

T2Kシンポジウムつくば

T2Kアプリケーション：  
システム運用技術研究の立場から

加藤和彦

筑波大学大学院 システム情報工学研究科

コンピュータサイエンス専攻

2008年4月7日

# 「材料」コスト激減

- ハードウェアコストの低下(1980年頃と比較)
  - ✓ マイクロプロセッサ性能は10,000倍以上  
(1MHz -> 3GHz, 64MB -> 4GB)
  - ✓ メモリ価格30,000分の一
  - ✓ エンドユーザ・ネットワークスピード10,000倍  
(1200bps -> 10Mbps)
- ソフトウェアコストの低下
  - ✓ ソフトウェアのコモディティ化
  - ✓ オープンソースソフトウェア財産の蓄積
  - ✓ 寡占化対抗としての「フリー化」
  - ✓ 無料サービスの台頭：広告ビジネスモデルの確立

# サーバを建てることは容易になってきたが. . .

- ハードウェア, ソフトウェアのcommodity化の進行により, 「材料」は安価に手に入る.
- 設定は容易になってきたが, 実は, 維持が難しい.
  - ✓ **セキュリティ**
  - ✓ **信頼性**
  - ✓ **成長性 (スケーラビリティ)**
- 気がついてみれば, あちこちにサーバが建っている.
  - ✓ (継続的) 管理が重荷に.

# JST CREST研究

## 次世代 広域分散基盤技術の開発

### OS&ミドルウェア

**仮想計算機, 分散性制御技術** 加藤和彦, 品川高廣, 新城靖, 板野肯三, (筑波大), 杉木章義 (JST), 大山恵弘 (電通大), 阿部洋丈, 廣津登志夫 (豊橋技科大), 須崎有康 (産総研)

### ネットワーク

**オーバーレイネットワーク** (東大 中尾彰宏)  
**仮想ネットワーク** (ソフトイーサ社 登大遊)

### データ インターオペラビリティ

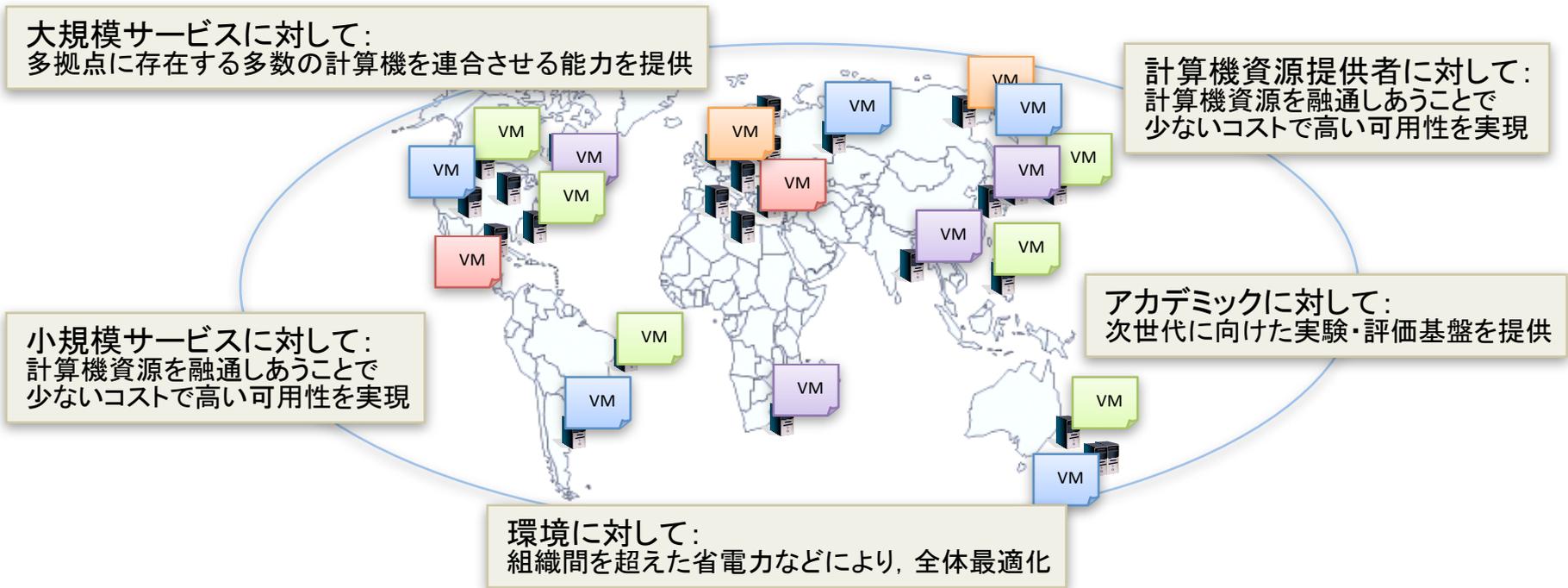
**拠点間データ処理技術** 北川博之, 天笠俊之, 森嶋厚行, 川島英之 (筑波大), 石川佳治 (名古屋大), 品川徳秀 (東京農工大), 渡邊陽介 (JST)

### プログラミング言語 検証

**データ交換のための検証技術** 井田哲雄, 南出靖彦 (筑波大)

# サステナブルサービス基盤

- 広域分散環境を想定した自律連合サービス基盤
  - Agile, loosely-coupled, collaborative なサービス基盤の実現
  - キーワード: 仮想化, 自律性と広域拡散

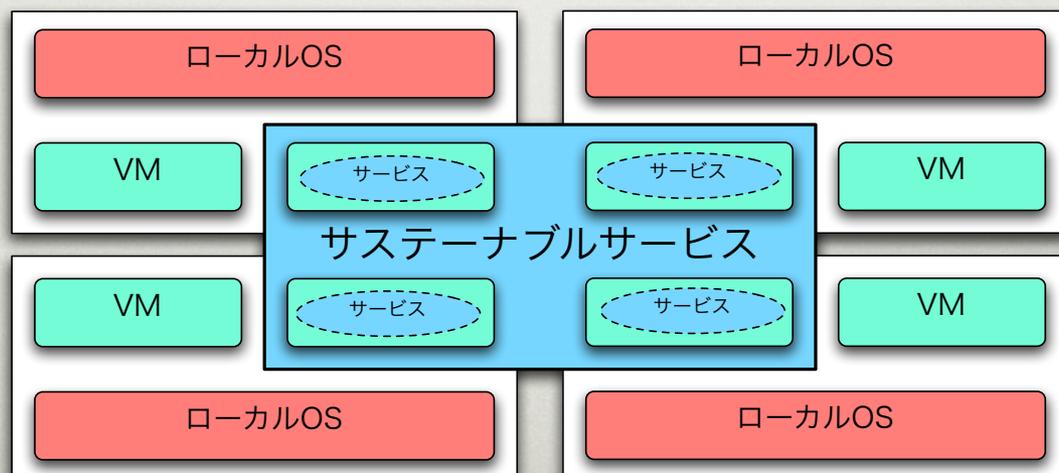


需要変動やさまざまな障害などの**予測不可能性**を自律と連合によって乗り越える

# サステナブルシステム

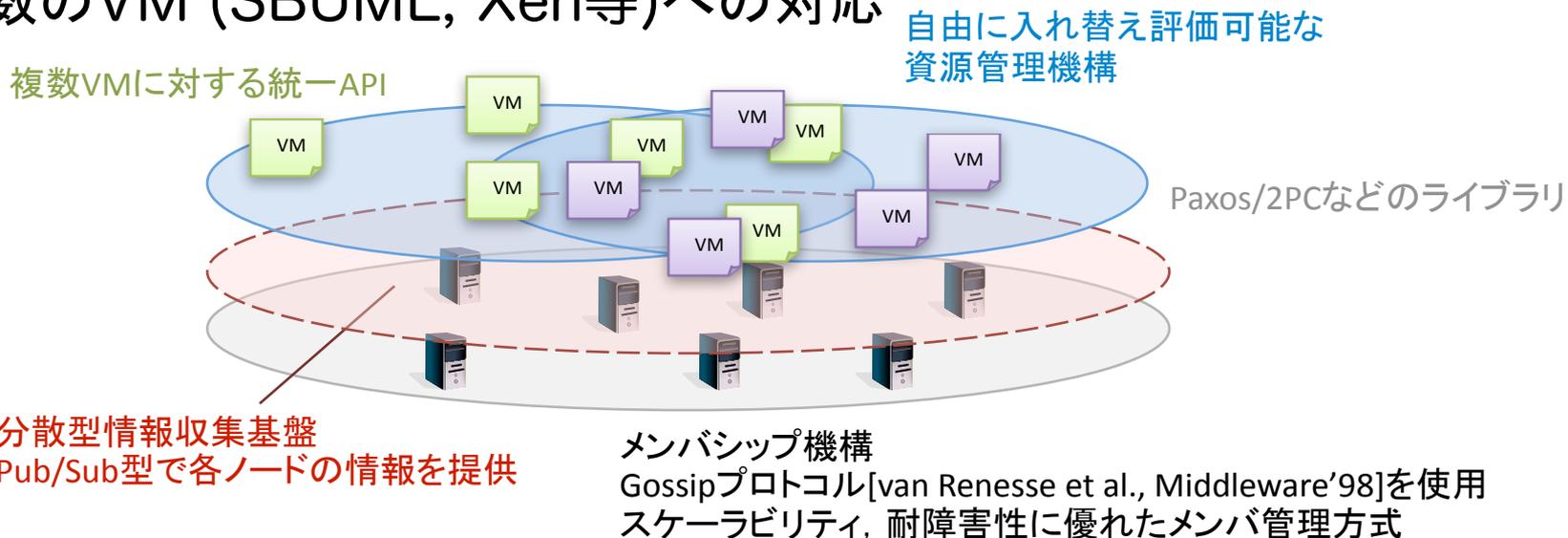
サステナブルシステム = 仮想機械 + 複製制御 + 一貫性制御

- **仮想機械**技術により，汎用化，配備（deployment）を容易化。
- **複製技術**と**一貫性制御**により，スケーラブル，リライアブルに。
  - 不具合（ハードウェア，ソフトウェア，ネットワーク等）を自律的に発見。



# サステナブル・コア基盤の機能

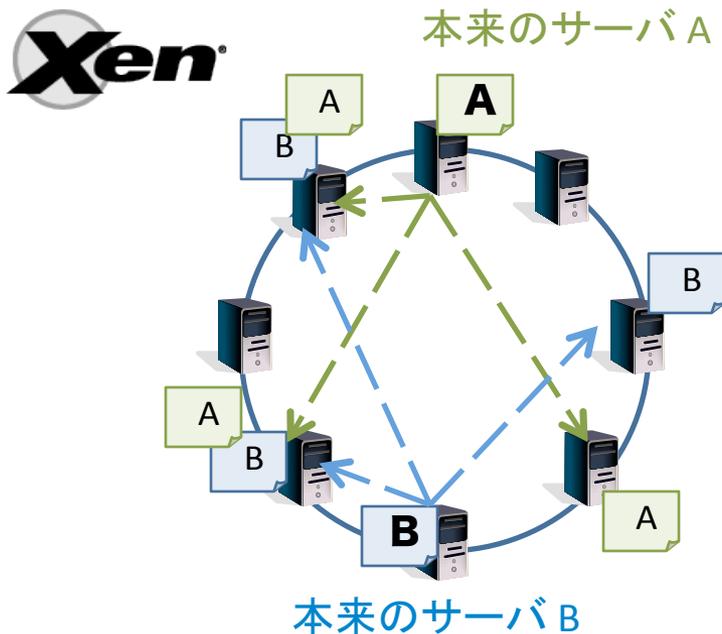
- 動的にサーバが参加・離脱する環境下でのサービス基盤
- 中央管理ノードが存在しない完全“非集中型”で構成
- 複数のVM (SBUML, Xen等)への対応



4月から17,000行記述し, さまざまなアルゴリズム・VMの評価を行う研究基盤を確立  
昨年10月のCRESTシンポジウムですでにデモを行う

# 応用例: サーバPeer-to-Peer

- 単一インスタンス・複数サービス
- 複数サーバ間で計算機資源を共有, 少ない資源で冗長化



サービス(VM)とホストを160bit SHAキーで管理

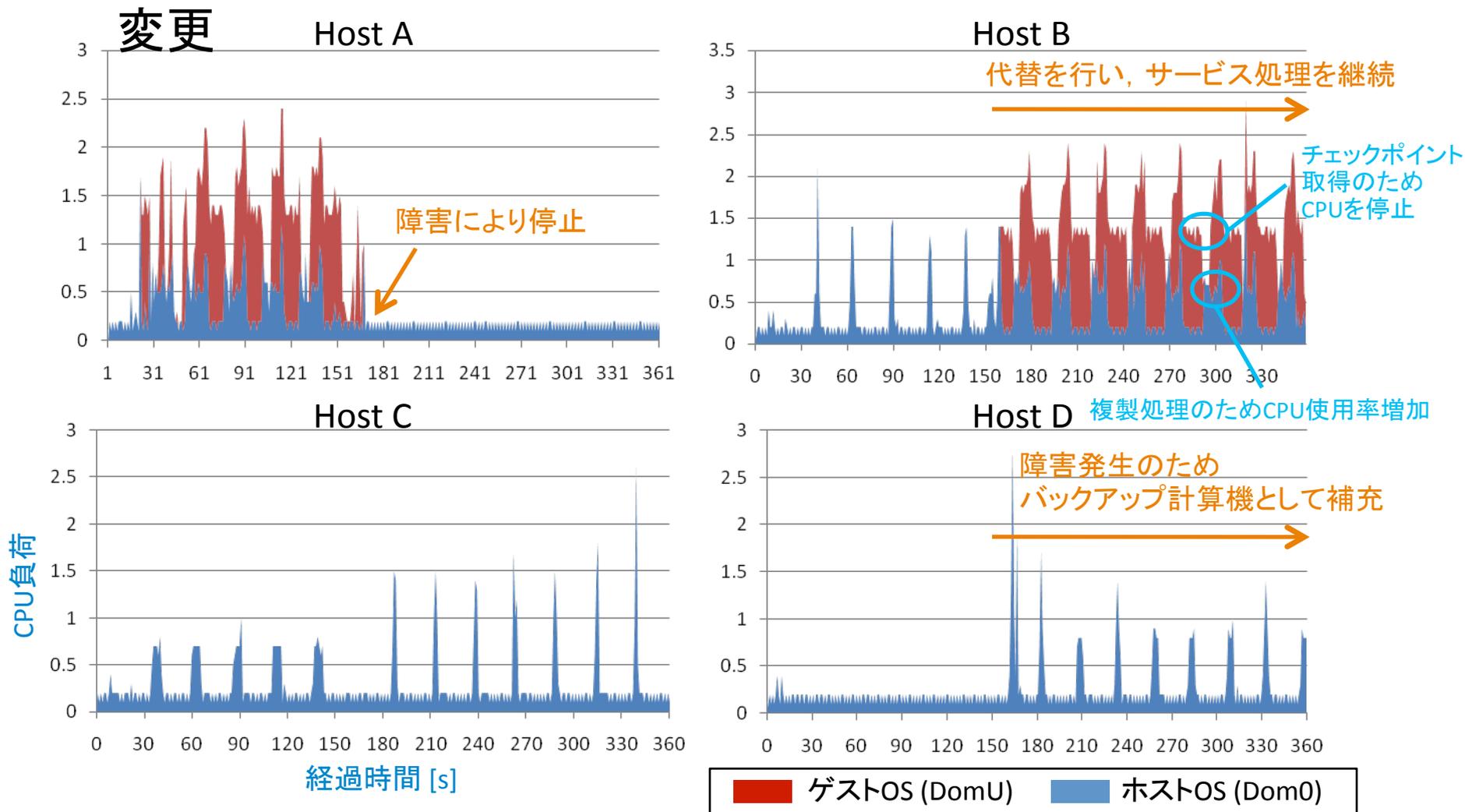
仮想的なリング型オーバーレイを構成

サービスのキーを起動時のホストと同一とすることで元のサーバに近いサーバでサービスを開始

障害発生時に他計算機が代替元の計算機が復帰すると、サービスも元のサーバへと戻る

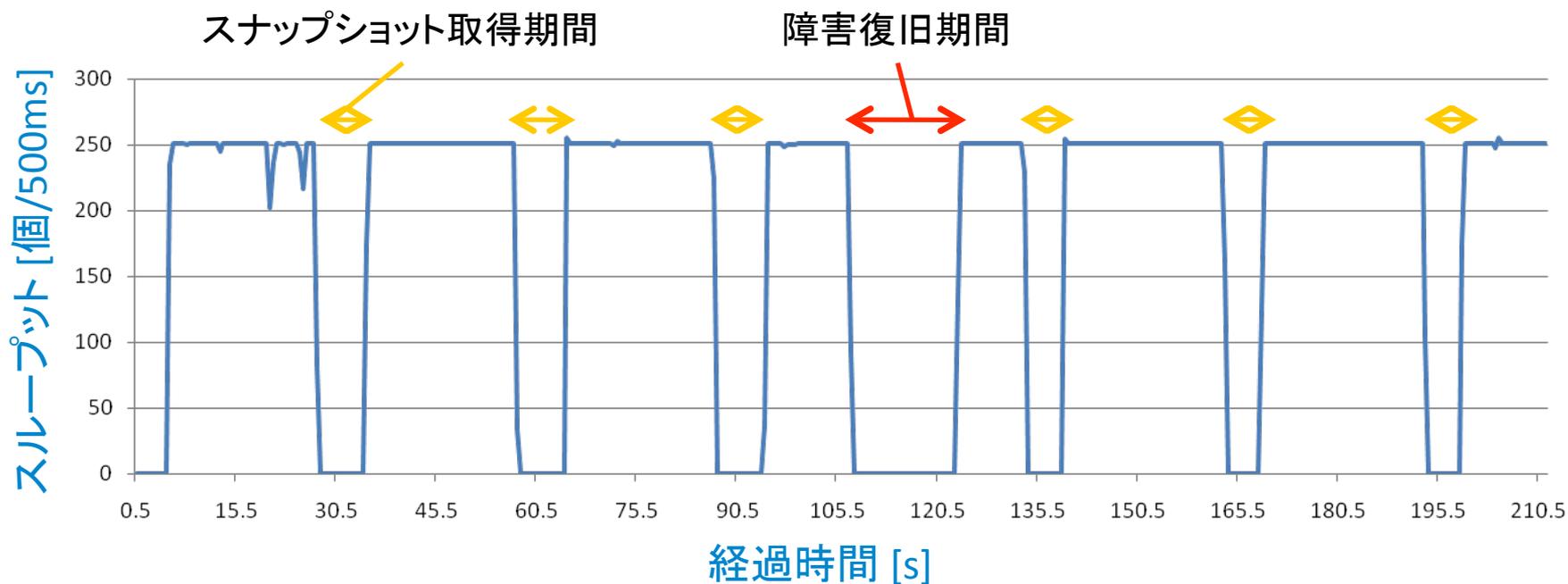
# 実験結果：各ホストのCPU使用率

- 障害を意図的に発生させ、ホストを {A,B,C} → {B,C,D} に



# 実験結果: クライアントからみたスループット

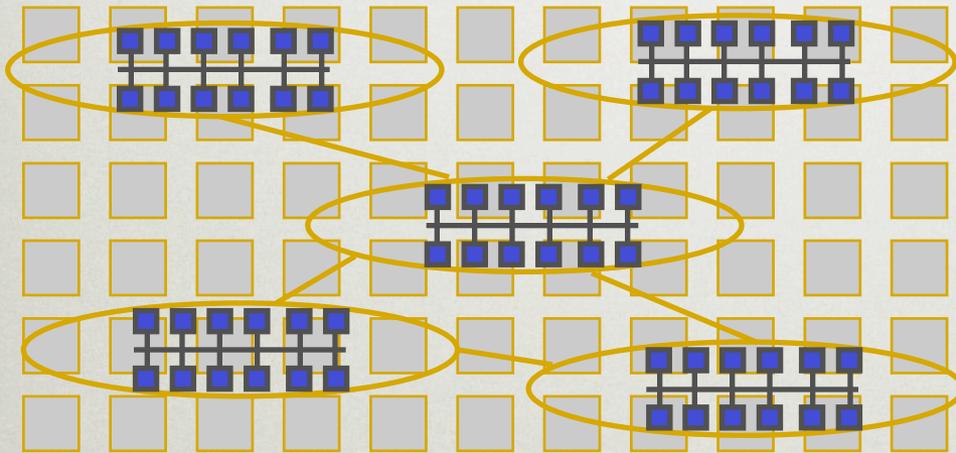
## ■ UDPのping-pongサーバで計測



スナップショット取得は約7-8秒  
障害からの復旧は約16秒

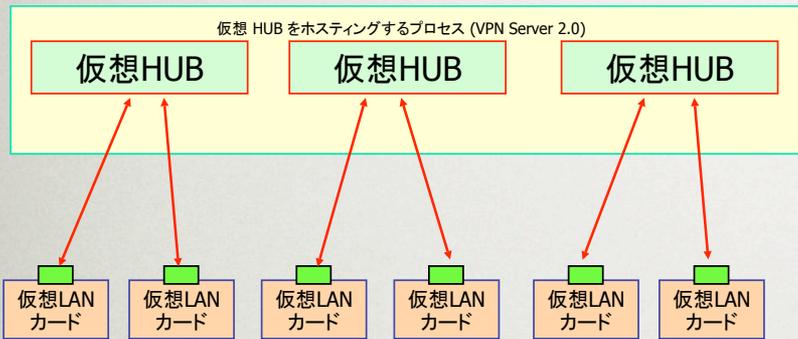
# 仮想インターネット

- 仮想計算環境+仮想ネットワーク+100CPUクラスタシステムにより1,000~10,000規模インターネット環境をエミュレート。

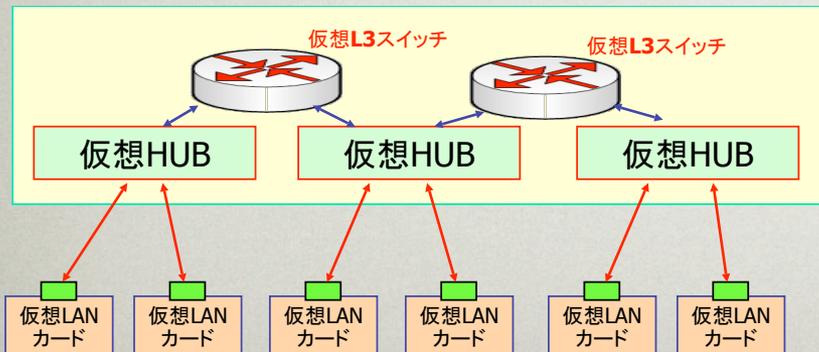


# 仮想ハブ／ルータ／インターネット

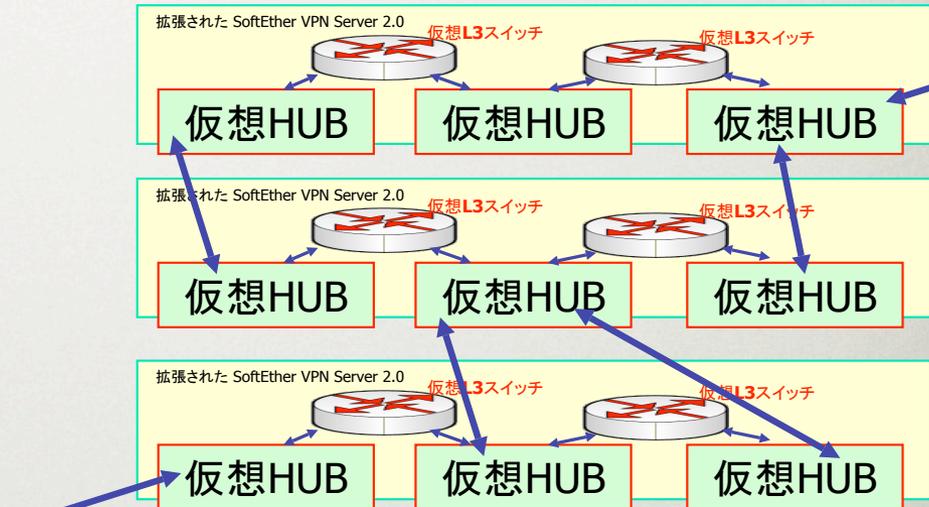
## 仮想ハブ



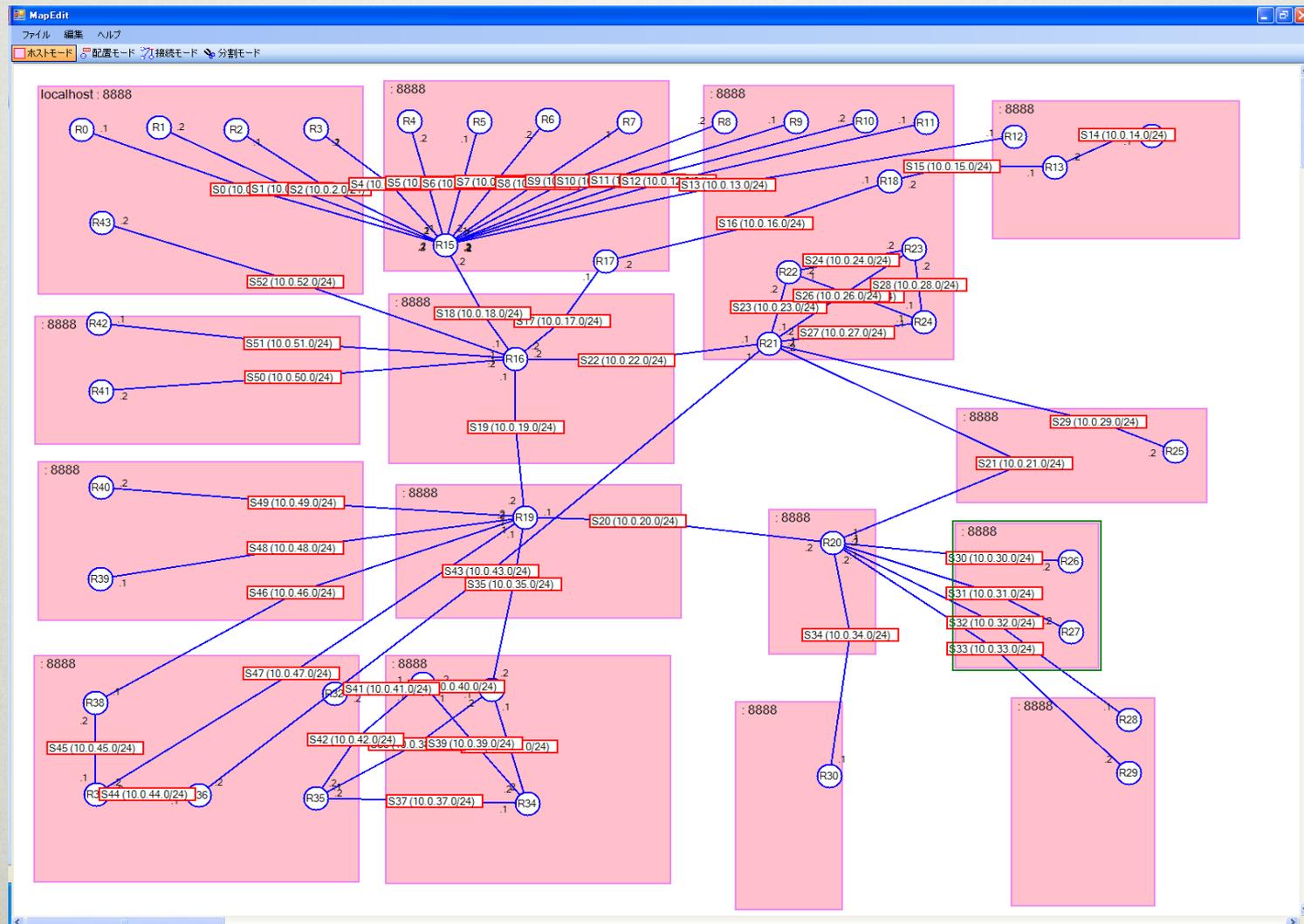
## 仮想ルータ



## 仮想インターネット

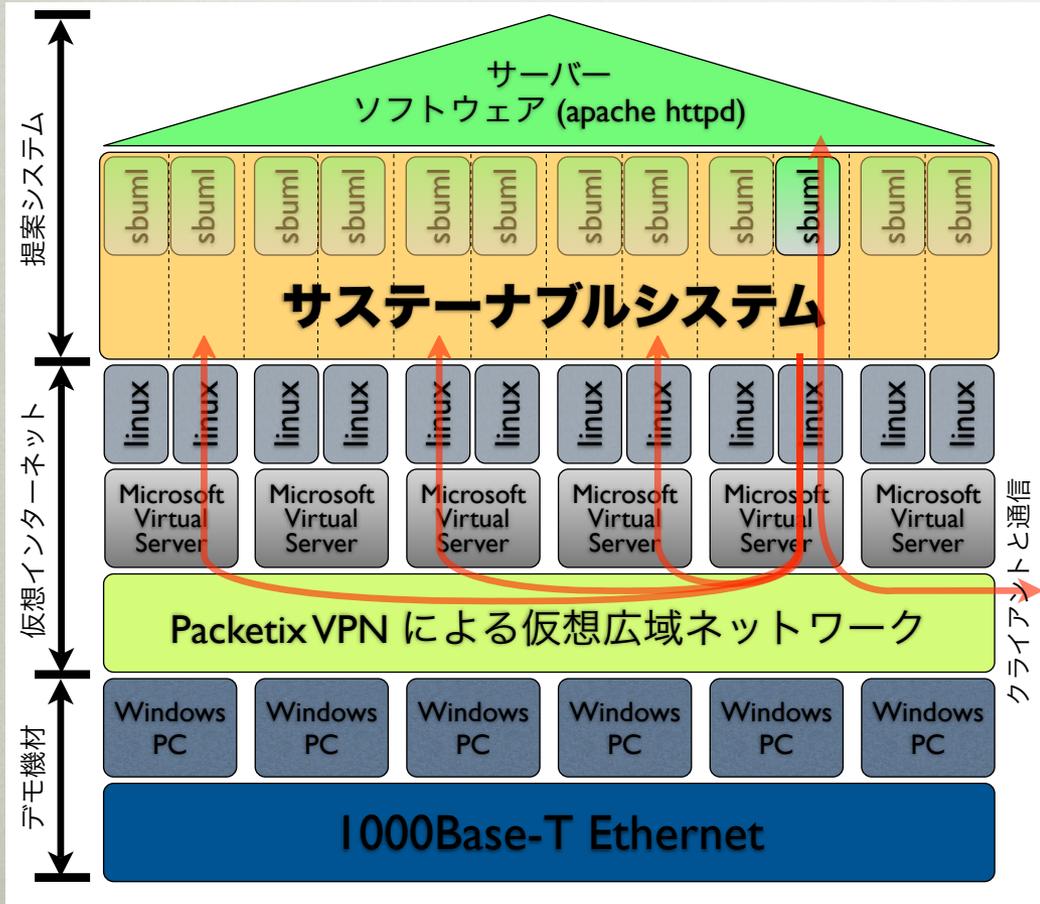


# ネットワークトポロジ編集ツール

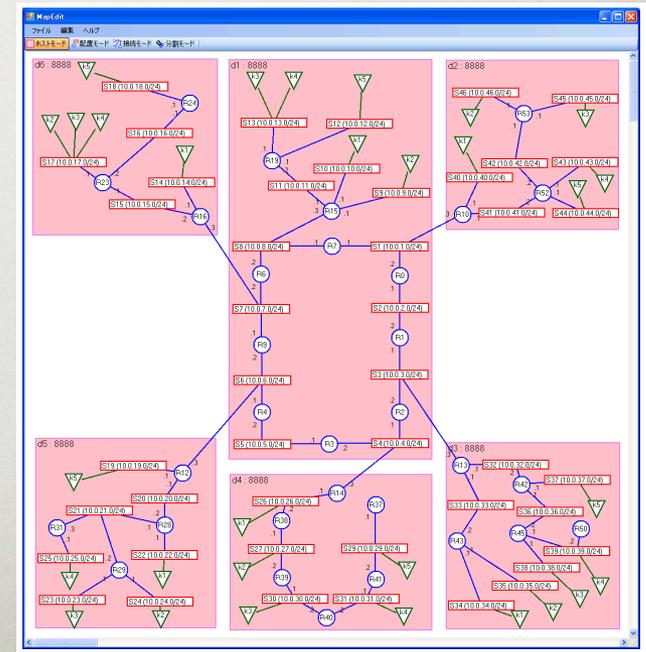


# 自律連合システムの実験環境

## システム構成



## 仮想インターネット



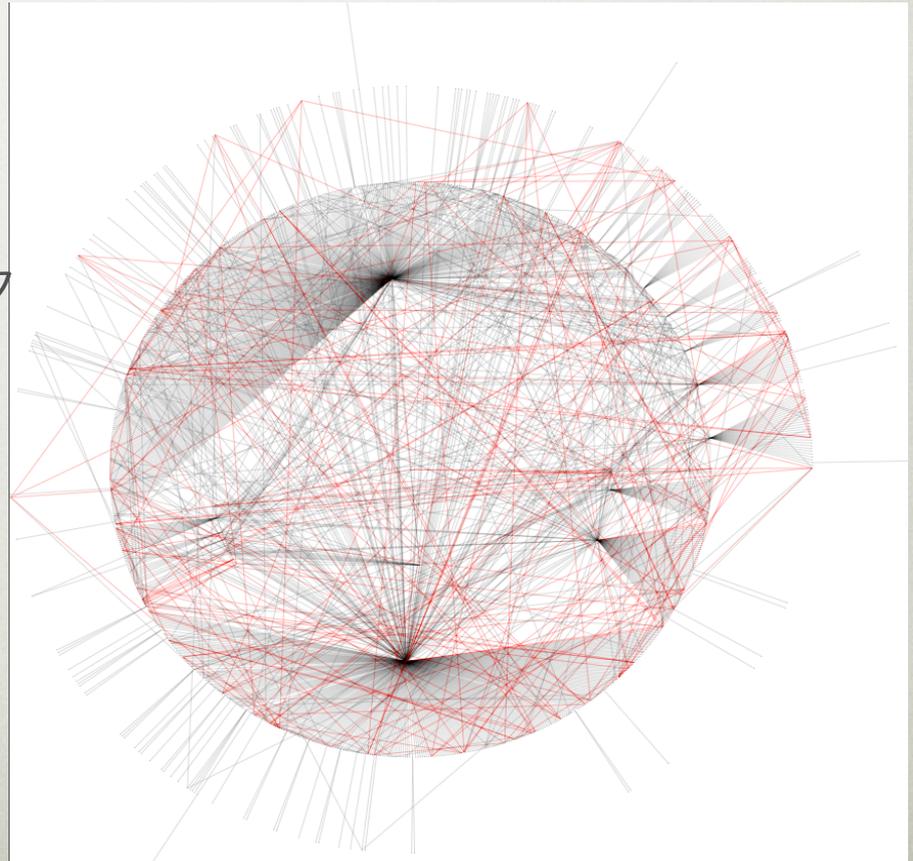
# インターネット シミュレーション

- ・ インターネットを模したネットワークトポロジーを計算により生成し，その上にオーバーレイネットワークを配置.
  - ✓ Generalized Linear Preference  
[Bu, INFOCOM02]
  - ✓ SSFNet ネットワーク  
シミュレータ

生成したトポロジーの一例

AS 数: 1000

Sustor ノード: 100



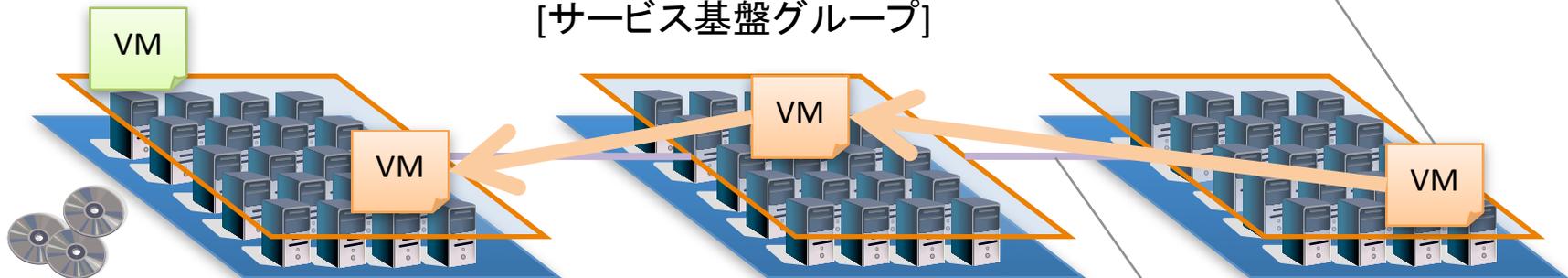
# 全体統合ビジョン

## 拠点間にもたがるスケーラビリティ&サステナビリティ

運用前の

VM内のサービスの静的検証

[言語と検証グループ]



サーバの自律連合基盤

[サービス基盤グループ]

データのストリーム処理

[データインターオペラビリティグループ]

Knoppixを利用したネットワークブートによる

サスティナブル基盤のインストール

[ネットワークOSグループ]

仮想ネットワークにより

拠点間・VM間を接続

[仮想ネットワークサブグループ]

広域ネットワークのサスティナブル化

[広域ネットワークグループ]

# 関連技術との比較

