

Xenoppix(x86/x86_64 of Xen 3.0.3)ハイブリッド版の詳細

須崎有康ⁱ, 八木 豊志樹ⁱ, 飯島 賢吾ⁱ, 塚本純一ⁱⁱ, 北川健司ⁱⁱⁱ

<http://unit.aist.go.jp/itri/knoppix/xen/>

1. Xenoppix とは

Xenoppix とは 1 CD Linux である KNOPPIX[1] に仮想マシンモニタ “Xen[2]” を収録し、1 CD ハイパーバイザーとして拡張したものである。Xen は Domain と呼ぶ仮想マシン環境を提供するが、そのうち 1 つを Domain0 と呼び、他の Domain をコントロールする権限を持つ特別な Domain として扱う。他の Domain にはカーネルのパッチを適用して動かす準仮想化の DomainU と、CPU の提供する仮想化命令を利用してカーネルパッチを適用せずに通常のカーネルが動く完全仮想化の HVM がある。

Xen は物理デバイスのドライバを提供しないが、Domain0 のカーネルの物理デバイスドライバを借用し、他の Domain へ提供する。KNOPPIX のデバイス認識/ドライバセットアップ機能 (AutoConfig) が優秀であり、Domain0 の OS として起動することで、Xen と相補的に働く。

最新の Xenoppix は KNOPPIX5.0.1 をベースに Xen3.0.3 を加えている。また、最新の CPU では仮想化命令や 64 ビット化対応しており、これらの機能を活用できるように拡張した。この詳細について述べる。

2. 最新の CPU への対応

Intel の Core2 や AMD の Athlon64 などの最新の CPU では仮想化命令と 64 ビット対応が盛り込まれている。

2.1 仮想化命令対応(完全仮想化)

Xen では最新の CPU が提供する仮想化命令 (Intel VT, AMD-V) を活用することで完全仮想化環境 HVM を提供し、ゲスト OS に変更を加えなくても動かすことを可能にしている。

最新の CPU では特権レベルとは異なる仮想化モード (VMX モード) が導入されている。VMX モードでは非特権だがセンシティブな命令 (CPU の状態遷移に影響を及ぼし、OS の仮想化で問題になる命令。x86 では PUSHF, SGTD など 17 個ある。) を実行する時にモードが変わり、仮想マシンモニタが処理を行なう。Xen では仮想化命令を活用した HVM 完全仮想化環境を提供することができる。

2.2 64 ビット化

最新 CPU には 64 ビットの機能 X86_64 (Intel 64, AMD64) が含まれているが、プレインストール OS が 32 ビットの場合は利用されずに、数値計算などを高速化できるのに、その機能が活かされていない。また、64 ビット化と仮想化命令は直交であり、それぞれを組み合わせられる。ただし、32 ビットで Xen を起動した場合は仮想化が 32 ビットになるため、64 ビット OS は動作できない。

Xenoppix 最新版では管理 OS となる Domain0 に「32 ビット版 Xen+x86 用 Linux カーネル 2.6.16」か「64 ビット版 Xen+x86_64 用カーネル 2.6.16」かをブート時に選択するようになっており、それぞれを活用できる。また、Xen を適用していない Linux カーネル 2.6.17 でも起動できる。こちらは HVM 上で起動するときにも使われる。

3. Xen3.0.3 の拡張機能

最新の Xenoppix では Xen3.0.3 を採用している。これにより、今まで以上の機能が利用可能になった。

3.1 SMP 構成変更

Xen3.0.3 になり、HVM の仮想 CPU の数を簡単に変更することができるようになった。これにより並列アプリケーションの実験が簡単にできるようになった。

3.2 ビデオカード問題の解決

Xen3.0.2 ではビデオカードがメインメモリを使用する UMA (Unified Memory Architecture) の場合に Xen の HVM が使用するメモリとぶつかる問

i) 産業技術総合研究所

ii) 東京大学

iii) アルファシステムズ

題が存在した。グラフィック統合型チップセットは UMA のため、多くのノート PC で HVM が実行できない問題が存在した。Xen3.0.3 ではこの問題が解決され、グラフィック統合型チップセットでも問題なく HVM を試すことができる。

3.3 HVM 上の起動制限の減少

Xen3.0.2 では Core/Core2 の HVM で起動する場合、いくつかのデバイス認識をキャンセルしなければならなかった。KNOPPIX の起動オプションで言えば `-nolpic - nolapic - nosmp - noapm` を付ける必要があった。Xen3.0.3 ではこのようなオプション無しでも HVM で問題なく OS を起動できるようになり、ブートメニューと通常のオプションなしと共用できるようになった。

このため、現在は x86, x86_64 カーネルそれぞれで Xen 付きと Xen なしの 4 つの GRUB メニューから選ぶ方式になった。

3.4 HVM 用のメモリサイズの確保が不要

HVM を使う上で domain0 のメモリサイズを予め指定する必要がなくなった。Xen3.0.2 ではメモリサイズを指定する必要があり、Xenoppix ではブートオプションに「指定無し」/「512MB に指定」の 2 つオプションを用意した。指定無しでは Xen が確保する以外の物理メモリを Domain0 が利用可能になり、DomainU には動的に割り当てていた。しかしながら、この状態では HVM にメモリが割り当てられない問題が存在した。このため、Domain0 にメモリ制限を加える「512MB に指定」を GRUB メニューに加えていた。Xen3.0.3 ではメモリサイズの指定がなくても HVM が起動できるようになり、このようなメモリサイズ制限のメニューを削除できた。

3.5 EFI パッチと HVM on AMD-CPU 問題

Xenoppix は GRUB に EFI パッチを適用して Intel Mac でも利用可能にしている。

Xen3.0.3 に上げることで逆に GRUB の EFI 対応パッチが AMD-CPU 上の HVM を起動できない問題が発生した。つまり、GRUB パッチの `wait_for_key` の処理で HVM が割り込み待ちの状態で止まる。このため EFI を採用する Intel Mac 対応と相容れない状況になった。Xen3.0.2 ではこの問題が発生しなかったため Xen3.0.3 により含まれた bug と思われる。この対応には「EFI パッチを当てる GRUB バージョン」と「当てない GRUB

バージョン」の 2 種類の DVD イメージを作成した。

3.6 PV(Para-Virtualized)ネットワーク

HVM では OS の変更無しで起動することができるが、ネットワークの性能がでないことがわかっている。Xen3.0.3 ではこの問題を PV (Para-Virtualized)ネットワークを提供すること解決することができる。ここでは HVM 上の OS に PV ネットワークドライバをインストールする。PV ネットワークドライバは準仮想化ドライバであり、Domain0 の offload Engine と連動して高速化が図られている。

4. 今後の展開

Xenoppix と平行して Internet Thin Client としてブートする HTTP-FUSE KNOPPIX[3]を開発している。Xenoppix と組み合わせた HTTP-FUSE Xenoppix[4]も作成しており、最新の Xenoppix も対応していきたい。

5. まとめ

Xenoppix(x86/x86_64 of Xen 3.0.3)ハイブリッド版の詳細について述べた。Xen を開発している XenSource からも 1CD で起動する DemoCD を提供しているが、こちらは DomainU のみの起動で Xen が可能にしている多くの機能を有効にしていない。また、ドライバは汎用なものを固定にしておき、本来の性能を活かすことができている。Xenoppix ではこれらの問題を解決して、より使いやすく性能のよいものを提供している。今後も KNOPPIX と Xen の両者を活かしたものを開発を継続するとともに、HTTP-FUSE KNOPPIX のようなネットワークブートにも適用範囲を広げていきたい。

参考文献

- [1] <http://unit.aist.go.jp/itri/knoppix/>
- [2] <http://www.cl.cam.ac.uk/research/srg/netos/xen/>
- [3] 須崎, 八木, 飯島, 丹, “HTTP-FUSE KNOPPIX”, ”.Linux Conference 2005. <http://lc.linux.or.jp/paper/lc2005/CP-02.pdf>
- [4] K.Suzaki, T.Yagi, K.Iijima, K.Kitagawa, and S.Tashiro, “HTTP-FUSE Xenoppix”, Ottawa Linux Symposium, 2006. Vol2. pp.379-391