

第2章 インターフェースの冗長化・負荷分散

```
no : /etc/resolv.confを変更しない.
```

- HWADDR = このインターフェースに割り当てられたMACアドレス。自動設定される。MACADDRとの併用は不可。
- IPADDR = このインターフェースのIPアドレス。
- MACADDR = HWADDRを上書き変更するMACアドレス。
- MASTER = このインターフェースを使用するチャンネル・ボンディング・インターフェース名。SLAVEと併用。
- SLAVE = 「MASTER」のチャンネル・ボンディング・インターフェースによって制御されるか否かの指定。
 - yes : このインターフェースはMASTERにより制御される。
 - no : このインターフェースはMASTERにより制御されない。
- NETWORK = ネットワークIPアドレス。
- NETMASK = ネットマスク値。
- ONBOOT = OS起動時にこのインターフェースを有効にする (yes) か有効にしない (no ; デフォルト) かの設定。
- SRCADDR = 送信パケットに強制設定する発信IPアドレス。
- USERCTL = 管理者 (root) 以外のユーザがこのインターフェースを制御できる (yes) かできない (no ; デフォルト) かの指定。

また、インターフェースを制御するドライバ・モジュール名やその制御のためのシステムの割り込みやメモリはカーネル・モジュール・ロード設定ファイル(「`module.conf`」や「`modprobe.conf`」)に設定します。カーネル・モジュール・ロード設定ファイルはOSによって異なりますが、CentOS 5.1では`modprobe.conf`です(リスト2.1)。

なお、一つの物理インターフェースに複数のIPアドレスを割り当てたり、逆に、複数の物理インターフェースに一つのIPアドレスを割り当てたりすることができます。後者が次項のEthernetチャンネル・ボンディングです(前者はIPエイリアス)。

2.2 Ethernetチャンネル・ボンディング

Linuxシステムでネットワーク・インターフェースのリンク障害を検出して回避策をとるフォールト・トレランス(HA; 高可用性)機能と、送受信トラフィックの負荷を分散する負荷分散(ロード・バランシング)機能を提供するものがLinuxのEthernetチャンネル・ボンディング(結合)です。

Ethernetチャンネル・ボンディングはホスト・レベルのHA/負荷分散対策です。ネットワーク・レベルでの対策はSpanning Treeなどスイッチや、ゲートウェイなどネットワーク機器や、クラスタリングなどで行うことになります。クラスタリング・システムについては姉妹書「Linux高信頼サーバ構築ガイド—クラスタリング編」で解説します。

Ethernetチャンネル・ボンディング(正式名称)は、「チャンネル・ボンディング(チャンネル結合)」とか「トランキンク」,「チーミング(チーム化)」などとも呼ばれます。

Ethernetチャンネル・ボンディング機能はLinuxカーネルに付属しているもので、CentOS 5.1/Linux Kernel 2.6.18-53.el5(更新版のLinux version 2.6.18-53.1.19.el5, Linux version 2.6.18-53.1.21.el5)には、Ethernetチャンネル・ボンディング・ドライバのv3.1.2(January 20, 2007)が内蔵されています(*2.1)。

2.2.1 Ethernetチャンネル・ボンディングの仕組み

Ethernetチャンネル・ボンディングは、図2.1のように、一つのシステムに装備されている複数の物理的なネットワーク・インターフェースをボンディングして(束ねて)、一つの論理的なネットワーク・インターフェース(「チャンネル・ボンディング・インターフェース」)として利用する仕組みです。一つのMACアドレスでEthernetフレームを処理するにもかかわらず、その構成要素である一つ一つの物理的インターフェースを切り替えながら、HA機能と負荷分散機能を提供します。

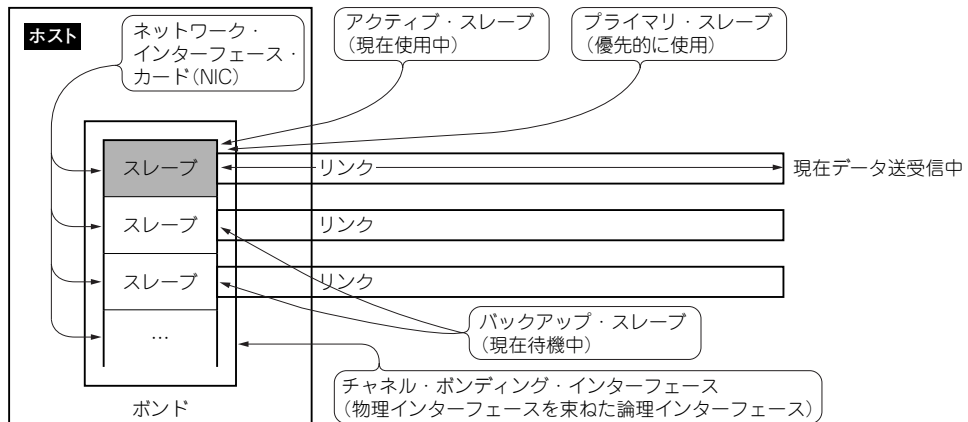


図2.1 Ethernetチャンネル・ボンディングの仕組み

(*2.1) Ethernetチャンネル・ボンディング・ドライバについては、詳細な解説文書「Linux Ethernet Bonding Driver HOWTO - Latest update: 12 November 2007」がLinuxカーネルの付属ファイル「bonding.txt」として同梱されている。

/Documentation/networking/bonding.txt

また、次のURLにもある。

<http://sourceforge.net/projects/bonding/>

http://sourceforge.net/project/downloading.php?group_id=24692&use_mirror=jaist&filename=bonding.txt&408104

一部の和訳は次のURLにある。

<http://www.linux.or.jp/JF/JFdocs/kernel-docs-2.6/networking/bonding.txt.html>

このとき、論理的なその一つのチャネル・ボンディング・インターフェースを「ボンド(bond)」, ボンドを構成する複数の物理的なネットワーク・インターフェースをそれぞれ、「スレーブ(slave)」と呼びます。また、そのボンディング・ホスト(ボンドを使用しているホスト)が送受信に使用している現在のスレーブを「アクティブ・スレーブ」、切り替えを待機しているスレーブを「バックアップ・スレーブ」と呼びます。さらに、スレーブの中で障害がなければ常にアクティブ・スレーブとして利用されるスレーブを特に、「プライマリ・スレーブ(primary slave)」といいますが、このプライマリ・スレーブは特定のボンディング動作方式(「active-backup」モード)でのみ設定可能です。

なお、ボンド数とボンド内のスレーブ数には制限がありません。

「ボンディング・モード」は、Ethernetチャネル・ボンディングのHA機能と負荷分散機能を提供する動作方式の設定で、この二つの機能を個別に、あるいは、組み合わせて提供する、七つのモードがあります。これらを、利用環境や利用用途に応じて設定して、ネットワーク送受信を利用することになります。

なお、ネットワーク・インターフェースを「リンク(link)」とも記述する場合があります。一般に、スレーブはデバイスの意味に重点を置く場合、リンクはLANケーブル接続やその接続の状態に重きを置く場合に使用されます。

Ethernet チャネル・ボンディング・ドライバには、送受信リンクの障害を検出するための機構として、送受信リンクの状態を監視する「リンク・モニタ(link monitor)」の機能も付随して装備されています。

2.2.2 Ethernetチャネル・ボンディング機能の利用環境

Ethernetチャネル・ボンディングのHA機能と負荷分散機能を利用する場合、一般的に、そのボンディング・ホストとボンド経由で接続するスイッチとの間の構成が異なります。

HA機能を利用する場合、ボンドを構成するリンクは対向するいくつか別々のスイッチに接続するかたちで[図2.2(a)], 負荷分散機能を利用する場合には、すべてのリンクが同一の対向スイッチに接続するかたちです[図2.2(b)]. もちろん、ボンディング・モードによってはこのようにならない場合もありますが、一般的にはこのようになります。

HA機能を利用する場合、一つの同じスイッチに接続していたのでは、そのスイッチ自体が停止してしまった場合、Ethernetチャネル・ボンディングのHA機能が無意味になってしまうためです。

一般的には上記のように、「1ホスト+ボンディング+1または複数のスイッチ」という構成ですが、実際の現場では、これら「スイッチの向こう側」あるいは、これら「スイッチ間のリンク」があるわけで、このあたりも考えなければなりません。この詳細については後述します。

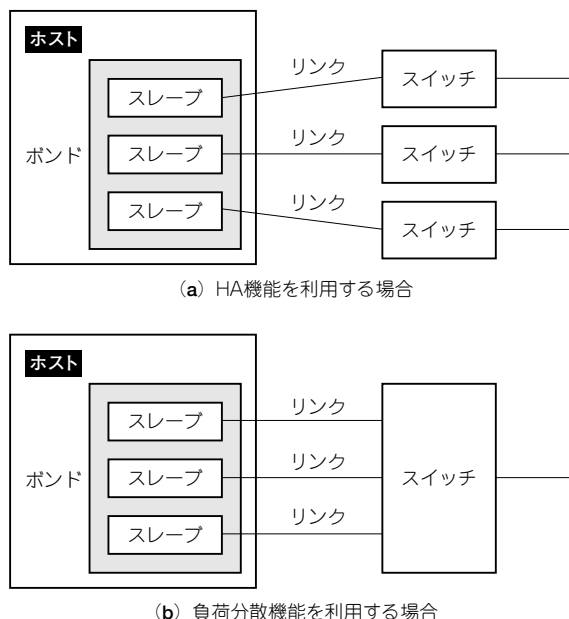


図2.2 Ethernetチャンネル・ボンディング機能の利用環境

2.2.3 リンク状態の監視

ボンディングのリンク状態を監視するリンク・モニタの方法には2種類あります。

より一般的なMII^(*2.2)規格のリンク・モニタ(MIIリンク・モニタ)と、ARP(Address Resolution Protocol)要求/応答パケットの送受信によるリンク・モニタ(ARPリンク・モニタ)があります。

MIIリンク・モニタはMII信号(イベント)を監視してリンク・ダウンを検出しますが〔図2.3(a)〕, ARPリンク・モニタでは設定された対地(MACアドレス)宛にARP要求パケットを送り、その応答でリンク(そこまでのリンク)の状態を監視します〔図2.3(b)〕。

これが、ARPリンク・モニタがMIIリンク・モニタとは異なる大きな点で、ARPパケットの送受信によるトラフィックがLAN上を流れることとなります。MIIリンク・モニタに比べてこのトラフィック負荷を増やす短所がありますが、一方で、MIIリンク・モニタがスイッチまでのリンク状態しか監視できないのに比して、ARPリンク・モニタでは、スイッチの向こう側のリンク(あえていえば、ARPパケットの対地までのリン

(*2.2) MII(Media Independent Interface；メディア非依存インターフェース)

IEEE 802.3(CSMA/CD)規格の中の、MAC層(メディア・アクセス制御層)とPHY層(物理層)との間で、伝送媒体(PHY層、ケーブル)に依存しないように相互接続させるためのインターフェース。10/100Mbps用で、ギガビットの場合はGMII(Gigabit Media Interface)。なお、10MbpsのMIIはオプションのため、サポートされないNICもある。