

Podhorszki Norbert és Vida Gábor
MTA SZTAKI, Párhuzamos és Elosztott Rendszerek Laboratórium

Alkalmazói programozási felület SETI-jellegű elosztott programokhoz és végrehajtó rendszer a BOINC infrastruktúrára

Az utóbbi években egyértelműen a legnagyobb méretű és kapacitású számítógép-rendszereket az Interneten keresztül összekapcsolódott személyi számítógépekből alakították ki. A nagy számú PC-k által alkotott elosztott rendszer (Desktop Grid) több teraflops teljesítményű erőforrást jelent olyan alkalmazások számára, amelyeknek megfelel ez a fajta felépítés. A Desktop Grid-ben a PC-k egymással nem kommunikálhatnak és így a végrehajtható programok is csak szekvenciálisak lehetnek (kivétel lehet esetleg a duálprocesszoros gépekre optimalizált többszálú program). A nagy számítási kapacitást így csak független részfeladatokra szétdarabolható problémák megoldására lehet használni.

A legjobb Desktop Grid példa a SETI@home, ahol az alkalmazás maga egy sok évig futó adatfeldolgozó projekt és a napi 35 GB adatot 350 KB-os csomagokban dolgozzák fel a PC-k. A SETI@home egy tudományos kísérlet, amely több millió Internetre kapcsolt számítógép kapacitását hasznosítja a Földönkívüli Intelligencia Kutatásában (SETI, Search for Extraterrestrial Intelligence). A SETI az egész kutatási terület átfogó neve, ennek egy technikai megközelítése a rádióhullámok vizsgálata, egy keskeny frekvenciatartományban sugárzott intelligens adás keresése. Az adatok feldolgozása lényegében egymástól független: két különböző égbolt tartomány jelei függetlenek és az egy helyről érkező különböző frekvenciájú jelek is a keresés szempontjából függetlennek tekinthetők.

A nyílt forrású BOINC-et a SETI@home fejlesztő csapata tervezi és implementálja azzal a céllal, hogy a SETI jellegű és a nagyközönség számára attraktív tudományos projektek nagy számítási igényű elosztott programjainak egy közös infrastruktúrát biztosítson. Ezzel lehetővé válik, hogy több millió felhasználó számítógépén ugyanaz a szoftver fusson, és minden egyes felhasználó eldönthesse, mely projektekben óhajt részt venni anélkül, hogy különféle szoftvereket kellene installálnia és letöltenie. A BOINC kliensszoftvert futtató PC-k többmilliós hálózata így a világ legerősebb szuperszámítógépévé válhat, amelyen már nem csak egyetlen alkalmazás futhat. A BOINC rendszer felállítása és használata felettébb kényelmetlen. Hozzáértő rendszergazdáknak kell telepítenie és konfigurálnia a szervereket. A programozáshoz pedig nem ad támogatást a BOINC, ugyanis a SETI fejlesztői elsősorban a maguk igényét akarták kielégíteni vele. Más programok fejlesztői nagy lendülettel és kitartással kénytelenek saját alkalmazásaikat a BOINC igényeihez és sajátosságaihoz szabni.

Az MTA SZTAKI-ban fejlesztett DC-API (Distributed Computing API) egy egyszerű elosztott programfejlesztési felületet definiál a SETI-hez hasonló ún. Master-Worker jellegű alkalmazások fejlesztésére. Az API elrejtje az aktuális végrehajtó rendszer jellegzetességeit és a programozó csak a feladat megoldására koncentrál, azaz részfeladatokat hoz létre és részeredményeket dolgoz fel. A végrehajtás megszervezése az API implementációjának dolga. Ezzel nagymértékben megkönnyíthető a szekvenciális programozáshoz szokott programozók dolga, valamint a kész program sem specifikus,

hanem végrehajtható minden olyan elosztott rendszeren (Desktop Grid, másféle Grid, klaszterek stb.) amelyhez az API-t illesztettük.