国立天文台における
Secure Firewall の利用

牧野淳一郎
国立天文台
理論研究部 教授
天文シミュレーションプロジェクト プロジェクト長

Secure Japan Forum 2008 2008/9/4

概要

- 国立天文台の紹介
- 天文シミュレーションプロジェクトの紹介
  - GRAPE(-DR) の話も少しだけ
  - 他の計算機の紹介
- アプリケーション例
  - ダークマター構造形成シミュレーション
  - 銀河形成シミュレーション
- ネットワークとセキュリティ
  - システム概要
  - 歴史、、、
  - 設計方式と実装
  - 問題と対応
  - 今後の課題

国立天文台の紹介

- 日本の天文学のナショナルセンター
- 地上観測施設
  - すばる望遠鏡
  - 野辺山宇宙電波観測所
  - その他、岡山観測所等
- JAXA と共同で宇宙からの観測も
- 暦、理科年表等の歴史的な機能
- 理論研究・シミュレーション研究による天文学

国立天文台の研究施設

国立天文台の研究・観測施設は日本各地にとどまらず、すばる望遠鏡や建設中のALMA（アルマ）のように海外にも進出しています。天文学の観測では、可視光、赤外線、電波、重力波などの観測手段と、太陽とそれ以外の宇宙などの観測対象に応じて、最適の観測条件と環境とが必要とされるからです。

野辺山エリア

野辺山宇宙電波観測所
Nobeyama Radio Observatory

日本の電波天文学を世界のトップレベルに押し上げた観測施設です。写真の45m電波望遠鏡（右）は、ミリ波で世界最大の望遠鏡で、新たな星間分子の発見や原始惑星系の回転ガス円盤の発見など、数々の画期的な成果を挙げています。

野辺山太陽電波観測所
Nobeyama Solar Radio Observatory

太陽面爆発を高精度で観測する干渉計システム「電波ヘリオグラフ」（写真下）で、太陽活動のモニターを行っています。

乗鞍岳エリア

太陽観測所・乗鞍コロナ観測所
Solar Observatory:Norikura Solar Observatory

北アルプス・乗鞍山系摩利支天岳（海抜2876m）の頂上に位置する太陽観測所です。太陽物理現象の精密な観測を行うために3台のコロナグラフが設置されています。

岡山エリア

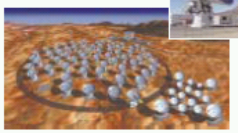
岡山天体物理観測所
Okayama Astrophysical Observatory

国内最大級の口径188cmの反射望遠鏡を中心に、銀河・恒星・太陽系天体などの光学赤外線観測研究の国内の拠点となっています。また、赤外線分光装置や赤外線超広視野カメラなど、宇宙を見る新しい目を次々と開発しています。

チリ・エリア

ALMA（アタカマ大型ミリ波サブミリ波干渉計）
ALMA (Atacama Large Millimeter/submillimeter Array) Project

ALMA（アルマ）は、日米欧が共同でチリの標高5000mの高原に建設中の巨大な電波望遠鏡群（イラスト）で、国立天文台が現在総力を挙げて取り組む大型プロジェクトです。



水沢エリア

水沢観測所
Mizusawa Astrogeodynamic Observatory

旧緯度観測所として長い歴史をもつ施設です。位置天文学・測地学の研究が進んで、日本の標準時を決める天文保時室などがあります。

江刺地球潮汐観測施設

レーザー光線を利用して地面の伸縮の変化を測るレーザー歪計です。潮汐による地球の微妙な変形をモニターします。

VERA観測所・水沢観測局
VERA Observatory：Mizusawa station

銀河系の3次元地図を作成するVERA観測局のひとつです。

三鷹エリア（本部）

三鷹キャンパス

三鷹キャンパスは、国立天文台の本部が置かれ、さまざまなプロジェクト、センター、研究部、事務部が集まっています。

VERA・4局
（水沢局を含む）

銀河系の3次元地図を作成します。

VERA観測所・入来観測局
VERA Observatory
：Iriki station

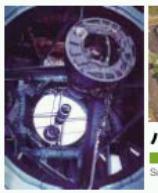
VERA観測所・石垣島観測局
VERA Observatory
：Ishigakijima station

VERA観測所・小笠原観測局
VERA Observatory
：Ogasawara station

銀河系の3次元地図を作成するVERA観測局のひとつです。

ハワイ・エリア

ハワイ観測所
Subaru Telescope



すばる望遠鏡

ハワイ島のマウナケア山頂（標高4200m）に設置された口径8.2mの世界最大級の可視・赤外線望遠鏡です。平成12年度から本格的な観測を始め、現在、世界最先端の研究成果を挙げつづけています。

ヒロ・オフィス（写真右上）

ハワイ島ヒロ市にあるハワイ観測所の本部です。「すばる望遠鏡」による観測研究の拠点となっています。

アメリカ合衆国
ハワイ州ハワイ島

すばる望遠鏡

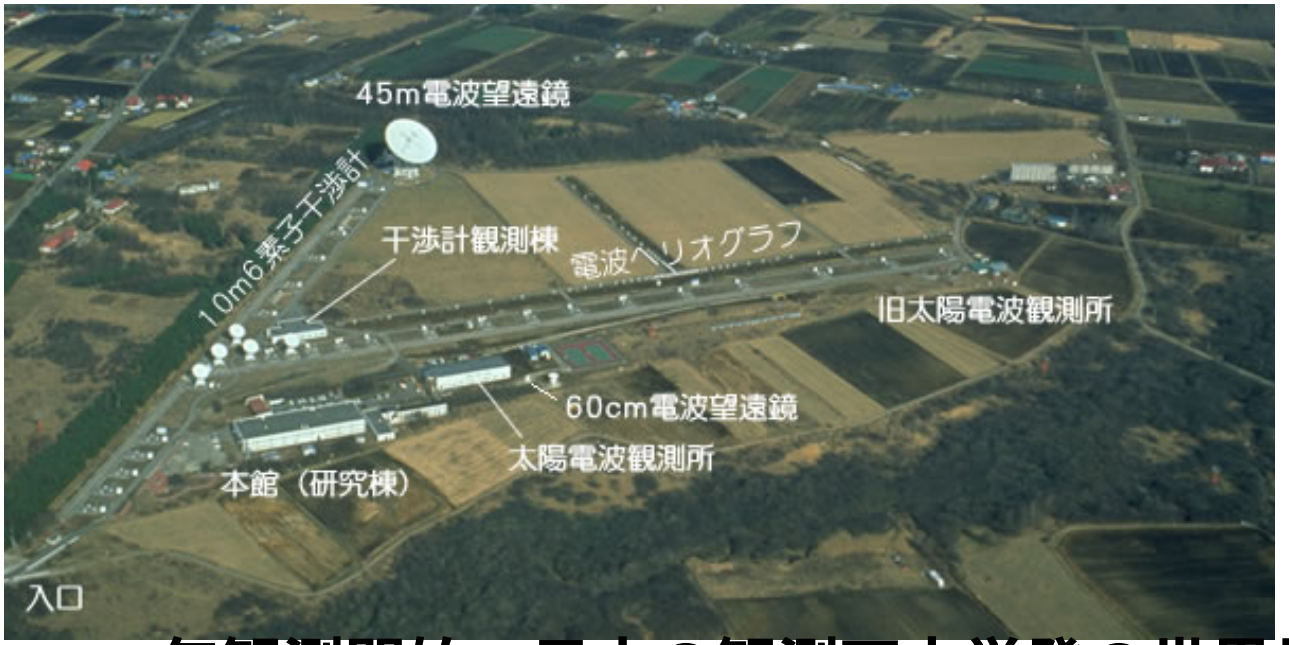
ハワイ・マウナケア山頂

主鏡直径 8.2m，世界最大級

大視野の主焦点カメラ (30分角)、ハッブル望遠鏡の100倍の視野

現在のところ、最遠方の銀河 (QSO) の10個のうち9個を発見

野辺山宇宙電波観測所



1982年観測開始。日本の観測天文学発の世界最先端装置
後継:日米欧共同プロジェクト ALMA 望遠鏡
(2009観測開始？)

その他主要プロジェクト

- ひので衛星
- 重力波観測装置 TAMA300
- VERA VLBI(超長基線電波干渉計) による銀河の構造・運動の観測
- VSOP-2 衛星による VLBI

天文シミュレーションプロジェクト

2 つの役割:

- 国立天文台の中の、理論・シミュレーション天文学研究グループ
- 国内外の天文学研究者のための計算機センター

歴史

1965年　人工衛星国内計算施設 発足

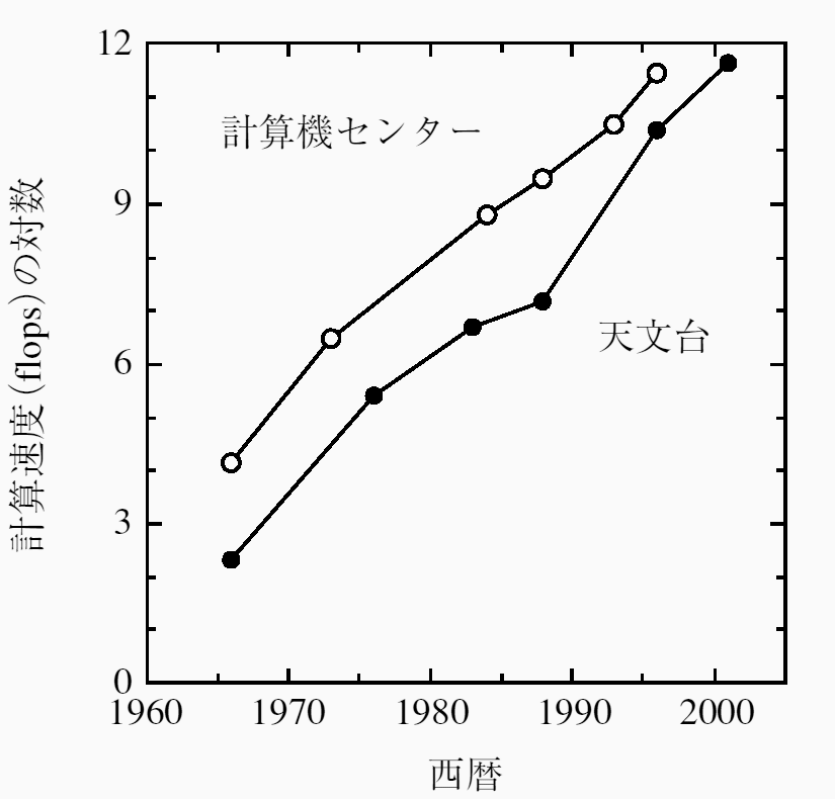
1988年　天文学データ解析計算センター 発足

同年　理論天文学研究系 発足

2006年　天文学データ解析計算センターを天文シミュレーションプロジェクトと天文データセンターに分離

「天文学理論・シミュレーションのための計算センター」として世界的にもユニークな位置を占める。

計算速度の進化



計算速度の対数

$(12 = 1\ \mathrm{Tflops})$

白◯: 東大センター

黒●: 国立天文台

大体東大の $1/10$ — $1/100$

OKITAC, UNIVAC **の他は**

富士通

2001-2007 のシステム

- VPP5000/60 600 Gflops
- GRAPE-5 (+GRAPE-6 (+GRAPE-7)) $>$ 10 Tflops
- WS/PC Cluster (Opteron 250 20CPUs, 2004) 96Gflops

VPP5000/60

60 ノード、600Gflops, 960GB メモリ, 12 TB ディスク, 60TB テープ

並列環境 t: MPI, VPP-Fortran

Up to 48 nodes/job

GRAPE hardwares

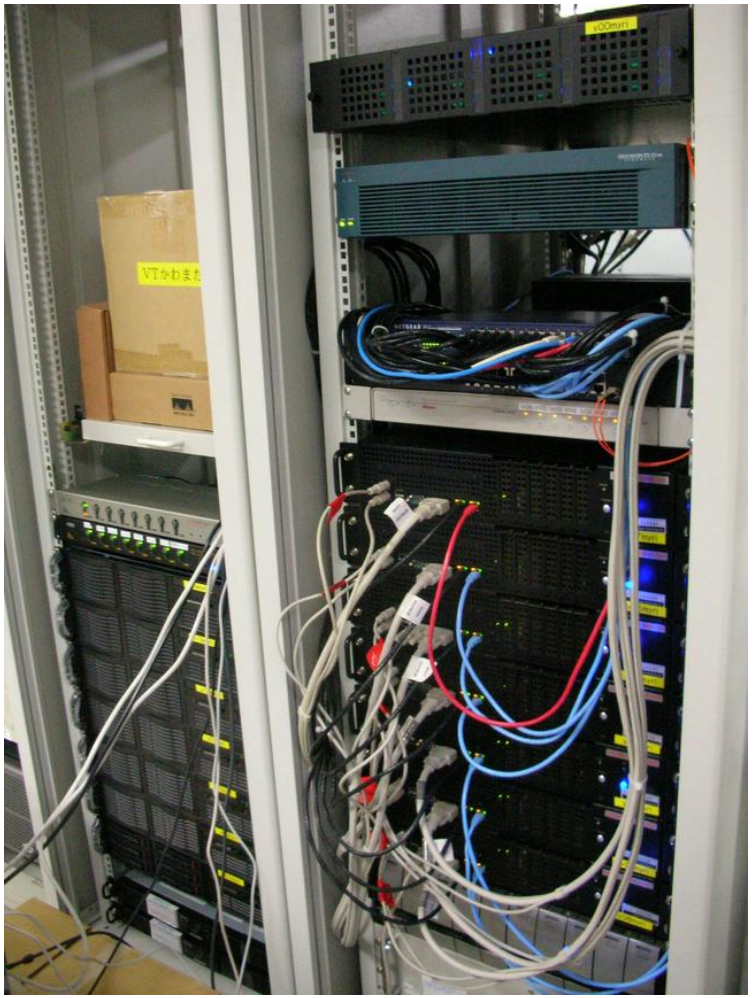
GRAPE-5 16 nodes (640Gflops peak)

GRAPE-6 8 nodes (8Tflops peak)

GRAPE-7 16 nodes (10Tflops peak)

No parallel job queue (yet)

PC Cluster



10-node, Dual-Opteron cluster (2004∼)

9.6Gflops/node

単一CPU での長時間計算に利用。

古典的「計算機センター」機能

結構利用者は多い

現行システム

- **ベクトル部分:** NEC SX-9 16CPU+4CPU
- **スカラー部分:** Cray XT4 9キャビネット (812演算ノード、28.6TF)
- GRAPE-6 +GRAPE-7 > 20 Tflops
- PC Cluster 920Gflops
- GRAPE-DR > 100 Tflops (2008/11)

Cray XT4

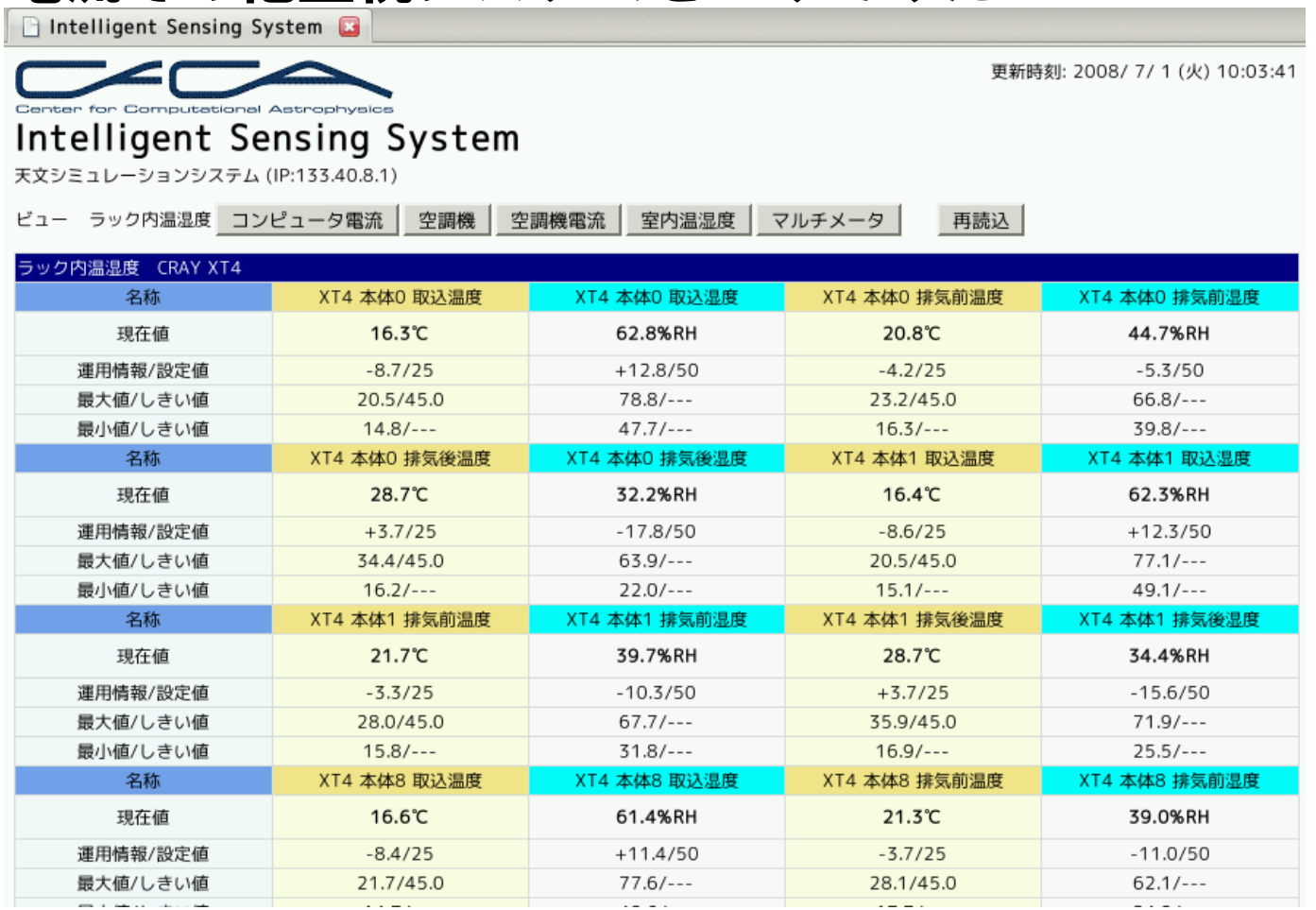
フロントパネル CG とデータ提供者

Cray になった理由

- 価格
- 消費電力
- 数千コアまでで実際に性能がでていること

空調

電流その他監視システムをつけてみた



ラック内温湿度 CRAY XT4

名称	XT4 本体0 取込温度	XT4 本体0 取込湿度	XT4 本体0 排気前温度	XT4 本体0 排気前湿度
現在値	16.3℃	62.8%RH	20.8℃	44.7%RH
運用情報/設定値	-8.7/25	+12.8/50	-4.2/25	-5.3/50
最大値/しきい値	20.5/45.0	78.8/---	23.2/45.0	66.8/---
最小値/しきい値	14.8/---	47.7/---	16.3/---	39.8/---
名称	XT4 本体0 排気後温度	XT4 本体0 排気後湿度	XT4 本体1 取込温度	XT4 本体1 取込湿度
現在値	28.7℃	32.2%RH	16.4℃	62.3%RH
運用情報/設定値	+3.7/25	-17.8/50	-8.6/25	+12.3/50
最大値/しきい値	34.4/45.0	63.9/---	20.5/45.0	77.1/---
最小値/しきい値	16.2/---	22.0/---	15.1/---	49.1/---
名称	XT4 本体1 排気前温度	XT4 本体1 排気前湿度	XT4 本体1 排気後温度	XT4 本体1 排気後湿度
現在値	21.7℃	39.7%RH	28.7℃	34.4%RH
運用情報/設定値	-3.3/25	-10.3/50	+3.7/25	-15.6/50
最大値/しきい値	28.0/45.0	67.7/---	35.9/45.0	71.9/---
最小値/しきい値	15.8/---	31.8/---	16.9/---	25.5/---
名称	XT4 本体8 取込温度	XT4 本体8 取込湿度	XT4 本体8 排気前温度	XT4 本体8 排気前湿度
現在値	16.6℃	61.4%RH	21.3℃	39.0%RH
運用情報/設定値	-8.4/25	+11.4/50	-3.7/25	-11.0/50
最大値/しきい値	21.7/45.0	77.6/---	28.1/45.0	62.1/---

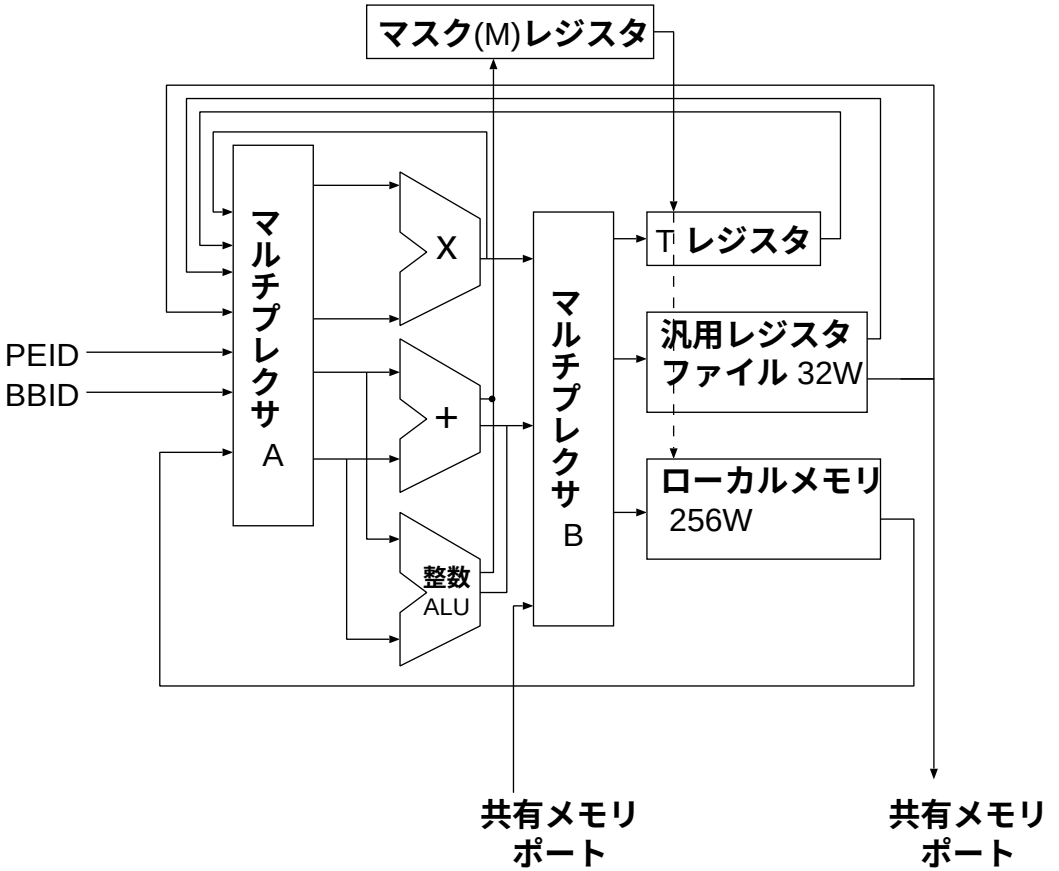
運用の感触

- 想定していたよりもずっと安定している
- 全体が落ちたのは 1 度だけ (ソフトウェア障害、対応済)
- 6 月末に CPU 交換 (B2→B3)
- そのまえからはほぼ 100% 稼働
- ユーザーの評判は大変良い
  - PC クラスタに比べてコア数ずっと高いところまで性能でる
  - I/O が速い

GRAPE-DR

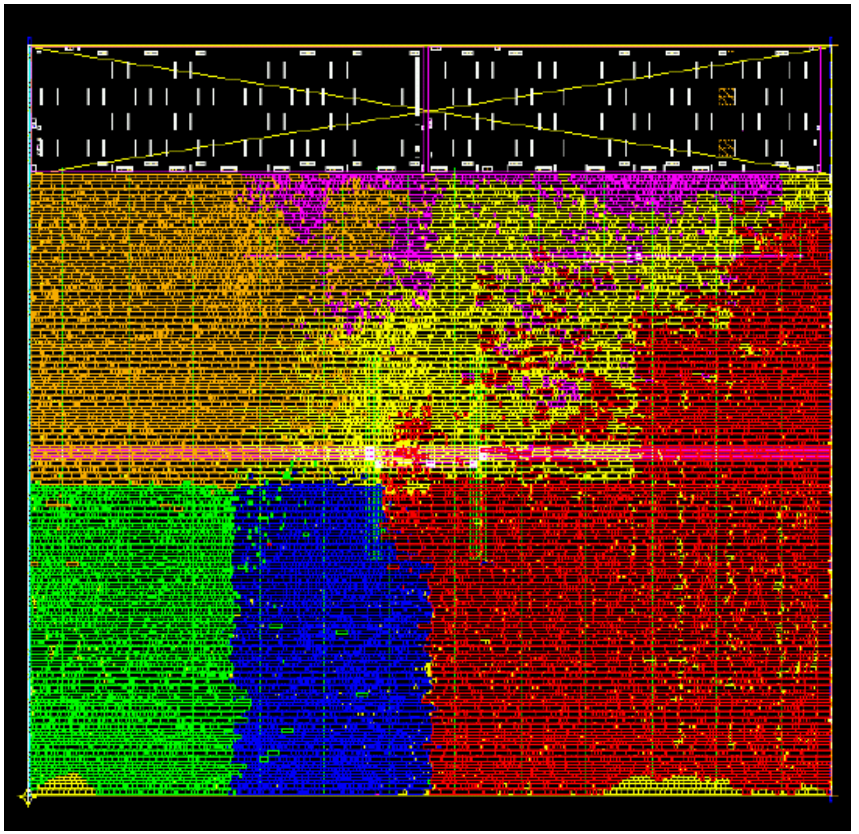
- SIMD 超並列プロセッサ
- 従来の重力専用計算機と違ってプログラム可能
- 今年度末完成、目標は 1Pflops (ちょっと厳しい)

PE の構造



- 浮動小数点演算器
- 整数演算器
- レジスタ
- メモリ 256 語 (K とか M ではない。)

PE レイアウト



0.7mm by 0.7mm

Black: Local Memory

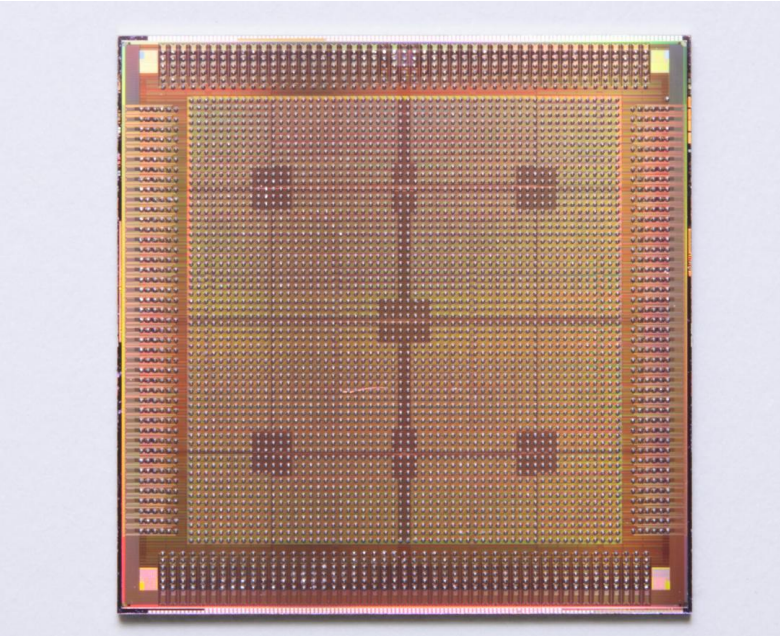
Red: Reg. File

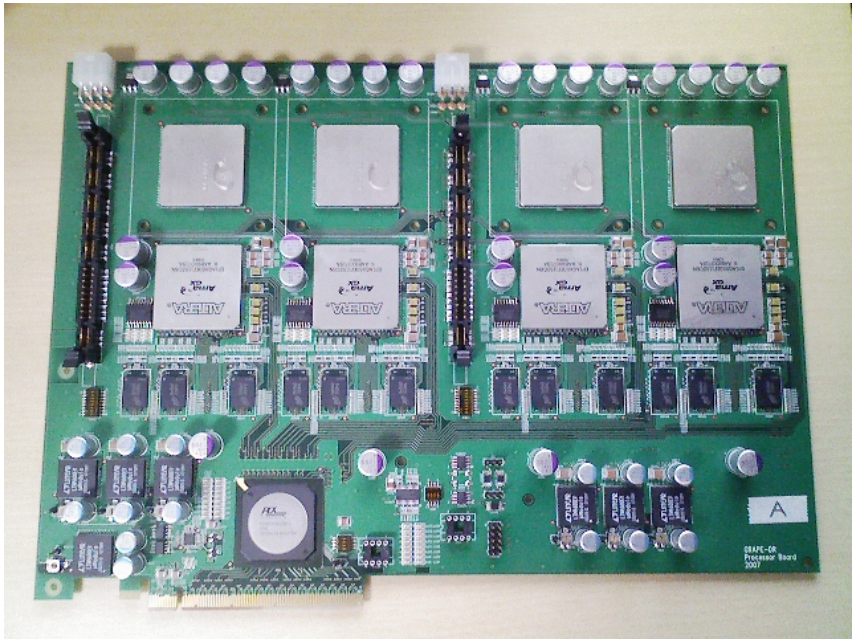
Orange: FMUL

Green: FADD

Blue: IALU

プロセッサチップとボード

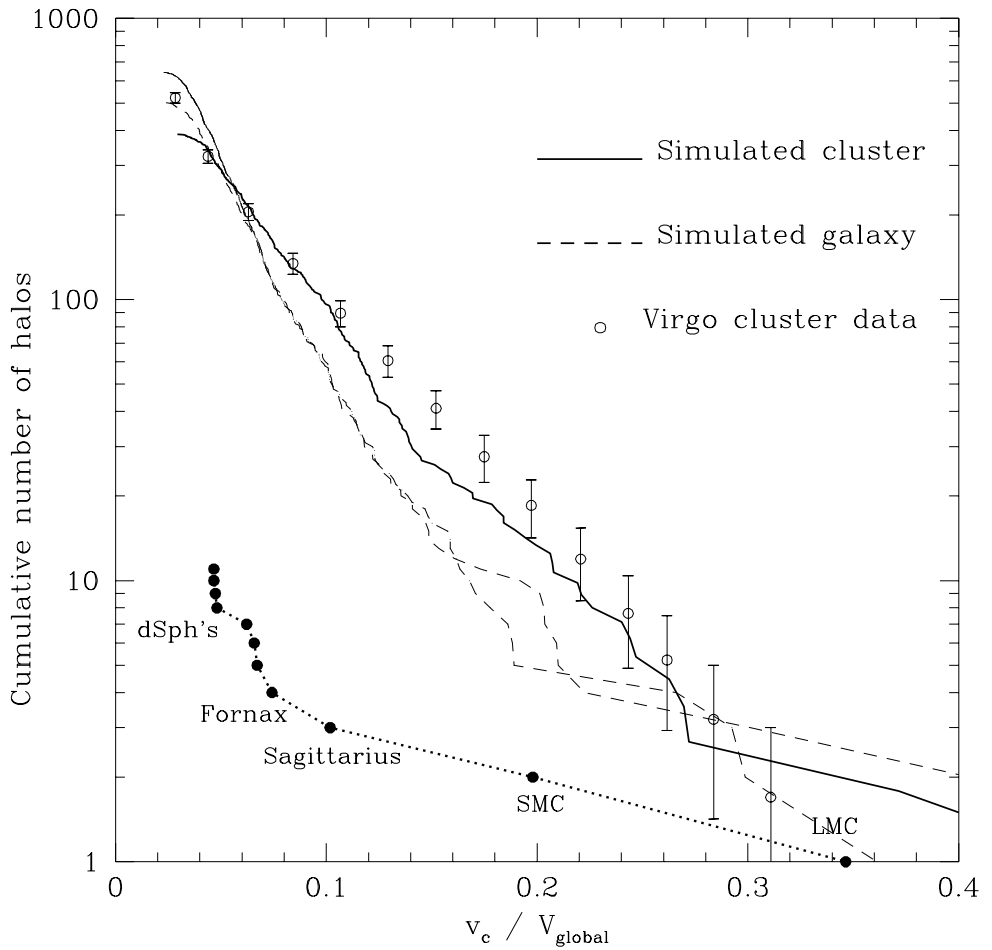




- PCI-Express カード (16 レーン、通信速度 2GB/s)
- 4 GRAPE-DR チップ
- 理論ピークスピード 1TP(DP), 2TF(SP)

アプリケーション例:大規模構造形成

考えた問題



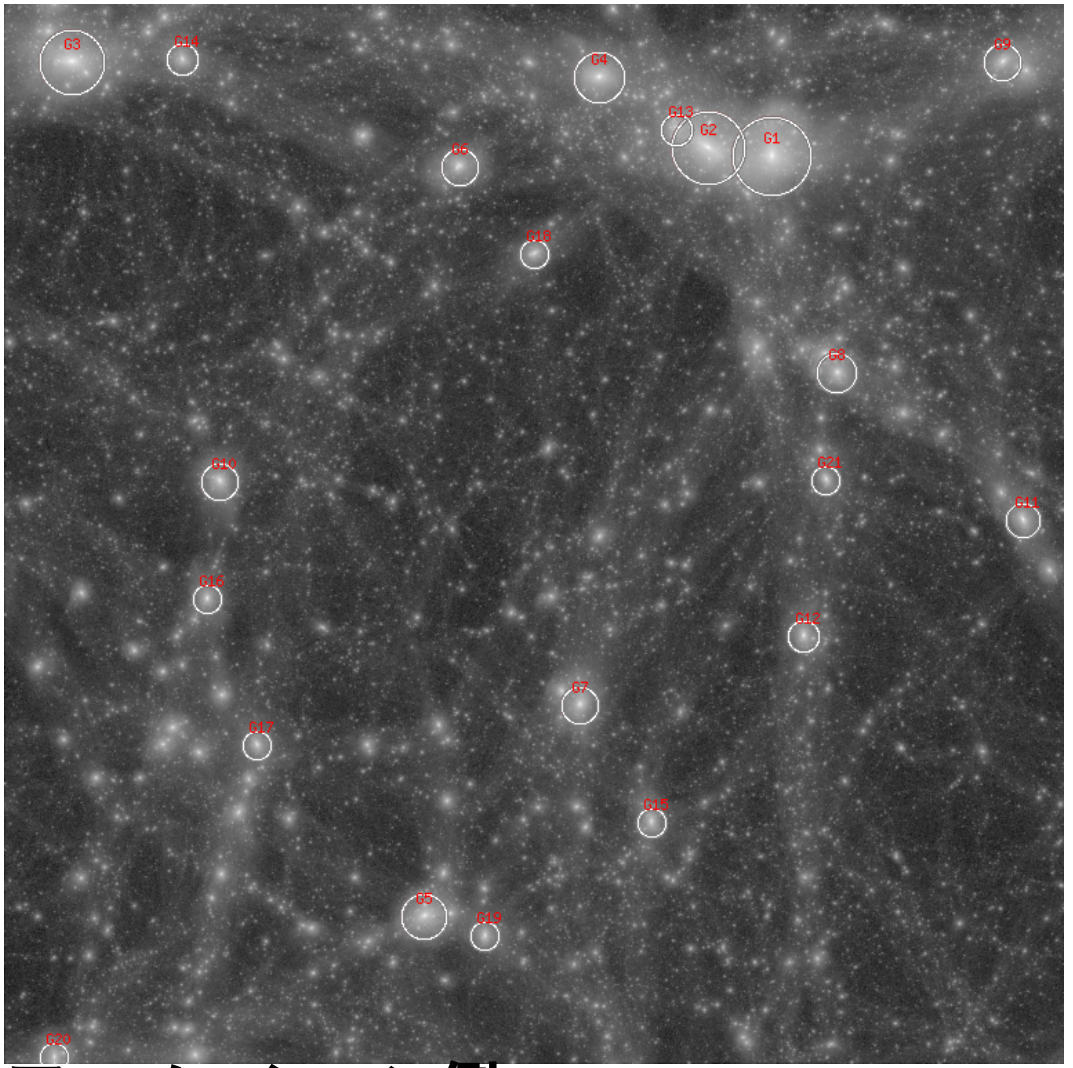
Moore et al 1999

- 銀河サイズの暗黒物質ハローには、小さいハローができすぎる。
- そんな多数の矮小銀河は見つかってない
- 何故？

我々の計算

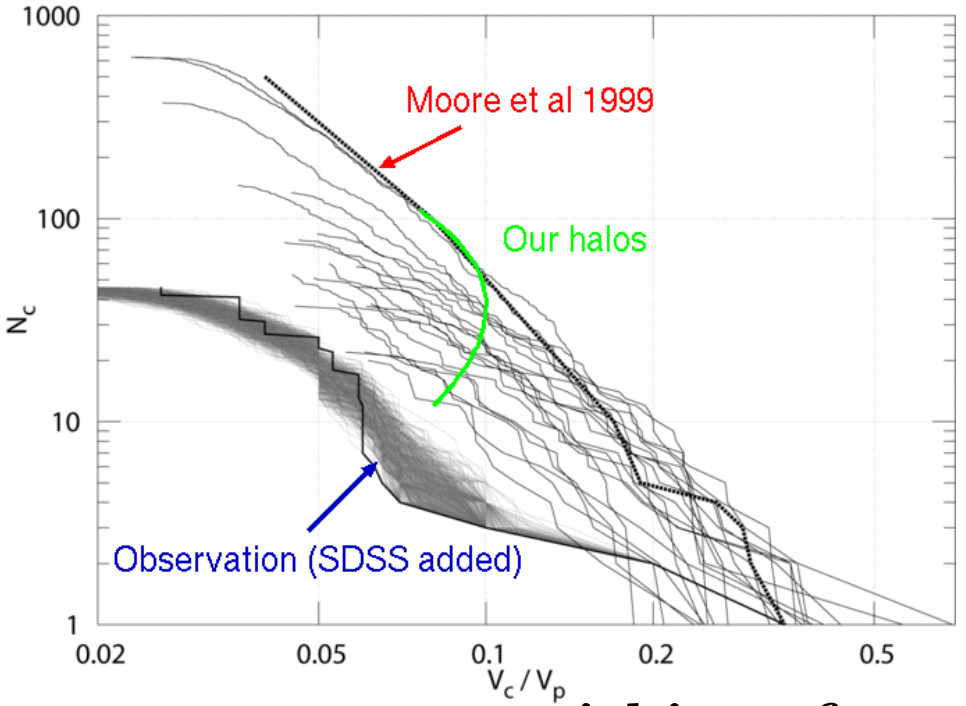
- ある領域での全部のハローを、無バイアスで「観測」
- GRAPE-6A クラスタ/PC クラスタ with IB/XT4
- 512^3 particles — 2048^3

512^3 計算結果



アニメーション例

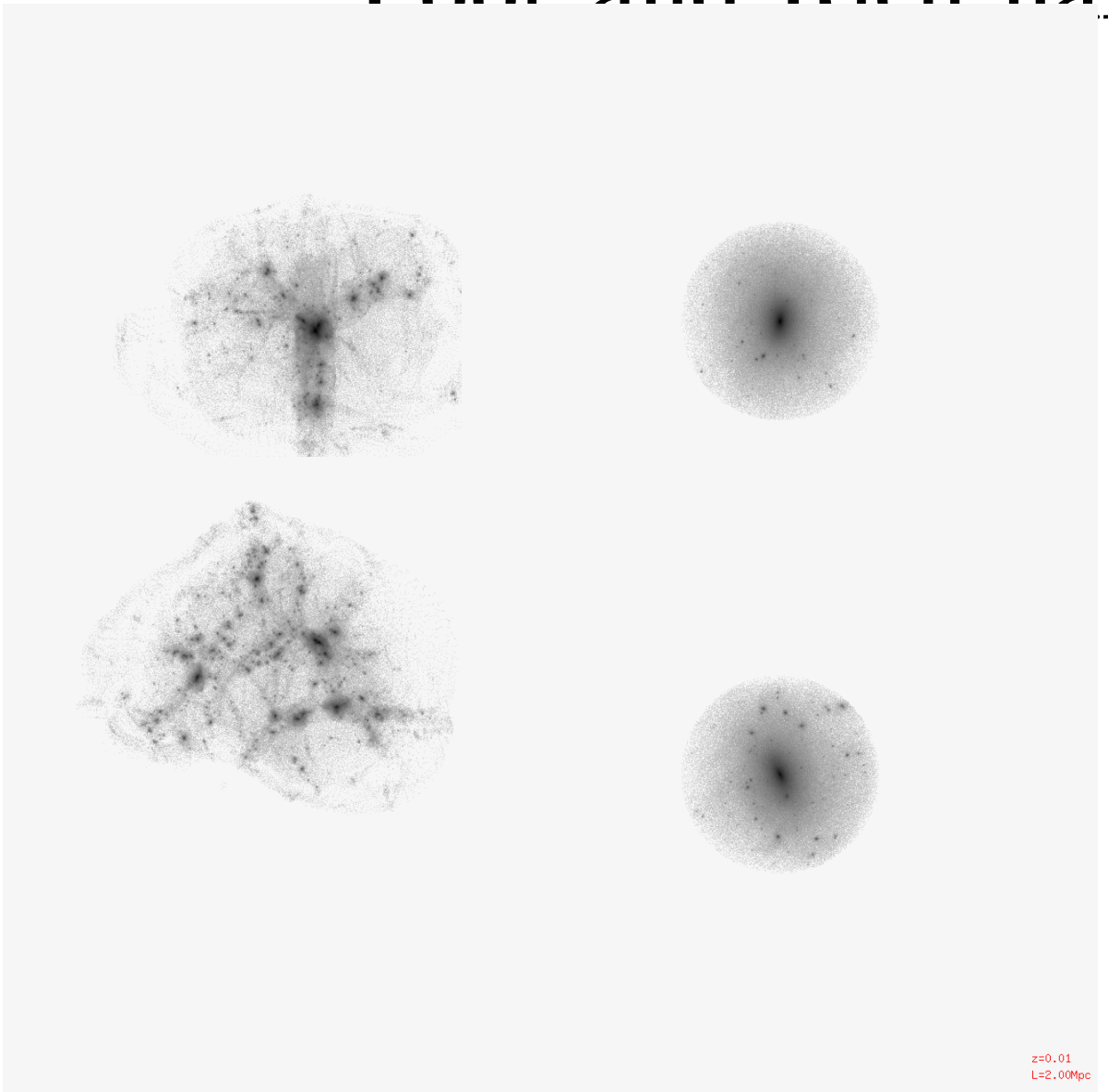
Result



- Large variation in number of subhalos
- The richest ones agree with Moore's result The poorest ones are within a factor of two with observations

= Dark CDM subhalos are not necessary

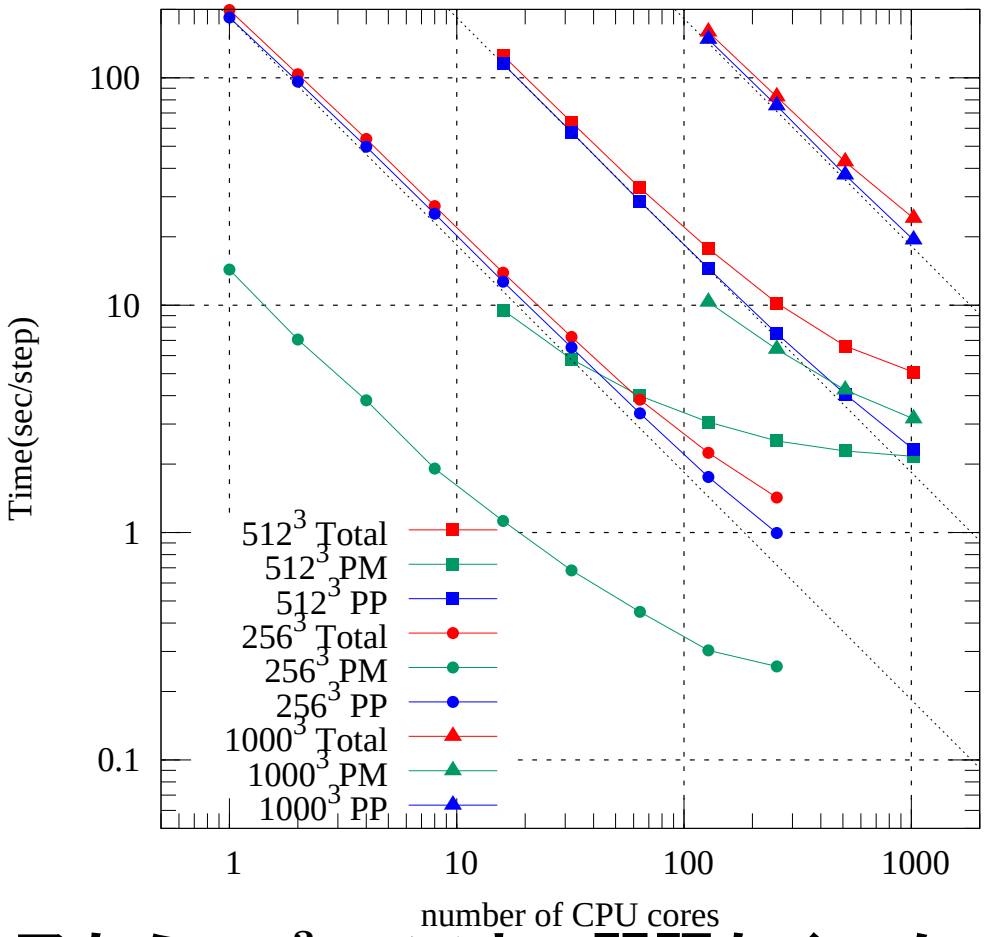
Poor and Rich halos



A poor halo at z=3 (left) and 0 (right)

A rich halo at z=3 (left) and 0 (right)

性能



10^9 粒子なら 10^3 コアまで問題なくスケールする。

銀河形成シミュレーション

アニメーション例
アニメーション例
まだあまりスケーラビリティはよくない (128 コア程度)
アルゴリズム改良中。目処はたっている。

グリッド計算実験

- NAOJ XT4 とアムステルダム大学の Power6 システムを 10G でつないで並列計算
- そのための異機種並列コード開発 (一応終了)
- ネットワーク接続実験: 2008/5/21
  - アムステルダム側は準備 (Power6 自体、、、) が間に合ってない
  - XT4 の2個の 10GbE インターフェースを使って、大陸間折り返し実験
  - コンスタントに 6Gbps (PCI-X インターフェースの限界) を実現
- 技術的可能性は実証出来た

ネットワークとセキュリティ

- アカデミックな大規模数値計算のユーザの計算機の使い方は 40 年前から同じ
  - 計算ジョブを投入する
  - 結果を「手元」に回収する
- とはいえ、計算機へのアクセス方法は変わった

  40 年前　パンチカードと磁気テープ

  30 年前　計算機センターの TSS 端末、ディスク、磁気テープ

  20 年前　ネットワークからのアクセス、 telnet, ftp

  10 年前　ネットワークからのアクセス、ssh, VPN
- 最近の変化:セキュリティ
- 「手元」の意味も変化。ネットワークが速ければ計算機センターのディスクも「手元」

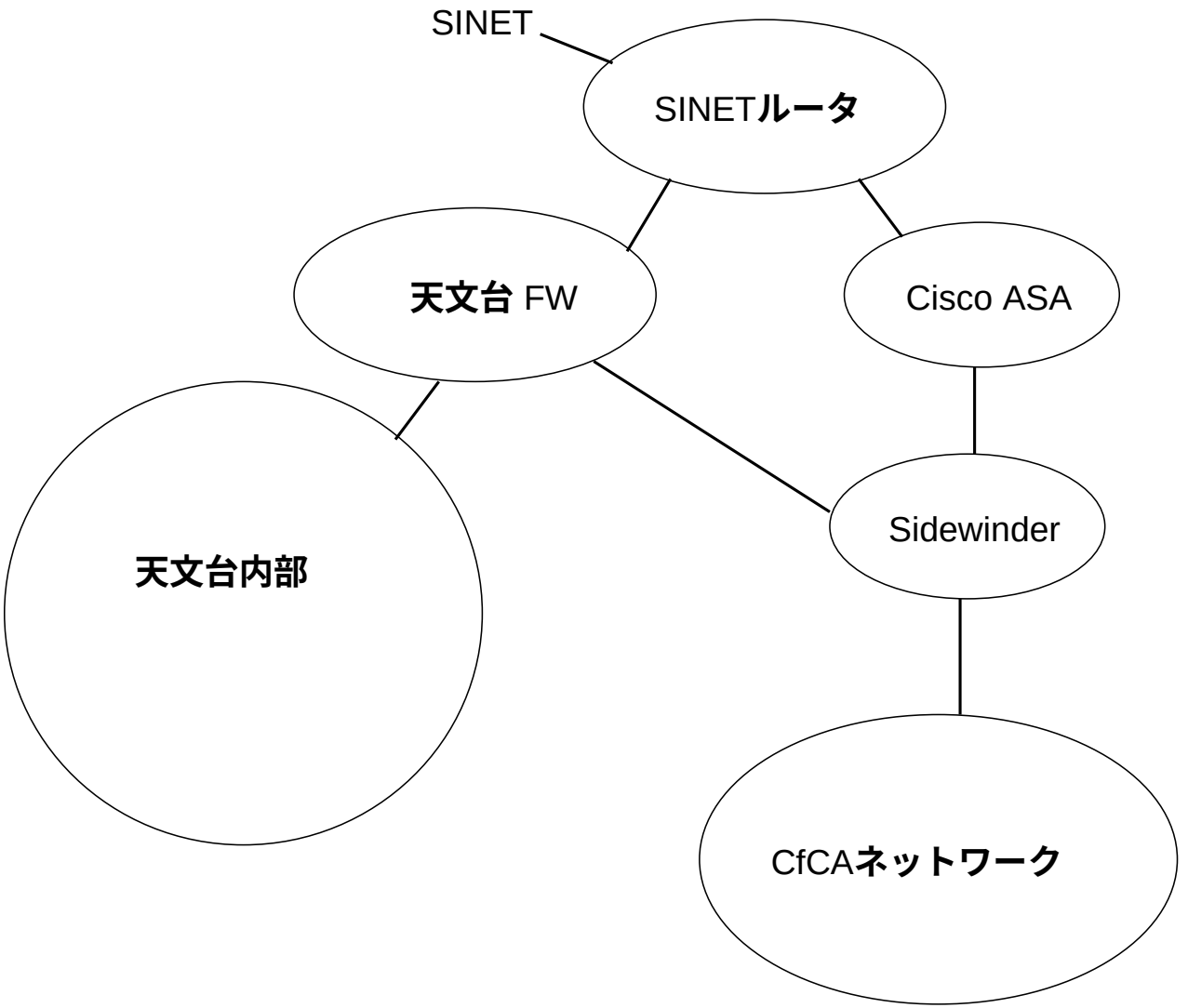
計算機センターとしてのセキュリティ

- 外部 (天文台内部含む) から侵入されない
- ユーザーレベルで侵入された時に他のユーザー/内部/外部になるべく迷惑をかけない

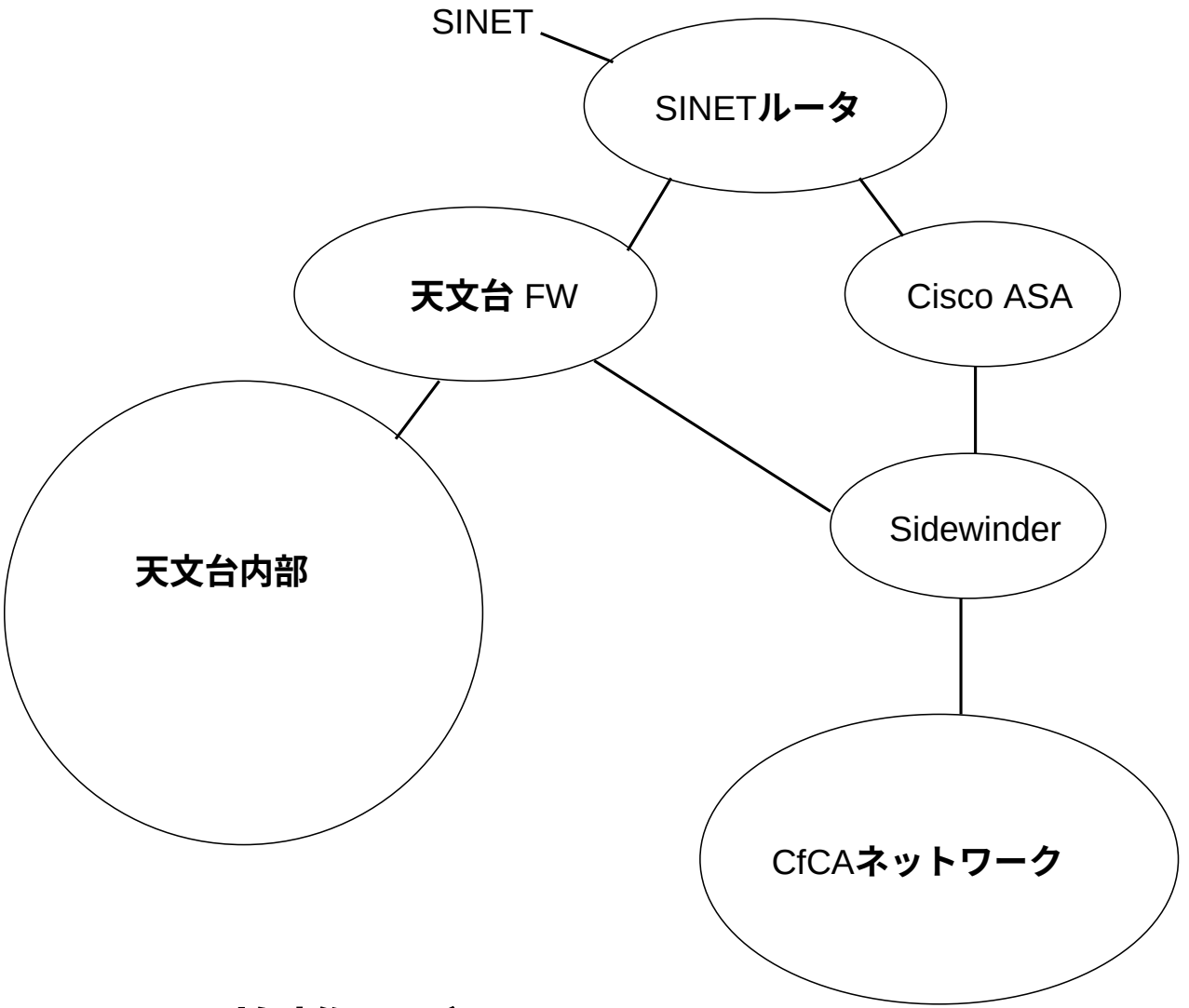
現在のシステム

- ユーザー認証は RSA SecurID によるワンタイムパスワード
- VPN は Cisco ASA 5510
- ファイアウォールに Sidewinder

概念図


SINET
SINETルータ
天文台 FW
Cisco ASA
天文台内部
Sidewinder
CfCAネットワーク

概念図



なんだか複雑すぎない？

複雑な理由

の少なくとも一部:

組織の構造に本質がある問題を技術的な方法で回避しようとしているため

東大等の国立大学、その (旧) 付置研にある程度共通の歴史的事情:

- **インターネット接続の歴史が長い。セキュリティが問題になるはるかに以前から**
- **インターネット接続がボトムアップな形で発展してきた**

東大の場合

- 1990年まで: UTNET以前
  - WIDE, JUNET 接続
  - 理学部(主に物理)の TISN
  - 工学部の「工学部LAN」
  - (教養学部:「3号館LAN」—吉村伸他)
- 1990-95 UTNET1
  - FDDI リング
  - 工学部LAN とは並存
- 1996-2000 UTNET2
- 2001- UTNET3

東大におけるネットワーク運用の複雑さ

- UTNET は全学のもの、各建物、キャンパス間を接続 (建物内は各学部の責任)
- 運用は学部で縦割、学部毎に全く方針が違う。
  - 理学部: 教室毎に論理的にも物理的にも別々
  - 教養学部: 建物毎に別なだけ
- 計算機センターも 3 組織 (現在、形式的には統合)
  - 大型計算機センター
  - 教育用計算機センター
  - NOC

セキュリティはどの部局の責任？

東大の話はこれくらいにして

天文台のネットワーク

- 歴史的には複雑
- 地理的に多数の観測所、そのいくつかは三鷹キャンパスより「大きい」
- 2003 年くらいまでネットワークの専門家がいなかった
- 現在も全台のネットワークを管理する人員は不足。常勤職員 2 名、助教と技術系職員 (昔の「技官」)1 名づつ
- その辺詳しくは以前 UNIX Magazine に連載されていた「国立天文台のネットワーク」を御覧下さい

天文台での外向け高速接続の必要性

- 観測では実はあんまりない
  - 望遠鏡:一晩に多くて数十枚の画像、大きくて 1GB
  - 衛星データ:ダウンリンクの速さで上限
  - VLBI (超長基線干渉計):リアルタイムの合成には Gbps 以上、現状はまだテープを輸送
- シミュレーションではいくらでも欲しい
  - 1 モデル 100TB くらいデータ出したい
  - Grid 利用なら 10Gbps でも不足

というわけで、要するに CfCA でしか使わない。

CfCA でのネットワーク設計の方針

- あまり全台ネットワーク担当に負担をかけない
- それなりの高速性と高い安全性を両立させる
- 予算は現実的な範囲で
- ユーザーからみて使いやすく、かつ安全に
- 管理の手間をなるべく少なく

全台ネットワーク担当に負担をかけない

- 全台ファイアウォールの設定は単純に
- 細かい設定、監視は CfCA のファイアウォールで
- CfCA の共同利用計算機ネットワークから、外部ユーザーは天文台内にアクセスできないように制御

高速性と安全性

速度

- 天文台と SINET3 の間は太い (20Gbps 程度まで利用可能)
- 現在この速度を利用するにはファイアウォールバイパスして接続先と VLAN にすることになる
- 近い将来にこの程度の速度をファイアウォール経由でも実現したい

安全性

- 実績あるもの
- 運用当初想定していなかった使い方にも対応できるとありがたい、、、

予算

- CfCA のスパコンレンタル料金は年間 2.5 億
- セキュリティ関係にはその 1% 程度を想定
  (実際にはもうちょっと、、、)
- 運用のための人的コストが少ないことは非常に重要
  - 法人化後も結局総定員法の影響下にあるため正規職員が増やせない
  - 増やせたとしてもネットワークの専門家を国家公務員の給料で雇えるか？

ユーザーからみた使いやすさと安全性

- ユーザーからみたら、セキュリティのための色々な仕掛けは単に「邪魔」
  - ログインの手順が面倒
  - ファイル転送が素直にできない、遅い
  - 画面を飛ばすことができない
  - etc etc……
- とはいえ、従来の ssh でだれでもどこからでも、というシステムではセキュリティ対策は困難
  - ユーザーのパスワード管理
  - ssh 自体のセキュリティホール

管理の手間

- 絶対的な人数が不足
- そんなに詳しい人がいるわけではない
- 最新のアプライアンス管理、運用経験のある人、、、

というわけで、優秀な人が膨大な時間と労力をかけて、というわけにはいかない

CfCA の現状の構成

現状の対応 (ベストというわけではない)

- ユーザーの外からのアクセスは IPsec-VPN が基本
  - Cisco ASA5510 を VPN 専用に利用
  - SSL-VPN も用意
- 認証は RSA SecurID
- VPN 以外の全てを Sidewinder G2 で一元化
- SSH サービスも用意 — Sidewinder を使ってみたかった理由の 1 つが、SSH トラフィックもみてくれること

半年間運用の後の感想

- Sidewinder は殆どトラブルなし (5 月に一度プライマリの DNS サービスがこけた)
- VPN、 SecurID による認証は色々あり

VPN/SecurID トラブル

Sidewinder とはあんまり関係ないですが、、、

- ブラウザからの認証トラブル
  - 何故か 1 度間違えると上限回数まで間違えたことになる？
- Linux クライアントの問題
  - Linux からの接続要求が多い: 研究室での数値計算やファイルサーバに利用しているケースが多いため？
  - Cisco VPN Client がサポートしてない Linux distribution ユーザーが多い。
- クライアント側のネットワークファイアウォール
  - 多くの機関で外向きの VPN が利用するポートもあけてない。
- その他色々細かいトラブル、、、

SSH サービス

- Sidewinder のパスポート認証と組み合わせた利用
- 認証した IP アドレスの機械からのみ接続可
- ssh アクセス先はゲートウェイ機械 (Linux の箱)。Cray や NEC 等のフロントエンドにはもう一度 ssh

大規模ファイル転送

- 現状では ssh/scp。
- あんまり速くない、、、
- CPU が速くなったので、ローカルディスクやネットワークの性能に比べて暗号化のオーバーヘッドは小さくなった
- おそらく長期的にも暗号化して十分。CPU の高速化 (並列度向上も入れると) はネットワーク高速化より速い。

今後の方向

- ユーザー側にファイル転送するのは現実的でなくなる。
  - 解析・可視化まで含めたトータルな環境を計算センター側で用意する必要あり
  - Grid ファイルシステム、ジョブ管理機能への対応も必要
- プログラム開発のためのネットワーク利用も多様化
  - 分散バージョン管理システム、BTS、プロジェクト管理システムへの対応
- セキュリティアプライアンスも多様な接続方法、利用方法に対応する必要

まとめ

- 国立研究所での大規模数値計算向け計算センターの対外ネットワーク構成は色々ややこしい問題を解決/迂回する必要がある。
- そのためには、一元管理が可能なセキュリティアプライアンスで多様な使い方に対応できる Sidewinder は有用である。
- これまで半年間の運用でも、ハードウェア、ソフトウェアともにノートラブルに近く、計算センターの安定運用に大きく貢献している。