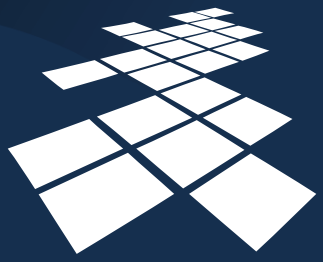
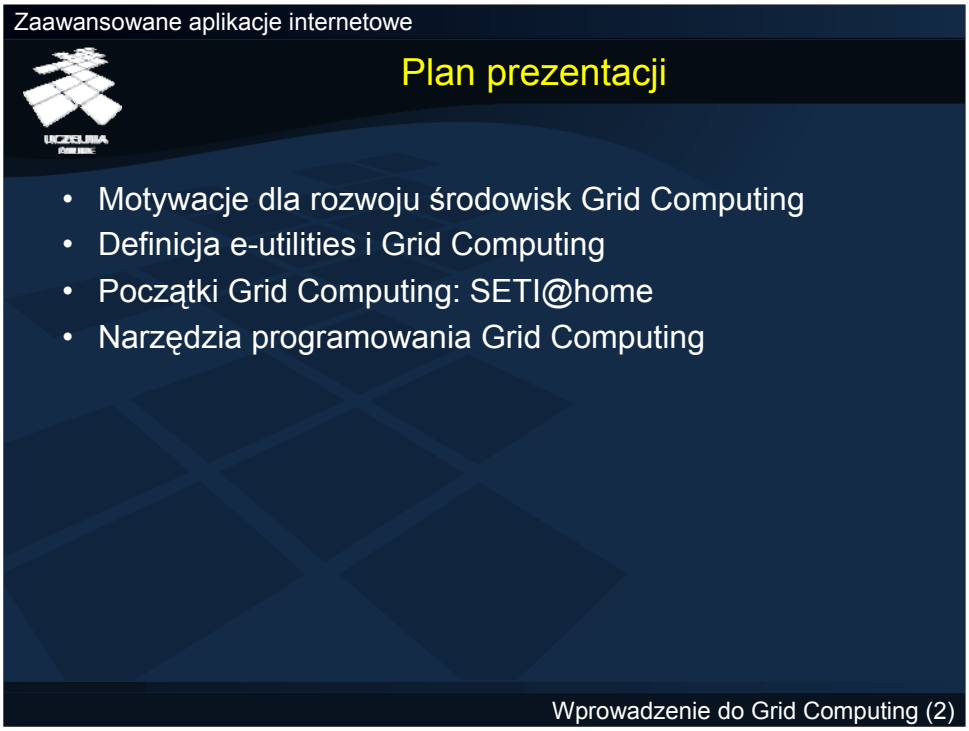
Wykład 9

Wprowadzenie do Grid Computing

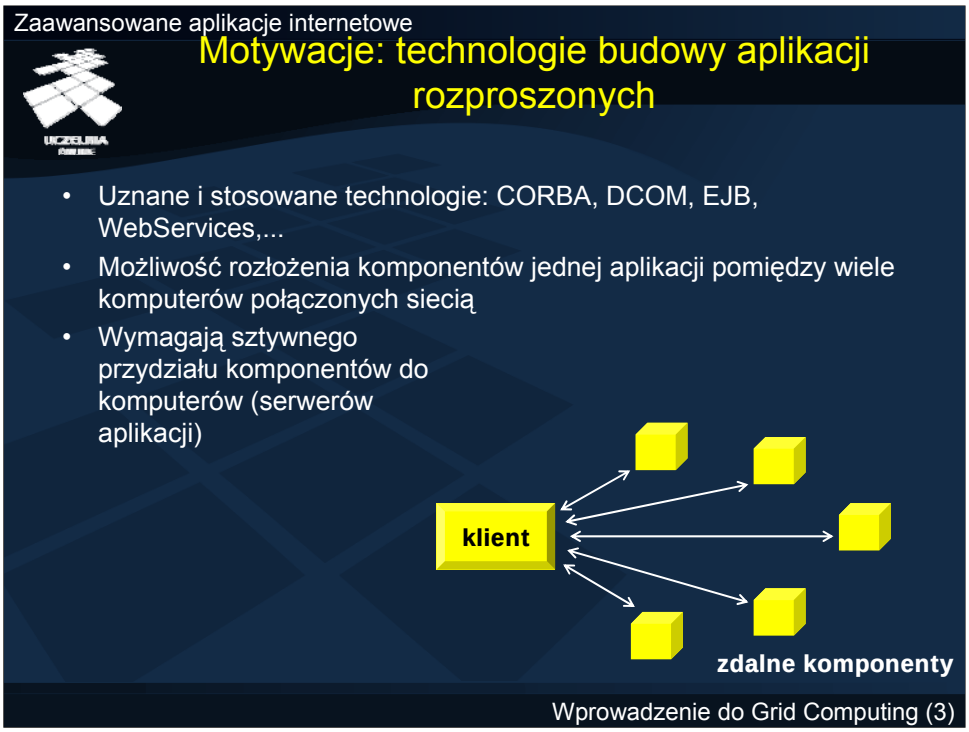
wykład prowadzi: Maciej Zakrzewicz



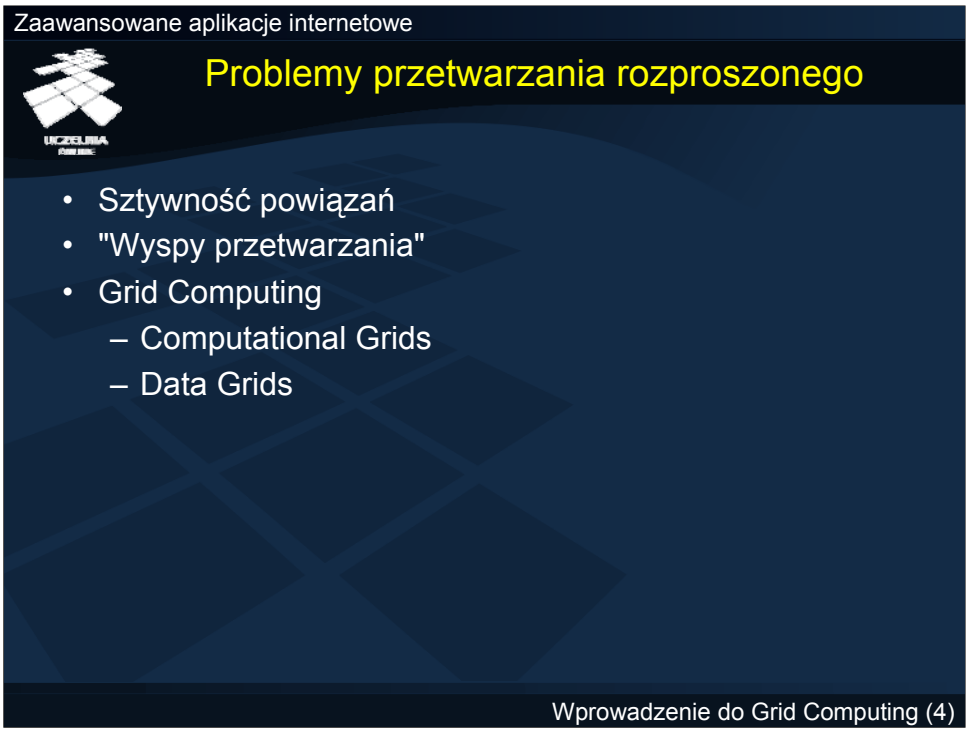
UCZELNIA
ONLINE



Celem wykładu jest przedstawienie koncepcji architektury Grid Computing. Omówione zostaną motywacje dla rozwoju środowisk Grid Computing, pierwsze zastosowania Grid Computing w projekcie SETI@home oraz narzędzia programowania tego typu środowisk.



Ogromny sukces zastosowań Internetu w dziedzinie globalnego upowszechniania informacji spowodował wzrost zainteresowania wykorzystaniem jego infrastruktury jako platformy dla realizacji aplikacji rozproszonych. W ciągu ostatnich kilkunastu lat zaproponowano szereg technologii, umożliwiających implementację komponentowych aplikacji rozproszonych, których elementy mogą być rozlokowane na wielu heterogenicznych węzłach połączonych siecią typu internet lub intranet: CORBA, DCOM, Enterprise JavaBeans (EJB), WebServices. Aplikacje realizowane z pomocą tych technologii składają się zwykle z pojedynczego modułu klienta, pracującego na komputerze użytkownika końcowego i odpowiadającego za obsługę graficznego interfejsu użytkownika oraz za sterowanie przetwarzaniem, oraz z wielu modułów zdalnych, osadzonych na serwerach aplikacji, odpowiadających za realizację właściwego przetwarzania danych. Charakterystyczne różnice pomiędzy wspomnianymi technologiami sprowadzają się do zakresu wspieranych języków programowania (np. EJB współpracuje wyłącznie z językiem Java), platform systemu operacyjnego (np. DCOM przeznaczony jest dla środowisk Microsoft Windows) oraz stosowanych protokołów komunikacji klienta z modułami zdalnymi (np. HTTP w WebServices, nastomiast CORBA i EJB korzystają z IIOP).



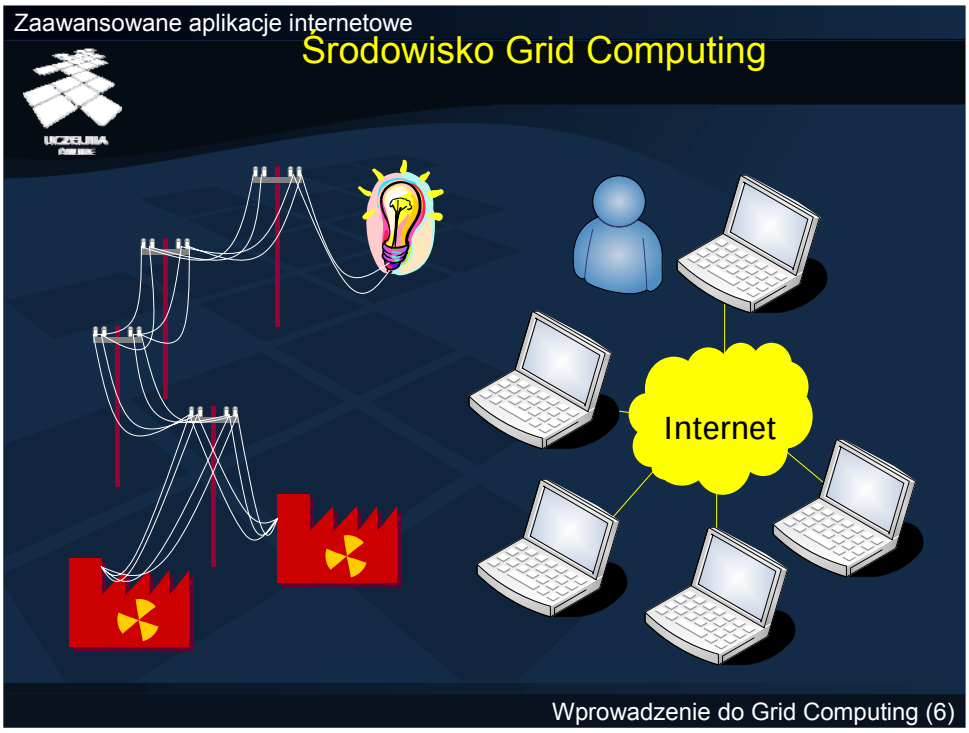
Pewnego rodzaju wadą wyżej wymienionych technologii budowy rozproszonych aplikacji komponentowych jest sztywność powiązania zdalnych modułów ze sprzętem lub serwerami aplikacji, do których zostały przydzielone. Osoby wdrażające rozproszone aplikacje komponentowe podejmują wiążące decyzje o fizycznej lokalizacji wszystkich elementów aplikacji. Działanie takie powoduje, że każdy system aplikacyjny posługuje się zbiorem dedykowanych komputerów, nazywanym często „wyspą przetwarzania danych” (island of computing), a zasoby sprzętowe przedsiębiorstwa są podzielone (najczęściej rozłącznie) pomiędzy posiadane systemy aplikacyjne. W codziennej eksploatacji może się zdarzać, że dochodzi do przeciążenia zasobów wykorzystywanych przez jeden system aplikacyjny, podczas gdy pozostałe systemy nie są w ogóle wykorzystywane przez użytkowników.

Problem sztywnego przydziału oprogramowania do sprzętu może zostać złagodzony dzięki zastosowaniu koncepcji przetwarzania Grid Computing (przetwarzanie siatkowe, przetwarzanie gridowe, przetwarzanie sieciowe). Grid Computing postuluje, aby traktować wszystkie posiadane zasoby sprzętowe jako jeden wielki komputer wirtualny, zdolny wykonywać wszystkie dotychczasowe aplikacje, automatycznie alokując je do poszczególnych maszyn w taki sposób, aby równoważyć obciążenie maszyn oraz redukować wpływ awarii na dostępność systemów aplikacji. Najczęściej mówi się o Computational Grids, które wirtualizują zasoby obliczeniowe (procesory, pamięci operacyjne) oraz o Data Grids, które wirtualizują pamięć masową (dyski twarde, napędy CD/DVD, streamery).

Wizja e-utilities

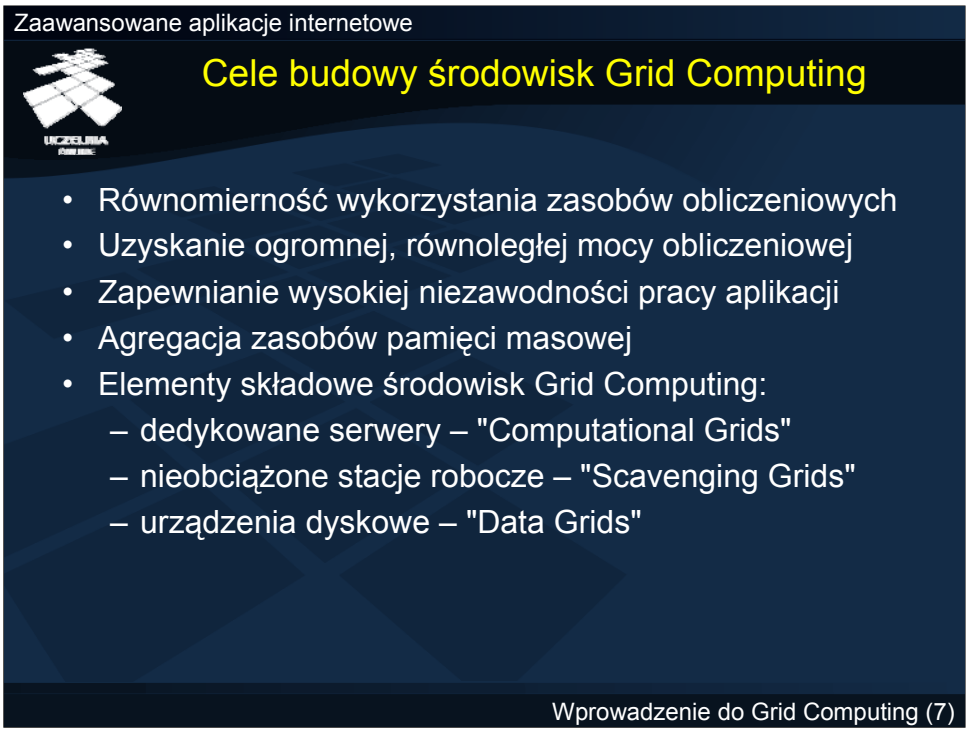
- Przechowywanie i przetwarzanie danych traktowane jak energia elektryczna, gaz i woda bieżąca
  - użytkownik nie musi być świadomy tego, skąd pochodzą i jaka jest architektura ich sieci dystrybucyjnych
  - użytkownik oczekuje gwarancji dostaw oraz zgodności parametrów z ustalonymi standardami (napięciem sieciowym, kalorycznością, ciśnieniem)

Często pisze się o Grid Computing w kontekście E-utilities, gdzie zwraca się uwagę na to, że zasoby służące do przechowywania i przetwarzania danych powinny być traktowane podobnie jak energia elektryczna, gaz i woda – użytkownik nie musi być świadomy tego, skąd pochodzą i jaka jest architektura ich sieci dystrybucyjnych, natomiast powinien oczekiwać gwarancji dostaw oraz zgodności parametrów z ustalonymi standardami (napięciem sieciowym, kalorycznością lub ciśnieniem). Użytkownik wykorzystywałby E-utilities powierzając do wykonania swoje zadania przetwarzania danych i otrzymując odpowiedzi w czasie zgodnym z ustaloną umową świadczenia usługi. Komunikacja ze środowiskiem E-utilities odbywałaby się poprzez ustalone standardowe interfejsy, natomiast usługi przetwarzania danych byłyby świadczone przez podmioty gospodarcze na zasadach podobnych do dzisiejszych Application Service Providers.



Przetwarzanie Grid Computing jest nową generacją przetwarzania rozproszonego, polegającą połączeniu dużej ilości heterogenicznych zasobów komputerowych (węzłów) w celu stworzenia iluzji posiadania komputera wirtualnego o bardzo dużej mocy obliczeniowej.

Najprostszym przykładem wykorzystania środowiska Grid Computing może być uruchamianie standardowej aplikacji na komputerze innym niż zwykle. Jeżeli komputer, na którym aplikacja jest zwykle uruchamiana, zostanie przeciążony, wtedy aplikacja może zostać uruchomiona na innym, nieobciążonym w danym momencie komputerze. Aby taki scenariusz mógł zostać zrealizowany, spełnione muszą być dwa warunki: (1) natura aplikacji musi pozwalać na wykonanie zdalne, np. z punktu widzenia interakcji z użytkownikiem i (2) zdalny komputer musi spełniać wymagania sprzętowe i systemowe stawiane przez uruchamianą aplikację.



W wielu przedsiębiorstwach występują ogromne ilości niewykorzystanych zasobów obliczeniowych. Większość stacji roboczych jest obciążana zaledwie w 5%. W niektórych przypadkach niedociążone są również serwery – dzieje się tak bądź za sprawą nadmiaru mocy w stosunku do rzeczywistego obciążenia, bądź ze względu na czasowe fluktuacje aktywności użytkowników. Budowa środowiska Grid Computing umożliwia wykorzystanie niedociążonych zasobów jako wsparcia dla tych zasobów komputerowych, które są chwilowo bądź permanentnie przeciążone. Reguły alokacji aplikacji lub zadań na węzłach środowiska Grid Computing są definiowane przez administratora systemu i mogą być przez niego dynamicznie zmieniane. Często dopuszcza się możliwość przenoszenia pomiędzy węzłami nawet tych zadań, które znajdują się w trakcie realizacji. Dzięki temu możliwe jest dynamiczne równoważenie obciążenia pomiędzy wszystkimi węzłami.

Uzyskanie ogromnej równoległej mocy przetwarzania jest jednym z najważniejszych atrybutów Grid Computing. Jest to odpowiedź na zapotrzebowanie formułowane nie tylko w ramach projektów naukowych, ale również przez technologie biomedyczne, modelowanie finansowe, technologie wydobycia ropy naftowej, animację komputerową, itd. Należy zwrócić uwagę na specyfikę aplikacji, które potrafią wykorzystać potencjał dużej liczby równoległych komputerów. Są to aplikacje wysoce aktywne obliczeniowo, ale podzielne na mniejsze fragmenty-podzadania, które mogą być niezależnie wykonywane na oddzielnych komputerach i nie muszą realizować intensywnej komunikacji z innymi podzadaniami. Potocznie o takich aplikacjach mówi się, że są "skalowalne", co w optymalnym przypadku oznacza, że np. dziesięciokrotne zwiększenie liczby procesorów spowoduje dziesięciokrotne skrócenie czasu pracy aplikacji. W praktyce tak dobra skalowalność nie występuje, głównie w związku z niemożliwością podziału algorytmów na dowolnie dużą liczbę podzadań lub w związku z koniecznością synchronizacji dostępu do współdzielonych zasobów. Oczywiście nie każdy rodzaj aplikacji może zostać efektywnie dostosowany do pracy w środowisku Grid Computing (grid-enabled applications). Przekształcenie istniejącej aplikacji w aplikację zdolną wykorzystywać własności środowiska Grid Computing musi zostać wykonane manualnie przez jej projektantów i programistów. Pomoc w realizacji tego procesu mogą stanowić ogólnodostępne biblioteki i środowiska programowania, np. Globus Toolkit.

Środowiska Grid Computing oferują także możliwości podnoszenia niezawodności aplikacji. Jeżeli awarii ulega węzeł realizujący jedno z podzadań aplikacji, wówczas podzadanie to może zostać ponownie uruchomione na innym, sprawnym węźle. Realizacja zadań może wręcz z założenia przebiegać z pewną redundancją tak, aby ewentualne awarie pojedynczych węzłów nie powodowały wzrostu czasu odpowiedzi krytycznych aplikacji czasu rzeczywistego. Jeżeli do awarii nie dojdzie, wtedy nadmiarowe wyniki zostaną „zniweczone”.

Agregacji mogą podlegać nie tylko zasoby obliczeniowe, ale również zasoby dyskowe. Często serwery lub stacje robocze dysponują bardzo dużymi, w znacznym stopniu niewykorzystanymi, zasobami pamięci masowej. Połączenie tych zasobów w jeden duży wirtualny zbiornik danych, nazywany Data Grid, umożliwia zwykle uzyskanie nie tylko zwiększonej pojemności, ale również pozwala na poprawę wydajności (striping) i niezawodności (mirroring i replikacja). Jeżeli zadanie obliczeniowe uruchamiane na jednym z węzłów środowiska Grid Computing wymaga intensywnego dostępu do danych, to dane te mogą zostać uprzednio skopiowane do pamięci masowej lokalnej w stosunku do komputera, któremu zostanie powierzone to zadanie (automatyczna replikacja). Zabieg taki spowoduje nie tylko poprawę wydajności aplikacji, ale także częściowo zabezpieczy dane na wypadek awarii (zniszczenia lub niedostępności) danych źródłowych.

Praktyczne środowiska Grid Computing są zwykle konstruowane z trzech głównych typów komponentów:

•Stacje robocze (scavenging grids): zakłada się, że użytkownicy komputerów osobistych nie wykorzystują permanentnie ich pełnej mocy obliczeniowej, w związku z czym może ona zostać "zagospodarowana" na potrzeby rozproszonego przetwarzania danych; zadania obliczeniowe są najczęściej realizowane w formie wygaszaczy ekranu, aktywowanych przez system operacyjny w czasie nieaktywności użytkownika; aplikacje uruchamiane w środowisku zbudowanym z dużej liczby stacji roboczych cechują się bardzo wysokim stopniem równoległości.

•Serwery (computational grids): dedykowane środowiska połączonych komputerów o dużej mocy obliczeniowej umożliwiają efektywną realizację bardzo złożonych zadań przetwarzania danych; zadania obliczeniowe pochodzące ze środowiska Grid Computing nie muszą rywalizować o dostęp do procesora z zadaniami "lokalnymi"; istnieją możliwości udostępniania specjalizowanych urządzeń na potrzeby wybranych zadań obliczeniowych.

•Dane (data grids): współdzielenie danych jest koniecznym warunkiem realizacji przetwarzania Grid Computing; aplikacje korzystają z wirtualnej bazy danych lub wirtualnego systemu plików, których lokalizacje są przezroczyste dla aplikacji; stosowane są mechanizmy automatycznej replikacji lub transferu danych w celu podniesienia wydajności operacji wejścia-wyjścia

Istotną funkcją środowiska Grid Computing jest alokacja aplikacji lub zadań na dostępnych zasobach. W najprostszych systemach, odpowiedzialność przydziału komputera do konkretnego zadania spoczywa na użytkowniku, który uruchamia zadanie. Bardziej zaawansowane systemy posiadają dedykowany moduł rozdziału zadań (scheduler, resource broker), który automatycznie odnajduje węzeł najbardziej odpowiedni do realizacji powierzonego zadania. Moduły rozdziału zadań potrafią adaptować się do aktualnej dostępności węzłów w środowisku. W środowiskach Grid Computing konstruowanych ze stacji roboczych, aktywną stroną są stacje robocze w stanie bezczynności, które zgłaszają się do modułu rozdziału zadań i pobierają nowe zadanie do wykonania. Niestety, nagłe pojawienie się na takich stacjach aktywności "lokalnej" zwykle powoduje zawieszenie lub opóźnienie realizacji powierzonego zadania, a to przekłada się na nieprzewidywalność czasu odpowiedzi całej aplikacji.

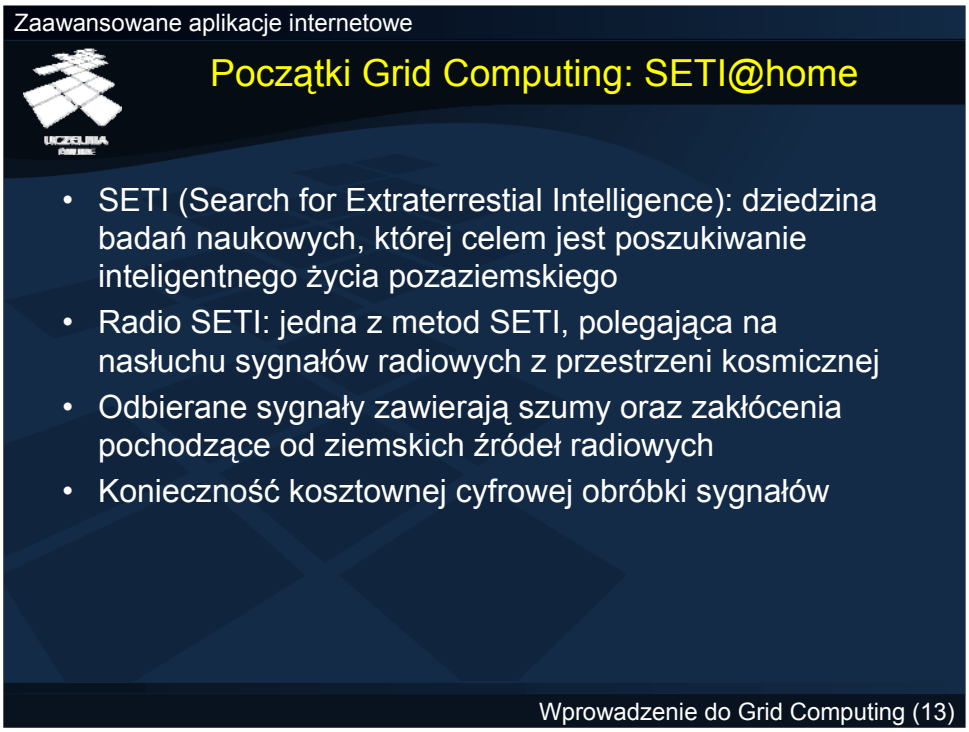
Poza dynamicznym rozdziałem zadań, środowisko Grid Computing powinno umożliwiać również wstępną rezerwację zasobów przez aplikację w celu zagwarantowania zadanej jakości usługi (quality of service). Przydział innych zadań do zasobów zarezerwowanych przez aplikację jest wówczas ograniczony i zwykle możliwy wyłącznie dla zadań o niskich priorytetach.

Problem dynamicznej alokacji zasobów omówimy z pomocą rysunku przedstawionego na slajdzie. Rysunek obrazuje grupę użytkowników wykorzystujących aplikacje pracujące w środowisku gridowym. Pięciu użytkowników jest przydzielonych do pierwszego komputera w sieci, trzech do drugiego. Załóżmy, że w tym momencie następuje awaria pierwszego komputera.

Awaria komputera powinna zostać wykryta przez oprogramowanie systemowe zajmujące się alokacją zasobów, które dokona automatycznej powtórnej alokacji, przydzielając użytkowników do innego komputera w sieci.

Po zakończeniu procesu realokacji zasobów użytkownicy mogą kontynuować swoją pracę. Dla większości z nich awaria pozostała niezauważona.

Należy zdawać sobie także sprawę z ewentualnych zagrożeń jakie stwarza wizja e-utilities w powiązaniu z mechanizmami dynamicznej realokacji zasobów. Może się zdarzyć, że po awarii jednego węzła sieci, jego dotychczasowe obciążenie jest realokowane na kolejny węzeł, który zostaje przeciążony i w wyniku tego ulega awarii. W tym momencie obciążenie z dwóch utraconych węzłów trafia na kolejny węzeł i doprowadza do kolejnej awarii wynikającej z przeciążenia. Taki kaskadowy proces mógłby doprowadzić do awarii całego systemu, analogicznej do spotykanych masowych awarii sieci energetycznych, np. w USA. W związku z tym niezwykle ważna jest jakość oprogramowania służącego do alokacji zasobów w sieci Grid, a w szczególności jego zdolność podtrzymywania stałego poziomu jakości i wydajności świadczonych usług obliczeniowych.



Jedną z pierwszych praktycznych realizacji wizji przetwarzania Grid Computing był projekt SETI@home, który umożliwił konstrukcję ogromnego, darmowego superkomputera dla prowadzenia obliczeń astronomicznych dla potrzeb SETI. SETI (Search for Extraterrestial Intelligence) jest dziedziną badań naukowych, której celem jest poszukiwanie inteligentnego życia pozaziemskiego. Jedno ze stosowanych w tej dziedzinie podejść, nazywane Radio SETI, wykorzystuje radioteleskopy astronomiczne do prowadzenia nasłuchu wąskiego pasma sygnałów radiowych pochodzących z przestrzeni kosmicznej. Odebranie takiego sygnału byłoby traktowane jako dowód istnienia technologii pozaziemskiej.

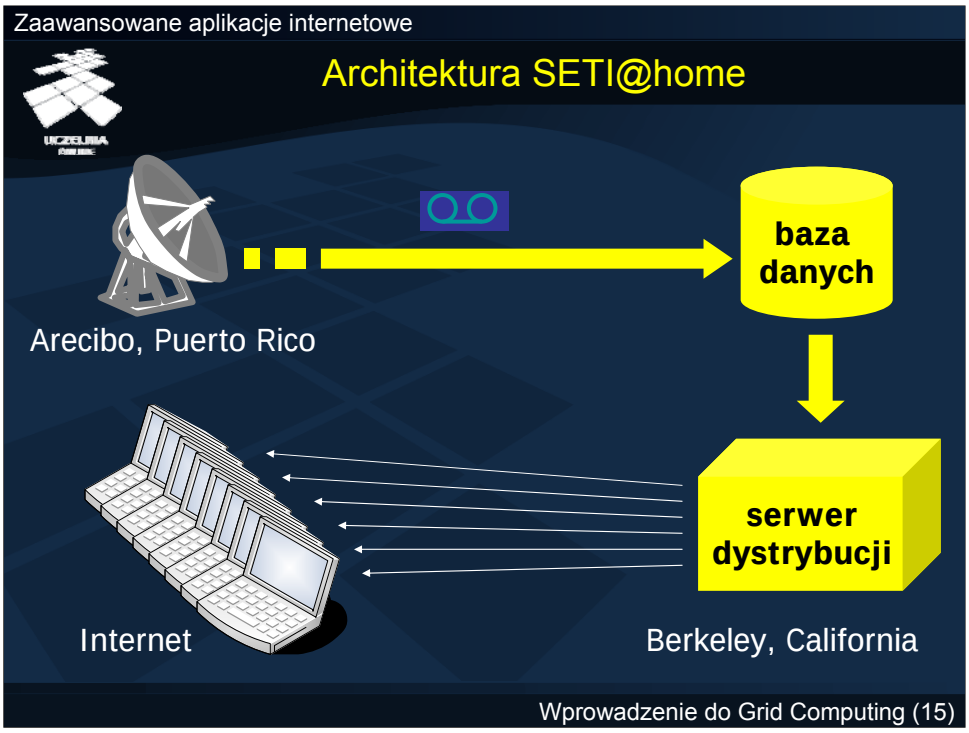
Niestety, radioteleskopy astronomiczne odbierają sygnały składające się w większości z szumów generowanych przez ziemskie urządzenia elektroniczne, a także z sygnałów telewizyjnych, satelitarnych i radarowych. W związku z tym konieczna jest cyfrowa analiza odbieranych sygnałów w celu wyeliminowania tych, które są pochodzenia ziemskiego. Analiza taka przebiega zwykle w trzech fazach: (1) dokonywany jest rozkład sygnału na częstotliwości składowe, (2) przy pomocy algorytmów rozpoznawania wzorców, znajdowane są sygnały-kandydaci do dalszej analizy, (3) eliminowane są te sygnały-kandydaci, które są najprawdopodobniej pochodzenia naturalnego lub ziemskiego. Cyfrowe przetwarzanie sygnałów odbieranych przez radioteleskopy jest bardzo kosztowne obliczeniowo, proporcjonalnie do szerokości i czułości nasłuchiwanego pasma radiowego oraz proporcjonalnie do stopnia pokrycia nieba przez radioteleskop. Powoduje to stale rosnące i niezaspokojone zapotrzebowanie na ogromną moc obliczeniową komputerów wykorzystywanych w projektach Radio SETI.

SETI@home

- Budowa wielkiego superkomputera wirtualnego, złożonego z dużej liczby pojedynczych maszyn połączonych siecią Internet
- Operatorzy pojedynczych maszyn uczestniczą w projekcie dobrowolnie i nieodpłatnie
- Wykorzystanie superkomputera wirtualnego do prowadzenia cyfrowej obróbki sygnałów radiowych odbieranych przez radioteleskop w Arecibo (Puerto Rico)

We wczesnych badaniach Radio SETI, do masowej analizy danych wykorzystywano superkomputery zlokalizowane w bliskości teleskopu. W roku 1995, David Gedye zaproponował realizację badań Radio SETI przy pomocy wirtualnego superkomputera zbudowanego z dużej liczby pojedynczych maszyn (głównie stacji roboczych) połączonych siecią Internet. Projekt został nazwany SETI@home, a dzięki jego spopularyzowaniu wśród użytkowników sieci Internet, do października roku 2003 dobrowolnie i nieodpłatnie zaangażowało się w niego ponad 4.5 miliona komputerów, realizując łącznie ponad 1.5 miliona lat obliczeń, odpowiadających realizacji 3.7×10^{21} operacji zmiennoprzecinkowych.

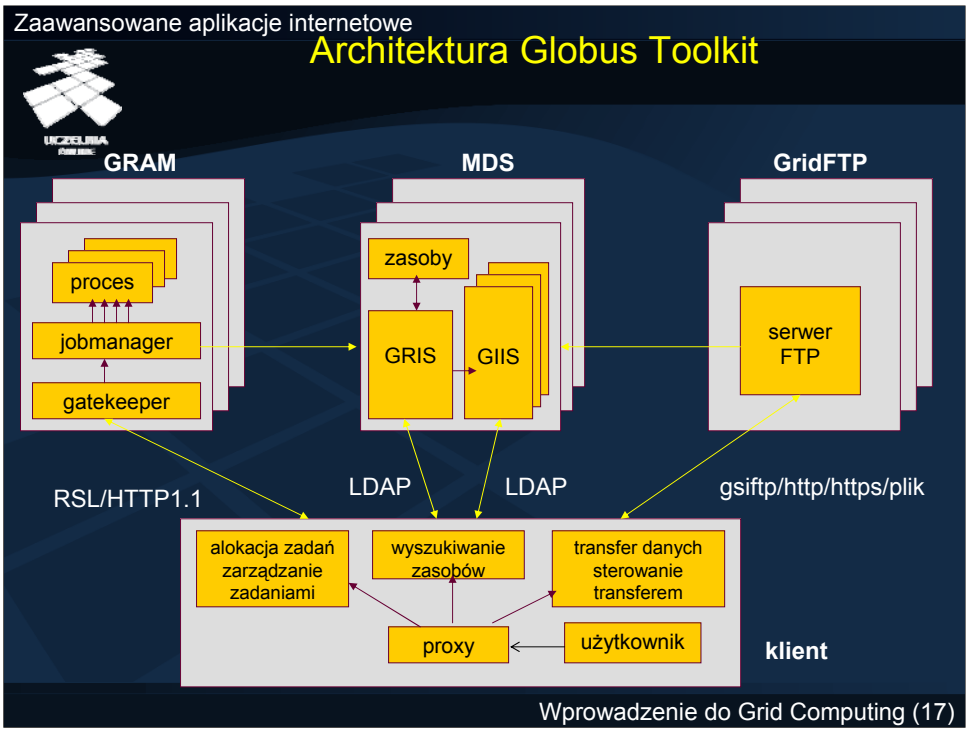
Fundament projektu SETI@home stanowi największy i najczulszy na świecie radioteleskop astronomiczny znajdujący się w Arecibo w Puerto Rico, którego antena pomocnicza jest źródłem danych pomiarowych przetwarzanych przez aplikacje SETI@home.



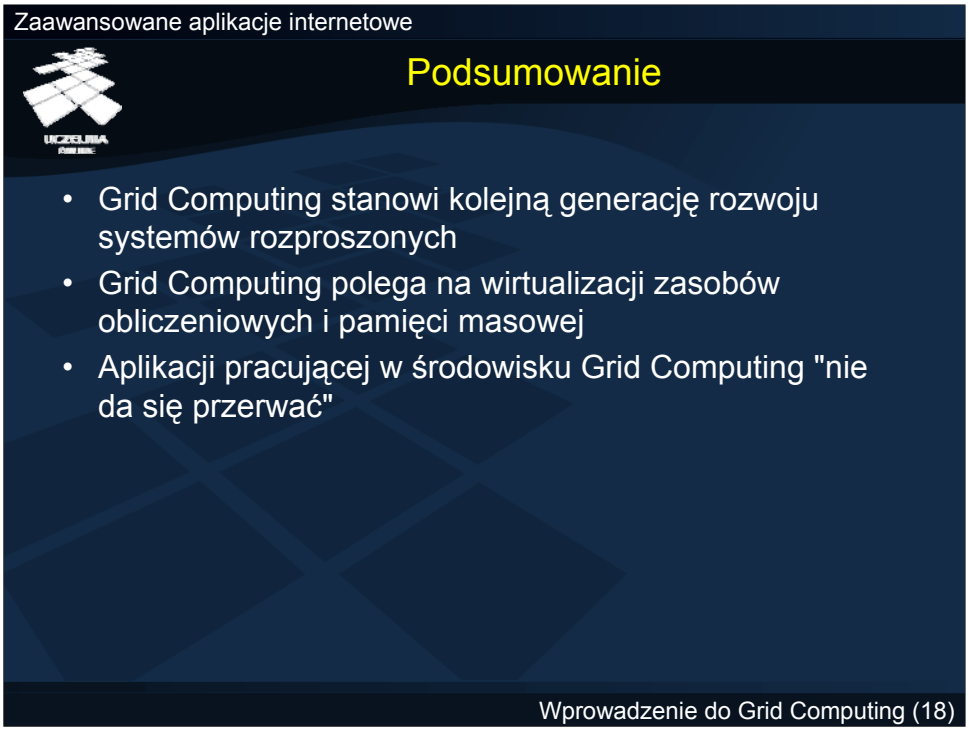
Dane źródłowe, zbierane z szybkością 5Mbps w paśmie 1.42 GHz (pasmo wodoru), są zapisywane na taśmach i transportowane do Uniwersytetu Berkeley w Kalifornii. Tam sygnał jest dzielony na fragmenty podpasma szerokości 1 kHz i długości 20s, nazywane jednostkami pracy (work unit), które są zapisywane w bazie danych, a następnie dystrybuowane wśród komputerów biorących udział w projekcie. Zainstalowana wcześniej na każdym z tych komputerów aplikacja klienta SETI@home pobiera jednostkę pracy, przetwarza ją, a wyniki odsyła przez Internet do Uniwersytetu Berkeley. Następnie aplikacja klienta może pobrać i przetwarzać kolejną jednostkę pracy. Aplikacje klientów nie komunikują się między sobą. Są zaimplementowane dla 175 platform systemowych w formie wygaszacza ekranu, procesu tła lub aplikacji graficznej. W przetwarzaniu jednostek pracy stosuje się także niewielką redundancję, w wyniku której pojedyncza jednostka jest przetwarzana dwa lub trzy razy przez różnych klientów. Celem redundancji jest wykrycie i wyeliminowanie wyników generowanych przez uszkodzonych lub fałszywych klientów.

Przetwarzanie sygnału przez aplikację klienta SETI@home ma na celu wykrycie jednego z czterech rodzajów wzorców: impulsów wzrostu mocy sygnału, wąskopasmowych sygnałów gaussowskich, sygnałów pulsujących lub potrójnych impulsów wzrostu mocy sygnału o tej samej częstotliwości (trypletów). Wzorce odkryte przez aplikacje klientów, po dostarczeniu do odbiorczej bazy danych Uniwersytetu Berkeley, poddawane są dalszej analizie, której celem jest eliminacja sygnałów pochodzenia ziemskiego. Przez wszystkie lata realizacji projektu nie znaleziono żadnego wzorca pochodzenia pozaziemskiego.

Rynek oferuje bardzo dużą liczbę narzędzi wspomagających konstrukcję, konfigurację oraz programowanie środowisk Grid Computing. Najpopularniejszymi z nich są: Globus Toolkit (www.globus.org), Oracle Globus Development Kit (www.oracle.com), IBM Grid Toolbox (www.ibm.com), Avaki (www.avaki.com), DataSynapse LiveCluster (www.datasynapse.com), Entropia DCGrid (www.entropia.com), Platform LSF, ActiveCluster oraz MultiCluster (www.platform.com), United Devices MetaProcessor Platform (www.ud.com). Z punktu widzenia funkcjonalności i liczby zastosowań, najbardziej godny uwagi jest pakiet Globus Toolkit.



Globus Toolkit zawiera trzy podstawowe komponenty do budowy środowisk Grid Computing: (1) moduł zarządzania zasobami, (2) moduł usług informacyjnych i (3) moduł zarządzania danymi. Moduł zarządzania zasobami odpowiada m.in. za alokację zadań do węzłów, zdalne uruchamianie zadań, zarządzanie stanem i postępem ich pracy. Zawiera wewnętrzne komponenty GRAM (Grid Resource Allocation Manager) i GASS (Global Access to Secondary Storage). Nie zawiera wbudowanego komponentu rozdzielania zadań, lecz oferuje interfejs pozwalający wykorzystywać zewnętrzny komponent odpowiedzialny za taką funkcjonalność. Żądania uruchomienia zadań (komenda globusrun), zapisane w języku RSL (Resource Specification Language), są przekazywane przez użytkowników do elementu Gatekeeper. Gatekeeper jest demonem służącym do autoryzacji użytkownika, powoływania procesu Jobmanager i przekazywania jemu dalszej komunikacji z klientem. Proces Jobmanager parsuje żądania wyrażone w języku RSL, alokuje zadania do lokalnych zasobów, a po zakończeniu zadań przekazuje klientom wyniki ich pracy. Moduł usług informacyjnych jest prostym serwerem LDAP, służącym do przechowywania informacji o strukturze całego środowiska Grid Computing (adresy IP, rozmiary pamięci, itp.) oraz statystyk opisujących obciążenie poszczególnych węzłów. Swoją funkcjonalność realizuje przy pomocy komponentów GRIS (Grid Resource Information Service) i GIIS (Grid Index Information Service), wspólnie nazywanych MDS (Monitoring and Discovery Service). Użytkownik uzyskuje dostęp do modułu usług informacyjnych przy pomocy komendy ldapsearch. Moduł zarządzania danymi oferuje mechanizmy transferu plików i zarządzania transferem plików pomiędzy węzłami Grid Computing. Jego kluczowym elementem jest komponent GridFTP, bazujący na standardowym FTP, lecz wzbogacający je o transmisję wielokanałową, automatyczne strojenie i zintegrowane mechanizmy bezpieczeństwa. Pełna funkcjonalność Globus Toolkit jest dostępna w formie komend wiersza poleceń oraz API dla popularnych języków programowania.



Przetwarzanie Grid Computing umożliwia efektywne wykorzystanie posiadanych zasobów komputerowych oraz pozwala na konstrukcję systemów o podwyższonej niezawodności. Dostępność gotowych, standardowych narzędzi programistycznych stanowi dla projektantów zachętę do rozważenia architektur Grid Computing w nowotworzonych systemach.

Materiały dodatkowe

- "Grid Computing - moc z sieci", www.linux-magazine.pl, Lipiec 2004
- "Introduction to Grid Computing with Globus", www.ibm.com/redbooks

•"Grid Computing - moc z sieci", www.linux-magazine.pl, Lipiec 2004
•"Introduction to Grid Computing with Globus", www.ibm.com/redbooks