重複したNATエンドポイントを使用したSD-WANハブトンネルのトラブルシューティング

内容

概要

前提条件

- 要件
- 使用するコンポーネント

背景説明

- コントロールプレーンNATラーニング
- データプレーンNATラーニングおよびVCONTROL
- ポートの保持、ポートオーバーラップ、およびポートのランダム化

トポロジの例

コリジョンが発生する理由

- 異なるVBondが同じパブリックエンドポイントを監視できる
- 同じアドバタイズメントが1つのハブに到達する
- ブランチアドレスが繰り返し発生するPrivate-as-Publicコリジョン
- NATエンドポイントの再割り当て

症状と影響

問題の診断

- ステップ 1：ソフトウェアおよびスポークのトランスポート状態の記録
- ステップ 2：OMPでの重複エンドポイントの検索
- ステップ 3：論理TLOCとハブセッション所有権の比較
- ステップ 4：NATゲートウェイの実際の動作の確認
- ステップ 5：WiresharkでのVCONTROLの検査
- 手順 6：影響を受ける経路の確認と証拠の保存

回避策と制限事項

- 一意の割り当てを使用した安定したエンドポイントに依存しないマッピングの使用
- 異なるスポーク間でのポートオーバーラップの無効化
- 小規模な制御グループにポートオフセットを使用する
- 送信元ポートの保持を無効にするか、ランダム割り当てを有効にする
- パブリックアドレスの分離またはポートブロックの予約
- 代替トランスポートの保持と調整の回復

関連ソフトウェア不具合

付録：IPv4 VCONTROLトレーラのデコード

関連情報

概要

このドキュメントでは、スポークが1つのパブリックNetwork Address Translation（NAT；ネットワークアドレス変換）エンドポイントを共有する場合に、ハブルトンネル障害を特定して軽減する方法について説明します。

前提条件

要件

次の項目に関する知識があることが推奨されます。

- Cisco Catalystソフトウェア定義型ワイドエリアネットワーク(SD-WAN)のアーキテクチャと制御接続
- Overlay Management Protocol(OMP)、Transport Locator(TLOC)、Bidirectional Forwarding Detection(BFD)、およびIP Security(IPsec)
- マルチプロトコルラベルスイッチング(MPLS)転送ネットワーク
- NATマッピングおよびフィルタ動作
- Wiresharkによる基本的なパケット分析

使用するコンポーネント

このドキュメントの情報は、次のソフトウェアとハードウェアのバージョンに基づくものです。

- Cisco IOS® XE Catalyst SD-WANルータ
- Cisco Catalyst SD-WANコントローラ(vSmart)
- Cisco Catalyst SD-WANバリデータ(vBond)

Cisco IOS XE 17.18.2では、ペアワイズキーが機能するハブのセッション所有権の障害が再現されました。コマンドの可用性と出力フィールドは、プラットフォームとリリースによって異なります。

例では、プライベートアドレス、またはRFC 5737で予約されているドキュメントアドレスを、汎用のデバイス名で使用しています。CLIの抜粋は短縮され、正規化された識別子を使用します。インターフェイス名と一部の出力フィールドは、プラットフォームとリリースによって異なります。

このドキュメントの情報は、特定のラボ環境にあるデバイスに基づいて作成されました。このドキュメントで使用するすべてのデバイスは、初期（デフォルト）設定の状態から起動しています。本稼働中のネットワークでは、各コマンドによって起こる可能性がある影響を十分確認してください。

⚠ 注意：設定の変更や接続のリセットによって、トラフィックが中断される場合があります。承認されたメンテナンスプランを使用する。初期診断コマンドは読み取り専用です。

背景説明

対称NATが正しく識別されるシナリオから始めます。スポークAはvBond 1に登録され、スポークBはvBond 2に登録されます。RFC 7857で許可されているように、NATゲートウェイは、これらの接続の宛先が異なる場合に、同じ変換済み送信元IPアドレスとポートを割り当てることができます。その結果、両方のスポークがNATタイプAを報告する場合でも、両方のスポークがパブリックインターネットTLOCに同じパブリックエンドポイントをアドバタイズできます。

サンプルトポロジでは、パブリックインターネットカラーを使用しています。同じシナリオで、ビジネスインターネットなどの他のパブリックカラーも使用できます。診断フィルタを実際に影響を受けるトランスポートカラーに適用します。public-internet colorリテラルは必須の条件ではありません。

注:Cisco Catalyst SD-WANは、同じパブリックIPアドレスおよびUDPポートを共通の受信エッジに向けてアドバタイズする、異なるサイトからのTLOCをサポートしません。このコリジョン状態におけるレシーバの動作は予測不可能です。これは、対称NATが正しく識別されている場合でも発生する可能性があります。 正しいNAT分類だけでは、アドバタイズされたエンドポイントが一意であるとは保証されません。

コントロールプレーンNATラーニング

Edgeは、vBondへの登録時にプライベートトランスポートアドレスとポートを送信します。vBondはその接続の送信元アドレスとポートを確認し、確認したエンドポイントをEdgeに返します。エッジは、OMPを使用して、パブリックアドレスやTLOCのポート属性（パブリックインターネットやビジネスインターネットなどのパブリックカラーのTLOCを含む）として、この情報をアドバタイズできます。

観察は、そのvBondへの接続を示します。宛先依存のNATでは、ハブへのマッピングが異なる場合があります。ルータは、マッピング動作を分類するためにNATの観測結果を比較します。show sdwan control local-propertiesでは、Aはアドレスおよびポートに依存するマッピングを意味し、Eはエンドポイントに依存しないマッピングを意味します。表示された分類はルータの決定を表し、NAT変換テーブルまたはパケットキャプチャはNATゲートウェイの実際の動作を確認します。

たとえば、show sdwan control local-propertiesを実行してトランスポートのNAT TYPE列を見つけます。この短縮された出力では、NATタイプの凡例と選択されたインターフェイスの列が、正規化されたアドレスとともに保持されます。

<#root>

Spoke-A# show sdwan control local-properties

```
...
 NAT TYPE: E -- indicates End-point independent mapping
           A -- indicates Address-port dependent mapping
           N -- indicates Not learned
           Note: Requires minimum two vbonds to learn the NAT type

                         PUBLIC         PUBLIC PRIVATE         PRIVATE                          NAT
INTERFACE                IPv4           PORT   IPv4            PORT    COLOR          STATE     TYPE
GigabitEthernet2         192.168.9.11    12346  192.168.9.11    12346   mpls           up        E
GigabitEthernet3         203.0.113.10    23012  192.168.11.11   12346   biz-internet   up        A
```

biz-internet transport(BITT)はAを報告し、パブリックエンドポイントとして203.0.113.10:23012をアドバタイズします。MPLSトランスポートはEを報告し、プライベートエンドポイントとパブリックエンドポイントは同じです。NAT TYPEから分類を読み取ります。アドレスまたはポートの等価性だけでは、NATタイプは識別されません。Nは、ルータが分類を学習していないことを意味します。

データプレーンNATラーニングおよびVCONTROL

ハブは、暗号化されたBFDパケットを処理する前に、対応するピアとトンネルの着信IPsecセキュリティアソシエーション(SA)を検出する必要があります。SAは、パケットの認証と復号化に必要な暗号キーとその他の状態を識別します。受信側は、エンドポイント情報とEncapsulating Security Payload(ESP)Security Parameters Index(SPI)を使用して、その状態を特定します。SPI単独は、グローバルに一意なスポーク識別子ではありません。

通常、対称NATの背後にあるスポークは、到達可能なハブへのトラフィックを開始します。これにより、その特定の宛先のマッピングが作成されます。ハブが認識する送信元アドレスまたはポートは、vBondを通じて以前に学習され、OMPでアドバタイズされたパブリックエンドポイントと異なる場合があります。その結果、受信した外部ヘッダーは、インストールされたトンネル状態のエンドポイントと必ずしも一致しません。ハブは、BFDを復号化する前に、この新しいオンワイヤエンドポイントを既知のリモートTLOCおよびそのIPsec状態に関連付ける方法を必要とします。

ここで説明するIPv4 BFD-over-IPsec形式では、SD-WANはエンドポイント情報を、暗号化されたBFDペイロードの外部にある、UDPペイロードの末尾の可視のVCONTROLトレーラに配置します。ハブは、この情報を復号化の前に読み取ることができます。VCONTROLは、暗号化キーではなくエンドポイント情報を伝送し、それ自体ではピアIDを確立しません。

たとえば、最初のトンネル確立時に、データプレーンの学習が完了する前に、次の手順を実行します。

```
Spoke B advertised through OMP:       203.0.113.10:12346
Spoke B observed at the hub:          203.0.113.10:45678
VCONTROL carried source endpoint:    203.0.113.10:12346
```

受信側は、最初に通常のエンドポイント/SPIルックアップを試行します。このルックアップでSAが見つからない場合、VCONTROL送信元エンドポイントがフォールバックを提供します。この例では、パケットが203.0.113.10:45678から到達した場合でも、ハブは搬送された203.0.113.10:12346を、ローカルの宛先コンテキストおよびSPIとともに使用して、既知のリモートTLOCに関連付けられた着信SAを探すことができます。

選択したSAが受信キーコンテキストを提供します。SPIは、そのコンテキスト内で適用可能なキー生成を選択します。次に、ハブはIPsec認証と復号化を実行します。必要なチェックを通過するパケットだけがBFD処理に使用できます。読み取り可能なトレーラまたは正常なルックアップだけでは、認証、復号化、またはBFD受信の成功は証明されません。

したがって、対称NATでは、観測された送信元ポートと搬送された送信元ポートの違いは正常です。ピア関係が機能するためには、ハブがパケットを正しいトンネルとキーの状態に関連付け、実際のリターンエンドポイントを学習する必要があります。この例では、正常に学習されると、ハブはリターントラフィックを203.0.113.10:45678に送信します。これは、vBondによるコントロールプレーンの学習とは別です。

VCONTROLは、既存のエンドポイント/SPIの一致をレシーバがオーバーライドすることを強制しません。エンドポイントの状態が重複して、パケットが誤ったピアまたはキー生成と関連付けられている場合、トレーラは回復を保証しません。ルックアップヒントを提供するため、正しい周辺トンネルと暗号化状態がないと、同じエンドポイントを伝送する2つのピアを区別できません。

NATタイプEが報告される場合、該当するBFDパケットにもVCONTROLが存在します。このフラグは対称NATフラグではなく、通常のIPSecデータパケットには含まれません。ここで説明するIPv4 BFD形式では、2つのプライベートカラーの間でpublic-TLOC情報が省略されます。

データプレーンを通じて学習されたピア固有のエンドポイントは、スポークがすべての宛先に対してOMPアドバタイズメントを新しいエンドポイントで置き換える必要があることを意味するものではありません。学習されたさまざまなマッピングは、対称NATトラバーサルの予期される部分です。

トンネルがアップ状態になると、そのトンネル用に独自に変換されたエンドポイントを学習した送信者は、そのエンドポイントを送信元情報として伝送できます。これは、送信者のトンネルごとのパブリックエンドポイントであり、リモートエンドポイントや新しいグローバルOMPアドバタイズメントではありません。常に、同じ観測期間のパケット、OMPアドバタイズメント、およびセッション状態を比較します。

ポートの保持、ポートオーバーラップ、およびポートのランダム化

次のようなさまざまな動作があります。

動作	意味
送信元ポートの保持	NATゲートウェイは、スポークの元のUDP送信元ポート（たとえば、12346）を変換後の送信元ポートとして保持しようとします。
宛先間でのポートオーバーラップ	NATゲートウェイは、異なる外部宛先と通信する異なる内部エンドポイントに対して、変換された1つのIP：ポートを再利用します。
ポートのランダム化	NATゲートウェイは、ランダム化された割り当て方式を使用して、変換されたポートを選択します。
エンドポイントに依存しないマッピング	マッピングが存在する間、1つの内部エンドポイントは、外部の宛先に対して同じ変換されたエンドポイントを保持します。

RFC 7857のセクション3では、宛先が異なる場合に、異なる内部エンドポイントに対して重複する外部ポートを許可しています。NATは5つのタプルを使用してフローを区別します。これにより、同じ宛先への2つの識別不可能なUDPフローを同時に実行することはできません。ランダムな割り当てと保持は別々の割り当て選択であり、どちらもエンドポイントの独立性を確立しません。

トポロジの例

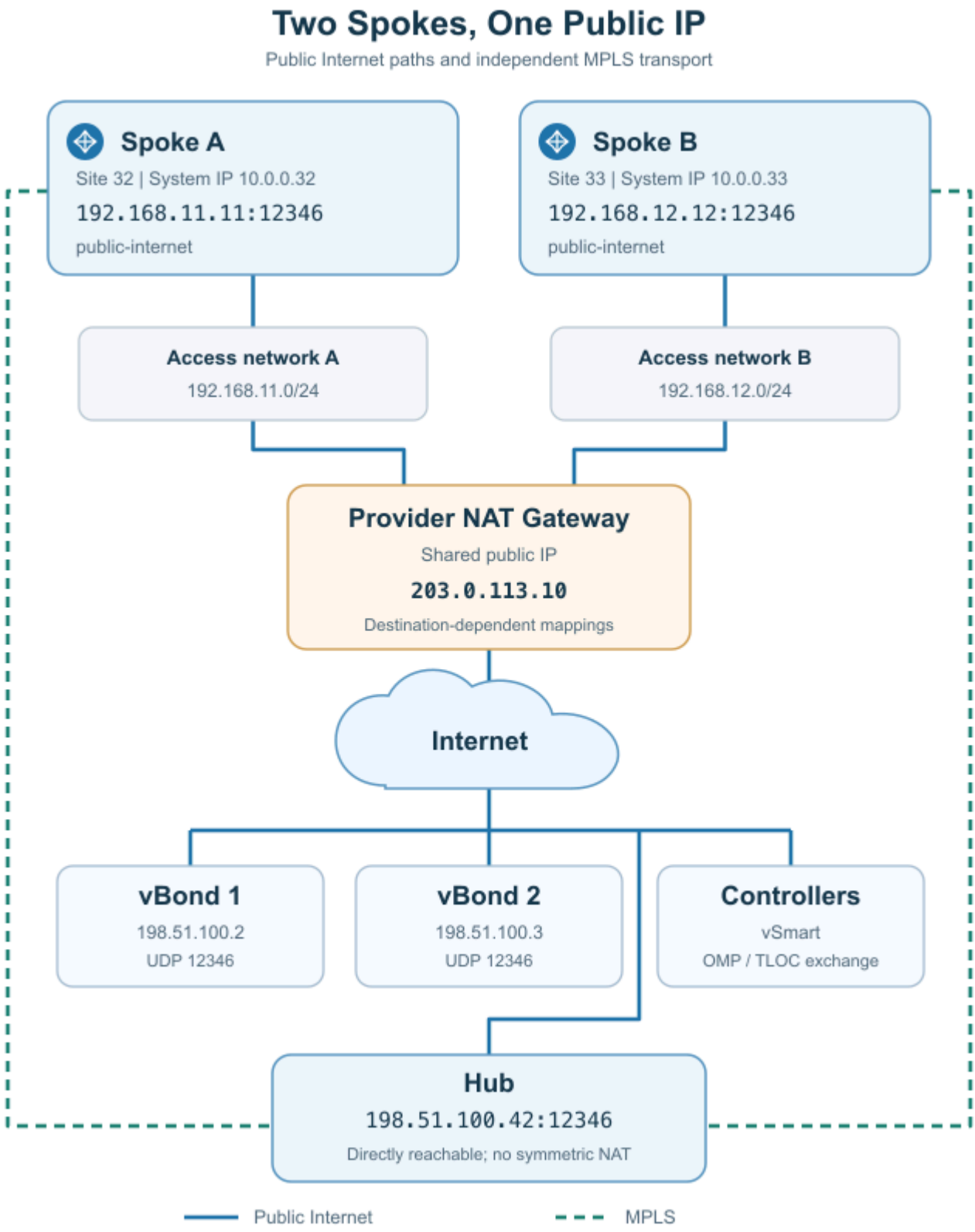

図 1.共有NATゲートウェイ、独立したスポークアクセスネットワーク、独立したMPLSトランスポート。

スポークAは192.168.11.11:12346を使用し、スポークBは192.168.12.12:12346を使用します。両方とも、NATゲートウェイを介してパブリックアドレス203.0.113.10を共有します。ハブには直接到達可能なアドレスがあり、対称NATの背後にはありません。2つのスポークは、プライベート

エンドポイントが異なっていても、同じパブリックIP:portをアドバタイズできます。

コリジョンが発生する理由

異なるVBondが同じパブリックエンドポイントを監視できる

次の同時UDPマッピングを検討します。

ピア	プライベートソース	翻訳済みソース	宛先
A	192.168.11.11:12346	203.0.113.10:12346	vBond 1、198.51.100.2:12346
B	192.168.12.12:12346	203.0.113.10:12346	vBond 2、198.51.100.3:12346

NATはリモートエンドポイントも確認するため、応答は正しく返されます。vBond 1からの応答はAのマッピングに属し、vBond 2からの応答はBのマッピングに属します。

また、両方のルータが他のvBondに接続し、宛先依存のマッピングを正しく検出することもできます。たとえば、各スポークのプライベートソースタプルを変更せずに、次の観測順序を考慮します。

ピア	最初に受け入れられたvBond観察	他のvBondとの比較	結果
A	vBond 1は203.0.113.10:12346を返します。	vBond 2は203.0.113.10:40001を返します。	NAT A、パブリックタプル203.0.113.10:12346を保持
B	vBond 2は203.0.113.10:12346を返します。	vBond 1は203.0.113.10:40002を返します。	NAT A、パブリックタプル203.0.113.10:12346を保持

vBondの観測値はルータごとに異なるため、NAT Aの検出は正しく行われます。ただし、最初に確認した内容は2台のルータ間で一致しています。分類Aでは、すべてのスポークに一意のパブリックポートが割り当てられるわけではなく、両方のvBondを訪問しても、その一意性は保証されません。

各スポークは、登録応答でvBondによって確認されたパブリックアドレスとポートを受け取り、保持された確認を使用してローカル転送プロパティを入力します。スポークは、OMPを介してSD-WANコントローラにアドバタイズするTLOC属性に、パブリックエンドポイントとそのプラ

イベートエンドポイントを含めます。コントローラは、該当する制御ポリシーに従ってTLOCを他のエッジに分散します。vBond自体はOMPアドバタイズメントを開始しません。

両方のスポークが同じ監視対象パブリックエンドポイントを保持する場合、OMPには次を含めることができます。

```
TLOC system IP   Color             Public endpoint        Private endpoint
10.0.0.32        public-internet   203.0.113.10:12346     192.168.11.11:12346
10.0.0.33        public-internet   203.0.113.10:12346     192.168.12.12:12346
```

これらは、システムIPが異なるため、論理TLOCは異なったままです。

同じアドバタイズメントが1つのハブに到達する

NATゲートウェイは、同じハブに接続する両方のスポークを区別する必要があります。考えられる割り当ての1つは次のとおりです。

ピア	ハブへの変換済みソース	宛先
A	203.0.113.10:12346	198.51.100.42:12346
B	203.0.113.10:45678	198.51.100.42:12346

ハブには、同じ宛先エンドポイントをアドバタイズする2つの論理ピアがありますが、パケットは異なる実際のエンドポイントから到達できます。通常の対称NAT操作では、データプレーン学習がリターンマッピングの確立に役立ちます。

ハブセッション所有権の制限は、次のトランスポートキーで選択されたセッション状態を異なるピアが共有する場合に発生します。

```
Hub-local IP + hub-local port + remote IP + remote port + protocol
```

両方のアドバタイズメントが同じキーを選択した場合、後のピアからのアップデートによって共有状態の所有権情報を置き換えることができます。あるピアのセッションを取り消したり、削除したり、再作成したりすると、他のピアのIPSec/BFD状態が乱れる可能性があります。データプレーン学習は、その状態が誤ったピアまたはフロー生成にすでに関連付けられている場合は、回復を保証しません。

可視トレーラは独立したピアIDではありません。両方のスポークは同じ送信元エンドポイントを伝送できます。ハブが誤ったピアまたはキー生成に属する既存のIPSec状態を選択すると、パケットがハブに到達しても認証とラーニングが失敗する可能性があります。特定のパケットに対してこのシーケンスを確立するには、Cisco Technical Assistance Center(TAC)は、選択したSAをピアと世代に関連付ける必要があります。パケットトレーラだけでは、それを証明できません。

同じパブリックIPアドレスだけではコリジョンは確立されません。通常、リモートポートが異なると、生成されるトランスポートキーも異なります。また、選択したハブローカルアドレス、ローカルポート、およびプロトコルを確認します。3番目のピアが同じエンドポイントをアドバタイズすると、そのキーを競合することもできます。障害パターンはイベントの順序によって異なります。

ブランチアドレスが繰り返し発生するPrivate-as-Publicコリジョン

Cisco Bug ID CSCwx12986により、NAT比較中の自動ソースポートホップの後、EdgeがパブリックTLOC属性でプライベートRFC 1918タプルをアドバタイズする可能性があります。異なるブランチが同じプライベートアドレスとポートを再利用すると、これらの誤ったパブリック属性がピア間で一致し、ハブセッション所有権の問題が発生する可能性があります。

たとえば、両方のスポークが別々のブランチファイアウォールの背後で192.168.11.11を使用できます。各ファイアウォールは、ブランチネットワークを個別に維持し、送信元アドレスを変換します。アップストリームプロバイダーのNATゲートウェイが再度変換を行う場合、パスには2つのNATステージ、つまり二重NATが存在します。ブランチアドレスの重複を分離しておくもう1つの方法は、独立したVirtual Routing and Forwarding(VRF)コンテキストです。

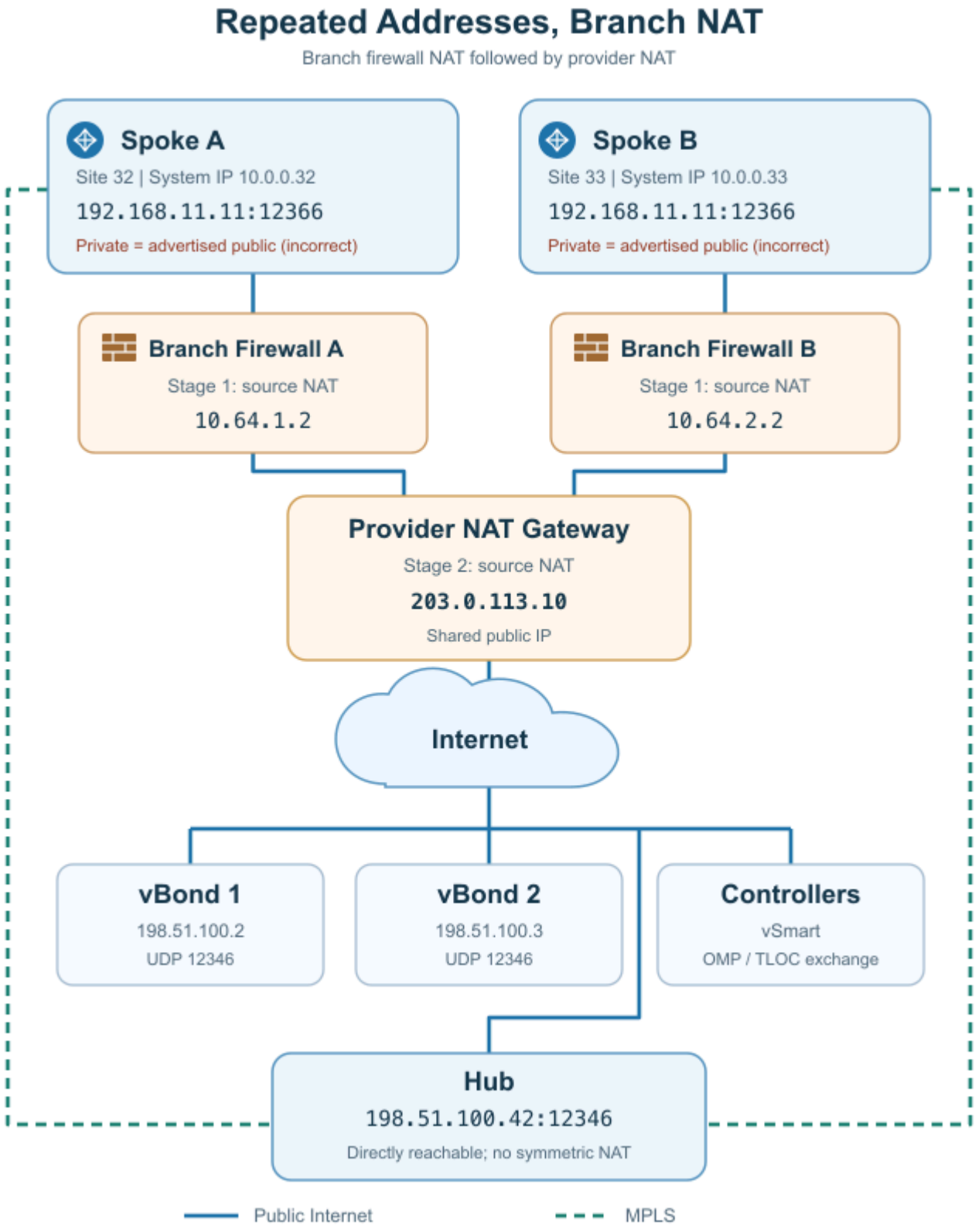

図 2：ブランチファイアウォールとプロバイダーのNAT変換。不正なパブリックTLOC属性は元のプライベートエンドポイントを保持しますが、パケットは変換された送信元アドレスを使用します。

両方のスポークがパブリックインターネットTLOCで不具合に遭遇し、同じポストホップ送信元ポートを使用する場合、ハブは次のアドバタイズメントを受信できます。

ピア	ポートホップ後のプライベートエンドポイント	不正なパブリックTLOCエンドポイント
A	192.168.11.11:12366	192.168.11.11:12366
B	192.168.11.11:12366	192.168.11.11:12366

注：繰り返されるプライベートアドレスとダブルNATは、Cisco Bug ID CSCwx12986自体の前提条件ではありません。追加のクロスピアコリジョンでは、アドバタイズされたIP：ポート値の照合と、同じハブローカルアドレス、ポート、およびプロトコルが必要です。アドバタイズされたポートが異なる同じプライベートIPでは、このコリジョンは確立されません。

パブリックアドレスとプライベートアドレスの同等は、NATを使用しないトランスポートでは正常です。パブリックフィールドのRFC 1918アドレスが疑わしいのは、意図したパスと観察された変換に矛盾する場合だけです。一意のブランチプライベートアドレスにより、この例での繰り返されるプライベートタプルの重複が回避されます。これにより、前述した個別の変換されたパブリックポートの再利用シナリオが妨げられることはありません。

NATエンドポイントの再割り当て

また、NATマッピングの期限が切れた後、ハブが古いピア状態を維持したままで、エンドポイントを別のスポークに再割り当てすることもできます。マッピング履歴とハブセッションのライフサイクルを比較して、新旧の所有者の重複を特定します。

症状と影響

一般的なインジケータは次のとおりです。

- 影響を受けたBFDパスがダウンしている間、または回復できない間、コントロール接続はUpのままになります。
- 2つの異なるシステムIPが同じパブリックIP:portをアドバタイズします。
- ハブは両方の論理TLOCを保持しますが、フィルタリングされたプラットフォームセッションビューに1つのピアが欠けているか、一貫性のないシステムIP、リモートカラー、識別子、トンネル、またはSA情報が表示されます。
- あるピアに対するトランスポートの回収または再作成は、同じエンドポイントを持つ別のピアを妨害します。
- IPSecの入力廃棄は、影響を受けるパスで増加する可能性があります。

障害が発生すると、影響を受けるトランスポートパスが使用できなくなる可能性があります。これがスポークに使用可能な唯一のトランスポートパスである場合、スポークはオーバーレイから

分離されます。大規模なハブでは、1つのNATアドレスを共有する多くのスポークが関連付けを必要とする人口を増加させますが、影響を受けるパスの数を測定する必要があります。この数は、そのアドレスの背後にあるサイトの数からは推測できません。

問題の診断

ステップ 1：ソフトウェアおよびスポークのトランスポート状態の記録

該当するCisco IOS XEエッジのそれぞれで、<affected-color>を実際のトランスポートカラー（biz-internetまたはpublic-internetなど）に置き換えます。

<#root>

show clock detail

show version | include Cisco IOS XE Software

show sdwan control local-properties

show sdwan control connections | include

show sdwan omp peers

プライベート/パブリックアドレスとポート、インターフェイス、色、NATタイプ、サイトID、および制御状態を記録します。最初の完全なローカルプロパティサンプルの後、このコンパクトなフィルタを使用して、インターフェイスヘッダーと選択したトランスポート行を保持します。TenGigabitEthernet0/1/0.10を該当するインターフェイスで置き換えます。

<#root>

show sdwan control local-properties "wan-interface-list" | include INTERFACE|PUBLIC|TenGigabitEthernet0/

プライマリシナリオでは、関連するフィールドは両方のスポークで読み取ることができます。これはフィールドの要約であり、ネイティブ列のレイアウトではありません。

```
Field                          Spoke A                 Spoke B
System IP                      10.0.0.32               10.0.0.33
Color                          public-internet         public-internet
Private endpoint               192.168.11.11:12346      192.168.12.12:12346
Public endpoint                203.0.113.10:12346       203.0.113.10:12346
NAT type                       A                       A
vSmart controls on transport   Up                      Up
```

アップコントロール接続は、コントローラの関係を確認します。スポークとハブ間のBFDパスや、パブリックエンドポイントの一意性は検証されません。

ステップ 2：OMPでの重複エンドポイントの検索

Cisco IOS XEハブ：

<#root>

show clock detail

show sdwan omp tlocs table | include 203.0.113.10

SD-WANコントローラ(vSmart)で、次の手順を実行します。

<#root>

show clock

show omp tlocs | tab | include 203.0.113.10

このコントローラの出力では、正規化された識別子が使用されています。IPv6および地域の列は省略されています。

<#root>

vSmart# show omp tlocs | tab | include 203.0.113.10

```
TLOC IP     COLOR            ENCAP FROM PEER  STATUS KEY PUBLIC IP     PORT  PRIVATE IP     PORT
10.0.0.32   public-internet ipsec 10.0.0.5   C,R    1   203.0.113.10 12346 192.168.11.11 12346
                                 10.0.0.32  C,I,R  1   203.0.113.10 12346 192.168.11.11 12346
```

```
10.0.0.33   public-internet ipsec 10.0.0.5   C,R    1   203.0.113.10 12346 192.168.12.12 12346
                            10.0.0.33  C,I,R  1   203.0.113.10 12346 192.168.12.12 12346
```

TLOCシステムのIP、色、カプセル化、パブリックIP、およびパブリックポートを比較します。異なるコントローラから同じ論理TLOCへの複数のパスは正常です。コリジョン候補は、同じ有効なエンドポイントを持つ2つの異なるピアアイデンティティです。

30 ~ 60秒後にフィルタ処理されたコマンドを繰り返し、両方のタイムスタンプを保持します。持続的なオーバーラップは有用な証拠であり、完全に同時のコマンドは必要ありません。一時的に短い場合は、TACでOMPイベントと対応するセッションライフサイクルレコードを収集します。

問題の疑われる2つのピアについて、それぞれのTLOCテーブルエントリを保持します。

<#root>

show sdwan omp tlocs table | include 10.0.0.32

show sdwan omp tlocs table | include 10.0.0.33

Cisco IOS XEのIPv4 TLOCの詳細を簡潔に表示するには、このフィルタを使用します。システムIPを調査中のピアと置き換え、ip <system-IP>セレクタを引用符で囲んだままにします。

<#root>

show sdwan omp tlocs "ip 10.0.0.32" | exclude not set| 0|None|efault|::

show sdwan omp tlocs "ip 10.0.0.33" | exclude not set| 0|None|efault|::

両方のピアのパブリックアドレス/ポート属性とプライベートアドレス/ポート属性を比較します。フィルタは、このビューから、unset、0値、default、およびIPv6の各行の照合を省略します。

ステップ 3：論理TLOCとハブセッション所有権の比較

FTMに配信されるTLOC状態を調べます。両方のピアのシステムIP、色、パブリックIPアドレス、およびパブリックUDPポートを比較します。

<#root>

show platform software sdwan ttmd tlocs

転送テーブルマネージャ(FTM)、BFD、およびプラットフォームセッションレコードを比較するには、次のコマンドを実行します。

<#root>

show sdwan ftm bfd tloc | include 10.0.0.32

show sdwan ftm bfd tloc | include 10.0.0.33

show sdwan bfd sessions | include 10.0.0.32

show sdwan bfd sessions | include 10.0.0.33

show platform software sdwan session dstip 203.0.113.10

show platform software sdwan session remote-sysip 10.0.0.32 detail

show platform software sdwan session remote-sysip 10.0.0.33 detail

次の例は、正規化された識別子を持つパターンを示しています。列ラベルは、わかりやすくするためにフィルタ処理されたFTM行に追加されます。これらは、フィルタ処理されたコマンド出力の一部ではありません。

<#root>

Hub# show sdwan ftm bfd tloc | include 10.0.0.32

```
System IP  Remote color     Encap  TLOC index  State
10.0.0.32  public-internet  IPSEC  32827       DOWN
```

Hub# show sdwan ftm bfd tloc | include 10.0.0.33

```
System IP  Remote color     Encap  TLOC index  State
10.0.0.33  public-internet  IPSEC  32828       DOWN
```

```
Hub# show platform software sdwan session dstip 203.0.113.10

RemoteSysIP Color        Proto SrcIp         SPort DstIp        DPort DPubIp       PPort BFD-LD TUN-ID
10.0.0.32   mpls         IPSEC 192.168.9.42  12346 203.0.113.10 12346 203.0.113.10 12346 20184  798
10.0.0.32   biz-internet IPSEC 192.168.10.42 12346 203.0.113.10 12346 203.0.113.10 12346 20183  799
```

OMPとFTMは両方の論理ピアIDを保持しますが、このセッションルックアップは共有宛先のスポークAのみを示します。スポークAにもダウンFTM状態があり、プラットフォームセッションテーブル内に存在しても正常なパスを示しません。FTM TLOCインデックスは、セッションテーブル内のBFDローカル識別子(BFD-LD)とは異なります。このパターンは、レイヤ間の所有権を調査する理由になります。CLIの行が欠落しているだけでは、パケット損失は発生しません。

この表では、Colorはローカルトランスポートを表しています。詳細ビューでRemote colorを使用して、リモートTLOCを識別します。DstIpとDPubIpも比較します。学習された実際の宛先は、セッションに関連付けられているパブリックエンドポイントと異なる場合があります。宛先専用フィルタが学習アドレスに一致しなくなった場合は、ピア固有のビューを使用します。

無効なネクストホップ、SA識別子0またはIn HW: FALSEは、一意のシグニチャではなく追加の証拠を提供します。これらのフィールドは、通常のセットアップまたはアンダーレイの障害時にも表示されます。

ステップ 4：NATゲートウェイの実際の動作の確認

NATまたはサービスプロバイダーのオペレータに、次のフィールドを持つ同時UDP変換レコードを問い合わせます。

Spoke	Inside source	Translated source	External destination
A	192.168.11.11:12346	203.0.113.10:12346	198.51.100.2:12346
B	192.168.12.12:12346	203.0.113.10:12346	198.51.100.3:12346
A	192.168.11.11:12346	203.0.113.10:12346	198.51.100.42:12346
B	192.168.12.12:12346	203.0.113.10:45678	198.51.100.42:12346

次に、説明のための変換テーブルを示します。使用しているNATプラットフォームに対応する診断コマンドを使用します。Cisco IOS XE NATゲートウェイでは、次で始めます。

<#root>

```
show ip nat translations | include 203.0.113.10
```

内側のソース列と宛先の列の両方を保持します。変換されたソースのみを含むリストでは、異なるvBondに対して再利用が有効である理由や、ハブマッピングを所有しているスポークを示すことはできません。

vBondとハブの両方について、NATの前後の影響を受けるUDPフローをキャプチャします。VCONTROLトレーラが切り捨てられないように、パケットの長さを完全に保持します。TACと合意した、限定スイッチドポートアナライザ(SPAN)またはEmbedded Packet Capture(EPC)コレクションを使用します。キャプチャポイント間で同じフローを比較します。同期されていないキャプチャクロックから遅延を推測しないでください。

ステップ 5：WiresharkでのVCONTROLの検査

ここで説明するIPv4 BFD/VCONTROL形式の付録にあるLuaデコーダを使用します。コードをsdwan_vctrl.luaとしてキャプチャと同じフォルダに保存します。この診断デコーダは、表示可能なフィールドを表示します。ESPの復号化や認証の検証は行いません。そのスコープは、予想されるBFD/VCONTROLフラグとトレーラ構造を持つ、完全でネストしていないIPv4/UDPパケットです。

デコーダとキャプチャが含まれているフォルダからWiresharkコマンドを実行します。

<#root>

```
wireshark -n \

  -X lua_script:sdwan_vctrl.lua \

  -d udp.port==12346,sdwan_vctrl \

  -d udp.port==45678,sdwan_vctrl \

  -r hub.pcap
```

サンプルポートをキャプチャ内のポートに置き換えます。WindowsまたはmacOSでは、Wireshark実行可能ファイルがPATHにない場合は、その実行可能ファイルへのフルパスを使用します。

または、Wiresharkの起動時にLuaスクリプトをロードし、関連するUDPパケットを選択して、Analyze > Decode Asに移動します。該当する送信元および宛先UDPポートエントリをSDWAN_VCTRL / SD-WAN VCTRL(IPv4)に割り当てます。 他方のポートの明示的な

UDPENCAPエントリが、このデコーダよりも優先される可能性があります。この分析セッションに変更を適用します。

次の表示フィルタを適用します。

```
sdwan_vctrl && ip.addr == 203.0.113.10
```

SD-WAN VCTRL > Visible VCONTROL trailerの順に展開し、次を比較します。

```
Outer IPv4 source:       203.0.113.10
Outer UDP source port:   45678
Carried source IPv4:     203.0.113.10
Carried source UDP port: 12346
Carried destination:     198.51.100.42:12346
```

次の図の値は、データプレーンの学習が完了する前のトンネルの初期確立を示しています。Bの実際のハブマッピングと伝送されたエンドポイントが表示されます。NATは外側のヘッダーを変更し、このトレーラ内のエンドポイント情報を書き換えません。このミスマッチだけでは障害は示されません。

コンパクトにエクスポートするには、次のようにTSharkコマンドを実行します。

<#root>

```
tshark -n -X lua_script:sdwan_vctrl.lua \

  -d udp.port==12346,sdwan_vctrl \

  -d udp.port==45678,sdwan_vctrl \

  -r hub.pcap -Y sdwan_vctrl -T fields \

  -e frame.number -e ip.src -e udp.srcport -e ip.dst -e udp.dstport \

  -e sdwan_vctrl.spi \

  -e sdwan_vctrl.src_ip -e sdwan_vctrl.src_port \
```

```
  -e sdwan_vctrl.dst_ip -e sdwan_vctrl.dst_port
```

カスタムデコーダを使用しない場合、ここで説明するIPv4形式では、UDPペイロードの末尾に17バイトのタイプ/長さ/値(TLV)トレーラが付加されます。リンク層のパディングがUDPペイロードの外部にある場合：

```
81 00 0f cb 00 71 0a 30 3a c6 33 64 2a 30 3a 00 11
|  |     |           |     |           |     |
|  |     |           |     |           |     +-- Complete trailer length: 17
|  |     |           |     |           +-------- Destination port: 12346
|  |     |           |     +-------------------- Destination: 198.51.100.42
|  |     |           +-------------------------- Source port: 12346
|  |     +-------------------------------------- Source: 203.0.113.10
|  +-------------------------------------------- TLV length: 15
+----------------------------------------------- Type: 0x81
```

この固定レイアウトは、IPv6、その他の制御フォーマット、切り捨てられたパケット、または任意の暗号化データには適用しないでください。空のデコーダフィルタは、パケットがサポートされている形式の範囲外であることを意味する場合があります。読み取り可能な送信元エンドポイントも、認証されたシステムIPラベルではありません。

注:Wiresharkでの一般的なESPシーケンスの警告は、ルータのアンチリプレイ診断ではありません。認証または所有権の結論を出す前に、パケットをルータカウンタおよびログと関連付けます。

手順 6：影響を受ける経路の確認と証拠の保存

ハブで、観察された障害の周囲にある2つのQuantumFlow Processor(QFP)カウンターサンプルを取得します。

<#root>

```
show clock detail
```

```
show platform hardware qfp active statistics drop | include Ipsec
```

```
show sdwan bfd sessions | include 10.0.0.32
```

```
show sdwan bfd sessions | include 10.0.0.33
```

絶対合計ではなく、カウンタの差異を使用します。集約IPSecのドロップには、他の原因がある可能性があります。影響を受けるトランスポートでパケットの到着を確認し、ピア、エンドポイント、BFD識別子、トンネル、SA、およびイベント時間と関連付けます。

2つのOMPスナップショット、スポークローカルのプロパティとコントロール、両方のピアのハブセッションの詳細、BFD状態、変換レコード、境界キャプチャ、およびソフトウェアバージョンをTACに提供します。すでに使用可能な場合は、関連する管理者テクノロジーと接続/取り消しの履歴を含めます。大規模なハブでテーブル全体を繰り返し収集することは避けてください。

サービスの到達可能性を個別にテストし、選択した実際のトランスポートを記録します。ゼロ秒のBFD Upが示されても、回復は持続しません。

回避策と制限事項

Cisco TACおよびNATオペレータを使用して、これらの緩和オプションを評価してください。この制限は、共通のハブに到達するアドバタイズされたエンドポイントの重複に関係します。これは、対称NAT、共有パブリックアドレス、または繰り返されるプライベートアドレスが一般にサポートされないことを意味するものではありません。回復手順と緩和効果は、導入によって異なります。

一意の割り当てを使用した安定したエンドポイントに依存しないマッピングの使用

図に示されているシーケンスのNAT側の防御オプションの1つは、vBond、コントローラ、およびハブの宛先に対して、個別のスポークの各トランスポートエンドポイントに、安定した一意の変換されたUDPエンドポイントを付与することです。スポーク間で、エンドポイントの独立性と一意性の両方を要求します。NATゲートウェイのフェールオーバーまたは状態の再作成後も、変換レコードとキャプチャの結果を確認します。

たとえば、Aは一貫して203.0.113.10:40001を使用し、Bは一貫して203.0.113.10:40002を使用できます。スタティック変換または適切な割り当てポリシーは、NATプラットフォームとアクセス設計に従って、この動作を提供できます。

エンドポイントに依存しないマッピングでは、無制限の着信アクセスは必要ありません。必要なファイアウォール制御を維持します。NATポリシーを修正しても、以前にアドバタイズまたはプログラムされたセッション状態は自動的に削除されません。新しい検出サイクルと回復をTACと調整してください。

異なるスポーク間でのポートオーバーラップの無効化

サポートされている場合は、関連するSD-WANスポーク間で、変換されたUDPエンドポイントを一意に予約します。宛先が異なるために再利用しないでください。これにより、NATゲートウェイが宛先依存の動作を保持している場合でも、前述の同時複製vBondマッピングが直接削除されます。

この設定では、より多くのポートを消費したり、追加のパブリックアドレスを必要としたりする可能性があります。このコマンドは、単独では、各スポークのエンドポイントに対するマッピングを独立させず、後でエンドポイントを再割り当てした後も古い受信者状態を解決しません。

複数のNATステージまたは冗長NATゲートウェイで、最終的なインターネット境界に表示される割り当てを確認します。1つのNATインスタンス内での一意性は、同じパブリックアドレスを共有する他のインスタンス間で一意性を確立しません。

小規模な制御グループにポートオフセットを使用する

port-offsetコマンドは0 ~ 19の値を提供します。たとえば、スポークAはオフセット0を保持でき、スポークBはオフセット1を使用できます。スポークBで次の設定コマンドを実行します。

<#root>

```
Spoke-B# config-transaction

Spoke-B(config)# system

Spoke-B(config-system)# port-offset 1

Spoke-B(config-system)# commit

Spoke-B(config-system)# end

Spoke-B# show sdwan control local-properties
```

変更後のアクティブな送信元ポートと変換後のポートを確認します。このシステム設定を変更すると、トランスポート間の制御およびデータ接続を再構築できます。これは中断のない診断手順ではありません。

20個のオフセットにより、20個の別個のベースポートグループが提供され、1つの選択肢としてオフセット0が提供されます。これらは、1つのパブリックIPの背後にあるSD-WANルータ20台と

いう一般的な制限ではありません。この特定の回避策では、20個の固有のオフセット選択肢のみが提供され、NATゲートウェイでこれらを保持して区別する必要があります。オフセットが再利用される場合、一意性スキームとして失敗します。ポートホッピングでは、オフセットアイデンティティは追加されません。

たとえば、124台のルータすべてが、1つの共有アドレスの背後で一意のオフセットを受信できるわけではありません。パブリックアドレスグループあたり最大20台のルータを使用するオフセットのみの方式では、少なくとも7つのグループが必要です。この計算は、NATの動作の検証の代わりとなるキャパシティの推奨事項ではありません。変換されたポートの書き換えやオーバーラップを行うNATゲートウェイでは、ルータの数が少ない場合でも、この方式を無効にすることができます。

送信元ポートの保持を無効にするか、ランダム割り当てを有効にする

保存設定のないポリシーを使用すると、多くのスポークが異なるvBondに対して:12346を受信する傾向を減らすことができます。ランダム化された割り当てにより、これらのマッピングをより大きなポートプールに分散できます。

ただし、いずれの変更もコリジョン回避を保証しません。ランダム化されたアロケータは、以前のハブの状態が残っている間も、引き続き、宛先を越えてポートを再利用したり、ハブへのマッピングを変更したり、期限切れのエンドポイントを再割り当てしたりできます。割り当てがサブスクライバと宛先で一意であるかどうかを確認します。ランダムなポートだけでは一意性は確立されません。

NATゲートウェイの送信元ポートの保持と割り当ての設定を、そのオペレータと一緒に確認します。該当する設定はNATプラットフォームによって異なります。変換レコードとパケットキャプチャを使用して、vBondとハブの両方に対するマッピング結果を確認します。

変換されたポートは、元のSD-WANベースポートの範囲外にある可能性があります。送信元ポートが変更されただけで正規の変換トラフィックが拒否されないように、ファイアウォールルールを確認します。

パブリックアドレスの分離またはポートブロックの予約

変換された専用のアドレス、またはスポークごとに予約された重複しないポートブロックにより、異なるアクティブスポークに対して別々にアドバタイズされたエンドポイントを提供できます。ポリシーは、ディスカバリパスとハブパス、および冗長NATノード全体に一貫して適用する必要があります。追加のアドレス空間とプロバイダーサポートが必要になる場合があります。マッピングの廃止後も動的な再利用が可能ですが、ハブ・セッションのライフサイクル管理を適切に行う必要があります。

代替トランスポートの保持と調整の回復

検証済みのMPLSまたはその他の適格なパスは、影響を受けるインターネットパスを調査する間、サービスを維持できます。共有NATアドレスをブロックするアンダーレイアクセスコントロールリスト(ACL)は、抑止のみを提供します。つまり、そのアドレスの背後にあるすべてのスポークのインターネット冗長性を削除でき、OMPアドバタイズメントを自動的に撤回することはできません。

コントロールクリア、インターフェイスのリセット、NATステートクリア、またはハブのリロードによって、状態が一時的に再作成される場合があります。「なし」は再発の防止を保証します。最初の診断ステップとしてブロード・クリアを実行しないでください。証拠を保存し、可能な限り割り当て条件を修正し、TACとの最小の回復措置に合意します。回復したパス上の継続したBFDアップおよび双方向トラフィックを検証します。

関連ソフトウェア不具合

Cisco Bug ID	この記事に関連する範囲
CSCww45867	異なるピアが同じパブリックTLOCエンドポイントをアドバタイズする場合のハブセッション所有権の制限に関するドキュメント。このレコードは、ソフトウェアの修正ではなく、ドキュメントの更新を追跡します。
CSCwx12986	自動ポートホップ/NATラーニングシーケンスでは、パブリックTLOCフィールドにプライベートタプルを残すことができます。ブランチが同じプライベートIP：ポートを再利用すると、誤ったパブリックエンドポイントがピア間で一致する可能性があります。このプロデューサ不具合を修正しても、有効な変換されたパブリックポートの再使用は妨げられません。
CSCwx48929	ディスカバリが同じvBondに戻ると、NAT分類が誤ってEになる可能性があります。正しい分類は、有効なdifferent-vBondポート再利用シナリオを妨げません。

Cisco Bug ID CSCwx12986およびCisco Bug ID CSCwx48929の修正では、アドバタイズされたパブリックエンドポイントが一意であることが保証されず、この修正だけでは、この衝突シナリオが排除されません。Cisco Bug ID CSCww45867文書レコードがクローズしても、ソフトウェア動作の変更は確立されません。Cisco TACで該当する制限と緩和策を確認します。

付録：IPv4 VCONTROLトレーラのデコード

このコードをsdwan_vctrl.luaという名前で保存します。診断デコーダでは、パケットの認証やペイロードの復号化は行われません。sdwan_vctrl表示フィルタを使用して、明示的なDecode As選

択を登録します。パケットキャプチャは変更されません。

```
-- Cisco SD-WAN IPv4 VCONTROL trailer decoder.
-- Explicit UDP "Decode As" only; does not decrypt ESP.
-- Scope: complete ordinary IPv4/UDP packets, one IP/UDP layer.

local p = Proto("sdwan_vctrl", "SD-WAN VCTRL (IPv4)")

local f = {
    spi       = ProtoField.uint32("sdwan_vctrl.spi",
                    "Wire SPI including control flags", base.HEX),
    sequence  = ProtoField.uint32("sdwan_vctrl.sequence",
                    "ESP sequence number", base.DEC),
    type      = ProtoField.uint8("sdwan_vctrl.type",
                    "Trailer type", base.HEX),
    tlv_len   = ProtoField.uint16("sdwan_vctrl.tlv_len",
                    "TLV length", base.DEC),
    src_ip    = ProtoField.ipv4("sdwan_vctrl.src_ip",
                    "Carried source IPv4"),
    src_port  = ProtoField.uint16("sdwan_vctrl.src_port",
                    "Carried source UDP port", base.DEC),
    dst_ip    = ProtoField.ipv4("sdwan_vctrl.dst_ip",
                    "Carried destination IPv4"),
    dst_port  = ProtoField.uint16("sdwan_vctrl.dst_port",
                    "Carried destination UDP port", base.DEC),
    total_len = ProtoField.uint16("sdwan_vctrl.total_len",
                    "Complete trailer length", base.DEC)
}
p.fields = f

-- Field extractors must be created outside the dissector.
local get_udp_length = Field.new("udp.length")
local get_ip_version = Field.new("ip.version")
local get_ip_proto   = Field.new("ip.proto")

function p.dissector(tvb, pinfo, tree)
    local lengths = { get_udp_length() }
    local versions = { get_ip_version() }
    local protocols = { get_ip_proto() }

    -- Deliberately skip nested/tunnelled protocol stacks.
    if #lengths ~= 1 or #versions ~= 1 or #protocols ~= 1 then
        return 0
    end
    if tonumber(versions[1].value) ~= 4 or
       tonumber(protocols[1].value) ~= 17 then
        return 0
    end

    -- UDP length includes its 8-byte header.
    local n = tonumber(lengths[1].value) - 8

    -- Require the full declared payload, not a truncated fragment.
    if n < 25 or tvb:reported_len() ~= n or tvb:len() ~= n then
        return 0
    end

    -- Exact BFD + VCTRL control-bit combination.
    local spi = tvb(0, 4):uint()
    if math.floor(spi / 0x10000000) ~= 0xA then
```

```
        return 0
    end

    -- Trailer is at the end of UDP payload, not Ethernet frame.
    local o = n - 17

    if tvb(o, 1):uint() ~= 0x81 or
       tvb(o + 1, 2):uint() ~= 15 or
       tvb(o + 15, 2):uint() ~= 17 then
        return 0
    end

    local root = tree:add(p, tvb())
    root:add(f.spi,      tvb(0, 4))
    root:add(f.sequence, tvb(4, 4))

    local trailer = root:add(tvb(o, 17),
                             "Visible VCONTROL trailer")
    trailer:add(f.type,      tvb(o,      1))
    trailer:add(f.tlv_len,   tvb(o + 1,  2))
    trailer:add(f.src_ip,    tvb(o + 3,  4))
    trailer:add(f.src_port,  tvb(o + 7,  2))
    trailer:add(f.dst_ip,    tvb(o + 9,  4))
    trailer:add(f.dst_port,  tvb(o + 13, 2))
    trailer:add(f.total_len, tvb(o + 15, 2))

    pinfo.cols.protocol = "SDWAN-VCTRL"
    return n
end

DissectorTable.get("udp.port"):add_for_decode_as(p)
```

関連情報

- RFC 7857セクション3：ポート重複動作
- RFC 4787：ユニキャストUDPに対するNATの動作要件
- RFC 5737：ドキュメント用に予約されたIPv4アドレスブロック
- Cisco Catalyst SD-WAN設計ガイド
- Cisco IOS XE Catalyst SD-WANシステムコマンド：port-offset
- Cisco SD-WANポートオフセットの設定
- Wiresharkコマンドラインリファレンス
- TSharkコマンドラインリファレンス
- テクニカル サポートとドキュメント – Cisco Systems

翻訳について
シスコは世界中のユーザにそれぞれの言語でサポート コンテンツを提供するために、機械と人による翻訳を組み合わせて、本ドキュメントを翻訳しています。ただし、最高度の機械翻訳であっても、専門家による翻訳のような正確性は確保されません。シスコは、これら翻訳の正確性について法的責任を負いません。原典である英語版（リンクからアクセス可能）もあわせて参照することを推奨します。