

(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
2025年1月23日(23.01.2025)



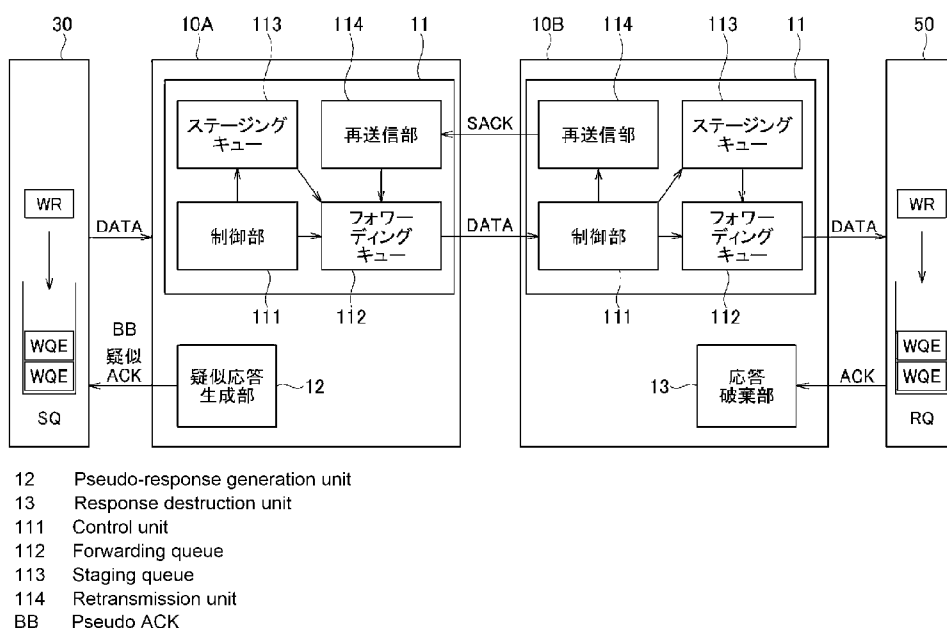
(10) 国際公開番号
WO 2025/017785 A1

- (51) 国際特許分類:
H04L 49/90 (2022.01) *H04L 1/1809* (2023.01)
G06F 15/173 (2006.01)
- (21) 国際出願番号: PCT/JP2023/026085
- (22) 国際出願日: 2023年7月14日(14.07.2023)
- (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: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 市川 潤 紀 (ICHIKAWA Junki); 〒1808585 東京都武蔵野市緑町3丁目9-

1 1 N T T 知的財産センタ内 Tokyo (JP). 西沢 秀樹(NISHIZAWA Hideki); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 尾花 和昭(OBANA Kazuaki); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 益谷 仁士(MASUTANI Hitoshi); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 小倉 毅(OGURA Tsuyoshi); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 釘本 健司(KUGIMOTO Takeshi); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 井上 綺泉(INOUE Kiwami); 〒1808585 東京都武

(54) Title: INTERMEDIATE DEVICE AND COMMUNICATION METHOD

(54) 発明の名称: 中間装置および通信方法



- 12 Pseudo-response generation unit
13 Response destruction unit
111 Control unit
112 Forwarding queue
113 Staging queue
114 Retransmission unit
BB Pseudo ACK

(57) Abstract: Intermediate nodes 10A, 10B are provided with a forwarding queue 112 for storing transferred packets, a staging queue 113 for storing unprocessed packets, and a retransmission unit 114 for performing retransmission processing of the packet from a transmission node 30 to a receiving node 50. When a packet loss is detected, the retransmission unit 114 requests retransmission of the lost packet. When receiving a retransmission packet after detecting the packet loss, the control unit 111 transfers the retransmission packet and stores it in the forwarding queue 112, and, after transferring the

蔵野市緑町 3 丁目 9 - 1 1 N T T 知的財産
センタ内 Tokyo (JP). 遊部 航希(YUBE Koki);
〒1808585 東京都武蔵野市緑町 3 丁目 9 - 1
1 N T T 知的財産センタ内 Tokyo (JP).

(74) 代理人: 三好 秀和, 外(MIYOSHI Hidekazu et al.); 〒1050001 東京都港区虎ノ門 1 丁目 2 番 8 号 虎ノ門琴平タワー 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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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:

— 国際調査報告 (条約第21条(3))

retransmission packet, transfers the packet stored in the staging queue 113 and stores it in the forwarding queue 112.

(57) 要約: 中間ノード 10A, 10B は、転送済のパケットを格納するフォワーディングキュー 112 と、未処理のパケットを格納するステージングキュー 113 と、送信ノード 30 から受信ノード 50 へ送パケットの再送処理を行う再送信部 114 を備える。再送信部 114 は、パケットロスが検出された場合に、ロスしたパケットの再送を要求する。制御部 111 は、パケットロスを検出後は、再送パケットを受信したときに、再送パケットを転送するとともにフォワーディングキュー 112 に格納し、再送パケットを転送後に、ステージングキュー 113 に格納されたパケットを転送するとともにフォワーディングキュー 112 に格納する。

明 細 書

発明の名称： 中間装置および通信方法

技術分野

[0001] 本開示は、中間装置および通信方法に関する。

背景技術

[0002] リモートダイレクトメモリアクセス（RDMA）は、遠隔の通信端末間で高速かつ高信頼なデータ転送を行う通信プロトコルである。RDMAは、送信側の装置のメモリ領域から受信側の装置のメモリ領域へ直接メモリアクセスするため、高速な通信が可能である。RDMAは、クレジットベースのフロー制御機能を持つほか、タイマーアウトやパケットロス検出に基づく再送制御を行うため、高信頼な通信が可能である。

先行技術文献

非特許文献

[0003] 非特許文献1： InfiniBand Architecture Specification Volume 1, Release 1.4, April 7, 2020

発明の概要

発明が解決しようとする課題

[0004] RDMAは、Send Queue（SQ）とReceive Queue（RQ）の組み合わせで構成されるQueue Pair（QP）を用いてデータ転送するモデルである。データの転送要求はWork Request（WR）と呼ばれる単位で行われ、これらの要求はSQやRQにWork Queue Entry（WQE）として格納される。QP間でのWQEの処理が完了すると、SQやRQのWQEが削除され、次のWRの処理が可能になる。

[0005] Connection型のサービスタイプでは、送信側のWQEを完了させるためには受信側からACKを受信する必要がある。送受信間に長距離ネットワークが存在する場合はリクエストとレスポンスのやり取りに時間がかかり、SQに未完了のWQEが滞留する。これにより、処理待ちのWRが増え、転送性能が低下するという問題があった。

[0006] 本開示は、上記に鑑みてなされたものであり、長距離ネットワークにおける転送性能の低下を抑制することを目的とする。

課題を解決するための手段

[0007] 本開示の一態様の中間装置は、リモートダイレクトメモリアクセスを用いてデータを転送する第1の装置と第2の装置の間に配置される中間装置であって、転送済の packets を格納する第1のキューと、未処理の packets を格納する第2のキューと、前記第1の装置から前記第2の装置へ送信される packets を受信して転送するとともに、当該 packets を前記第1のキューに格納する制御部と、 packets の再送処理を行う再送信部と、を備え、前記再送信部は、 packets ロスが検出された場合に、ロスした packets の再送を要求し、前記制御部は、 packets ロスを検出後は、再送 packets を受信するまでに受信した packets を前記第2のキューに格納し、再送 packets を受信したときに、当該再送 packets を転送するとともに前記第1のキューに格納し、前記再送 packets を転送後に、前記第2のキューに格納された packets を転送するとともに前記第1のキューに格納する。

発明の効果

[0008] 本開示によれば、長距離ネットワークにおける転送性能の低下を抑制することができる。

図面の簡単な説明

[0009] [図1] 図1は、通信システムの構成の一例を示す図である。

[図2] 図2は、転送処理の流れの一例を示すフローチャートである。

[図3] 図3は、転送処理の流れの一例を示すフローチャートである。

[図4] 図4は、 packets ロス時のキューの様子を説明するための図である。

[図5] 図5は、 packets ロス時のキューの様子を説明するための図である。

[図6] 図6は、 packets ロス時のキューの様子を説明するための図である。

[図7] 図7は、Go-Back-Nの再送処理の一例を説明するための図である。

[図8] 図8は、選択的再送処理の一例を説明するための図である。

[図9] 図9は、場所に応じて採用する再送処理を異ならせる一例を説明するた

めの図である。

[図10]図10は、中間ノードをNIC上に構成した一例を示す図である。

[図11]図11は、中間ノードのハードウェア構成の一例を示す図である。

発明を実施するための形態

[0010] 以下、本開示の実施の形態について図面を用いて説明する。

[0011] 図1を参照し、本実施形態の通信システムの構成の一例について説明する。同図に示す通信システムは、RDMAを用いてデータを転送する送信ノード30と受信ノード50、並びに、送信ノード30と受信ノード50の間に配置された中間ノード10A、10Bを備える。送信ノード30は、データをパケットに分割し、中間ノード10A、長距離ネットワーク、および中間ノード10Bを介して受信ノード50へパケットを送信する。送信ノード30と長距離ネットワークの間に中間ノード10Aが配置され、長距離ネットワークと受信ノード50の間に中間ノード10Bが配置される。中間ノード10A、10Bは、長距離ネットワークを介して接続される。送信ノード30と受信ノード50はそれぞれRequesterとResponderと呼ばれてもよい。

[0012] 送信ノード30から受信ノード50へデータを転送する際、送信ノード30は送信したいデータのメモリ領域をWQEに指定してSQに積み、受信ノード50はデータを受信したいメモリ領域をWQEに指定してRQに積む。SQやRQのキューサイズ分、WQEをSQやRQに積むことができる。QP間でのWRの処理が正常に完了すると、SQとRQからWQEが解放される。送信ノード30のWQEを完了させるためには、受信ノード50からACKを受信する必要がある。送信ノード30は、ACKを受信することで受信ノード50によるデータの正常な受信を確認できる。

[0013] 図1の通信システムでは、データ転送に際して、中間ノード10Aが送信ノード30へ疑似ACKを送信し、送信ノード30は疑似ACKに応じてSQのWQEを解放する。送信ノード30でのレスポンスの重複受信を防ぐために、中間ノード10Bは受信ノード50からのレスポンスを破棄する。

[0014] 中間ノード10Bがレスポンスを破棄すると、受信ノード50がパケット

の再送を要求しても、送信ノード30はパケットの再送要求を検知することができない。中間ノード10Bは、受信ノード50からACKを受信するまでパケットをキューに保持しておき、送信ノード30の代わりに受信ノード50のパケット再送要求に対応する。

[0015] また、疑似ACKにより送信ノード30からのデータ送信量が増大するので、RDMA仕様の再送方式であるGo-Back-Nを用いると、再送されるパケットが通信帯域を圧迫し、性能が低下する原因となる。Go-Back-Nは、パケットロスしたパケット以降の全てのパケットを再送する方式である。中間ノード10Bは、パケットロス以降の受信済みのパケットを破棄せずにキューに保持し、ロスしたパケットのみの再送を要求することで、再送パケット量を抑制する。

[0016] 以下、中間ノード10A、10Bの構成の一例について説明する。

[0017] 中間ノード10Aは、転送部11と疑似応答部12を備える。中間ノード10Bは、転送部11と応答破棄部13を備える。

[0018] 転送部11は、送信ノード30と受信ノード50との間で送受信されるパケットを転送する。中間ノード10Aの転送部11は、送信ノード30からパケットを受信して中間ノード10Bへ転送する。中間ノード10Bの転送部11は、中間ノード10Aからパケットを受信して受信ノード50へ転送する。中間ノード10A、10Bの転送部11はほぼ同じ機能を有するので、以下では中間ノード10A、10Bの転送部11の機能についてまとめて説明する。

[0019] 転送部11は、制御部111、フォワーディングキュー112、ステージングキュー113、および再送信部114を備える。

[0020] フォワーディングキュー112は転送済のパケットを一時的に保持するキューである。転送先のノードからACKを受信すると、フォワーディングキュー112のパケットは解放される。転送先のノードからパケット再送の要求を受信すると、フォワーディングキュー112の保持するパケットが再送される。転送先のノードとは、中間ノード10Aの場合は中間ノード10Bであり、中間ノード10Bの場合は受信ノード50である。

- [0021] ステージングキュー 1 1 3 は受信したが転送していない未処理のパケットを一時的に保持するキューである。パケットロスが検出された場合、ステージングキュー 1 1 3 は、ロスしたパケットが再送されるまで、パケットロスを検出した以降に受信したパケットを保持する。
- [0022] 制御部 1 1 1 は、受信したパケットに付与されたシーケンス番号に基づいてパケットを評価し、パケットの転送処理またはパケットの再送要求を行う。データは、連続したPacket Sequence Number (PSN) が付与されたパケットに分割され、PSNの順に送信される。制御部 1 1 1 は、受信したパケットのPSNが予想されるPSN（以下、ePSNと称する）と同じ場合は、受信したパケットをフォワーディングキュー 1 1 2 へ格納した後、次のノードへ転送する。受信したパケットのPSNがePSNと異なる場合は、受信したパケットをステージングキュー 1 1 3 に格納する。ステージングキュー 1 1 3 がパケットを保持する場合、制御部 1 1 1 は、ステージングキュー 1 1 3 に最後に格納されたパケットのPSN+1を未処理のPSN（以下、oPSNと称する）として管理する。oPSNは、再送パケットを受信するまでの間に到着するかもしれない通常パケットのPSNを示す。受信したパケットのPSNがePSNでもoPSNでもない場合、制御部 1 1 1 は、パケットロスが発生したと判断する。
- [0023] 再送パケットを受信した場合、制御部 1 1 1 は、再送パケットをフォワーディングキュー 1 1 2 へ格納した後、次のノードへ転送し、ステージングキュー 1 1 3 の保持するパケットをフォワーディングキュー 1 1 2 へ移動した後、次のノードへ転送する。
- [0024] 再送信部 1 1 4 は、ロスしたパケットの再送要求および再送要求に対するパケットの再送を行う。中間ノード 1 0 A の再送信部 1 1 4 は中間ノード 1 0 B からの再送要求に応じてパケットを再送する。中間ノード 1 0 B の再送信部 1 1 4 は中間ノード 1 0 A にパケットの再送を要求する。また、中間ノード 1 0 A の再送信部 1 1 4 は送信ノード 3 0 にパケットの再送を要求し、中間ノード 1 0 B の再送信部 1 1 4 は受信ノード 5 0 からの再送要求に応じてパケットを再送する。

- [0025] 再送信部 114 は、ロスしたパケットのみを再送要求する。具体的には、中間ノード 10B の再送信部 114 は、再送を要求するパケットの範囲を指定した Selective ACK (SACK) を生成して中間ノード 10A に送信する。SACK には Negative ACK (NAK) を拡張して用いる。SACK であることが分かるように AETH フィールドを拡張する。要求するパケットを指定する PSN フィールドには ePSN を入れ、MSN フィールドにはパケットロスを検出したパケットの PSN を入れる。
- [0026] 再送信部 114 は、SACK を受信すると、指定のパケットをキューから取り出して再送信する。具体的には、中間ノード 10A の再送信部 114 は、中間ノード 10B から SACK を受信すると、PSN から MSN-1 までのパケットをフォーワーディングキュー 112 から取得して中間ノード 10B に再送信する。
- [0027] 中間ノード 10A, 10B は、受信側と送信側のいずれにもなり得るので、中間ノード 10A, 10B の再送信部 114 は、パケット再送を要求する機能、パケットを再送する機能のいずれも備える。送信側と受信側で再送方式が異なる場合、再送信部 114 は再送方式に合った方法で再送処理を行う。例えば、送信ノード 30 と中間ノード 10A との間の再送方式が Go-Back-N の場合、再送信部 114 は Go-Back-N に従った方法で再送要求を送信ノード 30 へ送信する。受信ノード 50 と中間ノード 10B との間の再送方式が Go-Back-N の場合、再送信部 114 は Go-Back-N に従った方法で再送パケットを受信ノード 50 へ送信する。中間ノード 10A, 10B 間の再送方式が選択的再送の場合、再送信部 114 は上記の SACK を用いた再送方式で再送処理を行う。
- [0028] 疑似応答部 12 は、送信ノード 30 から受信ノード 50 へのリクエストを受信すると、疑似 ACK を生成して送信ノード 30 へ返却する。パケットロスを検出した場合は、中間ノード 10A は、送信ノード 30 へ再送を要求する。
- [0029] 応答破棄部 13 は、受信ノード 50 から送信ノード 30 へのレスポンスを受信すると、受信したレスポンスを破棄する。再送要求を受信した場合は、中間ノード 10B は、受信ノード 50 への再送処理を実行する。

- [0030] なお、疑似応答部 1 2 および応答破棄部 1 3 は転送部 1 1 による選択的再送の処理に必須の構成ではないが、中間ノード 1 0 A, 1 0 B が疑似応答部 1 2 と応答破棄部 1 3 を備えることにより発生する課題は、中間ノード 1 0 A, 1 0 B が上記の転送部 1 1 を備えることで解決できる。
- [0031] 図 2, 3 のフローチャートを参照し、パケットの転送処理の流れの一例について説明する。中間ノード 1 0 がパケットを受信すると、図 2, 3 の処理が実行される。ここでは中間ノード 1 0 A, 1 0 B を区別せずに中間ノード 1 0 とする。
- [0032] ステップ S 1 にて、中間ノード 1 0 は、再送中であるか否かを判定する。再送中とは、ロスしたパケットの再送を待っている状態のことである。中間ノード 1 0 は、ステージングキュー 1 1 3 にパケットを保持しているか否かに基づいて、再送中であるか否かを判定できる。再送中の場合、図 3 のフローチャートに処理を進める。
- [0033] 再送中でない場合、ステップ S 2 にて、中間ノード 1 0 は、受信したパケットの PSN が ePSN であるか否かを判定する。
- [0034] パケットの PSN が ePSN である場合、ステップ S 3 にて、中間ノード 1 0 は、パケットをフォワーディングキュー 1 1 2 に格納するとともに、パケットを次のノードへ転送する。中間ノード 1 0 は、ePSN を 1 増加する。中間ノード 1 0 は次のノードから ACK を受信するとフォワーディングキュー 1 1 2 に格納したパケットを解放する。
- [0035] パケットの PSN が ePSN と異なる場合、中間ノード 1 0 は、ステップ S 4 にて、パケットをステージングキュー 1 1 3 に格納し、ステップ S 5 にて、SACK を転送元のノードへ送信する。中間ノード 1 0 は、ステージングキュー 1 1 3 に格納したパケットの PSN+1 を oPSN に設定する。
- [0036] 再送中の場合、図 3 のステップ S 6 にて、中間ノード 1 0 は、再送されたパケットを受信したか否かを判定する。再送されたパケットを受信したか否かは、受信したパケットの PSN が ePSN であるか否かで判定できる。
- [0037] 再送パケットを受信した場合、ステップ S 7 にて、中間ノード 1 0 は、パ

ケットをフォワーディングキュー 1 1 2 に格納するとともに、パケットを次のノードへ転送する。

[0038] ステップ S 8 にて、中間ノード 1 0 は、ステージングキュー 1 1 3 に格納されたパケットをフォワーディングキュー 1 1 2 へ移動するとともに、移動したパケットを次のノードへ転送する。中間ノード 1 0 は、転送した最後のパケットの PSN+1 を ePSN に設定する。

[0039] なお、ステージングキュー 1 1 3 からフォワーディングキュー 1 1 2 へ移動するパケットは、受信したパケットの PSN+1 から抜けのない連続した PSN のパケットである。例えば、ステージングキュー 1 1 3 に PSN が 3, 4, 5, 7 のパケットが格納されている場合に、PSN が 2 のパケットを受信すると、PSN が 3, 4, 5 のパケットがステージングキュー 1 1 3 からフォワーディングキュー 1 1 2 へ移動される。PSN が 7 のパケットはステージングキュー 1 1 3 に残されたままである。

[0040] 中間ノード 1 0 は、ステップ S 7, S 8 で転送したパケットについて ACK を受信すると、フォワーディングキュー 1 1 2 に格納したパケットを解放する。

[0041] 受信したパケットが再送パケットでない場合、ステップ S 9 にて、パケットをステージングキュー 1 1 3 に格納する。中間ノード 1 0 は、ステージングキュー 1 1 3 に格納したパケットの PSN+1 を oPSN に設定する。

[0042] 中間ノード 1 0 は以上の処理を繰り返してパケットを次のノードに転送する。

[0043] 次に、図 4, 5, 6 を参照し、パケット転送時のフォワーディングキュー 1 1 2 とステージングキュー 1 1 3 の状態の一例について説明する。

[0044] まず、図 4 を参照し、パケット転送時にパケットロスした時の基本的な動作について説明する。

[0045] ステップ S 1 1 にて、中間ノード 1 0 A がリクエスト r 1 (以下、r 1 と称する) を送信し、中間ノード 1 0 B が r 1 を受信する。r 1 の PSN は 1 である。以下では r に続く数字で PSN を表すものとする。なお、中間ノード 1 0 B

のePSNは1である。中間ノード10Bのステージングキュー113は空であるからoPSNは-1とする。中間ノード10Bは、r1をフォワーディングキュー112に格納するとともに、r1を次のノードに転送する。中間ノード10BはePSNを2とする。

[0046] ステップS12にて、中間ノード10Aはr2を送信するが、r2はロスした。

[0047] ステップS13にて、中間ノード10Aがr3を送信し、中間ノード10Bがr3を受信する。r3のPSNとePSNが異なるので、中間ノード10Bはr3をステージングキュー113に格納する。中間ノード10Bは、oPSNを4とする。

[0048] ステップS14にて、中間ノード10BはPSNが2の packets の再送要求n2を中間ノード10Aへ送信する。

[0049] ステップS15にて、中間ノード10Aがr4を送信し、中間ノード10Bがr4を受信する。中間ノード10Bはr4をステージングキュー113に格納する。中間ノード10BはoPSNを5とする。

[0050] ステップS16にて、中間ノード10Aがr5を送信し、中間ノード10Bがr5を受信する。中間ノード10Bはr5をステージングキュー113に格納する。中間ノード10BはoPSNを6とする。

[0051] ステップS17にて、中間ノード10Aは再送要求n2に応じてr2を送信し、中間ノード10Bがr2を受信する。中間ノード10Bは、r2をフォワーディングキュー112に格納するとともに、r2を次のノードに転送する。中間ノード10Bは、ステージングキュー113に格納されたr3, r4, r5をフォワーディングキュー112に格納するとともに、r3, r4, r5を次のノードに転送する。中間ノード10BはePSNを6、oPSNを-1とする。

[0052] 以上の処理により、中間ノード10Aが送信したr1からr5が中間ノード10Bの次のノードに転送される。

[0053] 続いて、図5を参照し、連続してパケットロスが発生した時の動作につい

て説明する。

- [0054] ステップS 2 1 にて、中間ノード1 0 Aが r 1 を送信し、中間ノード1 0 Bが r 1 を受信する。中間ノード1 0 Bは、 r 1 をフォワーディングキュー1 1 2 に格納するとともに、 r 1 を次のノードに転送する。中間ノード1 0 BはePSNを2とする。
- [0055] ステップS 2 2, S 2 3, S 2 4 にて、中間ノード1 0 Aは r 2, r 3, r 4 を送信するが、 r 2, r 3, r 4 はロスした。
- [0056] ステップS 2 5 にて、中間ノード1 0 Aが r 5 を送信し、中間ノード1 0 Bが r 5 を受信する。 r 5 のPSNとePSNが異なるので、中間ノード1 0 Bは r 5 をステージングキュー1 1 3 に格納する。中間ノード1 0 BはoPSNを6とする。
- [0057] ステップS 2 6 にて、中間ノード1 0 Bは、PSNが2から4までのパケットの再送要求 n 2 を中間ノード1 0 Aへ送信する。具体的には、中間ノード1 0 Bは、PSNフィールドにePSNの2を入れ、MSNフィールドに r 5 のPSNの5を入れたSACKを生成して送信する。
- [0058] ステップS 2 7 にて、中間ノード1 0 Aは再送要求 n 2 に応じて r 2, r 3, r 4 を送信し、中間ノード1 0 Bが r 2, r 3, r 4 を受信する。中間ノード1 0 Bは、 r 2, r 3, r 4 をフォワーディングキュー1 1 2 に格納するとともに、 r 2, r 3, r 4 を次のノードに転送する。中間ノード1 0 Bは、ステージングキュー1 1 3 に格納された r 5 をフォワーディングキュー1 1 2 に格納するとともに、 r 5 を次のノードに転送する。中間ノード1 0 BはePSNを6、oPSNを-1とする。
- [0059] 以上の処理により、中間ノード1 0 Aが送信した r 1 から r 5 が中間ノード1 0 Bの次のノードに転送される。
- [0060] 続いて、図6を参照し、再送中に異なるパケットロスが発生した時の動作について説明する。
- [0061] ステップS 3 1 にて、中間ノード1 0 Aが r 1 を送信し、中間ノード1 0 Bが r 1 を受信する。中間ノード1 0 Bは、 r 1 をフォワーディングキュー

1 1 2 に格納するとともに、 r_1 を次のノードに転送する。中間ノード 1 0 B は ePSN を 2 とする。

[0062] ステップ S 3 2 にて、中間ノード 1 0 A は r_2 を送信するが、 r_2 はロスした。

[0063] ステップ S 3 3 にて、中間ノード 1 0 A が r_3 を送信し、中間ノード 1 0 B が r_3 を受信する。 r_3 の PSN と ePSN が異なるので、中間ノード 1 0 B は r_3 をステージングキュー 1 1 3 に格納する。中間ノード 1 0 B は oPSN を 4 とする。

[0064] ステップ S 3 4 にて、中間ノード 1 0 B は、PSN が 2 のパケットの再送要求 n_2 を中間ノード 1 0 A へ送信する。

[0065] ステップ S 3 5 にて、中間ノード 1 0 A は r_4 を送信するが、 r_4 はロスした。

[0066] ステップ S 3 6 にて、中間ノード 1 0 A が r_5 を送信し、中間ノード 1 0 B が r_5 を受信する。中間ノード 1 0 B は、 r_5 の PSN と ePSN が異なるので、 r_5 をステージングキュー 1 1 3 に格納する。中間ノード 1 0 B は、oPSN を 6 とする。

[0067] ステップ S 3 7 にて、中間ノード 1 0 B は、PSN が 4 のパケットの再送要求 n_4 を中間ノード 1 0 A へ送信する。

[0068] ステップ S 3 8 にて、中間ノード 1 0 A が r_6 を送信し、中間ノード 1 0 B が r_6 を受信する。 r_6 の PSN と ePSN が異なるので、中間ノード 1 0 B は r_6 をステージングキュー 1 1 3 に格納する。中間ノード 1 0 B は oPSN を 7 とする。

[0069] ステップ S 3 9 にて、中間ノード 1 0 A は再送要求 n_2 に応じて r_2 を送信し、中間ノード 1 0 B が r_2 を受信する。中間ノード 1 0 B は、 r_2 をフォワーディングキュー 1 1 2 に格納するとともに、 r_2 を次のノードに転送する。中間ノード 1 0 B は、ステージングキュー 1 1 3 に格納された r_3 をフォワーディングキュー 1 1 2 に格納するとともに、 r_3 を次のノードに転送する。ステージングキュー 1 1 3 に格納された r_5 、 r_6 は r_3 に連続し

ないPSNであるから対象外である。中間ノード10BはePSNを4とする。

[0070] ステップS40にて、中間ノード10Aがr7を送信し、中間ノード10Bがr7を受信する。r7のPSNとePSNが異なるので、中間ノード10Bはr7をステージングキュー113に格納する。中間ノード10BはoPSNを8とする。

[0071] ステップS41にて、中間ノード10Aは再送要求n4に応じてr4を送信し、中間ノード10Bがr4を受信する。中間ノード10Bは、r4をフォワーディングキュー112に格納するとともに、r4を次のノードに転送する。中間ノード10Bは、ステージングキュー113に格納されたr5, r6, r7をフォワーディングキュー112に格納するとともに、r5, r6, r7を次のノードに転送する。中間ノード10BはePSNを8、oPSNを-1とする。

[0072] 以上の処理により、中間ノード10Aが送信したr1からr7が中間ノード10Bの次のノードに転送される。

[0073] 次に、図7, 8を参照し、Go-Back-Nと選択的再送処理の違いについて説明する。

[0074] 図7を参照し、Go-Back-Nの再送処理の一例について説明する。

[0075] 送信ノード30から送信されたパケットは、中間ノード10Aから中間ノード10Bへ転送される。中間ノード10A, 10Bは再送用のキューを備えており、転送するパケットを一時的に保持する。再送用のキューは前述のフォワーディングキュー112である。

[0076] ステップS51にて、中間ノード10Bは、パケットロスを検出する。中間ノード10Bは、パケットロス検出後に受信したパケットは破棄する。

[0077] ステップS52にて、中間ノード10Bは、中間ノード10Aに対して、ロスしたPSNを指定してNegative ACK (NAK) を送信する。

[0078] ステップS53にて、中間ノード10Aは、指定されたPSN以降の全てのパケットを再送し、中間ノード10Bは、受信したパケットを受信ノード50へ転送する。

- [0079] Go-Back-Nでは、ロスしたパケット以降のパケットを全て再送するので、通常のパケット転送と再送が競合し、転送効率が低下する。特に、中間ノード10Aが疑似ACKを返信する場合、通常のパケット転送のRound Trip Timeが短くなり、通常のパケット転送と再送がより競合しやすくなる。
- [0080] 図8を参照し、選択的再送処理の一例について説明する。
- [0081] 送信ノード30から送信されたパケットは、中間ノード10Aから中間ノード10Bへ転送される。中間ノード10A、10Bは再送用のキューを備えており、転送するパケットを一時的に保持する。再送用のキューは前述のフォワーディングキュー112である。
- [0082] ステップS61にて、中間ノード10Bは、パケットロスを検出する。中間ノード10Bは、パケットロス検出後に受信したパケットはキューに格納しておく。
- [0083] ステップS62にて、中間ノード10Bは、中間ノード10Aに対して、ロスしたPSNを指定してSACKを送信する。
- [0084] ステップS63にて、中間ノード10Aは、指定されたPSNのパケットのみを再送し、中間ノード10Bは、受信したパケットを受信ノード50へ転送する。
- [0085] ステップS64にて、中間ノード10Bは、再送パケットに続くパケットをキューから受信ノード50へ転送する。
- [0086] 選択的再送では、ロスしたパケットのみを再送するので、再送パケットの量が抑制されて転送効率が改善し、再送時間を短縮できる。
- [0087] 次に、図9を参照し、パケットロスした箇所に応じて採用する再送処理を異ならせる例について説明する。
- [0088] 透過的動作を考慮し、例えば、送信ノード30の近傍および受信ノード50の近傍ではGo-Back-Nを採用し、中間ノード10A、10Bは再送コストが大きいため選択的再送を採用する。
- [0089] 送信ノード30と中間ノード10Aの間では、Go-Back-Nによる再送処理が行われる。

- [0090] ステップS 7 1 にて、中間ノード 1 0 A はOut-of-Sequenceを検知すると、NAKを送信ノード 3 0 へ送信する。送信ノード 3 0 は指定された番号以降の全てのパケットを再送する。中間ノード 1 0 A は、漏れなくパケットを受信できると、受信したパケットをキューに格納するとともに、中間ノード 1 0 B へ転送する。
- [0091] 中間ノード 1 0 A, 1 0 B 間では、選択的再送が行われる。
- [0092] ステップS 7 2 にて、中間ノード 1 0 B はパケットロスを検知する。パケットロス以降のパケットはキューに格納する。
- [0093] ステップS 7 3 にて、中間ノード 1 0 B はロスしたパケットの再送を要求するSACKを中間ノード 1 0 A へ送信する。
- [0094] ステップS 7 4 にて、中間ノード 1 0 A は指定された番号のパケットをキューから取出して送信し、中間ノード 1 0 B は再送パケットをキューに格納するとともに、受信ノード 5 0 へ転送する。
- [0095] ステップS 7 5 にて、中間ノード 1 0 B は再送パケットに続くパケットをキューから取出して受信ノード 5 0 へ転送する。
- [0096] 中間ノード 1 0 B と受信ノード 5 0 の間では、Go-Back-Nによる再送処理が行われる。
- [0097] ステップS 7 6 にて、受信ノード 5 0 はOut-of-Sequenceを検知すると、NAKを中間ノード 1 0 B へ送信する。中間ノード 1 0 B は指定された番号以降の全てのパケットをキューから取出して再送する。
- [0098] 以上説明したように、中間ノード 1 0 A, 1 0 B は、リモートダイレクトメモリアクセスを用いてデータを転送する第 1 の装置と第 2 の装置の間に配置される。中間ノード 1 0 A, 1 0 B は、転送済のパケットを格納するフォワーディングキュー 1 1 2 と、未処理のパケットを格納するステージングキュー 1 1 3 と、送信ノード 3 0 から受信ノード 5 0 へ送信されるパケットを受信して転送するとともに、当該パケットをフォワーディングキュー 1 1 2 に格納する制御部 1 1 1 と、パケットの再送処理を行う再送信部 1 1 4 を備える。再送信部 1 1 4 は、パケットロスが検出された場合に、ロスしたパケ

ットの再送を要求する。制御部１１１は、パケットロスを検出後は、再送パケットを受信するまでに受信したパケットをステージングキュー１１３に格納し、再送パケットを受信したときに、再送パケットを転送するとともにフォワーディングキュー１１２に格納し、再送パケットを転送後に、ステージングキュー１１３に格納されたパケットを転送するとともにフォワーディングキュー１１２に格納する。これにより、ロスしたパケットのみが再送されるので、再送パケット量が削減され、転送性能の低下を抑制できる。

- [0099] 再送信部１１４は、パケットの再送要求に対して、要求されたパケットをフォワーディングキュー１１２から取得して再送する。これにより、再送要求を受信したときに、中間ノード１０Ａ、１０Ｂでパケットの再送に対処可能となる。中間ノード１０Ａで疑似ACKを生成して送信し、中間ノード１０Ｂでレスポンスを破棄する場合も、受信ノード５０からの再送要求に対応することが可能となる。
- [0100] 以上では、送信ノード３０と受信ノード５０の間に中間ノード１０Ａ、１０Ｂを設置する構成で説明したが、図１０に示すように、送信ノード３０のNetwork Interface Card (NIC) 上に中間ノード１０Ａを構成し、受信ノード５０のNIC上に中間ノード１０Ｂを構成してもよい。
- [0101] また、中間ノード１０Ａ、１０Ｂは、物理サーバで構成してもよいし、仮想サーバで構成してもよい。スイッチまたはルータなどのネットワーク装置が中間ノード１０Ａ、１０Ｂの機能を備えてもよい。
- [0102] 上記説明した中間ノード１０Ａ、１０Ｂには、例えば、図１１に示すような、中央演算処理装置（CPU）９０１と、メモリ９０２と、ストレージ９０３と、通信装置９０４と、入力装置９０５と、出力装置９０６とを備える汎用的なコンピュータシステムを用いることができる。このコンピュータシステムにおいて、CPU ９０１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、中間ノード１０Ａ、１０Ｂが実現される。このプログラムは磁気ディスク、光ディスク、半導体メモリなどのコンピュータ読み取り可能な記録媒体に記録することも、ネットワークを介して配信す

ることもできる。

符号の説明

- [0103] 1 0 A, 1 0 B 中間ノード
- 1 1 転送部
- 1 1 1 制御部
- 1 1 2 フォワーディングキュー
- 1 1 3 ステージングキュー
- 1 1 4 再送信部
- 1 2 疑似応答部
- 1 3 応答破棄部
- 3 0 送信ノード
- 5 0 受信ノード

請求の範囲

- [請求項1] リモートダイレクトメモリアクセスを用いてデータを転送する第1の装置と第2の装置の間に配置される中間装置であって、
- 転送済の packets を格納する第1のキューと、
- 未処理の packets を格納する第2のキューと、
- 前記第1の装置から前記第2の装置へ送信される packets を受信して転送するとともに、当該 packets を前記第1のキューに格納する制御部と、
- packets の再送処理を行う再送信部と、を備え、
- 前記再送信部は、packet ロスが検出された場合に、ロスした packets の再送を要求し、
- 前記制御部は、packet ロスを検出後は、再送 packets を受信するまでに受信した packets を前記第2のキューに格納し、再送 packets を受信したときに、当該再送 packets を転送するとともに前記第1のキューに格納し、前記再送 packets を転送後に、前記第2のキューに格納された packets を転送するとともに前記第1のキューに格納する中間装置。
- [請求項2] 請求項1に記載の中間装置であって、
- 前記再送信部は、packets の再送要求に対して、要求された packets を前記第1のキューから取得して再送する中間装置。
- [請求項3] 請求項1に記載の中間装置であって、
- 前記制御部は、次に到着が予想される packets の第1のシーケンス番号と未処理の packets の最大の第2のシーケンス番号を管理し、受信した packets のシーケンス番号を前記第1のシーケンス番号および前記第2のシーケンス番号と比較し、packet ロスを検出する中間装置。
- [請求項4] 請求項1に記載の中間装置であって、

前記第 1 の装置から受信したパケットに対する疑似レスポンスを生成して前記第 1 の装置へ返却する疑似応答生成部を備える
中間装置。

[請求項5] 請求項 1 に記載の中間装置であって、
前記第 2 の装置から受信したレスポンスを破棄する応答破棄部を備える
中間装置。

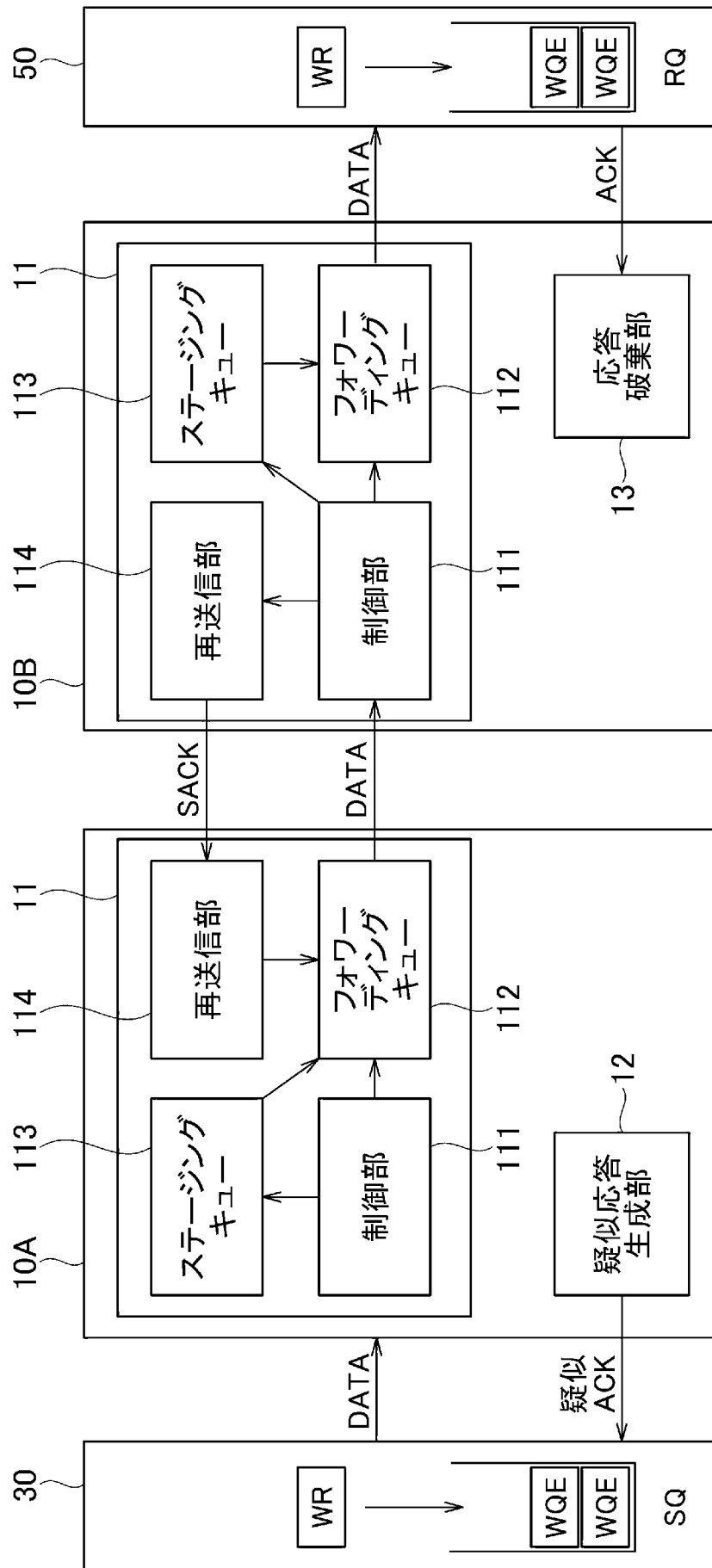
[請求項6] リモートダイレクトメモリアクセスを用いてデータを転送する第 1 の装置と第 2 の装置の間に配置される中間装置が実行する通信方法であって、

前記中間装置は、
転送済のパケットを格納する第 1 のキューと、
未処理のパケットを格納する第 2 のキューと、を備え、
前記中間装置は、
前記第 1 の装置から前記第 2 の装置へ送信されるパケットを受信して転送するとともに、当該パケットを前記第 1 のキューに格納し、
パケットロスを検出した場合に、ロスしたパケットの再送を要求し、
パケットロスを検出後は、再送パケットを受信するまでに受信したパケットを前記第 2 のキューに格納し、
再送パケットを受信したときに、当該再送パケットを転送するとともに前記第 1 のキューに格納し、
前記再送パケットを転送後に、前記第 2 のキューに格納されたパケットを転送するとともに前記第 1 のキューに格納する
通信方法。

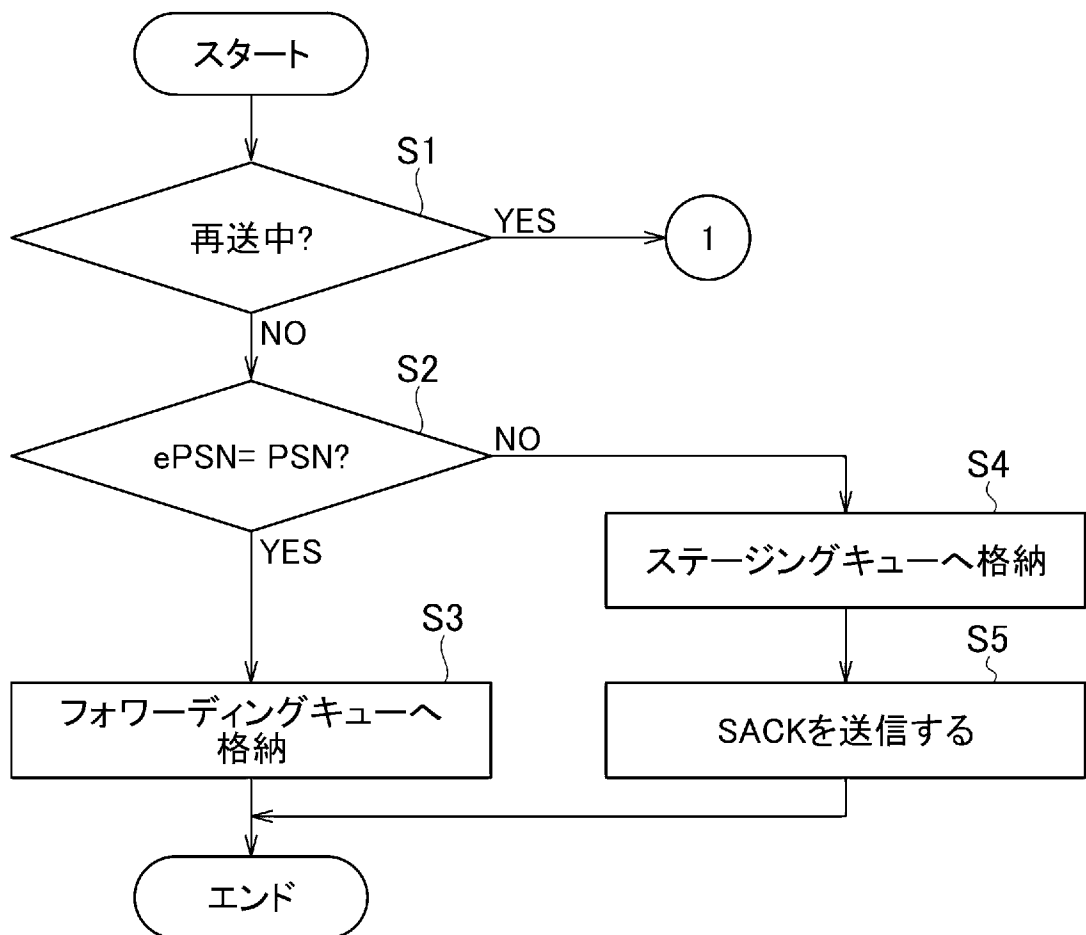
[請求項7] 請求項 6 に記載の通信方法であって、
前記中間装置は、パケットの再送要求に対して、要求されたパケットを前記第 1 のキューから取得して再送する

通信方法。

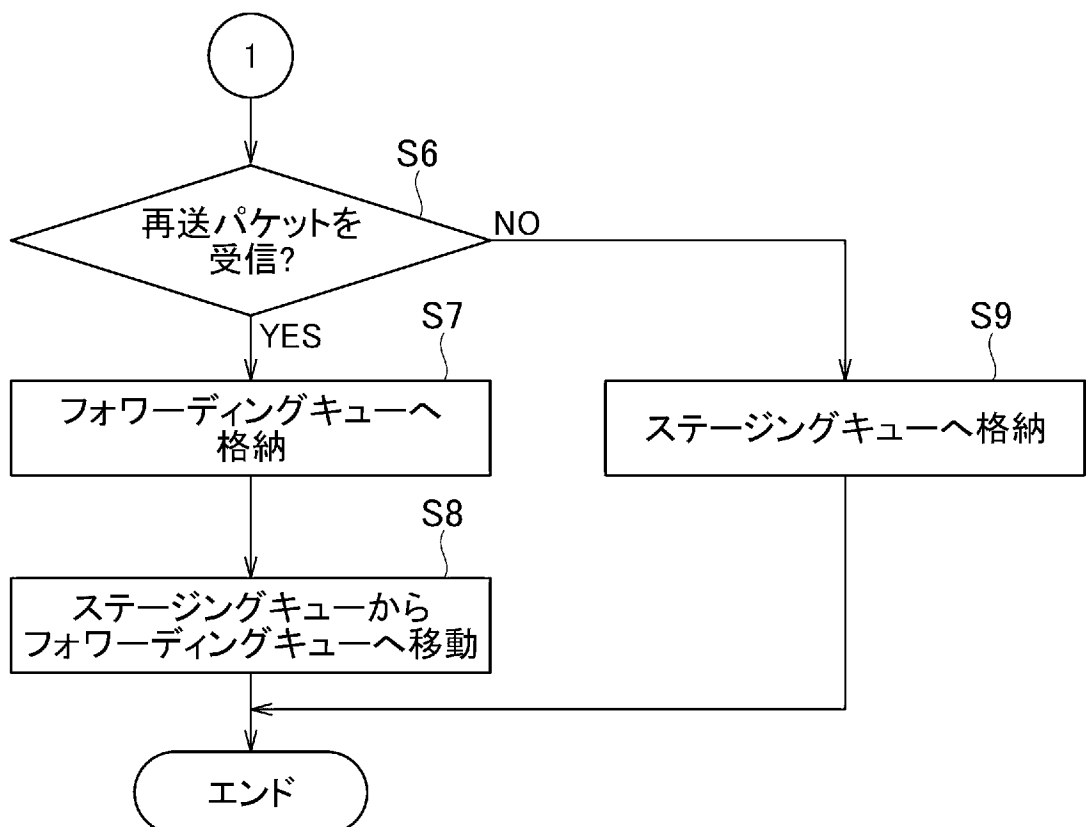
[図1]



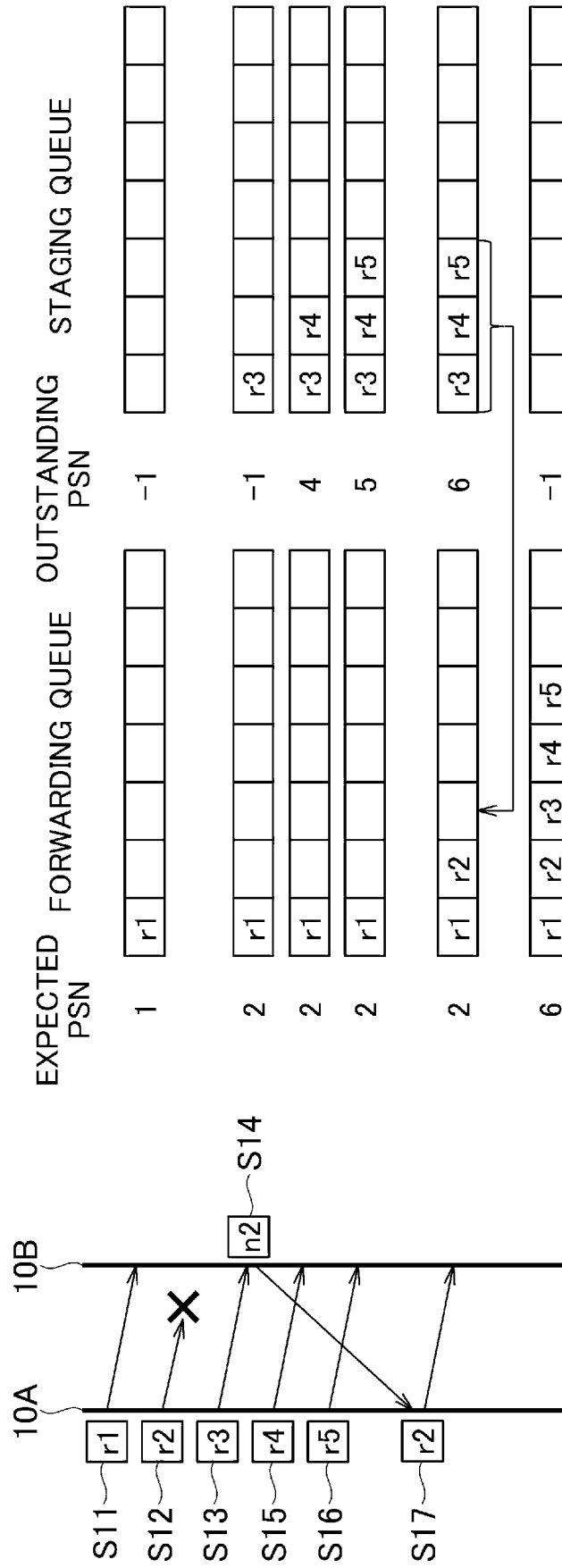
[図2]



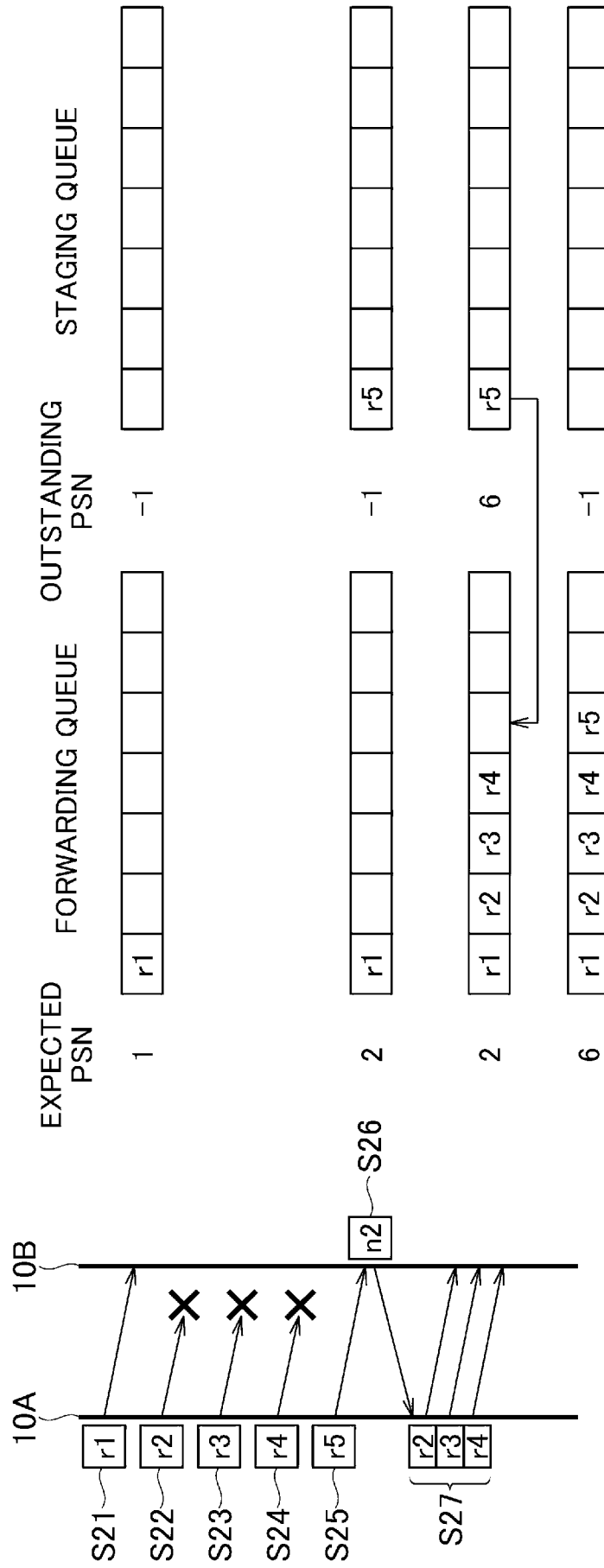
[図3]



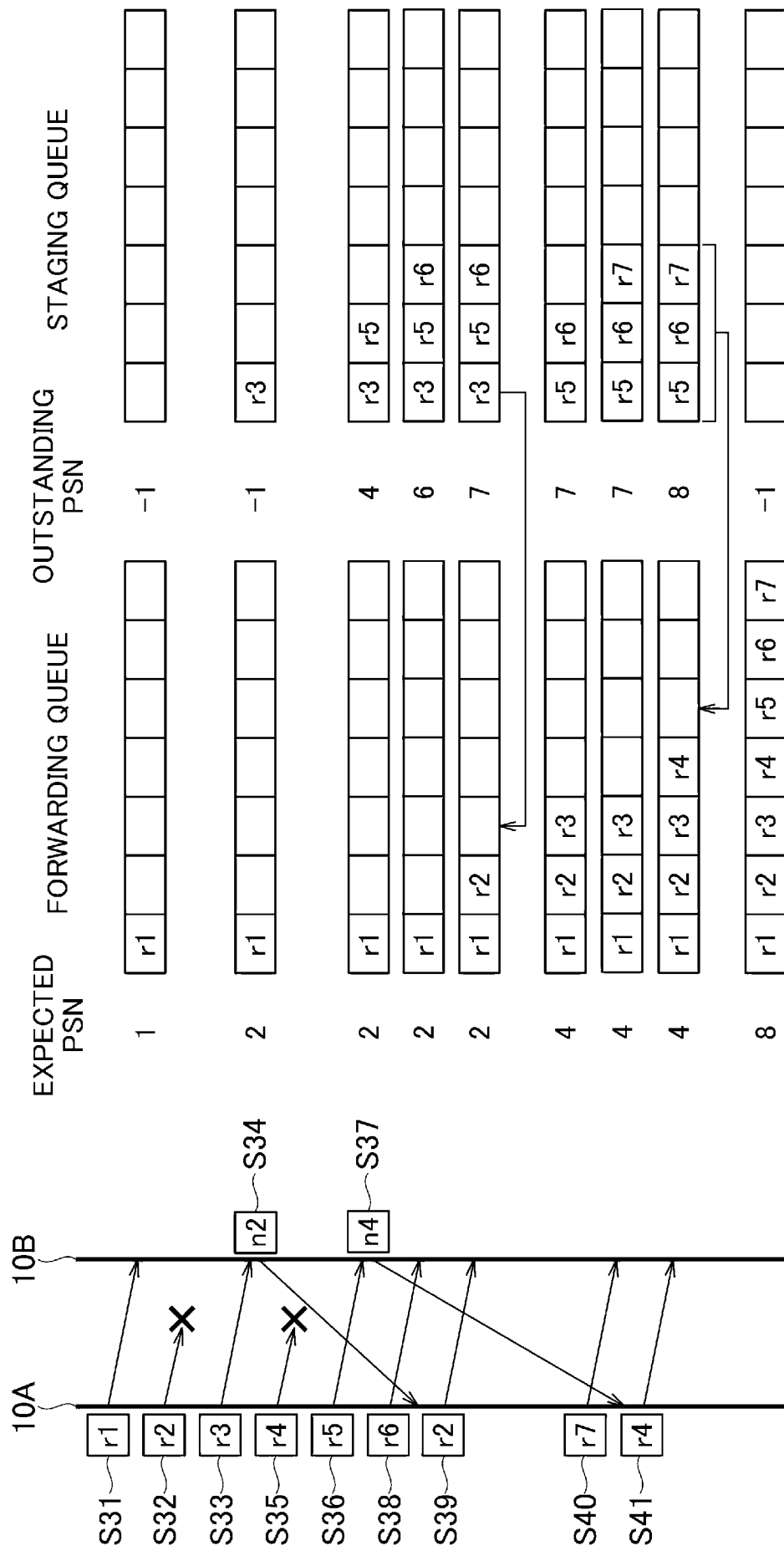
[図4]



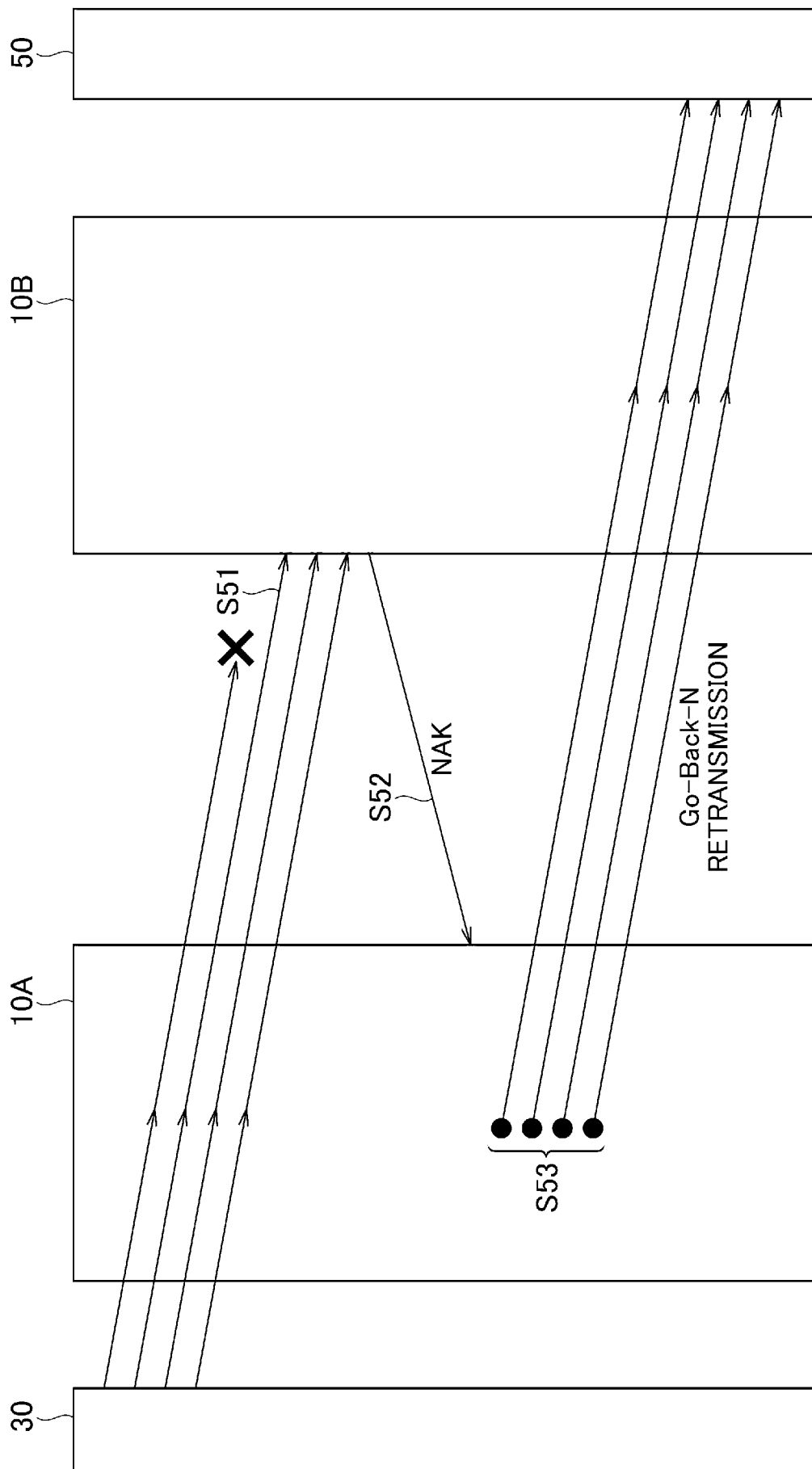
[図5]



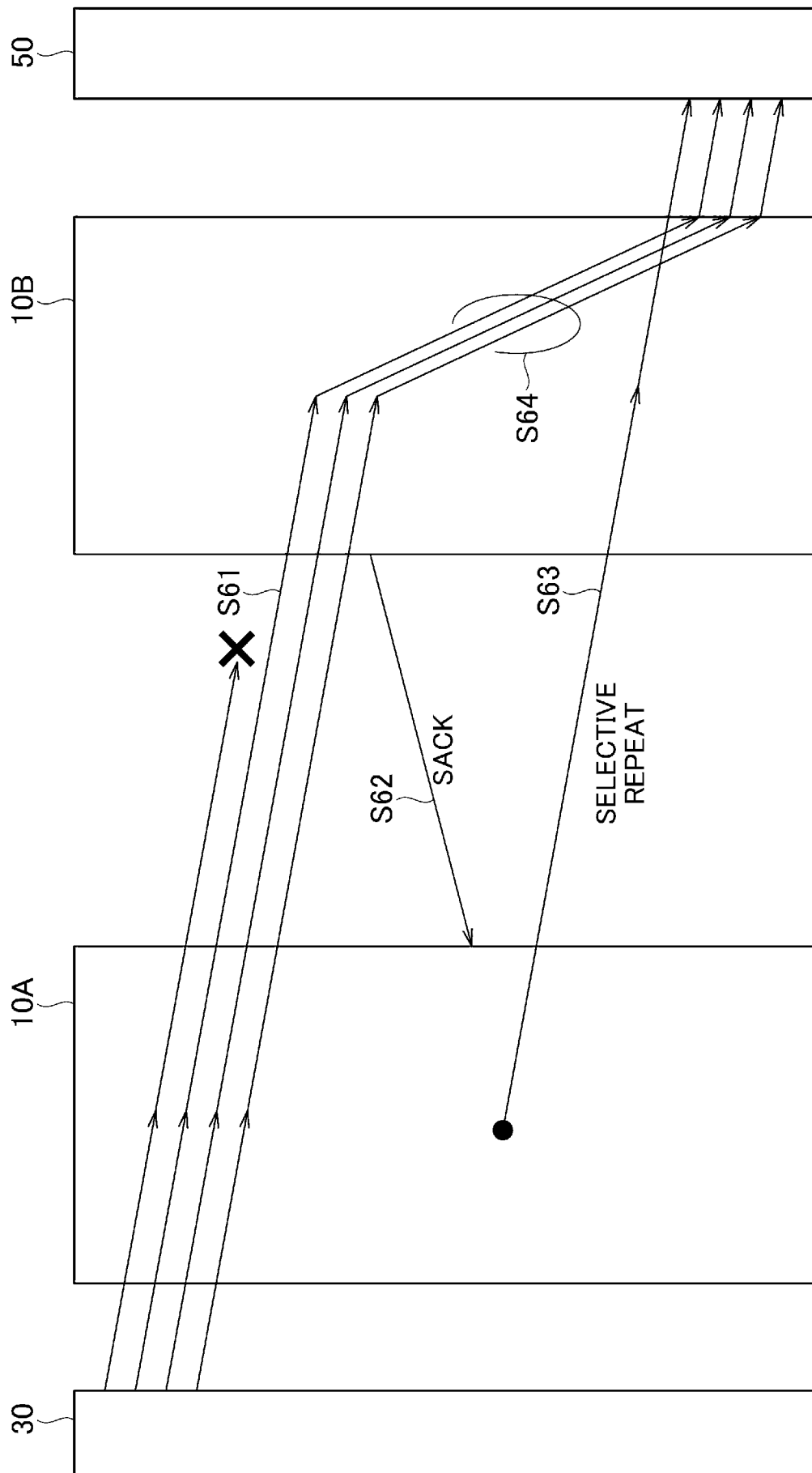
[図6]



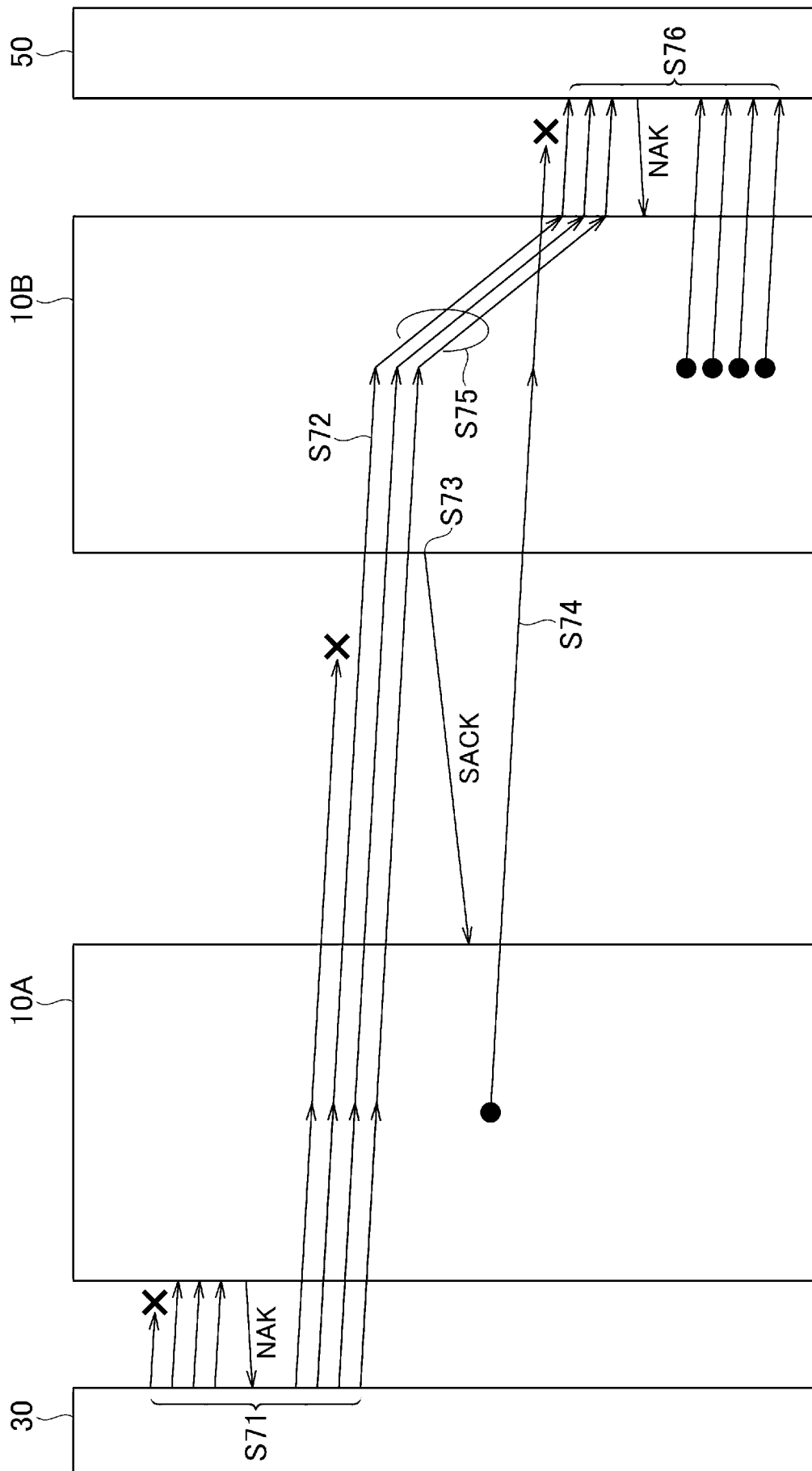
[図7]



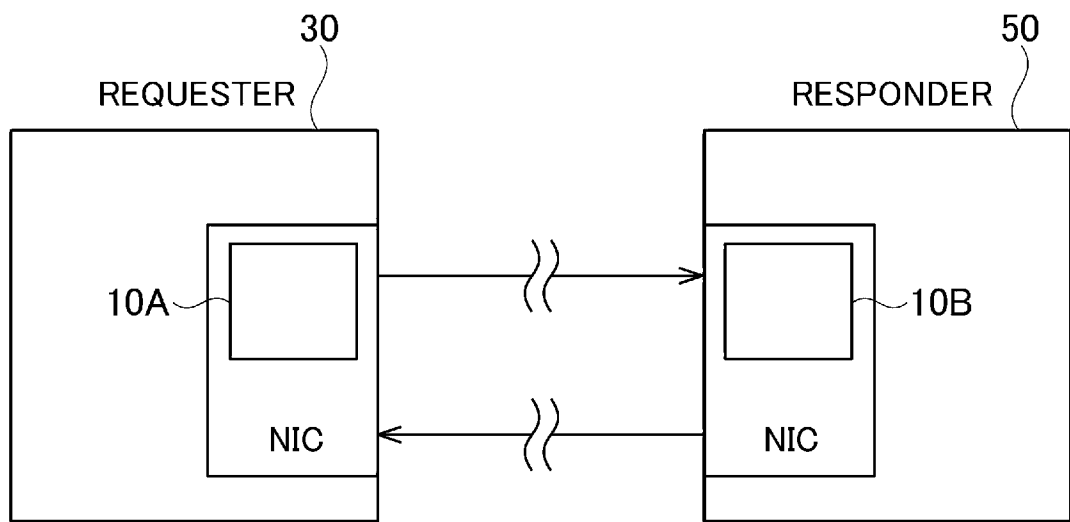
[図8]



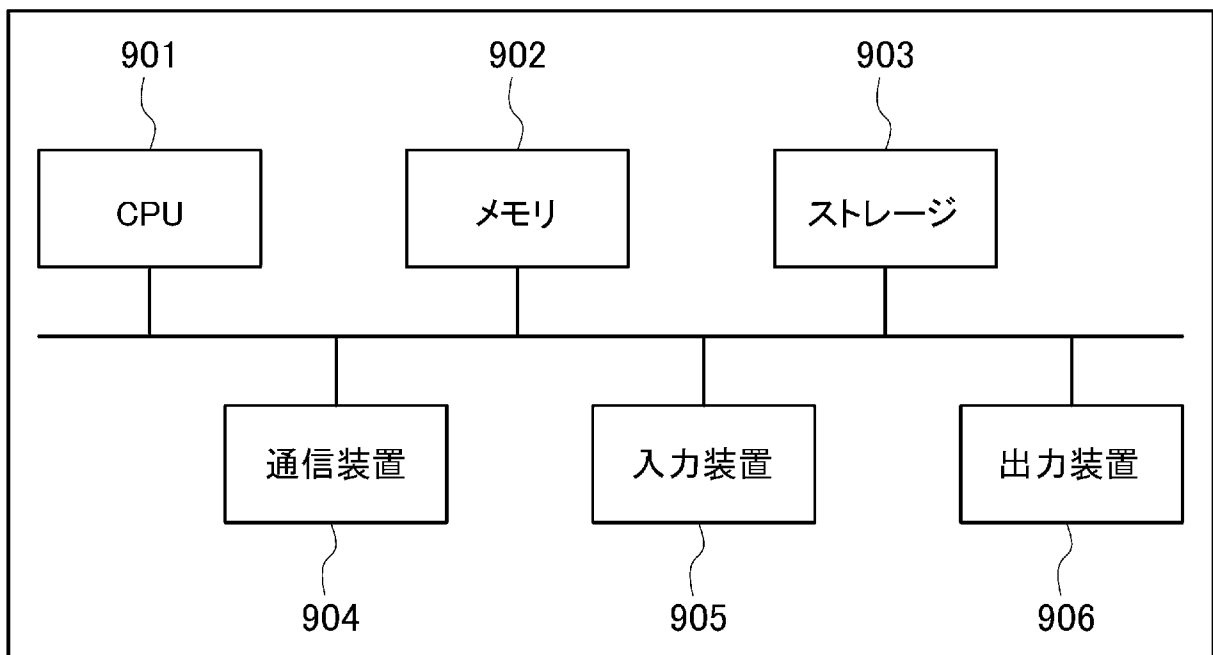
[図9]



[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/026085

A. CLASSIFICATION OF SUBJECT MATTER**H04L 49/90**(2022.01)i; **G06F 15/173**(2006.01)i; **H04L 1/1809**(2023.01)i

FI: H04L49/90; H04L1/1809; G06F15/173 665

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)

H04L49/90; G06F15/173; H04L1/1809

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 1971-2023
 Registered utility model specifications of Japan 1996-2023
 Published registered utility model applications of Japan 1994-2023

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.
Y	WO 2023/058232 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 13 April 2023 (2023-04-13) paragraphs [0022], [0028]-[0068], fig. 3, 5, 7	1, 2, 4-7
A		3
Y	JP 2013-247632 A (MITSUBISHI ELECTRIC CORPORATION) 09 December 2013 (2013-12-09) paragraphs [0025]-[0052], fig. 2-5	1, 2, 4-7
A		3
Y	CN 106936875 A (UNIV NANJING SCIENCE & TECH) 07 July 2017 (2017-07-07) paragraphs [0031], [0044], fig. 1	1, 2, 4-7

☐ 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

07 September 2023

Date of mailing of the international search report

19 September 2023

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/JP2023/026085

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO	2023/058232	A1	13 April 2023	(Family: none)	
JP	2013-247632	A	09 December 2013	(Family: none)	
CN	106936875	A	07 July 2017	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） H04L 49/90(2022.01)i; G06F 15/173(2006.01)i; H04L 1/1809(2023.01)i FI: H04L49/90; H04L1/1809; G06F15/173 665		
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）） H04L49/90; G06F15/173; H04L1/1809 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 1971-2023年 日本国実用新案登録公報 1996-2023年 日本国登録実用新案公報 1994-2023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
Y A	WO 2023/058232 A1（日本電信電話株式会社）13.04.2023（2023-04-13） [0022], [0028]-[0068], 図3,5,7	1,2,4-7 3
Y A	JP 2013-247632 A（三菱電機株式会社）09.12.2013（2013-12-09） [0025]-[0052], 図2-5	1,2,4-7 3
Y	CN 106936875 A（NANJING UNIVERSITY OF SCIENCE AND TECHNOLOGY）07.07.2017 （2017-07-07） [0031], [0044], 図1	1,2,4-7
☐ 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、一般的技术水準を示すもの “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 07.09.2023		国際調査報告の発送日 19.09.2023
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） 羽岡 さやか 5X 3149 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/JP2023/026085

引用文献			公表日	パテントファミリー文献	公表日
WO	2023/058232	A1	13.04.2023	(ファミリーなし)	
JP	2013-247632	A	09.12.2013	(ファミリーなし)	
CN	106936875	A	07.07.2017	(ファミリーなし)	