

佐藤豊

[y.sato@aist.go.jp](mailto:y.sato@aist.go.jp)

<http://delegate.org>

2009/10/14



DeleGateのこれまで | これから



# 内容

- DeleGateのこれまで
  - 15年と7ヶ月の歩み
- 今年度前期の進捗
  - 環境整備の日々                      夢の開発環境が！
  - 従来版v.9の仕上げ                  マルチスレッドなバグがぞろぞろと...
  - 新方式v.10の仕込み                10倍高速に！
- これからのDeleGate
  - 今年度後期～中長期計画

# 目的・目標・手法

- 広く納税者の直接の役に立つソフトウェア
  - 産学への基盤的ソフトウェア素材の提供
  - 海外への成果の発信・貢献
  - 国産基盤ソフトウェア技術の維持(種火)
- 
- コストパフォーマンスのよい研究開発
  - 税起源の研究資金にはなるべく頼らない
  - 主に海外からライセンス収入を得る
  - 長期的に安定した基盤を地道に維持提供
  - 臨機応変に社会的要請に応える

# 私とDeleGateのこれまで

## 1980 C/Unixと出会う

疎結合マルチプロセッサ(PDP11x2)  
分散OSの通信メカニズム  
プログラミング支援システムcosmosの開発

大型機→ミニコン

## 1987 電総研入所

ポストマスタ(JUNET/CSNET/ドメイン化)

インオペ大プロでOSIIに関わる  
ニュース・メールリーダーvin/cosmosの開発  
→NNTP/POP/X viabus,via-UIMS,onew  
→標準化に参加:MIME/ISO-2022-JP

ワークステーション

## 1994 DeleGate開始

科技庁出向(IMnet)

ワークステーション→PC  
DeleGateの開発開始 (←vin/cosmos,viabus,MIME)  
→Gopher/HTTP/NNTP/FTP/Telnet/SMTP/SSL

## 2001 産総研へ

パソコンUnixに転向(低価格高性能化)  
天久保サイトに光ファイバーがっ!  
Mac(OS X)ユーザに転向  
組み込み系への対応(Zaurus,KuroBox,WinCE)

携帯端末

## 2008 新ITRIへ

グループに所属

梅園サイトの再整備  
DeleGateを新開発カーネルベースに移行開始

7年周期に何かが...

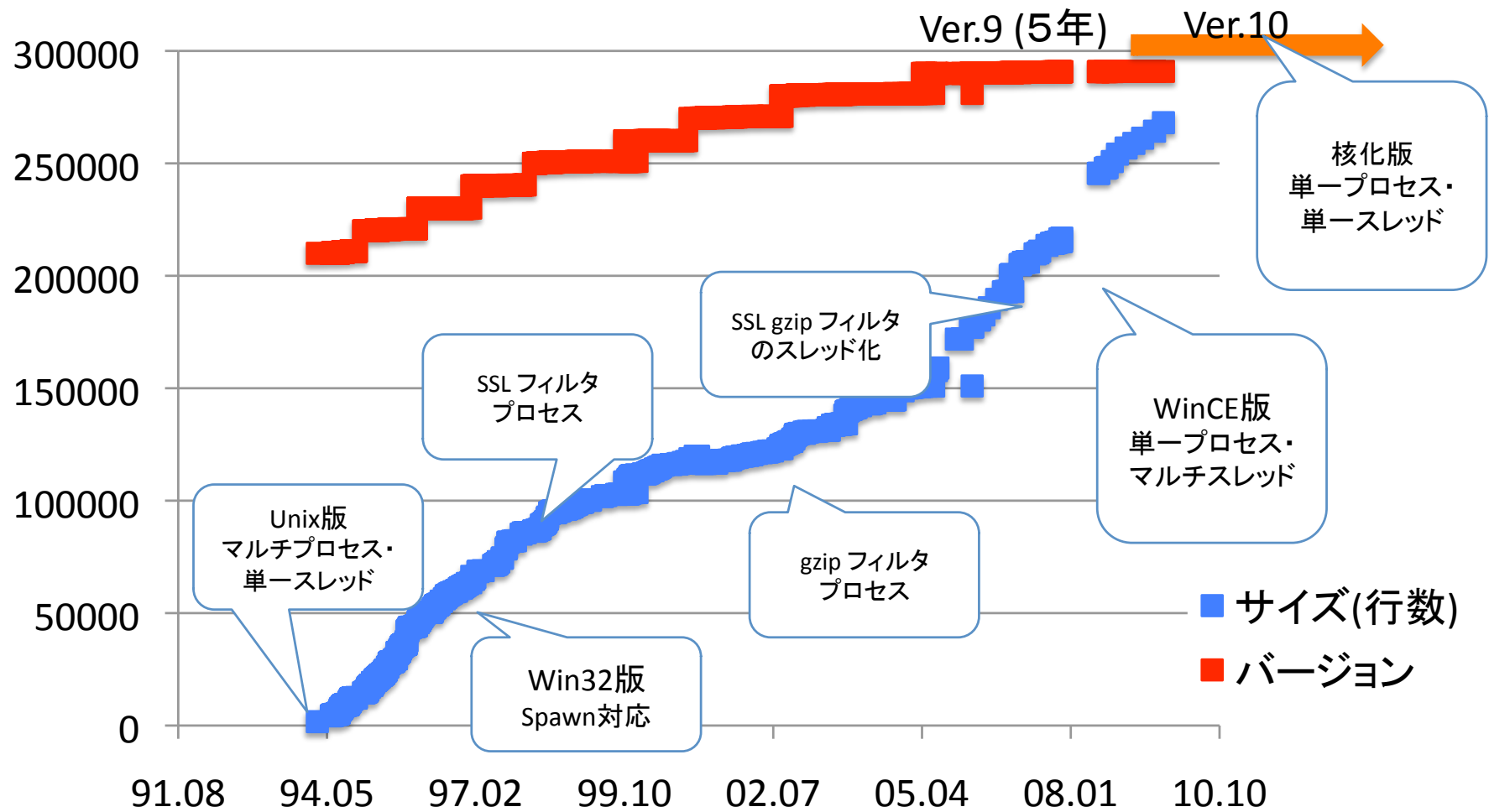


# DeleGateとは

- プロキシサーバ
  - アプリケーションレベルゲートウェイ
- 多目的
  - 高速化、安定化、安全化、匿名化、適合化、盗聴、...
- マルチプロトコル対応
  - HTTP,FTP,Telnet,NNTP,SMTP,POP,SOCKS,DNS,SSL, ...
- マルチプラットフォーム対応
  - Unix, Windows, WindowsCE
- 持続的成長 15年と7ヶ月 (1994年～)
- 中規模プログラム 25万行/C
- 無償配布先ドメイン数 4万/182ヶ国
- 国内外に商用ライセンス (2004年～)

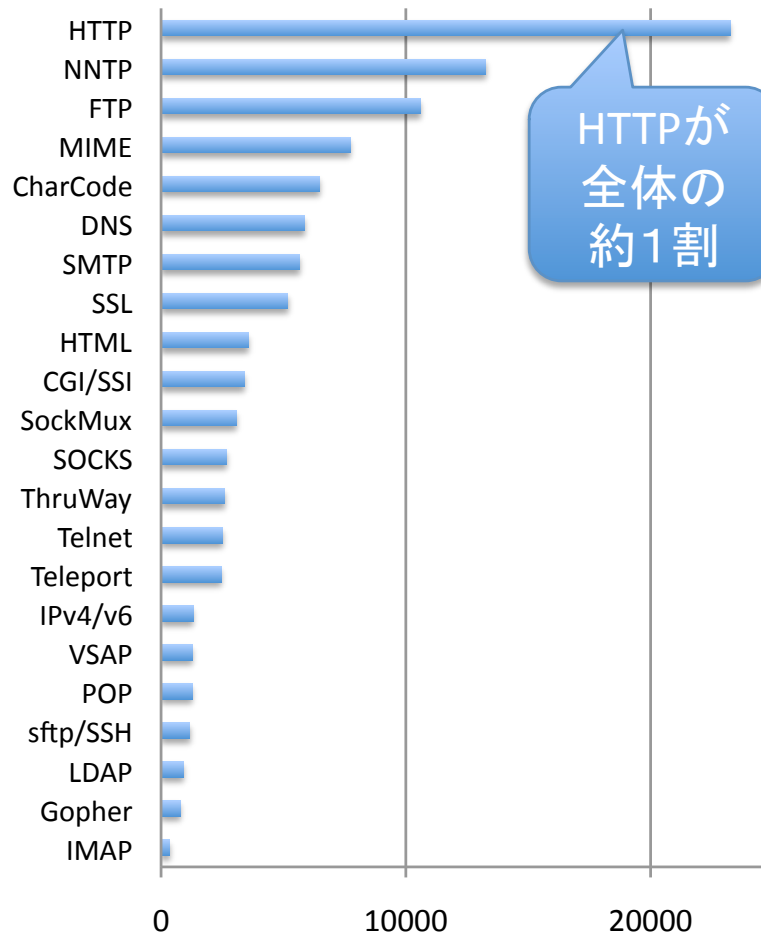


# DeleGate各版のプログラムサイズ

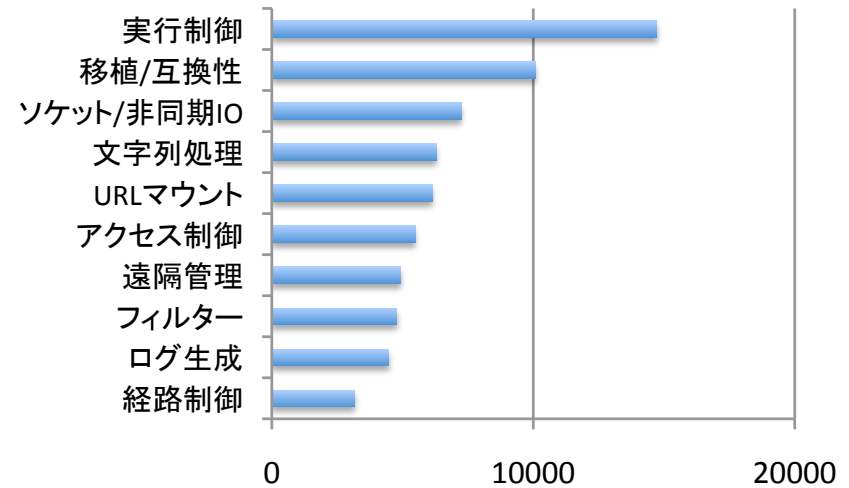


# プログラムコードの内訳

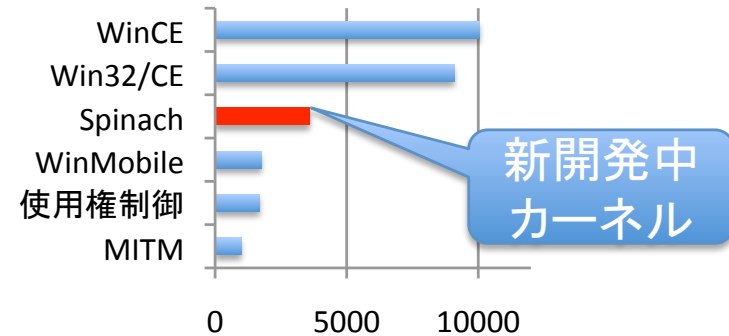
## プロトコル／データ形式処理



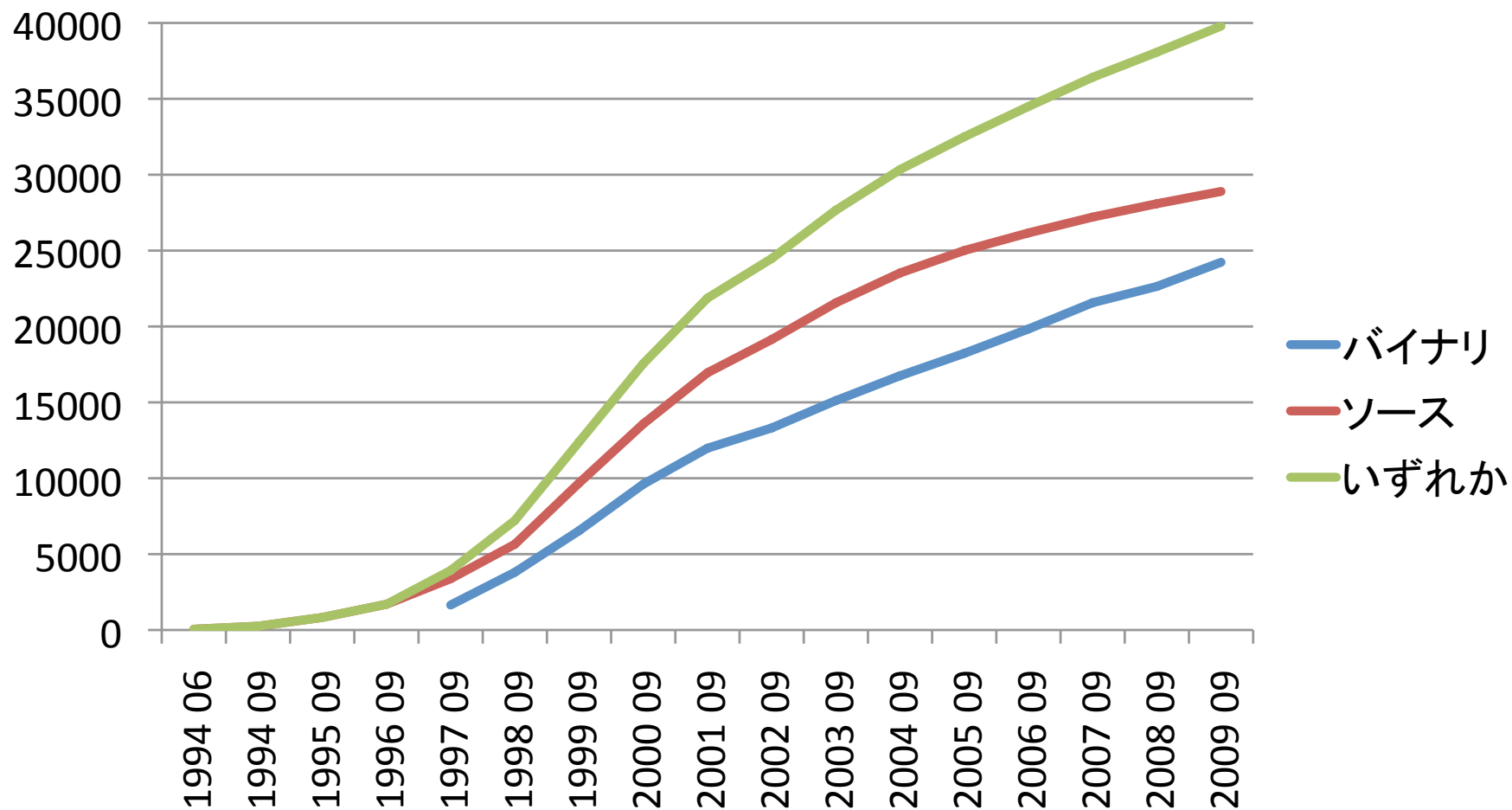
## 共通基盤的コード



## 非／未公開コード



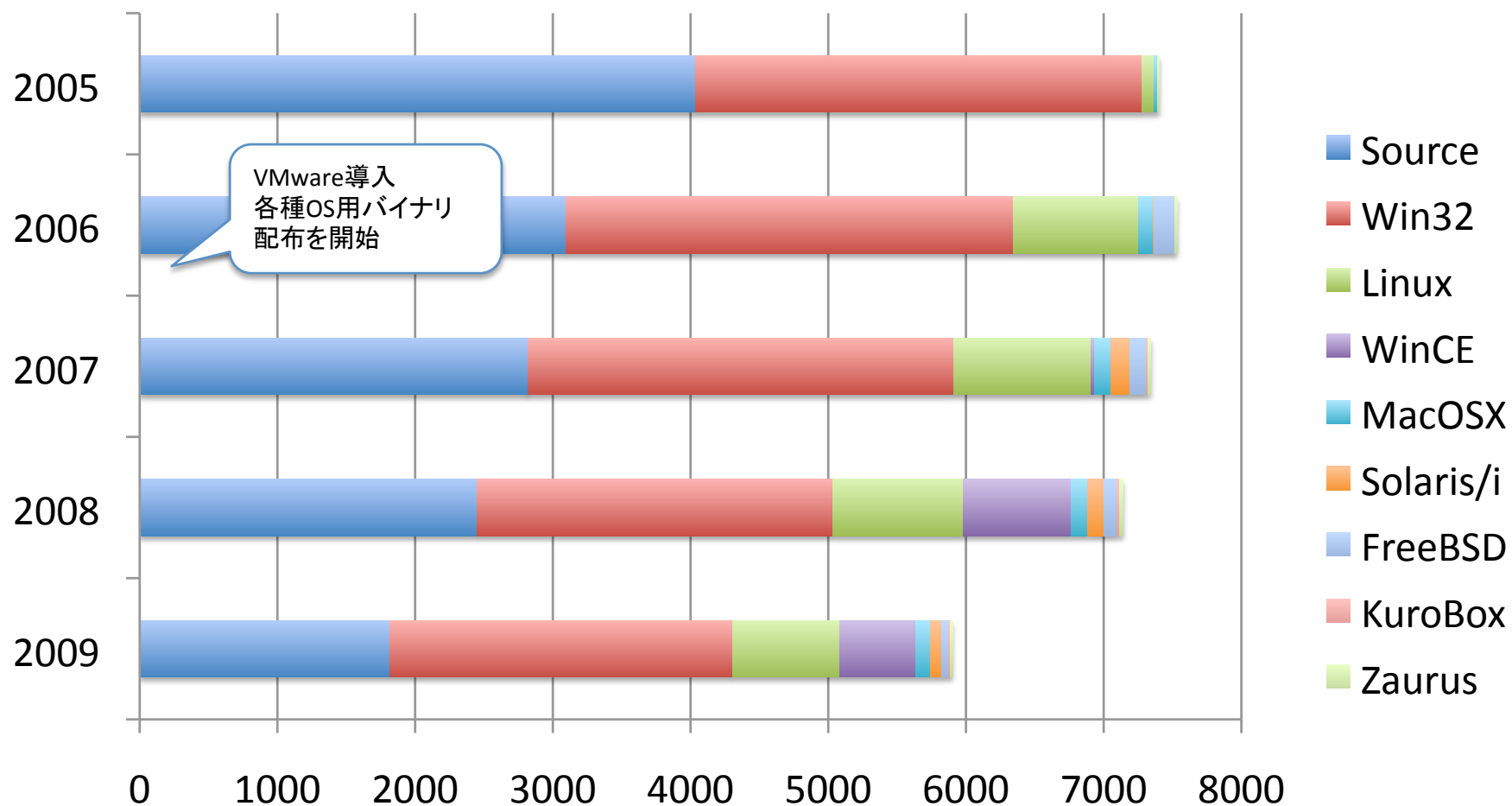
# ソース・バイナリ配布先サイト数



ソースで配布する時代の終焉

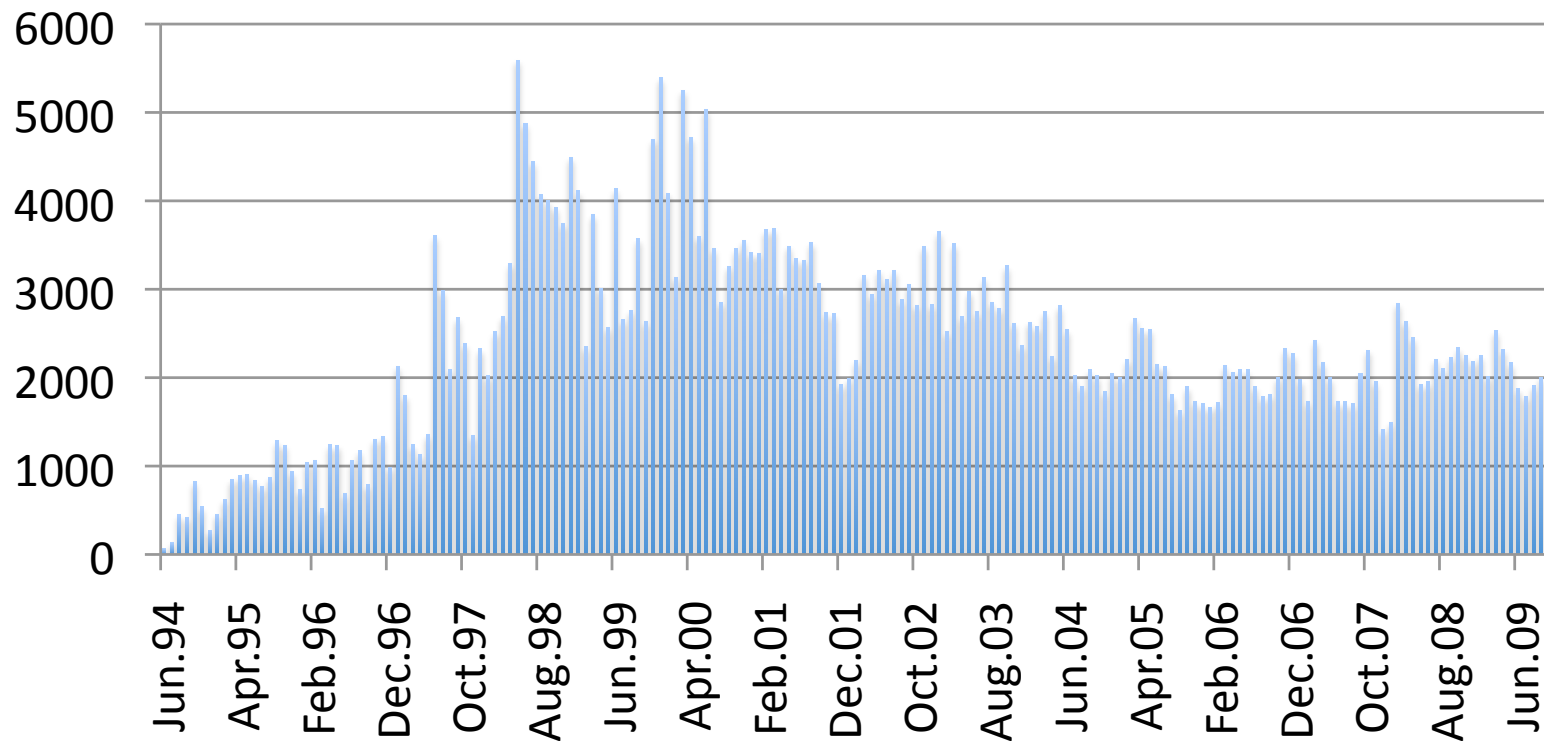


# ソース／バイナリ利用者数の変化



設定済・即利用可形式での配布を

# 月別ダウンロード数

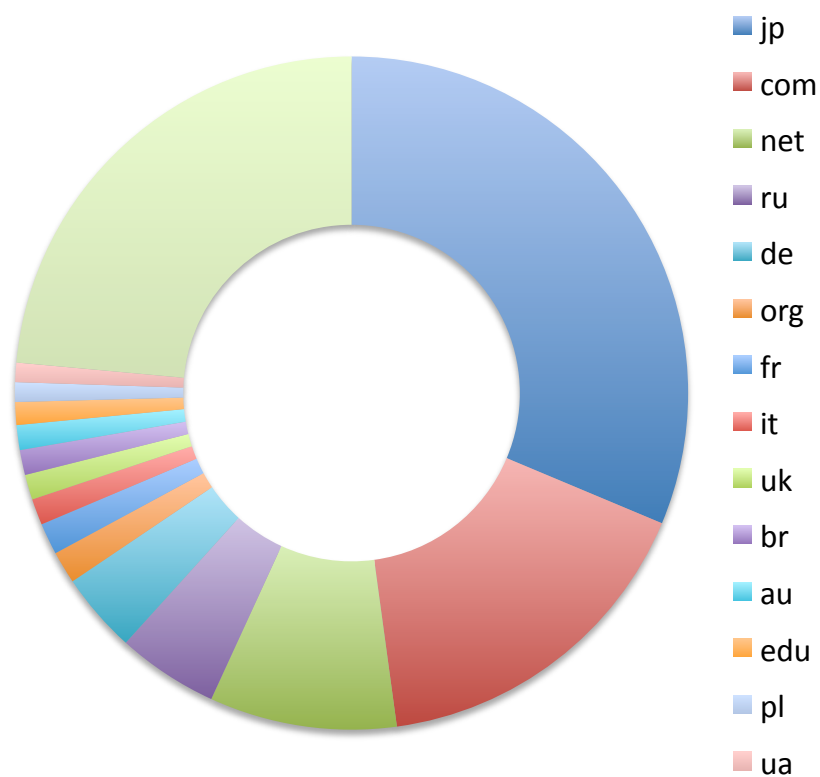


## 実質的ダウンロード数

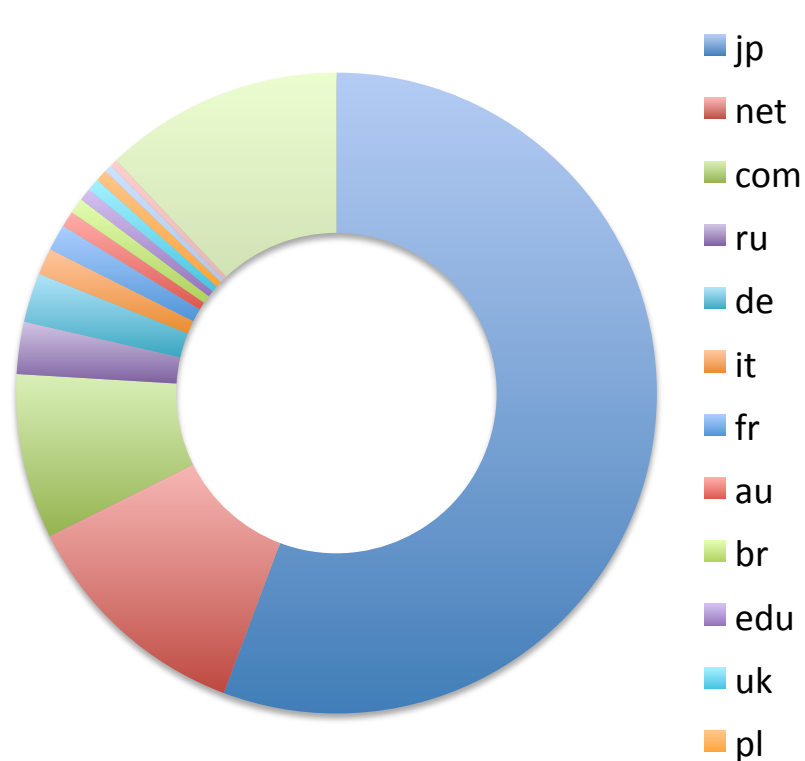
- 同一ホスト・ユーザからの連続・近接ダウンロード除外 -- 70%除外
- ミラーサイト、ロボットサイトからのダウンロード除外 -- 残り30%内の3%除外

# DeleGateの国別利用者の分布

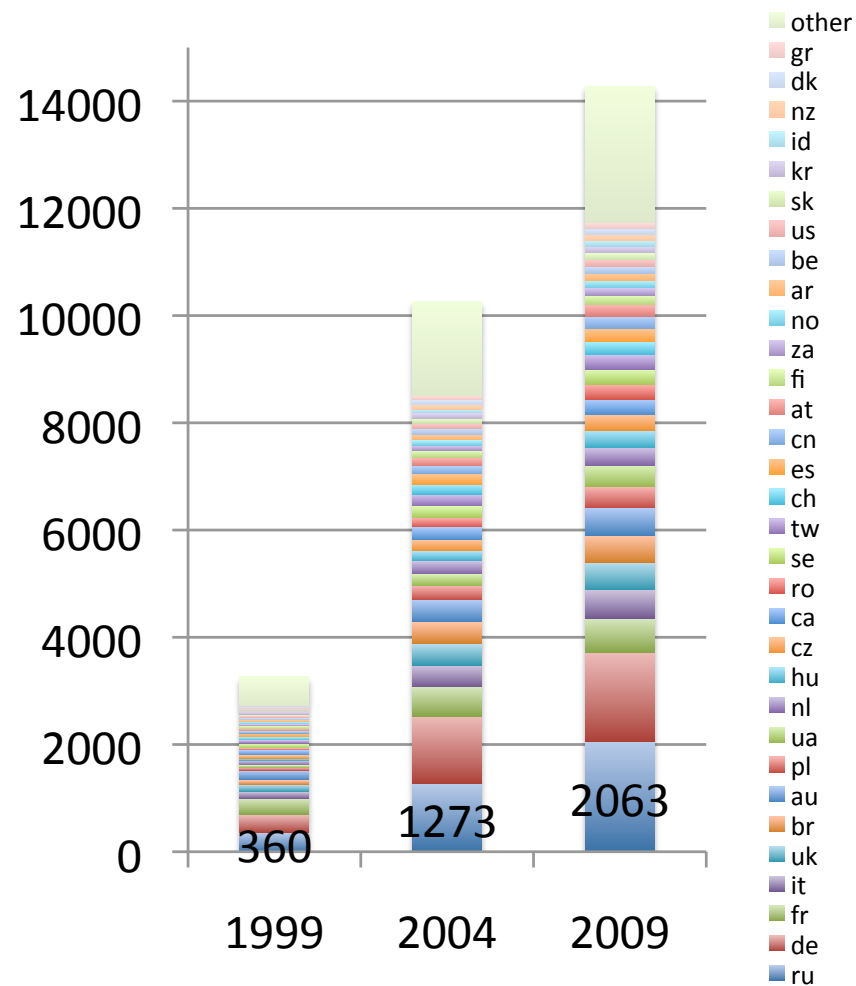
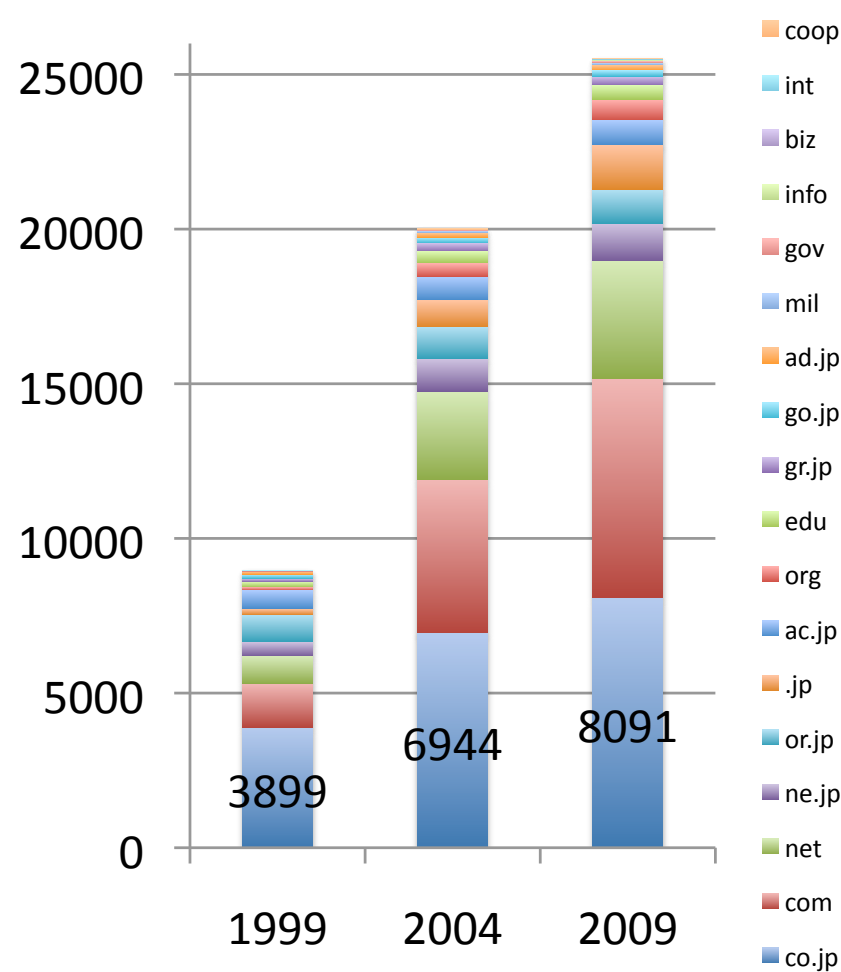
## サイト数



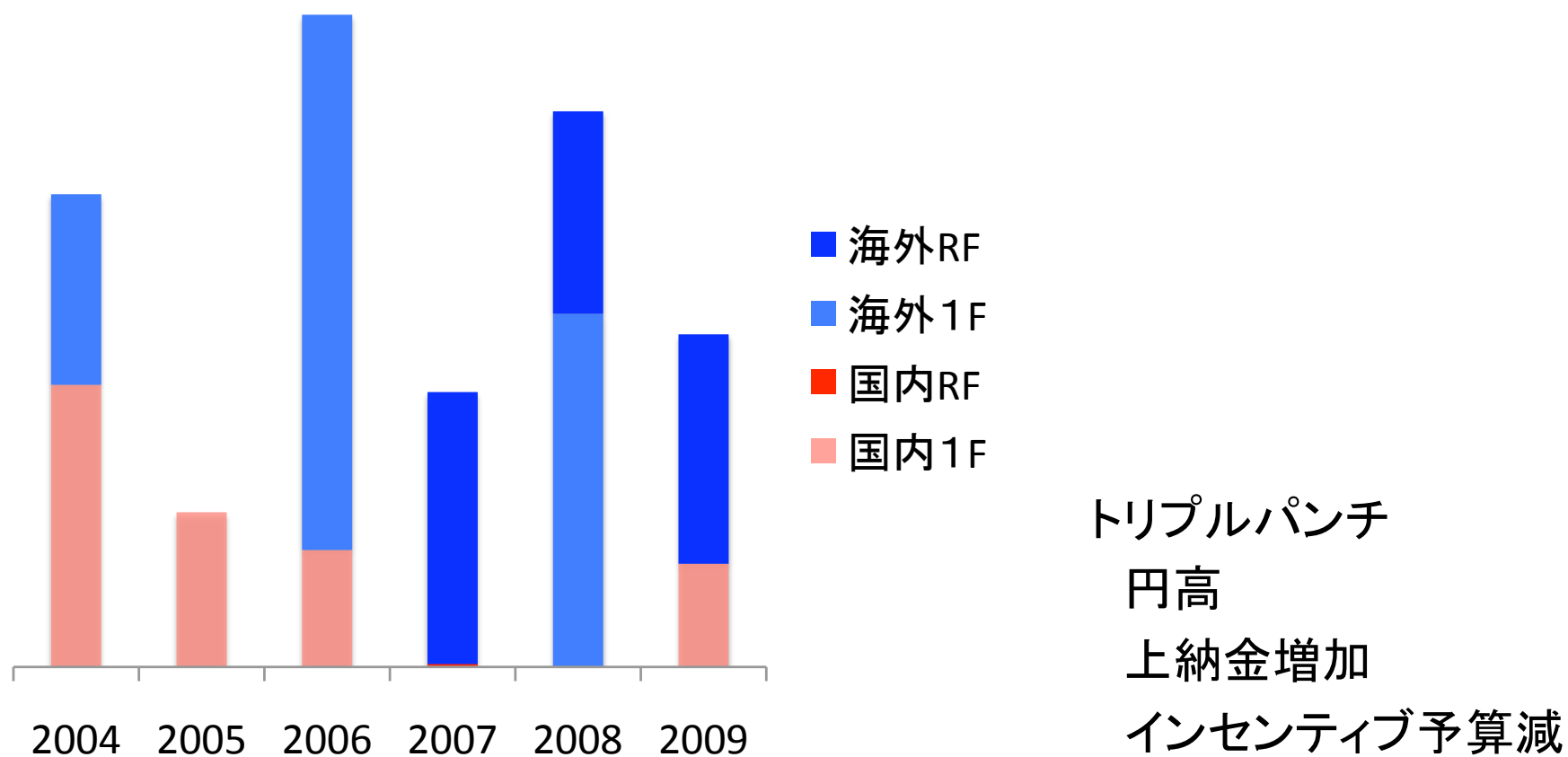
## ホスト数



# DeleGate利用サイトのドメイン内訳



# DeleGateライセンス収入の推移



# 普及の現状

- 日本人による日本人のための、から脱却しきれていない
- 納税者に奉仕する事が第一義なので、それで良いのかも
- ライセンス収入が殆ど海外からなのも、それで良いのかも
- 利用者が多い割には、国内からの技術的フィードバックが少ないのは残念

# DeleGateの2009年度前期

- 開発、テスト、配布環境の整備
- 第9版 従来構造(1994～2009)の仕上げ
- 第10版 新規構造(2009～)の仕込み・試作



# DeleGate10で新機軸へ

- 従来の実装が行き詰まっているわけでは無い
- とは言え、大きな新展開もない
- このまま更に10年間続けても面白くはない
- 従来と異なる使い道を開拓したい
- これまでの実装方法では実現が難しかった事
  - 超小型化・軽量化・高速化 →省エネ/用途拡大
  - 高度な安定性・信頼性向上 →商用拡大
- 全く新しく作り直そう！
- 新設計10年安心的超小型高速通信処理核開発
- 従来の実装は、分解して新核上に移設



# プロキシサーバの基本的処理

- 応用プロトコル／DeleGate

- ステートレス

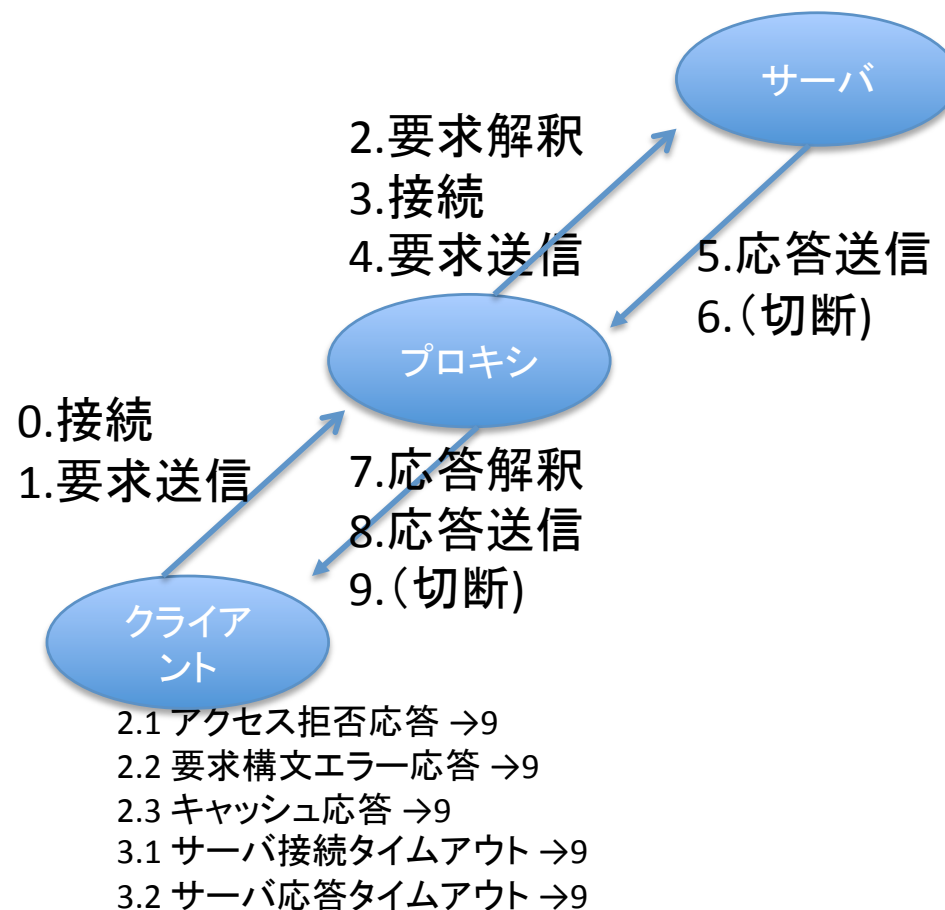
- HTTP 仮想セッション
- Gopher
- DNS (無接続)

- ステートフル

- FTP
- SMTP
- POP
- IMAP
- NNTP

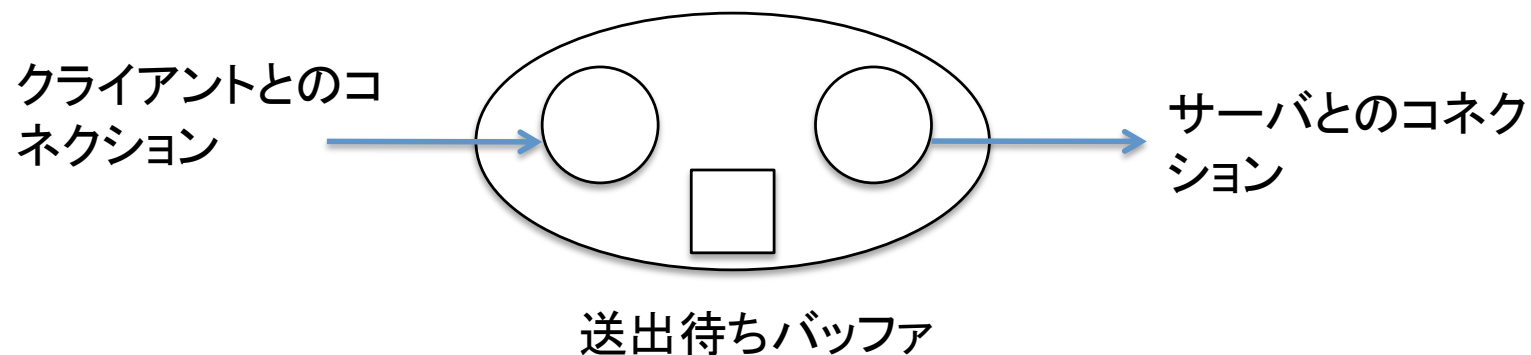
- 特殊

- Telnet (双方向対等)
- SOCKS (透過中継)
- SSL (TLS) (双方向/暗号化)



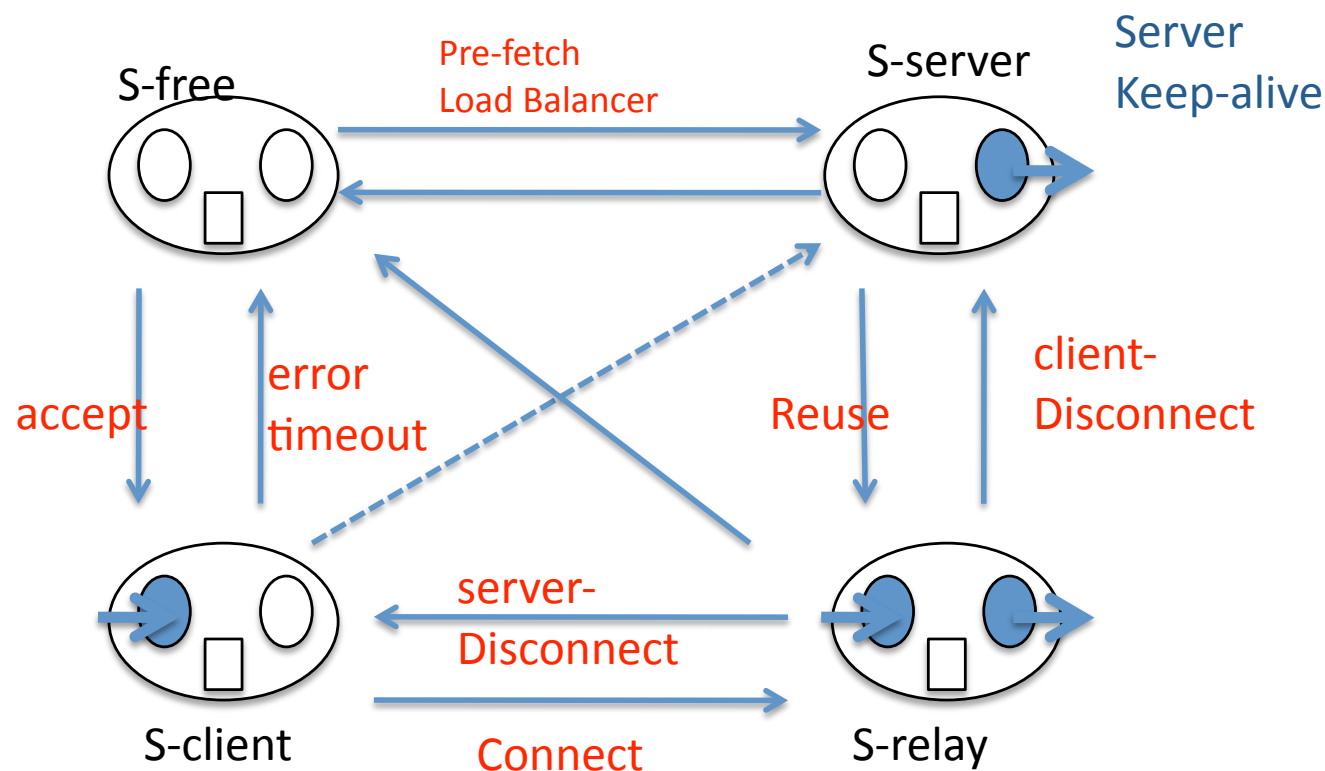
# 新方式の基本構造

- プロキシ単位の基本構造



- 基本的状態
  - クライアントとのコネクションの有無
  - サーバとのコネクションの有無
  - 送信待ちデータの有無と方向
  - 現在の状態を開始した時刻

# 新方式の基本的状態遷移



- タイムアウトによる状態遷移(強制切断)の一元管理
- HTTPサーバ接続の共有再利用 (S-client + S-server)が簡単軽量に実現
- 先行接続等による高速化の可能性

# 新方式核の試作版

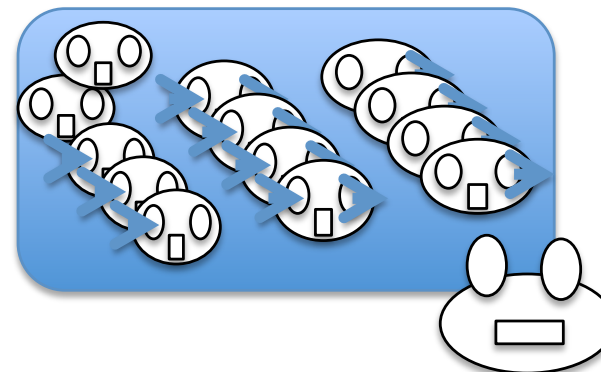
## 開発コード

Spinach – 実行制御用オートマトンのインタプリタ

ソケットの監視、受理、接続、データ中継

Pilsner – 入出力処理用オートマトンのインタプリタ

入力のパターンマッチ／出力データへの変換



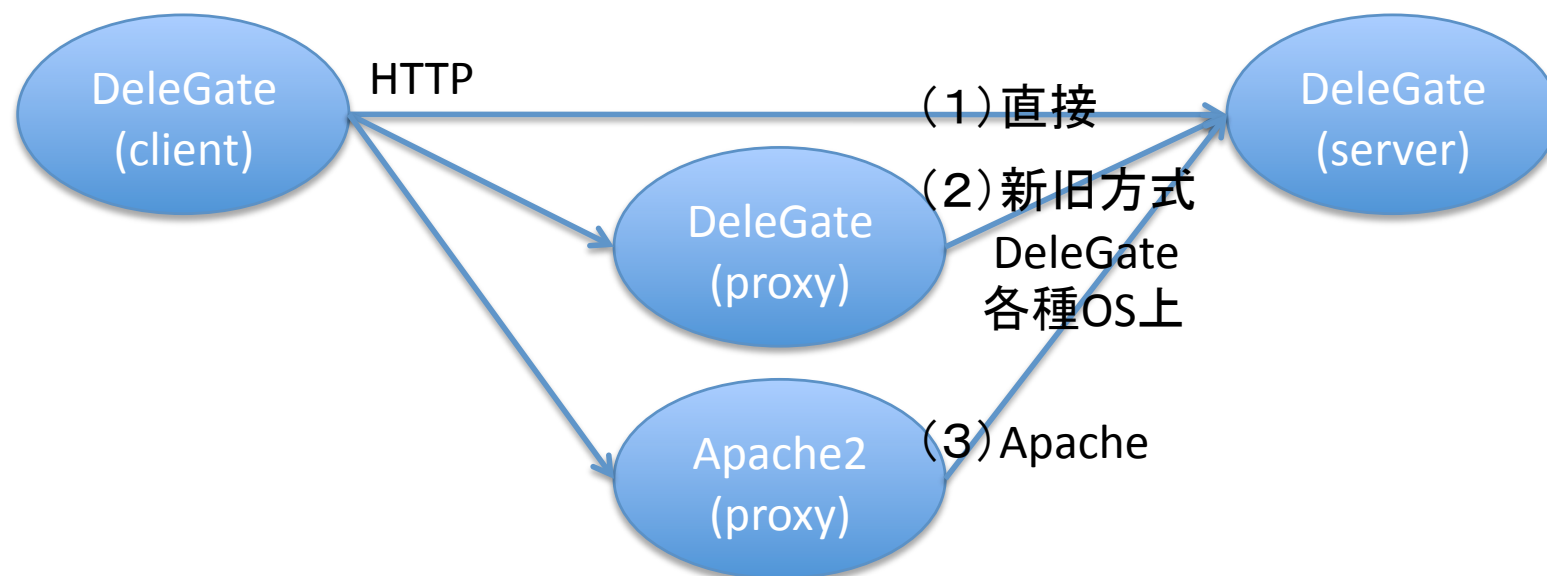
## – 試作版の現状

- Spinachのみ実装、Pilsnerはまだ無い(手書き)
  - 機能はHTTPの中継+FTP/HTTPゲートウェイのみ
  - 約3000行 (バイナリで数十KB)
    - 真核部分1000行
    - 補助的/外部的/例外的処理1000行
    - 手書きHTTP処理1000行(Pilsner化予定)
  - ほとんどの処理がこの中で完了
  - ログ、キャッシュ、経路制御等の、従来版の機能は皆無
- WindowsCE版のバイナリ配布デフォルト設定はこれでリプレース
- ダイアルアップ機能

# 高速化したい部分と達成したい性能

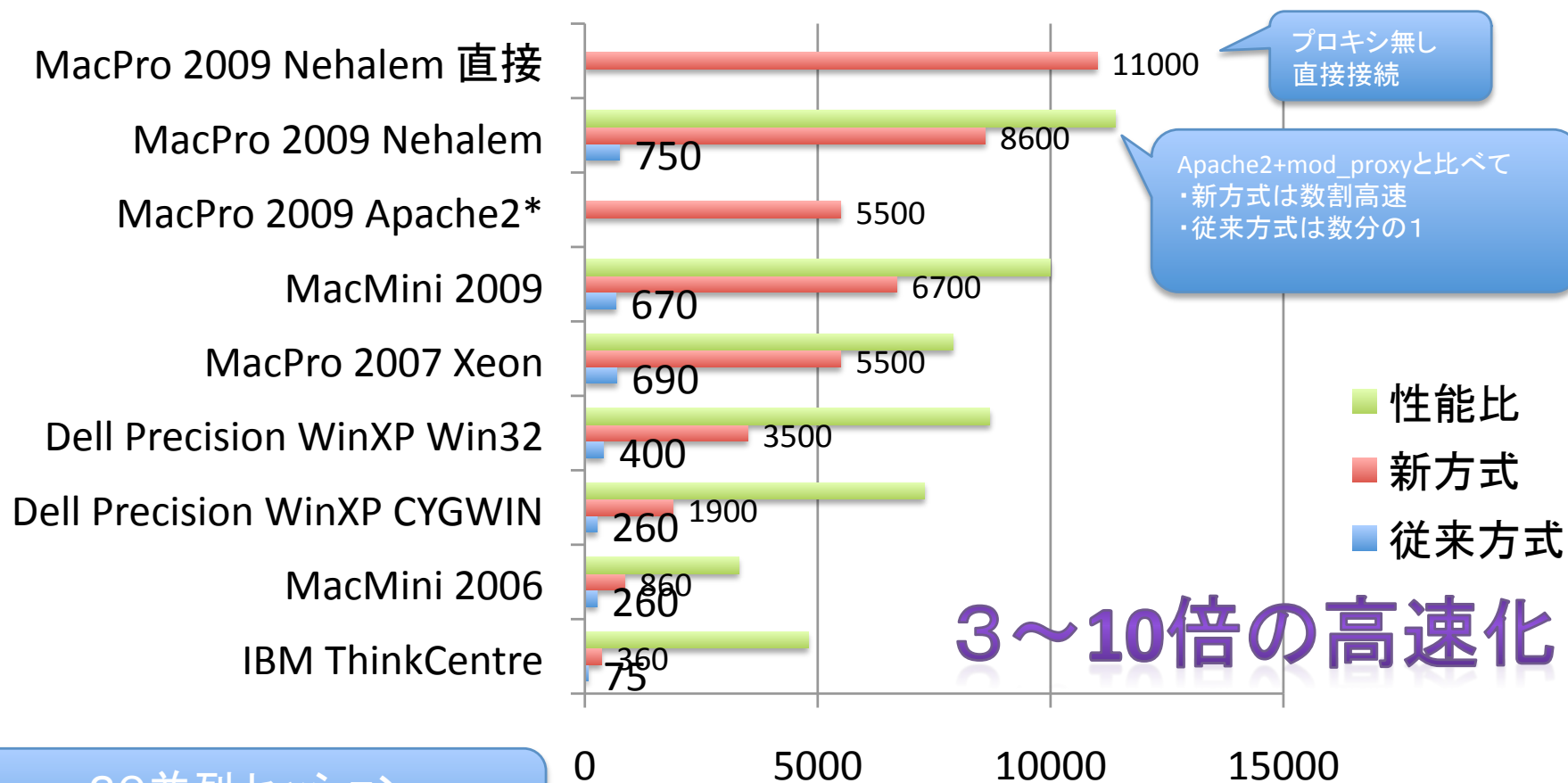
- スループットは実装による差が小さいので置いておく(中継性能)
  - 単にrecv() / send()の繰り返しの性能... のはず
- プロキシの苦手な小型データ多数中継が課題(応答性能)
  - 大量の小さなインラインイメージ
  - 大量の304応答(リロード不要確認応答)の中継など
- 最近のありがちなWebの1ページをさくっと
  - 100インラインイメージ／ページ
  - トータル1MB／ページ
- 当面の性能目標の目安
  - モバイル個人用: 100HTTP/s 1MB/s
  - 組込～小規模用: 1000HTTP/s 10MB/s
  - 本格サーバ向け: 10000HTTP/s 100MB/s

# 測定対象と測定条件



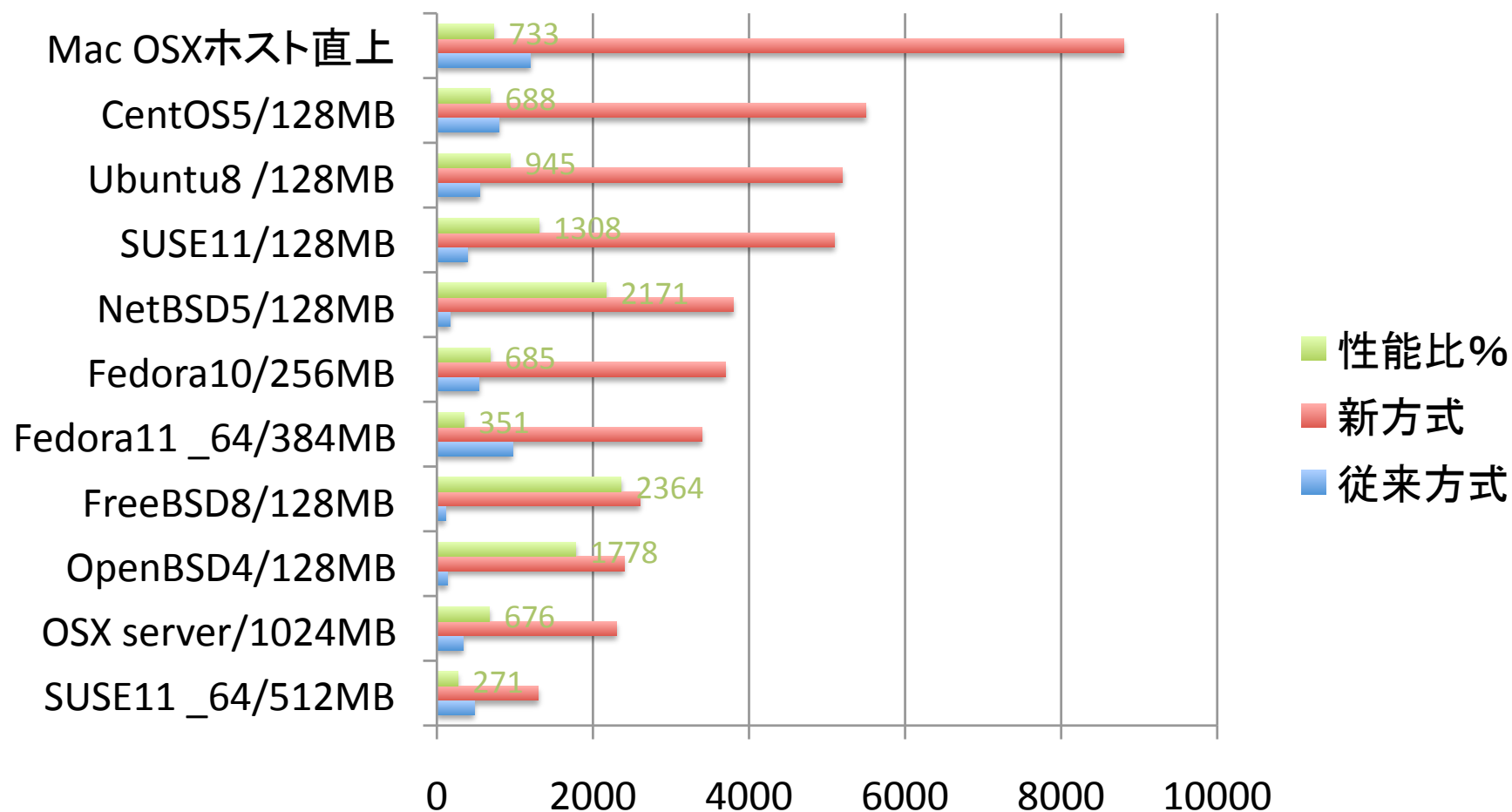
- クライアントとサーバはDeleGate / MacPro2009 (8core Nehalem Xeon)
- Apacheも同ホスト上
- 使用した仮想マシンはVmware Fusion2
- 接続はマシン内部またはギガビットハブ経由
- HTTP/1.1、キャッシュ無し、ログ無し、圧縮無し
- 応答性能の測定： 要求／応答ともに約150バイト
- 中継速度の測定： 約100MBのバイナリファイルの中継

# 新旧方式の応答性能: デスクトップ



最近のマシンほど改善効果が大 ... ネットワーク性能？

# 新旧方式の応答性能:VM上

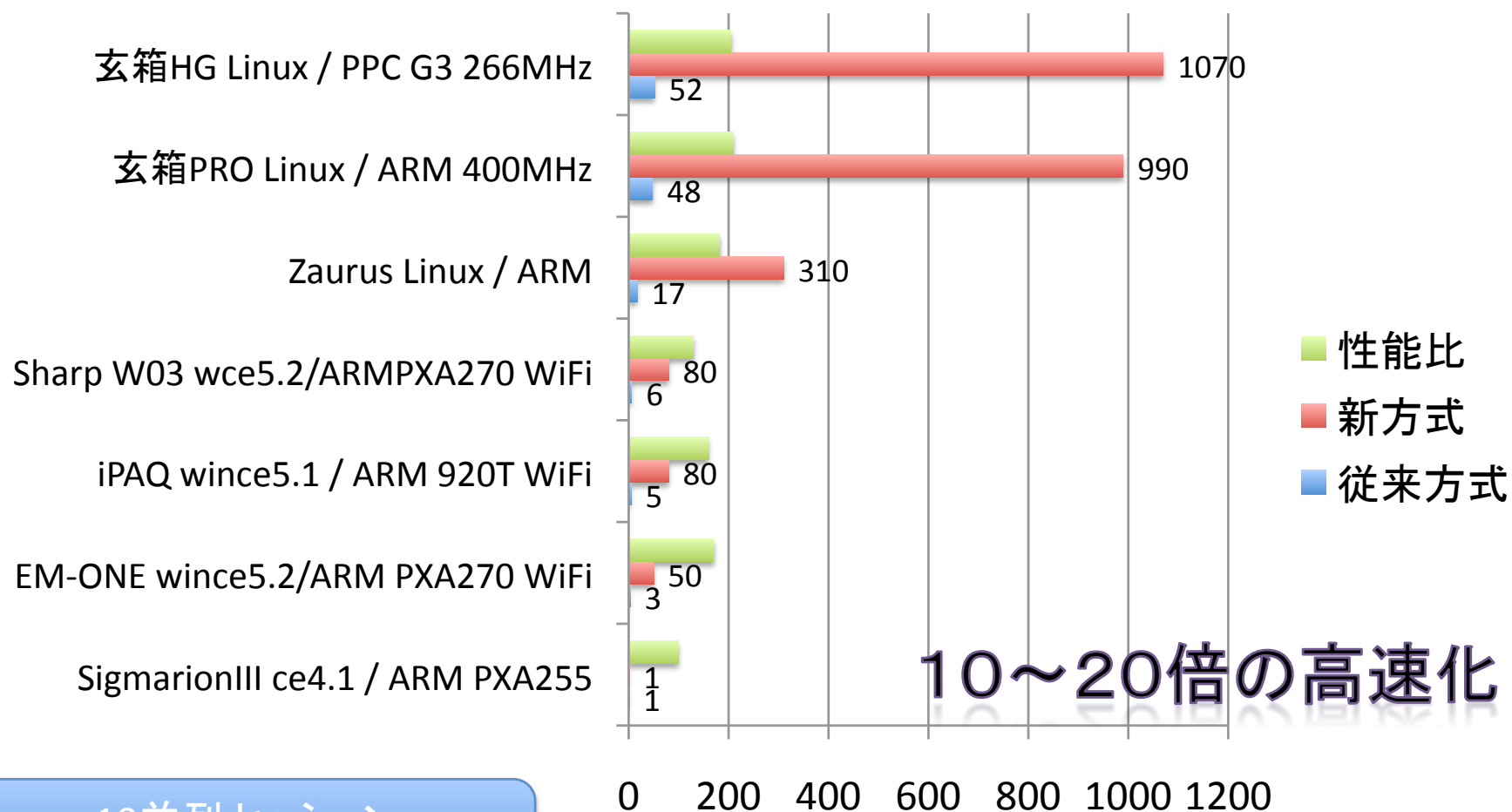


10並列セッション  
各セッション10リクエスト

同一ハード上なのに性能差大 ... proxy実行、VMとの相性



# 新旧方式の応答性能：組込系

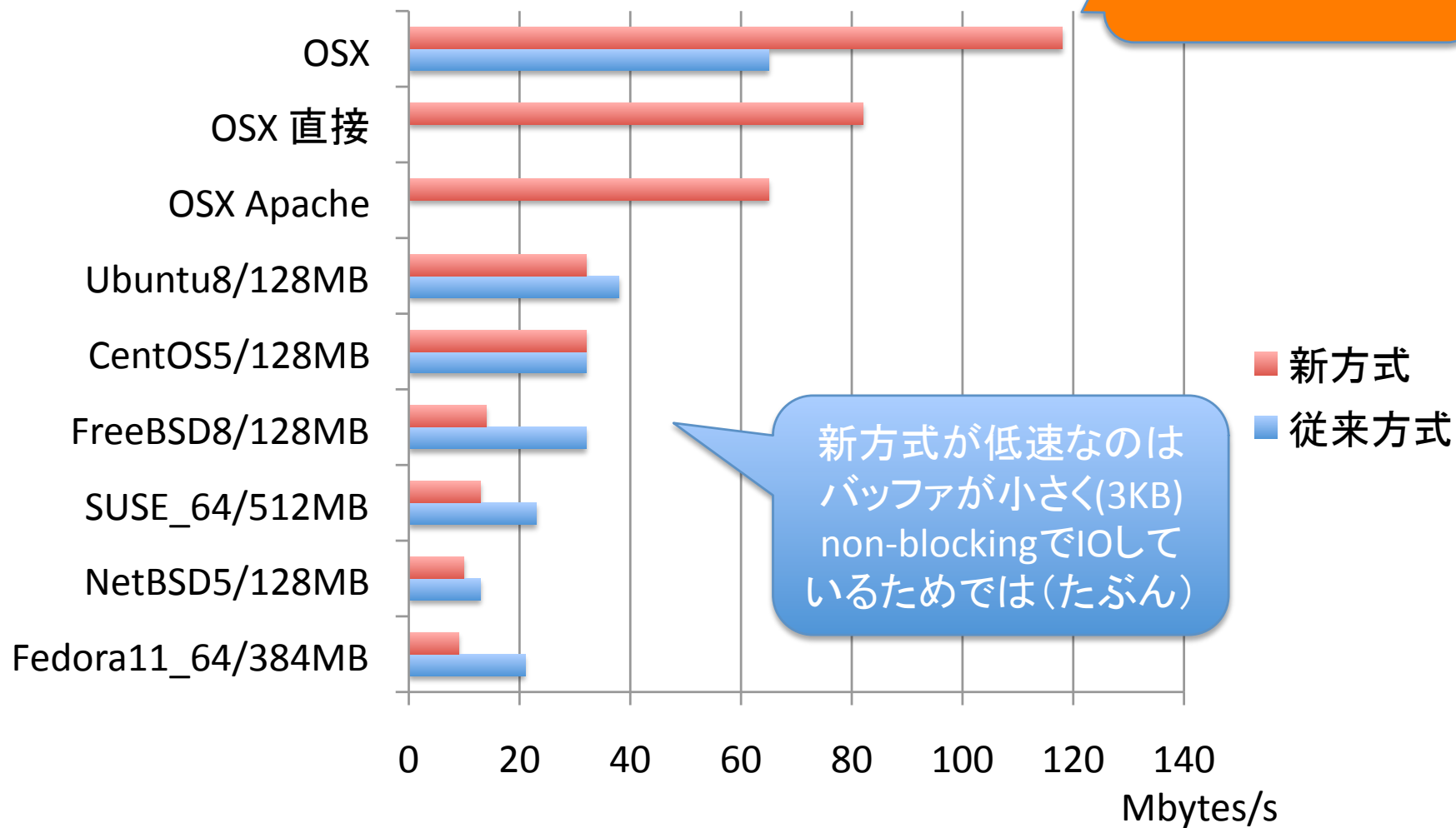


10並列セッション  
各セッション10リクエスト

効果大 ... 小型化で小さなCPUキャッシュが有効に

# 中継速度の比較

なぜ直接より高速？  
ネハレム効果か？



# バッファサイズを大きくしてみると

## VM上のUnixでの性能向上

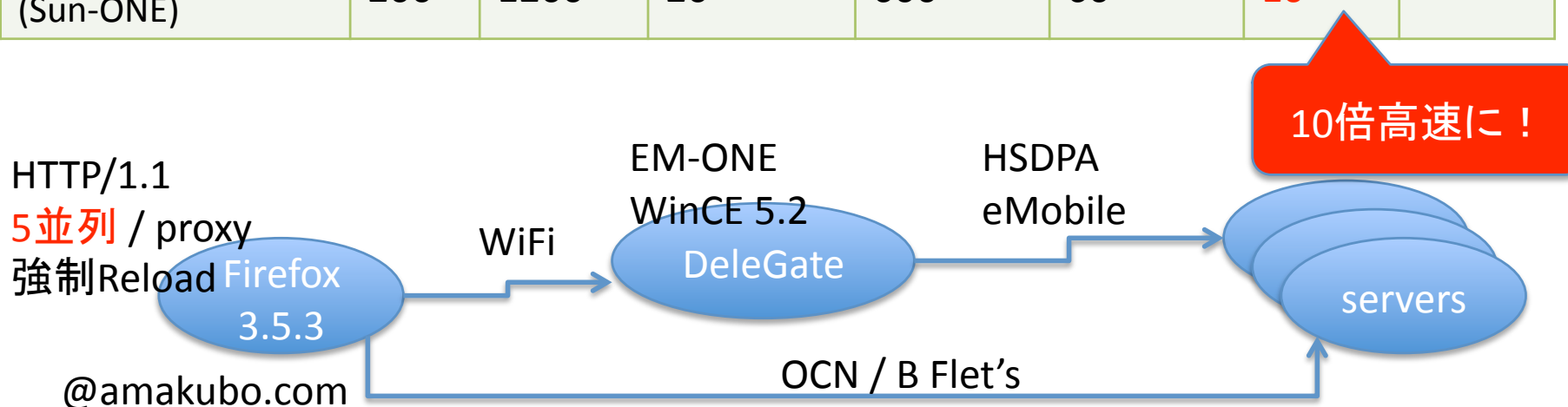
- (未測定)

## Mac OS X 直上での性能向上

- 6Kバイトあたりでサチる
- 最大220MB/s (2Gbps近く)
- バッファを大きくすると、性能が極めて不安定に
- CPUキャッシュ効果か？
- ???

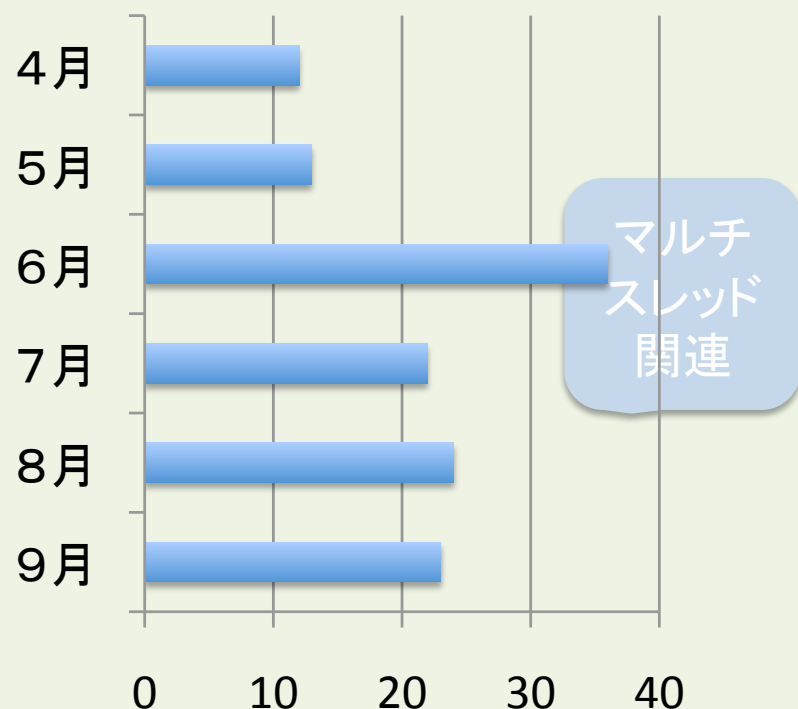
# 実際のサイトへの中継速度測定

	ページ要素数	要素サイズ合計(KB)	最短所要時間 (Flet's光)	従来方式所要時間 (秒)	新方式所要時間 (秒)	性能向上比	16並列新方式 (秒)
<a href="http://www.yahoo.co.jp">www.yahoo.co.jp</a> (keep-aliveせず)	30	230	4	36	16	2	12
<a href="http://www.aist.go.jp">www.aist.go.jp</a> (DeleGate/8.X)	60	250	2	150	13	12	-
store.apple.com/jp	100	500	6	200	30	7	-
<a href="http://www.mlb.com">www.mlb.com</a> (Sun-ONE)	200	1200	20	600	60	10	-



# 従来版のバグ取り作業

バグ修正数



## 今年度前半修正の主要なバグ

- マルチスレッド関連
  - 割り込みによる破壊
  - OSにより様々な再現性
- 64ビットCPU固有
  - 主にスレッドID関連
- WinCE, Win2K, WinNT各固有
  - 主にスレッド同期関連
- OpenBSD, NetBSD, CentOS各固有
  - 動的ライブラリ関連
- Win32上固有低速キャッシュ
- 夏時間/Win32関連
- FTPキャッシュ同時更新破壊

# 並列性起因バグの修正

- 問題
  - マルチスレッド実行中(SSLやgzipスレッドの活動中)に強制終了割り込み(再起動含む)が起こると様々な異常が発生
- 症状
  - 無限ループ発生 →CPU100%
  - 排他制御破綻、デッドロック発生 →再起動失敗
  - ヒープ破壊、セグメント例外発生 →異常終了
- 再現性
  - 数百回、1万回、10万回の再起動に一度発生
  - 発生パターンがOSや版に依存
- 原因
  - クリティカルセクション中で割り込み処理発生
  - 同時アクセスによるデータ破壊
  - 解放されないセマフォ
- 対処
  - クリティカルセクションを割り込み禁止に
  - そもそもどこにCSが？ →ひたすら再現実験 →数十ヶ所発見 (所要1人月以上)
- 高性能MacProとマルチプラットフォーム開発用VMの大活躍

# 並列性起因バグの実例

- スレッドAがCGI用に環境変数配列を入れ替えている間に、スレッドBが環境変数にアクセスし、解放されたヒープ領域へのアクセスが起こり、不定の結果に...
- FreeBSD7でのみ高頻度で発生し、捕捉に成功

[元の版]

```
extern char *environ[];  
...  
feev(environ);  
environ = newenviron;
```

解放処理中に  
アクセスがあ  
ると...

[修正版]

```
extern char *environ[];  
char *oldenviron;  
...  
oldenviron = environ;  
environ = newenviron;  
feev(oldenviron);
```

# DeleGateの開発環境@2009

@amakubo.com

OCN/Flet'B

Asahi-net/Flet's B

Wakwak/Flet's ADSL

EM-ONEa/eMobile

Internet

@UMEZONO-SITE

2-1D 4F (2000W)

Linux RedHat7 ... DeleGate.ORG

iMac / PPCG4

PowerMac/PPCG5 ... VM host

2 x WindowsXP ... VM host

Windows2K

12 x VM guests (Linux, BSD, Solaris)

Zaurus, SigmarionIII, EM-ONE

2 x MacPro/i86 ... VM host

2 x WindowsXP ... VM host

40 x VM guests

3 x MacMini

2 x iMac

3 x Kuro-Box

DecAlpha, SparcU1



# マルチプラットフォーム開発環境

- Vmware上の仮想マシン約60[台]
  - Linux -- 2.6/2.4, 32bits/64bits
    - Ubuntu SUSE Fedora Vine Debian Mandriva Turbo CentOS RedHat
  - BSD系
    - FreeBSD NetBSD OpenBSD
  - Windows
    - XPsp1,sp2 2K NT Me
- 実機約20台
  - Mac OS X
  - DecAlpha OSF/1 ...
  - Sun Sparc U1 (故障中) Solaris5
  - Zaurus, KuroBox
  - EM-ONE, iPAQ, SigmarionIII, ...
  - WindowsXP,2K
    - Win32 Cygwin



# DeleGateのこれから

## これまで

- どのような有用な機能ができるか？の探求(どのように作るか、性能は二の次)
- テストは手作業、ユーザの利用によるバグ出し
- SOHO小規模利用／デスクトップには必要十分な性能

## これから

- どのような実装で高速化・小型化できるか？
- 部分的にも自動検証可能な実装は可能か？
- 組込～大規模利用に耐える小型化／高性能化

# 2009年度後期計画(1)

- 新方式による従来機能の実装
  - 上位プロキシへの中継
  - キャッシュ機能(仮想サーバとしてのキャッシュ)
- 新方式の適用範囲の拡大
  - SOCKS、DNS、FTP
- 新方式の性能向上
  - マルチコアを有効利用するマルチスレッド化
  - リゾルバ用の独立スレッドの導入
  - ハードウェア化に適した構造の考察

# 2009年度後期計画(2)

- 精密な性能測定
  - Apache, Squid 等との性能比較
  - 標準ベンチマークによる比較
  - ワーキングセットサイズと性能の相関の調査
  - OSとVMの性能差、相性の原因究明、最適OSの選定方法を考察
- 引き続き従来版(ver.9)の安定性向上
- AISTリバースプロキシ<[www.aist.go.jp](http://www.aist.go.jp)> をver.9に ...
  - ver.8固有の性能上の問題 (keep-alive, gzipスレッド等)

# 新たな配布・普及方法

- ソース配布
- バイナリ配布
- 各OS用パッケージ配布
  - Rpm/Linux, dmg/MacOSX, zip/Windows, ...
  - 「DeleGate専用設定OS」のISOイメージによる配布
- VMによる「DeleGate専用機」の配布(あるいは「販売」)
  - 設定不要で即使用可能
    - 必要なライブラリ・フィルタ、インストール済
    - DeleGateの全機能の標準的な設定済
    - セキュリティ設定済(Firewall等) 悪用防止設定済
    - 不要ソフトは全削除
    - 64MB程度にできると良いが
  - 環境を判断して自律的・自動的に設定
  - ライセンス条件に合わせて事前設定(改変不可)
  
  - BSD系はVmwareと相性が悪いのかも？
  - VMware + Ubuntu ?
    - VMイメージでの配布条件とか

# 高齢化社会への寄与

- 老人に優しいユーザインターフェイスの開発
  - Web等の応用プロトコル経由で提供されるUI
  - DeleGateのデータ加工フィルタ機能の応用
    - 表示内容の絞り込み、必要最少化
    - 複数ページのマージ／状態遷移の単純化
    - フォントや色調の選択
- 昔のソフトを使い続けたい中高年の支援
  - 古いソフトの動態保存と維持、提供
  - マルチプラットフォーム開発環境の活用

# 各種エビでんすの収集

- 多様な評価基準
  - 利用者数(登録ユーザ制の導入)
  - 有償ライセンス
  - 論文等からの引用
    - Google Scholar で [“www.delegate.org”](http://www.delegate.org)
  - 特許中での引用
  - ランキング／レーティングサイト
  - 解説サイト
  - 用語辞典サイト
  - Wiki

2009.10.14

**完**