

どこからともなくブートする OS — HTTP からブートする Linux —

1. 背景

現在の OS はブートする際に固定のデバイス(ハードディスクなど)にあるルートファイルシステムを指定しなければならない。ネットワークブートでもホストとなる固定サーバが必要である。この方法ではルートファイルシステムがあるデバイスやホストが壊れた場合に OS をブートすることができない。多くのクライアントマシンが一斉にネットワークブートした場合にサーバへのネットワーク負荷が過大となり、ブートできない問題がある。

2. 目的

「どこからともなくブートする OS」では、ブートデバイスやサーバに固定されず、ハードディスクが故障してもブート可能とり、信頼性の向上が図れる。また、OS の乗換えが再インストールを意味せず、ただ単に対象ルートファイルシステムの指定を変えるだけになり、気軽に行なうことができる。これにより、ハードディスクに一旦インストールして変更しづらくなる問題(OS のロックイン)を解決する。

3. 開発の内容

開発は1CDLinux の“KNOPPIX”をベースに行った。KNOPPIX はブート時にデバイス自動認識/ドライバの組み込み機能(Autoconfig)が優れており、不特定多数のユーザが個々の PC から Internet ブートするのに必要な条件を備えている。ルートファイルシステムも圧縮ループバックデバイス(cloop: compressed loop back device)に格納され、1ファイルとして扱える。しかし、現状の cloop では 700MB の cloop ファイルを一括でダウンロードしなければならない。この問題を解決するために、本開発では cloop を拡張し、必要な部分をオンデマンドでダウンロードしてブート可能な「分割圧縮ブロックファイルによるループバックデバイス(fuse-cloop)」を開発した。改良のポイントは下記(図1)である。

- 64KB 毎にブロックを切り出し、圧縮したデータを”ファイル”に格納する(分割圧縮ブロックファイル)。
- 分割圧縮ブロックファイルをまとめて、ループバックデバイスとして扱えるようにする。
- ループバックデバイスでは、必要な時に必要な分割圧縮ブロックファイルを取得すればいいようにする。つまり、全ての分割圧縮ブロックファイルが揃わなくてもループバックデバイスとして扱えるようにする。
- 分割圧縮ブロックファイルのファイル名は checksum(Ex: md5)の値とし、ファイル名で内容確認ができるようにする。同一内容の場合、checksum が同じになりファイル数を減らすことができる。

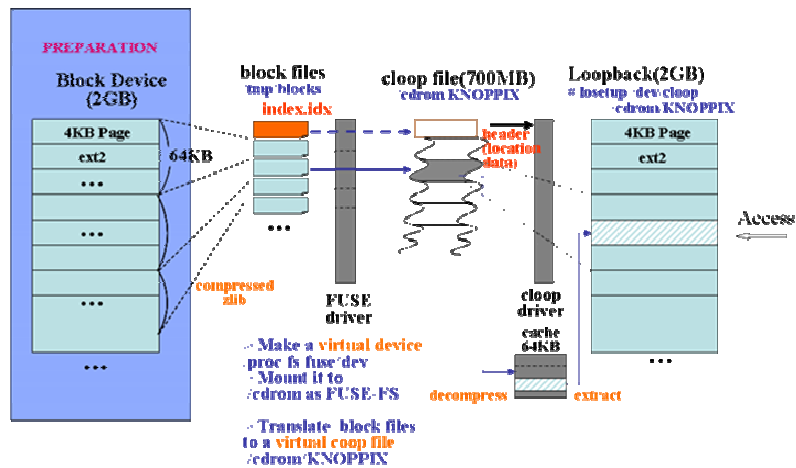
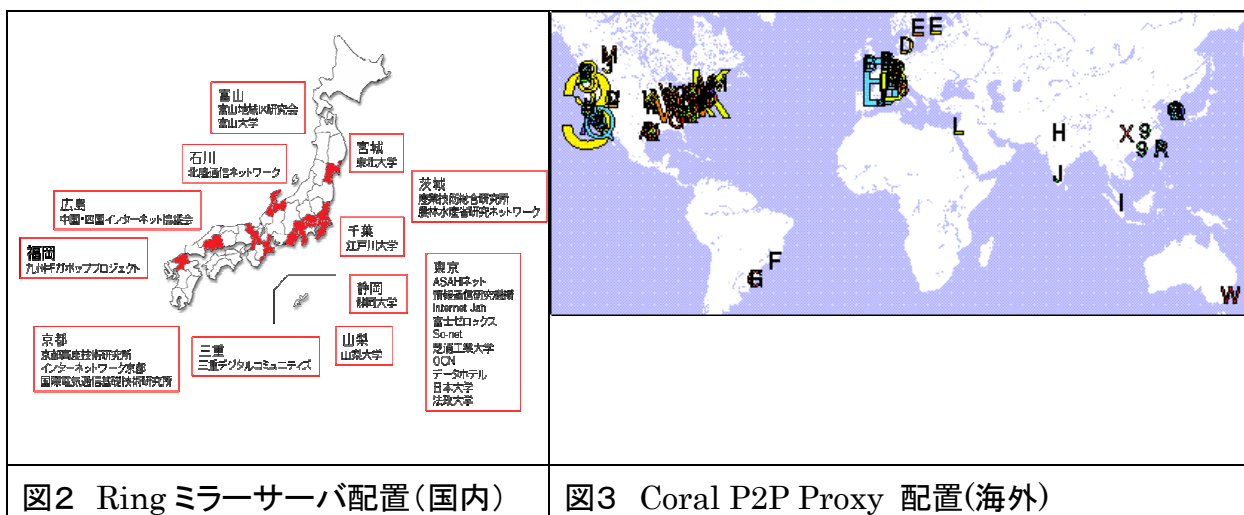


図1 分割圧縮ファイルの構成

今回の実装では分割圧縮ブロックファイルをまとめる手段として仮想ファイルシステム FUSE(Filesystem in Userspace)を利用した。

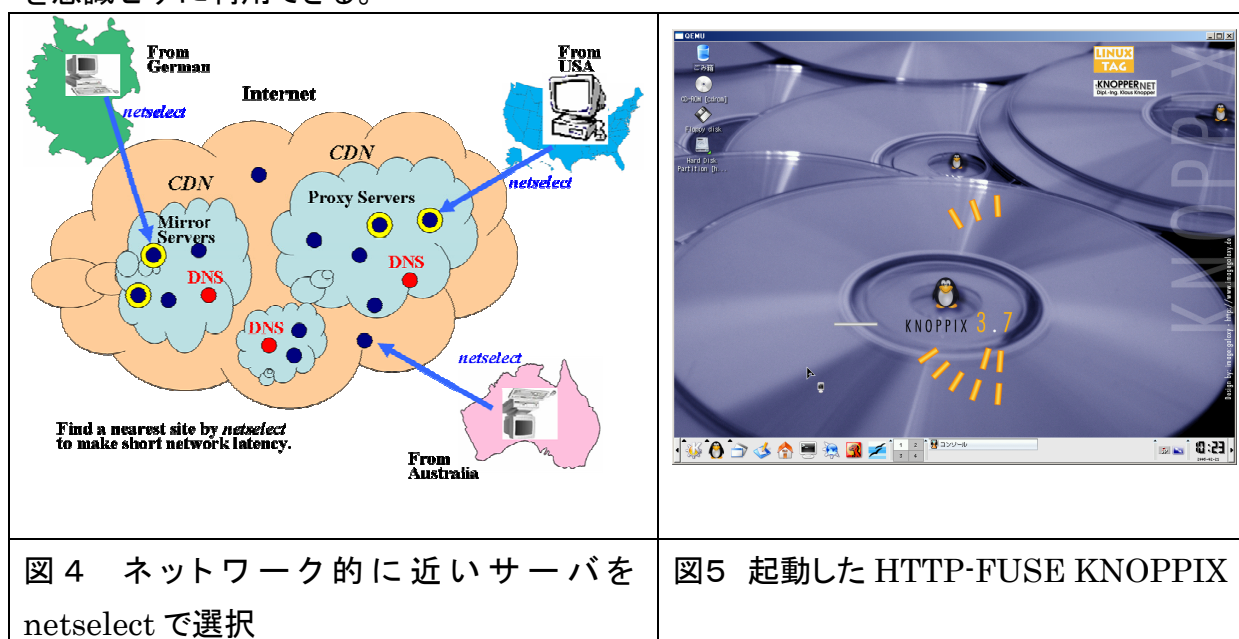
<http://fuse.sourceforge.net/>

分割圧縮ブロックファイルは HTTP により、ミラーサーバや Proxy サーバによって配布される(図2、3)。現在、国内では HTTP ミラーサーバの ring プロジェクト (<http://www.ring.gr.jp>)を分割圧縮ブロックファイルが配信されている。また、海外ではニューヨーク大学で開発されている P2P Proxy の coral プロジェクト (<http://www.coralcdn.org/>)により配信を行えるようになっている。これらのサーバ郡を利用することで、サーバによるボトルネックを分散することが可能になった。



さらに、HTTP-FUSE KNOPPIX はこれらの配信サーバから最も近いものを netselect で選択して、高速に起動する仕組みを加えた。これにより、ユーザには世界中どこの HTTP サーバからかは意識させずに KNOPPIX をできるだけ高速起動できるようになった(図4)。起動画面は図5に示す。通常の KNOPPIX と同様に使え、HTTP からのブート

を意識せずに利用できる。



4. 従来の技術(または機能)との相違

米国ベル研究所 (Lucent Technologies) で開発されている次世代 OS の Plan9 (<http://www.cs.bell-labs.com/plan9dist/>) の archival block storage server である “Venti” は、データの読み書きをデバイス固有のアドレスではなく、書き込むブロックの中身についてユニークなハッシュ識別子を基に行う。この構造は、HTTP-FUSE KNOPPIX でもブロック内容の MD5 ハッシュによるストレージ管理を行っている点で似ているが、ブロックの保存先がファイルである点が大きく異なる。ファイルにすることで HTTP 配信が可能となり、Venti で用意している特殊なプロトコルは不要である。FUSE-cloop ではファイル配信さえ行えればよく、サーバ側に特殊なソフトウェアを想定しない。また、ファイル形式ならコピーも簡単に作成でき、proxy によるファイルコピーを使った Internet 配信を活用することができる。

5. 期待される効果

本開発により、ルートファイルシステムを Internet から取得してブートできる Linux (KNOPPIX) を開発した。LAN 環境では同様なことができたが、Internet を介してブートする OS は聞いたことがない。ベースにした KNOPPIX においても Internet から直接ブートでき、透過的にブロックファイルを利用できる KNOPPIX は世界初である。

KNOPPIX では、一部のアプリケーションの更新でも全体の作り直しをする必要があったが、今回の開発で部分更新が可能になった。これにより今まで KNOPPIX のカスタマイズを行っても 700MB のダウンロードサーバが手配できないことがあったが、その問題が解決された。

6. 普及(または活用)の見通し

開発成果は下記 URL で公開している。

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

この成果は既に Linux&UNIX 関連雑誌数誌で紹介され、開発ソフトウェアの添付もされている。海外ではドイツの KNOPPIX フォーラムに開発を紹介したところ、幾つかの利用者の応答があり、交流がつついている

<http://www.knoppix.net/forum/viewtopic.php?t=18285>

また、Linux Conference 2005 の論文「HTTP-FUSE KNOPPIX」としても採択されている。

<http://lc.linux.or.jp/lc2005/01.html>

7. 開発者名(所属)

須崎有康 (産業技術総合研究所 情報技術研究部門)

岡島純 (デジタルインフラ)