社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＱＡＩ－25（4）
2007／10／25

[パネル討論]
開発者の視点から見たサーバアーキテクチャの現状と今後の課題

河合 栄治

奈良先端科学技術大学院大学 情報科学研究科
〒630-0192 奈良県生駒市高山町 8916 番地の 5

E-mail: eiji-ka@is.naist.jp

あらまし 近年，サーバホストへの複数コアの搭載，ネットワークインタフェースの高速化，仮想化技術の導入などにより，サーバ実装技術および運用環境が大きく変化しようとしている．本発表では，これらの変化についてサーバ開発者の視点から整理し，今後のサーバアーキテクチャについて考察を加える．
キーワード ネットワークサーバ，アーキテクチャ，高速化，仮想化

[Panel Discussion]
Server Architectures from a Developer's Perspective

Eiji KAWAI

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-0192 Japan

E-mail: eiji-ka@is.naist.jp

Abstract A wide variety of emerging technologies such as multi-core processors, high bandwidth network interfaces, and virtual machines have a great impact on server implementation techniques and server operation environments. In this presentation, we discuss those issues from the viewpoint of server developers and give consideration of the future server architectures.
Keyword network server, architecture, high-speed service, virtualization

開発者の視点から見たサーバアーキテクチャの現状と今後の課題

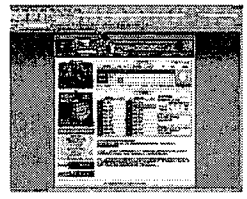
奈良先端科学技術大学院大学
河合栄治

サーバアーキテクチャの現状

サーバを取り巻く環境の変化

- クライアントの圧倒的な増加
  - 家庭へのインターネット接続の普及・広帯域化
  - 携帯端末の普及、定額課金化
- サービス技術の変化
  - flashのデファクト化
  - Ajax的サービスの出現
- サービスの変化
  - 滞在型Webサービスの普及
  - コンテンツの作りも変化

滞在型Webコンテンツの一例：朝日放送甲子園2007Webサイト



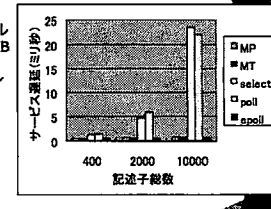
- 基本構成
  - 全体ページ
  - スコアボード
    - Flashにより5秒おきに更新
  - キャプチャ映像
    - JavaScriptで5秒おきに更新
- このページを上げておけば、いつでも最新情報を閲覧できる

サーバの実装論＝サーバアーキテクチャ

- （レガシーな）アーキテクチャ論
  - 要求処理の多重化をどう実装するか
    - I/Oブロックを回避する仕組みが必要
  - 基本形
    - MP/MT（マルチプロセス・スレッド）
    - ポーリング
    - イベント駆動
- 性能・スケーラビリティ
  - MP/MT < ポーリング < イベント駆動
    - 本当か？
    - ApacheはMP/MTだけど、遅いのか？

マイクロベンチマークによる比較

- マイクロベンチマーク
  - HTTPライクな通信モデル
    - サイズ：128Byte/10KB
    - 要求レート：2000/秒
  - MP/MTではコネクション数の分だけ生成
- 結果
  - ポーリングは遅い
    - もはや使う意味はない
  - MP/MTも十分遅い
    - O(1)スケジューラ
    - しかし、10000コネクションでも不十分な局面も…



最近のサーバアーキテクチャ

- サーバのカーネル実装
  - 例：Linux TUXサーバ（khttpd）
  - U-Kオーバヘッドの削減
  - セキュリティの問題
    - 単純なファイル転送向き
- 非対称スレッド実装
  - 例：Chamomile（甲子園用オリジナルサーバ）
  - 処理単位をモジュール化し、スレッドを割り当てる
  - フィードバック機構による処理レートの制御も可能
  - 実装フレームワークが複雑
    - モジュール化されても、モジュール間のインタラクション管理が複雑に

サーバアーキテクチャの今後の課題

下位レイヤの大幅な変化

- ハードウエアの高度化
  - プロセッサのMulti-core化、Many-core化
  - 10GbE対応
- 仮想化技術の導入
  - ハードウエアによる仮想化
  - ソフトウエアによる仮想化
- 動的な追従性のさらなる向上
  - 単なるレイヤ破壊ではなく
  - 単なるパラメータチューニングではなく
  - 堅牢性と柔軟性の両立

ネットワークの高速化における課題

- 次のボトルネックはメモリI/O？
  - Unixのソケット実装では、4Gbps程度で限界？
    - メモリアクセスが数回発生
    - メモリ帯域とアクセス回数でスループット上限が決定される
  - キャッシュフレンドリーなプロトコルスタック実装
  - まだまだメモリ帯域は向上する？
- Infiniband/RDMA技術
  - Zero-copyの徹底
    - I/Oにタグ付けをし、U-K間の共有メモリを実現
  - 使用例：iVisto
  - 今後、サーバプログラミングモデルが大きく変化していく？

仮想化技術の課題

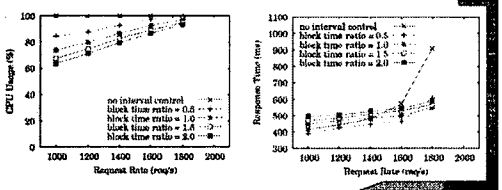
- 仮想化技術で性能問題を解決するのは難しい
  - あらゆる計算資源が仮想化される
    - プロセッサ、メモリ、ネットワーク、ストレージ
  - しかも動的に変化する
    - プロセッサやメモリが増減する
  - 処理コストに関する常識が大きく変化する
    - MMUの操作は非常に高価に
  - 資源管理ポリシーの衝突
    - 例：メモリ管理
- 資源の動的な変化に追従できるシステムが必要
  - 高性能システム＝ハードウエア資源の極限利用
  - 急激な変化についていけない

省電力化の課題

- サーバの消費電力問題
  - coreが増えると、消費電力が増える
  - 例：甲子園サーバ
    - 1Uサーバ、2.8GHz 2core x 2で250W程度消費（実測値）
    - 8台で電力1系統（20A）ほぼ使い切る
  - 甲子園サーバでは、12kコネクションを抱えていても、CPUサイクルの90%程度はアイドル
- サーバプログラムから電力消費を制御する仕組みが欲しい
  - ユーザ遅延1ms増加で、25%消費電力抑制とか
  - OSの汎用的な機能として実装するには？

省電力化の課題（続き）

- サーバでsleepしたら速くなった例も



OSは変化から取り残されている？

- ハードウエア技術開発のサイクルとサービスのライフサイクルが乖離
  - サーバにはレガシーなプラットホームを使い続けたい？
  - レイヤギャップがどんどん拡大
  - 機能は実装されても、性能管理は困難に
    - 特に仮想環境では性能管理の問題が顕著に

まとめ

- サービス技術も変化している
  - サービスの実装方法やユーザの振るまい
  - サーバにとってはどんどんつらい状況に…
- サーバ実装における課題
  - NW高速化
  - 仮想化
  - 省電力化
- 最大の課題はOSの進化論をどう扱うか
  - サーバにとってはレガシーなままでよいのか？