

Vajda Ferenc

Tudományos kutatás és együttműködés informatikai bázison

„A jövőt egyszerűbb megteremteni, mint megjósolni.”
Alon Key

Bevezetés

Az elmúlt időben a tudomány jelentős változáson ment keresztül. Bár különböző tudományterületek és ágak jellemzői különbözőek, de általánosságban megállapíthatjuk, hogy a tudósok egyre bonyolultabb problémákkal foglalkoznak és a szükséges eszközök és berendezések is egyre költségesebbek. A tudományos kutatás folyamata nagyszámú technikai, szociális és eljárásokkal kapcsolatos tevékenységet foglal magában, azonban ezek mindegyike információval kapcsolatos. Információt gyűjtenek, analizálnak, kapcsolnak össze egymással, vannak le következtetéseket belőlük, illetve osztanak szét. Ezen tevékenységek szinte mindegyike számítógépeket, és újabban, számítógép-hálózatokat igényel.

A tudományos kutatás rendkívül gyorsan fejlődött az elmúlt időszakban. Ennek illusztrálására álljon itt csak két szám. A tudományos kutatással kapcsolatos információmennyiség jelenleg kb. 12 év alatt megduplázódik és legalább 90 százaléka a valaha is élt tudósoknak még most is életben van.¹ A kutatás fejlődése két — egymásnak látszólag ellentmondó — irányban halad. Egyrészt a bonyolultság specializálódáshoz vezetett (a tudós egyre többet tud egyre kevesebbről), másrészt a problémák komplexitása ugyanakkor előtérbe hozta az interdiszciplinaritást (a tudós egyre kevesebbet tud egyre többről).

Tudományos kutatók voltak a számítógépek első használói és a számítástechnika fejlődésének az igények oldaláról meghatározó tényezői. A kommunikáció és számítástechnika bázisán kialakított hálózatok (lokális és globális, intra, extra és inter) is gyorsan a tudományos kutatás infrastruktúrája integrált részeivé váltak. A személyes kommunikációt megkönnyítő elektronikus posta (E-mail) után az elosztott adatbázisokon való keresést, a rendkívül leegyszerűsített böngészést lehetővé tevő világháló (World Wide Web) is a kutatás igényeinek kielégítésére és aktív közreműködésével (a CERN kezdeményezésére) jött létre. A kutatás igényei miatt alakultak ki valamilyen szempontból azonos érdeklődésűek levelező listái vagy hirdetőtáblái (Bulletin boards). Ugyancsak a hálózat bázisán folynak ma már a publikálással kapcsolatos különböző tevékenységek. Így a szöveg, illetve a hozzá tartozó ábrák, táblázatok, képletek stb. előállítását nagymértékben megkönnyítő komplex (szö-

vegyszerkesztők, a teljes információt egységesen kezelő kódolási módok (például ún. postscript kódolás), a kiadvány-szerkesztés, bírálat, végleges nyomdakész (camera ready) változat előállítás, illetve a kiadvány vagy egyedi cikkek, absztraktok szétosztása vagy az ahhoz való hozzáférés biztosítása. Ugyanez érvényes a folyóiratokban megjelenő cikkekre, amelyek a hálózat segítségével elérhető tartalomjegyzék alapján a hálózaton keresztül megrendelhetők. A hivatkozásokra vonatkozóan álljon itt egy példa: az IEEE Internet Computing normál folyóirat 1997 március-áprilisi számában megjelent cikk (38–46. oldal) szöveggörnyezetben kilenc helyen hivatkozik URL (Universal Resource Locator), ill. http (Hyper Text Transfer Protocol) címre, míg a hivatkozási jegyzék 16 tétele közül 10 a hálózaton keresztül is elérhető. Az információ összekapcsolását biztosító hypertext szövegek hálózaton való átvitelét a HTML nyelv (Hypertext Markup Language) definiálja.

Az előbbieken érintett alapvető és ma már általánosan használt hálózati alkalmazásokon túlmenően olyan megoldásokat kívánunk ebben a rövid ismertetésben bemutatni, amelyek a tudományos kutatás speciális infrastrukturális hátterét alkotják és fogják még kiterjedtebben alkotni a jövőben.

A tudományos együttműködés új környezete: a kollaboratórium

A tudományos kutatás által megoldandó feladatok egyre növekvő bonyolultsága szükségszerűen felértékelte a kutatók együttműködésének jelentőségét. Az együttműködés környezetét, amely a kutatás eszközeinek és technológiáinak közös használatát biztosítja — függetlenül azok fizikai, földrajzi elhelyezkedésétől — egy új fogalommal, a kollaboratóriummal szokták újabban azonosítani. Maga a kifejezés (eredeti formájában: collaboratory) az együttműködés (kollaboráció — collaboration) és a laboratórium (laboratory) szavak összekapcsolásából származik.² Jelentése egy olyan falak nélküli centrum, amely földrajzi elhelyezkedéstől függetlenül lehetővé teszi az együttműködő kutatók számára, hogy kollégáikkal — párbeszéd formájában — kapcsolatba kerüljenek, berendezésekhez férjenek hozzá, számítógépes erőforrásokat és adatokat megosztva használjanak és digitális könyvtárak információit elérhessék. Maga a koncepció annyiban új az eddigi alapvetően hálózati bázisú lehetőségekhez képest, hogy a számítógéppel segített együttműködés egy új környezetet biztosítja. A kollaboratóriumok által nyújtott lehetőségek a következő csoportokba sorolhatók:

- Adatok megosztott használata, amely lehetővé teszi a különböző helyeken, egy közös feladaton dolgozó kutatóknak, hogy gyorsan és egyszerűen férjenek hozzá adatokhoz mind egy adott adatbázisban, mind adatbázisokon keresztül.

- Programok megosztott használata, amely egymástól távol dolgozó kutatók számára lehetővé teszi adatok analizálását, megjelenítését, modellezését stb., illetve az ezeket biztosító szoftverekhez való kényelmes hozzáférést.

- Távoli berendezések vezérlése, amely nem csupán az adott helyen rendelkezésre nem álló berendezések, hanem nehezen elérhető helyeken (például a Földön nehezen megközelíthető területen vagy a világűrben) levő berendezések használatát is biztosítja.

- Távolban levő kollégákkal való kommunikáció, ahol az időbeli (időzónák miatti) eltérés sokszor nagyobb nehézséget okoz, mint a térbeli távolság.

Az említett négy feladatcsoportnak megfelelő szolgáltatást nyújtó kollaboratóriumok kialakítása számos technikai és ezen túlmenő, de nem elhanyagolható fontosságú szociológiai és szervezési feladat megoldását is kívánja. Ezek szisztematikus felsorolása helyett itt csupán néhány jellemző problémára térünk ki.

Egymással összekapcsolt adatbázisok felépítését és használatát nagymértékben megkönnyítette a szabványos irányba fejlődő világhálózat (World Wide Web). A meg-

osztott programhasználat technológiája már régebben kialakult a csoportos használat eszközeinek formájában (például: groupware). Természetesen ezen programok nyújtotta funkciók dokumentálása, implementálása és működtetése a speciális követelményeknek kell, hogy megfeleljen. Távoli berendezések használata nemcsak a hálózat sebességére vonatkozóan szab meg korlátokat, hanem a távoli vezérlést biztosító telemetria szabványosítása terén is követelményeket támaszt. A kutatók kapcsolatának hagyományos eszközei mellett egyre nagyobb szerepet kapnak a különböző multimédia bázisú eszközök vagy a videokonferencia hálózat-alapú megoldásai. A csomagkapcsolt hálózat ma már lehetővé teszi mind audio-, mind videókapcsolat létesítését. Az interaktív videó jelentős szerepet játszhat kísérletek távoli megfigyelésében is. A kapcsolat biztosítására olyan eszközöket is felhasználnak, mint a megosztott képernyő (beszélgetés: talk üzemmód), vagy az elektronikus fehér tábla.

A kollaboratórium számos területen sikeresen alkalmazásra került. (Lásd például az IEEE Computer Graphics and Application folyóirat 1997 március-áprilisi számát: megosztott, három dimenziós virtuális környezetről vagy multimédia típusú alkalmazásokról³). A kollaboratóriumok gyakorlati megvalósítása napról napra fejlődik és maga a megoldás a jövő számára is újabb lehetőségeket és alkalmazási területeket kínál. Az amerikai National Academy of Science szervezett formában foglalkozik a nemzeti kollaboratóriumok kialakításával és már jelentős gyakorlati eredményekkel is rendelkezik, például a molekuláris biológia területén.

Hálózati számítástechnika, munkaállomás farmok és szuperszámítástechnika hálózati bázison

A különböző szintű hálózatok (lokális-globális, intranet-supernet-internet) megjelenésével és nagymértékű elterjedésével az elosztott (distributed) számítástechnika is egyre elterjedtebbé vált. A szlogenek, mint például a számítástechnika és kommunikáció „házassága” vagy a „hálózat maga a számítógép” jól jellemzik ezt a folyamatot.

Hogy lehet egy adott munkát gyorsabban elvégezni? Vagy keményebben kell dolgozni, vagy ötletesebben, illetve kérhetjük, hogy valaki segítsen. A három lehetséges megoldás erre a számítástechnikában a processzor sebességének növelése, jobb algoritmusok használata, illetve a párhuzamos feldolgozás.

A számítógép farmok (klaszterek) tulajdonképpen egy régen ismert paradigma — az elosztott feldolgozás (distributed processing) — új formában való megjelenése. Míg a párhuzamos feldolgozás hagyományos eszközei (és így a szuperszámítógépek is) azonos típusú processzorokkal működnek (homogének és szimmetrikusak), a farmok különböző típusú számítógépek (munkaállomások) együttműködését biztosítják (heterogének). A számítógép (munkaállomás) farmok, a különböző földrajzi helyen lévő, különböző csoportok vagy egyének által használt, különböző gyártótól származó, különböző szoftverrel felszerelt gépek együttes alkalmazását teszik lehetővé egy adott feladat megoldása során. A farmok a nagy számban használt és nagysebességű hálózattal összekötött munkaállomások közös használatát különböző programozási modelleken alapuló, erre a célra kifejlesztett támogató szoftverrendszerekkel (például MPI: Message Passing Interface) teszik lehetővé mind tudományos, mind adatfeldolgozási feladatok megoldására.

A hálózatcentrikus számítástechnika egyik új gyermeke a Java nyelv. Először a Java nyelven alapuló virtuális számítógépek a hálózaton való navigálást biztosították a (különböző, heterogén platformokon) megoldandó feladatok programozási nyelveként. Az együttes feladatmegoldást és programozást lehetővé tevő szoftver (groupware: csoport-szoftver) mintájára kialakított új terminológia a több felhasználó együttes, együttműködő munkáját lehetővé tevő szoftverfejlesztési környezet. A Java

ezen a téren is ígéretesnek bizonyult, biztosítva az együttműködés átlátszóságát, ami lehetővé teszi a közös (együttműködő) használatot olyan alkalmazásoknál, amelyek eredetileg csak egyetlen felhasználó számára fejlesztettek ki. A szakmai területen már megszokott terminológiai burjánzást egy további kifejezés, a közbülső — az elosztott rendszer és az alkalmazások között megfelelő támogatást nyújtó — szoftver, az ún. middleware is demonstrálja.

Az Internet a maga sok százezres központi (host) számítógépével és sok milliós terminál számítógépével rendkívül nagy számítástechnikai erőforrást reprezentál. A legnagyobb szuperszámítógépek teljesítménye is csak eltörpül ezen számítási kapacitás mellett. A hálózati alapú szuper számítástechnika merész tervekkel³ és már gyakorlati eredményekkel is büszkélkedhet.

A hagyományos szuperszámítástechnika új tagja az amerikai Energiaügyi Minisztérium (DOE) által létrehozott rendszer, amely 9200, egyenként több, mint 100 millió lebegőpontos művelet (megaFLOP) másodpercenkénti végzésére képes processzor összehangolásával valósult meg (Accelerated Strategic Computing Initiative). Az amerikai HPCC: High Performance Computing and Communication⁵ (Nagyteljesítményű számítástechnikai és kommunikáció) program, illetve az elnök és alelnök közvetlen támogatása rendkívül széles körű és nagyjelentőségű eredményeket ígér.

Kiszámítástudomány (Computing Science)

Ezen szakterület pontos magyar megnevezése és szabatos definíciója még várat magára. Maga a megnevezés számítási módszereket jelent, amelyek olyan kísérletek elvégzését teszik lehetővé, amiket nem lehet ténylegesen elvégezni, vagy pedig túl költségesek lennének. Más szóval, a számítógépes modellezés és a folyamatok szimulációjának egy jól definiált feladatsoportjáról van szó. A kiszámítástudomány lényegében négy alapvető szakterületre épül. Ezek közül az első a környezet, amely a kutatás általános módszerét jelenti. Maga a módszer egyszerűsített formában a következő módon értelmezhető:

- Fizikai alapelvek alapján egy M modellt származtatnak (a modell természetesen előzőleg létrehozott részmodelleket is tartalmazhat).

- Az M modellt numerikus módszerekkel többféle számítási modellé (C_i) fejlesztik tovább.

- A C_i számítási modellek szolgálnak a kísérletek alapjául. Különböző számítási eszközök és a vizualizációs eljárások segítségével végzik az M modell jellemzőinek feltárását és érvényességének meghatározását. Itt van lehetőség különböző kísérleti eredményekkel való összevetésre is.

- Egy ponton a C_i számítási modellek a vizsgált rendszer fizikai tulajdonságairól is információt szolgáltatnak. Az M modellt ezután a kiszámítástudományi kísérlet eredményei alapján lehet tovább finomítani.

Az ismertetett módszer hatékony alkalmazásának alapfeltétele, hogy a modell megfogalmazásánál a számítási szempontokat is figyelembe kell venni (véges aritmetika, numerikus algoritmusok, architektúra, programszerkezet stb.).

A második terület a modellezés módszereire koncentrál. Az önmagában is összetett részterület ismertetése helyett csak a legfontosabb alapelveket említjük. Ezek a lehető fizikai pontosság és a kiszámíthatóság, illetve az előzetes hibabeccslés és annak végső ellenőrzése.

A harmadik részterület a tudományos alkalmazás és az algoritmusok és architektúrák kapcsolatát definiálja. Elosztott rendszerek hatékony alkalmazásának feltétele hatékony elosztott algoritmusok használata.

³ Lásd például a petaflops tanulmányt.⁴ A FLOP: Floating Point Operation — lebegőpontos művelet — a számítógépek teljesítményének szokványos mérőszáma, és a petaFLOP teljesítményű számítógép 1 millió milliárd, vagyis 10^{15} lebegőpontos műveletet fog másodpercenként elvégezni.

A negyedik részterület a módszer helyességével, illetve verifikációjával kapcsolatos problémákkal foglalkozik.

A számítógépes szimuláció nagyon sok területen bizonyult rendkívül termékenynek. A kísérlet és az elmélet kölcsönhatása a modern tudomány elismert jellemzője. A szimuláció, a modellezés érdekes eredményeket szolgáltathat a fizikai folyamatokról. A kiszámítástudomány fizikai rendszerek számítási szempontból végrehajtható modelljeire koncentrálnak, algoritmusokat dolgoz ki a modellel kapcsolatos problémák megoldására. Nagyon fontos a számítógép-architektúrákhoz illeszkedő algoritmusok megválasztása és olyan környezet és eszköztár biztosítása, amely az alacsony szintű programozástól mentesíti a tudósokat és mérnököket, figyelembe véve a természetszerűleg interdiszciplináris jellegét. A modellezésnél alapvető fontosságú a fizikai korrektség, a kiszámíthatóság és a hibák korlátossága. Az eredmény megjelenítése a megoldás fontos, integráns része.

A számítástudomány (computer science) a kiszámítástudomány természetes háttere, azonban közvetlenül nem biztosítja annak minden szempontból megfelelő támogatását. Maga az oktatás, illetve annak tematikája is ludas ebben, ahol a feladatok általában más formában és háttérrel vannak megfogalmazva. A két terület közvetlenebb kölcsönhatására van szükség.

Tanulás—tanítás és a hálózat

A kutatás és az oktatás szoros kapcsolata indokolja, hogy néhány oktatással kapcsolatos problémáról is szójunk. A távoktatás eszközei és módszerei hosszú ideje az érdeklődés homlokterében vannak. Különösen fontos szerepet játszhat a távoktatás azon szakmai területeken, ahol a szakmai tudnivalók gyors változása, fejlődése napi szükségletté teszi az élethosszig tartó tanulást, továbbképzést. A távoktatás hagyományos eszközei az egyszerű levelező oktatástól a tv-adással és audió válaszadással rendelkező rendszereken át egészen a specializált videokonferencia eszköztáráig terjednek. A legutóbbi időszak fejlődése új lehetőségeket nyitott a távoktatás számára a hálózat (világhálózat) felhasználásával. Hozzájárult ehhez első sorban a kommunikáció sebességének elért és várható fejlődése (különösen az ún. ATM: Asynchronous Transfer Mode alkalmazása), a korszerű processzorokkal és speciális áramkörökkel támogatott multimédia (nagy felbontású képek megjelenítése, több irányú audió és videó szolgáltatás), valamint az elektronikus könyvtárakhoz való egyszerű és hatékony hozzáférés. Ezek tették lehetővé az ún. virtuális osztályterem megteremtését, ahol a korszerű eszközöket a hallgatók megosztva használhatják. A távoktatás új lehetőségei, korszerű eszközei és módszerei napról napra fejlődnek és maga a módszer egyre növekvő jelentőséggel bír.

HIVATKOZÁSOK:

- 1 The Unpredictable Certainty: Information Infrastructure Through 2000. National Academy Press, Washington D.C., 1996. <http://www.nap.edu/readingroom/books/unpredictable/>
- 2 National Collaboratories, National Academy of Sciences, 1996: <http://www2.nap.edu/htbin/docpage/>
- 3 W. Lanotte et al.: Visinet: Collaborative 3D Visualization and VR over ATM Networks. IEEE Computer Graphics and Applications, March–April 1997, pp. 66–75.
- 4 G.C. Fox and W. Furmanski: PETAFLIPS and EXAFLIPS: Supercomputing on the WEB. IEEE Internet Computing, March–April 1997, pp. 38–46. <http://computer.org/Internet/>
- 5 From desktop to Terallop: Exploiting the U.S. Lead in High Performance Computing NSF Report, August 1997.