたがってRDMA用のメモリを確保するなどの設定を実行し、設定情報をRDMA通信部11に送出する（S16）。これにより、RDMA通信路が設定される。RDMA通信部11は、RDMA通信路の設定情報を含む設定応答を、制御信号送受信部15および制御通信路3を介して計算機1に送信する（S17）。
- [0048] また、計算機2のRDMA通信部11は、S14の設定要求に含まれる主信号通信路4の設定情報に基づいて主信号通信路4を設定する。RDMA通信部11は、主信号通信路4の設定情報を含む設定応答を、制御信号送受信部15および制御通信路3を介して計算機1に送信する（S17）。
- [0049] GMPLSの分散制御を用いた場合、計算機2が送信する主信号通信路4の設定情報（設定応答）には、メッセージ識別子(Session ID)、主信号通信路4のローカル端点のノード識別子、主信号通信路4のリモート端点のノード識別子、隣接ホップ/端点識別子、伝送容量、主信号通信路4として確定した隣接ホップ/端点識別子および波長ラベル識別子などが含まれる。
- [0050] SDNの集中制御を用いた場合、集中制御サーバを介して計算機1に送信される主信号通信路4の設定情報（設定応答）には、メッセージ識別子(Session

ID)、主信号通信路4のローカル端点のノード識別子、主信号通信路4のローカル端点としてのインターフェース/端点識別子、主信号通信路のリモート端点のノード識別子、波長ラベル識別子、伝送容量などが含まれる。

[0051] 計算機1のRDMA通信部11は、RDMA通信路の設定応答と、主信号通信路4の設定応答とを、それぞれ受信する。計算機1のRDMA通信部11は、主信号通信路4の設定応答を受信すると、主信号送受信部16に主信号通信路4の起動を指示する(S18)。計算機2のRDMA通信部11は、主信号通信路4の設定応答を送信後、主信号送受信部16に主信号通信路4の起動を指示する(S19)。

[0052] 計算機1、2の主信号送受信部16は、主信号通信路4の設定要求および設定応答の設定情報に基づいて、主信号通信路4を確立する(S20)。各主信号送受信部16が起動指示を受け付けてから、実際に主信号通信路4が確立されるまでの時間は、数分程度との長い時間を要する。

[0053] 計算機1、2の各主信号送受信部16は、主信号通信路4の確立後、起動指示に対する応答をRDMA通信部11に送出する(S21、S22)。これにより、主信号通信路4上にRDMA通信路が確立され(S23)、RDMA通信が開始される。本実施形態では、主信号通信路4が確立する前にRDMA通信路は設定されているため、S20の主信号通信路4の確立から、S23のRDMA通信路の設定までの時間を短縮することができる。

[0054] 図5は、図4のS12～S17のRDMA通信部11とRDMA制御部14による、RDMA通信路の設定処理の詳細を示すシーケンス図である。

[0055] 計算機1のRDMA通信部11は、送信元アドレスおよび送信先アドレスを取得して(S12-1)、RDMA通信路の設定指示を、RDMA制御部14に送出する(S12-2)。設定指示(例えば“open”)には、以下の指示が含まれる。

[0056] ・SQ(Send Queue)、CQ(Completion Queue)およびQP(Queue Pair)の作成
・仮想メモリアドレスと物理メモリアドレスの対応づけ
・メモリの割当て

RDMA制御部14は、設定指示にしたがってRDMA用のメモリを確保するなどの設定を実行し（S13-1）、設定情報（設定結果）をRDMA通信部11に送出する（S13-2）。

[0057] RDMA通信部11は、RDMA通信路の設定情報を含む設定要求を、制御通信路3を介して計算機2に送信する（S14）。設定情報には、設定指示にしたがって設定された設定情報が含まれる。RDMA通信路の設定要求には、例えば、サービスタイプがRC（Reliable Connection(RC)の場合は“Connect”で、サービスタイプがUD（Unreliable Datagram）の場合は“mmap”が用いられる。本実施形態の設定要求には、主信号通信路4の設定情報も含まれる。

[0058] 計算機2のRDMA通信部11は、制御通信路3を介して設定要求を受信すると、RDMA通信路の設定指示を、RDMA制御部14に送出する（S15）。設定指示（例えば“open”）には、以下の指示が含まれる。

[0059] ・ RQ(Receive Queue)、CQ(Completion Queue)およびQP(Queue Pair)の作成
・ 仮想メモリアドレスと物理メモリアドレスの対応づけ
・ メモリの割当て

RDMA制御部14は、設定指示にしたがってRDMA用のメモリを確保するなどの設定を実行し（S16-1）、設定情報をRDMA通信部11に送出する（S16-2）。

[0060] RDMA通信部11は、制御通信路3を介して、RDMA通信路の設定情報を含む設定応答を、計算機1に送信する（S17）。これにより、RDMA通信路が設定される。

[0061] 図6は、図4のS18およびS21（またはS19およびS22）のRDMA通信部11と主信号送受信部16との間のメッセージを説明する説明図である。

[0062] RDMA通信部11は、S18、S19起動指示として、以下の項目を含むメッセージを主信号送受信部16に送出する。

[0063] ・ インターフェース 起動
・ レーザon

- ・ 波長制御on
- ・ 変調方式（例 16QAM/32QAM、QPSK）
- ・ エラー訂正方式(Reed Solomon)

なお、主信号通信路4の停止指示の場合は、上記メッセージは、「インターフェース 停止」、「レーザoff」、「波長制御off」となる。

[0064] 主信号送受信部16は、S21、S22の応答として、主信号通信路4の確立に成功または失敗したかの処理結果を示すメッセージを送出する。

[0065] 図4に示す本実施形態では、計算機1のRDMA通信部11は、RDMA通信路の設定要求と、主信号通信路4の設定要求とを、1つの設定要求として一体化して、接続先の計算機2に送信する。これにより、RDMA通信路の設定要求を、主信号通信路4が確立してから送信する場合に比べて、RDMA通信が開始されるまでの時間を短くすることができる。

[0066] 一方、図7は、図2に示す比較例の通信路の設定処理を示すシーケンス図である。図示するシーケンス図では、S12～S13で主信号通信路4を設定するための設定要求を送受信し（処理A）、S14～S18で主信号通信路4を設定し（処理B）、S19～S21でRDMA通信路を設定する（処理C）。比較例におけるRDMA通信路の設定要求（S19）は、主信号通信路4の確立後に、主信号通信路4を介して送信される。したがって、RDMA通信路が確立されるまでの時間は、図4に示す本実施形態より長い時間を要する。

[0067] <作用効果>

以上説明した本実施形態の通信システムは、光ネットワークに接続される複数の計算機1、2を備える通信システムであって、前記光ネットワークは、主信号通信路4と、制御通信路3とを含み、各計算機1、2は、RDMA通信路を設定するRDMA制御部14と、RDMA制御部14により設定されたRDMA通信路の設定情報と、主信号通信路4を設定するための設定情報とを含む設定要求を、制御通信路3を介して接続先の計算機に送信するRDMA通信部11と、主信号通信路4の設定情報に基づいて、主信号通信路4を確立する主信号送受信部16と、を備える。

[0068] このように本実施形態では、RDMA通信路の設定要求と、主信号通信路4の設定要求とを統合して、制御通信路3を介して一括して送信する。これにより、本実施形態では、比較例の通信方式のようにレイヤ1の主信号通信路の確立を待つことなくRDMA通信路の設定を進めることができるため、短い時間でRDMAのデータ通信の開始が可能となる。すなわち、RDMA通信の開始までに要する時間を短縮することができる。

[0069] <ハードウェア構成>

上記説明した計算機1、2は、例えば、図8に示すような汎用的なコンピュータシステムを用いることができる。図示するコンピュータシステムは、CPU (Central Processing Unit、プロセッサ) 901と、メモリ902と、ストレージ903 (HDD: Hard Disk Drive、SSD: Solid State Drive) と、通信装置904と、入力装置905と、出力装置906とを備える。メモリ902およびストレージ903は、記憶装置である。このコンピュータシステムにおいて、CPU901がメモリ902上にロードされた所定のプログラムを実行することにより、各計算機1、2の各機能が実現される。

[0070] また、計算機1、2は、1つのコンピュータで実装されてもよく、あるいは複数のコンピュータで実装されても良い。また、計算機1、2は、コンピュータに実装される仮想マシンであっても良い。計算機用のプログラムは、HDD、SSD、USB (Universal Serial Bus) メモリ、CD (Compact Disc)、DVD (Digital Versatile Disc)などのコンピュータ読取り可能な記録媒体に記憶することも、ネットワークを介して配信することもできる。

[0071] なお、本発明は上記実施形態に限定されるものではなく、その要旨の範囲内で数々の変形が可能である。

符号の説明

[0072] 1、2 : 計算機

11 : RDMA通信部 (通信部)

12 : 通信切替部

13 : データ記憶部

1 4 : RDMA制御部 (制御部)

1 5 : 制御信号送受信部

1 6 : 主信号送受信部

3 : 制御通信路

4 : 主信号通信路

請求の範囲

- [請求項1] 光ネットワークに接続される複数の計算機を備える通信システムであって、
前記光ネットワークは、主信号通信路と、制御通信路とを含み、
各計算機は、
RDMA (Remote Direct Memory Access) 通信路を設定する制御部と、
前記制御部により設定された前記RDMA通信路の設定情報と、前記主信号通信路を設定するための設定情報とを含む設定要求を、前記制御通信路を介して接続先の計算機に送信する通信部と、
前記主信号通信路の設定情報に基づいて、前記主信号通信路を確立する主信号送受信部と、を備える
通信システム。
- [請求項2] 前記通信部は、前記RDMA通信路の設定を前記制御部に指示し、
前記制御部は、前記指示に基づいてRDMA通信用のメモリ領域を確保し、確保した前記メモリ領域を含む設定情報を、前記通信部に通知する
請求項1に記載の通信システム。
- [請求項3] 接続先の計算機と光ネットワークで接続される計算機であって、
前記光ネットワークは、主信号通信路と、制御通信路とを含み、
前記計算機は、
RDMA (Remote Direct Memory Access) 通信路を設定する制御部と、
前記制御部により設定された前記RDMA通信路の設定情報と、前記主信号通信路を設定するための設定情報とを含む設定要求を、前記制御通信路を介して接続先の計算機に送信する通信部と、
前記主信号通信路の設定情報に基づいて、前記主信号通信路を確立する主信号送受信部と、を備える

計算機。

[請求項4] 光ネットワークに接続される複数の計算機を備える通信システムが行う通信方法であって、

前記光ネットワークは、主信号通信路と、制御通信路とを含み、
各計算機は、

RDMA (Remote Direct Memory Access) 通信路を設定するステップと、

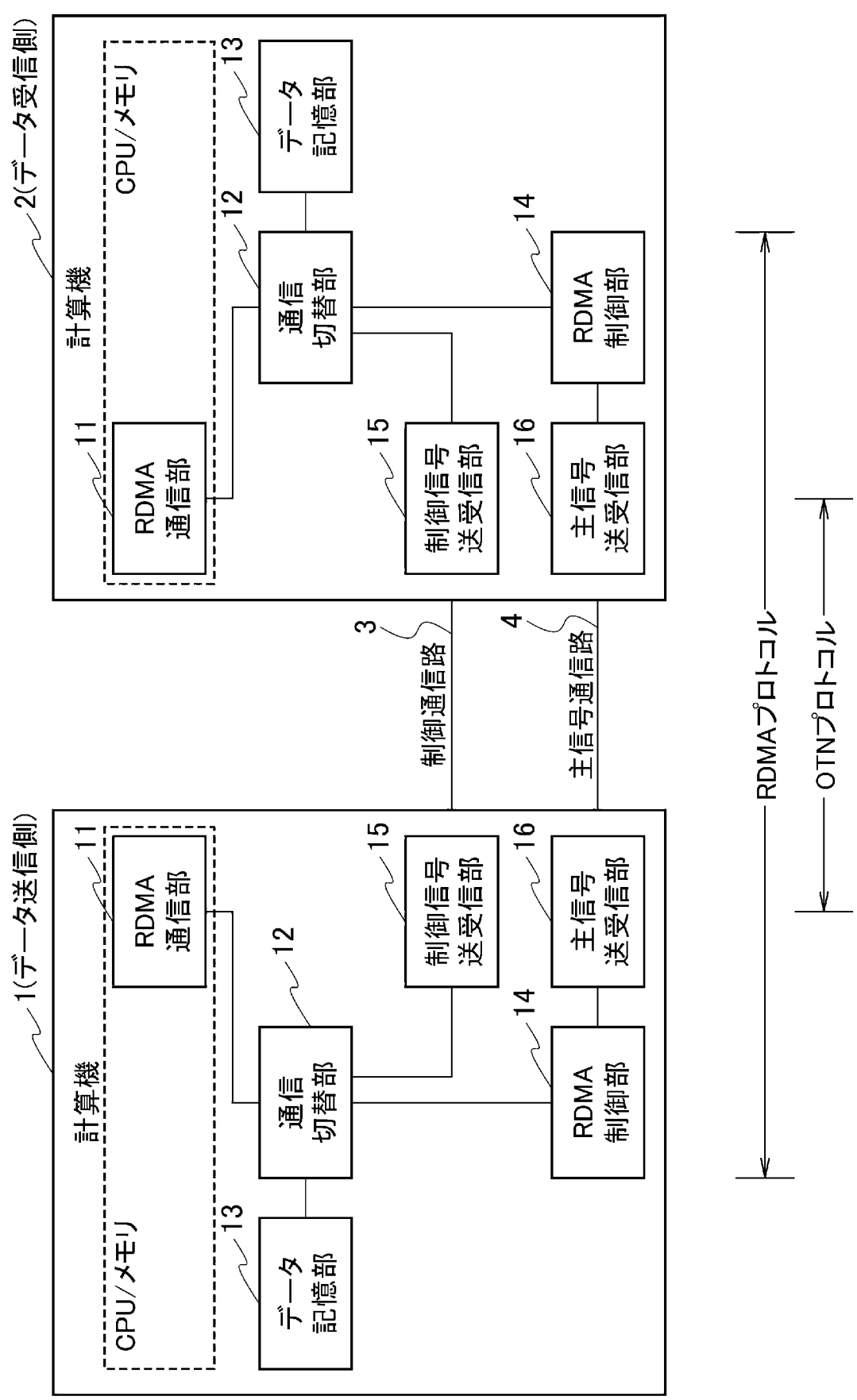
前記設定するステップで設定された設定情報と、前記主信号通信路を設定するための設定情報とを含む設定要求を、前記制御通信路を介して接続先の計算機に送信するステップと、

前記主信号通信路の設定情報に基づいて、前記主信号通信路を確立するステップと、を行う

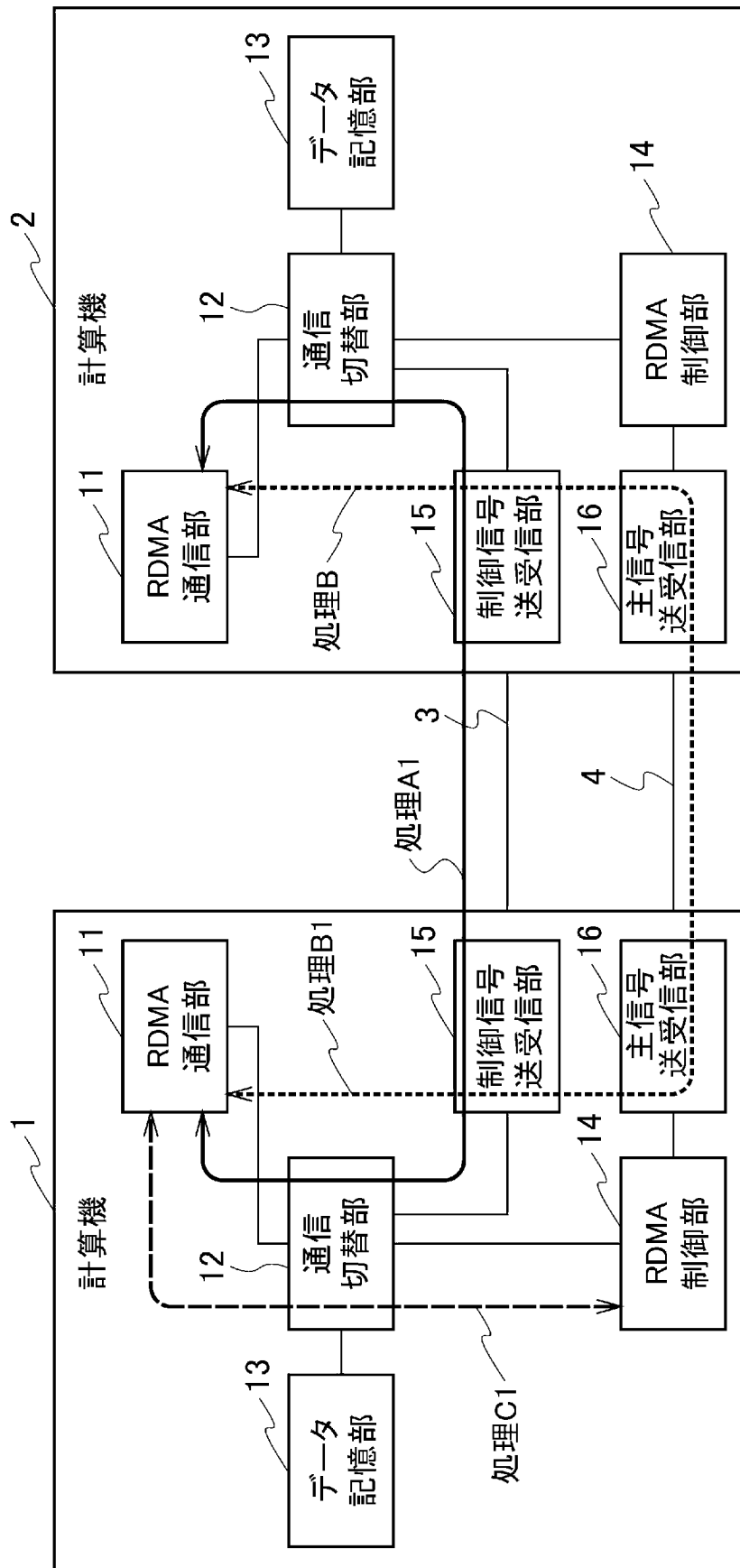
通信方法。

[請求項5] 請求項3に記載の計算機としてコンピュータを機能させるプログラム。

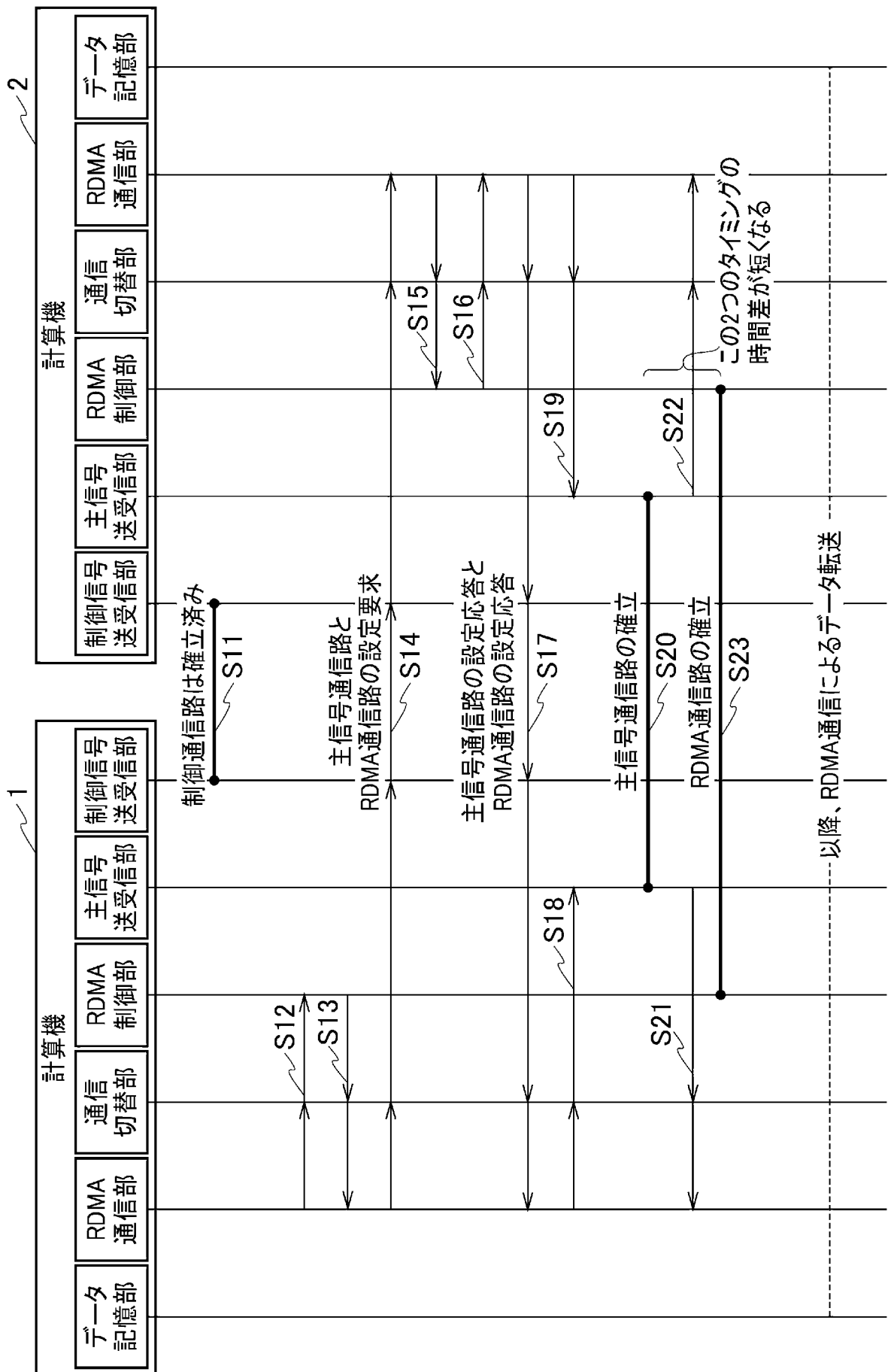
[図1]



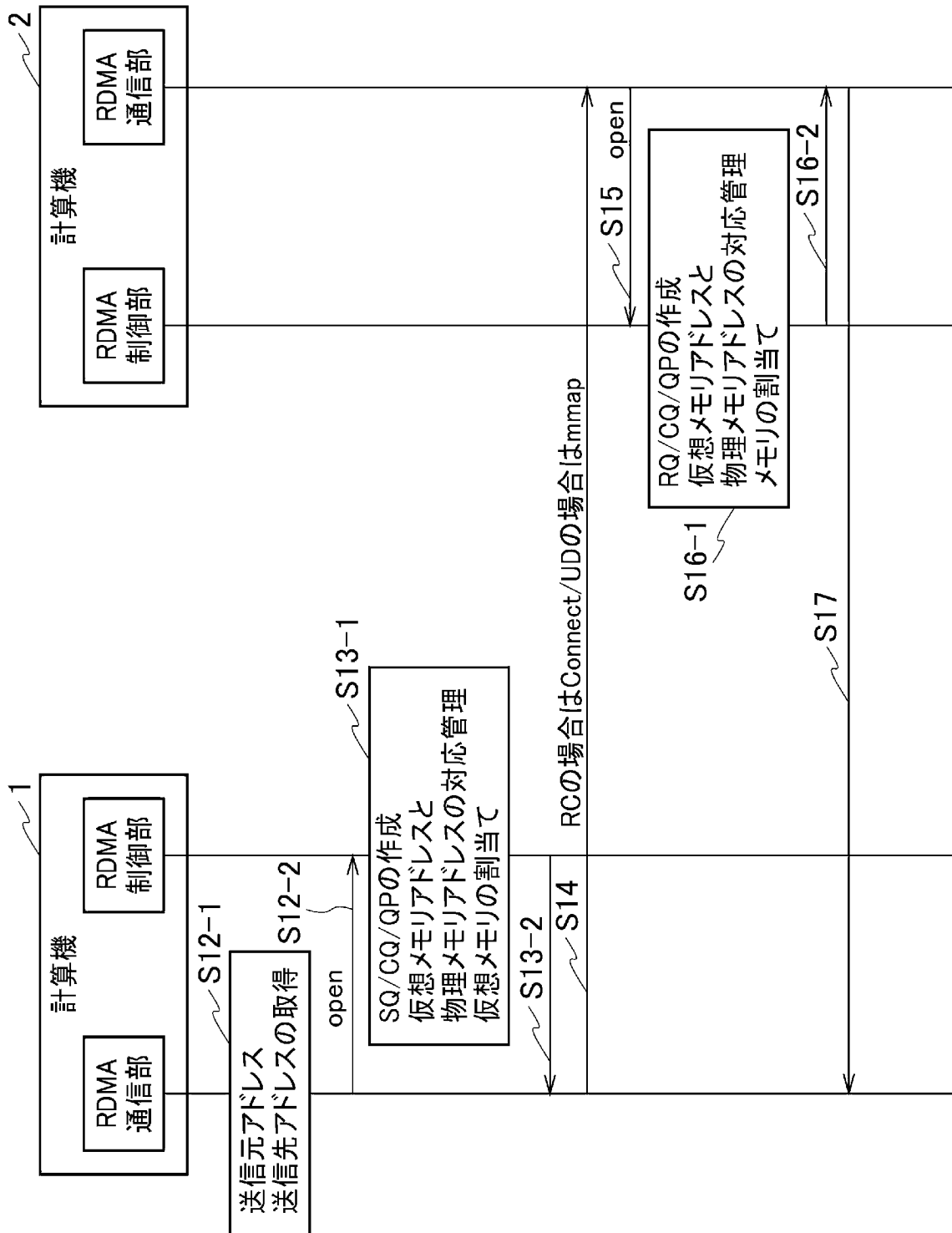
[図3]



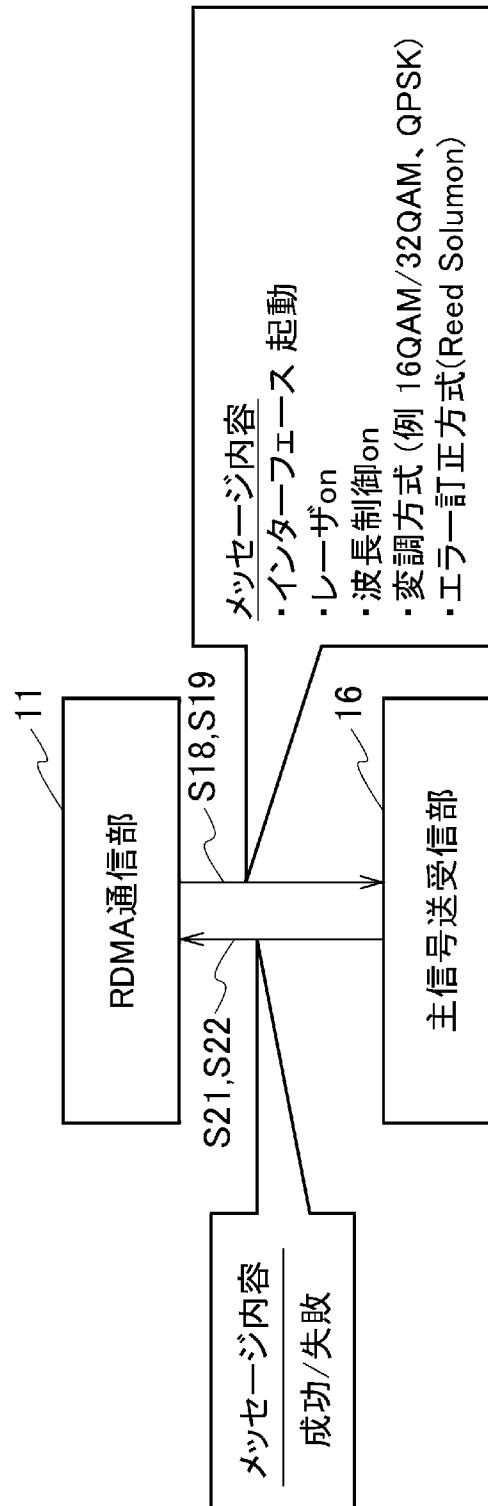
[図4]



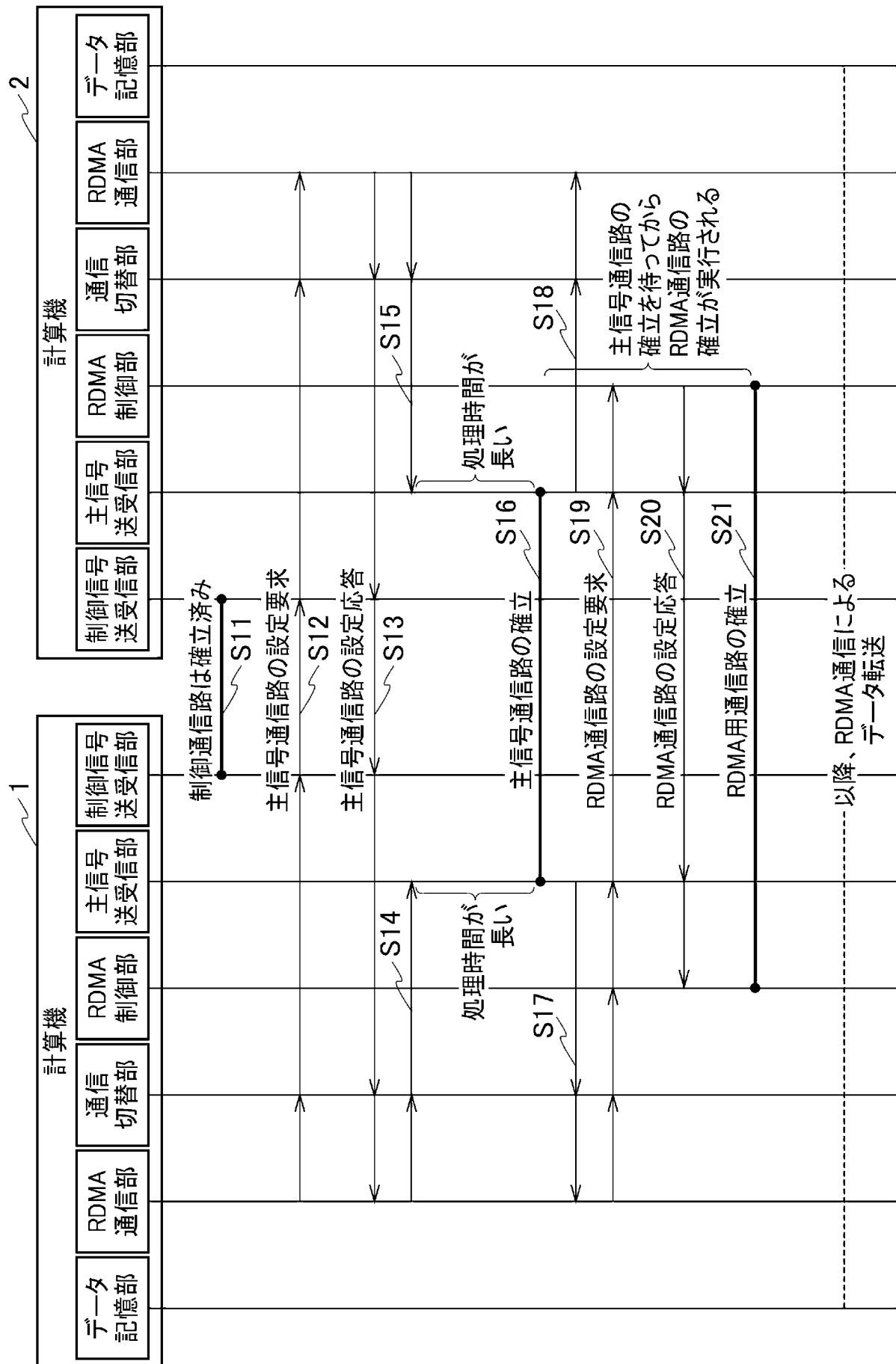
[図5]



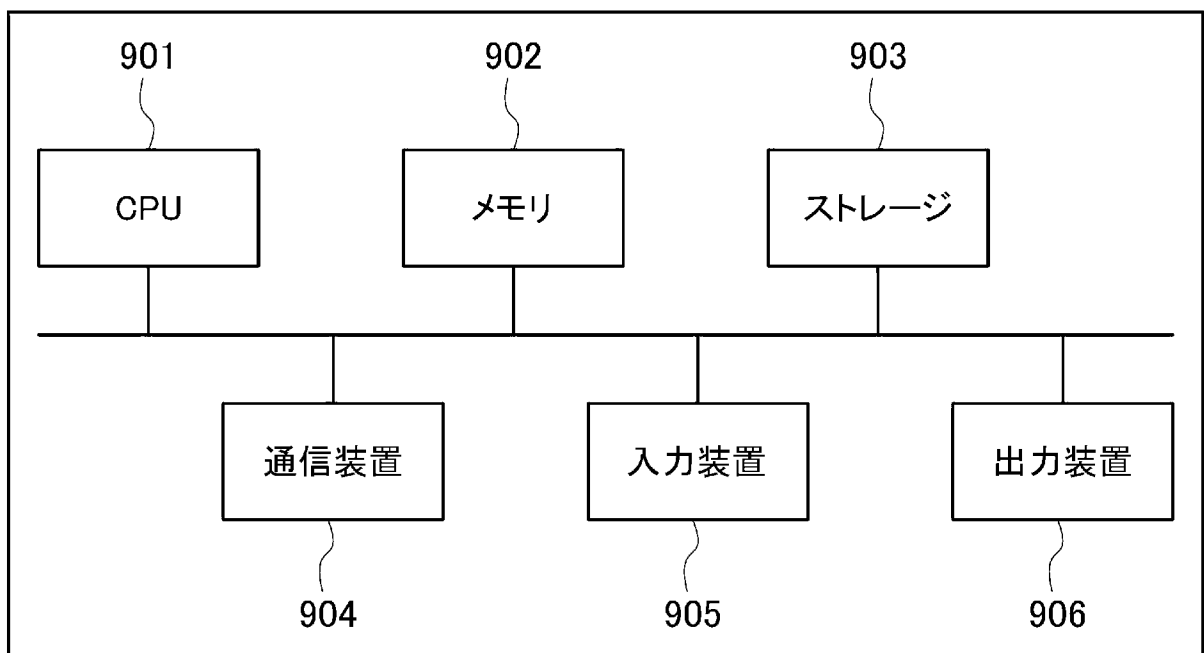
[図6]



[図7]



[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/032312

A. CLASSIFICATION OF SUBJECT MATTER**H04L 29/08**(2006.01)i

FI: H04L13/00 307A

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)

H04L29/08

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

Published examined utility model applications of Japan 1922-1996
 Published unexamined utility model applications of Japan 1971-2021
 Registered utility model specifications of Japan 1996-2021
 Published registered utility model applications of Japan 1994-2021

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	高橋 一志 TAKAHASHI, Kazushi. R D M Aを用いた仮想マシンイメージストレージに向けて, 情発処理学会 シンポジウム コンピュータシステム・シンポジウム2014 [online], 12 November 2014, pp. 55-62 chapters 3, 4.2, non-official translation (Toward virtual machine image storage using RDMA. Information Processing Society of Japan Computer System Symposium 2014 [online].)	1, 3-5
A		2
A	JP 2007-235674 A (NEC CORP) 13 September 2007 (2007-09-13) paragraph [0006]	1-5
A	JP 2012-65281 A (FUJITSU LTD) 29 March 2012 (2012-03-29) fig. 2	1-5
A	JP 2016-48500 A (FUJITSU LTD) 07 April 2016 (2016-04-07) paragraphs [0084]-[0092]	1-5



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

11 October 2021

Date of mailing of the international search report

26 October 2021

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)
 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915
 Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2021/032312

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
JP	2007-235674	A	13 September 2007	US	2007/0206598	A1	
				paragraph [0008]			
JP	2012-65281	A	29 March 2012	US	2012/0072607	A1	
				fig. 2			
				EP	2431885	A1	
JP	2016-48500	A	07 April 2016	US	2016/0062943	A1	
				paragraphs [0098]-[0107]			

A. 発明の属する分野の分類（国際特許分類（IPC）） H04L 29/08(2006.01)i FI: H04L13/00 307A		
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（IPC）） H04L29/08		
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 1922-1996年 日本国公開実用新案公報 1971-2021年 日本国実用新案登録公報 1996-2021年 日本国登録実用新案公報 1994-2021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	高橋 一志 KAZUSHI TAKAHASHI, RDMAを用いた仮想マシンイメージストレージに向けて、情報処理学会 シンポジウム コンピュータシステム・シンポジウム 2014 [online], 2014.11.12, pp. 55-62 第3, 第4.2章	1, 3-5
A		2
A	JP 2007-235674 A（日本電気株式会社）13.09.2007（2007-09-13） 段落[0006]	1-5
A	JP 2012-65281 A（富士通株式会社）29.03.2012（2012-03-29） 第2図	1-5
A	JP 2016-48500 A（富士通株式会社）07.04.2016（2016-04-07） 段落[0084]-[0092]	1-5
☐ 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抵触するものではなく、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 11.10.2021		国際調査報告の発送日 26.10.2021
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号		権限のある職員（特許庁審査官） 佐々木 洋 5X 3362 電話番号 03-3581-1101 内線 3596

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2021/032312

引用文献			公表日	パテントファミリー文献			公表日
JP	2007-235674	A	13.09.2007	US	2007/0206598	A1	
				段落[0008]			
JP	2012-65281	A	29.03.2012	US	2012/0072607	A1	
				第2図			
				EP	2431885	A1	
JP	2016-48500	A	07.04.2016	US	2016/0062943	A1	
				段落[0098]-[0107]			