

Wprowadzenie

Wykład prowadzą:

Jerzy Brzeziński
Jacek Kobusiński



Plan wykładu

- Wprowadzenie
- Podstawowe definicje
- Charakterystyka środowiska przetwarzania rozproszonego
- Przykłady środowisk
 - Internet
 - GRID
 - Przetwarzanie rozproszone ...@Home
 - Google



Cechy systemów rozproszonych

- duża wydajność
- duża efektywność inwestowania
- wysoka sprawność wykorzystania zasobów
- skalowalność
- wysoka niezawodność
- otwartość funkcjonalna



Problemy związane z konstrukcją systemów rozproszonych

- optymalne zrównoleglenie algorytmów przetwarzania
- ocena poprawności i efektywności algorytmów rozproszonych
- alokacja zasobów rozproszonych
- synchronizacja procesów
- ocena globalnego stanu przetwarzania
- realizacja zaawansowanych modeli przetwarzania
- niezawodność
- bezpieczeństwo



Motywy

- ogromne rzeczywiste zapotrzebowanie na systemy rozproszone
- dostępność środków technicznych i praktyczne możliwości realizacji systemów rozproszonych
- różnorodność otwartych problemów związanych z konstrukcją i zarządzaniem systemami rozproszonymi

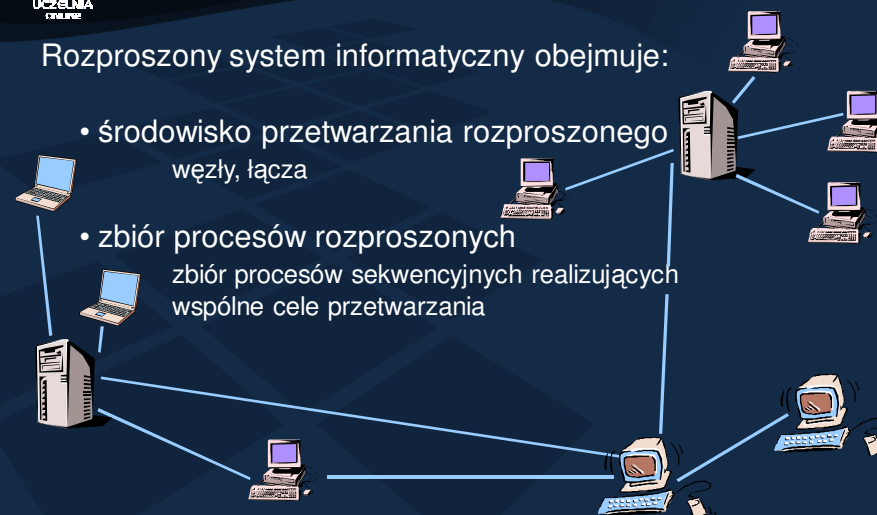
Wprowadzenie (5)



Rozproszony system informatyczny

Rozproszony system informatyczny obejmuje:

- środowisko przetwarzania rozproszonego
węzły, łącza
- zbiór procesów rozproszonych
zbiór procesów sekwencyjnych realizujących wspólne cele przetwarzania



Wprowadzenie (6)



Środowisko przetwarzania rozproszonego

Środowisko przetwarzania rozproszonego jest zbiorem $\mathcal{N}$ autonomicznych jednostek przetwarzających N_i (węzłów), zintegrowanych siecią komunikacyjną (środowiskiem komunikacyjnym, łączami komunikacyjnymi, łączami transmisyjnymi).

Wprowadzenie (7)



Komunikacja w środowisku przetwarzania rozproszonego

Komunikacja między węzłami możliwa jest tylko przez transmisję **pakietów informacji** (wiadomości, komunikatów) łączami komunikacyjnymi.

Wprowadzenie (8)



Zegary w środowisku przetwarzania rozproszonego

Jednostki przetwarzające realizują przetwarzanie z prędkością narzucaną przez **lokalne zegary**.

- zegary są **niezależne** → węzły działają **asynchronicznie**.
- zegary są **zsynchronizowane** lub istnieje **wspólny zegar globalny** dla wszystkich węzłów → węzły działają **synchronicznie**.



Węzeł

Jednostka przetwarzająca $N_i \in \mathcal{N}$ (węzeł) jest elementem środowiska przetwarzania rozproszonego obejmującym:

- procesor
- lokalną pamięć operacyjną
- interfejs komunikacyjny



Łącze komunikacyjne

Łącze komunikacyjne jest elementem umożliwiającym transmisję informacji między interfejsami odległych węzłów. Wyróżnia się łącza jedno- i dwu-kierunkowe.



Bufory łącza

Wyposażone są one w bufory o określonej pojemności (ang. *links capacity*).

Jeżeli łącze nie posiada buforów (jego pojemność jest równa zero), to mówimy o łączu **nie buforowanym**, w przeciwnym razie – o **buforowanym**.



Kolejność odbierania komunikatów

- **Łącze FIFO** – kolejność odbierania komunikatów wysyłanych z danego węzła jest zgodna z kolejnością ich wysłania
- **Łącze nonFIFO** – w przeciwnym przypadku



Niezawodność łącza

Łącza mogą gwarantować również, w sposób niewidoczny dla użytkownika, że żadna wiadomość nie jest tracona, duplikowana lub zmieniana – są to tzw. **łącza niezawodne** (ang. *reliable, lossless, duplicate free, error free, uncorrupted, no spurious*).



Czas transmisji

Czas transmisji w łączu niezawodnym (ang. *transmission delay, in-transit time*) może być ograniczony lub jedynie określony jako **skończony** lecz **nieprzewidywalny**.



Struktura środowiska przetwarzania

Często przedstawiana jako graf:

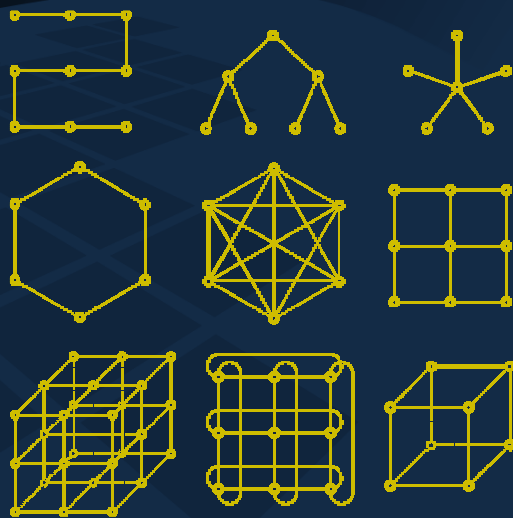
$$\mathcal{G} = \langle \mathcal{V}, \mathcal{A} \rangle$$

w którym :

- wierzchołki grafu $V_i \in \mathcal{V}$ reprezentują jednostki przetwarzające $N_i \in \mathcal{N}$,
- krawędzie $(V_i, V_j) \in \mathcal{A}$, $\mathcal{A} \subseteq \mathcal{V} \times \mathcal{V}$, grafu niezorientowanego lub łuki $\langle V_i, V_j \rangle \in \mathcal{A}$ grafu zorientowanego, reprezentują odpowiednio łącza dwu- lub jedno-kierunkowe.



Przykłady topologii



Wprowadzenie (17)



Skala przetwarzania

- Systemy końcowe (ang. *end systems*)
- Klastry (ang. *clusters*)
- Sieci intranetowe (ang. *intranets*)
- Internet

Wprowadzenie (18)



Klasy zastosowań

- Aplikacje wykorzystujące przetwarzanie rozproszone na wielu jednostkach obliczeniowych (ang. *distributed supercomputing*)
- Aplikacje wymagające dużej przepustowości (ang. *high throughput*)
- Aplikacje „na żądanie” (ang. *on demand*)
- Aplikacje intensywnie przetwarzające dane (ang. *data intensive*)
- Aplikacje umożliwiające współpracę (ang. *collaborative*)

Wprowadzenie (19)



TOP500

- Obecnie (czerwiec 2008) na I miejscu :
IBM - Roadrunner – 122400 CPU,
1026 TFlop/s,
98 TB RAM
- Prawo Moor’a na temat szybkości komputerów jest nadal prawdziwe !
- Ostatni komputer na liście ma moc przekraczającą 5 TFlop/s

Wprowadzenie (20)



Internet – środowisko rozproszone



Internet – liczba użytkowników sieci

Liczba użytkowników sieci Internet [mln]			
1995	2002	2005	? (2015)
45	445.9	1 080	2 000

Źródło: Computer Industry Almanac 01.2006



Internet – dostęp do sieci

Dostęp do sieci Internet wg krajów [mln osób]	
USA	197.8
Chiny	119.5
Japonia	86.3
Indie	50.6
Niemcy	46.3
Polska	10.6

Źródło: Computer Industry Almanac 01.2006



GRID

„A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities ”

/ Ian Foster /



GRID – cechy

dependable consistent pervasive inexpensive

usługa wiarygodna

użytkownicy żądają pewności, że otrzymają przewidywalny, nieprzerwany poziom wydajności dzięki różnym elementom tworzącym GRID.

usługa powszechnie dostępna (wszechobecna)

usługa zawsze powinna być dostępna, niezależnie od tego gdzie znajduje się użytkownik tej usługi.

usługa spójna

potrzebny jest standardowy serwis, dostępny poprzez standardowe interfejsy, pracujący ze standardowymi parametrami.

usługa relatywnie tania (opłacalna)

dostęp do usługi powinien być relatywnie tani, tak by korzystanie z takiej usługi było atrakcyjne także z ekonomicznego punktu widzenia.



Trzy spojrzenia na GRID

Użytkownik

Wirtualny komputer, który minimalizuje czas wykonania obliczeń oraz zapewnia dostęp do zasobów

Programista

Zestaw narzędzi i interfejsów zapewniających przezroczysty dostęp do danych

Administrator

Środowisko umożliwiające monitorowanie, administrowanie i bezpieczne używanie rozproszonych zasobów obliczeniowych, dyskowych oraz sieciowych



Czego GRID nie może ...

- naruszać bezpieczeństwa i autonomii poszczególnych jednostek wchodzących w jego skład
- powodować konfliktów w działaniu z istniejącym już oprogramowaniem
- narzucać użytkownikom języków programowania, narzędzi, bibliotek do programowania równoległego, itp.



Co GRID powinien ...

- umożliwiać rozproszenie geograficzne zasobów
- obsługiwać heterogeniczność sprzętową i programową
- być połączony poprzez heterogeniczną sieć
- korzystać z ogólnie dostępnych, standardowych protokołów i interfejsów
- być odporny na zawodny sprzęt
- pozwalać na dynamikę dostępu do sprzętu
- zrzeszać różne organizacje (wirtualne) z ich własnymi politykami bezpieczeństwa i dostępu do zasobów



Przykładowe projekty przetwarzania rozproszonego

- SETI
- Cure Cancer
- Fight Anthrax
- Prime Numbers
- Distributed.net
- GIMPS
- FreeDB.org
- The Internet Movie Database
- The Distributed Chess Project
- Wikipedia
- Dmoz – Open Directory Project
- ClimatePrediction.net
- Lifemapper



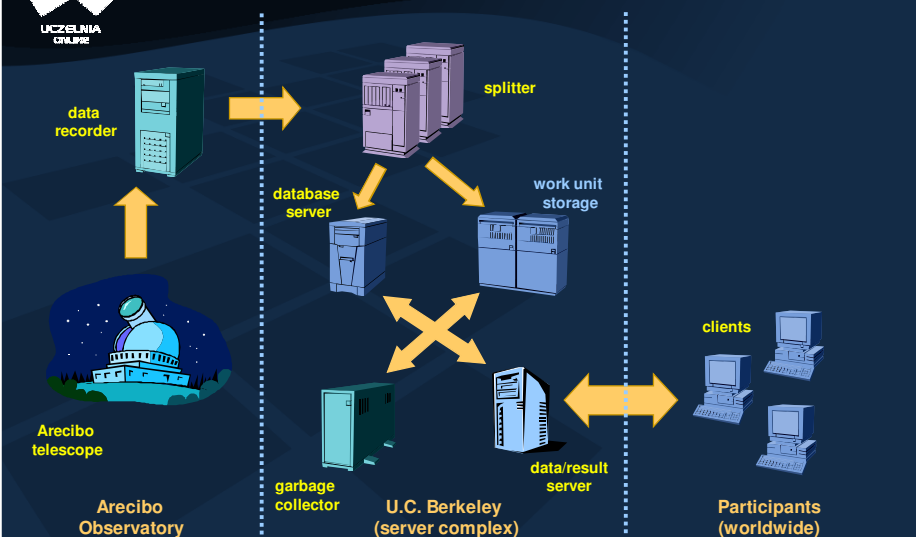
SETI@HOME „Czy jest tam ktoś ?”

Poszukiwania cywilizacji pozaziemskich
(ang. **S**earch for **E**xtra-**T**errestrial **I**ntelligence)

- 1998 - plany realizacji projektu SETI@home
400 000 chętnych wstępnie rejestruje się w ciągu następnego roku
- 1999 - pierwsza wersja oprogramowania
200 000 pobrań w ciągu tygodnia !!!
- Luty 2005 liczba uczestników – 5 361 313 osób, (z 226 krajów)
- Wykonano $6.523404e+21$ operacji zmiennoprzecinkowych
- Średnia moc (24h) - 250 TeraFLOPs/sec



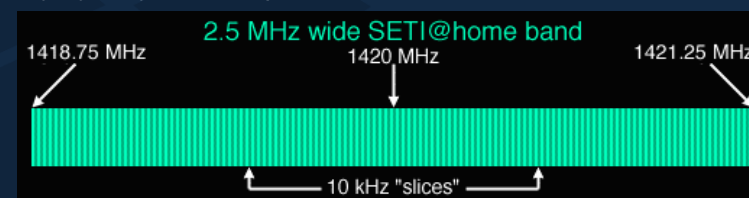
SETI@HOME : jak to działa ?



SETI@HOME – „Work-units”

Źródło [SETI@home](#) 28 lutego 2005

- Dane są dzielone na porcje po ok. 0,25 MB (tzn. „work-units”), które są wysyłane do uczestników.
[107 sekund, 10 kHz (20kbps) danych $\approx$ 2mln bitów czyli 0,25 MB]
- Dotychczas wygenerowano 218 553 918 próbek z danych uzyskanych za pomocą radioteleskopu w Arecibo. Dane te były przesłane na 1 344 taśmach, czyli średnio na jednej taśmie znajduje się 162 885 próbek.





- Koszty leczenia: 37 mld \$ bezpośrednio koszty medyczne + 11 mld \$ koszty pośrednie
- Dotychczasowe leki są niedoskonałe, działania uboczne, mało skuteczne

Znaleźć molekuły, które będą najbardziej obiecującym punktem startowym w poszukiwaniu leku na raka.

- I etap analizy został zakończony, trwa II etap

Łączny czas przetwarzania CPU (l:d:g:m:s)	219 998y 22d 10h 29m 28s
Zwrócono wyników	138 310 239
Liczba uczestników	918 238
Liczba komputerów	1 932 678



UCZELNIA
CZM. BPS

Vascular Endothelial Growth Factor	Superoxide Dismutase	Insulin Tyrosine Kinase	RAS Proteins
Cyclooxygenase	C-ABL Tyrosine Kinase	Fibroblast Growth Factor Receptor	CDK-2
RAF	Farnesyltransferase	Protein-Tyrosine- Phosphatase 1B	VEGFr1



Fight Anthrax

Zadanie: Przeanalizować i wykluczyć możliwie jak największą liczbę molekuł, które nie będą skutecznym lekiem do walki z węglikiem.

Wyniki: Z początkowej liczby molekuł (3,57 mld) udało się wyodrębnić ok. 300 000 cząstek, które warto analizować w kolejnych etapach badań. Całość została wykonana w ciągu 24 dni (!!!) przy założeniu 5-krotnej weryfikacji danych.

Łączny czas przetwarzania CPU (l:d:g:m:s)	6 989l 164d 06h 35m 38s
Zwrócono wyników	2 909 726

Wprowadzenie (36)



Google – początki

(Uniwersytet Stanford, Backrub project, 1998):

- Sun Ultra II Dual 200MHz, 256MB of RAM, 3 x 9GB HDD i 6 x 4GB HDD
- 2 x 300 MHz Dual Pentium II, 512MB RAM, 9 x 9GB HDD
- 8 x 9GB HDD (ofiarowany przez IBM).
- 10 x 9GB HDD (własnoręcznie wykonany)



Google – dzisiaj

- Ponad 450,000 serwerów (533MHz Intel Celeron – 1,4GHz Intel Pentium III)
- Jeden lub więcej 80GB HDD w każdym serwerze
- 2 – 4 GB RAM
- 5 farm serwerów (Kalifornia, Wirginia, Oregon), dokładnie dane nie są znane.
 - ok. 6000 procesorów
 - 12000 HDD
 - Połączenie ze światem OC-48 (2488Mbit/s)
 - Połączenia pomiędzy farmami OC-12 (622Mbit/s)



Google w liczbach ...

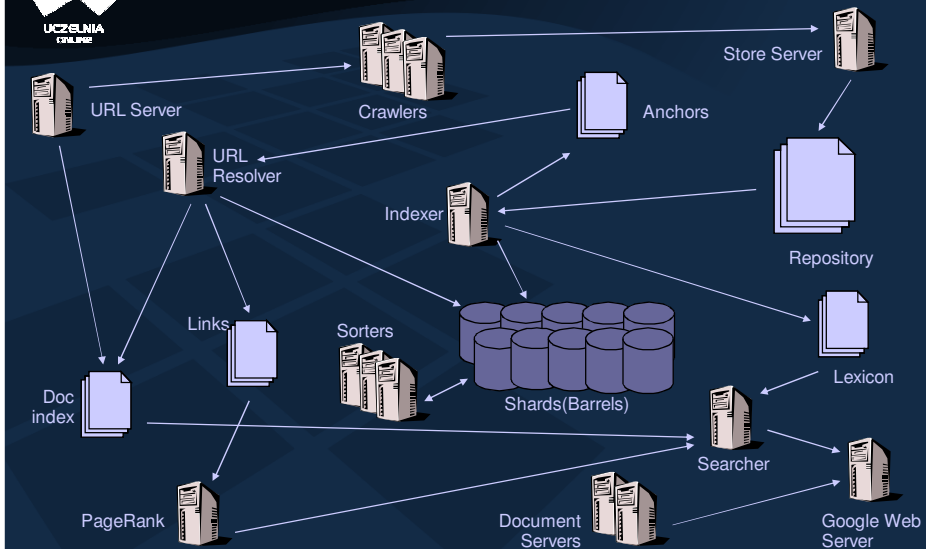
- 112 międzynarodowych domen
- 5680 pracowników zatrudnionych na pełen etat
- obsługuje 104 języki
- 4 mld stron x 10kB = 40 TB
- 380 mln wizyt dziennie
- 2,9 mld zapytań , średnio 1000 żądań/sek
- 100 dok/s jest pobieranych przez roboty wyszukujące
- ~300 Teraflops - moc obliczeniowa



Główne zadania

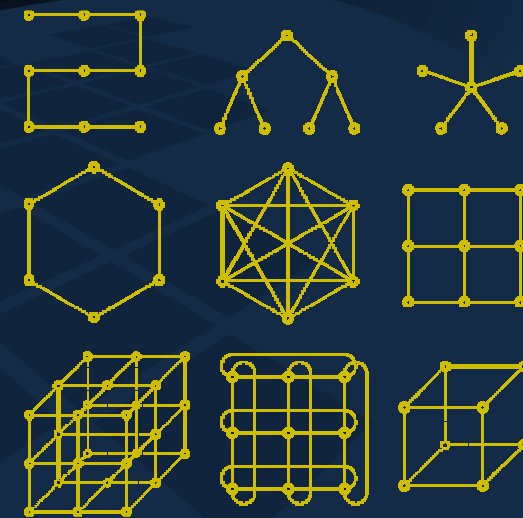
1. Aktualizacja zawartości
2. Indeksowanie przechowywanej zawartości
3. Obsługa żądań użytkowników

Schemat architektury Google



Wprowadzenie (41)

Przykłady topologii



Wprowadzenie (17)