

# Stochastyczne Algorytmy Obliczeniowe

## Zastosowanie algorytmu genetycznego do rozwiązywania problem replikacji danych w środowisku rozproszonym

Andrzej Kaczmarczyk

Marcin Łoś

### 1 Wstęp

Celem projektu było wybranie jednego problemu obliczeniowego, zapoznanie się z istniejącymi jego rozwiązaniami, oraz próba ulepszenia któregoś z nich. Nasz wybór padł na problem replikacji danych w systemach rozproszonych, oraz rozwiązanie oparte na algorytmie genetycznym, opisane w [1].

### 2 Opis problemu

Dany jest system rozproszony, składający się z  $M$  hostów  $\{H_i\}$ , o pojemnościach  $s_i$ , połączonych siecią komunikacyjną tak, że koszt przesyłu jednostki danych pomiędzy  $H_i$  i  $H_j$  wynosi  $C(i, j)$ . Istnieje także  $N$  obiektów  $\{O_i\}$ , o rozmiarach  $o_i$ . Host  $H_i$  wykonuje na obiekcie  $O_k$  odpowiednio  $r_k^{(i)}$  operacji odczytu, i  $w_k^{(i)}$  operacji zapisu. Każdy obiekt  $O_i$  znajduje się pierwotnie na jednym hoście,  $SP_i$ .

Obiekty mogą zostać zreplikowane na inne hosty, tak, że suma rozmiarów obiektów zreplikowanych na hoście  $H_i$  nie przekracza jego pojemności  $s_i$ . Wszystkie hosty mają pełną wiedzę o replikach obiektów. Operacja odczytu obiektu  $O_k$  z hosta  $H_i$  przebiega w ten sposób, że obiekt jest wysyłany do hosta  $H_i$  z najbliższego mu hosta zawierającego replikę  $O_k$ . Operacja zapisu natomiast odbywa się w ten sposób, że host  $H_i$  wysyła nowy stan obiektu do hosta  $SP_k$  (pierwotnego miejsca zawierającego  $O_k$ ), a ten rozsyła informację o zmianie do pozostałych hostów zawierających repliki  $O_k$ .

Koszt sumaryczny przy danym rozłożeniu replik to suma kosztów przesyłania obiektów spowodowanego operacjami odczytu i zapisu, przebiegającymi w opisany powyżej sposób. Koszt pojedynczego przesyłu to iloczyn ilości przesyłanych danych (rozmiaru obiektu) i kosztu jednostkowego przesyłu między hostami (danego przez  $C(i, j)$ ). Problem polega na znalezieniu replikacji minimalizującej koszt sumaryczny.

Nieco bardziej szczegółowy opis, wraz z wzorami na całkowity koszt znaleźć można w [1].

### **3 Istniejące rozwiązania**

### **4 Rozwiązanie bazowe**

### **5 Wprowadzone zmiany**

### **6 Podsumowanie**

### **Literatura**

- [1] T. Loukopoulos I. Ahmad. *Static and Adaptive Data Replication Algorithms for Fast Information Access in Large Distributed Systems*