

インターネットコンピューティング

担当 佐藤

概要

- ◆ 前回：
 - 巨大な計算能力を持つスーパーコンピュータは、現在の科学には欠くことのできない手段になっている。コンピュータを用いて行う計算科学の技術、成果について、解説する。
- ◆ インターネットコンピューティング
 - 広域のネットワークを利用して行う、グリッドコンピューティング、P2Pコンピューティングなどの最新動向について、解説し、これからのインターネット技術について展望する。

もくじ

- ◆ インターネットの基礎知識
- ◆ Web技術とWebサービス

- ◆ P2Pコンピューティング
- ◆ グリッドコンピューティング

インターネットの仕組み

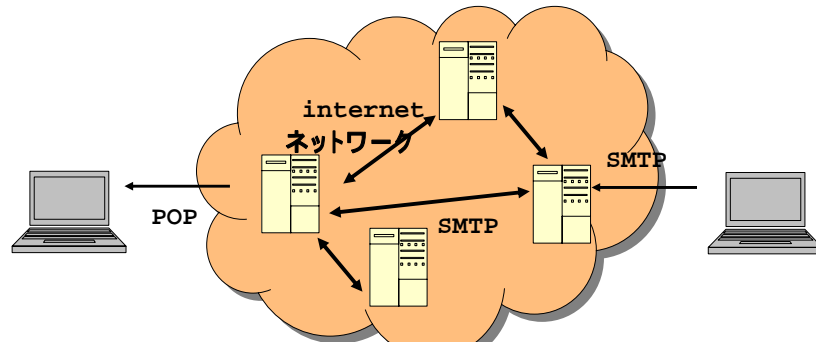
- ◆ IPアドレス (Internet Protocol address)
 - 通信を届けるためのインターネット上の番地
 - 通信を行うためには、IPアドレスを持つ必要がある
 - サーバーはIPアドレスを持つ
 - 一般のPCは、接続したときにIPアドレスをもらう (DHCP: Dynamic Host Configuration Protocol)
 - ドメインネーム IPアドレスに対応した名前
 - DNS (Domain Name Server) IPアドレスとドメインネームの対応



インターネットのサービス

◆ E-mailサービス

- SMTP (Simple Mail Transfer Protocol)
 - メールを配送するためのプロトコル
- POP (Post Office Protocol)
 - メールサーバからメールを取り出すプロトコル

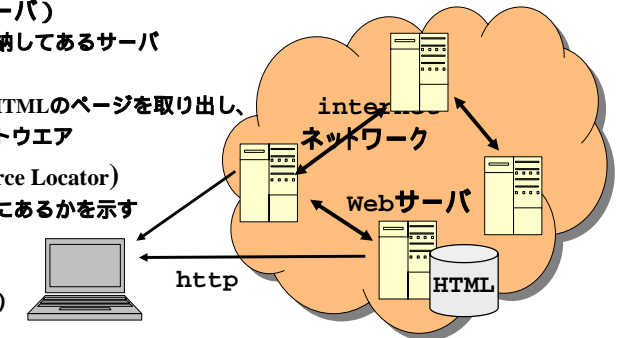


インターネットのサービス

◆ WWW (or web)

- World-Wide Web (蜘蛛の巣のように関係が張り巡らされているイメージ)
- HTML (Hyper Text Markup Language)
 - Webのページを書くための言語。文書フォーマットの記述ができる。HyperLink (他のページへの参照)を持つ。
- HTTP (Hyper text transfer protocol)
- HTTPサーバ(webサーバ)
 - Webのページを格納してあるサーバ
- Web ブラウザ
 - HTTPサーバからHTMLのページを取り出し、表示するためのソフトウェア
- URL(Uniform Resource Locator)
 - Webページがどこにあるかを示す

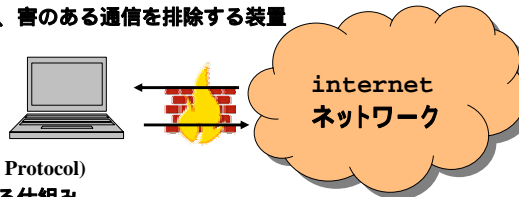
Webブラウザ (IE, firefox,...)



インターネットの基礎知識

◆ Firewall

- 特定の通信だけを通すようにし、害のある通信を排除する装置



◆ DHCP(Dynamic Host Configuration Protocol)

- 一時的なIPアドレスを割り当てる仕組み

◆ DOS攻撃(Denial of Services)

- 相手のコンピュータやルータなどに不正なデータを送信して使用不能に陥らせたり、トラフィックを増大させて相手のネットワークを麻痺させる攻撃

◆ Spamメール

「SPAM」とは、Hormel Foods社の味付け豚肉の缶詰の商品名のこと。イギリスのコメディイアン「Monty Python's Flying Circus」(モンティ・パイソンの空飛ぶサーカス)の有名なコントに次のようなものがある。レストランに夫婦が入ってきてメニューを覗いていると、近くに座っているパイキングの一行が「SPAM, SPAM, SPAM!」と大声で歌いだす。次第に店員も「SPAM」を連呼しだし、最初は嫌がっていた夫婦も最後には困してSPAMを注文せざるを得なくなる、という筋書き。ほしくもないのに大量に送りつけられてくる広告メールから、このコントでしくく連呼される「SPAM」を連想したのが由来とされている。

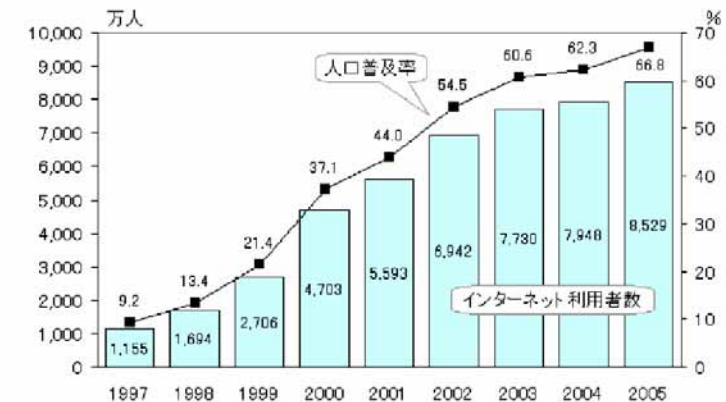
◆ コンピュータウイルス

- 他人のコンピュータに勝手に入り込んで悪さをするプログラム

<http://www2.ttcn.ne.jp/honkawa/6210.html>

インターネットの普及

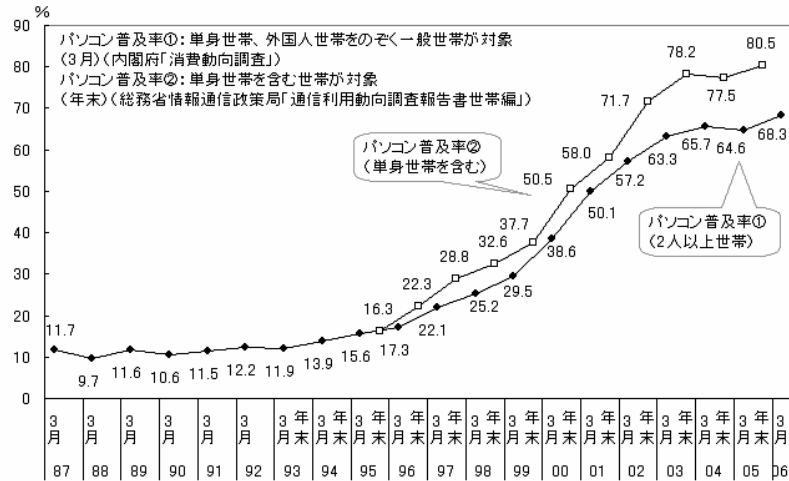
インターネット利用者数・人口普及率



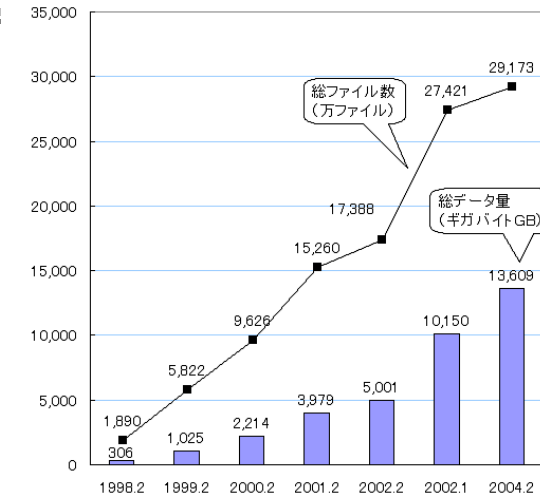
(注) 年末の推計。インターネット利用者数は、パソコン、携帯電話、ゲーム機等のいずれかでの利用者。対象年齢は1999年まで15～69歳、2000年末15～79歳、2001年以降6歳以上。
(資料) 総務省「通信利用動向調査」

PCの普及

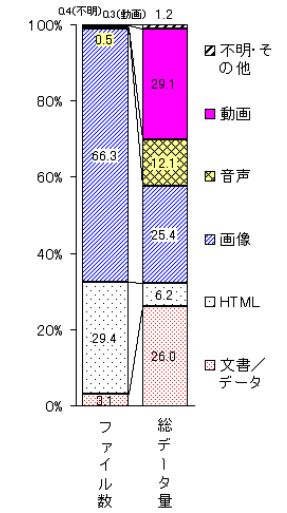
パソコン世帯普及率



わが国のウェブ上の総コンテンツ量の推移(推計)



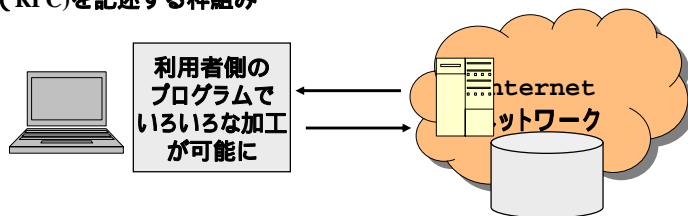
(ファイルタイプ別比率)



(注) JPDメインのインターネットコンテンツ量。ファイルタイプ別比率は2004年2月調査。
 (資料) 総務省情報通信統計データベース

Web サービス

- ◆ “ Web ” (ブラウザ)を使ったサービスではない!
- ◆ Webの拡張として、PC上のプログラムからwebの情報が使えるようになる技術
 - 利用者側で、いろいろな拡張ができる
 - サーバ間でいろいろな情報を交換できる(企業同士の取引など)
- ◆ 仕組み
 - 決められたフォーマットを通じて、webサーバ側のプログラムを呼び出すことができるようにした。
 - WSDL (Web Service Description Language)を媒介として、サービス (RPC)を記述する枠組み



Google API

- ◆ <http://www.google.com/apis/index.html>

「米Googleは11日、「Google Web API」を公開した。これはSOAP1.1およびWSDLに基づいたインターフェイスで、Googleがこれまでに収集した20億ものWebページのデータを自分の好きなようにプログラミングして利用することができる。この試みについては多くの開発者が「新たな革命」「Netscapeのリリース以来の出来事」と賞賛している。」(2002年4月)

- ◆ Webサービスで公開されている
 - ただし、一日で検索できる件数は限られている
 - Javaだけでなく、.NET, PHP, Perl, PythonからもOK
- ◆ Amazon.com, yahoo, などいまではいろいろな企業がwebからのWebを提供している。

このGoogleAPIがインターネット全体に及ぼす影響は計り知れないものだ。20億ページのデータを自在にプログラミングできることから、どのようなアイデアが生まれてくるのか現在のところ想像もつかないが、GoogleではそのためにいくつかのアイデアをFAQのなかで示している。

例えば、「定期的に特定のキーワードについて新しいページが登録されていないかを走査、通知してくれるプログラム」、「Webの中にある情報の量に応じて市場にどのような需要があるのかを分析する市場調査プログラム」、「コマンドラインや携帯電話などHTML以外のインターフェイスを通してGoogleを利用するプログラム」、「Webにある情報を使った“おもしろい”ゲーム」などを挙げている。

Googleでは、このGoogleAPIの議論のためにニュースグループを設置。開発者たちがアイデアや意見を交換できるようにしており、自分が開発したプログラムを公開することも歓迎するとしている。こうしたアイデアは後々Googleの新たなビジネス戦略にも用いられるものと思われる。

GoogleAPIを利用するためには、開発者キットをダウンロードしなければならない。これには開発者に必要となるJavaライブラリ、WSDLファイル、その他サンプルコードなどが含まれており700KB弱のファイルである。ダウンロードの後、新たに設けられた「Google Account」に登録することでライセンスキーを取得、それによってGoogleAPIを試すことができる。GoogleAPIがサポートしているプログラミング言語は、Java、Perl、Ruby、Microsoft VisualStudio.NETだが、GoogleではWebサービスをサポートしているプログラミング言語ならうまく動作するのではないかと示唆している。

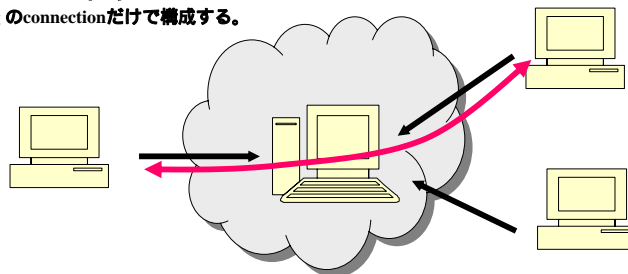
Webをめぐるあたらしい動き

- ◆ 「巨人」 google
 - Google map, google earth, ...
 - Goole mail (gmail), application service(表計算、ワード)
 - 動画情報サイト youtubeを買収

- ◆ Web 2.0
 - 明確な定義はない (buzzword!)
 - Tim O'reilly氏の論文「What is Web 2.0」
 - ユーザ側からの関与、参加型の情報サービス
 - ブログ (Blog)
 - SNS (Social Networking Site/Services)
 - Wikipedia web上の百科事典

P2P技術とは

- ◆ P2P (peer to peer)コンピューティング
- ◆ 固定のIPアドレスを持たないIPeer (個人)とPeerを (直接) 通信できるようにする技術
 - IPアドレスは接続時に一時的に取得
 - Firewallの中にあることもある
 - ファイル交換ソフト(Gnudella,...)で注目
 - 各PCが自立的につながる形態
 - 各PCは対等の関係(?)
 - out-going のconnectionだけで構成する。



P2Pコンピューティングの目的

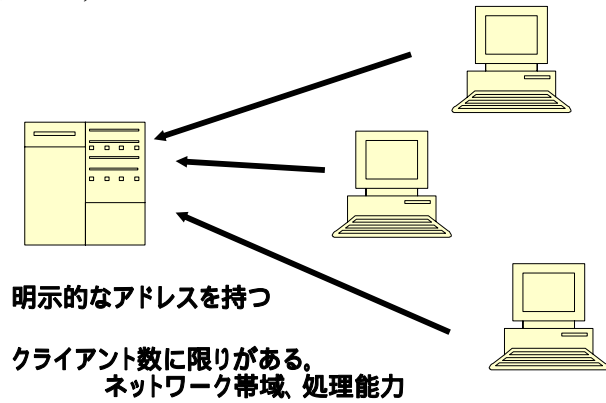
- ◆ 潜在的には世の中に数百万台 (以上?) のコンピュータがあり、これがネットワークに接続されている。
 - 社内、大学でもフルに活用されていない資源がたくさんある
 - 計算資源としてもちいることができないか?
 - ファイルを分散して格納できないか?

- ◆ 目的
 - ファイル共有
 - インスタントメッセージ
 - Skype
 - コラボレーションツール
 - 分散コンピューティング

従来の分散コンピューティングのモデル

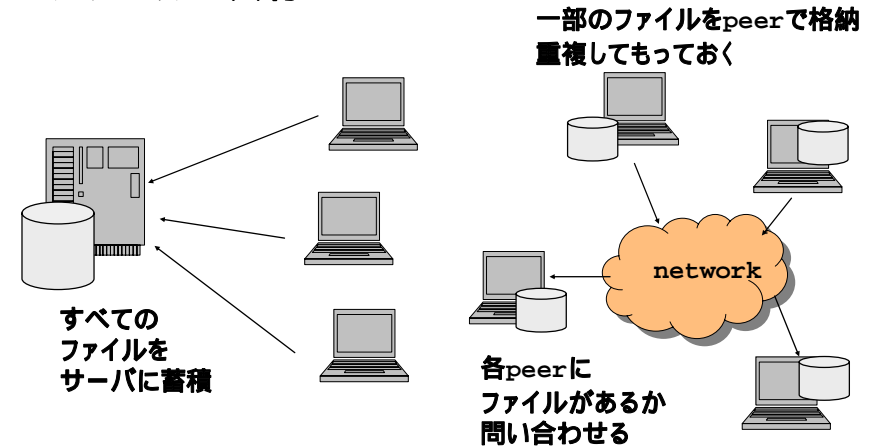
◆ クライアント・サーバモデル

- WWW、FTP、e-mail



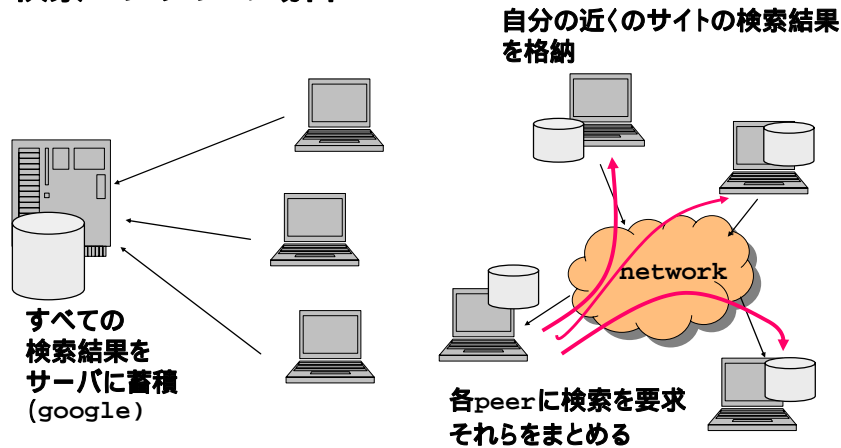
P2Pで何ができるか (1)

◆ ファイルの共有



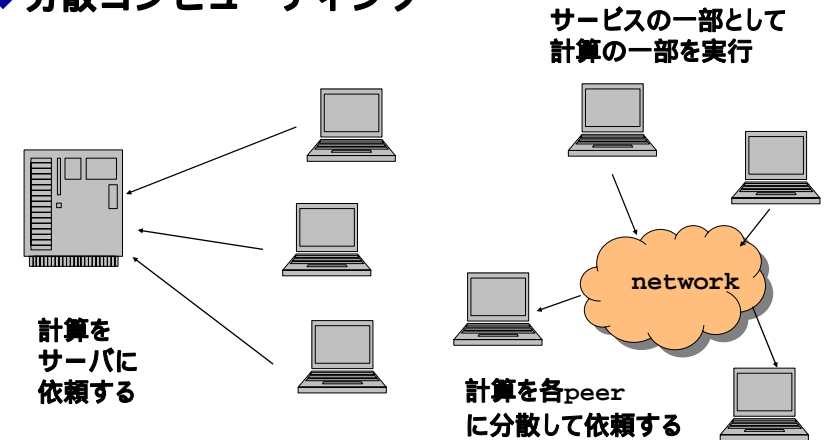
P2Pで何ができるか (2)

◆ 検索エンジンの場合



P2Pで何ができるか (3)

◆ 分散コンピューティング



P2P技術によるファイル共有

- ◆ Napstar, Gnutella,... Winny
- ◆ 個人同士であれば、ファイルをやり取りしてもいい？
 - 著作権の問題
- ◆ Winnyの問題
 - PCを仲介する中央サーバを必要とせず、暗号化されたファイルを「ノード」と呼ぶ各ユーザーのPCを次々と経由させることでダウンロードさせる仕組みを採用。ダウンロード要求を出したPCと、該当ファイルを持つPCがダイレクトに1対1で接続されることがないため、違法ファイルを持つユーザーが特定されにくい匿名性の高さが最大の特徴
 - 作者の意見「個人的な意見ですけど、P2P技術が出てきたことで著作権などの従来の概念が既に崩れはじめている時代に突入しているのだと思います。お上の圧力で規制するというのも一つの手ですが、技術的に可能であれば誰かがこの壁に穴あけてしまって後ろに戻れなくなるはず。最終的には崩れるだけで、将来的には今とは別の著作権の概念が必要になると思います。どうせ戻れないのなら押しちゃってもいいかって所もありますね」

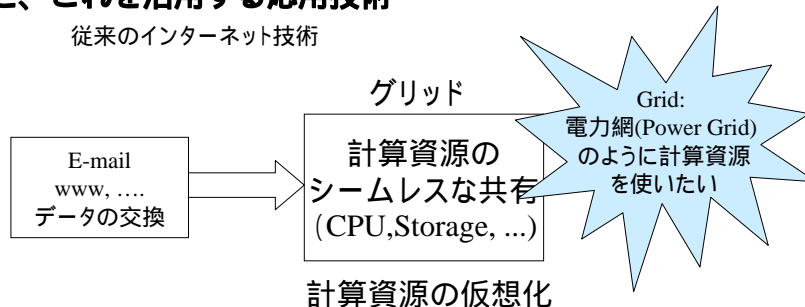
グリッドとは

- ◆ 世界中のPCをつないで計算 seti@home,...
 - これだけではない！
- ◆ スーパーコンピュータをつないで大規模計算
 - これだけではない！

グリッドコンピューティングとは

- ◆ グリッド技術とは広域の高速ネットワーク上において、**「安全に」**大量のデータ、計算資源、貴重な装置等を共有し、協調作業、資源の有効活用するネットワーク基盤技術（ソフトウェア、ネットワーク、ハードウェア）と、これを活用する応用技術

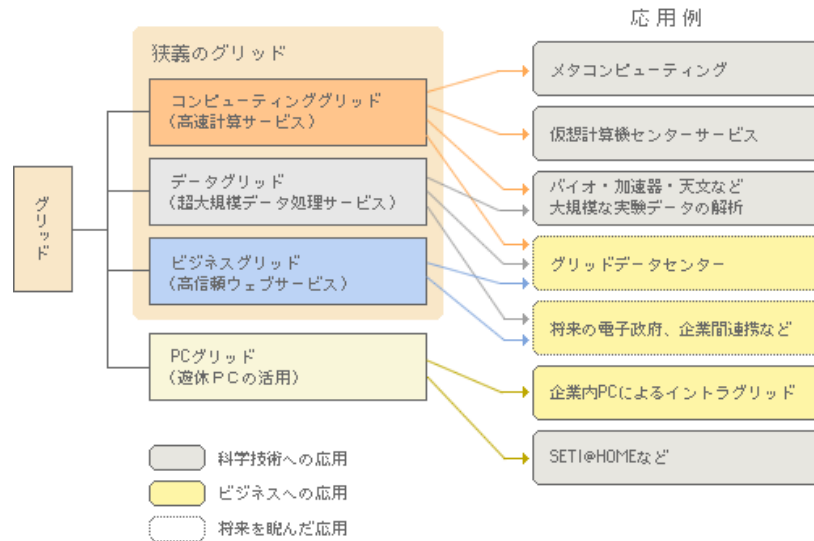
従来のインターネット技術



グリッドコンピューティングとは

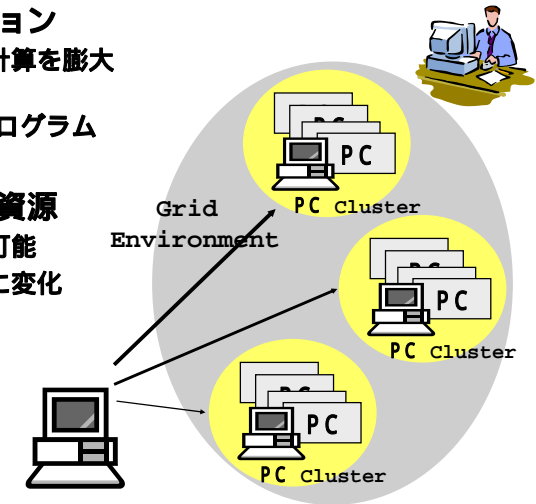
- ◆ なにが変わったのか？
 - 以前から、分散コンピューティングの研究はあった。
 - インターネットの急激な進歩、普及
 - 全世界で共通な基盤を作ろうとしている！
- ◆ 何につかえるのか？
 - 計算資源（スーパーコンピュータ）を共有し、大規模計算を行う(meta-computing, MPI-G, ..., ITBL?)
 - 遠隔の計算資源と手元のPC/WSをシームレスに結合(computing portal, GridPRC)
 - 大量のデータの処理(data intensive computing, gfarm for ATLAS)
 - 高価な装置の遠隔共有(電子顕微鏡、衛星データ、加速器?)
 - 共同作業のサポート(電子会議システム、AccessGrid)
 - 遊休の計算機の利用、SETI@home, P2P, ..., XtermWeb

グリッドコンピューティングの分類



グリッドと並列アプリケーション

- ◆ 典型的なアプリケーション
 - パラメータ検索：同じ計算を膨大な計算資源で実行
 - master-slave型の並列プログラム
- ◆ 典型的なグリッド計算資源
 - 複数のクラスタが利用可能
 - 計算資源の状況が動的に変化

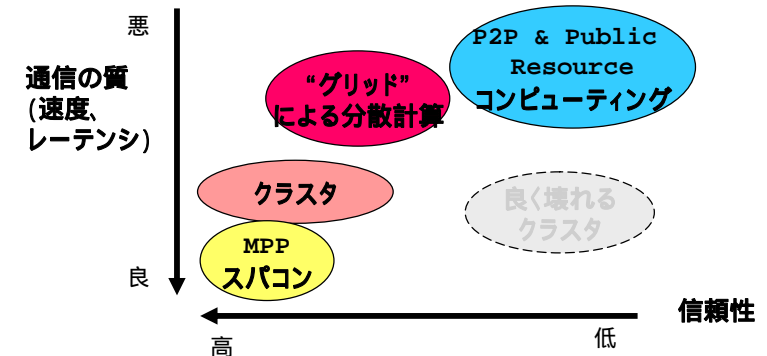


Public Resource コンピューティング

- ◆ PCが自発的に計算に参加して行うボランティアによる大規模計算プロジェクト
- ◆ SETI@HOME
 - Search for Extraterrestrial Intelligence
 - http://www.planetary.or.jp/setiathome/home_japanese.html
 - インターネットにつながっているコンピュータを使って地球外知的生命体の探査(SETI)を行なう科学実験です。無料のプログラムをダウンロードして電波望遠鏡のデータを分析することで、あなたも参加することができます。
- ◆ Folding@HOME
 - <http://folding.stanford.edu/>
 - タンパク質の折り畳み、異常折り畳み及び関連の疾病の理解

グリッドコンピューティングで何ができるか

- ◆ 大量の計算
- ◆ 大規模なデータの処理
- ◆ グリッドは、スパコンにとって変わるものではない！
 - アプリケーションをえらぶ。coarse-grain...
 - 何をさせるかは問題！...でも、パラメータサーチはたくさんある



グリッドでの大規模計算

- ◆ 潜在的な計算資源は、たくさんある！
 - P2Pならばオフィス・家庭にあるPCも、... 1万台のPCがあれば、
 - 1GFlopsのPCがあれば、10TFlops
 - 50GBのディスクがあれば、500TB

◆ 従来の計算

- ある条件で、どのような現象が現れるか。



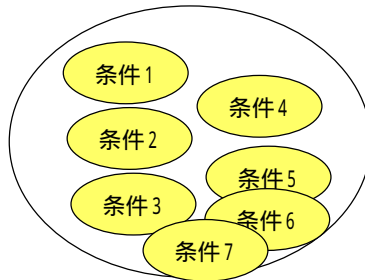
◆ 膨大な計算リソースが使えれば、...

- ある性質を持つ条件は何か。
- 設計

パラメータサーチ
アンサンブル計算

性質

パラメータ
サーチ



データグリッド

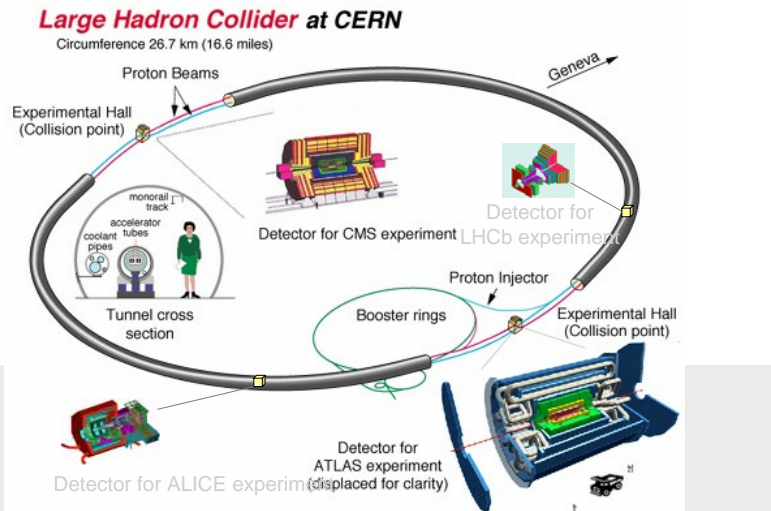
◆ 様々な電子的なデータが蓄積しつつある

- 加速器
- 衛星データ
- 天体観測
- センサーデータ
- ゲノム

◆ データマイニング技術による発見

◆ 分散されたデータの統合

Example: Large Hadron Collider (LHC) Accelerator at CERN



P2Pコンピューティングとグリッドの違い

◆ グリッド

- 複数の組織間の計算資源の共有 (VO: virtual Organization)
 - 組織間の相互認証
- 信頼できる計算資源
- 高速かつ信頼性の高いネットワーク
 - サイト間では、IP reachableなことが多い (VPNもあり)
- 数千台規模

◆ P2P

- 個人利用のPCも含む雑多な計算資源 (heterogeneity)
- 集中管理ができない, しない (decentralized)
- 信頼できない計算資源 (volatile)
 - 壊れる、... 電源が切られる
- 複雑なネットワーク構成
 - internet, firewall, NAT
- 数万台規模 (large scale)