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

2006－ＱＡＩ－18（6）
2006／1／20

[招待講演] PlanetLab and beyond

中尾 彰宏

東京大学大学院情報学環 〒113-0033 東京都文京区本郷7丁目3－1
E-mail: nakao@iii.u-tokyo.ac.jp

あらまし　広域分散システム構築のための地球規模のテストベッドであるPlanetLabの技術動向について紹介する.

キーワード　オーバレイネットワーク，広域分散システム，ユビキタスコンピューティング

PlanetLab and beyond

Akihiro Nakao

Graduate School of Interdisciplinary Information Studies
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033 JAPAN
E-mail: nakao@iii.u-tokyo.ac.jp

Abstract　An introduction to the technical challenges in PlanetLab, an open overlay network platform for developing, deploying, and accessing planetary-scale network services is presented.

Keyword　overlay network, distributed system, ubiquitous computing

PlanetLab

Akihiro NAKAO
Interfaculty Initiatives in Information Sciences,
Graduation School of the University of Tokyo

1

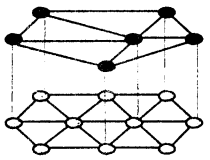
広域分散システム

- 地理的に広域に分散したシステム
  - "The Internet"
  - P2P VoIP (IP電話)
  - P2P ファイル共有
  - Robust Routing
  - Content Distribution Networks (CDN)
  - Scalable Network Embedded Storage
  - Scalable Event Propagation
- "Overlay Network"として実装

2

Overlay Networks

- インターネット自体を変更しないで新しいネットワークサービスやネットワークプロトコルを導入する手法
- "実ネットワーク"上に"仮想ネットワーク"を "overlay"
- 新しいネットワークサービスの研究に有効な方法
  - 広域分散アプリケーション
  - 次世代インターネット・アーキテクチャ (IPvN, GENI)



3

広域分散システム研究の背景

- 安価な資源
  - Network
  - PC's
- 計算機の矮小化に伴う膨大な情報量の処理
  - Ubiquitous computing
  - Sensor Network
- 情報発信の地理分散による障害回避と負荷分散
  - "Point of Presence"
- 効率的なデータ転送のための経路制御
- 次世代検索エンジン・新しい情報収集の仕組み
- 次世代インターネットアーキテクチャの提案
  - "Ossified(硬化)"したインターネット
  - 次世代インターネット・アーキテクチャ

4

PlanetLab

- A "Planetary-Scale" Overlay Network
- 広域分散システム構築のためのテストベッド
  - インターネット環境での評価を可能にする
  - 長時間持続の試験運用による、実際の配置運用へのシームレスなマイグレーション
    → "Dual Use" モデルが最大の特徴
- 従来のテストベッド
  - 実験室クラスターでのネットワークのエミュレーションが主流
  - 実際の配置運用へのマイグレーションが弱い

5

PlanetLab's Scale

- 現在30ヶ国280施設に分散する約600ノード
- 目標は1,000ノード

Related Work

- NetBed(Emulab)
- ABONE
- Grid Computing: e.g. Globus
- Internet2
- XBONE

7

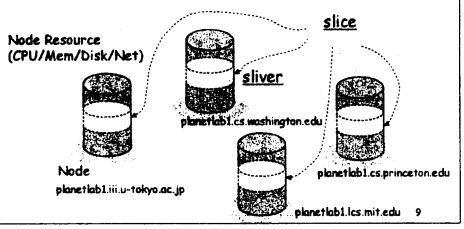
PlanetLab Consortium

- 少数の大学が管理するConsortiumで運営
  - Princeton University
  - University of Washington
  - Intel Research at Berkeley
- Consortium参加者によるノードの提供
  - 1,000ノード規模のテストベッドの占有は非合理
  - 参加者がリソースを供出
  - 持ち寄ったリソースを参加者が共有

8

PlanetLab Architecture

- Service：分散システム全体が提供する機能
- Slice：全てのノードのリソースを各サービスに割当てる単位
- Sliver：sliceをノード当たりに分割したの単位



9

Node Level Architecture

- Sliver = 仮想マシン(VM)
- Slice = 仮想マシンの集合 (a set of VM's)
- 各ノードは仮想マシンモニタ(VMM)を実装する
  - PlanetLabはLinux VServer + αをVMMに採用
  - 複数のsliceをサポートするのに十分なスケーラビリティ
  - Unix system callレベルでの仮想化
  - Chroot + Security Context

10

Sliceの実際

- 各Serviceは割当てられたSlice上で実行
- Serviceを実行するためにはVMにログイン

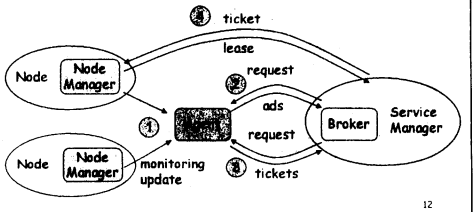
例) 東京大学でユビキタスの実験サービスを評価したい

・Slice名：utokyo_ubiq　が割り当てられる
・Nodeの一つplanetlab1.iii.u-tokyo.ac.jpのSliverである
　VMを利用して分散サービスを実行する
　　ssh utokyo_ubiq@planetlab1.iii.u-tokyo.ac.jp

11

Network Level Architecture

- Node Manager
- Agent
- Service Manager (+Broker)
- SSL + XML RPC



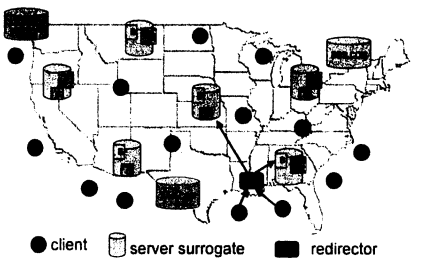
12

Applications

- CoDeeN
- Network Telescope
- OceanStore
- CoDNS
- OpenDHT
- PlanetSeer
- End-System Multicast
- And more... 450 slices running today !

15

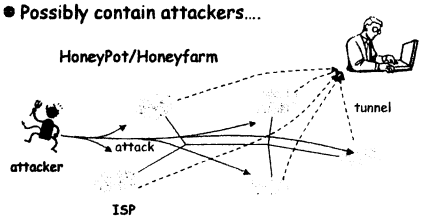
CoDeeN: Partial Replication CDN



14

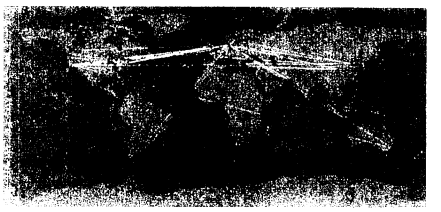
HoneyPot/HoneyFarm

- Observe attackers when they attack honeypots
- Identify signatures of attacks
- Possibly contain attackers....



15

March 2003 Nimda and Code Red



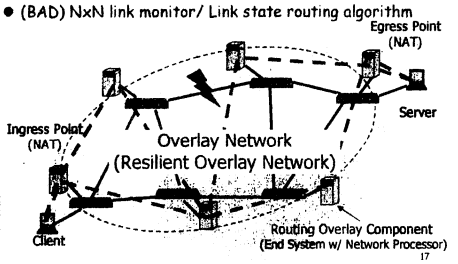
Intel: Netbait

- Detect and track Internet worms globally
- Potential: Spread Defenses Faster than the worms
- Stop a contagion in its tracks

16

Robust Routing Overlay

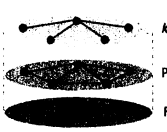
- Resilience (recover link failures quicker than BGP)
- Performance (optimize application specific metrics)
- (BAD) NxN link monitor/ Link state routing algorithm



17

PLUTO

- PLUTO-Mesh reduces overhead traffic
  - Reduction depends on underlying topology
  - Further overhead reduction would degrade routing quality
- Evaluate a sparser mesh on PLUTO-Mesh
  - *k*-MST ($\alpha \sim 2k/N$, $P = O(k)$, $R = O(k, N)$)
  - Smaller *k* gracefully degrades routing quality



k-MST — $\alpha \sim 2k/N$, $P = O(k)$, $R = O(k, N)$; smaller *k* gracefully degrades routing; PLUTO improves diversity (AS, router)

PLUTO-Mesh — $\alpha = 0.6$, $\Delta P/P \sim \Delta R/R \sim 0.4$; marginal negative impact on routing

Fully Connected Mesh — $\alpha = 1$

18

OpenDHT

- Distributed Hash Table (DHT)
  - Load Balance
    - O(K/N), K: # of keys, N: # of nodes
  - Scalability
    - Each node maintains info about only O(logN) other nodes
    - Look-up requires O(logN) messages
  - Availability (Failover)
    - Node join/leave affects O(1/N) fractions of the keys
- Why do not we make an open source DHT ?

19

Future of PlanetLab

- Federation
  - PLC (PlanetLab Central)
  - Regional : PLE (Europe), PLA(Asia), PLJ(Japan)
- Private PlanetLab
  - OneLab (INRIA)
  - MUON (Utokyo)
- Cluster
  - Rocks
- VMM
  - XEN
  - VMWare

20

PlanetLab Japan

- PlanetLab Japanの発足
  - 日本の様々な分野の研究者のためのツール
  - Regional PlanetLab ?
- Private PLs in Japan
  - 将来的に日本独自の構想を取り入れる
  - Ubiquitous Computing
  - High Bandwidth
  - Wireless/Mobility
- Several possible scopes
  - The public PlanetLab
  - Private PL
  - Consortium-wide PL
  - Reservation-based (GRID)

21

Conclusion

- PlanetLab is "Self-Scaling"
  - 参加者が多い程より有用なシステムとなる
- "Game Change"
  - 広域分散システムをデザインし評価する環境
  - Even a grad student can change the world!

連絡先:

http://www.planet-lab-jp.org
nakao@iii.u-tokyo.ac.jp

22