カーネル 2.5 の世界

- 次世代 Linux カーネルが目指すもの -

後藤 正徳
GOTO, Masanori
gotom@debian.org

Linux カーネル

- Linus Torvalds 氏が開発
- Unix ベース モノリシックカーネル
- BSD, SYSV, POSIX, X/Open 互換
- GPL
- 今年誕生10周年
- ここ数年で大きく普及 (全世界に何百万以上のユーザ)

Linux カーネルの歴史 (1)

1990 ... Linus 氏が MINIX をベースに開発を開始
1991 ... comp.os.minix にアナウンス
1993 ... Linux 0.99.x
1994 ... Linux 1.0
Linux 1.1
1995 ... Linux 1.2
Linux 1.3
1996 ... Linux 2.0
Linux 2.1

Linux カーネルの歴史 (2)

1999 ... Linux 2.2
Linux 2.3
2001 ... Linux 2.4
Linux 2.5 開発スタート?
200x ... Linux 2.6? 3.0?

※バージョン番号の下一桁が
▽偶数なら安定版
▽奇数なら開発版

Linux カーネルの歴史 (3)

カーネル 2.4 の特徴

▽新しいデバイス (特にホットプラグデバイス)
　の実装と統合 (pcmcia, usb, ieee1394, ...)
▽SMP の性能向上
▽バッファキャッシュの改善
▽新アーキテクチャ対応 (SH, MIPS64, IA-64)
▽intel PAE (64GB メモリ) 対応

▽→ 大規模システムでも使用に耐えうる

最近のカーネル開発状況

バージョン: 2.2.19/2.4.10

▽-pre がつかないバージョンが正式リリース版

▽-pre がついたバージョンは pre release (テスト版)

-ac kernel: 2.2.20ac-pre-10/2.4.9-ac16

▽Alan Cox 氏がメンテナンスしているバージョン

▽"安定版" メンテナンスバージョン

▽Linus 氏が取り込んでいないパッチも含んでい

カーネル 2.5 (1)

カーネル 2.6/3.0 に向けての開発版

- ▽今年 3月に kernel hacker がアメリカに集まって議論
- ▽予定では7月末に登場予定だった.
- ▽まだ登場していない (!)
- ▽カーネル 2.4 が安定した時点でリリースされる
- ▽2.4 が VM などに問題を抱えているため、いつになるか不透明

本発表は...

- カーネル 2.5 で実装されると思われる主要な機能・特徴をまとめる
- 各 Mailing List, Project でのアナウンス、ToDo List などを元に作成
- 既に古い内容が含まれているかもしれない
- いくつかの項目は、先に 2.4 へ入ってしまうかもしれない?
  - ext3
  - ATA/133
- 今後のカーネル・コンピュータ動向の参考になる

カーネル 2.5 が目指すもの

▽より高いスケーラビリティ
▽さらなる大規模システムへの対応
▽リリース間隔の短縮
▽新しい機能・デバイスの追加
▽(ユーザ利用性が更に向上)

CPUアーキテクチャ

2.4 サポートCPU:

▽alpha, arm, cris, i386, ia64, m68k, mips, mips64, parisc, ppc, s390, s390x, sh, sparc, sparc64

2.5 では...

▽mips:　1ツリーへの統合

▽better support for integrating 32/64bit.

今後対応予定のCPU:

▽ppc64, x86_64 (-ac ツリーには既に存在)

▽um (ユーザモードLinux)

カーネル

- "軽量" セマフォの採用
  - 高速なロック
  - POSIX process shared
- スケジューラの更なる改善 (?)
  - 既にいくつかの改善パッチが出ている
- プリエンプティブ・カーネル化 (?)

メモリマネージメント (VM)

- NUMA (Non-Uniform Memory Access) サポート
  - 現在は SGI NUMA マシン用パッチのみ
- reverse mapping (2.4 に入る?)
- HIGHMEM zero-bounce
- Large Page サポート
  - intel arch では 4KB/4MB ページ粒度
- スラッシング時の改良

ファイルシステム

ジャーナリングファイルシステム
▽ext3 (マージ間近?)
▽XFS (SGI)
▽JFS (IBM)
▽Tux2 (開発中)
ネットワークファイルシステム
▽OpenGFS
dcache スレッド化

ネットワーク (1)

SCTP (Stream Control Transmission Protocol)

▽TCP, UDP の利点をあわせ持つ次世代ストリーミング向けプロトコル

▽RFC2960 に記載

802.1Q VLAN

▽1物理イーサネットポートに複数のVLANを割り当てる機能

IPSec

▽暗号化輸出規制の関係で、そのまま入るのは難しい情勢

ネットワーク (2)

高負荷時への対処

▽現在の実装では、高負荷耐性が低いため

▽割り込み・パケット制御を改善へ

Zero-Copy TCP/UDP

▽ユーザ空間から直接デバイスドライバのバッファへデータを送るよう変更

▽性能を引き上げる

▽実現できるか不明

10 GbE

デバイス名

- デバイス番号 16bit → 32bit
  - 8 major 8 minor → 12 major 20 minor へ拡張
  - 扱えるデバイス数上限が大きく増える
- Documentation/devices.txt 改訂凍結騒動
  - 固定番号で管理は限界? (例: Solaris /devices)
  - 名前の方が分かりやすい
  - ホットプラグデバイス対応では動的割当てが望ましい
- 候補
  - devfs (device filesystem) → 反対意見も多い
- 2.5 で最も激しく議論されるであろうテーマ

ブロックデバイス

- 64bit ブロックデバイスサポート
  - 32 bit (2TB) → 64 bit への移行
- I/O ロックの細粒度化
  - スケーラビリティ改善
- Raw I/O の改善
  - 大きなブロックへの対応
- ジャーナリングファイルシステムへの対応
  - prepare/write commit
- iSCSI
  - RFC Draft IPS (IP-Storage)
- LVM (Logical Volume Manager) 改良

非同期I/O (AIO)

- I/O処理を非同期に発行して高速化する技術
- 現在の実装は glibc で pthread を使用してプロセスベースによるエミュレーションを行っている
- 2.5 に入る予定
  - POSIX 似になるか独自実装になるか不明

デバイスドライバ

新しいハードウェアへの対応

- Serial ATA, Serial SCSI
- 3GIO
- Infini Band
- 10GbE
- Bluetooth/無線LAN
- …

サウンド

OSS-Free から ALSA へ

OSS-Free:

- ▽Open Sound System (OSS) のフリー版
- ▽OSS 自体は商用 (4 Front Technologies)
- ▽サポートされていないカードは OSS を使用しなければならない

ALSA:

- ▽OSS-Free をリライト
- ▽OSS インタフェースと互換性がある
- ▽仕様が公開されているほぼ全てのカードをサポート

グラフィックス

フレームバッファ
▽複数デバイス対応

V4L から V4L2 へ?
▽V4L (Video For Linux): ビデオキャプチャ, ラジオ, カメラなどをサポート
▽V4L2: V4L を完全に書き直し
▽V4L とアプリケーションレベルで互換性あり

電源管理 (Power Management)

ACPIサポート

▽現在の実装は完成には程遠い

▽今後の改良が必須

▽Intel パッチで改善に向かいつつある

ドライバ電源管理機能

▽2.5 で統一した仕組みがサポート (?)

▽そもそも電源管理を意識していないドライバが多い

ホットプラグ

▽USB, IEEE1394, PCI, SCSI, PCMCIA サポート済

▽これまでは個々のデバイスが対応 → 統合へ

▽統合に際して名前空間 (デバイス番号) の解決が課題

ビルドシステム

CML2

- CML2 言語を使用して記述
- 現在のカーネル Makefile を改善
- ユーザは make config ; make でモジュールも含めてビルドできるようになる
- 2.5 で採用 (?)

ソース管理

- 新しい管理方式へ移行
- システムの候補:
  - Subversion
  - BitKeeper
- いずれも cvs を大きく改善
- はたして移行するのか?

新しいインターフェース

▽Austin Group (SUSv3), POSIX への対応

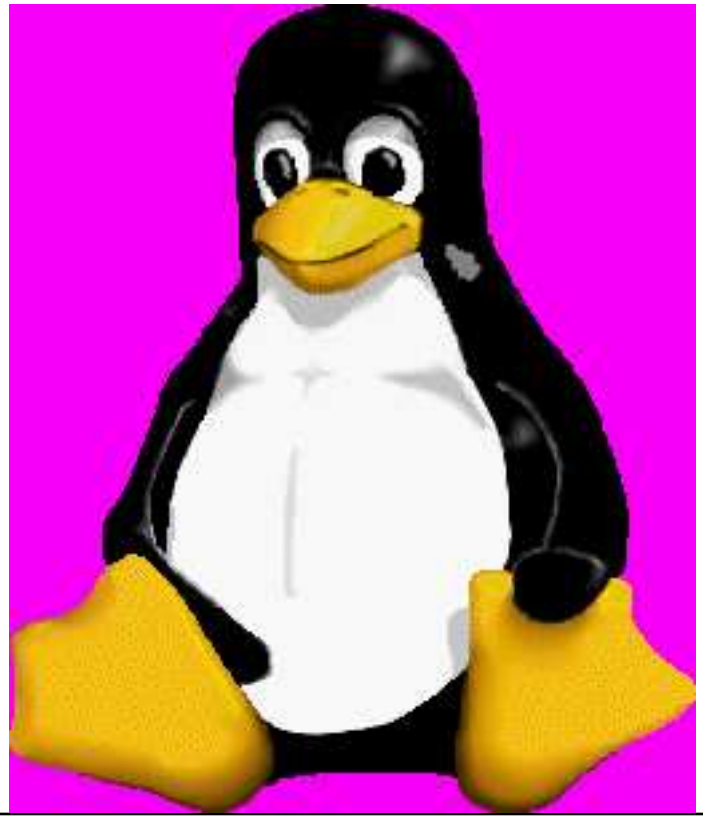
▽/dev/poll

まとめ

カーネル 2.5 は...

▽さらに大規模システム対応を目指す

▽決してエンタープライズだけのものではない。
通常用途のサポートにも要注目

▽(今年中に出るか?)

Happy Linux-ing!