613-001736 Rev.A 120817

マルチレイヤー・モジュラー・スイッチ

SwitchBlade® x8100

テクニカルドキュメント

Allied Telesis

はじめに

このたびは、SwitchBlade x8100 をお買いあげいただき、誠にありがとうございます。

SwitchBlade x8100 は、高さ 7U の筐体に 10 個のラインカード用スロットと 2 個のコントロールファブリックカード用スロットを装備したマルチレイヤー・モジュラー・スイッチです。

本ドキュメントでは、シャーシ内部でラインカードとコントロールファブリックカードがどのように接続されているか、通常のトラフィックとシャーシ内部の制御トラフィックがどのように扱われるか、各種情報やファームウェアがどのように同期されるかなど、SwitchBlade x8100 の設計・内部構造についての概略を説明しています。SwitchBlade x8100 を活用するための参考資料としてご利用ください。

なお、本ドキュメントは SwitchBlade x8100 の取扱方法や設定手順を述べるものではありません。SwitchBlade x8100 のご使用にあたっては、かならず下記のマニュアルをご参照ください。

- 機器の設置・接続・取り外しなどに関する具体的な手順や注意事項については、「取扱説明書」をご覧ください。
- ファームウェアの制限事項やマニュアルの誤記訂正・補足事項などについては、「リリースノート」をご覧ください。
- 各種機能や設定コマンドの詳細については、「コマンドリファレンス」をご覧ください。

1　バックプレーン

SwitchBlade x8100 のバックプレーンは、「データプレーン」と「コントロールプレーン」という 2 つの独立した系統に分かれています。

- データプレーン

  シャーシ外部とやりとりされる通常のトラフィックを運ぶためのバックプレーン

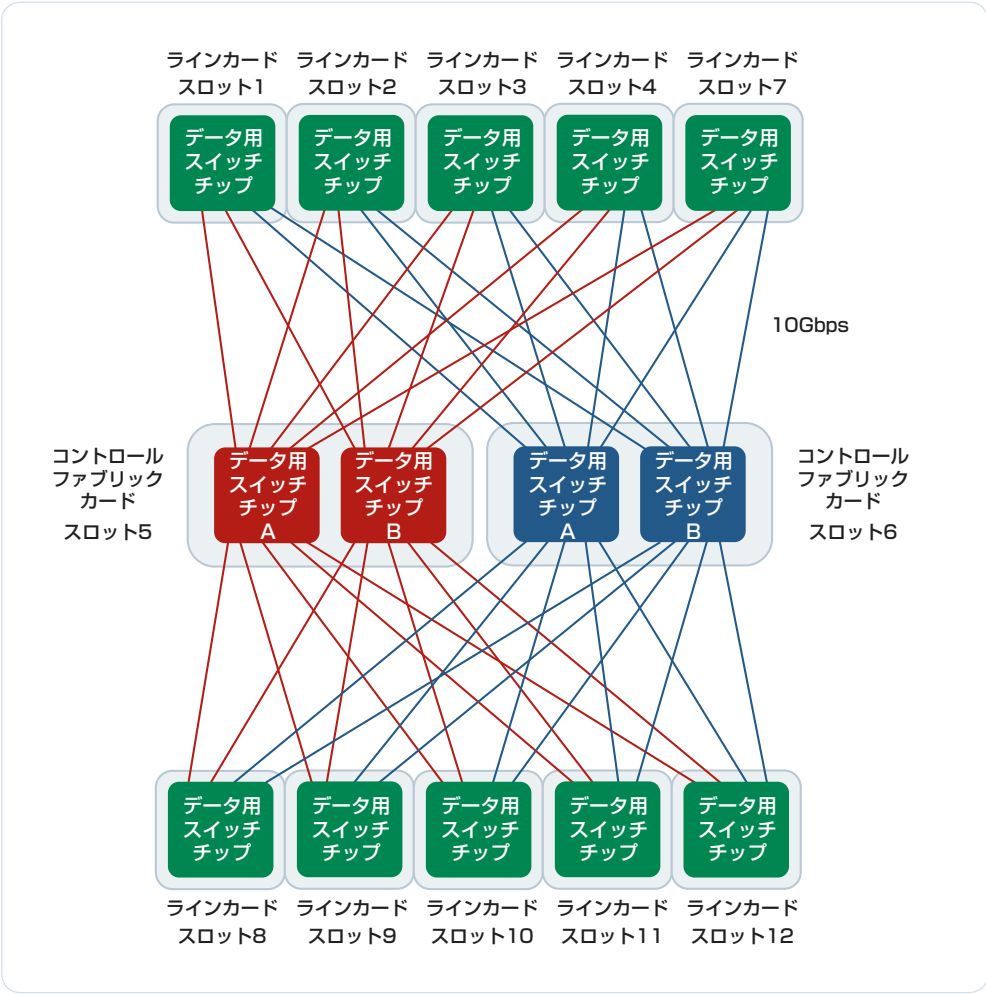
- コントロールプレーン

  シャーシ内部（コントロールファブリックカード・ラインカード間）の制御トラフィックを運ぶためのバックプレーン

1.1　データプレーン

データプレーンは、シャーシの外部とやりとりされるトラフィック、すなわち、通常のデータトラフィックと各種プロトコルの制御パケットを運ぶためのバックプレーンです。

下図に示すように、データプレーンは各ラインカードのスイッチチップと、コントロールファブリックカードに備えられた 2 個のスイッチチップを、それぞれ 10Gbps のリンクで接続した構成になっています。



前記の接続により、ラインカードスロット 1 スロットあたりのデータプレーン帯域は、コントロールファブリックカード 1 台装着時が 20Gbps、2 台装着時は 40Gbps となります。

また、ここから、各種ラインカードから他のラインカードへの通信性能は下記のとおりとなります。

ラインカード	コントロールファブリックカード	
	1 台装着時	2 台装着時
AT-SBx81GT24 （10/100/1000BASE-T ポート× 24）	ブロッキング （6:5）	ノンブロッキング
AT-SBx81GP24 （10/100/1000BASE-T PoE ポート× 24）	ブロッキング （6:5）	ノンブロッキング
AT-SBx81GS24a （SFP スロット× 24）	ブロッキング （6:5）	ノンブロッキング
AT-SBx81XS6 （SFP+ スロット× 6）	ブロッキング （3:1）	ブロッキング （3:2）

1.1.1 負荷分散

ラインカードとコントロールファブリックカードの間は、コントロールファブリックカードの台数によって 2 本または 4 本の 10Gbps リンクで結ばれます。

これらの 10Gbps リンクは、全体として 1 つの仮想的な内部ポート（リンクアグリゲーショングループ）と見なされ、そこを通るデータトラフィックはハッシュアルゴリズムによっていずれかのメンバーポート（10Gbps リンク）に負荷分散されます。負荷分散は、パケットの送信元・宛先 MAC アドレスと始点・終点 IP アドレスにもとづいて行われます。

1.1.2 パケットの転送経路

SwitchBlade x8100 では、各ラインカードにスイッチチップが搭載されています。

データプレーン上を流れるパケットの転送判断に必要なテーブル（FDB、ARP テーブル、IP ルートデータベース）は、すべてのコントロールファブリックカードとラインカードで同期されているため、どのラインカードでパケットを受信したとしても、各ラインカードが参照するテーブルの内容は同一です。

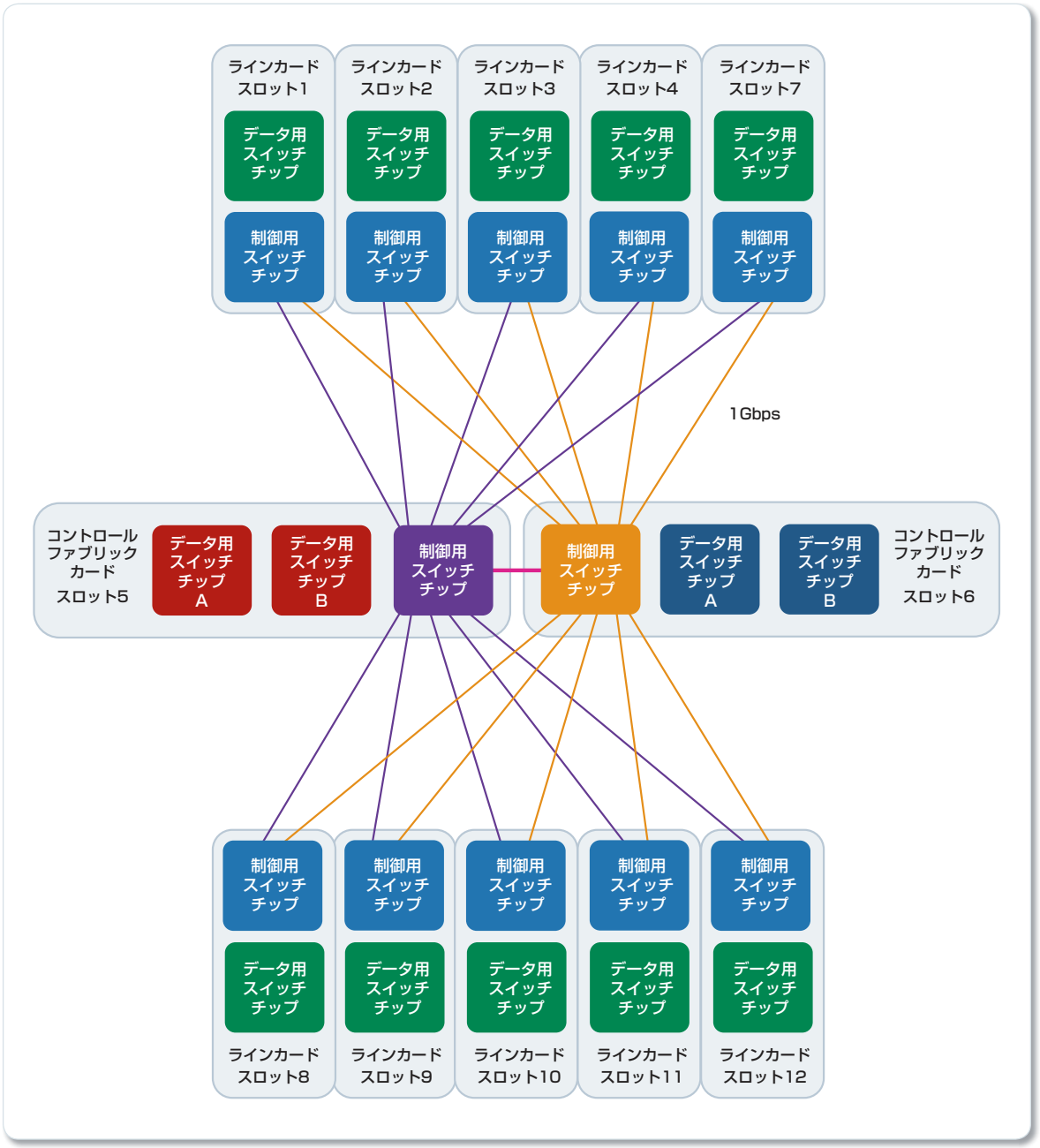
パケットの入力ポートと出力ポートが同じラインカード上にある場合、該当パケットの転送処理はそのラインカード上のスイッチチップで完結します。

一方、パケットの入力ポートと出力ポートが異なるラインカード上に存在する場合、該当パケットはコントロールファブリックカード経由で出力ポートのあるラインカードに転送され、送出されます。

1.2　コントロールプレーン

コントロールプレーンは、シャーシ内部の制御トラフィックを運ぶためのバックプレーンです。各種情報の同期やファームウェアイメージファイルの転送など、コントロールファブリックカード・ラインカード間の通信は、すべてコントロールプレーン経由で行われます。

下図に示すように、コントロールプレーンは、コントロールファブリックカードとラインカードに備えられた制御用スイッチチップを 1Gbps のリンクで相互接続した構成になっています。

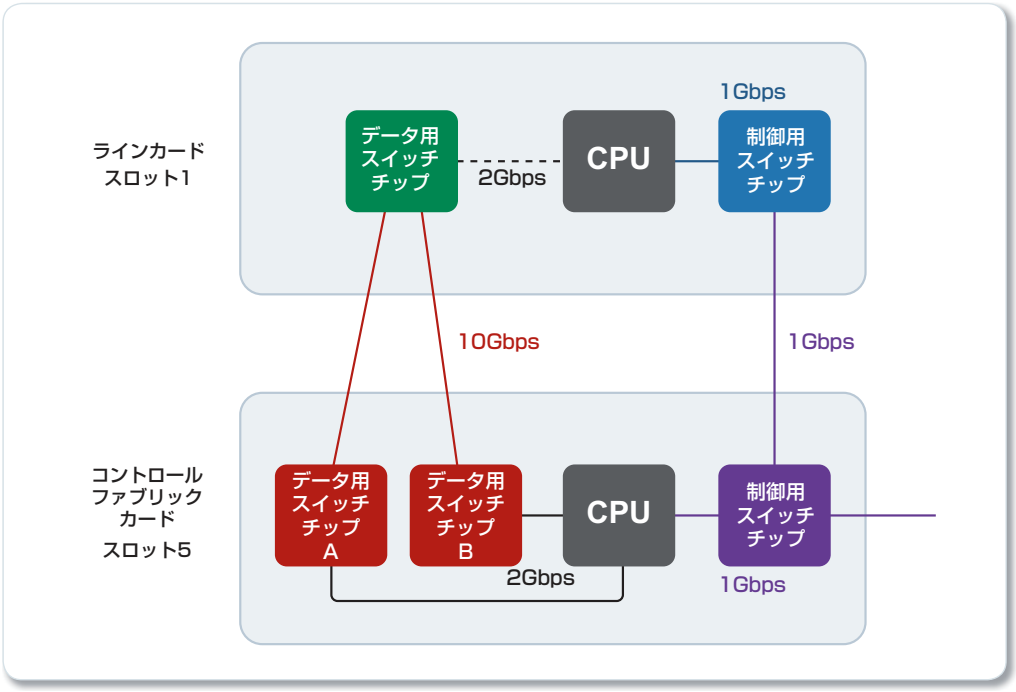


2 CPU 宛てトラフィック

コントロールプレーン上のトラフィックはすべて、コントロールファブリックカード・ラインカード間の通信であり、各カードの CPU によって処理されます。コントロールプレーンから CPU にパケットを転送するため、CPU と制御用スイッチチップは 1Gbps のリンクで接続されています。

また、データプレーン上のトラフィックにも、EPSR、xSTP、LACP、OSPF、ARP のように CPU 処理を必要とするものがありますが、これらのプロトコル制御パケットはアクティブなコントロールファブリックカードの CPU によって処理されます。データプレーンから CPU にパケットを転送するため、CPU とデータ用スイッチチップの間も 2Gbps のリンクで接続されています。

次図は、コントロールプレーン、データプレーンと CPU がどのように接続されているかを示します。



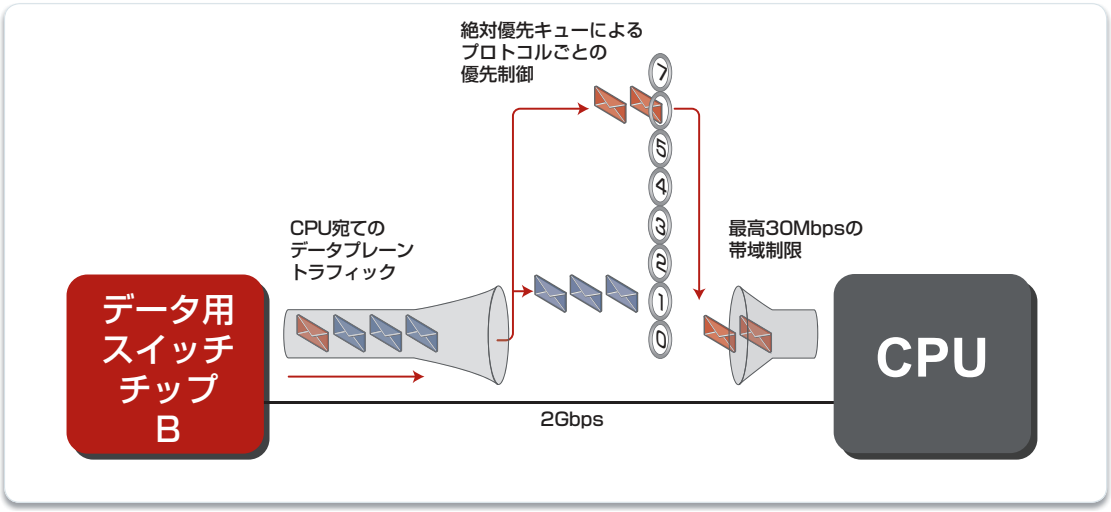
2.1 CPU 宛てデータプレーントラフィックの優先制御と帯域制限

データプレーンを流れるパケットの大部分は、各ラインカードに搭載されたデータ用スイッチチップによって処理されます。

しかし、データプレーン上のパケットの中にも、EPSR、xSTP、LACP、OSPF、ARP など、アクティブなコントロールファブリックカードの CPU で処理しなくてはならないものが一部含まれています。

CPU 処理を必要とするデータプレーン上のパケットは、パケットを受信したラインカードから、データプレーンを通ってアクティブなコントロールファブリックカードのデータ用スイッチチップに転送され、そこから 2Gbps のリンクを経由して CPU に送られます。

CPU が各パケットを適切なタイミングで処理できるよう、データ用スイッチチップから CPU に転送されるデータプレーントラフィックに対しては、プロトコルごとの優先制御が行われます。また、CPU の過負荷を避けるため、CPU 宛てのデータプレーントラフィックには 30Mbps の帯域制限がかけられます。



CPU 宛てのデータプレーントラフィックは 8 つの送信キュー 0 ～ 7 のいずれかに格納されます。キューは番号の大きいほうが上位（高優先度）であり、上位のキューにパケットがある間、下位キューのパケットは送信されません。

次に、データプレーンから CPU に送られる各プロトコルパケットがどのキューに格納されるかを示します。

プロトコル	キュー
該当なし	7
L2 Control Packets / EPSR / Loop Detect	6
DHCP Snooped Packets	5
該当なし	4
Link-local multicast protocols (OSPF / RIPv2 / PIM / VRRP / IGMP)	3
Default / ARP Reply / ARP Request / L2 unregistered Multicast	2
Unicast IPv4 Interface route	1
Unicast IPv4 Default route / L3 unregistered Multicast	0

3 コントロールファブリックカード

3.1 アクティブ・スタンバイと各種情報の同期

コントロールファブリックカードを 2 台装着している場合は、1 台がアクティブ、もう 1 台がスタンバイとなります。

本体宛てのトラフィックを処理し、シャーシを制御するのは、アクティブなコントロールファブリックカードの役目です。スタンバイのコントロールファブリックカードは、2 個のデータ用スイッチチップを通じて、データプレーン帯域を 2 倍（ラインカードスロット 1 スロット当たり 20Gbps → 40Gbps）に増強しますが、スタンバイの CPU が本体宛てトラフィックを処理したり、シャーシを制御したりすることはありません。

ただし、スタンバイのコントロールファブリックカードの CPU でもネットワークプロトコルモジュールはすべて稼働しており、また、アクティブなコントロールファブリックカードとの間で各種情報の同期をとることで、フェイルオーバーに備えています。

次にアクティブとスタンバイの両コントロールファブリックカード間で同期される情報の一覧を示します。

- FDB
- ARP テーブル
- IP ルートデータベース
- RIP ルートデータベース
- OSPF ネイバー / ルートデータベース
- VRRP の状態
- IGMP Snooping
- IGMP マルチキャストグループテーブル
- PIM-SM/DM マルチキャストルートデータベース
- ローカル RADIUS サーバーの認証情報
- DHCP サーバーの IP アドレスリース情報
- EPSR の状態
- ポート認証の情報

上記一覧のうち、FDB、ARP テーブル、IP ルートデータベースの 3 つは、コントロールファブリックカード間だけでなく、コントロールファブリックカード・ラインカード間でも同期されます。これにより、各ラインカードは、受信したパケットの転送先を判断するときに自分自身が持つテーブルを参照することができます。

3.2 起動時のアクティブ決定プロセス

コントロールファブリックカードを 2 台装着した状態でシャーシを起動すると、左スロット（スロット 5）に装着したコントロールファブリックカードが優先的にアクティブになります。これは CLI コマンドや電源の入れ直しなどによってシャーシを再起動した場合も同様です。

左スロット（スロット 5）にコントロールファブリックカードが装着されていないか、左スロットのコントロールファブリックカードが故障している場合は、右スロット（スロット 6）のコントロールファブリックカードがアクティブになります。

3.3 フェイルオーバー

障害やホットスワップによりアクティブなコントロールファブリックカードが不在となった場合は、スタンバイのコントロールファブリックカードがアクティブとなって、システムの制御を引き継ぎます。

フェイルオーバーの時間を短くするため、スタンバイのコントロールファブリックカードはすべてのネットワークプロトコルモジュールを稼働させており、また、アクティブなコントロールファブリックカードとの間で各種情報を同期しています。

3.4 ホットスワップ

スタンバイのコントロールファブリックカードをホットスワップで取り外しても、一時的にデータプレーン帯域が半分になり、また、予備のコントロールファブリックカードが存在しない状態となる以外に、システムへの直接的な影響はありません。空いたスロットに再度コントロールファブリックカードを取り付ければ、データプレーン帯域は元通り 2 倍となり、アクティブなコントロールファブリックカードと各種情報の同期が行われて、フェイルオーバーへの準備が整います。

アクティブなコントロールファブリックカードをホットスワップで取り外した場合はフェイルオーバーが発生し、スタンバイのコントロールファブリックカードがアクティブとなって、システムの制御を引き継ぎます。その後、空いたスロットにコントロールファブリックカードを再度取り付けると、そのカードはスタンバイとなって、アクティブなコントロールファブリックカードと各種情報を同期し、データプレーン帯域を 2 倍に増強します。

シャーシ起動時とは異なり、アクティブなコントロールファブリックカードが稼働している状態で取り付けたコントロールファブリックカードは、たとえ左スロット（スロット 5）に装着したとしても、取り付けと同時にアクティブにはなりません。ただし、シャーシを再起動すれば、起動時のアクティブ決定プロセス（p.9）にしたがって、左スロットのコントロールファブリックカードが再びアクティブになります。

3.5 ファームウェア

本製品ではすべてのコントロールファブリックカードとラインカードに CPU が搭載されているため、同じファームウェアをすべてのカードにロードする必要があります。

しかし、ファームウェアはアクティブなコントロールファブリックカードから他のカードへ自動的に配布されるため、ファームウェアのバージョンアップ時に必要なのは、アクティブなコントロールファブリックカードのフラッシュメモリーに新しいファームウェアのイメージファイルを保存し、これを通常用ファームウェアとして指定し、シャーシを再起動することだけです。

これにより、アクティブなコントロールファブリックカードは新しいファームウェアで起動し、コントロールプレーン上の内部ネットワークを通じて他のカードに新しいファームウェアを配布・同期するようになります。

3.5.1 ファームウェアバージョンの同期

すべてのコントロールファブリックカードとラインカードは同じファームウェアを実行している必要があります。これが確実に行われるようにするのは、アクティブなコントロールファブリックカードの役割です

ラインカードは、毎回アクティブなコントロールファブリックカードからファームウェアをロードして起動する仕組み（p.13）になっているため、原則的にバージョンの不一致は起こりえません。

一方、コントロールファブリックカードは、デフォルトでは自身のフラッシュメモリーからファームウェアをロードするため、同一シャーシに装着された 2 台のコントロールファブリックカードが異なるバージョンのファームウェアで起動する可能性があります。その場合はアクティブなコントロールファブリックカードが、スタンバイのコントロールファブリックカードのファームウェアを自動的に更新し、自身と同じバージョンになるようにします。

ファームウェアバージョンの同期方法には、2 台のファームウェアバージョンがどのように異なっているかによって次に示す 2 とおりの方法があります。

1. マイナーバージョンが異なる場合

   アクティブ・スタンバイのコントロールファブリックカード間でマイナーバージョンに差異がある場合、たとえば、一方のファームウェアバージョンが 5.4.2-3.6 で、もう一方が 5.4.2-3.7 の場合は、コントロールファブリックカード間での通信が可能です。

   ※ 5.4.2-3.7 は説明上使用している架空のファームウェアです。本書執筆時点では実在しません。

   この場合、アクティブなコントロールファブリックカードは、ファイル同期の仕組みを用いて自身のファームウェアイメージファイルをスタンバイのコントロールファブリックカードに転送し、そのファームウェアでスタンバイを再起動することで、バージョンの不一致を解消します。

2. メジャーバージョンが異なる場合

   アクティブ・スタンバイのコントロールファブリックカード間でメジャーバージョンに差異がある場合、たとえば、一方のファームウェアバージョンが 5.4.2-3.6 で、もう一方が 5.4.3-0.1 の場合は、ラインカードにファームウェアを配布するときと同じ仕組み（p.13）を使用します。

   ※ 5.4.3-0.1 は説明上使用している架空のファームウェアです。本書執筆時点では実在しません。

   この場合、アクティブなコントロールファブリックカードは、スタンバイのコントロールファブリックカードのフラッシュメモリーに特殊なファイルを書き込んだ上で、スタンバイを再起動します。

   スタンバイのコントロールファブリックカードのブートローダーは、起動時にこのファイルを見つけると、フラッシュメモリーからファームウェアをロードするのではなく、コントロールプレーン上の内部向け TFTP サーバーからファームウェアをロードしようと試みます。

   具体的には、BOOTP リクエストを送信して内部ネットワーク用の IP アドレスとファームウェアイメージファイルの場所を取得し、TFTP によってアクティブなコントロールファブリックカードからファームウェアをロードします。

   いったんスタンバイのコントロールファブリックカードに同一バージョンのファームウェアがロードされると、その後はファイル同期の仕組みによってフラッシュメモリーの内容を同期し、バージョンの不一致を解消します。

3.5.2 ファームウェアバージョンの同期に失敗した場合

なんらかの理由により、コントロールファブリックカード間でファームウェアバージョンの不一致が解消できないと、スタンバイのコントロールファブリックカードは使用不能な状態に陥ります。

ファームウェアバージョンの同期に失敗した場合、シャーシはアクティブなコントロールファブリックカードだけで稼働することになります。これはすなわち、データプレーン帯域が半分となり、アクティブなコントロールファブリックカードに障害が発生した場合に役割を引き継ぐカードが存在しないということになります。この状態はただちに解決する必要があるため、ファームウェアバージョンの同期に失敗した場合は、これを通知するためのログメッセージが出力されます。

なお、ファームウェアバージョンの同期に失敗するのは、スタンバイのコントロールファブリックカードのフラッシュメモリー空き容量が不足しており、アクティブなコントロールファブリックカードで動作しているファームウェアのイメージファイルを保存できない場合です。

通常、アクティブなコントロールファブリックカードは、スタンバイのコントロールファブリックカードのフラッシュメモリーに空き容量が不足している場合、古いファームウェアイメージファイルを削除して空きを作ります。しかし、ファームウェアイメージファイル以外のファイル（たとえば、プロセスが異常終了したときに作られるコアダンプファイル）が多数存在している場合、新しいファームウェアファイルを格納するスペースを作れない可能性があります。

バージョンアップなどのため、アクティブなコントロールファブリックカードに新しいファームウェアをインストールするときは、アクティブ、スタンバイ両方のフラッシュメモリーに十分な空き容量があることを確認してからインストールしてください。具体的な確認方法については、コマンドリファレンス「運用・管理」/「システム」の「ファームウェアの更新手順」をご覧ください。

4 ラインカード

4.1 ラインカードの起動プロセス

SwitchBlade x8100 では、コントロールファブリックカードだけでなく、ラインカードにも CPU が搭載されています。そのためラインカードにも、アクティブなコントロールファブリックカードと同じファームウェアをロードする必要がありますが、それは次の手順にしたがって自動的に行われます。

1. シャーシの電源を入れると、最初にコントロールファブリックカードが起動します。コントロールファブリックカードは、デフォルトではフラッシュメモリーからファームウェアをロードし、その後コントロールプレーン上で内部向けの DHCP サーバーと TFTP サーバーを起動し、ラインカードがファームウェアをロードできるようにします。
2. コントロールファブリックカードは、実行中のファームウェアイメージファイルを、内部向け TFTP サーバーの公開ディレクトリーにコピーします。
3. ラインカードのブートローダーは、起動時にコントロールプレーン上で BOOTP リクエストをブロードキャストします。
4. コントロールファブリックカード上の内部向け DHCP サーバーは、ラインカードの BOOTP リクエストに応じて、内部向けの IP アドレスをラインカードに付与し、ファームウェアイメージファイルの場所（内部向け TFTP サーバーの IP アドレスとイメージファイルのファイル名）を通知します。
5. ラインカードは、内部向け DHCP サーバーから付与された IP アドレスとイメージファイルの情報をもとに、内部向け TFTP サーバーにアクセスし、コントロールプレーンを通じてファームウェアイメージファイルをロードします。これでラインカードの起動は完了です。

4.2 ホットスワップ

ラインカードはホットスワップ可能です。ホットスワップ時に他のラインカードの動作に影響を与えることはありません。新しく装着されたラインカードは、前項で述べた手順にしたがって、アクティブなコントロールファブリックカードからファームウェアをロードして起動します。

ご注意

- 本書に関する著作権などの知的財産権は、アライドテレシス株式会社（弊社）の親会社であるアライドテレシスホールディングス株式会社が所有しています。アライドテレシスホールディングス株式会社の同意を得ることなく本書の全体または一部をコピーまたは転載しないでください。
- 弊社は、予告なく本書の一部または全体を修正、変更することがあります。
- 弊社は、改良のため製品の仕様を予告なく変更することがあります。

商標について

SwitchBlade はアライドテレシスホールディングス株式会社の登録商標です。

その他、この文書に掲載している製品等の名称は各メーカーの商標または登録商標です。

マニュアルバージョン

2012 年 8 月 17 日 Rev.A

アライドテレシス株式会社