

RV3000

HITACHI
Inspire the Next

HITACHI Fibre Channel アダプタ

ユーザーズガイド
(高速系切替支援機能編)

重要なお知らせ

本書の内容の一部、または全部を無断で転載、複製することは固くお断わりします。

本書の内容について、改良のため予告なしに変更することがあります。

本書の内容については万全を期しておりますが、万一ご不審な点や誤りなど、お気づきのことがありましたら、お買い求め先へご一報くださいますようお願いいたします。

本書に準じないで本製品を運用した結果については責任を負いかねますので、あらかじめご了承ください。

規制・対策などについて

輸出規制について

本製品を輸出される場合には、外国為替および外国貿易法並びに米国の輸出管理関連法規などの規制をご確認のうえ、必要な手続きをお取りください。なお、ご不明の場合はお買い求め先にお問い合わせください。

登録商標・商標について

Linux は、Linus Torvalds 氏の米国およびその他の国における登録商標あるいは商標です。

HPE は、米国およびその他の国における Hewlett Packard Enterprise Company の商標または登録商標です。

Red Hat は、Red Hat Inc.の米国およびその他の国における登録商標あるいは商標です。

VMware , VMware vSphere, ESXi は、米国およびその他の国における VMware, Inc.の登録商標または商標です。

Marvell は、米国およびその他の国における Marvell Technology Group の登録商標あるいは商標です。

その他、本マニュアル中の製品名および会社名は、各社の商標または登録商標です。

著作権について

このマニュアルの内容はすべて著作権によって保護されています。このマニュアルの内容の一部または全部を、無断で記載することは禁じられています。

All rights reserved. Copyright© 2021, 2024, Hitachi, Ltd.

Licensed Material of Hitachi, Ltd.

Reproduction, use, modification or disclosure otherwise than permitted in the License Agreement is strictly prohibited.

はじめに

本書は、32Gbps HITACHI Fibre Channel アダプタが提供する「高速系切替支援機能」について説明したものです。

お取り扱いいただく前に本書の内容をよくお読みください。

マニュアルの表記

マークについて

マニュアル内で使用しているマークの意味は次のとおりです。

注意	これは、装置の重大な損傷*、または周囲の財物の損傷もしくはデータの喪失を引き起こすおそれのある潜在的な危険の存在を示すのに用います。 「装置の重大な損傷」とは、システム停止に至る装置の損傷をさします。
 制限	装置の故障や障害の発生を防止し、正常に動作させるための事項を示します。
 補足	装置を活用するためのアドバイスを示します。

オペレーティングシステム（OS）の略称について

本マニュアルでは、次の OS 名称を省略して表記します。

Red Hat Linux

Red Hat Enterprise Linux 8 Server

（以下 Red Hat Enterprise Linux 8 或いは RHEL8）

尚、本書では上記全ての Red Hat Linux を Linux と略記します。

VMware

VMware ESXi™ 7.x

（以下、VMware ESXi 7.x）

尚、本書では上記全ての VMware を VMware と略記します。

用語の定義

#	用語	内容
1.	アダプタドライバ	HITACHI Gigabit Fibre Channel アダプタのデバイスドライバを示します。OS や製品を限定した表記がない限り、デバイスドライバに下記「HFC-PCM-2 PE」と「HFC-PCM-2 EE」の2製品を含んだものを表します。
2.	HFC-PCM-2	「HFC-PCM-2 PE」と「HFC-PCM-2 EE」の2製品の総称です。
3.	HFC-PCM-2 PE	Hitachi Fibre Channel – Path Control Manager 2 Premium Edition for Linux の略です。文中で【HFC-PCM-2 PE】記号を使用している箇所は、HFC-PCM-2 PE 特有の説明です。また【HFC-PCM-2】記号を使用している箇所は、HFC-PCM-2 PE を使用している場合でも該当する説明です。
4.	HFC-PCM-2 EE	Hitachi Fibre Channel – Path Control Manager 2 Enterprise Edition for Linux の略です。また【HFC-PCM-2】、【HFC-PCM-2 EE】記号を使用している箇所は、HFC-PCM-2 EE を使用している場合でも該当する説明です。
5.	FC	Fibre Channel の略です。ホストと周辺装置をつなぐインタフェースであり 1Gbps、2Gbps、4Gbps、8Gbps などの速度があります。
6.	HBA	Host Bus Adapter の略。各種の物理的なアダプタカードを指すもので、Fibre Channel Adapter は HBA の一種です。
7.	OS	オペレーティングシステムの略語です。
8.	エラーログ	Linux で標準的に利用されるログで、このログの中にハードウェアやソフトウェアの異常を記録します。
9.	LU	Logical Unit の略記で、Hitachi Storage を論理的に分割してホストに一台の物理的なディスクのように見せる単位です。LU に割り当てられた番号が LUN(Logical Unit Number)です。
10.	論理デバイスファイル	Linux 環境において、各 LU に対して付けられる論理ファイルです。通常、/dev/sda~となります。
11.	LU パス	HBA と LU とを結ぶパスのことを示します。
12.	Target パス	HBA と IO 側接続ポートとを結ぶパスのことを示します。
13.	パスグループ	本製品では、複数の Target パスをグループ化し、オペレーティングシステムには 1 本のパスのように見せます。この Target パスの集合を 1 つのパスグループとして定義します。
14.	オフライン	論理的に使用不可（論理閉塞）とした状態です。HFC-PCM-2 の Failback 機能での回復可能です。
15.	閉塞	物理的な閉塞状態。パスを使用不可とし、HBA ポートを光断状態とする。
16.	SCSI	Small Computer System Interface の略です。ホストと周辺装置をつなぐインタフェースです。Hitachi Fibre Channel では SCSI プロトコルを FC プロトコルにマッピングしています。よって、Hitachi Fibre Channel よりソフトウェア的に上位のレイヤーである Linux カーネルやアプリケーションはサーバに接続されるデバイスを SCSI デバイスと認識できます。
17.	FC Switch	Fibre Channel Switch の略です。複数のサーバからディスク装置の FC インタフェースを共有する場合などに使用します。
18.	I/O（または IO）	本書では以下のいずれかを表します。 <ul style="list-style-type: none"> ・ ディスク装置や FC Switch などの HBA と接続するデバイス ・ サーバと上記デバイスを接続する HBA や光ケーブルなどの物理的なインタフェース ・ 上記デバイスや物理的インタフェースの総称 ・ アプリケーションや OS カーネルなどのソフトウェアが HBA を通じて上記デバイスに対してデータ転送を要求するトランザクション <p>本書では特にディスク装置に対するトランザクションをさして「ディスク I/O」と呼ぶ場合もあります。</p>

19.	Link	FC インタフェースで接続されたデバイス同士間で互いのデバイスを認識しあい、データ転送可能である状態を表すのに使用します。データ転送可能な状態のことを Linkup またはリンクアップ、デバイスや FC インタフェースの障害などにより、データ転送不能な状態のことを LinkDown またはリンクダウンとよびます。
20.	CRC エラー	Cyclic Redundancy Check エラーの略です。FC のトランザクションを構成する FC フレームはフレーム終端部に 32 ビットの誤り検出符号である CRC を持ちます。この CRC 部の値が正常な値でなければ当該 FC フレームは CRC エラーとされ、転送エラーが発生したと判定されます。
21.	【Linux】 【VMware】	OS により機能が異なる箇所をこの記号で示します。この記号がない記載については Linux と VMware で共通する説明です。
22.	Basic	仮想化プラットフォームを使用していない構成を指します。

目次

重要なお知らせ	2
規制・対策などについて	2
輸出規制について	2
登録商標・商標について	2
著作権について	2
はじめに	3
マークについて	3
オペレーティングシステム（OS）の略称について	4
来歴	5
用語の定義	6
目次	8
1 本書の構成	11
2 関連ドキュメント	12
HITACHI Fibre Channel アダプタ ユーザーズ・ガイド の構成	12
3 お使いになる前に	14
注意事項	14
サポートバージョン	15
サポートバージョン(障害閾値管理機能)	15
対象 OS	15
対象 HBA	16
対象 HBA（RV3000）	16
4 システムにおける I/O 障害によるリスク	17
I/O 障害によるシステムへの影響	17
障害がシステム全体へ影響を及ぼすケース	19
瞬断(短時間リンクダウン障害)が繰り返し発生したケース	19
FC インタフェース障害が繰り返し発生したケース	20
無応答障害が繰り返し発生したケース	21
5 高速系切替支援機能が提供する機能	22

障害閾値管理機能	22
監視対象の障害の定義	23
障害閾値及び監視時間の設定	25
障害の計測単位及び閉塞単位	27
HBA ポート閉塞機能	31
障害閾値管理機能の適用ケース	33
6 本機能使用上の注意事項	36

安全に関する注意事項は、下に示す見出しによって表示されます。これは安全注意シンボルと「警告」および「注意」という見出し語を組み合わせたものです。



これは、安全注意シンボルです。人への危害を引き起こす潜在的な危険に注意を喚起するために用います。起こりうる傷害または死を回避するために、このシンボルのあとに続く安全に関するメッセージにしたがってください。



これは、死亡または重大な傷害を引き起こすかもしれない潜在的な危険の存在を示すのに用います。



これは、軽度の傷害、あるいは中程度の傷害を引き起こすおそれのある潜在的な危険の存在を示すのに用います。

注意

これは、装置の重大な損傷*、または周囲の財物の損傷もしくはデータの喪失を引き起こすおそれのある潜在的な危険の存在を示すのに用います。

* 「装置の重大な損傷」とは、システム停止に至る装置の損傷をさします。



【表記例 1】感電注意

△の図記号は注意していただきたいことを示し、△の中に「感電注意」などの注意事項の絵が描かれています。



【表記例 2】分解禁止

○の図記号は行ってはいけないことを示し、○の中に「分解禁止」などの禁止事項の絵が描かれています。



【表記例 3】電源プラグをコンセントから抜け

●の図記号は行っていただきたいことを示し、●の中に「電源プラグをコンセントから抜け」などの強制事項の絵が描かれています。

安全に関する共通的な注意について

次に述べられている安全上の説明をよく読み、十分理解してください。

- 操作は、このマニュアル内の指示、手順に従って行ってください。
 - 装置やマニュアルに表示されている注意事項は必ず守ってください。
- これを怠ると、けが、火災や装置の破損を引き起こすおそれがあります。

操作や動作は

マニュアルに記載されている以外の操作や動作は行わないでください。

装置について何か問題がある場合は、電源を切り、電源プラグをコンセントから抜いたあと、お買い求め先にご連絡いただくか保守員をお呼びください。

自分自身でもご注意を

装置やマニュアルに表示されている注意事項は、十分検討されたものです。それでも、予測を超えた事態が起こることが考えられます。操作に当たっては、指示に従うだけでなく、常に自分自身でも注意するようにしてください。

1

本書の構成

本章では、本書の構成、及び内容について説明します。

表 1 本書の構成

章番号	章題	内容
1	本書の構成	本書の内容について説明します。
2	関連ドキュメント	本書の関連マニュアルについて説明します。
3	お使いになる前に	高速系切替支援機能を使用する前に知っておいていただきたい注意事項、適用範囲について説明します。
4	システムにおけるI/O障害によるリスク	I/O 障害が発生した場合におけるシステムへの影響について説明します。
5	高速系切替支援機能が提供する機能	4 章で説明した I/O 障害が与えるシステムへの影響を極力抑えることを目的とした「高速系切替支援機能」について説明します。
6	本機能使用上の注意事項	「高速系切替支援機能」を使用する際の注意点について説明します。

2

関連ドキュメント

この章では、本書の関連マニュアルについて説明します。

HITACHI Fibre Channel アダプタ ユーザーズ・ガイド の構成

32Gbps HITACHI Fibre Channel アダプタのユーザーズガイドは、「表 2 ドキュメント一覧」のとおり分冊されており、それぞれの内容は以下ようになります。

表 2 ドキュメント一覧

#	ドキュメント名称	内容
1.	HITACHI Fibre Channel アダプタ ユーザーズ・ガイド (VMwareドライバ編)	アダプタの VMware ドライバのインストール及びアップデート方法、エラーログ情報、及びドライバパラメータの一覧について記載しています。
2.	HITACHI Fibre Channel アダプタ ユーザーズ・ガイド (サポートマトリクス編)	ドライバの機能・OS のバージョンと、その機能をサポートしたドライババージョンの対応について説明しています。 更に、ファームウェア機能と、その機能をサポートしたファームウェアバージョンについても記載しています。
3.	HITACHI Fibre Channel アダプタ ユーザーズ・ガイド (VMwareユーティリティソフト編)	VMware ESXi 7 以降における HBA 設定ユーティリティである、CIM プロバイダ及び CIM クライアントのインストール方法や操作方法を説明しています。
4.	HITACHI Fibre Channel アダプタ ユーザーズ・ガイド (高速系切替支援機能編)	本書 高速系切替支援機能の説明、使用方法、注意事項などについて記載しています。
5.	HITACHI Fibre Channel アダプタ ユーザーズ・ガイド(EFI編)	アダプタパラメータ及び Storage Area Network (SAN) からブートするための設定の設定方法について記載しています。
6.	Hitachi Fibre Channel – Path Control Manager 2 Premium Edition for Linux Hitachi Fibre Channel – Path Control Manager 2 Enterprise Edition for Linux ユーザーズガイド	HFC-PCM-2 PE/EE の機能説明、インストール及びアップデート方法、ログ情報などについて記載しています。
7.	Hitachi Fibre Channel –Path Control Manager 2 Premium Edition for Linux Hitachi Fibre Channel –Path Control Manager 2 Enterprise Edition for Linux ユーザーズ・ガイド(ユーティリティソフト編)	HFC-PCM-2 PE/EE 用の HBA 設定ユーティリティのインストール方法や操作方法を説明しています。
8.	Hitachi Fibre Channel –Path Control Manager 2 Premium Edition for Linux	HFC-PCM-2 PE/EE のドライバの機能・OS のバージョンと、その機能をサポートしたドライババージョンの対応について説明しています。

#	ドキュメント名称	内容
	Hitachi Fibre Channel –Path Control Manager 2 Enterprise Edition for Linux ユーザーズ・ガイド(サポートマトリクス編)	更に、ファームウェア機能と、その機能をサポートしたファームウェアバージョンについても記載しています。

3

お使いになる前に

本章では、高速系切替支援機能を使用する前にご承知いただきたい内容について説明します。ご使用前にお読みください

注意事項

高速系切替支援機能を正しくご利用いただくためには、必ずお使いになる前に以下の注意事項をお読み頂き、本機能の採用可否を判断頂く必要があります。なお、高速系切替支援機能の採用を決定後、本機能を使用する場合、使用上の注意事項がございます。こちらの注意事項については、表 13 使用上の注意事項をお読みください。

(1) 高速系切替機能をご使用される場合【OS 共通】

- 本機能はマルチパス管理ソフトによるパス冗長化構成で使用して下さい。残っている有効なパスが1つである状態で本機能を使用されている場合、そのパスで障害閾値超過が発生すると当該 HBA ポートが閉塞し、OS がダウンする（サーバからストレージへのアクセス不可となる）可能性があります。また、高速系切替支援機能のパラメータ設定によっては、同一ストレージに対してパス冗長化構成で使用していても、一方のパスに影響があるストレージ側の故障に起因して、冗長パス側でも障害閾値超過が発生して当該 HBA ポートが閉塞し、OS がダウンする（サーバからストレージへのアクセス不可となる）可能性があります。本機能をご使用の際にはストレージ仕様をご確認頂いた上でパラメータ設計をお願いすると共に弊社サポートへご相談をお願い致します。
- 本機能をサポートしているアダプタのドライバ、対応ユーティリティ、HBA ファームウェアが本書の「サポートバージョン」に記載したバージョンになっているか確認してください。
- 仮想化プラットフォーム(VMware など)を使用した環境で、本機能により HBA ポートを閉塞した場合、物理 HBA ポートを閉塞(光断)します。当該 HBA ポートを共有している全ゲスト OS で当該ポートからディスク装置にアクセスできなくなります。ゲスト OS がマルチパス構成ならばパス交代が発生し、シングルパス構成ならばアプリケーションエラーが発生する可能性があります。

サポートバージョン

高速系切替支援機能をサポートするドライバ、ファームウェアバージョンについて説明します。

サポートバージョン(障害閾値管理機能)

高速系切替支援機能で提供している「障害閾値管理機能」を使用する場合、全てのドライバ、ファームウェアを下記表のバージョンにアップデートしてください。

表 3 サポートバージョン(障害閾値管理機能)

#	製品名	搭載装置	サポートバージョン
1.	Linux ドライバ	RV3000	RHEL8 以降：全バージョン
2.	VMware ドライバ (Native ドライバ)	RV3000	VMware ESXi 7.0：全バージョン

対象 OS

高速系切替支援機能のサポート OS については、「表 4 サポート OS 一覧」に記載します。

表 4 サポート OS 一覧

#	OS	プラットフォーム	サポート有無 ○：サポート ×：未サポート
1.	Red Hat Enterprise Linux 8	x86_64	○
2.	VMware ESXi 7.0	x86_64	○

対象 HBA

表 5 に高速系切替支援機能で提供している「障害閾値管理機能」をサポートする HBA の一覧を示します。閉塞に要する最大時間は、「障害閾値管理機能」が提供する「HBA ポート閉塞機能」で、アダプタドライバが HBA に対して閉塞指示を行ってから、実際に HBA ポートが閉塞(光断)されるまでの時間を示しています。

対象 HBA (RV3000)

表 5 対象 HBA 一覧 (RV3000)

#	製品名称	形名	サポート有無【OS 共通(*1)】 ○：サポート(*2) ×：未サポート	閉塞に要する最大時間
1.	Fibre Channel ボード (32Gbps、 2Port)	TD-***-R2E09A TDS***-R2E09A	○ (全バージョン)	2s

(*1)[Linux, VMWare] 共通

(*2) ()内は「障害閾値管理機能」をサポートする HBA ファームウェアバージョンを示します。

4

システムにおける I/O 障害によるリスク

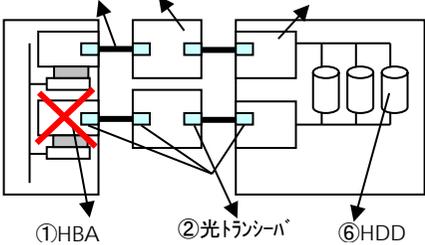
本章では I/O 障害が発生した場合におけるシステムへの影響を説明します。

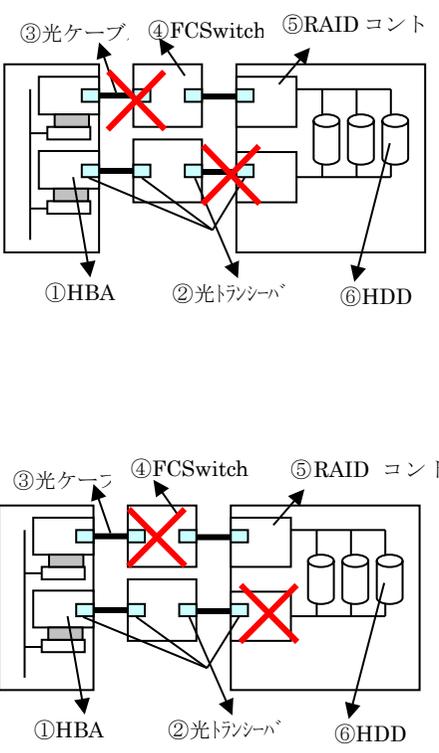
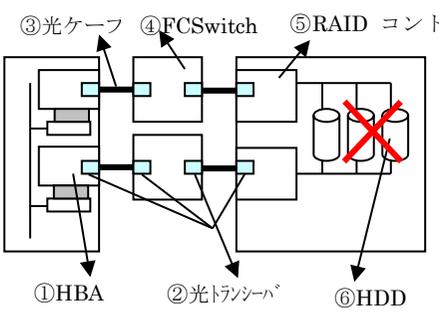
I/O 障害によるシステムへの影響

ディスク障害もしくはサーバとディスク装置を接続するインターフェース上に障害が発生した場合、特定コマンドの応答が返ってこない、プロセスを強制終了できない、システム全体がスローダウンするなどの予期せぬ挙動を示すことがあります。更に、ディスク周りのコンポーネントを冗長化しているにもかかわらず、片系障害の影響がシステム全体に及ぶケースがあります。

SCSI コマンド処理中に発生する障害は、リトライや冗長バスへの交代などによりシステムへの影響を少なくできます。しかし、一部の無応答障害や一時的な障害が繰返し発生した場合では、前述の「予期せぬ挙動」を引き起こす可能性があります。「表 6 障害部位により発生する I/O 障害の種類」に I/O 障害の種別毎に、システムへの影響についてまとめます。

表 6 障害部位により発生する I/O 障害の種類

No	障害部位	障害名	システムへの影響
1	 <p>①HBA ②光トランシーバ ⑥HDD</p>	HBA 障害	<p>■ケース 1 (障害割込み発生)</p> <p>HBA ハードウェアより、ハードウェア故障を示す割込みがドライバに発行されます。数秒の HBA ハードウェアのリカバリ処理の後、ディスクドライバまたはアダプタドライバのリトライにより SCSI コマンドが正常終了します。VMware の場合バス交代も発生しません。HFC-PCM-2 の場合はバス交代します。</p> <p>■ケース 2 (SCSI コマンドタイムアウト)</p> <p>SCSI コマンドのタイムアウトを検知後、HBA ハードウェアの診断を実施し HBA ハードウェアの故障を認識します。その後、数秒の HBA ハードウェアのリカバリ処理の後、ディスクドライバまたはアダプタドライバのリトライにより SCSI コマンドが正常終了し、バス交代も発生しません。</p>

2	 <p>③光ケーブル ④FCSwitch ⑤RAID コント</p> <p>①HBA ②光トランシーバ ⑥HDD</p>	<p>光ケーブル 又は 光トランシーバ 障害</p> <p>FCSwitch 又は RAID コ ントロー ラ障害</p> <p>■ケース 1 (LinkDown : 瞬断) 数秒で Link が回復する障害。この場合、ディスクドライブまたはアダプタドライブのリトライにより SCSI コマンドが正常終了し、パス交代も発生しません。</p> <p>■ケース 2 (LinkDown : 継続的な光断) 十数秒(デフォルトでは 30 秒前後)で、パス交代が発生します。SCSI コマンドを交代パスでリトライし、正常終了します。</p> <p>■ケース 3 (FC インタフェース障害) ディスクドライブまたはアダプタドライブのリトライにより SCSI コマンドが正常終了し、パス交代も発生しません。</p> <p>■ケース 4 (無応答障害 : 一時的) ディスクドライブまたはアダプタドライブのリトライにより SCSI コマンドが正常終了し、パス交代も発生しません。(SCSI コマンド監視時間×リトライ回数)後に正常終了します。 (*1)</p> <p>■ケース 5 (無応答障害 : 継続的) ディスクドライブまたはアダプタドライブのリトライにより SCSI コマンドが正常終了することがありません。リトライ数分の時間が経過してからパス交代が発生します。負荷状況、LU 数にも依存しますが、アプリケーションで数分~数十分待たされる場合があります。(*1) SCSI コマンドを交代パスでリトライし、正常終了します。</p>
3	 <p>③光ケーブル ④FCSwitch ⑤RAID コント</p> <p>①HBA ②光トランシーバ ⑥HDD</p>	<p>ディスク装置内部 障害</p> <p>■ケース 1 (ディスク装置内部で検知した障害) ディスク装置内部でリカバリされ、サーバ側は障害を検知しません。パス交代は発生します。</p> <p>■ケース 2 (SCSI レベル障害) ディスクドライブまたはアダプタドライブのリトライにより SCSI コマンドが正常終了し、パス交代も発生しません。</p> <p>■ケース 3 (無応答障害) No2、ケース 4、5 と同じです。</p>

(*1) アプリケーションで当該時間より短い時間で監視を行っている場合、アプリケーションでエラー検知します。

(*2) SCSI コマンド監視時間は、OS や各構成要素のバージョン、設定値などにより変わります。

障害がシステム全体へ影響を及ぼすケース

「表 6 障害部位により発生する I/O 障害の種類」に示す障害が発生した際のシステムへの影響を説明します。ここで記載するリトライ回数は OS やバス管理ソフトなど構成要素のバージョン/設定などによって変わる可能性があります。

瞬断(短時間リンクダウン障害)が繰り返し発生したケース

リンクダウンが発生し LinkDownTime(*1)以内にリンクアップする(瞬断)ケースを、短時間リンクダウンと呼びます。アダプタ光モジュール障害や、FC 経路の障害などで短時間リンクダウンが繰り返し発生するケースでは、バス交代が発生せず、SCSI コマンド発行の度に LinkDownTime 以内の遅延が発生し、システムがスローダウンする場合があります。(「図 1 瞬断が繰り返し発生したケース」)

(*1) LinkDownTime :

VMware をご使用の場合は「HITACHI Fibre Channel アダプタ ユーザーズガイド (VMware ユーティリティソフト編)」の「qlport_down_retry」を参照してください。

HFC-PCM-2 PE、HFC-PCM-2 EE をご使用の場合は「HFC-PCM-2 ユーザーズ・ガイド (ユーティリティソフト編)」の「Link Down Time」を参照してください。

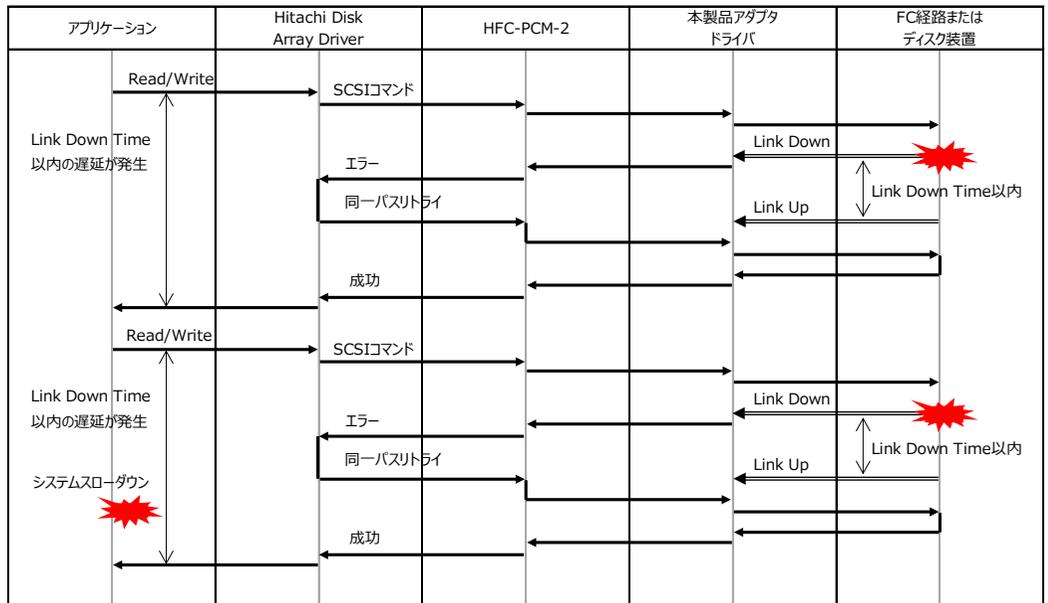


図 1 瞬断が繰り返し発生したケース

FC インタフェース障害が繰り返し発生したケース

FC ケーブル障害や、FC 経路上などの障害で FC インタフェース障害が一時的に発生すると、ディスクドライバによるリトライで救済されパス交代も発生しません。このため一時的な FC インタフェース障害が繰り返し発生するケースでも、パス交代が発生せず、SCSI コマンド発行の度に数秒(一般的に 1 秒以内)の遅延が発生し、システムがスローダウンする場合があります。(「図 2 FC インタフェース障害が繰り返し発生したケース」)

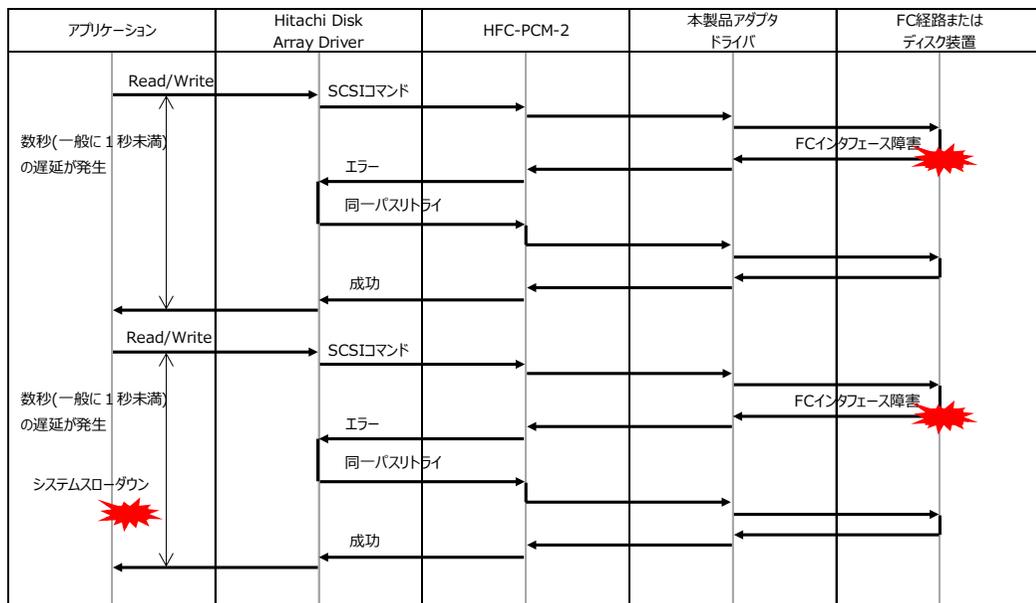


図 2 FC インタフェース障害が繰り返し発生したケース

無応答障害が繰り返し発生したケース

ディスクやFC経路の障害で、一時的な無応答障害が発生すると、SCSIコマンドタイムアウトが発生しますが、ディスクドライバによるリトライで救済されパス交代が発生しません。一時的な無応答障害が繰り返し発生するケースでは、SCSIコマンド発行の度にSCSIコマンド監視時間(*1)の遅延が発生し、パス交代が発生せず、システムがスローダウンする場合があります。(図3 無応答障害が繰り返し発生したケース)。

(*1) SCSIコマンド監視時間は、OSや各構成要素のバージョン、設定値などにより変わります。

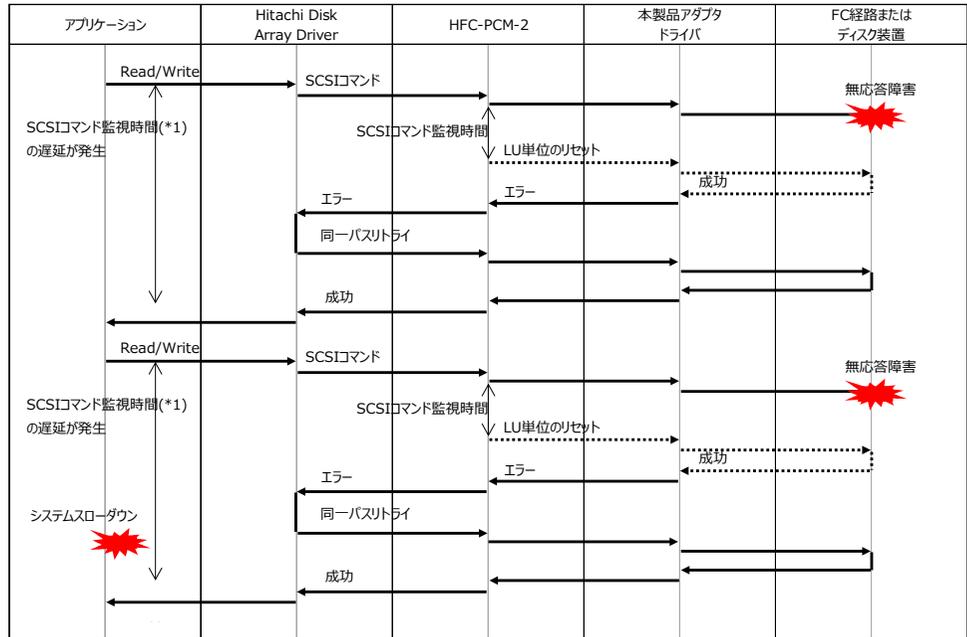


図3 無応答障害が繰り返し発生したケース

5

高速系切替支援機能が提供する機能

4章では、I/O 障害が与えるシステムへの影響について説明しました。本章では、I/O 障害が発生した場合のシステムへの影響を極力抑えることを目的とした「高速系切替支援機能」について説明します。

障害閾値管理機能

「システムにおける I/O 障害によるリスク」に説明した障害の発生回数を HBA ポート単位で監視し、閾値を超えて障害が発生した場合には、当該 HBA ポートを閉塞させる機能を提供します。閉塞した HBA ポートに対しては、ユーザが閉塞状態を解除、または OS をリポートしない限り、ドライバは SCSI コマンドを発行することはありません。

表 7 障害閾値管理機能一覧

#	機能	説明
1	障害の監視	障害毎に監視する閾値を設定します。障害発生回数が閾値を超過した場合（障害閾値超過）、当該 HBA ポートは閉塞（光断）されます。コマンドラインより障害監視を一時的に停止・再開することも可能性です。
2	HBA ポート閉塞機能	障害閾値を超えた場合及び、ユーザからの指示によって当該 HBA ポートを閉塞します。これにより、障害が頻発した場合でも、当該 HBA ポートを即座に閉塞し、バス交代時間又は HA 系切替え時間を短時間でいきます。コマンドにより閉塞状態から回復することも可能です。
3	タイムアウト障害発生時のチューニング機能	タイムアウト障害が発生した場合のリトライ回数や、タイムアウト障害後に実行されるリセット系コマンドの監視時間、リトライ回数を設定可能とします。これにより、タイムアウトした場合の最大処理時間を詳細に設計することが可能となります。

監視対象の障害の定義

障害閾値管理機能で監視が可能な障害を、「表 8 監視対象の障害の定義」に示します。

本機能では表に示す障害種毎の発生回数が、障害毎に設定する閾値を超過すると、その障害が発生した HBA ポートを閉塞状態にします。閾値については「障害閾値及び監視時間の設定」を参照下さい。

表 8 監視対象の障害の定義

No	障害名称	障害の定義	「表 6 障害部位により発生する I/O 障害の種類」の障害
1	HBA 障害	HBA HW 障害が発生。	No.1 ケース 1,2
2	短時間 LinkDown		
	HBA-ディスク装置間(*1) HBA-FC Switch 間(*1)	Link Down Time(*2)内に再度リンクアップしたケース。瞬断。 FC Switch 構成の場合、HBA-FC Switch 間で発生した障害を表す。	No.2 ケース 1
	FC Switch-ディスク装置間(*1) FC Switch-FC Switch 間(*1)	Link Down Time(*2)内に再度リンクアップしたケース。瞬断。 FC Switch 構成のみ発生し、FC Switch-ディスク装置間、又は FC Switch-FC Switch 間で発生した障害を表す。	上記と同様
3	FC インタフェース障害	SCSI コマンドの応答として報告される CRC エラーなど。	No.2 ケース 3
4	SCSI タイムアウト障害	SCSI コマンドのタイムアウト。(*3)	No.1 ケース 2 No.2 ケース 4,5 No.3 ケース 3
5	SCSI コマンドタイムアウト後のリセットコマンド障害	SCSI コマンドがタイムアウトした後のリセットコマンドが失敗。	上記と同様

(*1)LinkDown の発生箇所。詳細は、図 5～図 7 参照。

(*2)LinkDown を検知してからドライバが LinkDown 状態に遷移するまでの時間(デフォルト：VMware 10 秒、HFC-PCM-2 PE/HFC-PCM-2 EE 15 秒)です。本時間内は、LinkUp を待ち続けます。

(*3)本障害の対象は、ドライバにより以下の通りとなります。
リセットコマンド以外の SCSI コマンドのタイムアウト障害

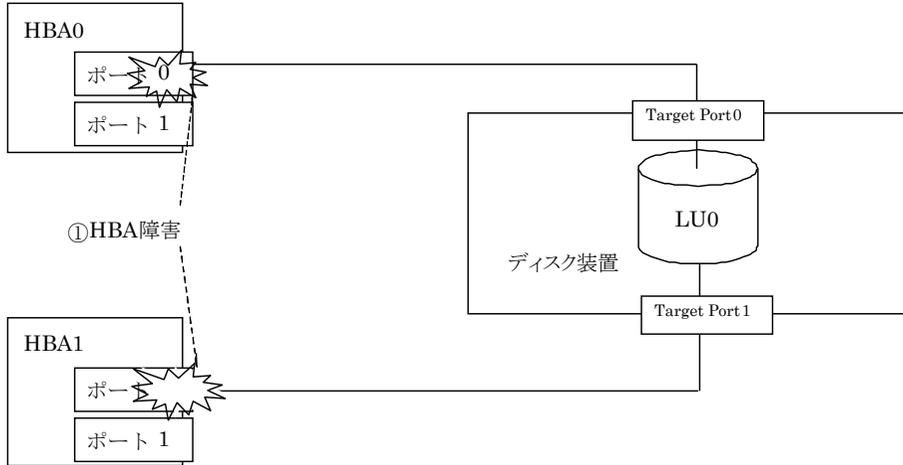


図 4 監視対象の HBA 障害

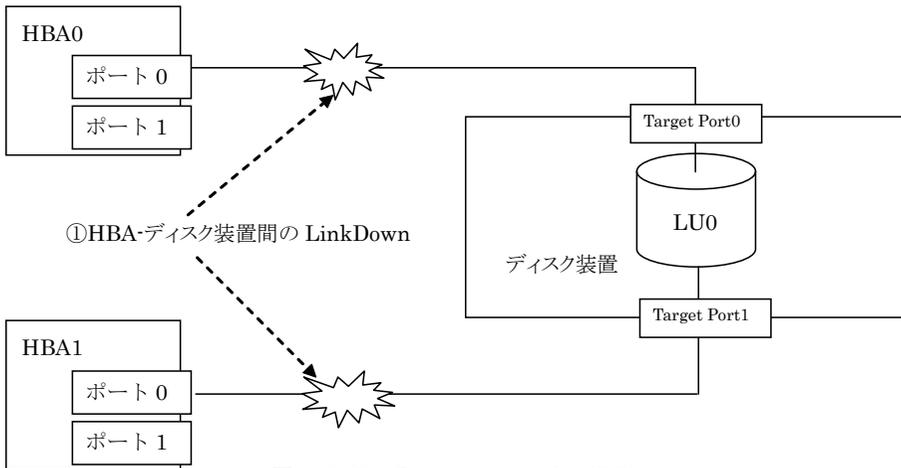


図 5 監視対象の LinkDown(直結構成)

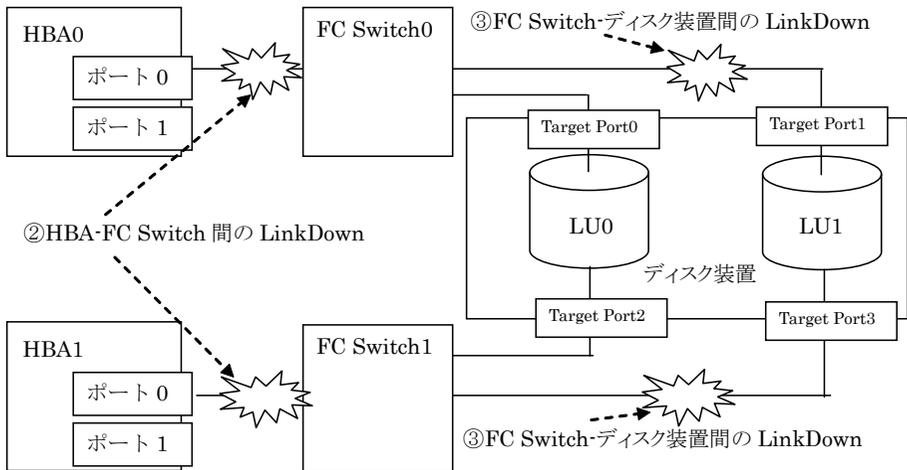


図 6 監視対象の LinkDown(FC Switch 構成)

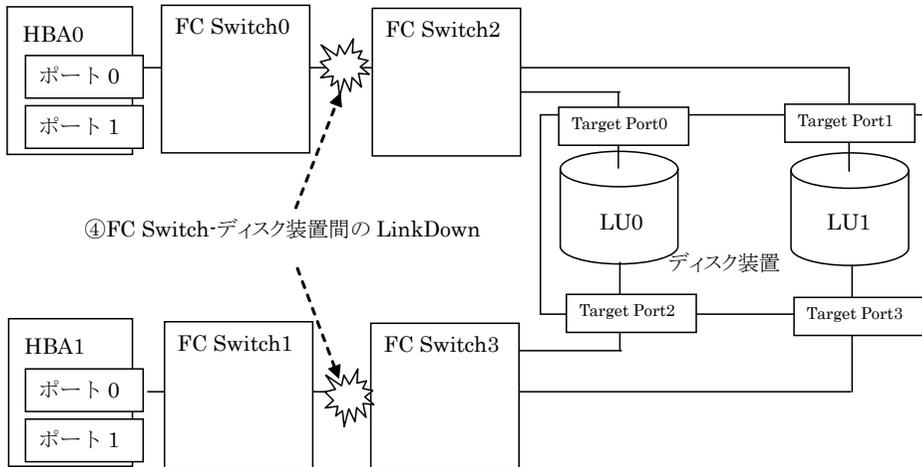


図 7 監視対象の LinkDown(FC Switch 構成(カスケード接続))

障害閾値及び監視時間の設定

障害閾値管理機能では、「表 9 障害閾値の設定項目と設定範囲」に示す設定項目をコマンドラインより設定します。設定、及び表示のコマンドについては、VMware をご使用の場合、「Hitachi Fibre Channel アダプタ ユーザーズガイド(VMware ユーティリティソフト編)」を参照してください。HFC-PCM-2 PE、HFC-PCM-2 EE をご使用の場合、「HFC-PCM-2 ユーザーズガイド(ユーティリティソフト編)」を参照して下さい。設定項目のうち、HFC-PCM-2 PE、HFC-PCM-2 EE は障害の監視時間を設定できませんが、VMware ドライバは OS 起動後から OS シャットダウン・リブートまでの障害発生回数の通算になります。

表 9 障害閾値の設定項目と設定範囲

No	設定項目	初期値	設定範囲	備考
1	HBA 障害			
	①HBA 障害閉塞閾値	0 (回) (閾値管理無し)	0~10 (回)	閉塞状態に遷移する HBA 障害閾値。"0"を設定した場合は障害の監視を行いません。(*4)
2	LinkDown 障害設定(*1)(*2)			
	①障害監視時間 【HFC-PCM-2】	30 (分)	1~60 (分)	LinkDown 障害の回数を計測する時間。本時間を超過する度に計測した障害発生回数をクリアします。
	②LinkDown 障害閉塞閾値 (短時間 LinkDown)	0 (回) (閾値管理無し)	0~30 (回)	閉塞状態に遷移する LinkDown 障害閾値。"0"を設定した場合は障害の監視を行いません。(*4)
3	FC インタフェース障害設定			
	①障害監視時間 【HFC-PCM-2】	30 (分)	1~60 (分)	FC インタフェース障害の回数を計測する時間。本時間を超過する度に計測した障害発生回数をクリアします。
	②FC インタフェース障害閉塞閾値	0 (回) (閾値管理無し)	0~2048 (回)	FC インタフェース障害閾値。"0"を設定した場合は障害の監視を行いません。
4	SCSI コマンドタイムアウト障害設定			

	①障害監視時間 【HFC-PCM-2】	30 (分)	1~60 (分)	SCSI コマンドタイムアウト障害の回数を計測する時間。本時間を超過する度に計測した障害発生回数をクリアします。
	②SCSI コマンドタイムアウト障害閉塞閾値	0 (回) (閾値管理無し)	0~2048 (回)	SCSI コマンドタイムアウト障害閾値。"0"を設定した障害の監視を行いません。(*3)
5	SCSI コマンドタイムアウト後のリセットコマンド障害監視(*5)			
	①SCSI コマンドタイムアウト後のリセット障害監視	0 (回) (閾値管理無し)	0~2048 (回)	SCSI コマンドタイムアウト後のリセットコマンドが失敗の障害閾値。"0"を設定した障害の監視を行いません。(*4)

(*1)LinkDown 検出後、LinkUp を待ち続ける時間 Link Down Time(デフォルト：VMware 10 秒、HFC-PCM-2 PE/HFC-PCM-2 EE 15 秒)以内にリンクアップした場合をカウントします。Link Down Time の設定に関しては、VMware をご使用の場合は「HITACHI Fibre Channel アダプタ ユーザーズガイド (VMware ユーティリティソフト編)」の「qlport_down_retry」、HFC-PCM-2 PE、HFC-PCM-2 EE をご使用の場合は「HFC-PCM-2 ユーザーズ・ガイド (ユーティリティソフト編)」を参照して下さい。

ストレージ側の故障によって、2 秒以下の LinkDown が発生する場合があります、Link Down Time を変更する場合、3 秒以上に設定してください。

(*2) LinkDown 障害閾値、及び監視時間は、HBA-ディスク装置間、HBA-FC Switch 間と FC Switch-ディスク装置間、FCSW-FC Switch 間で共通です。

(*3) VMware では、タイムアウト監視時間が非常に短い SCSI コマンドがあります。タイムアウト監視時間が 30 秒以上の SCSI コマンドタイムアウトを監視します。

(*4) 障害閾値を指定した場合、同一ストレージに対してパス冗長化構成で使用していても、一方のパスに影響があるストレージ側の故障に起因して、冗長パス側でも障害閾値超過が発生して当該 HBA ポートが閉塞し、OS がダウンする (サーバからストレージへのアクセス不可となる) 可能性があります。本パラメータを設定する際にはストレージ仕様をご確認頂いた上でパラメータ設計をお願いすると共に弊社サポートへご相談をお願い致します。

(*5) VMware では、SCSI コマンドタイムアウト後のリセット障害監視は非サポートです。HFC-PCM-2 では、SCSI コマンドタイムアウト後のリセットコマンドが失敗した場合はパス交代しますが、閾値を設定しない場合障害の発生したパスに対してリセットをエスカレーションして回復を試みます。回復した場合、該当パスは使用可能状態となります。リセット障害閾値を指定した場合は、障害閾値超過した時点で該当 HBA ポートを閉塞します。

障害の計測単位及び閉塞単位

障害閾値管理機能では、「表 10 障害の計測単位及び閉塞単位」に示す「計測単位」で障害発生回数をカウントし、閾値超過時には「閉塞単位」に示す単位で HBA ポートを閉塞します。短時間 LinkDown では、HFC-PCM-2 では HBA ポート単位に障害発生回数を管理し、VMware では HBA ポート単位及び Target ポート単位に個別に障害発生回数を管理し、LinkDown 発生時に「図 9 LinkDown 障害カウンタの管理単位(VMware HBA-I/O 間の LinkDown 発生時)」、「図 10 LinkDown 障害カウンタの管理単位(VMware FC Switch-HBA 間、FC Switch-I/O 間の LinkDown 発生時)」に示すようにカウントアップします。閾値は HBA ポート単位、Target ポート単位で共通の値です(表 9 障害閾値の設定項目と設定範囲(*2)参照)。短時間 LinkDown では HBA ポート単位の障害のカウントと Target ポート単位の障害のカウントのいずれかが障害閾値を超過した場合、ドライバは HBA ポートを閉塞します。計測した障害のカウントのクリア契機については「表 11 障害カウンタのクリア契機」に示します。

表 10 障害の計測単位及び閉塞単位

#	障害	計測単位		閉塞単位
		HFC-PCM-2	VMware	
1	HBA 障害	HBA ポート	HBA ポート	HBA ポート
2	短時間 LinkDown			
	HBA- ディスク装置間、 HBA-FC Switch 間 FC Switch- ディスク装置間 FC Switch-FC Switch 間	HBA ポート	HBA ポート Target ポート(*1)	HBA ポート
3	FC インタフェース障害	HBA ポート	HBA ポート	HBA ポート
4	SCSI コマンド T.O 障害	HBA ポート	HBA ポート	HBA ポート
5	リセットコマンド障害	HBA ポート	- (非サポート)	HBA ポート

(*1)Target ポート単位に発生する障害でも HBA ポートを閉塞することにより、システムから障害の要因と考えられる部位(HBA、FC Switch、Target ポート)を即座に取り除き、早期にパス交代することを目的としています。

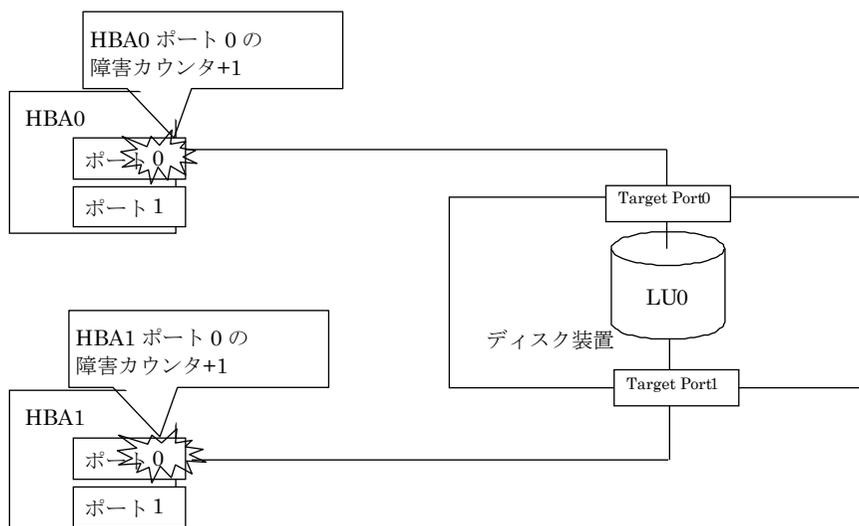


図 8 監視対象の HBA 障害

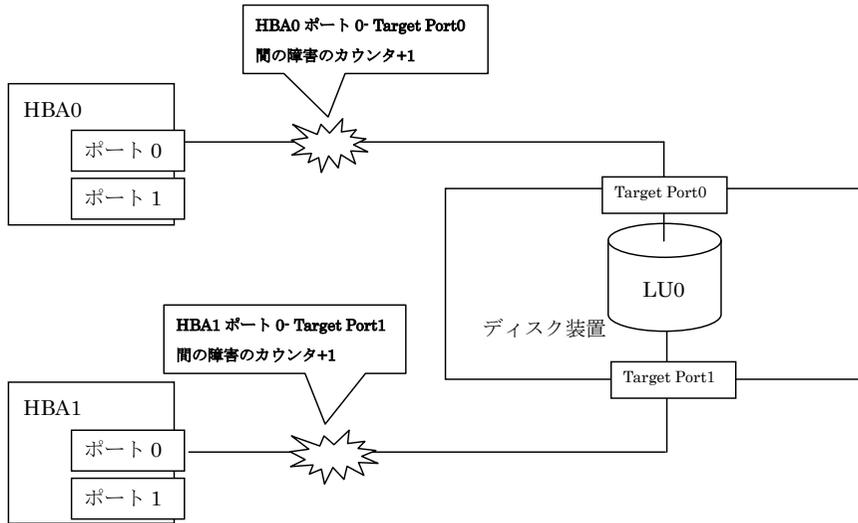


図 9 LinkDown 障害カウンタの管理単位(VMware HBA-I/O 間の LinkDown 発生時)

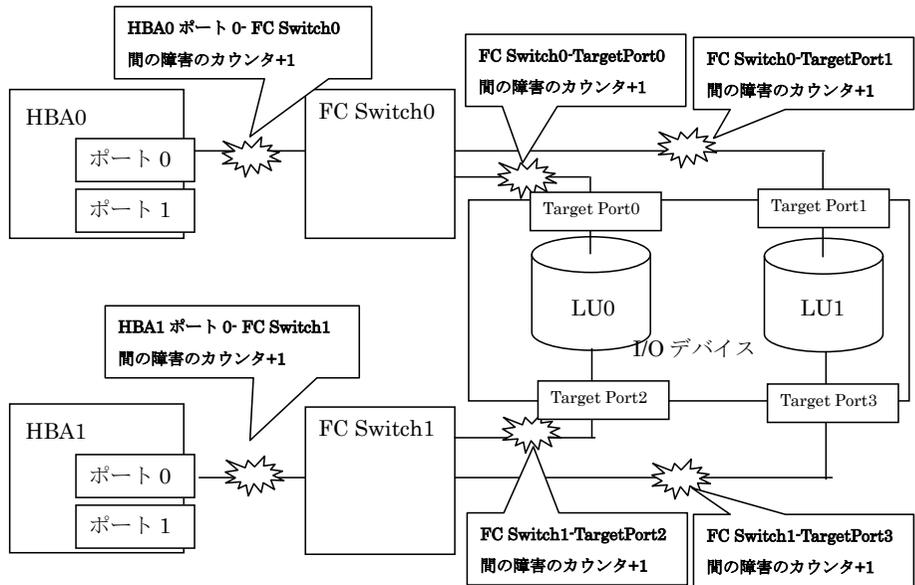


図 10 LinkDown 障害カウンタの管理単位(VMware FC Switch-HBA 間、FC Switch-I/O 間の LinkDown 発生時)

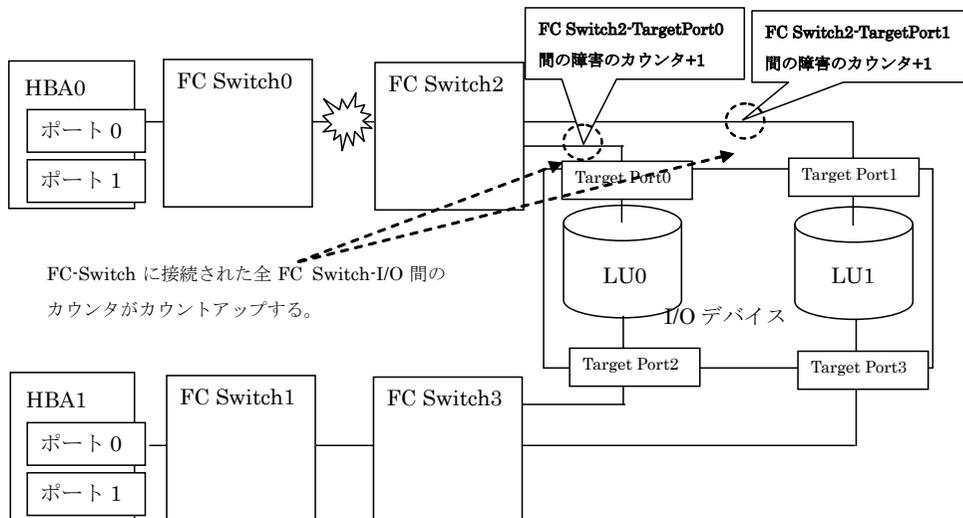


図 11 LinkDown 障害カウンタの管理単位(VMware FC Switch-FC Switch 間の LinkDown 発生時)

障害カウンタをクリアするコマンドはありません。障害カウンタをクリアする契機を表 11 に示します。No.1 で示す障害閾値管理機能を一旦停止後し、再度開始すれば、カウンタをクリアします。

障害閾値管理機能を停止するコマンド、再度再開するコマンドは以下です。(HFC-PCM-2 の例)

```
hfcmgr2_mp -is {on|off} force
```

on: 障害閾値管理機能 ON (default) (開始)

off: 障害閾値管理機能 OFF (停止)

force : 確認メッセージ無しで実行する。

表 11 障害カウンタのクリア契機

No	障害カウンタクリア契機	クリア対象の障害カウンタ
1	障害閾値管理機能を停止後再度開始	全 HBA ポート、全 Target ポート
2	HBA ポート閉塞解除	閉塞解除で指定した HBA ポート、指定した HBA ポートと接続している Target ポート
3	OS リブート	全 HBA ポート、全 Target ポート

HBA ポート閉塞機能

障害閾値管理機能により障害閾値を超過した場合にアダプタドライバが HBA ポートを閉塞します。また、コマンドラインから HBA ポートを閉塞、閉塞解除することも可能です。閉塞させることで、障害発生頻度の高い HBA ポートへの SCSI コマンドを遮断し、早期のパス交代又は HA 系切替えを促します(ただし、HA 系切替えを行うには、HA クラスタ構成など HA 系切替えの機能をシステムへ作り込む必要があります)。また、閉塞状態からの回復は、ユーザからの指示または OS リブートによって回復可能です。(OS リブートにより閉塞状態が解除されないケースがあります。詳細は「表 12 HBA ポート閉塞機能一覧」を参照して下さい。)

「図 12 障害による HBA ポート閉塞」のようなパスが冗長化されている構成で、HBA0 ポート 0 から LU0 への SCSI コマンドにおいて継続的な無応答障害が発生した場合、LU0 への SCSI コマンドの応答に遅延が発生します。

さらに、その無応答障害の要因が HBA0 ポート 0 にある場合、LU1 への SCSI コマンドも同様に無応答障害が発生する可能性があります。

閉塞機能を用いて、障害のある HBA ポートを閉塞すると、HBA0 ポート 0 と同一の物理ディスクに接続された HBA1 ポート 0 から SCSI コマンドを発行するようにマルチパス管理ソフトウェアが経路を変更します。

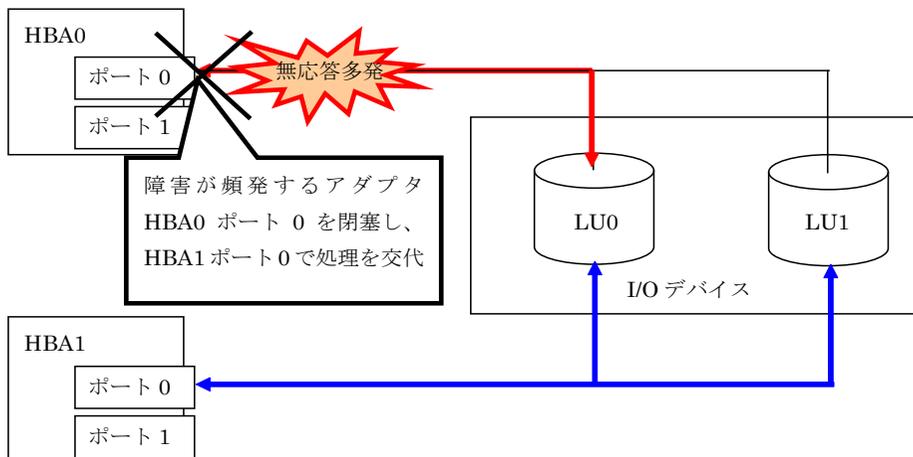


図 12 障害による HBA ポート閉塞

HBA ポート閉塞機能がサポートする機能一覧を「表 12 HBA ポート閉塞機能一覧」に示します。HBA ポート閉塞機能により、閉塞/閉塞解除を行った場合、エラーログ/イベントログが採取されません。

表 12 HBA ポート閉塞機能一覧

No	機能	内容	閉塞契機	閉塞解除契機	エラーログ
1	障害による HBA ポート閉塞	「障害閾値管理機能」と連動し、障害発生回数が閾値を超過したら、HBA ポートを自動的に閉塞する機能です。	障害発生回数が閾値を超過	①リポート ②ユーザによる閉塞解除指示	HFC2_ISOLERR
2	HBA ポート強制閉塞	ユーザがコマンドインタフェースを用いて、指定した HBA ポートを閉塞状態にする機能です。	ユーザによる HBA ポート強制閉塞指示	①リポート(*1) ②ユーザによる閉塞解除指示	HFC2_ISOLUSR
3	閉塞状態の解除	ユーザがコマンドインタフェースを用いて閉塞状態の解除を行う機能です。			HFC2_ISOLRCV

(*1) OS リポート後も閉塞状態を維持する機能(「HITACHI Fibre Channel アダプタ ユーザーズガイド (VMware ユーティリティソフト編)」、「HFC-PCM-2 ユーザーズ・ガイド (ユーティリティソフト編)」の「障害閾値管理機能(動作状態の確認)」を参照)が動作するため、閉塞解除コマンドを実行するまでは、OS リポートしても閉塞解除されません。

(1) HBA ポート閉塞に要する時間

HBA ポートを閉塞するまでの時間とは、アダプタドライバが HBA に対して閉塞指示を行ってから、実際に HBA ポートが閉塞(光断)されるまでの時間になります。

(2) OS リポート後も閉塞状態を維持する機能

下記手順を実施することで、OS リポート後も閉塞状態を維持することが可能です。

VMware の場合

hfcvmutil2 で、HBA ポートに対して「障害閾値管理機能」を「save」に設定すると、ブート時にドライバが当該ポートを閉塞した状態で立ち上げます。障害閾値管理機能で閉塞させたポートに対して、「障害閾値管理機能」を「save」に設定することにより、OS リポート後も閉塞を維持し、リポート後の障害再発を防止することが可能です。パラメータの設定方法については、「HITACHI Fibre Channel アダプタ ユーザーズガイド (VMware ユーティリティソフト編)」の「障害閾値管理機能(動作状態の確認)」をご参照ください。

HFC-PCM-2 の場合

RAMDISK イメージを更新して下さい。RAMDISK イメージ更新方法は「HFC-PCM-2 ユーザーズ・ガイド (ユーティリティソフト編)」の「RAMDISK イメージの更新手順」をご参照ください。

ブートパスの HBA ポートに対して閉塞状態維持に設定した場合、OS 起動に失敗する可能性があります。ブートパスには本機能を設定しないでください。

障害閾値管理機能の適用ケース

「システムにおける I/O 障害によるリスク」で説明した I/O 障害によりシステムがスローダウンするケースに対して、障害閾値管理機能を適用した場合の効果を説明します。ここで記載するリトライの発行元やリトライ回数は OS やバス管理ソフトなど構成要素のバージョン/設定などによって変わる可能性があります。

(1) HBA 障害閾値設定

Linux の場合 HBA 障害閾値を設定しなくても初回にパス交代します。VMware について説明します。障害閾値管理機能の HBA 障害の監視閾値を 2 に設定した場合、HBA 障害が 2 回発生した時点で、パス交代が発生し、HBA[障害によるシステムのスローダウンを回避できます。

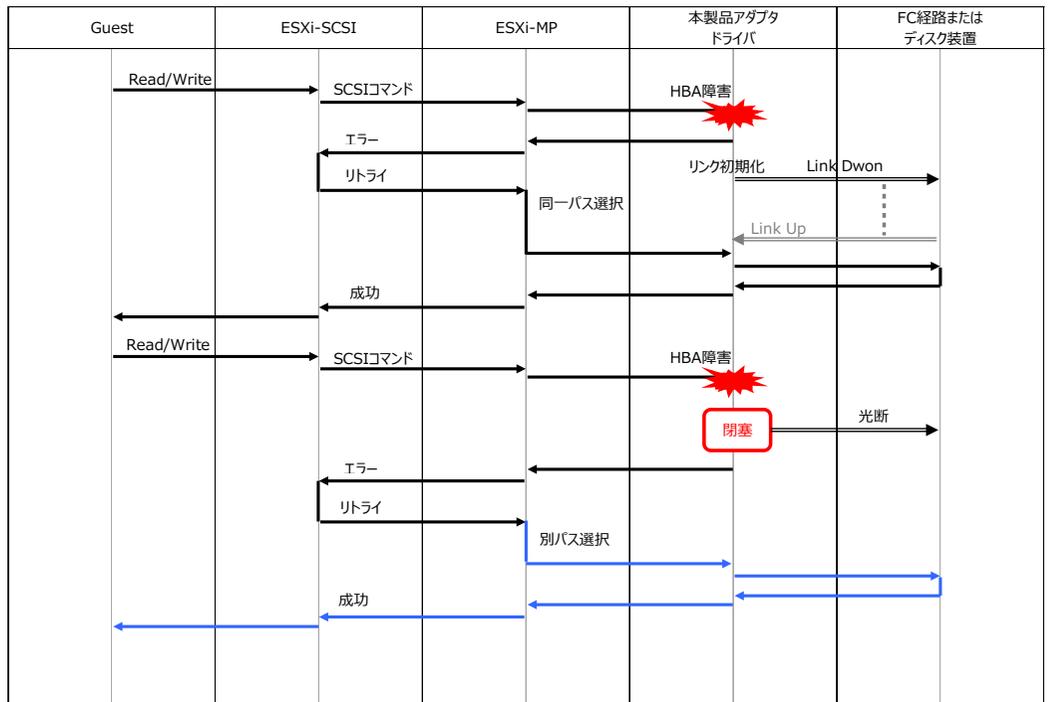


図 13 HBA 障害の閾値設定

(2) 短時間リンクダウン障害閾値設定

短時間リンクダウンを繰り返すケース(「瞬断(短時間リンクダウン障害)が繰り返し発生したケース」参照)で、障害閾値管理機能の短時間リンクダウン障害の監視閾値を2に設定した場合、短時間リンクダウンが2回発生した時点で、バス交代が発生し、リンクダウン障害によるシステムのスローダウンを回避できます。

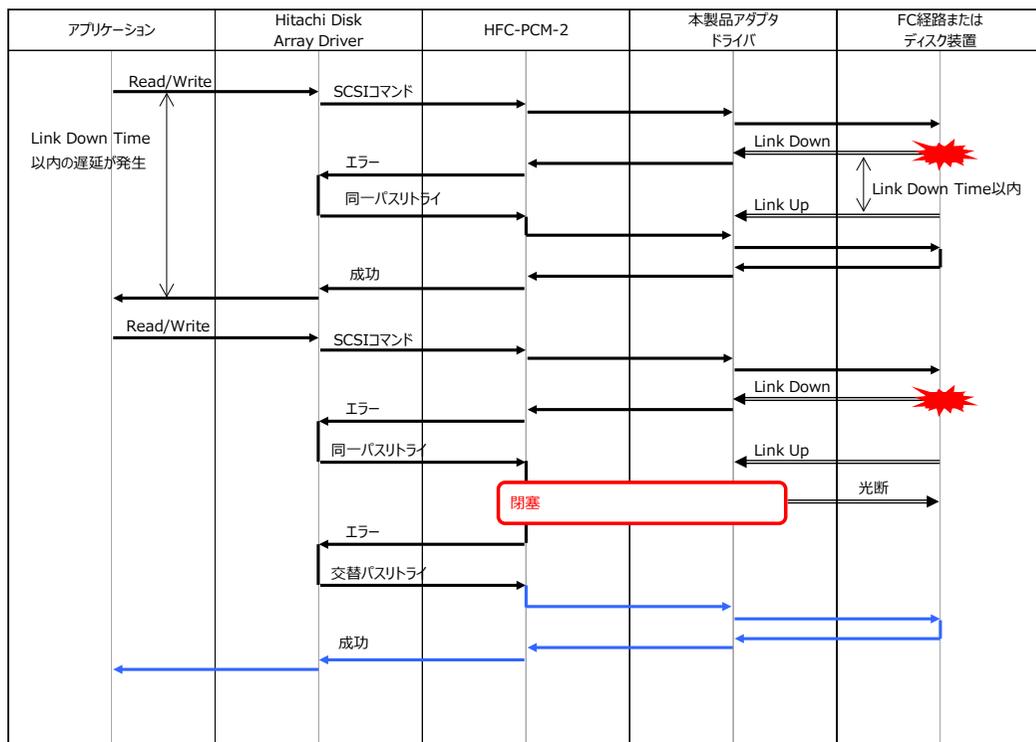


図 14 短時間リンクダウン障害の閾値設定

(3) FC インタフェース障害閾値設定

一時的な FC インタフェース障害を繰り返すケース(「FC インタフェース障害が繰り返し発生したケース」参照)で障害閾値管理機能の FC インタフェース障害の監視閾値を 1 に設定した場合、FC インタフェース障害が 1 回発生した時点で、パス交代が発生しシステムのスローダウンを回避できます。

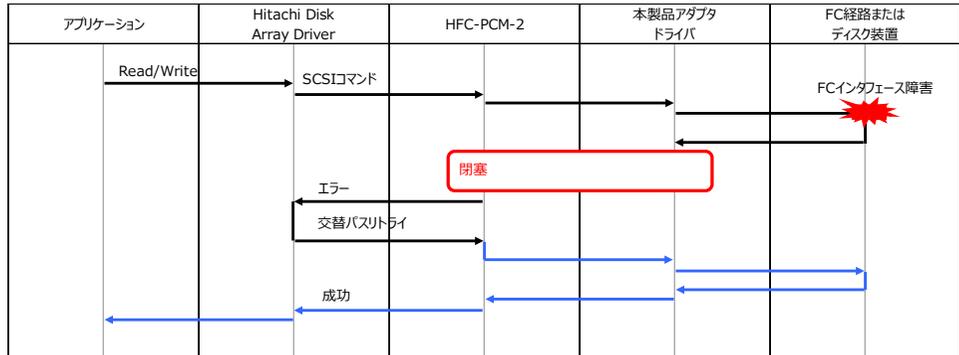


図 15 FC インタフェース障害の閾値設定

(4) タイムアウト障害閾値設定

無応答障害が繰り返し発生するケース(「無応答障害が繰り返し発生したケース」参照)で、障害閾値管理機能のタイムアウト障害の監視閾値を 1 に設定した場合、タイムアウト障害が 1 回発生した時点で、パス交代が発生しシステムのスローダウンを回避できます。

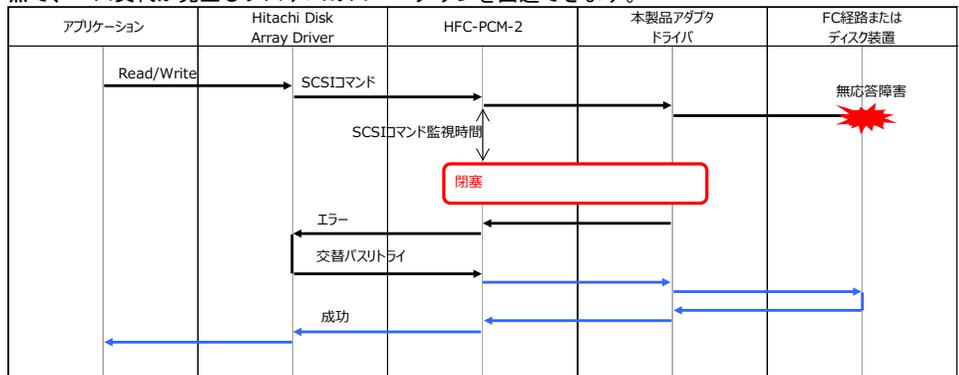


図 16 タイムアウト障害の閾値設定

6

本機能使用上の注意事項

本機能を使用する際には以下の点にご注意下さい。なお、高速系切替支援機能を正しくご利用いただくためには、事前に 3 章 お使いになる前に — 注意事項をお読み頂き、本機能の採用可否を判断頂く必要があります。

表 13 使用上の注意事項

#	対象	注意事項
1.	VMware	hfcmgr2 で障害閾値管理機能の設定する場合に、HBA ポートに対して閉塞状態を不揮発に設定すると、ブート時にドライバが当該ポートを閉塞した状態で立ち上げます。ブートパスの HBA ポートに対して閉塞状態を不揮発に設定した場合、OS 起動に失敗する可能性があります。ブートパスには閉塞状態を揮発に設定してください。
2.	HFC-PCM-2	コマンドにより閉塞した HBA ポートに対して閉塞解除コマンドを実施した場合、HBA ポートの状態は回復しますが、HBA ポートに接続された LU パスの状態は閉塞(offline(C))状態を維持します。パスの状態を回復する場合は、LU パスの状態をオンラインにしてください。コマンドについては、HFC-PCM-2 ユーザーズガイド(ユーティリティソフト編)「4.hfcmgr2_mp コマンド」の「LU パス状態の表示/変更/追加/削除」を参照してください。
3.	HFC-PCM-2	OS 起動時に強制閉塞状態で立ち上がった HBA ポートに対して、閉塞解除コマンドを実施したときに、接続されている LU を認識するためには LU 追加コマンドが必要になります。OS 稼働中の LU 追加については Hitachi Disk Array Driver for Linux / Hitachi Disk Array Driver Mirroring Edition for Linux のユーザーズガイドを参照してください。

HITACHI Fibre Channel アダプタ
ユーザーズガイド
(高速系切替支援機能編)
第5版

2024年5月

無断転載を禁止します。

 株式会社 日立製作所

〒100-8280 東京都千代田区丸の内一丁目6番6号

<http://www.hitachi.co.jp>