

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6600959号
(P6600959)

(45) 発行日 令和1年11月6日(2019.11.6)

(24) 登録日 令和1年10月18日(2019.10.18)

(51) Int. Cl. F I
G 0 6 F 13/38 (2006.01) G O 6 F 13/38 3 4 O C
G 0 6 F 13/42 (2006.01) G O 6 F 13/42 3 2 O Z

請求項の数 7 (全 25 頁)

<p>(21) 出願番号 特願2015-68011 (P2015-68011) (22) 出願日 平成27年3月30日 (2015. 3. 30) (65) 公開番号 特開2016-189057 (P2016-189057A) (43) 公開日 平成28年11月4日 (2016. 11. 4) 審査請求日 平成30年2月15日 (2018. 2. 15)</p>	<p>(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 100109313 弁理士 机 昌彦 (74) 代理人 100124154 弁理士 下坂 直樹 (72) 発明者 妙嶋 慎二郎 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 田名網 忠雄</p>
--	---

最終頁に続く

(54) 【発明の名称】 障害処理装置、その装置を使用する情報処理装置、障害処理方法及び障害処理プログラム

(57) 【特許請求の範囲】

【請求項1】

上位装置から受信した指令情報群のうちの各指令情報に割り当てられた識別子に関連づけられた第一の記憶領域に、前記識別子が使用中であるか否かを示す情報を記憶する識別子記憶手段と、

前記第一の記憶領域に記憶された各前記識別子に関連づけられた第二の記憶領域に、前記識別子が割り当てられた前記指令情報が下位装置へ送信された時刻を起点とする経過時間を記憶する経過時間記憶手段と、

前記識別子が使用中でない場合には、当該識別子を前記指令情報に割り当てることを許可し、

前記識別子が使用中である場合には、当該識別子を前記指令情報に割り当てることを禁止し、

前記識別子が使用中であり、且つ当該識別子に関連づけられた前記経過時間が予め定められた閾値に達していない場合には、当該識別子を割り当てられた前記指令情報の前記下位装置への送信に対する応答を受信したときに、当該応答の前記上位装置への送信、及び当該識別子に関連づけられた前記第一の記憶領域のリセットを行い、

前記識別子が使用中であり、且つ当該識別子に関連づけられた前記経過時間が前記閾値に達している場合には、当該識別子を割り当てられた前記指令情報の前記下位装置への送信に対する応答を受信したときに、当該応答の破棄、及び当該識別子に関連づけられた前記第一の記憶領域のリセットを行う

制御手段と
を備える障害処理装置。

【請求項 2】

上位装置が出力した指令に対する下位装置からの応答情報を一時記憶する応答情報緩衝記憶装置と前記下位装置が出力する前記上位装置に対する要求情報を一時記憶する要求情報緩衝記憶装置とを更に備え、前記制御手段が前記応答情報緩衝記憶装置及び前記要求情報緩衝記憶装置から前記応答情報及び前記要求情報を順次読み出す請求項 1 の障害処理装置。

【請求項 3】

請求項 1 乃至 2 のいずれかに記載の障害処理装置と、
P C I e の規格を採用して情報の伝達を行う情報処理手段と
P C I e 規格の信号を各々の周辺機器のインターフェースに変換して接続する P C I e カードと、
前記情報処理手段から前記 P C I e カードとの接続先を切り替える P C I e 切り替え手段と、
各々の P C I e カードに接続された周辺機器と
を包含する障害処理システム。

10

【請求項 4】

上位装置から受信した指令情報群のうちの各指令情報に割り当てられた識別子に関連づけられた第一の記憶領域に、前記識別子が使用中であるか否かを示す情報を記憶する識別子記憶手段と、

20

前記第一の記憶領域に記憶された各前記識別子に関連づけられた第二の記憶領域に、前記識別子が割り当てられた前記指令情報が下位装置へ送信された時刻を起点とする経過時間を記憶する経過時間記憶手段と

を有する障害処理装置における障害処理方法であって、

前記識別子が使用中でない場合には、当該識別子を前記指令情報に割り当てることを許可し、

前記識別子が使用中である場合には、当該識別子を前記指令情報に割り当てることを禁止し、

前記識別子が使用中であり、且つ当該識別子に関連づけられた前記経過時間が予め定められた閾値に達していない場合には、当該識別子を割り当てられた前記指令情報の前記下位装置への送信に対する応答を受信したときに、当該応答の前記上位装置への送信、及び当該識別子に関連づけられた前記第一の記憶領域のリセットを行い、

30

前記識別子が使用中であり、且つ当該識別子に関連づけられた前記経過時間が前記閾値に達している場合には、当該識別子を割り当てられた前記指令情報の前記下位装置への送信に対する応答を受信したときに、当該応答の破棄、及び当該識別子に関連づけられた前記第一の記憶領域のリセットを行う

障害処理方法。

【請求項 5】

上位装置が出力した指令に対する下位装置からの応答情報を一時記憶し、前記下位装置が出力する前記上位装置に対する要求情報を一時記憶し、記憶した前記応答情報と前記要求情報とを順次読み出す請求項 4 の障害処理方法。

40

【請求項 6】

上位装置から受信した指令情報群のうちの各指令情報に割り当てられた識別子に関連づけられた第一の記憶領域に、前記識別子が使用中であるか否かを示す情報を記憶する識別子記憶手段と、

前記第一の記憶領域に記憶された各前記識別子に関連づけられた第二の記憶領域に、前記識別子が割り当てられた前記指令情報が下位装置へ送信された時刻を起点とする経過時間を記憶する経過時間記憶手段と

を有する障害処理装置が備えるコンピュータに、

50

前記識別子が使用中でない場合には、当該識別子を前記指令情報に割り当てることを許可し、

前記識別子が使用中である場合には、当該識別子を前記指令情報に割り当てることを禁止し、

前記識別子が使用中であり、且つ当該識別子に関連づけられた前記経過時間が予め定められた閾値に達していない場合には、当該識別子を割り当てられた前記指令情報の前記下位装置への送信に対する応答を受信したときに、当該応答の前記上位装置への送信、及び当該識別子に関連づけられた前記第一の記憶領域のリセットを行い、

前記識別子が使用中であり、且つ当該識別子に関連づけられた前記経過時間が前記閾値に達している場合には、当該識別子を割り当てられた前記指令情報の前記下位装置への送信に対する応答を受信したときに、当該応答の破棄、及び当該識別子に関連づけられた前記第一の記憶領域のリセットを行う

処理を実行させる障害処理プログラム。

【請求項7】

上位装置が出力した指令に対する下位装置からの応答情報を一時記憶し、前記下位装置が出力する前記上位装置に対する要求情報を一時記憶し、記憶した前記応答情報と前記要求情報とを順次読み出す処理を更に前記コンピュータに実行させる請求項6の障害処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、障害処理装置、その装置を使用する情報処理装置、障害処理方法及び障害処理プログラムに関する。

【背景技術】

【0002】

高速インターフェース、とりわけシリアル伝送では、情報をパケット化して伝達することが良く用いられている。パケットを用いた伝送方式は周辺装置との接続に用いられ、さらに情報処理の機器内でプロセッサと周辺装置との接続にも用いられる。情報処理装置の機器内の接続方式は、例えばPCI Express（登録商標）（以後PCIeと記述する）のバスアーキテクチャがある。

【0003】

PCIeは高速動作を実現できることから、一般的な情報処理装置に用いられている。しかしながら何等かの障害により、パケット送信に対する応答の受信が正常に完結しない場合に、障害の処理が正確に行われずに上位の情報処理装置のホスト等が、後から受け取った誤った情報を基に処理を続行する。結果としてシステムが不安定になる欠点が存在する。

【0004】

このPCIeは、汎用機を始めとして、種々の情報処理装置に広く用いられており、近年ではシステムの安定動作が重要視されるメインフレームにも用いられることが多くなった。システムの安定稼働が必要なメインフレームにPCIeを適用するには、前述のPCIeの欠点を克服してシステムの安定動作を確保することが求められている。

【0005】

例えば、特許文献1には、PCIeにおいて上位装置からの要求パケットに対して下位装置から規定の時間内に応答がない場合に、下位装置に対するインターフェースを行っているインターフェース装置がタイムアウトであると判断して上位装置に通知する。上位装置は、その通知を受け取ることでタイムアウトした要求パケットに代わる新たなパケットの再送を行う。この時に、上位装置は、再送の為に新たに作成したパケット自体に、最初に要求したパケットとは異なるパケットであることを示すフラグを書き込む。この書き込まれた情報により、再送が実行された後に下位装置から最初の要求パケットに対する応答パケットが送出されたとしても、インターフェース装置は、受け取ったパケットのフラグ

10

20

30

40

50

を見て古いパケットであるか否かを認識する。インターフェース装置が、当該パケットが古い要求に対する下位装置からの応答であると認識した場合、このパケットを無視し、新たに再送したパケットを正規のパケットと認識する。

【0006】

また、特許文献2に記載されている技術において、上位装置に接続した下位装置の正常性を確認する為に、入出力接続装置は検査専用のパケットを生成して対象となる下位装置に送出する。入出力接続装置は、その応答が検査用パケットの応答である場合に受信パケットを解析して、当該下位装置が異常であるか否かを判定する。また、検査用のパケットと同一のタグを有するパケットに対して当該下位装置の検査が終了するまで、入出力接続装置は、ホストから当該下位装置に対するパケットを抑止する。

10

【先行技術文献】

【特許文献】

【0007】

【特許文献1】特開2007-249646号公報

【特許文献2】特開2011-192216号公報

【発明の概要】

【発明が解決しようとする課題】

【0008】

特許文献1の技術は、PCIeの仕様として明確には規定されていないパケットを構成するヘッダーの予備の情報記載領域に、再送した旨の情報(前述のフラグ)を記述し、この情報を基にパケットの識別を行う技術である。しかしながら、ヘッダーの予備の情報記載領域に記載した当該情報は、PCIeの規格には明確には規定されていない。従って、当該情報がこれらの機器に予期しない動作を引き起こす、あるいは不正の情報であるとして動作を停止させる等の不具合を発生させる懸念がある。また、上位装置においても、PCIeの仕様に準拠することを前提に作成された各種のソフトウェアが、予期しない情報を検出したとして異常終了するなど、意図しない動作を引き起こす懸念がある。したがって、安定稼働が必要なメインフレームに特許文献1の技術を適用することはできないという問題がある。

20

【0009】

特許文献2には、下位装置からの応答が遅れて、タイムアウトを判断して応答の遅れたパケットを廃棄した後に当該パケットを受け取った場合に発生する障害を回避する技術の記載がない。したがって、下位装置から応答が遅れ、その結果上位装置において下位装置からの応答等の情報入手待ちによる動作の大幅な遅延が生じる場合には、受け取った情報が不正であることから発生する各種の障害を解消することができないという問題がある。

30

【0010】

本発明の目的は、上述した問題点を解決する障害処理装置、その装置を使用する情報処理装置、障害処理方法及び障害処理プログラムを提供することにある。

【0011】

具体的に、本発明は、上位装置が要求した指令に対する下位装置からの応答が遅延する等の障害が原因で発生した誤ったデータが、上位装置および上位装置で稼働するソフトウェアによって読み出されることを防止することを主たる目的とする。

40

【課題を解決するための手段】

【0012】

障害処理装置は、情報群の中から当該情報を一意に区別する為に付与する識別子の一覧と前記識別子に関連づけられた第一の記憶領域とを備え、前記第一の記憶領域は、前記識別子を付与した前記当該情報を特徴付ける情報を前記当該情報自体から抽出した第二の情報と、当該識別子が使用中であるか否かを示す情報とを記憶する識別子記憶手段と、前記識別子記憶手段の前記識別子の一覧に記載された各々の識別子と関連づけられた第二の記憶領域を備え、前記第二の記憶領域は、当該前記識別子を付与した時刻を起点として測定した経過時間と、前記経過時間が予め定められた閾値に達したか否かを示す情報と、を記

50

憶する経過時間記憶手段と、前記識別子が未使用の場合には前記当該識別子を他の情報の識別子として割り当てることを許可し、前記識別子が使用中であり且つ当該前記識別子に関連づけられた前記経過時間が前記閾値に達した場合には当該識別子を他の情報の識別子として割り当てることを許可しない制御手段と、を備える。

【0013】

障害処理方法は、情報群の中から当該情報を一意に区別する為に付与する識別子の一覧と前記識別子に関連づけられた第一の記憶領域とを備え、前記第一の記憶領域に前記識別子を付与した前記当該情報を特徴付ける情報を前記当該情報自体から抽出した第二の情報と、当該識別子が使用中であるか否かを示す情報とを記憶し、前記識別子記憶手段の前記識別子の一覧に記載された各々の識別子と関連づけられた第二の記憶領域を備え、前記第二の記憶領域に当該前記識別子を付与した時刻を起点として測定した経過時間と、前記経過時間が予め定められた閾値に達したか否かを示す情報と、を記憶し、前記識別子が未使用の場合には前記当該識別子を他の情報の識別子として割り当てることを許可し、前記識別子が使用中であり且つ当該前記識別子に関連づけられた前記経過時間が前記閾値に達した場合には当該識別子を他の情報の識別子として割り当てることを許可しない。

10

【0014】

障害処理プログラムは、情報群の中から当該情報を一意に区別する為に付与する識別子の一覧と前記識別子に関連づけられた第一の記憶領域とを備え、前記第一の記憶領域に前記識別子を付与した前記当該情報を特徴付ける情報を前記当該情報自体から抽出した第二の情報と、当該識別子が使用中であるか否かを示す情報とを記憶し、前記識別子記憶手段の前記識別子の一覧に記載された各々の識別子と関連づけられた第二の記憶領域を備え、前記第二の記憶領域に当該前記識別子を付与した時刻を起点として測定した経過時間と、前記経過時間が予め定められた閾値に達したか否かを示す情報と、を記憶し、前記識別子が未使用の場合には前記当該識別子を他の情報の識別子として割り当てることを許可し、前記識別子が使用中であり且つ当該前記識別子に関連づけられた前記経過時間が前記閾値に達した場合には当該識別子を他の情報の識別子として割り当てることを許可する処理をコンピュータに実行させる。

20

【発明の効果】

【0015】

本発明は、上位装置が要求した指令に対する下位装置からの応答が遅延する等の障害が原因で発生した誤ったデータが、上位装置および上位装置で稼働するソフトウェアによって読み出されることを防止できる。

30

【図面の簡単な説明】

【0016】

【図1】本発明の実施の第1の実施の形態における障害処理装置を示すブロック図である。

【図2】本発明の実施の第2の実施の形態における障害処理装置を示すブロック図である。

【図3】識別子（タグエントリー）記憶部20を示す図である。

【図4】経過時間（タイマーエントリー）記憶部30を示す図である。

40

【図5】リード処理を示すフロー図である。

【図6】応答処理のフロー図である。

【図7】タイムアウト処理のフロー図である。

【図8】メモリリードの動作を説明する図である。

【図9】応答タイムアウトを検出した時の動作を説明する図である。

【図10】応答タイムアウトにより障害が発生した時の動作を説明する図である。

【図11】識別子（タグエントリー）記憶部20を示す図である。

【図12】応答処理フロー図である。

【図13】タイムアウト処理フロー図である。

【図14】タグエントリー解放処理フロー図である。

50

【図15】メモリリードが応答タイムアウトした後に後続のメモリリードを行った場合の動作を説明する図である。

【図16】64ビットアドレス形式のメモリアクセスリクエストで使用されるTLPヘッダーの図である。

【図17】32ビットアドレス形式のメモリアクセスリクエストで使用されるTLPヘッダーの図である。

【図18】応答アクセスリクエストで使用されるTLPヘッダーの図である。

【図19】メモリリードリクエストや構成アクセスリクエストに対する応答のTLPヘッダーの図である。

【図20】本発明の実施の第3の実施の形態における障害処理システムを示すブロック図である。 10

【図21】本発明の障害処理装置101の機能ブロック図である。

【図22】識別子(タグエントリー)記憶部22を示す図である。

【図23】経過時間(タイマーエントリー)記憶部32を示す図である。

【発明を実施するための形態】

【0017】

[第1の実施の形態]

本発明の第1の実施の形態について、図面を参照して詳細に説明する。

【0018】

図1は本発明の第1の実施の形態の障害処理装置180を示すブロック図である。 20

【0019】

障害処理装置180は、制御部6と、識別子記憶部21と、経過時間記憶部31とを備えている。識別子記憶部21と、経過時間記憶部31は、制御部6に接続されている。制御部6は上位装置2から指令情報を受け取り、それを下位装置3へ出力する。制御部6は下位装置3からの応答情報を受け取り、上位装置2へ出力する。本発明は図1の構成に限定されるものではない。

【0020】

識別子記憶部21は識別子が付された記憶領域を含む。識別子記憶部21の上位装置2が送出した指令(指令情報)に関連付けした識別子の記憶領域は、当該指令情報から抽出した下位装置3を特徴付ける情報と、当該指令情報に関連する識別子が使用中であるか否かを示す識別子使用フラグとを記憶する。 30

【0021】

経過時間記憶部31は識別子記憶部21に配置された各々の識別子と関連づけられて、経過時間を記憶するカウンターを含む。すなわち、識別子ごとにカウンターが設定されている。経過時間記憶部31のカウンターは、当該識別子が付与された指令情報が制御部1から下位装置3に送出された時を起点としてカウントされる経過時間を、識別子ごとに記憶する。

【0022】

制御部6は、上位装置2から下位装置3に向けた指令情報を受け取ると、識別子記憶部21から、使用フラグがOFF、すなわち未使用である識別子を選択する。制御部6は、当該指令情報の一部を、識別子記憶部21の、選択した識別子の記憶領域に書き込む。さらに制御部6は、下位装置3に、選択した識別子が付与された指令情報を出力すると同時に、選択した識別子に関連付けられた経過時間記憶部31のカウンターを、一定時間経過するごとに経過した時間が加算された数値に更新する。 40

【0023】

経過時間記憶部31の当該識別子のカウンターが予め定めた閾値と一致した場合に、制御部6が当該識別子のカウンターがタイムアウトしたと判断する。

【0024】

経過時間記憶部31の当該識別子のカウンターがタイムアウトしていない期間に制御部6が下位装置3から、指令情報に対する応答情報を受け取ると、当該応答情報に付与され 50

ている識別子が識別子記憶部 2 1 の記憶領域に存在するか否かを検索する。その検索の結果、識別子記憶部 2 1 に記憶されている当該識別子使用フラグが有効である場合、制御部 6 は受け取った応答情報（以降、単に「応答」ともいう）を上位装置 2 に送信し、識別子記憶装置 2 1 の当該識別子の記憶領域をリセットする。この操作で、制御部 6 は、当該識別子を新たに発生した上位装置 2 からの次の処理の識別のために割り当てることが可能となる。

【 0 0 2 5 】

次に、経過時間記憶部 3 1 の当該識別子に関連づけられた当該経過時間の数値がタイムアウトの閾値を超えた場合を説明する。

【 0 0 2 6 】

制御部 6 は、当該識別子に関連づけられた当該経過時間の値が別に定める閾値に達すると、当該識別子の指令に対する応答待ちが規定の時間を経過したと認識する。この場合、制御部 6 は、識別子記憶部 2 1 の当該識別子の記憶領域に配置されたタイムアウトフラグを ON に設定して、タイムアウトであることを記録する。さらに制御部 6 は、当該識別子に関連する指令情報に対する応答がなくタイムアウトしたことを上位装置 2 に通知する。この処理により、当該識別子は、当該指令情報に対する処理が完了するまで保持され、他の指令に割り当てされない。上位装置 2 は当該識別子の処理が当該タイムアウトの通知を受け取り、タイムアウトしたことを認識する。

【 0 0 2 7 】

さらに、当該識別子のカウンターがタイムアウトした後に、制御部 6 が下位装置 3 から当該識別子に対応する応答を受け取った場合に、制御部 6 が受け取った当該応答の情報を廃棄する。制御部 6 が、識別子記憶部 2 1 経過時間記憶部 3 1 の当該識別子に対応する記憶領域を消去して、記憶内容のリセットを行う。この処理により、上位装置 2 からの新たな指令に対して、前述のタイムアウトした識別子が割り当てられることを防止する。

【 0 0 2 8 】

以上説明した様に、本発明の第 1 の実施の形態によれば、タイムアウトした識別子が別の指令に割り当てられて、先に割り当てた時点の指令の応答が、後に割り当てた同じ識別子の情報であると誤認識されることを防止することが出来る。なぜなら、制御部 6 は、タイムアウトした識別子に対して、当該識別子の再割り当てを禁止し、前に割り当てられた識別子の応答が完結した後に当該識別子の再割り当てを許可するからである。

【 0 0 2 9 】

[第 2 の実施の形態]

第 2 の実施の形態について説明する。第 2 の実施の形態では、本発明の障害処理装置を P C I e に適用した場合を説明する。

【 0 0 3 0 】

P C I e ではパケットベースのプロトコルが採用されている。また互いに関連・依存する複数の処理をまとめ、一体不可分の処理単位として扱うことをトランザクション処理と呼び、そのような処理単位をトランザクションと呼ぶ。P C I e におけるプロトコル階層においてもトランザクション層があり、その層は、トランザクション層パケットの構築と処理を行っている。トランザクション層のパケットは T r a n s a c t i o n _ L a y e r _ P a c k e t (以下 T L P) と呼ばれ、その T L P のヘッダーに、接続した周辺機器に対する要求の情報が格納されている。

【 0 0 3 1 】

図 2、図 2 2 及び図 2 3 を参照して本発明の第 2 の実施の形態について説明する。図 2 は本発明の第 2 の実施の形態における障害処理装置 1 5 0 を示すブロック図である。

【 0 0 3 2 】

障害処理装置 1 5 0 は、制御部 1 と、識別子記憶部 2 2 (以後タグエントリー記憶部 2 2 と呼ぶ)と、経過時間記憶部 3 2 (以後タイマーエントリー記憶部 3 2 と呼ぶ)とを備えている。

【 0 0 3 3 】

タグエントリー記憶部 2 2 と、タイマーエントリー記憶部 3 2 は制御部 1 に接続されている。制御部 1 はリクエストバッファ 4 と応答バッファ 5 を含む。制御部 1 は上位装置 2 から指令情報を受け取り、下位装置 3 へ出力する。制御部 1 は下位装置 3 からの応答情報を受け取り上位装置 2 へ出力する。本発明は図 2 の構成に限定されるものではない。

【 0 0 3 4 】

なお、本例では上位装置 2 はプロセッサあるいはプロセッサを搭載したカードである。下位装置 3 は P C I e スイッチ等に代表される周辺機器との接続を行うインターフェース装置である。

【 0 0 3 5 】

図 2 2 に示すタグエントリー記憶部 2 2 は、上位装置 2 から下位装置 3 に送出する指令情報の識別子であるタグエントリー番号 (0 、 1 、 2 、 3 、 ・ ・ ・ 、 n ; n は整数) と、その番号が付与された記憶領域 2 2 - 0 から記憶領域 2 2 - n を含む記憶装置である。各々の記憶領域 2 2 - 0 から記憶領域 2 2 - n は、上位装置 2 が下位装置 3 に送出する当該指令情報から抽出した下位装置 3 を特徴付ける情報、を記憶するデータ領域 2 2 - n - m と、当該タグエントリー番号が使用中であるか否かを示す有効フラグ 2 2 - n - 1 と、当該タグエントリー番号が付与された応答待ちでタイムアウトが発生したことを示すタイムアウトフラグ 2 2 - n - 4 を含む。

【 0 0 3 6 】

図 2 3 に示すタイマーエントリー記憶部 3 2 はタグエントリー記憶部 2 2 と関連づけられた当該タグエントリー番号と、当該タグエントリー番号が付与されたタイマー領域 3 2 - 0 からタイマー領域 3 2 - n を含む記憶装置である。タイマー領域 3 2 - 0 からタイマー領域 3 2 - n は、制御部 1 が下位装置 3 に指令を送出した時刻を起点として計る経過時間を記憶するタイマーカウント (タイマーカウント値と呼ぶ) 3 2 - n - 2 と当該タイマーカウント 3 2 - n - 2 が有効か否かを示す有効フラグ 3 2 - n - 1 とを含む。

【 0 0 3 7 】

図 2 に示すリクエストバッファ 4 は、下位装置 3 から上位装置 2 に対するリクエスト T L P を一時記憶する。図 2 に示す応答バッファ 5 は、下位装置 3 から上位装置 2 に対する応答 T L P を一時記憶する。前述のリクエストバッファ 4 および応答バッファ 5 に一時記憶された各々の情報は、その後制御部 1 により順次読み出される。

【 0 0 3 8 】

制御部 1 は上位装置 2 から下位装置 3 に向けた指令情報を受け取ると、タグエントリー記憶部 2 2 の記憶領域を検索し、有効フラグ 2 2 - n - 1 が未使用であるタグエントリー番号を選択する。制御部 1 は、当該指令情報の一部を当該タグエントリー番号が付与された記憶領域に書き込む。制御部 1 は、タグエントリー番号を付与した指令情報を下位装置 3 へ出力する。更に、制御部 1 は、前述のタグエントリー番号と関連づけられた当該タイマーエントリー記憶部 3 2 にある当該タイマー領域 3 2 - n の記憶内容を、一定時間経過するごとに経過した時間を加算した数値に更新する。

【 0 0 3 9 】

制御部 1 は、当該タグエントリー番号に関連づけられたタイマー領域の当該経過時間を記憶するタイマーカウンターの数値が、予め定めた閾値を超えていない期間中に、制御部 1 が下位装置 3 から応答情報を受け取ると、当該応答情報に付与されているタグエントリー番号と同じ番号がタグエントリー記憶部 2 2 の記憶領域に存在するか否かを検索する。

【 0 0 4 0 】

制御部 1 が検索した結果、タグエントリー記憶部 2 2 に記憶されている当該タグエントリー番号の記憶領域にある有効フラグ 2 2 - n - 1 が有効である場合は、制御部 1 が受け取った応答情報を上位装置 2 に送信する。

【 0 0 4 1 】

さらに、制御部 1 が、タグエントリー記憶部 2 2 の当該タグエントリー番号の記憶領域の情報をリセットする。この操作で、制御部 1 が当該タグエントリー番号は新たに発生した上位装置 2 からの次の指令の識別に当該タグエントリー番号を割り当てすることを可能

10

20

30

40

50

にする。

【 0 0 4 2 】

次に、タイマーエントリー記憶部 3 2 の当該タグエントリー番号に関連づけられたタイマー領域のタイマーカウントの値がタイムアウトの閾値を超えた場合を説明する。

【 0 0 4 3 】

制御部 1 は、当該タグエントリー番号に関連づけられたタイマー領域のタイマーカウントの値が別に定める閾値に達すると、当該タグエントリー番号の指令に対する応答待ちが規定の時間を経過したと認識する。制御部 1 が、タグエントリー記憶部 2 2 の当該タグエントリー番号の記憶領域に配置されたタイムアウトフラグを ON に設定して、タイムアウトであることを表示する。制御部 1 が上位装置 2 に当該タグエントリー番号の指令に対する応答がタイムアウトしたことを通知する。この処理により、当該タグエントリー番号は、当該指令に対する処理が完了するまで保持され、他の指令に割り当てされない。上位装置 2 は当該タグエントリー番号の処理が、当該タイムアウトの通知を受け取り、タイムアウトしたことを認識する。

10

【 0 0 4 4 】

さらに、当該タグエントリー番号に関連づけられたタイマー領域のタイマーカウントがタイムアウトした後に、制御部 1 が、下位装置 3 から当該タグエントリー番号に対応する応答を受け取った場合は、制御部 1 が、下位装置 3 から受け取った当該応答の情報を廃棄する。制御部 1 が、タグエントリー記憶部 2 2 及びタイマーエントリー記憶部 3 2 の当該タグエントリー番号に対応する記憶領域およびタイマー領域の記憶を消去する。この処理により、制御部 1 が上位装置 2 からの新たな指令に対して、前述のタイムアウトしたタグエントリー番号を割り当てられることを防止できる。なお、本例ではリクエストバッファ 4 と応答バッファ 5 が制御部 1 の内部に配置した例を説明したが、制御部 1 の外部に配置した構成にしてもよい。

20

【 0 0 4 5 】

以上説明した様に、本発明の第 2 の実施形態によれば、パケットの応答が何等かの理由で遅れたことで、当該トランザクション処理がタイムアウトにより完結しない場合に、タグエントリー番号に関連づけられた記憶領域の記憶内容を廃棄し、廃棄前の記憶領域に記憶した応答情報を上位装置が誤って受け取ることによる誤動作を防止することが出来る。なぜなら、タイムアウトしたタグエントリー番号に対して、当該タグエントリー番号の再割り当てを禁止し、前に割り当てられたタグエントリー番号が付与された指令の応答が完結した後に当該タグエントリー番号の再割り当てを許可するからである。

30

【 0 0 4 6 】

[第 3 の実施の形態]

本発明の第 3 の実施の形態について図を参照して説明する。図 2 0 は情報処理装置 1 0 0 の本発明の障害処理装置を適用した情報処理装置のブロック図である。

【 0 0 4 7 】

情報処理装置 1 0 0 0 は、中央処理装置 3 0 0、デバイス (D E V) 1 3 0、P C I e スイッチ 1 2 0、P C I e カード 1 4 0 ~ P C I e カード 1 4 3、および外部記憶装置 2 0 0 0、外部記憶装置 3 0 0 0 を含む。

40

【 0 0 4 8 】

中央処理装置 3 0 0 は、演算処理を行うプロセッサ 1 0 0、障害処理装置 1 0 1、障害処理装置 1 0 2 及びデータを格納するメモリ (M E M) 1 1 0 を含む。

【 0 0 4 9 】

プログラム 2 0 0 はプロセッサ 1 0 0 の上に展開され、プロセッサ 1 0 0 を制御して各種処理を行う。(プログラム 2 0 0 は図示していない。)

図 2 1 は障害処理装置 1 0 1 の機能ブロック図である。障害処理装置 (以後 P C I e コントローラと呼ぶ) 1 0 1 は、制御部 1 と、識別子記憶部 2 0 (以後、タグエントリー記憶部 2 0 と呼ぶ) と、経過時間記憶部 3 0 (以後、タイマーエントリー記憶部 3 0 と呼ぶ) とを備えている。

50

【 0 0 5 0 】

タグエントリー記憶部 2 0 と、タイマーエントリー記憶部 3 0 は、制御部 1 に接続されている。制御部 1 はリクエストバッファ 4 と応答バッファ 5 を含む。制御部 1 はプロセッサ 1 0 0 から指令情報を受け取り、P C I e スイッチ 1 2 0 へ出力する。

【 0 0 5 1 】

図 3 を用いてタグエントリー記憶部 2 0 を説明する。タグエントリー記憶部 2 0 は、上位装置であるプロセッサ 1 0 0 から下位装置 3 である P C I e スイッチ 1 2 0 に送出する指令情報の識別子であるタグエントリー番号と、その番号が付与された複数の記憶領域 2 0 - 0 から記憶領域 2 0 - n を含む記憶装置である。前述の記憶領域は、当該記憶領域 2 0 - n が使用され有効であることを示す有効フラグ 2 0 - n - 1 と、プロセッサ 1 0 0 から指示された要求情報を格納するリクエスト情報部 2 0 - n - 2 と、プロセッサ 1 0 0 からの指示に従って作成し P C I e スイッチ 1 2 0 に送信した T L P の H e a d e r を格納する T L P ヘッダ部 2 0 - n - 3 を含む。

10

【 0 0 5 2 】

図 4 を用いてタイマーエントリー記憶部 3 0 を説明する。タイマーエントリー記憶部 3 0 は、制御部 1 が P C I e スイッチ 1 2 0 へのリクエストに対する応答のタイムアウト監視を行うための情報を記憶する記憶装置である。タイマーエントリー記憶部 3 0 は、前述のタグエントリー番号と、そのタグエントリー番号に関連づけられたタイマー領域 3 0 - 0 からタイマー領域 3 0 - n を含む。タイマーエントリー記憶部 3 0 は、タイマー領域 3 0 - n のタイマーが有効化どうかを示す有効フラグ 3 0 - n - 1 と、タイマーカウント（タイマーカウント値と呼ぶ） 3 0 - n - 2 を含む。

20

【 0 0 5 3 】

制御部 1 は P C I e スイッチ 1 2 0 からの応答情報を受け取りプロセッサ 1 0 0 へ出力する。

【 0 0 5 4 】

P C I e スイッチ 1 2 0 は、障害処理装置 1 0 1 を経由したプロセッサ 1 0 0 と、ネットワーク（図示していない）、P C I e カード 1 4 0 から P C I e カード 1 4 3 の各々に接続された外部記憶装置 2 0 0 0 及び外部記憶装置 3 0 0 0 との接続の切り替えを行う。

【 0 0 5 5 】

P C I e コントローラ 1 0 1 は、プロセッサ 1 0 0 と P C I e スイッチとの中間に位置し、プロセッサ 1 0 0 と P C I e カード 1 4 0 から P C I e カード 1 4 3 との packet 通信で発生する障害の処理を行う。

30

【 0 0 5 6 】

障害処理装置（P C I e コントローラと呼ぶ） 1 0 2 は、P C I e スイッチ 1 2 0 を経由しない特定の処理を行うデバイス 1 3 0 とプロセッサ 1 0 0 とを接続するとともに、当該装置間の packet 通信で生じる障害の処理を行う。

【 0 0 5 7 】

図 2 0 において、外部記憶装置 2 0 0 0 及び外部記憶装置 3 0 0 0 は P C I e カードと各々に接続した磁気ディスク装置などであり、データを記憶し、プロセッサ 1 0 0 からの指令に基づきデータの書き込み、読出しを行う。

40

【 0 0 5 8 】

P C I e カード 1 4 0 から P C I e カード 1 4 3 は、packet ベースのシリアルインターフェースの信号を、接続する先の周辺装置等各々のインターフェース信号及びコネクタ形状に変換する。

【 0 0 5 9 】

P C I e コントローラ 1 0 2 とデバイス 1 3 0 は P C I e __ I / F 1 3 1 で接続する。P C I e コントローラ 1 0 1 と P C I e スイッチ 1 2 0 は P C I e __ I / F 1 2 1 で接続する。P C I e スイッチ 1 2 0 とデバイス 1 3 0 は P C I e __ I / F 1 2 2 で接続する。P C I e スイッチ 1 2 0 と P C I e カード 1 4 0 は P C I e __ I / F 1 2 3 で接続する。P C I e スイッチ 1 2 0 と P C I e カード 1 4 1 は P C I e __ I / F 1 2 4 で接続する。

50

PCIeスイッチ120とPCIeカード142はPCIe__I/F125で接続する。
 PCIeスイッチ120とPCIeカード143はPCIe__I/F126で接続する。
 PCIeカード140と外部記憶装置2000はインターフェース2001で接続する。
 PCIeカード141と外部記憶装置2000はインターフェース2002で接続する。
 PCIeカード142と外部記憶装置3000はインターフェース3001で接続する。
 PCIeカード143と外部記憶装置2000はインターフェース3002で接続する。

【0060】

入出力デバイスとのデータ交換の方法について説明する。

【0061】

上位装置2から外部記憶装置2000又は外部記憶装置3000に指令を出力する場合、PCIeカード140からPCIeカード143のいずれかを經由して指令が伝達される。

10

【0062】

プロセッサ100からPCIeカード140からPCIeカード143に対する指令はプロセッサ100が入出力デバイスにアクセスする為の命令をMEM110にアクセスする為の命令と同じアドレス空間で扱うことで入出力デバイスに対する操作を実現する。この方法は、Memory__Mapped__Input__Output(以後MMIOと呼ぶ)と呼ばれる。以後、本発明の動作説明では、このMMIOを前提として説明する。

【0063】

ここで、PCIeのTLPヘッダーの内容を図16、図17、図18及び図19に示す。

20

【0064】

ある装置から他の装置に対して行う指令をリクエストとする。またその指令に対する他の装置からの出力を応答とする。TLPの内容は、要求内容やリクエストか応答かに応じて形式がそれぞれ異なっている。TLPヘッダーの詳細やTLPを受信したときの動作はPCIeスペックに記載されているので、本発明に関係があることについてのみ説明する。

【0065】

図16、17は主にメモリアクセスリクエストで使用されるTLPヘッダーで、図16が64bitアドレス形式で図17が32bitアドレス形式となっている。図16、17及び18に記載のFmt(フォーマット)とType(タイプ)が要求内容を示し、アドレスが要求先を示している。図18は構成アクセスリクエストで使用されるTLPヘッダーで、要求先がBus__Number(バス番号)とDevice__Number(デバイス番号)、Function__Number(機能番号)、Register__Number(レジスタ番号)によって指定されている。

30

【0066】

図17のリクエスターID(IDentifier)41、図18のリクエスターID51及び図19のリクエスターID61は要求元を示す識別子を示す。図17のタグ42、図18のタグ52及び図19のタグ62はパケットに付与されたタグエントリー番号を示す。このリクエスターIDとタグの情報がそれぞれ抽出されて、図17、図18、図19に示すトランザクションID70を形成し、応答がリクエストに対応するものかどうかを制御部1が判断する際に使用される。

40

【0067】

なお、本例のトランザクションID70は、前述のリクエスターIDとタグとを組み合わせた情報である。制御部1が、各種のリクエストのTLPや応答のTLPに含まれるリクエスターIDや関連するタグを索引して認識する。しかし、制御部1が、トランザクションID70用に専用に記憶領域を準備して、その領域に前述のリクエストのTLPや応答のTLPから必要な情報を抽出して書き込み、その情報を利用することでもよい。

【0068】

図19はメモリアクセスリクエストや構成アクセスリクエストに対する応答のTLPヘッ

50

ダーの詳細を示す図である。応答のTLPヘッダーは、制御部1が応答に対応するリクエストを参照して、制御部1が識別するためのリクエスターID61とタグ62、応答元を示す応答IDや要求の結果を示す応答ステータスを持っている。リクエスターID61とタグ62は、応答に対応するリクエストのTLPに格納されていたリクエスターID41、リクエスターID51とタグ42、タグ52の情報をコピーして、持ちまわったものである。PCIeコントローラ101が、リクエストと応答のトランザクションID70を比較することでリクエストに対する応答が応答されたかどうかを判断する。

【0069】

次に、情報処理装置1000における、PCIeのインターフェースの動作について説明する。

10

【0070】

まず、図8を用いて、PCIeカード140～PCIeカード143からのメモリアドレスの動作を説明する。

【0071】

プロセッサ100上で動作するソフトウェア200が、PCIeカード140～PCIeカード143のレジスタなどを読み出す場合、プロセッサ100の読出し命令をPCIeカード140～PCIeカード143のレジスタを示すアドレスを指定して、読出し指令を実行する(a-1)。

【0072】

プロセッサ100は、読出し命令を実行されるとPCIeコントローラ101に前記アドレスを指定してメモリアドレスのTLPコマンドの送信を指示する(a-2)。PCIeコントローラ101は、メモリアドレスを指示されるとタグ番号をタグエントリ記憶部20から取得し(a-3)、リードリクエストTLPを作成してPCIe_I/F121にTLPコマンドを送信する(a-4)。この時、リクエスターIDにはPCIeコントローラを識別する値が格納され、リクエストに対する応答を制御部1が識別する。その為、トランザクションIDを含むリクエストTLPが、タグエントリ記憶部20の当該タグエントリ番号の記憶領域に記憶される。また、応答受信時に要求元のプロセッサ100に応答を返すためプロセッサ100の要求情報が、タグエントリ記憶部20の当該タグエントリ番号の記憶領域に保持される。

20

【0073】

次に、PCIe_I/F121に送信されたリードリクエストTLPのTLPコマンドがPCIeスイッチ120を介して目的のPCIeカード140～PCIeカード143のいずれかに送信される(a-5)。

30

【0074】

PCIeカード140～PCIeカード143はリードリクエストTLPを受信すると対応する処理を行い、応答TLPを作成して読み出したデータを付けてPCIeコントローラ101に応答TLPを送信する(a-6)。この時、トランザクションIDはリードリクエストTLPのものが持ちまわされ、応答IDにはPCIeカード140～PCIeカード143のいずれであるかを識別する値が格納される。

【0075】

40

PCIeコントローラ101はPCIeスイッチ120を介してPCIeカード140～PCIeカード143からの応答TLPを受信すると記憶していたリクエストTLPと応答TLPのトランザクションIDを比較する。前述の比較した結果が、一致した場合はリクエストに対する応答であると判断する。PCIeコントローラ101は、付随していた応答データをプロセッサ100に送信し(a-9)、リクエストで使用していたタグエントリ番号を解放する(a-8)。この解放処理によりタグエントリ記憶部20の当該タグエントリ番号は他の処理に割り当てることが出来るようになる。また、PCIeコントローラ101が、リクエストTLPの記憶などに使用していた記憶領域の情報を消去する。そして、プロセッサ100はPCIeコントローラ101からの応答情報を受けると、読出し命令を完了してソフトウェア200に読み出したデータを返す(a-10)

50

)。

【0076】

次に、図9を用いて、PCIeカード140～PCIeカード143が何らかの理由によりリードリクエストTLPに対して応答TLPを返さない、あるいは応答が遅くなる等の場合の動作を説明する。

【0077】

ソフトウェア200がPCIeカード140～PCIeカード143のレジスタなどを読み出すために読み出し命令を実行し、PCIeカード140～PCIeカード143にリードリクエストTLPが送信されるまでは図8に示す動作と同じである(b-1～b-5)。

10

【0078】

ここで、PCIeカード140～PCIeカード143が一定時間内に応答TLPを送信しないとPCIeコントローラ101がタイムアウトを検出する(b-10)。PCIeコントローラ101が記憶していたプロセッサ100からの要求情報に基づいて、プロセッサ100にエラーデータを送信し(b-8)、確保していたタグエントリー番号及びタグエントリー番号に関連づけられた記憶領域、タイマー領域を解放する(b-11)。この解放処理によりタグエントリー記憶部20の当該タグエントリー番号は他の処理に割り当てることが出来るようになる。また、同時にリクエストTLP等の記憶に使用していた当該タグエントリー番号に関連づけられた記憶領域の情報が消去される。プロセッサ100はPCIeコントローラ101よりエラーデータを受信するとソフトウェア200

20

【0079】

その後、PCIeカード140～PCIeカード143が遅れて応答TLPを返した場合(b-6、b-7)、PCIeコントローラ101がその応答TLPを受信する。PCIeコントローラ101が、応答TLPに付与されたタグエントリー番号がタグエントリー記憶部20の各々の記憶領域に存在するかを検索する。しかし、応答TLPに付与されたタグエントリー番号に対応するトランザクションIDを持つタグエントリー番号が存在しないため、無効なTLPと判断し、予期しない応答として、制御部1が受け取った応答TLPを廃棄する等の処理を行う(b-12)。

30

【0080】

以上が、応答が正常に出力され、定常的にデータの授受が行われた場合の動作である。

【0081】

次に、図5のフローチャートを用いて、プロセッサ100からリードの指令を実行した場合の動作を説明する。

【0082】

制御部1はプロセッサ100からメモリリードを指示されるとリード処理(ステップ300)を実行する。

【0083】

リード処理(ステップ300)では、制御部1が、タグエントリー記憶部20より記憶領域20-nの各々を読み出す(ステップ301)。制御部1が、各々のタグエントリー20-nの有効フラグ20-n-1を参照して未使用の記憶領域20-nがあるか否かを判断する(ステップ302)。制御部1が、有効フラグ20-n-1を設定して記憶領域20-nを確保する(ステップ303)。

40

【0084】

そして、制御部1が、プロセッサ100からのメモリリードの要求情報に従ってリードリクエストTLPを作成する(ステップ304)。この時、制御部1が、トランザクションIDのタグ番号に記憶領域20-nの番号nを格納する。

【0085】

その後、プロセッサ100からのメモリリードの要求情報を、制御部1が記憶領域20-nのリクエスト情報部20-n-2に格納する。制御部1が、作成したリードリクエス

50

ト T L P の T L P ヘッダーを記憶領域 2 0 - n の T L P ヘッダー部 2 0 - n - 3 に格納する (ステップ 3 0 5)。

【 0 0 8 6 】

また、制御部 1 が、記憶領域 2 0 - n に対応するタイマー領域 3 0 - n の有効フラグ 3 0 - n - 1 とタイマーカウント値 3 0 - n - 2 を設定する。制御部 1 が、タイマーを起動する (ステップ 3 0 6)。制御部 1 が、最後に作成したリードリクエストの T L P コマンドを P C I e スイッチ 1 2 0 に送信して処理を終える (ステップ 3 0 7、ステップ 3 0 8)。

【 0 0 8 7 】

これにより、P C I e カード 1 4 0 ~ P C I e カード 1 4 3 がリードリクエスト T L P を受信すると、読み出されたデータを付けて応答 T L P を P C I e コントローラ 1 0 1 に対して送信する。

10

【 0 0 8 8 】

次に、図 6 のフローチャートを用いて応答処理の動作を説明する。P C I e コントローラ 1 0 1 の制御部 1 が応答 T L P を受信すると応答の処理を実行する。

【 0 0 8 9 】

応答処理 (ステップ 4 0 0) では、制御部 1 が、応答バッファ 5 から受信した応答 T L P を読み出し (ステップ 4 0 1)、制御部 1 が、タグエントリー 2 0 - n の各々を読み出す (ステップ 4 0 1)。制御部 1 が、応答 T L P と同じトランザクション I D を持つ有効な記憶領域 2 0 - n があるか否かを判断 (ステップ 4 0 3)。

20

【 0 0 9 0 】

ここで同じトランザクション I D を持つ記憶領域 2 0 - n が無い場合は、制御部 1 が、応答 T L P を予期しない応答として処理し (ステップ 4 0 8)、処理を終える (ステップ 4 0 9)。

【 0 0 9 1 】

応答 T L P と同じトランザクション I D を持つ記憶領域 2 0 - n (タグエントリーの記憶領域) がある場合 (ステップ 4 0 3 : Y E S)、その記憶領域 2 0 - n に格納されたリクエストに対する応答に対する処理になる。この場合、制御部 1 が、記憶領域 2 0 - n に対応するタイマー領域 3 0 - n の有効フラグ 3 0 - n - 1 をクリアしてタイマーを停止し (ステップ 4 0 4)、記憶領域 2 0 - n のリクエスト情報部 2 0 - n - 2 に従って応答 T L P のデータをプロセッサ 1 0 0 に送信する (ステップ 4 0 5)。

30

【 0 0 9 2 】

そして、制御部 1 が、記憶領域 2 0 - n の有効フラグ 2 0 - n - 1 をクリアして記憶領域 2 0 - n を解放して (ステップ 4 0 6)、処理を終える (ステップ 4 0 7)。

【 0 0 9 3 】

次に、図 7 のフローチャートを用いて、P C I e カードからの応答が遅延したことによって応答タイムアウトになった場合の動作を説明する。

【 0 0 9 4 】

応答タイムアウトは、制御部 1 が、有効なタイマー領域 3 0 - n のタイマーカウント値 3 0 - n - 2 の数値を減算してその数値が 0 になった時に、発生させる。

40

【 0 0 9 5 】

タイムアウト処理 (5 0 0) では、制御部 1 が、該当するタイマー領域 3 0 - n の有効フラグ 3 0 - n - 1 をクリアしてタイマーの停止と解放を行い (ステップ 5 0 1)、タイマー領域 3 0 - n に対応する記憶領域 2 0 - n のリクエスト情報部 2 0 - n - 2 に従ってエラーデータをプロセッサ 1 0 0 に送信する (ステップ 5 0 2)。その後、制御部 1 が、記憶領域 2 0 - n の有効フラグ 2 0 - n - 1 をクリアすることで当該 2 0 - n のタグエントリー番号を解放し (ステップ 5 0 3)、処理を終える (ステップ 5 0 4)。

【 0 0 9 6 】

次に、図 1 0 を用いて、本発明の障害処理装置の動作を行わない場合に、情報処理装置 1 0 0 0 に障害が発生する状況を説明する。図 1 0 は、P C I e カード 1 4 0 ~ P C I e

50

カード143に対して複数のメモリリードが行われている状態で任意のメモリリードが応答タイムアウトになった場合の動作を示す。

【0097】

ソフトウェア200が、PCIeカード140～PCIeカード143のメモリリードを行う為に、プロセッサ100に読み出し指令を実行する(c-1)。プロセッサ100がPCIeコントローラにコマンドを送信する(c-2)。PCIeコントローラ101がリードリクエストのTLPコマンドをPCIeスイッチ120に送信する(c-3)。PCIeスイッチ120がTLPコマンドをPCIeカード140～PCIeカード143にTLPコマンドを送信してPCIeカード140～PCIeカード143がそれを受信する(c-4)。しかし、PCIeカード140～PCIeカード143が何らかの理由により応答の送信が遅れると、PCIeコントローラ101は応答タイムアウトを検出して(c-17)タグエントリ番号を解放し(c-18)、プロセッサ100にエラーデータ(c-5)を返す。ソフトウェア200は、エラーデータを受け取り、読み出しが失敗したことが分かる(c-6)。

10

【0098】

しかし、この後、ソフトウェア200がPCIeカード140～PCIeカード143のメモリリードを行うことがある。この場合に、PCIeコントローラ101が応答タイムアウトになったリクエストと同じタグエントリ番号を取得してしまう。そして、PCIeカード140～PCIeカード143にリードリクエストTLPを発行した後(c-7からc-10)に、前回のリードリクエストTLPに対する応答TLPが遅れてPCIeコントローラ101に到着する現象(c-11、c-12)が発生する。この結果、後続のリードリクエストTLPにおけるトランザクションIDと応答タイムアウトになった応答TLPのトランザクションIDが一致してしまう。従って、この場合には、PCIeコントローラ101は先行のリードリクエストTLPに対する応答TLPを後続のリードリクエストTLPに対する応答TLPとして誤って処理してしまう。そして、制御部1がプロセッサ100に間違った応答データを出力して(c-13)、ソフトウェア200に間違った読み出しデータを渡してしまう(c-14)。これは、制御部1が、ソフトウェア200に間違ったデータを渡すことになり、ソフトウェアが誤動作する原因となる。

20

【0099】

次に、図15及び図11を用いて、本発明の障害処理装置を動作させて、ソフトウェア200が誤動作することを回避する動作について説明する。本例は、PCIeカード140～PCIeカード143のメモリリードが応答タイムアウトした後に後続のメモリリードを行った場合で説明する。

30

【0100】

まず、図21のPCIeコントローラ101の制御部1において、記憶領域20-nに応答タイムアウトが発生したことを示すタイムアウトフラグ20-n-4(図11参照)を設ける。

【0101】

最初にソフトウェア200がPCIeカード140～PCIeカード143のメモリリードを行ってPCIeカード140～PCIeカード143の応答が遅れることによって応答タイムアウトが発生する状況である。この状況は、図15において、状態(d-1～d-4、d-20)である。この状態は、図10の状態(c-1～c-4、c-17、)による説明と同じである。

40

【0102】

ここで、PCIeコントローラ101が応答タイムアウトを検出したとき、制御部1はタイムアウト処理(ステップ700)を実行する。

【0103】

図13のタイムアウト処理のフローチャートを参照すると、タイムアウト処理(700)では、制御部1が、タイマー領域30-nのタイマーを停止する。制御部1が、有効フラグ30-n-1をクリアしてタイマー領域30-nを解放する(ステップ701)。制

50

御部 1 が、タイマー領域 30 - n に対応する記憶領域 20 - n のリクエスト情報部 20 - n - 2 の要求情報に従い、エラーデータをプロセッサ 100 に送信する（ステップ 702）、状態（d - 5）。そして、制御部 1 が、記憶領域 20 - n のタイムアウトフラグ 20 - n - 4 を設定して処理を終える（ステップ 703、ステップ 704）、状態（d - 20）。これにより、応答タイムアウトになったリクエスト TLP で使用されていたタグエントリー番号は保持されたままとなり、応答タイムアウトでは、タグエントリー番号は解放されない。

【0104】

次に、ソフトウェア 200 が後続のメモリリードを行って P C I e コントローラ 101 がその指示を受けたとき、制御部 1 は図 5 のリード処理（300）を実行する。これは図 15 における状態（d - 7、d - 8）の動作である。

【0105】

このとき、応答タイムアウトになったリクエスト TLP で使用された記憶領域 20 - n は確保されたままなので、制御部 1 は、保留したタグエントリー番号以外の記憶領域 20 - n が選択される。これは、状態（d - 21）である。これにより、後続のメモリリードのリードリクエスト TLP におけるトランザクション ID は先行の応答タイムアウトになったリードリクエスト TLP とは異なったものになる。これは、状態（d - 9、d - 10）である。

【0106】

ここで、先行のリードリクエスト TLP に対する応答 TLP を P C I e コントローラが受けた場合には、図 12 のフローチャートに示す応答処理（600）が実行される。制御部 1 が、応答バッファ 5 から受信した応答 TLP を取り出す（ステップ 601）。状態（d - 11、d - 12）。制御部 1 が、タグエントリー 20 - n の各々を読み出す（602）。制御部 1 が、応答 TLP と同じトランザクション ID を持つタグエントリー 20 - n があるか判断する（ステップ 603）。

【0107】

このとき、対応する記憶領域 20 - n は確保されたままなのでトランザクション ID は一致する（ステップ 603：Yes）。制御部 1 が記憶領域 20 - n のタイムアウトフラグ 20 - n - 4 が設定されていることを確認する（ステップ 604）。制御部 1 が、タイムアウトフラグ 20 - n - 4 が設定されていることを認識して、応答 TLP は予期しない応答であると判断（ステップ 605：Yes）。そして予期しない応答に対する処理を行う（ステップ 612）。その後、当該処理を終了する（ステップ 613）。

【0108】

これにより、応答タイムアウトになったリードリクエスト TLP に対する応答 TLP は、予期しない応答として破棄され、プログラムに間違ったデータを送信することが防止される。状態（d - 22）。

【0109】

そして、P C I e カード 140 ~ P C I e カード 143 から後続のリードリクエスト TLP に対する応答 TLP が送信される。これは状態（d - 13）である。P C I e コントローラ 101 がその情報を受信すると、状態（d - 14）、制御部 1 は図 12 の応答処理（600）を実行する。制御部 1 が、応答バッファ 5 から応答 TLP を取り出し（ステップ 601）、タグエントリー記憶部 20 より同じトランザクション ID を持つタグエントリー 20 - n があるか判断する（ステップ 603）。

【0110】

このとき、対応する記憶領域 20 - n は確保されているので、制御部 1 が、記憶領域 20 - n のタイムアウトフラグ 20 - n - 4 の内容を確認する（ステップ 604）。

【0111】

制御部 1 が、タイムアウトフラグ 20 - n - 4 が設定されていないことを認識して、応答がタイムアウトになっていないと判断する（ステップ 605：No）。そして、制御部 1 が、記憶領域 20 - n に対応するタイマー領域 30 - n のタイマーを停止する。制御部

10

20

30

40

50

1が、有効フラグ30-n-1をクリアしてタイマー領域30-nを解放し(ステップ606)、記憶領域20-nのリクエスト情報部20-n-3の要求情報に従って応答TLPの応答データをプロセッサ100に送信する(ステップ607)。状態(d-15)

その後、制御部1が、記憶領域20-nの有効フラグ20-n-1をクリアしてタグエントリー20-nを解放して処理を終える(ステップ608、ステップ609)。状態(d-23)。

【0112】

PCIeコントローラ101から応答データを受信したプロセッサ100は読出し命令を完了させてプログラムに受信した読出しデータを渡す。状態(d-16)。これにより、プログラム200はPCIeカード140~PCIeカード143のメモリリードに対応したデータを受信することができる。

10

【0113】

その後さらに、プログラム200はPCIeカード140~PCIeカード143をリセットするなどして障害処理を行い、PCIeカード140~PCIeカード143から応答タイムアウトになったリードリクエストTLPに対応するトランザクションIDが使用されないことを保証する。その後、プログラム200は、PCIeコントローラ101の特定のレジスタに書き込みを行う為に、書き込み指令を発行する。状態(d-17)プロセッサ100は、PCIeコントローラ101に対して、応答タイムアウトで保持したままとなっている記憶領域20-nの解放を指示する解放コマンドを送信する。状態(d-18)。

20

【0114】

PCIeコントローラ101は、前述のソフトウェア200からの指示を制御部1にお伝達し、制御部1が、図14のフローに示すタグエントリー解放処理(800)を行う。

【0115】

タグエントリー解放処理(800)では、記憶領域20のタイムアウトフラグ20-n-4が設定されている全タグエントリー番号の有効フラグ20-n-1とタイムアウトフラグ10-n-4をクリアし、解放を行う(ステップ801)。

【0116】

この操作により、応答タイムアウトONにさり、確保されたままとなっていたタグエントリーが、以後のパケットを生成する作業に用いるタグエントリーに使用することが可能になる。状態(d-24)。

30

【0117】

以上説明した様に、本発明の第3の実施形態によれば、PCIeのデバイスからの応答がタイムアウトしてトランザクション処理が完結しない場合に、情報処理装置のプログラム等が、誤ったデータの影響をうけて動作が不安定になることを防止できる。なぜなら、制御部1が、PCIeのパケット処理で、タイムアウトとなったパケットのタグエントリー番号を使用しない様に無効化して、障害が解消された時に有効化することで、同じタグエントリー番号が別の指令に対して、誤って割り当てられることを防止できるからである。

【符号の説明】

40

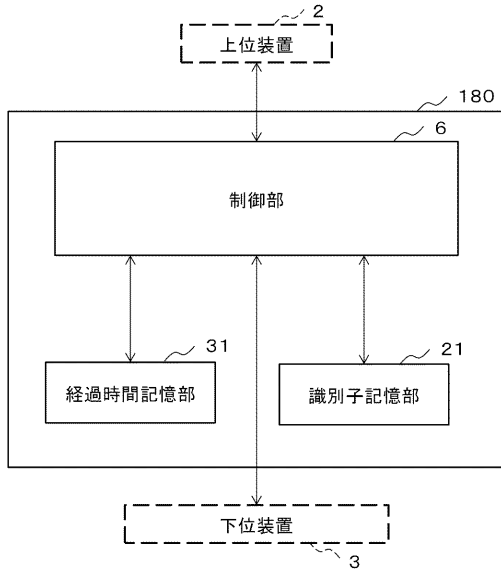
【0118】

- 1 制御部
- 2 上位装置
- 3 下位装置
- 4 リクエストバッファ
- 5 応答バッファ
- 6 制御部
- 20 識別子記憶部
- 21 識別子記憶部
- 22 識別子記憶部

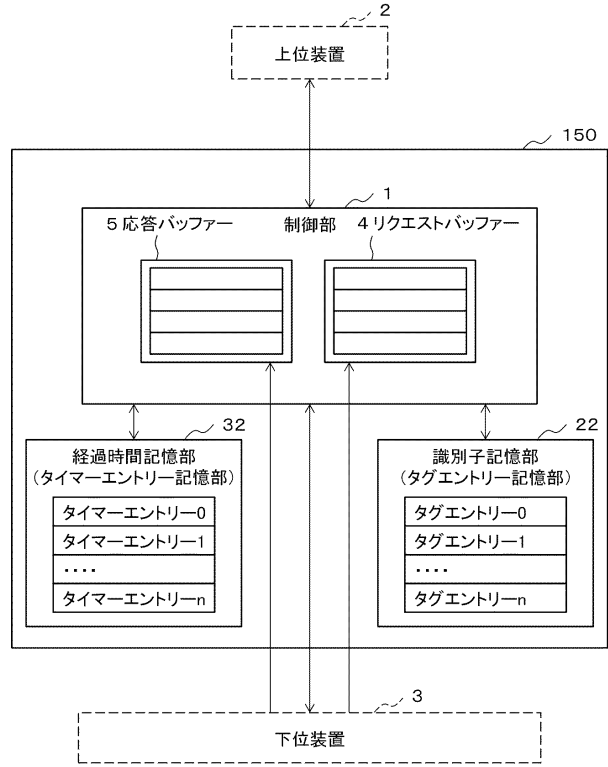
50

3 0	経過時間記憶部	
3 1	経過時間記憶部	
3 2	経過時間記憶部	
1 0 0	プロセッサ	
1 0 1	障害処理装置	
1 0 2	障害処理装置	
1 2 0	P C I e スイッチ	
1 2 1	P C I e _ I / F	
1 2 2	P C I e _ I / F	
1 2 3	P C I e _ I / F	10
1 2 4	P C I e _ I / F	
1 2 5	P C I e _ I / F	
1 2 6	P C I e _ I / F	
1 3 0	デバイス	
1 3 1	P C I e _ I / F	
1 4 0	P C I e カード	
1 4 1	P C I e カード	
1 4 2	P C I e カード	
1 4 3	P C I e カード	
1 5 0	障害処理装置	20
1 8 0	障害処理装置	
2 0 0	プログラム	
3 0 0	中央処理装置	
1 0 0 0	情報処理装置	
2 0 0 0	外部記憶装置	
2 0 0 1	インターフェース	
2 0 0 2	インターフェース	
3 0 0 0	外部記憶装置	
3 0 0 1	インターフェース	
3 0 0 2	インターフェース	30

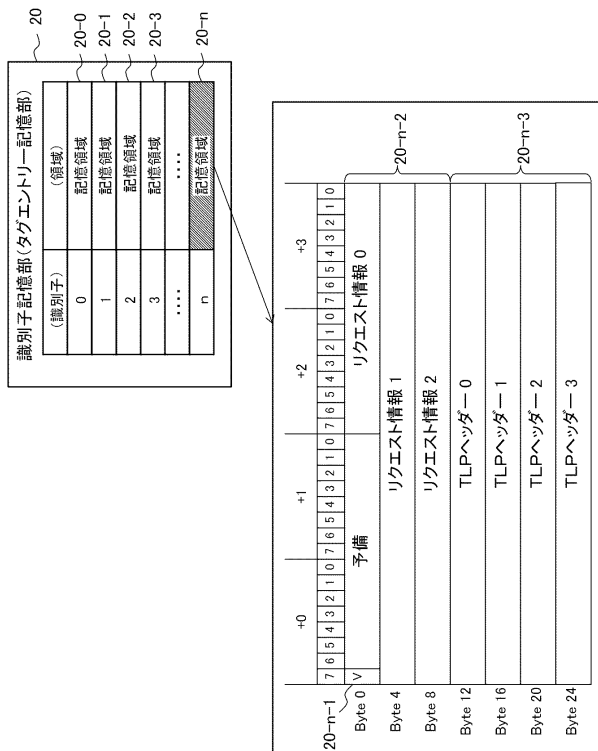
【図1】



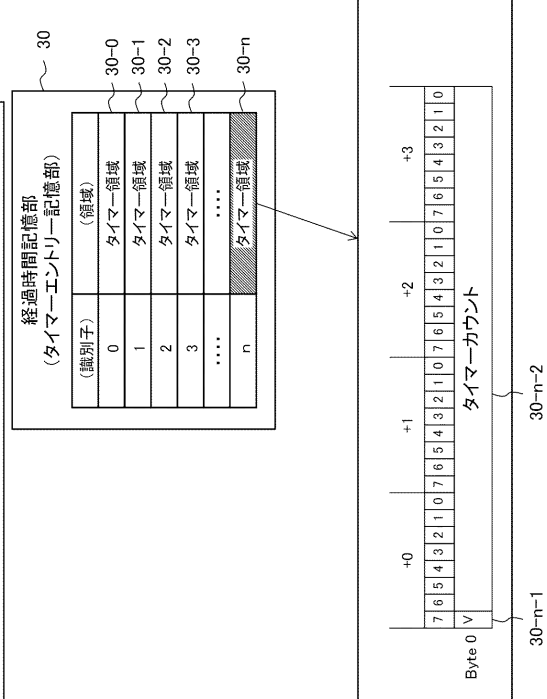
【図2】



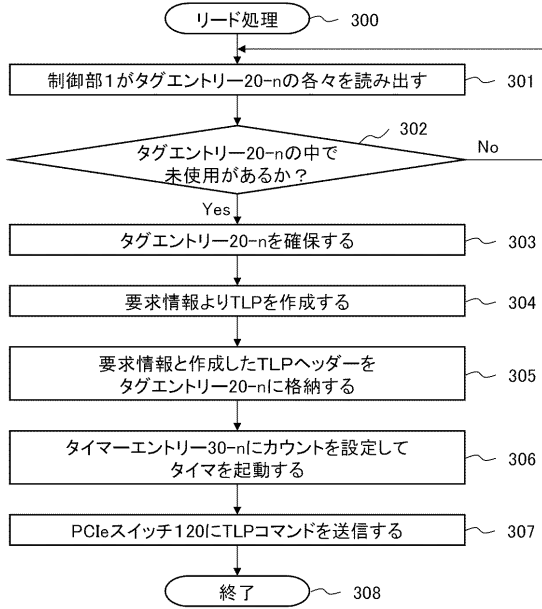
【図3】



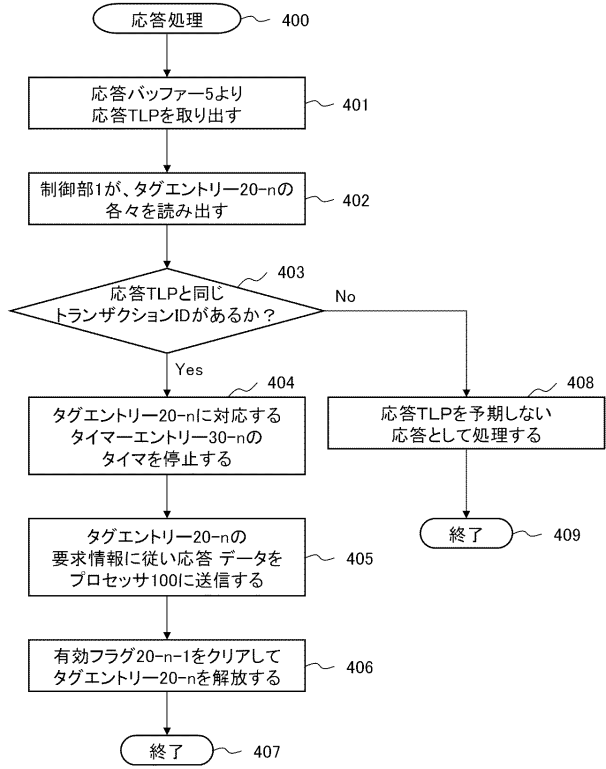
【図4】



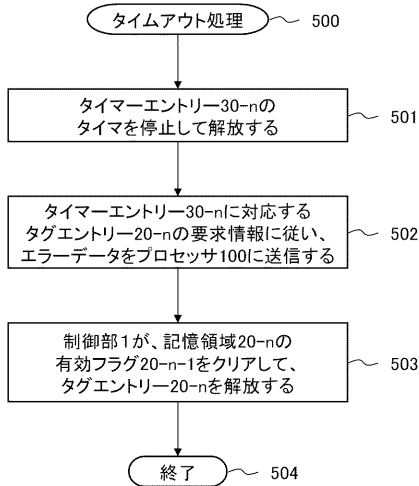
【図5】



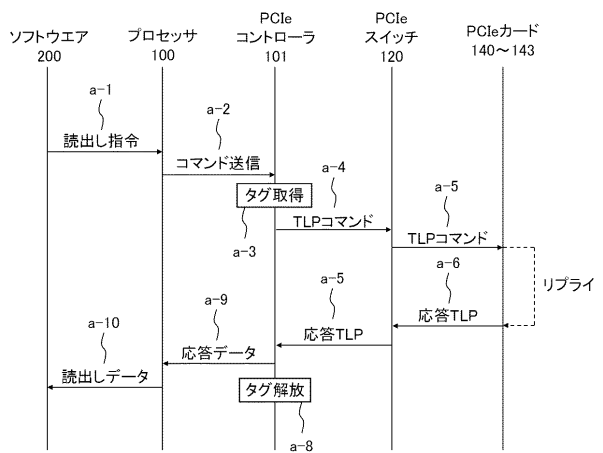
【図6】



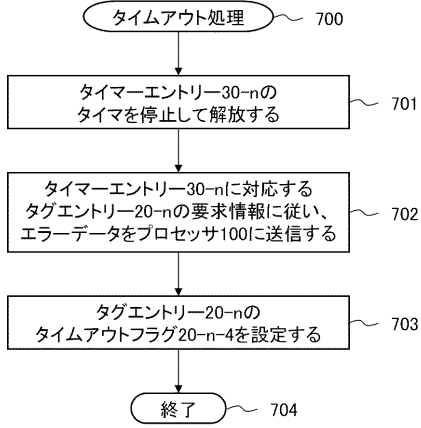
【図7】



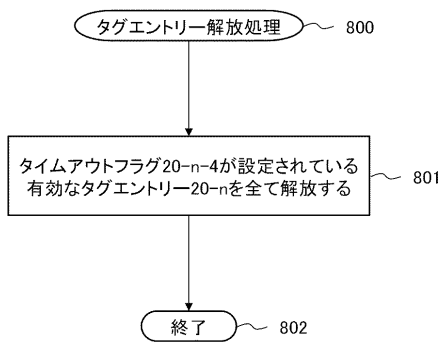
【図8】



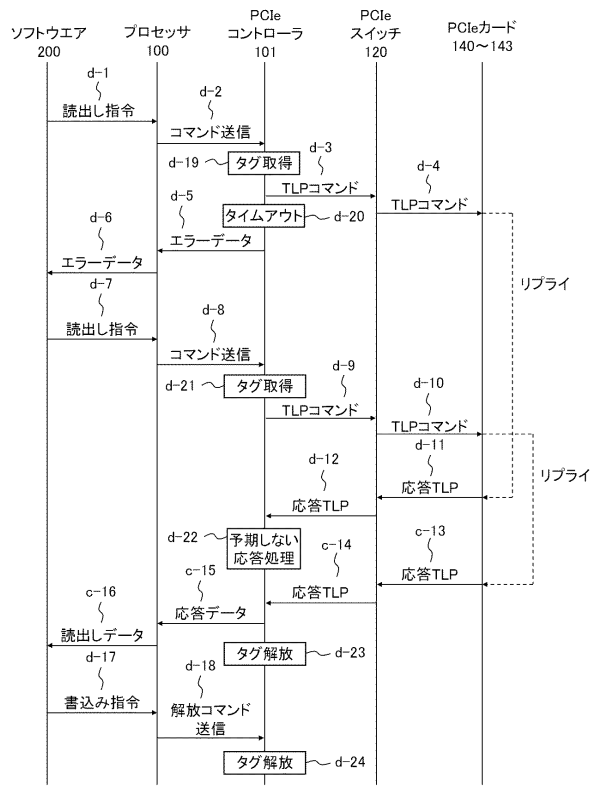
【図13】



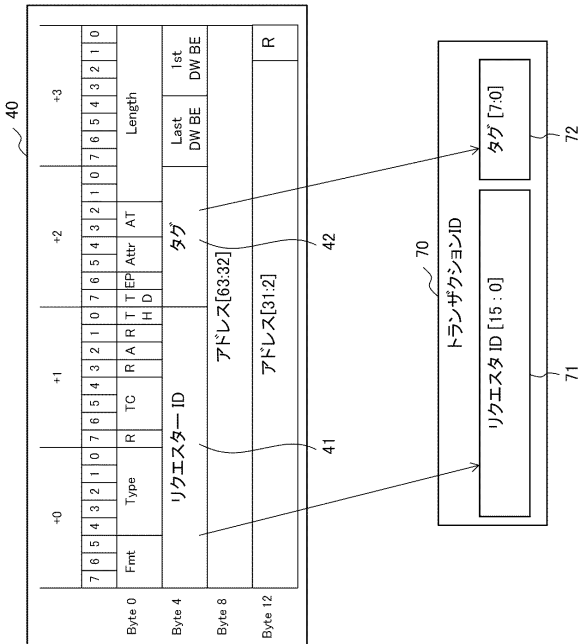
【図14】



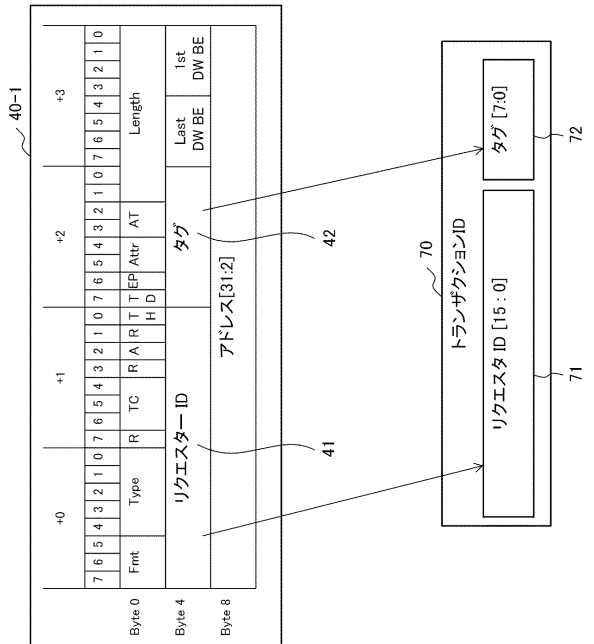
【図15】



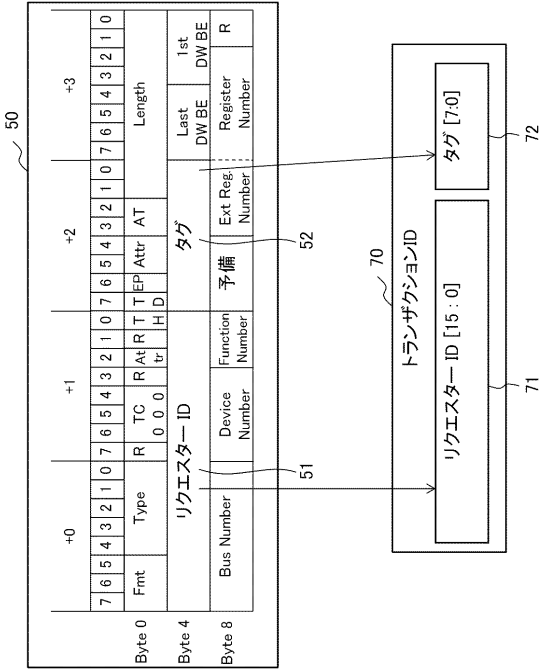
【図16】



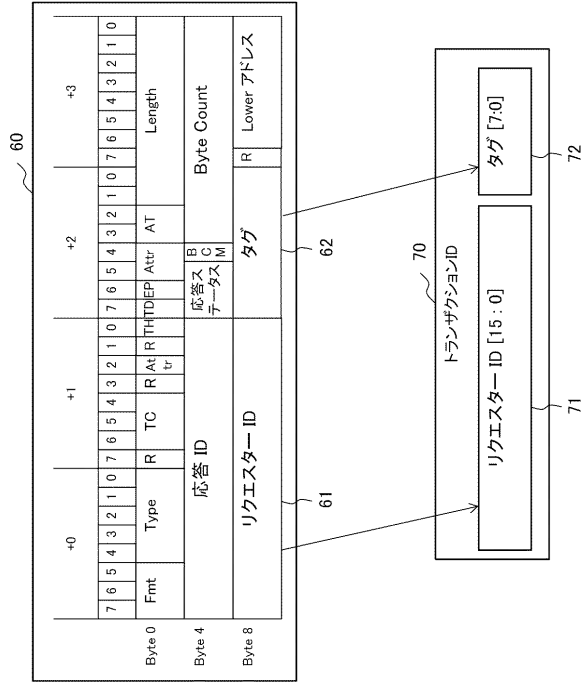
【図17】



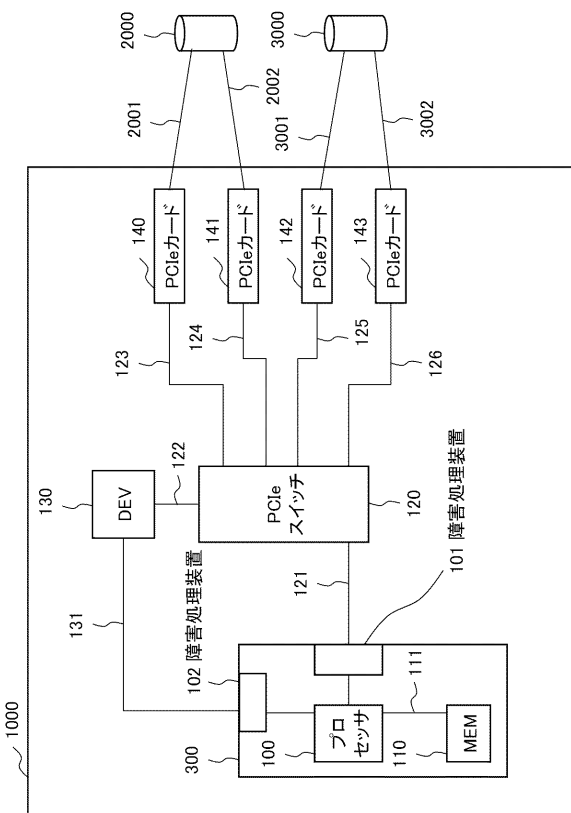
【図18】



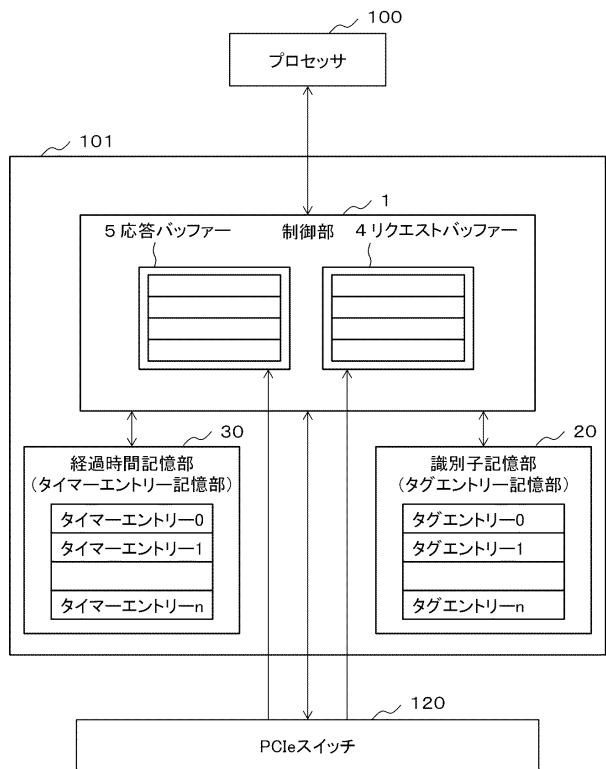
【図19】



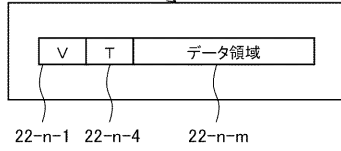
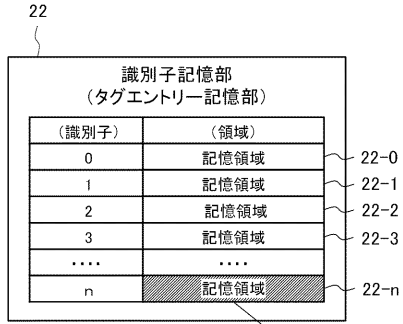
【図20】



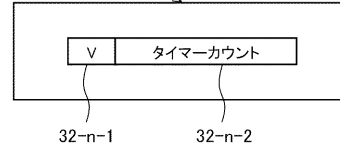
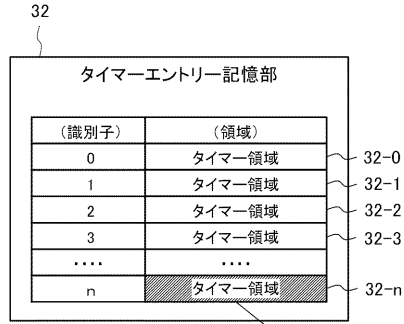
【図21】



【図 2 2】



【図 2 3】



フロントページの続き

(56)参考文献 米国特許出願公開第2009/0204974 (US, A1)

特開2013-054414 (JP, A)

特開2011-248759 (JP, A)

特開平06-083742 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 13/00

G06F 13/38 - 13/42