さくらのクラウドにおける L2ネットワークの仕様について

～DSR構成のLVS構築時にお気をつけいただくために～

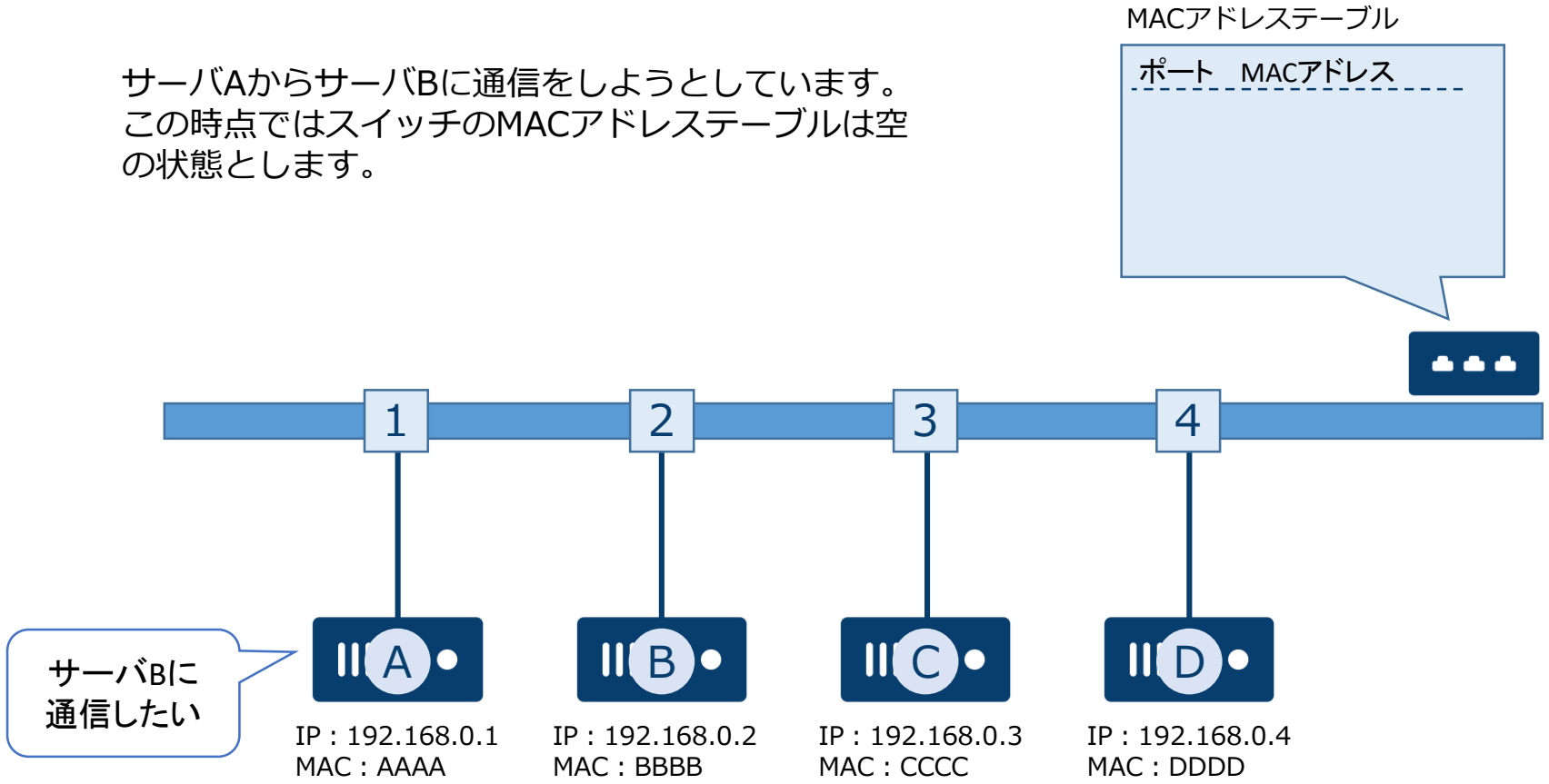
DAY
2018/9

COMPANY
さくらインターネット株式会社

DEPARTMENT
クラウドチーム

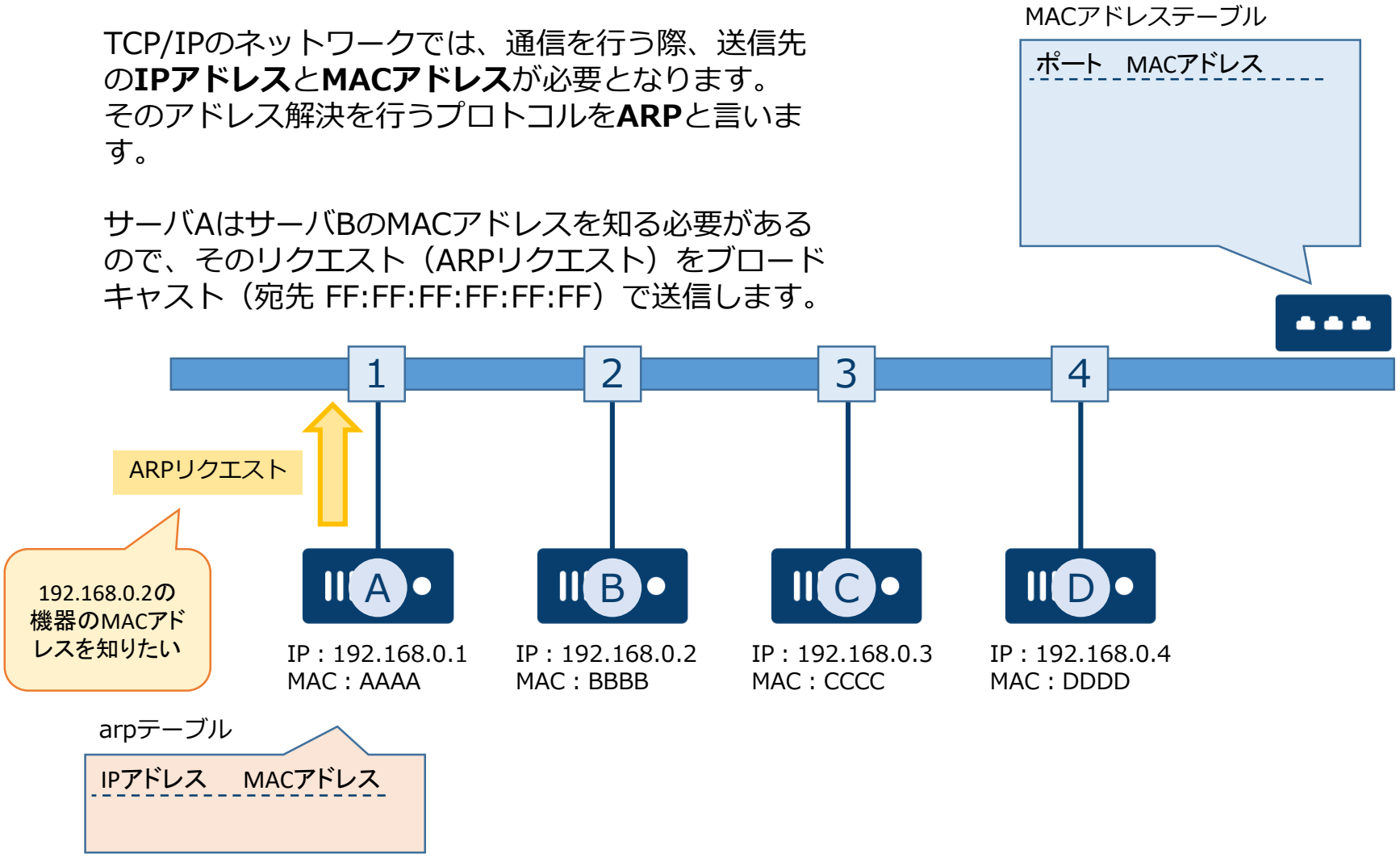
NAME

サーバAからサーバBに通信をしようとしています。
この時点ではスイッチのMACアドレステーブルは空の状態とします。

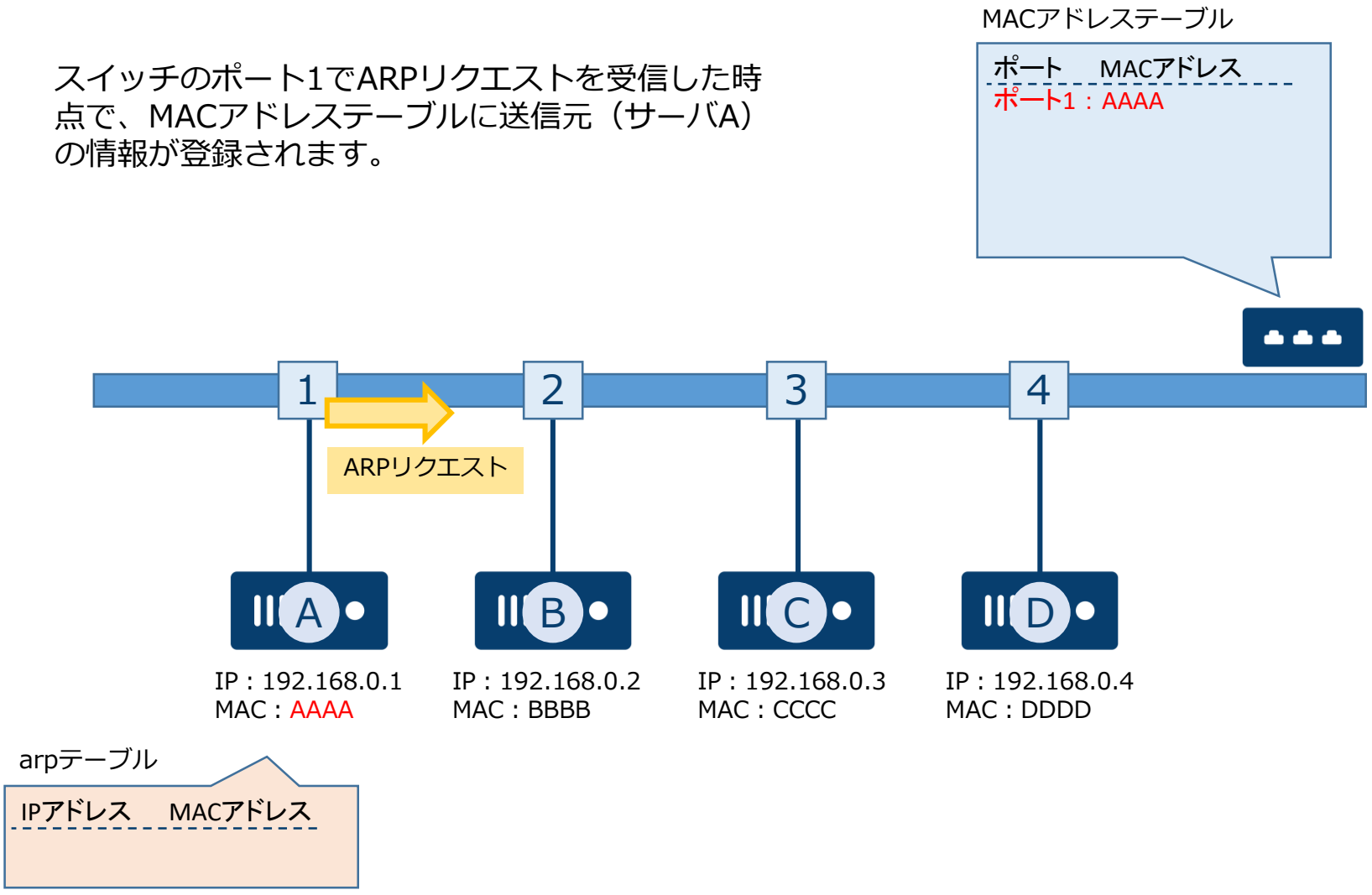


TCP/IPのネットワークでは、通信を行う際、送信先の**IPアドレス**と**MACアドレス**が必要となります。
そのアドレス解決を行うプロトコルを**ARP**と言います。

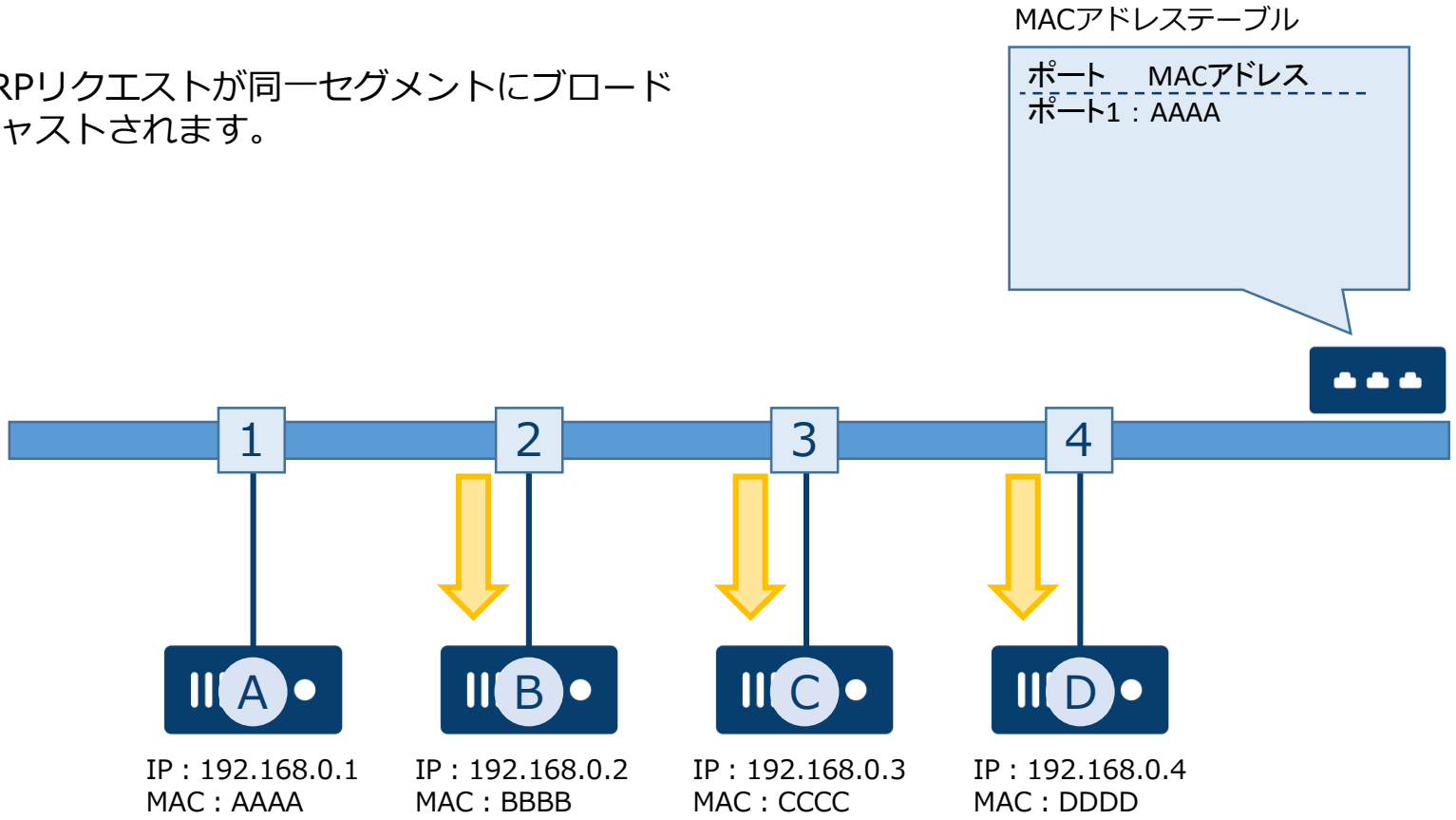
サーバAはサーバBのMACアドレスを知る必要があるので、そのリクエスト（ARPリクエスト）をブロードキャスト（宛先 FF:FF:FF:FF:FF:FF）で送信します。



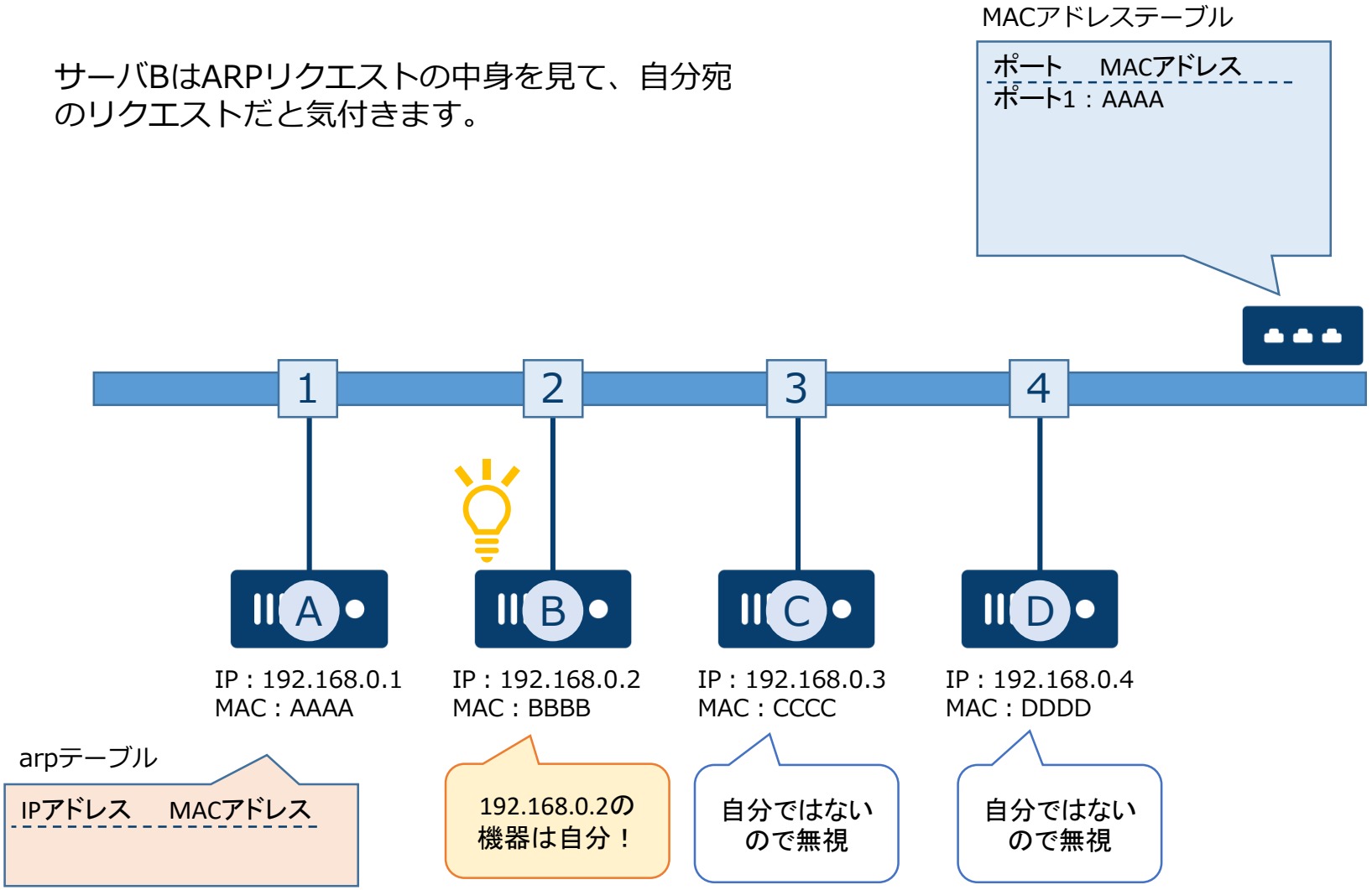
スイッチのポート1でARPリクエストを受信した時点で、MACアドレステーブルに送信元（サーバA）の情報が登録されます。



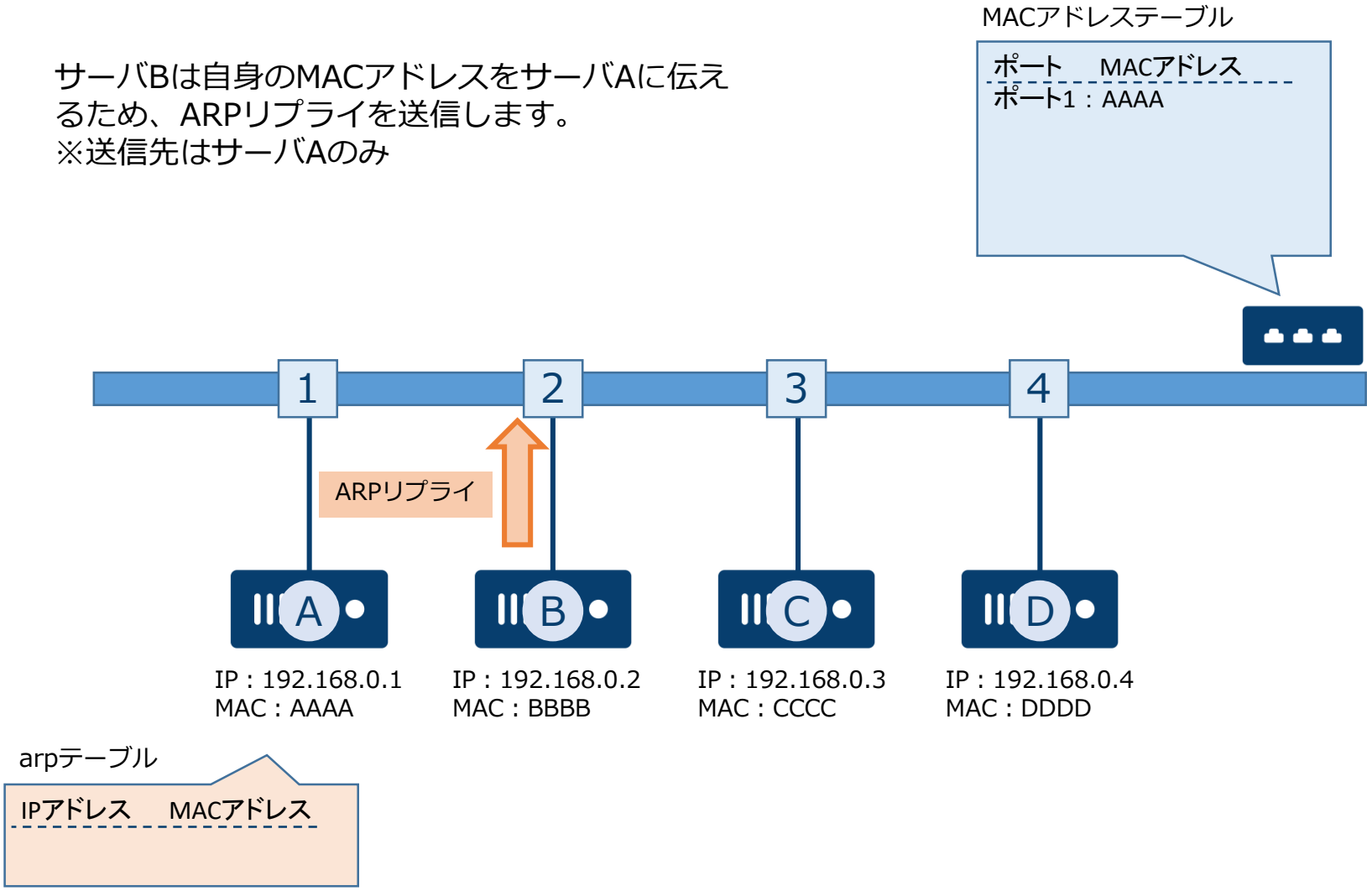
ARPリクエストが同一セグメントにブロードキャストされます。



サーバBはARPリクエストの中身を見て、自分宛のリクエストだと気付きます。



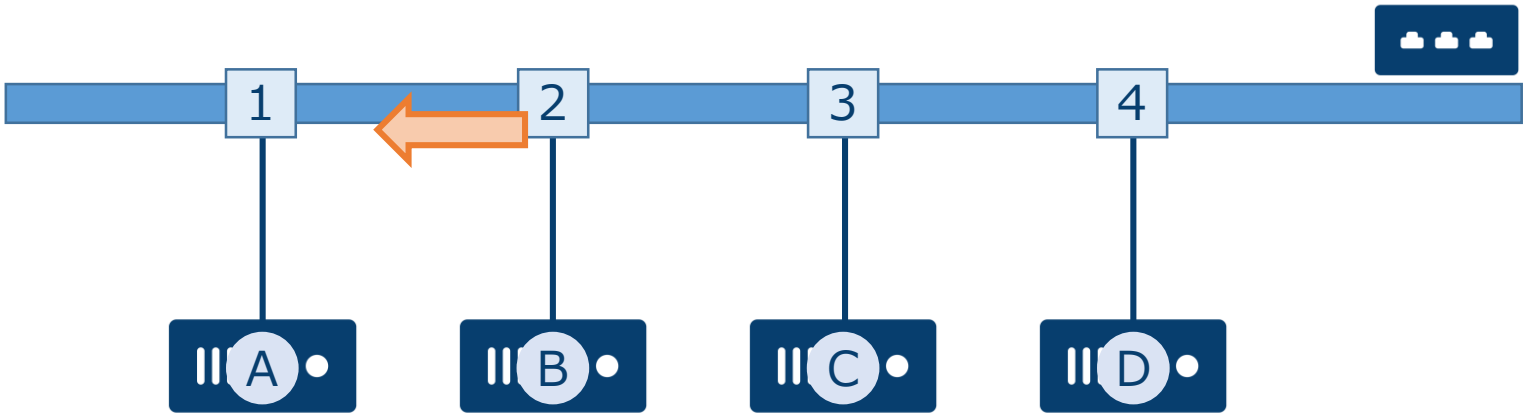
サーバBは自身のMACアドレスをサーバAに伝えるため、ARPリプライを送信します。
※送信先はサーバAのみ



スイッチのポート2でARPリプライを受信した時点で、MACアドレステーブルに送信元（サーバB）の情報が登録されます。

MACアドレステーブル

ポート	MACアドレス
ポート1	AAAA
ポート2	BBBB



IP：192.168.0.1
MAC：AAAA

IP：192.168.0.2
MAC：BBBB

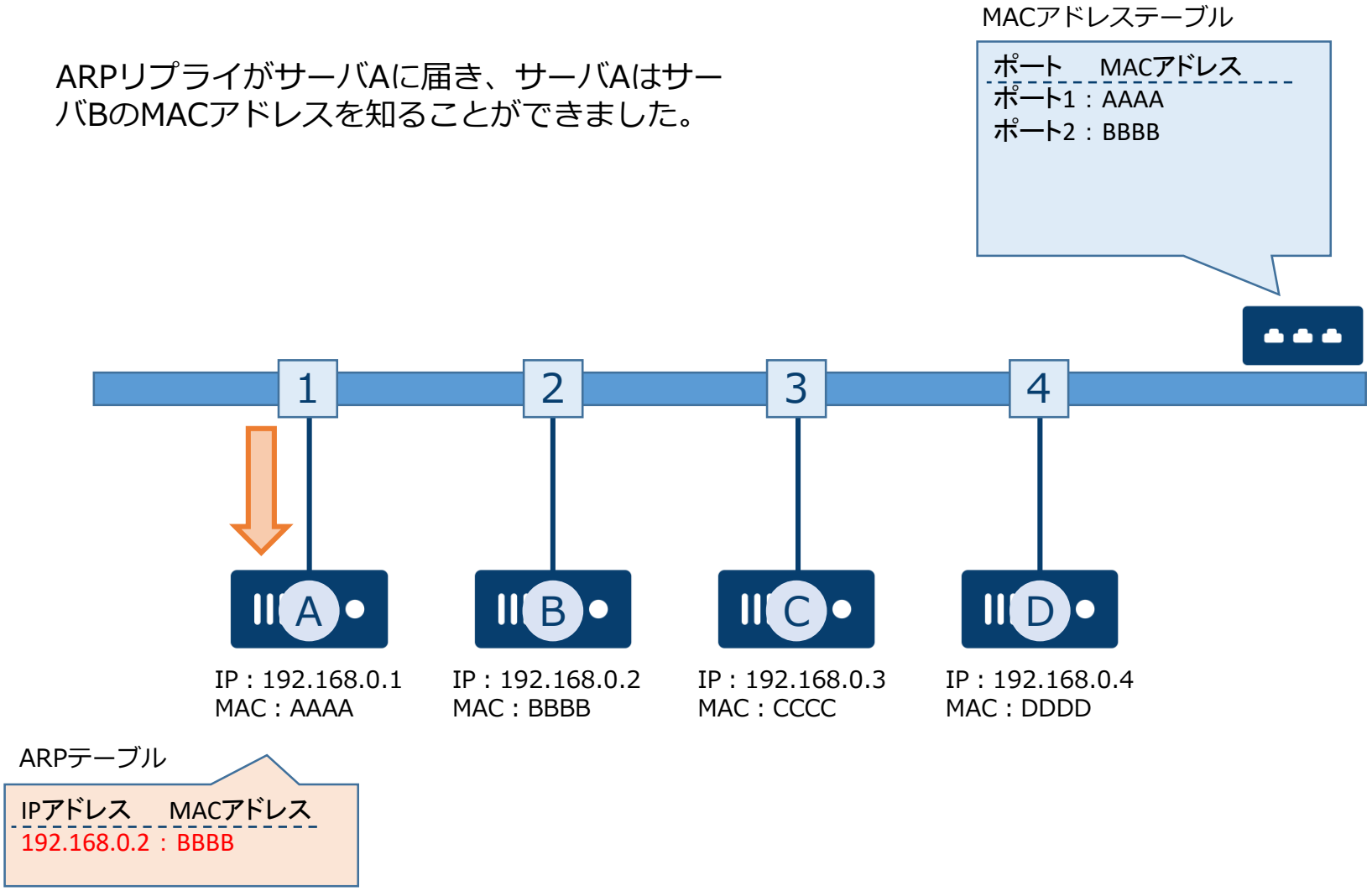
IP：192.168.0.3
MAC：CCCC

IP：192.168.0.4
MAC：DDDD

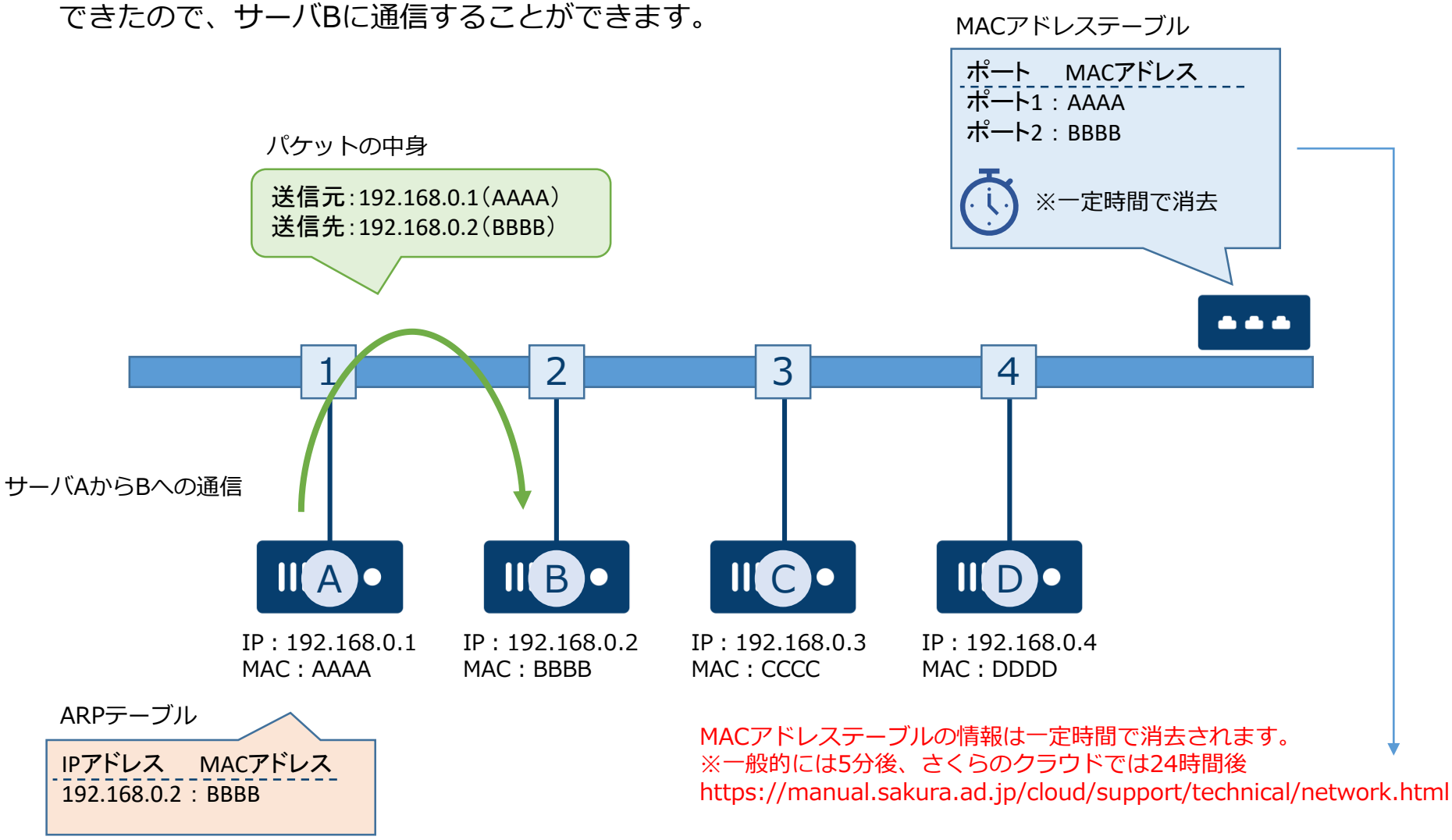
arpテーブル

IPアドレス	MACアドレス

ARPリプライがサーバAに届き、サーバAはサーバBのMACアドレスを知ることができました。



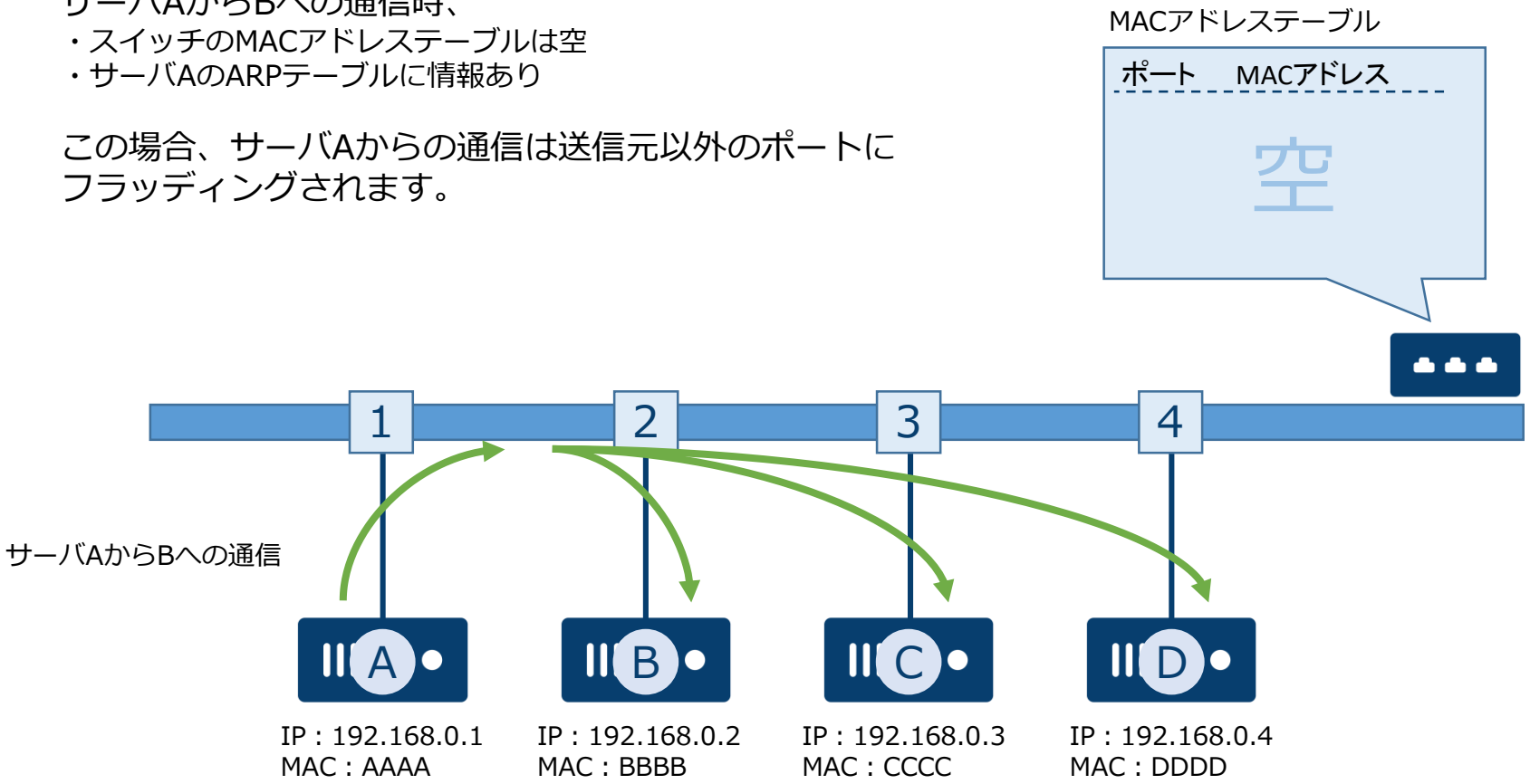
サーバAはサーバB（送信先）のMACアドレスを知ることが
できたので、サーバBに通信することができます。



MACアドレステーブルの情報は一定時間で消去されます。
※一般的には5分後、さくらのクラウドでは24時間後
https://manual.sakura.ad.jp/cloud/support/technical/network.html

サーバAからBへの通信時、
・スイッチのMACアドレステーブルは空
・サーバAのARPテーブルに情報あり

この場合、サーバAからの通信は送信元以外のポートに
フラッディングされます。



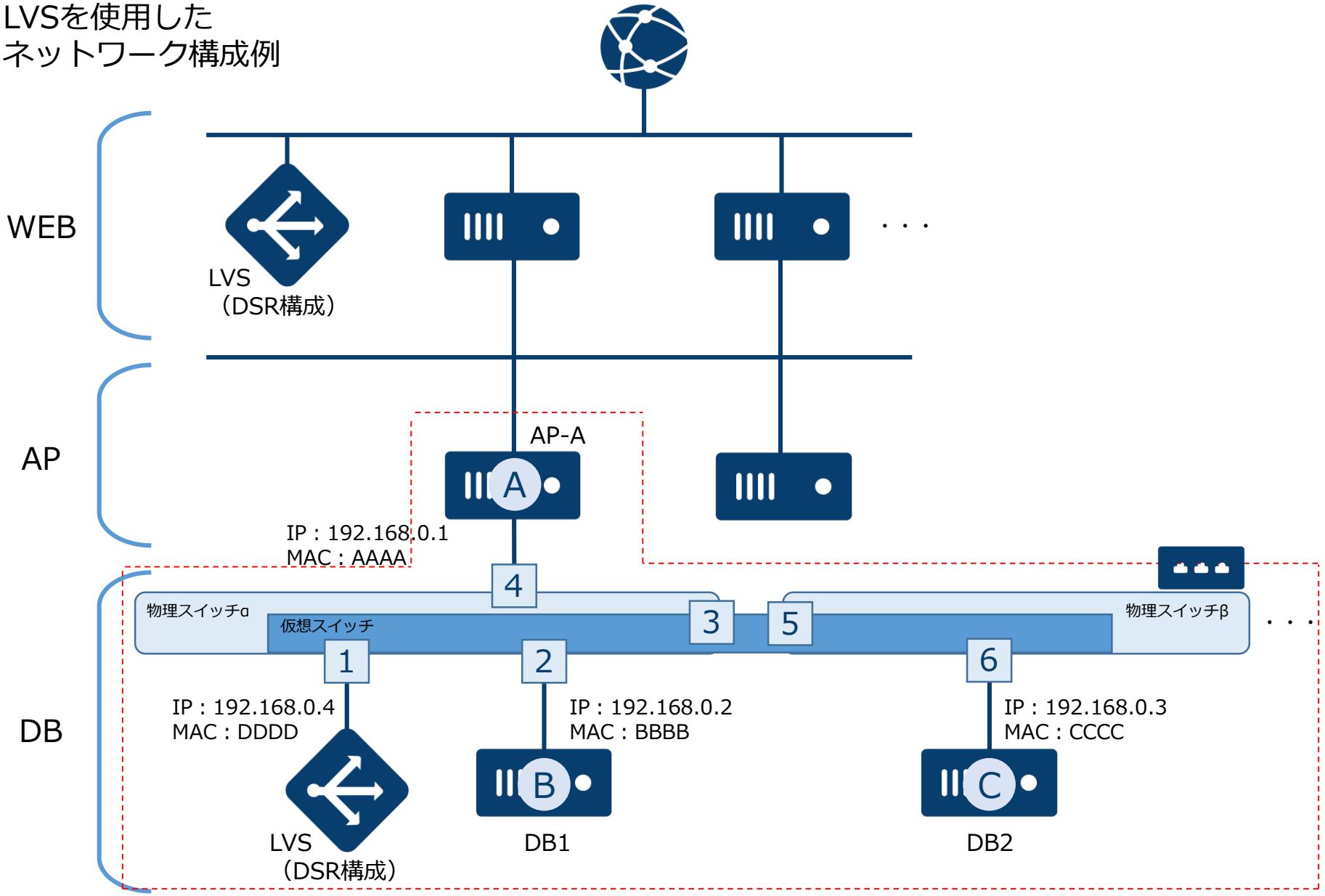
ARPテーブル

IPアドレス	MACアドレス
192.168.0.2	BBBB

フラッディングの通信（1対多）が発生した場合、回線が輻輳し他のお客様へ影響を与える可能性があります。そのため、さくらのクラウドではフラッディングパケットの総量帯域が制限されています。
https://manual.sakura.ad.jp/cloud/support/technical/network-settings.html#dsr

DSR構成のLVSを構築するケース

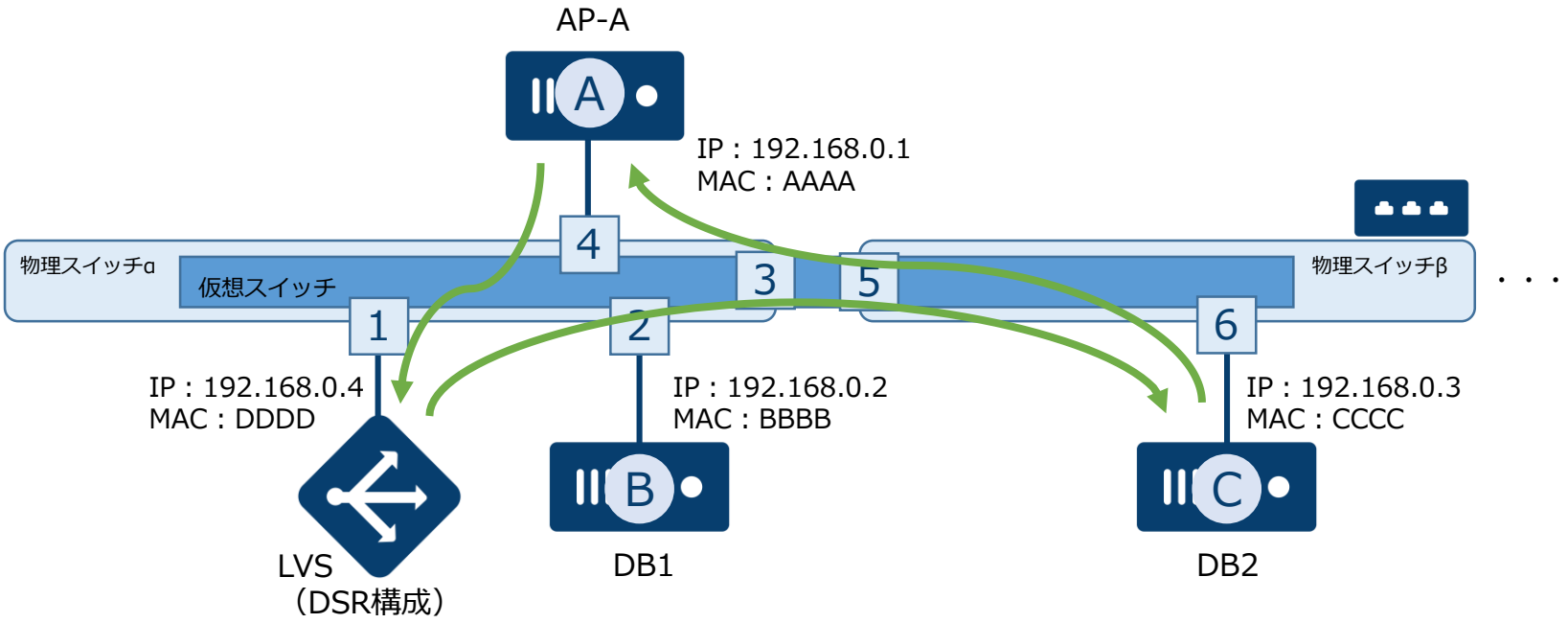
LVSを使用した
ネットワーク構成例



APサーバからDBへの通信をLVSで分散する部分を抽出した図になります。

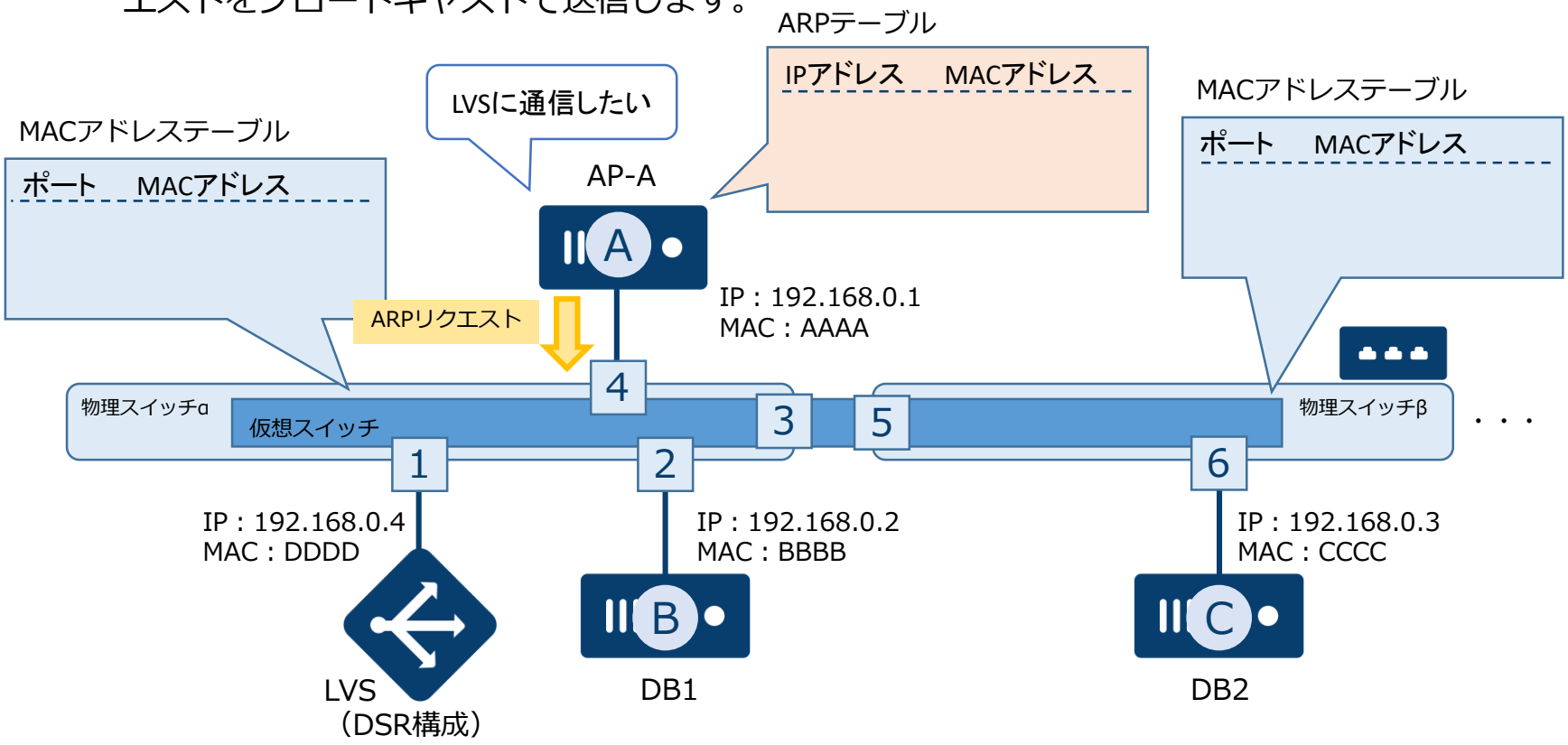
なお、今回はDSR構成ですので、DBサーバに分散された通信は、LVSを経由せず直接APサーバに返ります。

※以下は、LVSがサーバCを分散先としたケース



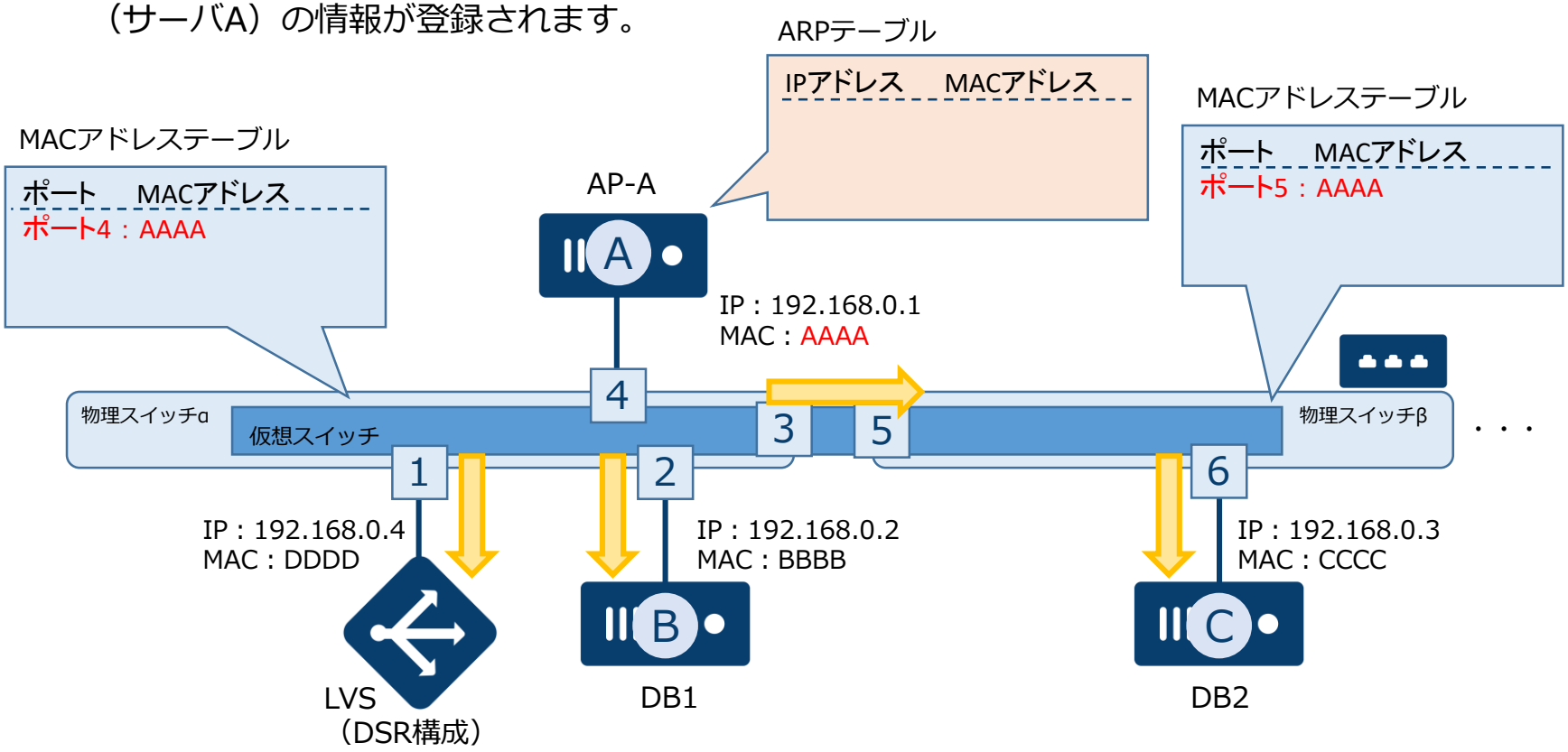
通信初期の段階では、サーバAのARPテーブル、
各物理スイッチのMACアドレステーブルは空の
状態とします。

まず、サーバAはLVSに通信するためにARPリク
エストをブロードキャストで送信します。

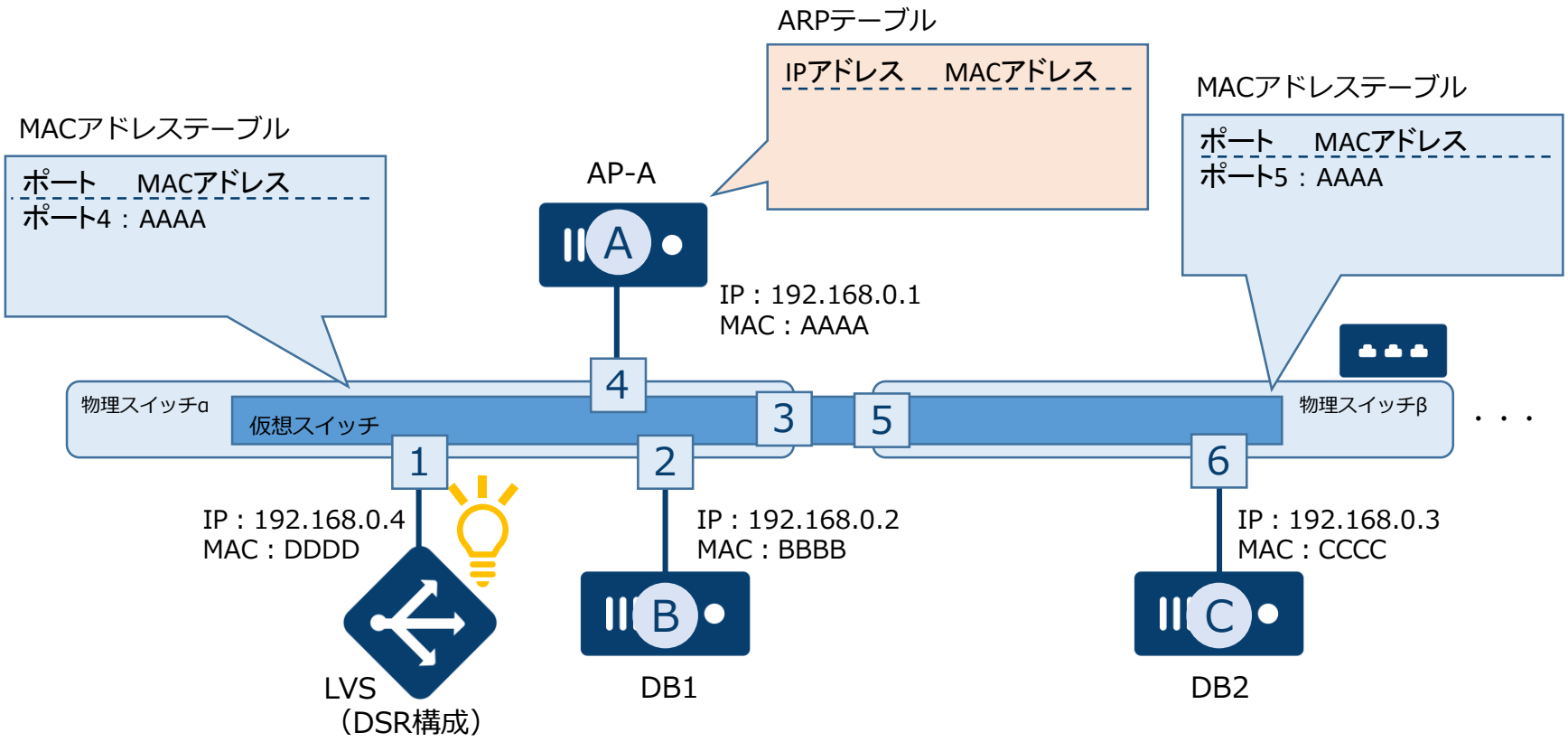


ARPリクエストが同一セグメントにブロードキャストされます。

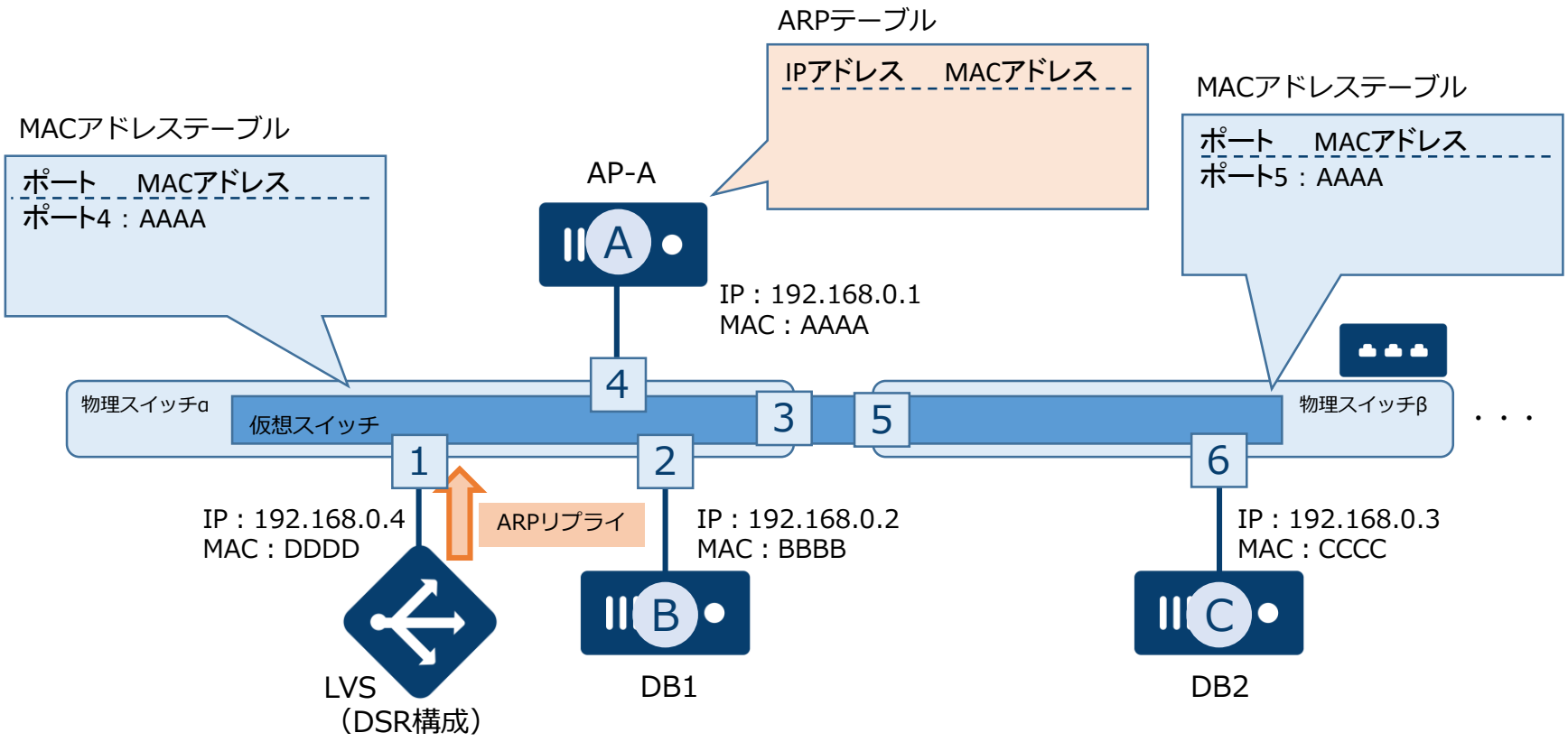
スイッチの各ポートでARPリクエストを受信した時点で、MACアドレステーブルに送信元（サーバA）の情報が登録されます。



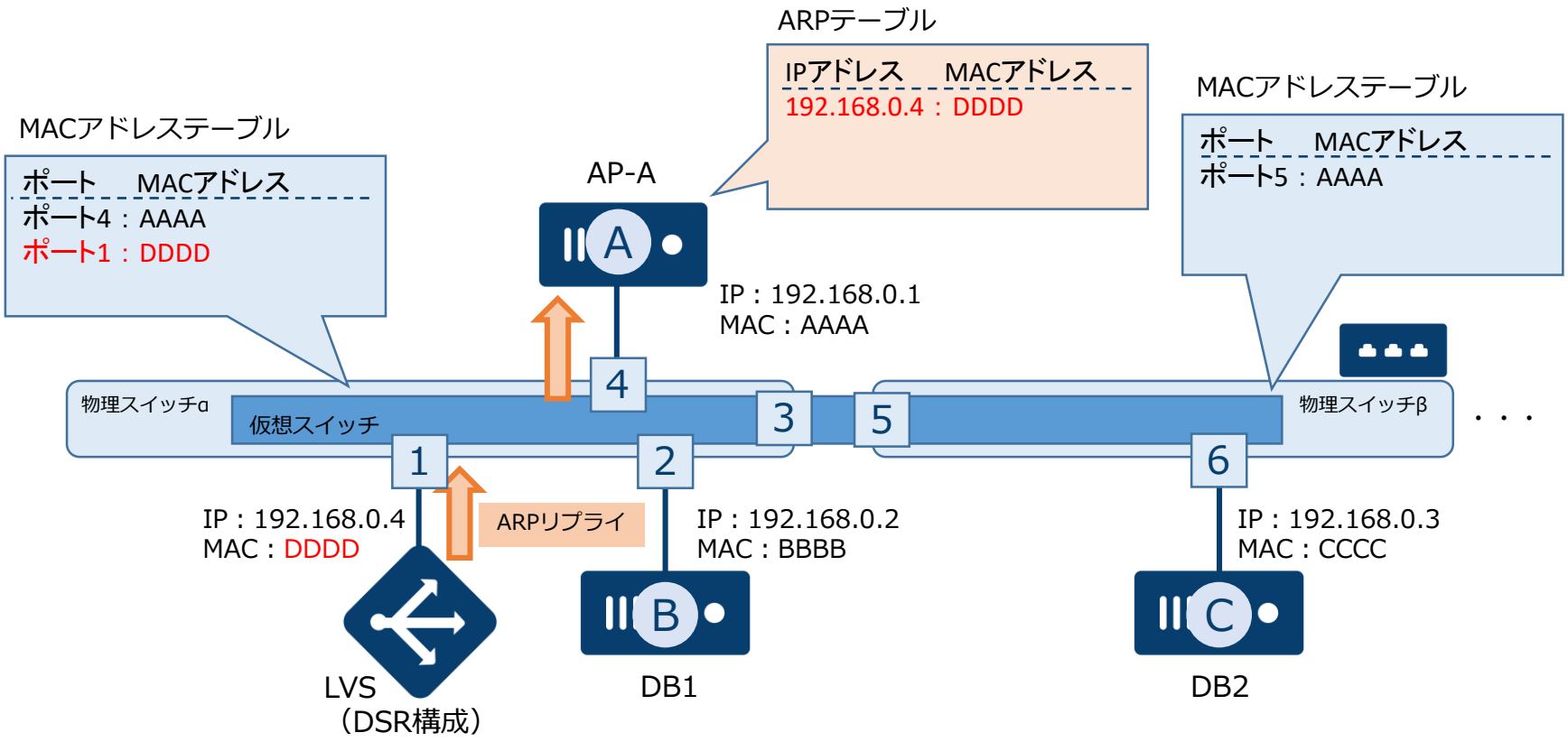
LVSはARPリクエストの中身を見て、自分宛の
リクエストだと気付きます。



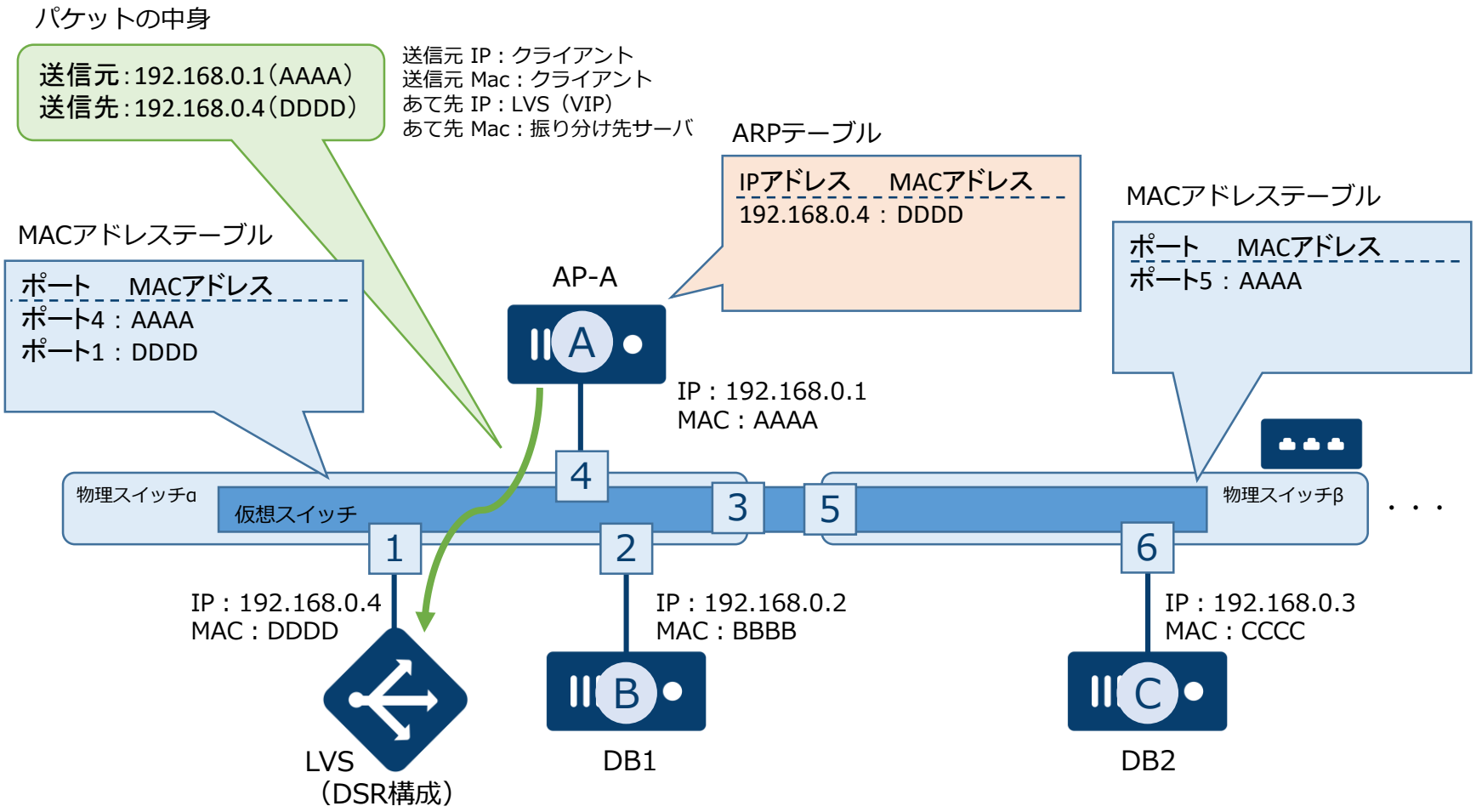
LVSは自身のサーバAにARPリプライを送信します。
※送信先はサーバAのみ



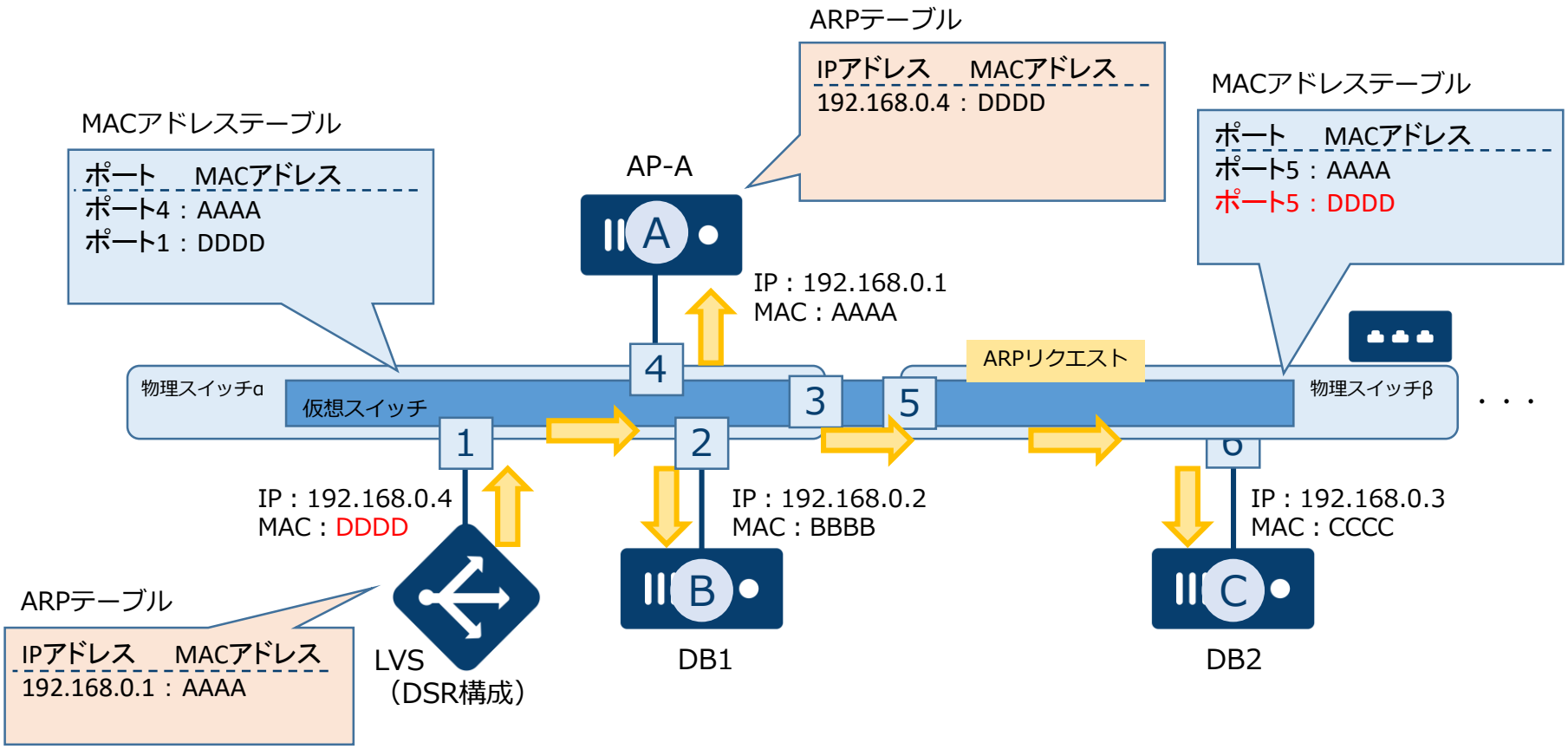
ARPリプライがサーバAに届き、サーバAはLVS
のMACアドレスを知ることができました。



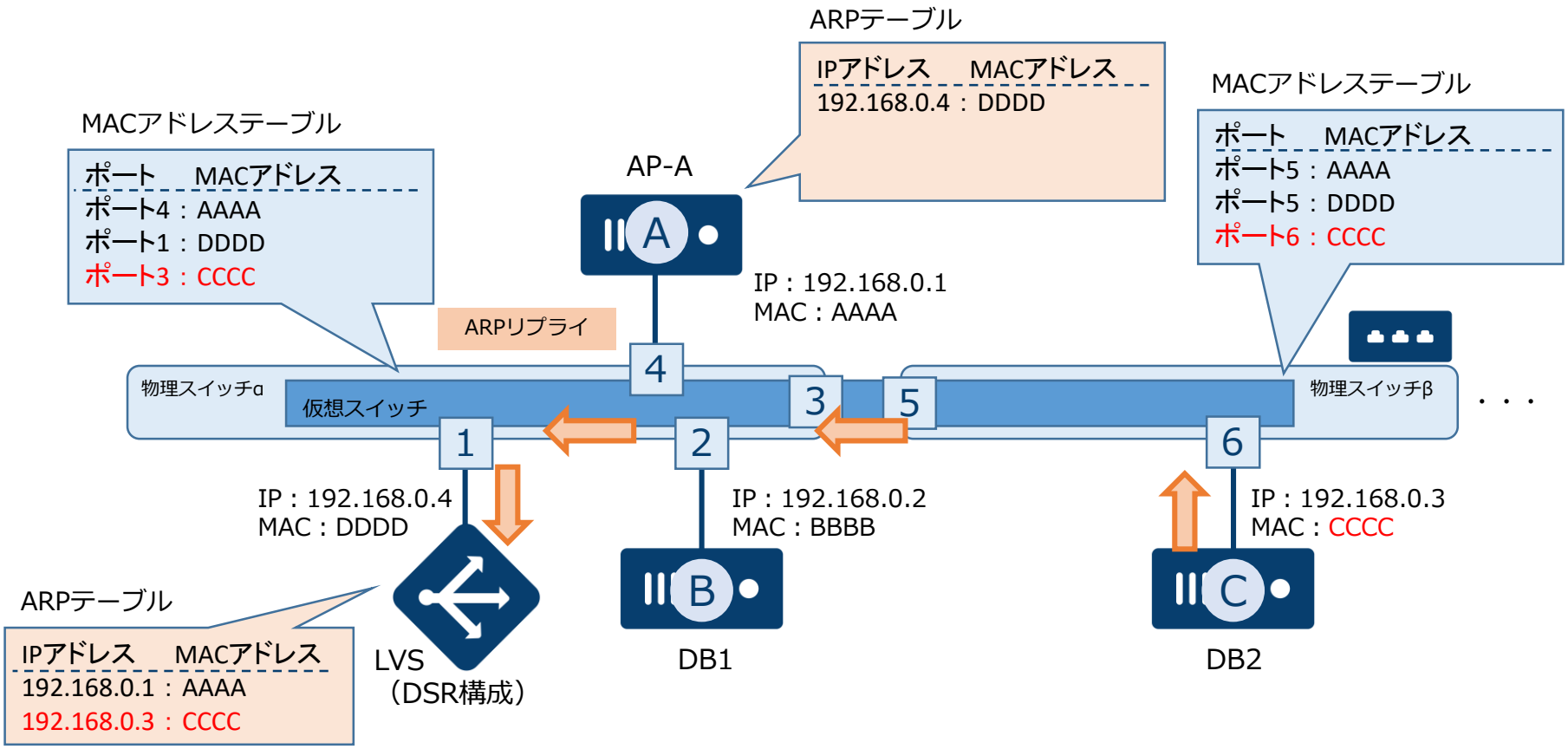
サーバAはLVSのMACアドレスを知ることができたので、
LVSに通信することができます。



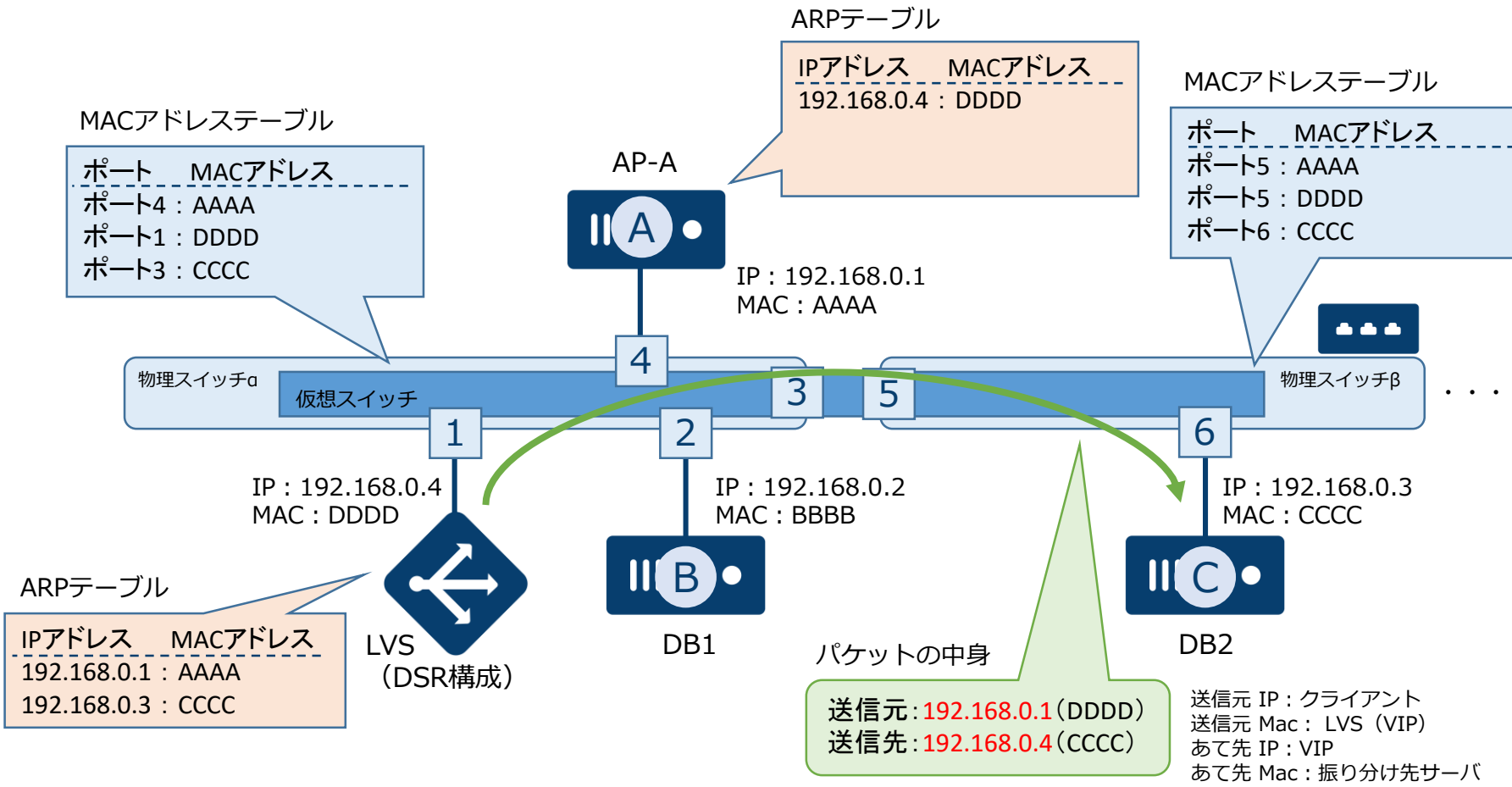
LVSは分散先のサーバとしてサーバCを選択したとします。
LVSはサーバCのMACアドレスを調べるため、ARPリクエストを送信します。



ARPリプライがLVSに届き、LVSはサーバCの
MACアドレスを知ることができました。

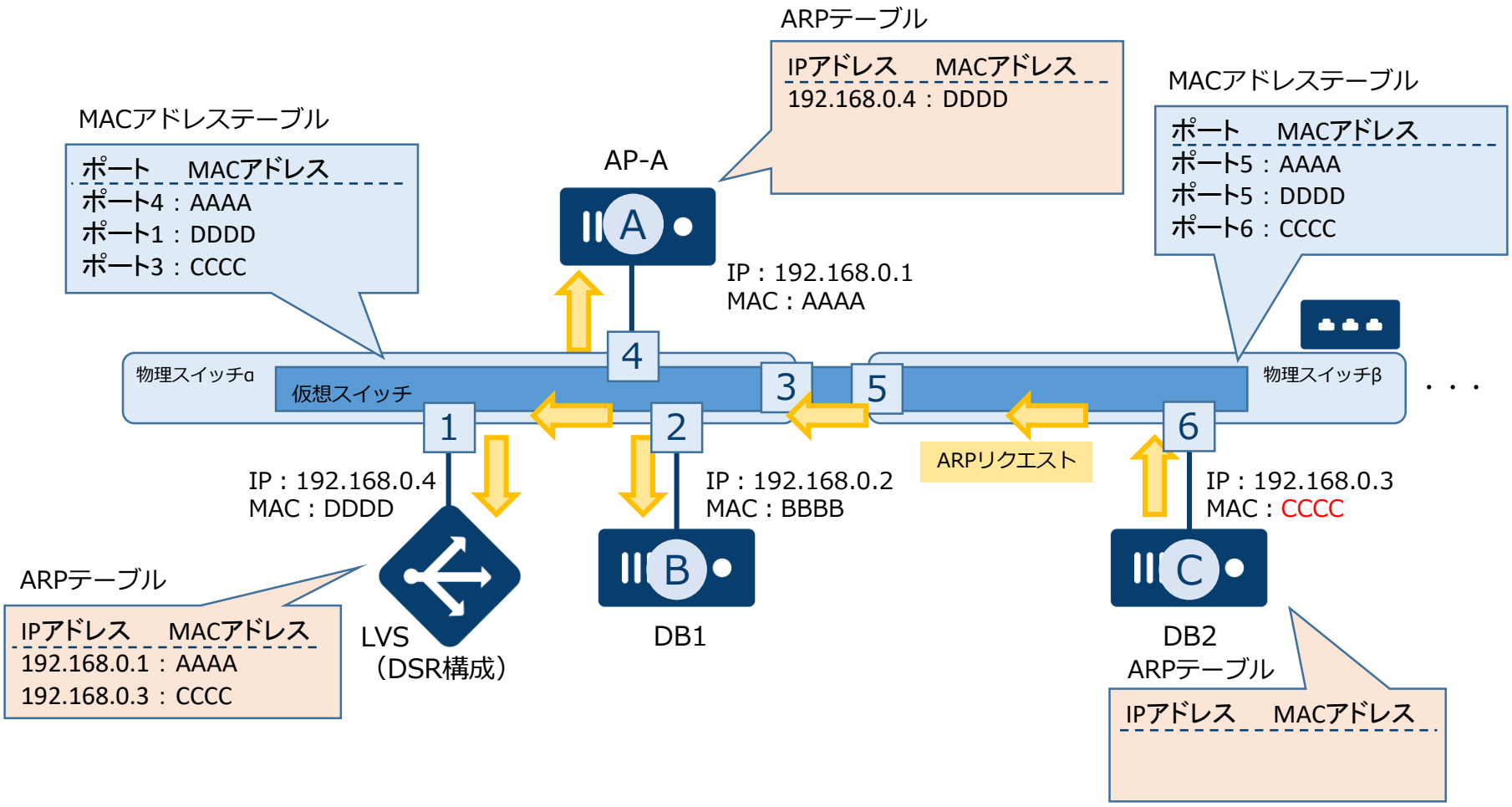


LVSはサーバCのMACアドレスを知ることができたので、
サーバCに通信することができます。

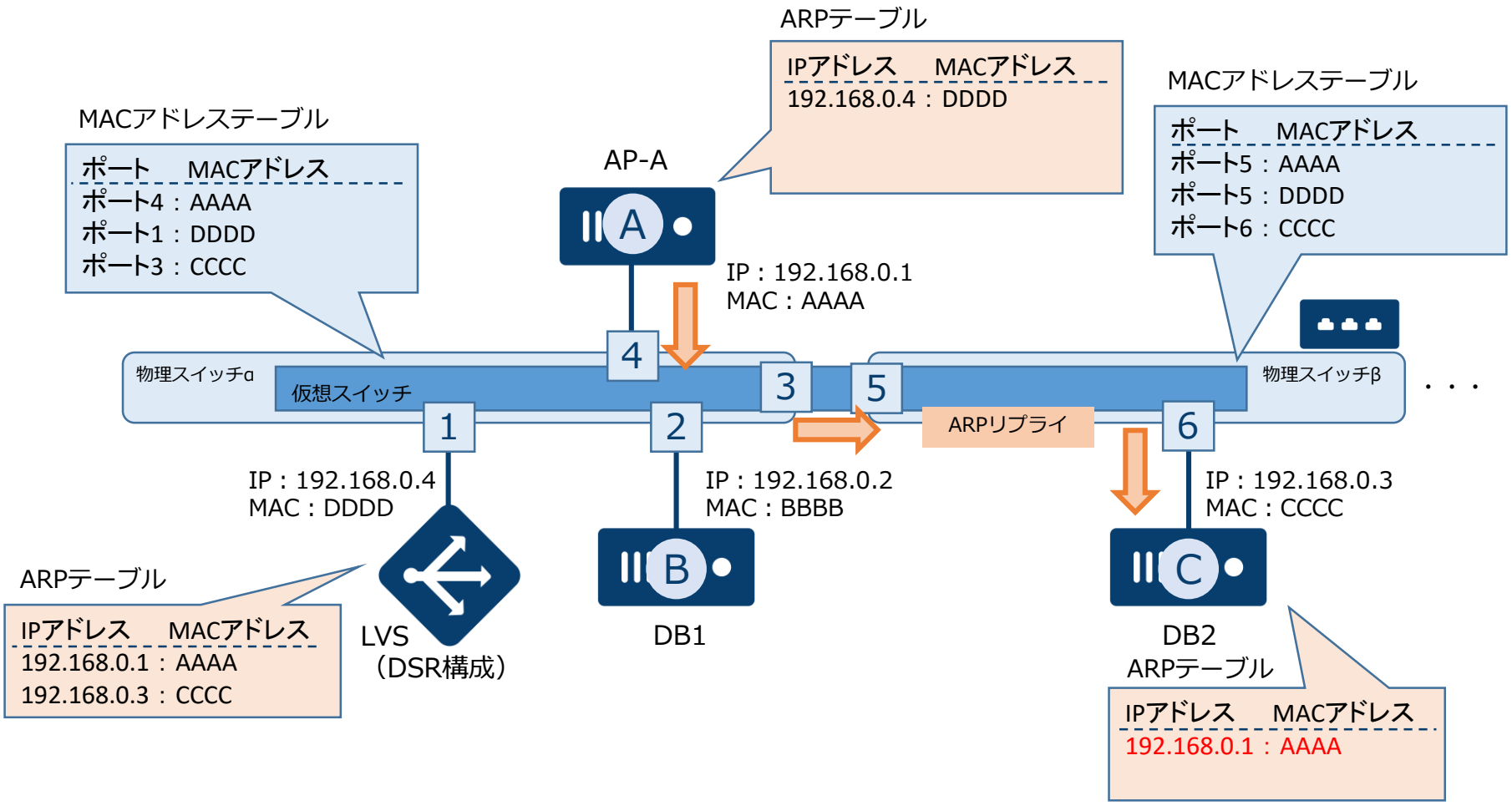


DSR構成のため、サーバCからの戻りの通信は、LVSを経由せず直接サーバAに送信されます。

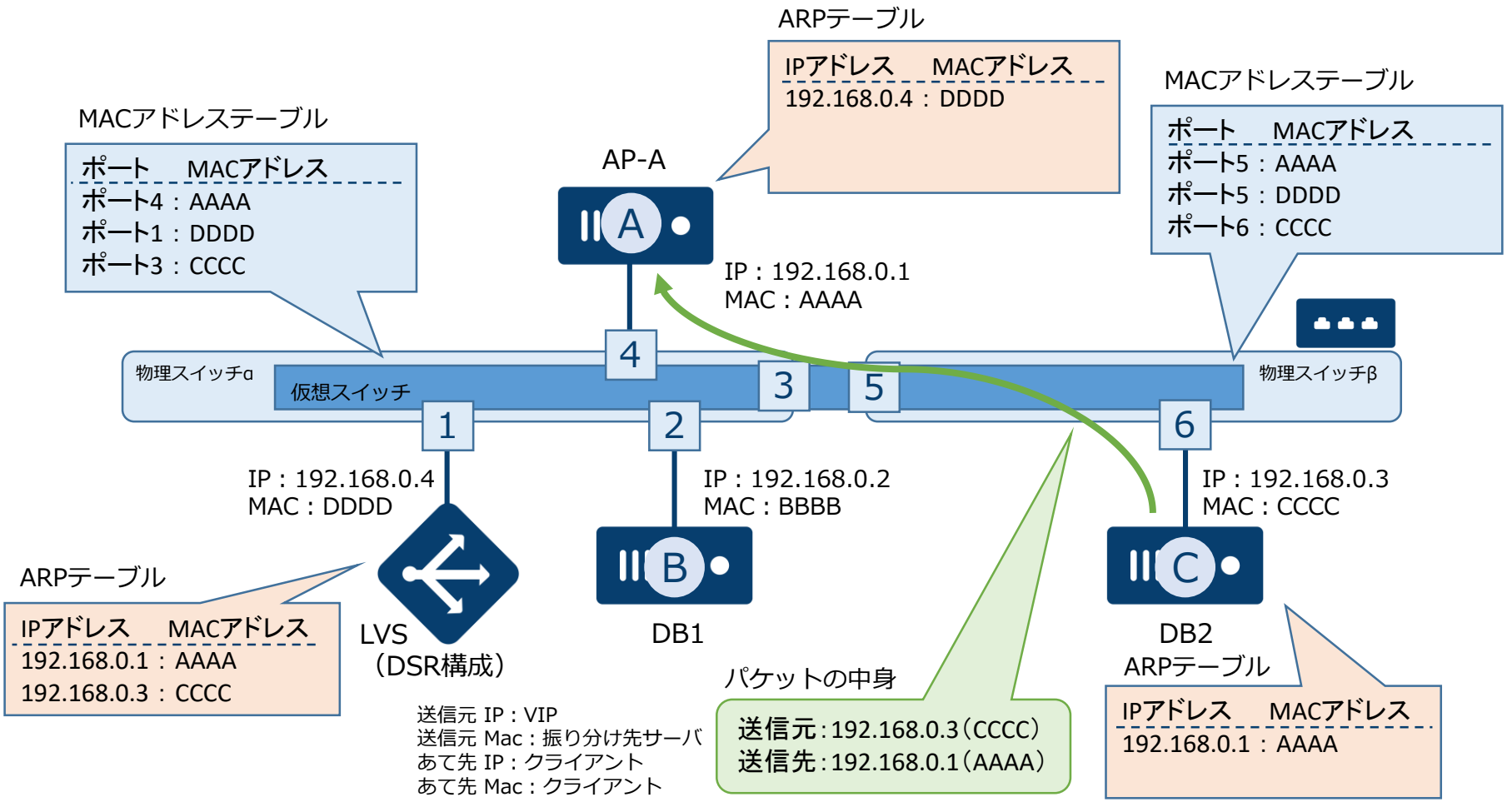
サーバCはサーバAのMACアドレスを調べるため、ARPリクエストを送信します。



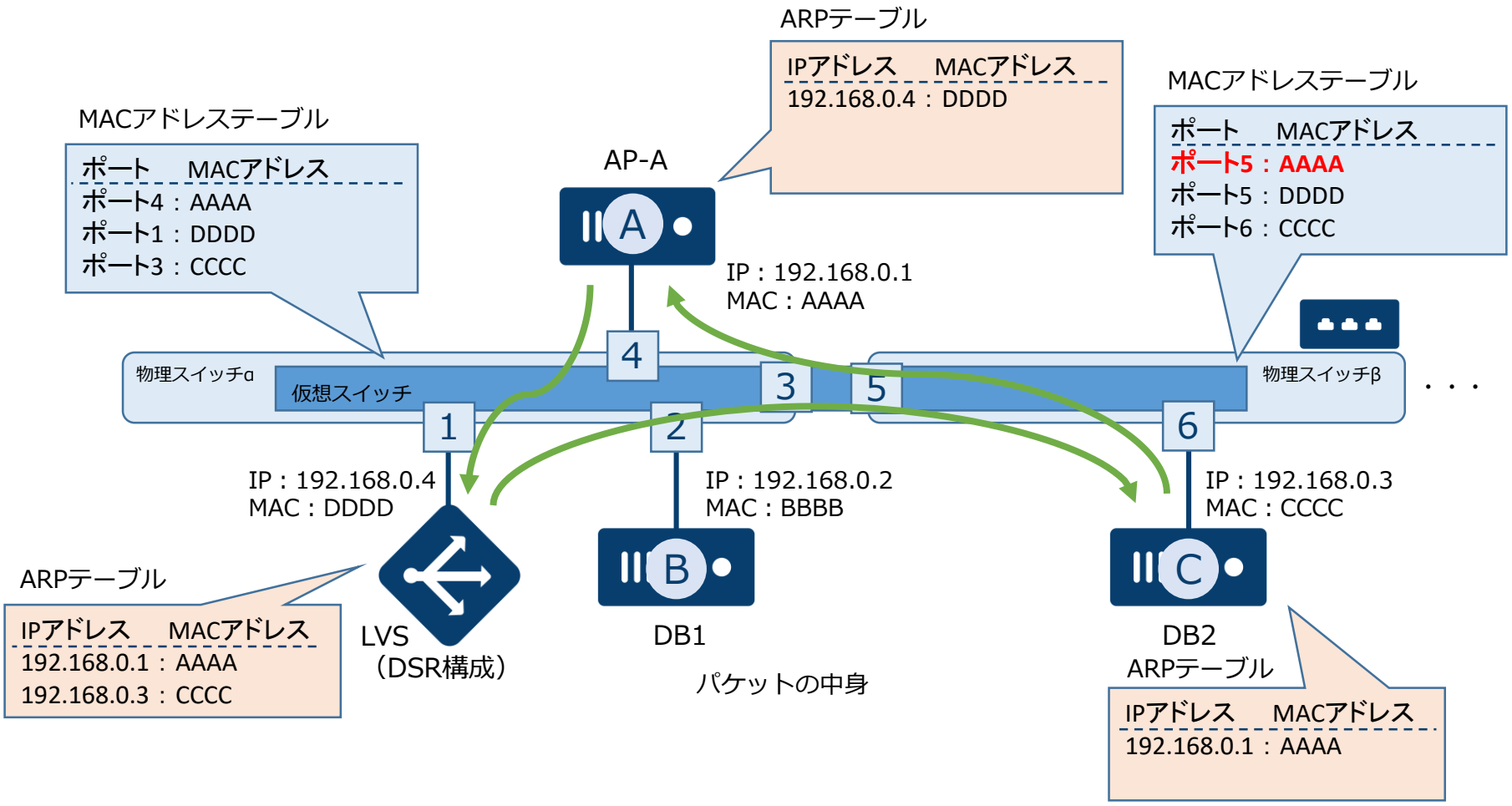
ARPリプライがサーバCに届き、サーバCはサーバAの
MACアドレスを知ることができました。



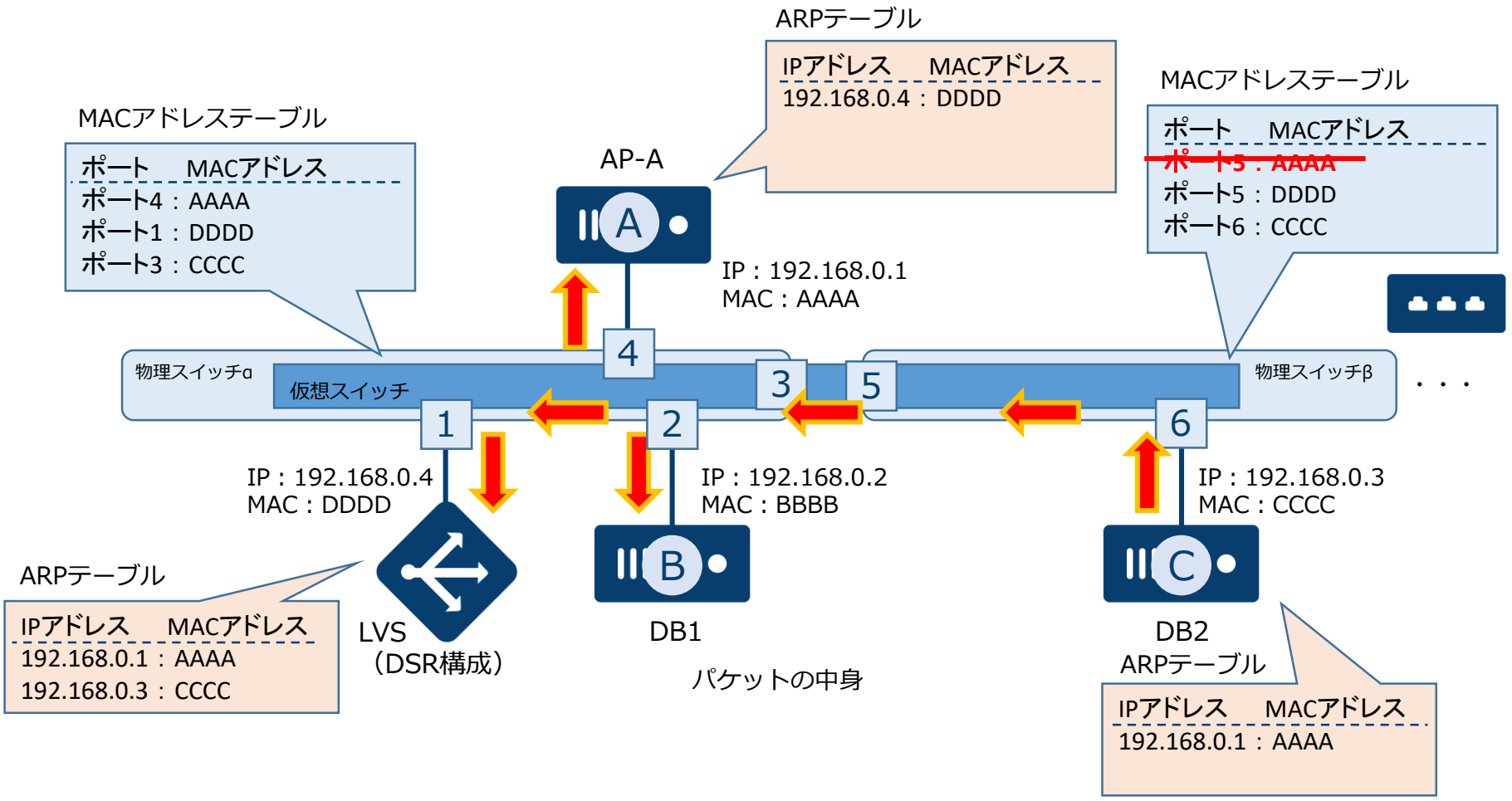
サーバCはサーバAのMACアドレスを知ることができたので、
サーバAに通信することができます。



このDSR構成ですとサーバAとサーバCによるARPリクエスト（ブロードキャスト）以外、「サーバA⇒サーバC」の通信が発生ません。また、LinuxサーバのARPテーブルは仕様上保持される傾向があり、サーバからARPリクエストは送信されません。
つまり、物理スイッチβはサーバAのMACアドレスを、改めて知る機会をほぼ失ったことになります。



この状態が続き、物理スイッチβのMACアドレステーブルの情報が消去（エージング）されると、サーバCからサーバA宛ての通信は、以下のようにフラッディングされます。



■ 対策例

各サーバから定期的なpingを打ち相互にMACアドレスの学習を促す仕組みの実装
(弊社のロードバランサー仮想アプライアンスは元々対策を実装済みです)