Direct Server Return の設定

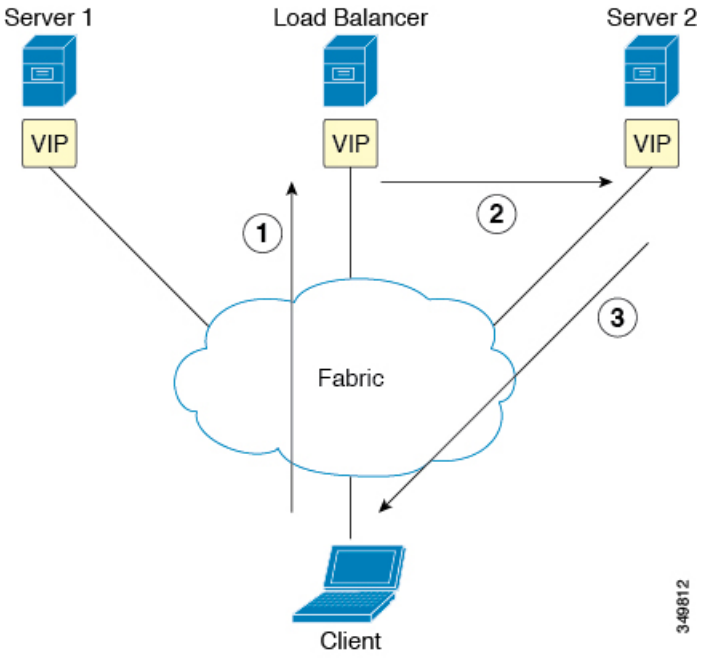
Direct Server Return について

Didrect Server Return 機能により、サーバはロード バランサを通過する必要なく、クライアントに直接応答できます。これにより、サーバからクライアントへのパスにおけるボトルネックが解消されます。従来のロードバランサの導入では、ロードバランサは、クライアントとサーバとの通信のパス（クライアントからサーバへの要求パスとサーバからクライアントへの応答パスの両方）に存在します。クライアントからサーバ方向の要求内のデータの量は比較的少ないものの、サーバからクライアントへの応答トラフィックはかなり大きく、クライアントからサーバへの要求データの約10倍になります。この大量の応答トラフィックがあるパス内のロード バランサがボトルネックになり、通信に悪影響を及ぼします。

Direct Server Return の導入では、ロード バランサとサーバとで仮想 IP アドレスが共有されます。クライアントは、ロード バランサに到達することを目的とした仮想 IP アドレスに常に要求を送信し、また、サーバからクライアントへの直接応答ではこの仮想 IP アドレスを送信元アドレスとして使用します。IP 送信元アドレスのデータパスの取得が有効になっているCisco Application Centric Infrastructure（ACI）は、サーバからクライアントへのトラフィックの仮想 IP アドレスを取得する際に問題を引き起こし、クライアントからロード バランサへの要求トラフィックを途絶させることになります。Direct Server Return の導入を適切に動作させるには、ACI ファブリックは通信中のエンドポイント間の要求と応答のトラフィックを目的の宛先に正しく配信されるようにする必要があります。これには、リーフ上でのデータパス IP アドレスの取得を、クライアントからロード バランサへのトラフィック、ロード バランサからサーバへのトラフィック、およびサーバからクライアントへのトラフィックに割り込みを生じさせないように制御することが必要です。

次の図に、Direct Server Return の導入のデータ パスを示します。

図 *1 : Direct Server Return* の全体的なフロー



1. ロード バランサとすべてのバックエンド サーバが仮想 IP アドレスで設定されています。ロード バランサのみが、この仮想 IP アドレス宛の Address Resolution Protocol（ARP）要求に応答します。クライアント要求のロード バランシング後に、ロード バランサはパケット内の宛先 MAC アドレスを書き換えて、その MAC アドレスをバックエンド サーバの 1 つに転送します。
2. 仮想 IP アドレスはバックエンド サーバ上に設定されますが、ARP が無効になっているため、この仮想 IP アドレス宛の ARP 要求にバックエンド サーバは応答できません。
3. サーバはリターン トラフィックをクライアントに直接送信してロード バランサをバイパスします。

レイヤ 2 の Direct Server Return

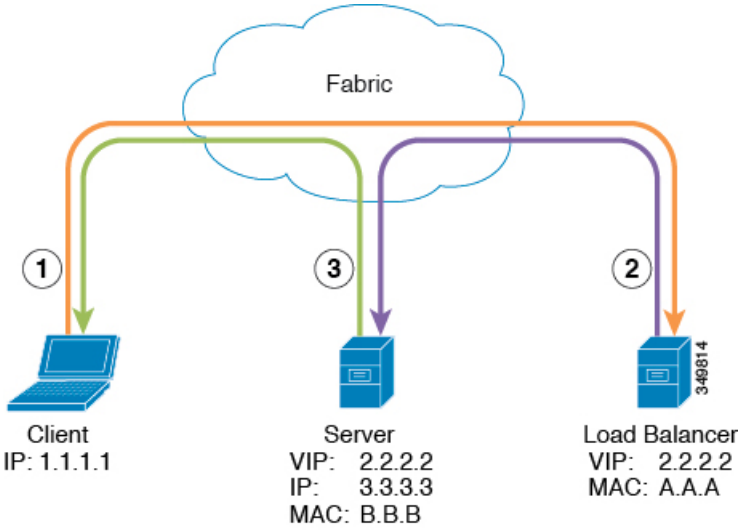
レイヤ 2 の Direct Server Return は一般的な導入または従来型の導入であり、ダイレクト ルーティング、SwitchBack、または nPath とも呼ばれます。この導入では、ロード バランサとサーバで仮想 IP アドレスが共有されます。ロード バランサとサーバはレイヤ 2 隣接である必要があります。レイヤ 2 の Direct Server Return の導入には、次の制限があります。

- サーバ配置の柔軟性が失われる
- クライアントの仮想 IP アドレス要求への Address Resolution Protocol（ARP）応答を抑制するために、追加のサーバ設定が必要になる

- ポート選択はレイヤ 3 で行われ、プロトコルに依存する。ポート選択はレイヤ 2（サーバ通信に対するロード バランサ）で行われない

レイヤ 2 の Direct Server Return の導入には、次のトラフィック フローがあります。

図 *2 :* レイヤ *2* の *Direct Server Return* のトラフィック フロー



1. クライアントからロード バランサへ

Source IP Address	1.1.1.1
Destination IP Address	2.2.2.2
宛先 MAC アドレス	A.A.A

2. ロード バランサからサーバへ

Source IP Address	1.1.1.1
Destination IP Address	2.2.2.2
宛先 MAC アドレス	B.B.B

3. サーバからクライアントへ

Source IP Address	2.2.2.2
Destination IP Address	1.1.1.1
宛先 MAC アドレス	デフォルト ゲートウェイの MAC アドレス

でのレイヤ2DirectServerReturnの導入についてCiscoApplicationCentric Infrastructure

次の情報は、Cisco Application Centric Infrastructure（ACI）でのレイヤ 2 Direct Server Return の導入に当てはまります。

- 仮想 IP アドレス（2.2.2.2）は ACI ファブリック内を移動する
  - 同じ送信元仮想 IP アドレス（2.2.2.2）を持つロード バランサからサーバおよびサーバからクライアントへのトラフィック
  - サーバからクライアントへのトラフィックはルーティングされ、トラフィックはファブリック内のゲートウェイ MAC アドレス宛になる
  - サーバからの送信元 IP アドレスのデータ パスの取得はファブリック内の仮想 IP アドレスに移動する
- 異なる送信元から表示されるクライアント IP アドレス（1.1.1.1）についての問題はない
  - クライアント IP アドレスはファブリック内のクライアントとロード バランサの両方からの送信元 IP アドレスとして表示される
  - ロード バランサとサーバは、レイヤ 2 隣接であり、ロード バランサからサーバへのトラフィックはレイヤ 2 に転送される
  - ファブリック内のレイヤ 2 転送トラフィックからのデータパス IP アドレスの取得はない
  - クライアント IP アドレスがファブリック内のロード バランサからの送信元 IP アドレスとして表示された場合も、クライアント IP アドレスは取得されない

Direct Server Return の設定に関する注意事項と制約事項

Direct Server Return を展開する際には、次の注意事項と制約事項に従ってください:

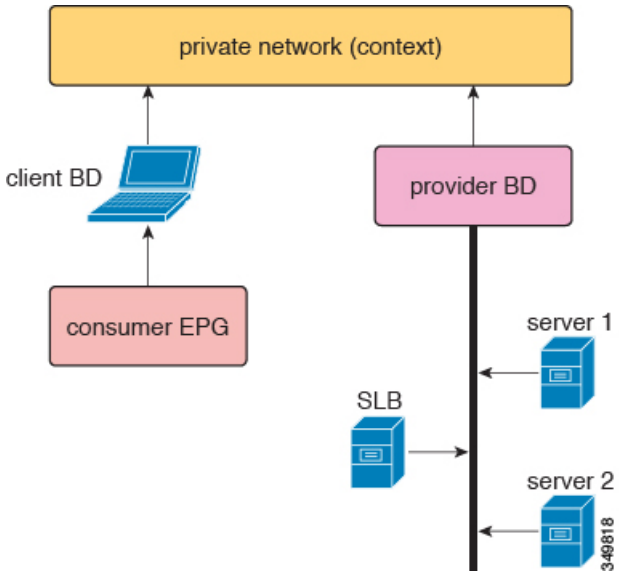
- VRF (VIP が展開される) は、「強制」モードに設定する必要があります。
- VRF は、「入力」の強制を設定する必要があります。
- 共有サービスは、この設定ではサポートされていません。
- EP 移動検出モード: GARP ベースの検出をブリッジ ドメインで有効にする必要があります。
- ユニキャスト ルーティングをブリッジ ドメインで有効にする必要があります。
- VIP がある EPG には、それに関連付けられている契約が必要です (契約はハードウェアの設定を進めます)。

- VRF 下の VZAny 契約は L4 ～ L7 VIP をプログラムしません。契約は EPG 下で引き続き許可されます。
- クライアントから VIP へのトラフィックは、必ずプロキシ スパインを通る必要があります。
- ロード バランサはワンアーム モードにする必要があります。
- サーバとロード バランサ EPG を同じデバイス上に配置するか、ロード バランサ EPG をすべてのサーバ EPG ToR に展開する必要があります。
- サーバ EPG とロード バランサ EPG は、同じブリッジ ドメインにある必要があります。

サポートされている Direct Server Return の設定

次の図に、サポートされている Direct Server Return の設定を示します。

図 *3* : サポートされている *Direct Server Return* の設定



サポートされている設定に次の情報が適用されます。

- サーバ ロード バランサとサーバは同じサブネットとブリッジ ドメインにある
- サーバ ロード バランサは 1 ARM モードで動作する必要があり、サーバ ロード バランサの内部レッグと外部レッグは同じブリッジ ドメインを指している必要がある
- コンシューマ エンドポイント グループとプロバイダー エンドポイント グループは、同じプライベート ネットワークの下にある必要がある。共有サービス設定はサポートされていない

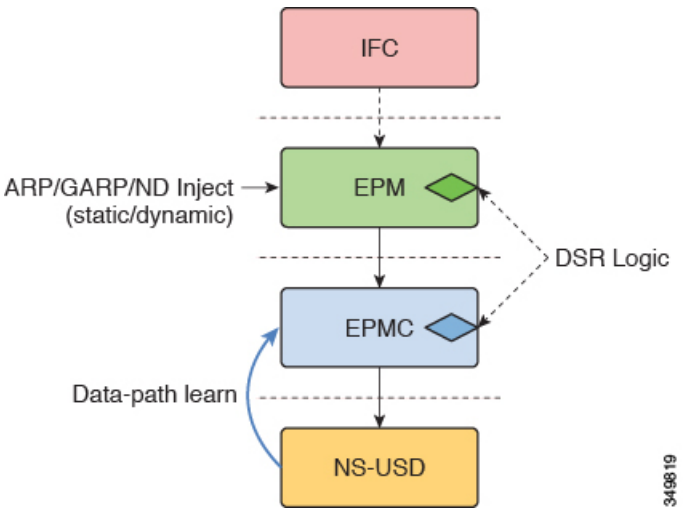
Direct Server Return のアーキテクチャ

ロードバランサとサーバが共有する仮想 IP アドレスをテナントの Virtual Routing and Forwarding（VRF）のファブリック内で静的に設定し、レイヤ 2 Direct Server Return を有効にすることができます。仮想 IP アドレスを静的にすることで、仮想 IP アドレスのデータ パスの取得が阻止されます。スイッチ側のエンドポイント マネージャは、具象モデルの {VRF、VIP、S クラス} タプル形式のポリシー エンジンからの静的設定を処理します。エンドポイント マネージャ（EPM）、エンドポイント マネージャ クライアント（EPMC）および転送シリコンにアーキテクチャは変更されていませんが、一方でこれらのレイヤはすべて変更されており、仮想 IP アドレスの静的設定が許可・維持されています。このアーキテクチャでは、仮想 IPv4 と仮想 IPv6 の両方に対して静的設定が可能です。アドレス解析および初期セットアップ以外は、アーキテクチャ上および設計全体にわたって、仮想 IPv4 と仮想 IPv6 の両方が同じコード パスで処理されます。

Direct Server Return の設計フローと設定フローは EPM/EPMC/USD のフローの一部です。このためのフローは、ノースバウンドからサウスバウンド、つまり、**[policy engine]** > **[EPM]** > **[EPMC]** > **[forwarding silicon]**の場合にのみ存在します。この場合の転送シリコンは North Star です。これは、この目的に対応しているのは North Star のローカル ステーション テーブルに限られることが理由です。同じエンドポイントの作成/変更/削除フローを、追加の静的 IP アドレス エンドポイント フラグと共に使用します。

新しい IP アドレス エンドポイントのすべての取得要求は、静的仮想 IP アドレス エンドポイント チェックを受けます。取得/処理要求がすでに存在する {VRF、VIP、S クラス} タプルに対するものである場合は、Direct Server Return の前処理コードによって変更前処理が行われ、適切なフラグを使用した一般的なエンドポイント処理に戻されます。

図 *4:* スイッチ側の処理



次のリストに、Direct Server Return の設計ポイントに関する概要を示します。

- すべての仮想 IPv4 および IPv6 の追加/変更/削除設定は、エンドポイント マネージャによって処理される

  - Direct Server Return は、プレフィクスではなく、完全な仮想 IP アドレス（/32、/128）を使用する
  - アドレス ファミリおよびアドレスの最上位レベルのセットアップ以外、コード パスは Direct Server Return 用にマージされる

- EPM と EPMC は完全な（/32 または /128）仮想 IP アドレスを North Star ローカル ステーション テーブルの送信元アドレスにインストールする
  - キー/データは、設定された 3 タプル {VRF、VIP、S クラス} 情報から取得する
  - ローカル ステーション テーブルの送信元アドレスに「静的」としてエントリが挿入される

- EPM は ARP/GARP/ND IP-MAC バインディングと MAC の取得を通じてロード バランサのエンドポイントを検出できる
- EPM と EPMC は、{VRF、VIP} タプルの疑似 North Star データパスの取得を阻止する
- EPM と EPMC では、取得したエントリの S クラスがポリシー エンジンで設定した {VRF、VIP、S クラス} タプルと一致しなければ、IP-MAC バインディング アソシエーションを許可しない。これは、ARP/GARP/ND パスとデータパス取得パスの両方に適用される
- EPM は、COOP へのロード バランサの（ARP/GARP/ND を通じた）検出伝播を代替しない
- エントリの S クラスが一致しない場合、EPM と EPMC は既存の取得済みエントリ（ARP/GARP/ND）を設定時にクリーンアップする
- 仮想 IP アドレスを設定すると、EPM と EPMC は、データパスの取得を通じて作成された同じ {VRF、VIP} タプルの既存のエントリを常にクリーンアップする
- ARP/ND/MAC エージングはこれらの変更に対応していないが、EPM と EPMC が、{VRF、VIP} タプルの設定が削除されない限り、ローカル ステーション テーブルに維持されている静的エントリは削除されない
- この機能を実装する際に、既存のエントリを削除するのではなく、同じエントリ設定をポリシー エンジンから取得する場合、ARP/GARP/ND を通じて取得した {VRF、VIP} タプルを保持するアプローチを取る。これは、既存のエントリの S クラスが設定されたエントリの S クラスと同じである場合に限る。このアプローチにより、エントリの削除が原因で発生するファブリック全体にわたる大規模な変動を回避する
- S クラスとエントリに関連するその他の情報は IP アドレス情報の一部として保持される。つまり、情報はエンドポイント レベルではなく、エンドポイントの IP アドレスレベルで保持される
- {BD、仮想 IP のプレフィックス、S クラス} タプルとポリシー エンジンで設定した {VRF、VIP、S クラス} タプル間に重複がある場合は、{VRF、仮想 IP、S クラス} タプルが優先される

静的なサービス導入のための Direct Server Return の XML POST の例

次に、Direct Server Return の静的なサービス導入の例を示します。

```
<fvAp name="dev">
      <fvAEPg name="loadbalancer">
        <fvRsDomAtt tDn="uni/phys-{{tenantName}}"/>
        <fvRsBd tnFvBDName="lab"/>
        <fvVip addr="121.0.0.{{net}}" />
        <fvRsPathAtt tDn="topology/pod-1/paths-104/pathep-[eth1/1]" encap="vlan-33" />
        <fvRsProv tnVzBrCPName="loadBalancer"/>
        <fvRsCons tnVzBrCPName="webServer"/>
      </fvAEPg>
        <fvAEPg name="webServer">
          <fvRsDomAtt tDn="uni/phys-{{tenantName}}"/>
            <fvRsBd tnFvBDName="lab"/>
          <fvRsPathAtt tDn="topology/pod-1/paths-101/pathep-[eth1/1]" encap="vlan-34"/>

            <fvRsProv tnVzBrCPName="webServer"/>
        </fvAEPg>
      <fvAEPg name="client">
        <fvRsDomAtt tDn="uni/phys-{{tenantName}}"/>
        <fvRsBd tnFvBDName="lab"/>
        <fvRsPathAtt tDn="topology/pod-1/paths-103/pathep-[eth1/4]" encap="vlan-1114"/>
        <fvRsCons tnVzBrCPName="loadBalancer"/>
      </fvAEPg>
</fvAp>
```

L4-L7 VIP の EPG が展開されているか、コントラクトの方向に関わらず L4-L7 VIP の EPG とのコントラクトを持つ EPG が展開されているすべての top-of-rack スイッチ（ToR）にダイレクト サーバ リターン設定がダウンロードされます。この例では、ダイレクト サーバ リターン仮想 IP アドレス設定が ToR ノード 101、103、104 にダウンロードされます。ノード 104 には設定された L4-L7 VIP のロードバランサ EPG があり、ノード 101 および 103 には Web サーバまたはクライアント EPG があり、ロードバランサ EPG へのコントラクトを有します。

ダウンロードされたダイレクト サーバ リターン設定を持つすべての ToR は、データパスから L4-L7 VIP アドレスを学習せず、その他の EPG から L4-L7 VIP アドレスを学習しません（ARP/GARP/ND の場合でも）。たとえば、L4 - L7 VIP アドレスは、コントロール プレーンを介してロードバランサ EPG からのみ学習します。Web サーバ EPG から誤って L4 L7 VIP を学習してしまうことを防ぐのに役立ちます（たとえば web サーバで ARP を抑制することを忘れた場合）。

静的なサービス導入のための Direct Server Return

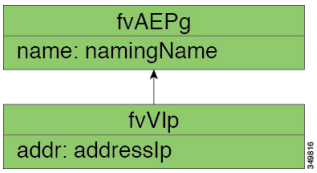
静的なサービス導入モードでは、適切なアプリケーション エンドポイント グループとコントラクトをホップごとに作成することによって、サービス フローを設定します。

静的なサービス導入の論理モデル用の Direct Server Return

アプリケーション エンドポイント グループ（fvAEPg）の下に fvVip オブジェクトを使用することによって、ロード バランサが使用する仮想 IP アドレスを設定できます。

次の図に、静的なサービス導入の論理モデルを示します。

図 *5*：静的なサービス導入の論理モデル



サービス グラフを挿入するための Direct Server Return

Cisco Application Centric Infrastructure（ACI）は、ベンダー パッケージとサービス グラフを使用してサービスの挿入を自動化します。このモードでは、サービスデバイスのレッグに対して作成されるエンドポイント グループ（内部および外部エンドポイント グループなど）が、オペレータによる設定を必要とせずに、ACI によって作成されます。

サービス グラフの挿入では、次の XML POST の例に示すように、サービス デバイスの適切な論理インターフェイス コンテキストの下に仮想 IP アドレスを設定する必要があります。

```
<vnsLDevCtx ctrctNameOrLbl="webCtrct"
                        graphNameOrLbl="G1"
                        nodeNameOrLbl="SLB">

    <vnsRsLDevCtxToLDev tDn="uni/tn-coke/lDevVip-InsiemeCluster"/>

    <vnsLIfCtx connNameOrLbl="inside">
        <vnsRsLIfCtxToBD tDn="uni/tn-coke/BD-cokeBD1"/>
        <vnsRsLIfCtxToLIf tDn="uni/tn-coke/lDevVip-InsiemeCluster/lIf-inside"/>
    </vnsLIfCtx>

    <vnsLIfCtx connNameOrLbl="outside">
        <vnsRsLIfCtxToBD tDn="uni/tn-coke/BD-cokeBD1"/>
        <vnsRsLIfCtxToLIf tDn="uni/tn-coke/lDevVip-InsiemeCluster/lIf-outside"/>
        <vnsSvcVip addr="9.9.9.9" />
        <vnsSvcVip addr="11.11.11.11" />
    </vnsLIfCtx>
</vnsLDevCtx>
```

この要求の例では、2 つの仮想 IP アドレス（9.9.9.9 と 11.11.11.11）をサーバ ロード バランサの外部レッグ上に設定します。仮想 IP アドレスの定義は、静的な Direct Server Return 設定と同様に、エンドポイント グループの下ではなく、LIfCtx の下になります。これは、静的サービスの導入の場合とは異なり、サービス グラフの場合は、オペレータにデバイス レッグのエンドポイント グループへの直接アクセス権がないためです。

Direct Server Return 共有レイヤ 4 ～ レイヤ 7 サービスの設定

サービスデバイスを共通のテナントまたは管理テナントに設定した場合、暗黙モデルには若干の違いがあります。vnsEPpInfo の代わりに、サービス仮想 IP アドレスの更新管理対象オブジェクトが vnsREPpInfo の子として作成されます。1 つの vnsSvcEpgCont の管理対象オブジェクトが vnsRsEPpInfo ごとに作成されて複数のテナント間で共有 SvcVip を追跡します。

Direct Server Return 用の Citrix サーバ ロード バランサの設定

次に、Direct Server Return 用に Citrix サーバ ロード バランサを設定する方法の概要を示した手順を説明します。

ステップ 1 バックエンド サーバがパケットを受け入れるようにバックエンド サーバのループバックに仮想 IP アドレスを設定します。

ステップ 2 バックエンド サーバの仮想 IP アドレスに対する Address Resolution Protocol（ARP）応答を無効にします。

ステップ 3 必要に応じて、ロードバランシング仮想サーバにバインドされたサービスのプロキシ ポートを無効にします。プロキシ ポートはデフォルトで無効になっています。

ステップ 4 ロード バランシング仮想サーバの m パラメータを「MAC」に設定します。

ステップ 5 グローバルか、またはサービスごとに USIP モードを有効にします。

ステップ 6 「L3」モード、「USNIP」モード、および「MBF」モードを有効にします。

ステップ 7 バックエンド サーバのルートを直接インターネットに到達できるように設定します。

Direct Server Return 用の Linux サーバの設定

次に、Direct Server Return 用に Linux サーバを設定する方法の概要を示した手順を説明します。

ステップ 1 次のコンテンツを使用し、Centos 内に /etc/sysconfig/network-scripts/ifcfg-lo ファイルを作成して、ループバック インターフェイス上に仮想 IP アドレスを設定します。

```
DEVICE=lo:1
IPADDRESS=10.10.10.99
NETMASK=255.255.255.255
NETWORK=10.10.10.99
BROADCAST=10.10.10.99
ONBOOT=yes
NAME=loopback
```

この例では、10.10.10.99 が仮想 IP アドレスです。

ステップ 2 クライアント要求への応答に使用するサーバ インターフェイスの arp_ignore と arp_announce の値を設定します。

```
echo 1 > /proc/sys/net/ipv4/conf/eth1/arp_ignore
echo 2 > /proc/sys/net/ipv4/conf/eth1/arp_announce
```

この例では、`eth1` がクライアント要求への応答に使用するサーバ インターフェイスです。

ARP の設定の詳細については、次の Linux 仮想サーバの Wiki ページを参照してください。

http://kb.linuxvirtualserver.org/wiki/Using_arp_announce/arp_ignore_to_disable_ARP