

A GRID technológia és kutatás hazai és nemzetközi eredményei

Dr. Szeberényi Imre, szebi@ik.bme.hu
BME IIT, BME IK

1. Bevezetés

A globális méretű elosztott informatikai rendszerek kialakulását a robbanásszerűen fejlődő adatátviteli megoldások tették ill. teszik lehetővé. Ezen rendszerekben megoldhatóvá válik az informatikai erőforrások (pl. számítási, tárolási kapacitás) és szolgáltatások (pl. kereső ill. feldolgozó szolgáltatások) igény szerinti felhasználása, ami a jelenleginél jóval hatékonyabb erőforrás-gazdálkodáshoz vezethet. Kezdetben elsősorban a tudományos kutatási feladatok megoldása során jelentkeztek az erőforrások megosztására, hatékonyabb kihasználására irányuló igények. Ennek oka az, hogy a kutatási feladatok megoldása gyakran kooperációban történik, melyben természetesnek és logikusnak tűnik a kutatóhelyeken elosztottan megjelenő erőforrások hatékony kihasználása. Ezek a követelmények és igények indították el a Grid kutatást, illetve az ezzel kapcsolatos világméretű kísérleteket.

Mára már az élet minden területén megjelent az informatikai erőforrások elosztott, dinamikus, igény szerinti kihasználására vonatkozó követelmény, ami abból az egyszerű felismerésből fakad, hogy a felhasználói tömegek számára az informatikai alkalmazások szolgáltatásai pontosan olyan szolgáltatásként jelennek meg, mint az elektromos hálózat, a csatornahálózat vagy éppen a mosodai szolgáltatások. Az informatikai rendszerek felhasználói rétegének kiszélesedésével egyre több az olyan felhasználó, akit a technikai megoldások részletei nem vagy kevésbé érdekelnek.

A vezető informatikai nagyvállalatok szakemberei a napjainkban is forrongó, folyamatosan átalakuló Grid technológiai kutatásoktól várnak megoldásokat számos erőforrásigényes kérdésre. A kutatási eredmények és prognózisok a Gridtől hasonló forradalmi áttörést várnak, mint amilyen áttörés volt annak idején a World Wide Web megjelenése.

2. Technológiai háttér, történelmi áttekintés

A számítástechnika fejlődését kezdetektől fogva az jellemzi, hogy a technológiai fejlesztések során elért kapacitásnövekedést, legyen az számítási, tárolási vagy adatátviteli, a felhasználói igények szinte azonnal meghaladták. A számítási kapacitás növelésének egyik lehetséges útját sokan a párhuzamosításban ill. az erőforrások megosztásában látták szinte a kezdetektől fogva. A 90-es évek elejére több gyártó kínált speciális párhuzamos architektúrájú gépet. Ugyanakkor az asztali munkaállomások elterjedésével és a hálózati technológiák fejlődésével egyesek elkezdtek olcsó munkaállomásokból nagyméretű farmokat építeni, melyek teljesítménye egyszerű párhuzamosítási ill. feladatmegosztási technikák alkalmazásával egyes alkalmazási területeken felvette a versenyt a szuperszámítógépekkel.

Az operációs rendszerbe épülő vagy azt kiegészítő modulok segítségével egységes felület nyújtó, elsősorban szuperszámítógépek összekapcsolásával létrejött ún. *metaszámítógépek* a felhasználók (alkalmazók) számára egységes számítási erőforrásként jelentek meg, miközben a rendszer egyes elemei a hálózati technológiák rohamos fejlődése következtében fizikailag

egyre távolabb kerülhettek egymástól. A kezdetben szigorúan homogén rendszereket a 90-es évek második felére felváltották a heterogén architektúrájú rendszerek, melyek már farmokat és szuperszámítógépeket is tartalmaztak. Ezzel egy időben több kutatói közösségben merült fel az az igény, hogy ne csak számítási, hanem tárolási ill. egyéb erőforrások is elérhetőek ill. megoszthatóak legyenek az ilyen rendszerekben.

A 90-es évek közepéről két jelentős amerikai metaszámítógépes projektet érdemes kiemelni, a FAFNER-t (Factoring via Network-Enabled Recursion) és az I-WAY-t (Information Wide Area Year), melyek igen jelentősen befolyásolták a mai Grid rendszerek kialakulását. Így pl. a FAFNER projektben alkalmazott feladatmegosztási technológiát alkalmazta később a *seti@home* projekt is. Az I-WAY projekt pedig jelentős hatással volt a Globus Toolkit megoldásaira, mely megoldásokat a legtöbb Grid rendszerben ma is alkalmaznak [9].

A következő legfontosabb mérföldkőnek a Grid elnevezés megjelenése tekinthető. Az elnevezés az elektromos hálózat analógiája alapján egy olyan erőforrás-hálózatra utal, amely a világot behálózva tetszőleges számítógéphez kapcsolható erőforrás-elérést tesz lehetővé, legyen az számítási, tárolási erőforrás vagy akár egy speciális berendezés vagy érzékelő.

Valójában az erőforrások megosztásának igénye ill. lehetősége már a 60-as évek második felében megjelent a számítástechnikában a többfelhasználós operációs rendszerekkel, így kis túlzással azt is mondhatnánk, hogy a Grid a kerék újra feltalálása. Az „új kerék” azonban sokban különbözik a régitől: míg a többfelhasználós rendszerek egy központi gép igény szerinti erőforrásmegosztását biztosítják, addig a Grid rendszerek földrajzilag elosztott erőforrások elérését és igény szerint azok közös cél érdekében történő működtetését biztosítják. Így tehát nem csak elosztani képesek, hanem összefogni is, ami már igen lényeges különbség.

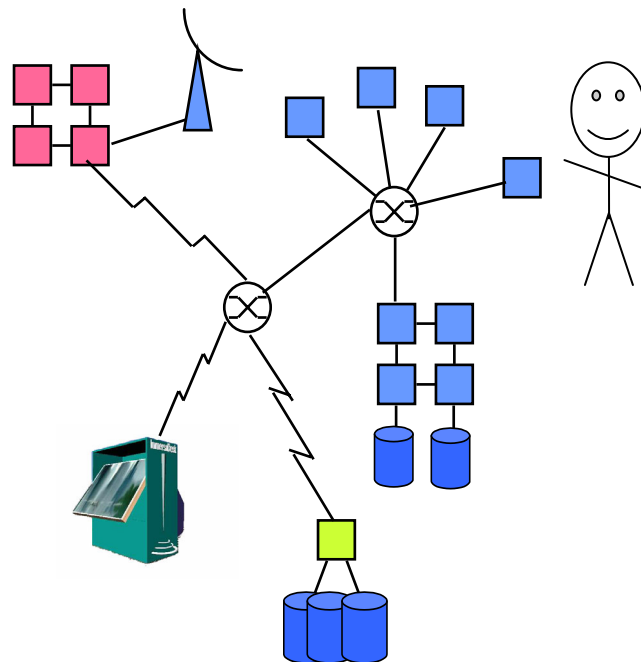
A Grid elnevezés 1997-ben keletkezett az amerikai Argonne National Laboratory-ban. Magát a koncepciót Ian Foster és Carl Kesselman foglalta össze a Grid bibliájának is nevezett könyvben. [1] Alapjaiban az itt vázolt koncepciót használják a mai rendszerek is, melyek azért hatalmas fejlődésen mentek keresztül az elmúlt közel 10 évben.

3. Mi a Grid valójában, hogyan működik?

A Grid rendszerek olyan számítógépes szolgáltatást valósítanak meg, amellyel a rendszer számítógépeihez kapcsolt erőforrások földrajzi helytől függetlenül, egységes módon, jogosultságokkal szabályozva elérhetőek a rendszer felhasználói ill. azok programjai számára (1. ábra). Míg a Web technológia segítségével alapvetően információ osztható meg, addig a Grid rendszerek az erőforrások megosztására illetőleg összefogására, hatékony újraelosztására alkalmasak.

Gyakran hibásan a Grid rendszereket a szabad CPU kapacitásokat felhasználó *seti@home* típusú rendszerekkel azonosítják. Ez a fajta működés a Grid rendszereknek csupán egy kis szeletét fedi le. A Grid technológia ettől jóval többet nyújt: lehetővé teszi akár földrajzilag egymástól igen távol levő, tetszőleges erőforrások egységes kezelését, felhasználását, megbízható, biztonságos módon történő újraelosztását.

Másik hibás megközelítés az, amikor a Grid rendszereket összekeverik a számítógépes farmokkal (*cluster*). Természetesen a farmok is képesek hasonló számítási, tárolási vagy akár információs szolgáltatásokat nyújtani, de a hagyományos farmok esetében egy felhasználót a farm minden egyes komponensének „ismernie” kell, azaz ahhoz, hogy egy adott szolgáltatást a farm felhasználója igénybe vegye, a farm minden komponensén rendelkeznie kell a felhasználónak felhasználói azonosítóval. Gyakran ezt az azonosítót használva magának a felhasználónak kell gondoskodnia a programoknak és a hozzájuk tartozó adatoknak a megfelelő farm komponensekre történő eljuttatásáról. A Grid rendszerekben komponens szinten nem beszélhetünk felhasználóról. Ha a Grid felhasználója jogosult egy erőforrás vagy szolgáltatás igénybevételére, akkor azt elérheti függetlenül attól, hogy az adott erőforráson van-e felhasználói azonosítója ill. hozzáférése.



1. ábra: Egy Grid rendszer vázlatos felépítése

Az elektromos hálózat hasonlatával élve a TV akkor is ráköthető az elektromos hálózatra, ha nincs belépőkártyánk az áramot előállító erőműbe, sőt azt sem kell tudnunk, hogy milyen energiahordozóból keletkezett a TV által felhasznált elektromos áram. Csupán az kell, hogy az áramszolgáltatóval megállapodjunk a szolgáltatás igénybevételéről. A Grid rendszerek esetében a szolgáltatóval történő megállapodás elektronikus aláírással hitelesített módon történik. A Grid szolgáltató ellenőrzi a megállapodást, valamint az aláírás hitelességét, majd a megállapodásnak megfelelően megadja a szükséges jogosultságokat a rendszer erőforrásainak igénybevételére.

A Grid nem egyetlen program, hanem számos egymásra épülő komponens, mely réteges felépítésével biztosítja, hogy az egyes komponensek könnyen lecserélhetők, továbbfejleszthetők és a meglévő rendszerekhez illeszthetők legyenek.

A legfontosabb ilyen elem az ún. köztesréteg (*middleware*). A köztesréteg a felhasználói réteg és az erőforrások fizikai rétege között helyezkedik el, és olyan fontos szolgáltatásokat biztosít, mint pl.:

- a felhasználók azonosítása és jogosultságának ellenőrzése,
- egyszeri bejelentkezés (single sign-on) megvalósítása,
- egységes, a helyi ütemezőtől független felület biztosítása a feladatok indításához és megállításához,
- adatátviteli szolgáltatások, ide tartozik a programok és azok adatainak a futási helyre vitele is,
- információgyűjtés az erőforrásokról és állapotukról,
- erőforrások el- és kiosztása (*brokering*).

A Grid rendszerek fontos szerepet játszanak az ún. e-kutatás, vagy e-tudomány kialakulásában is, ami nagyon leegyszerűsítve olyan virtuális közösségek kialakulását jelenti, melyek azonos kutatási feladatokon dolgoznak, eredményeiket, adataikat egymással megosztják, ugyanakkor földrajzilag, szervezetileg messze esnek egymástól, esetleg más földrészen vannak.

A virtualizálás fontos eleme a Grid rendszereknek. Ezért kézenfekvő a virtuális laboratóriumok, közösségek létrejötte.

4. Néhány jelentős Grid projekt

A teljesség igénye nélkül a következőkben bemutatok néhány olyan hazai és nemzetközi Grid projektet, amely véleményem szerint alapvetően meghatározta ill. meghatározza a hazai Grid kutatást.

4.1 Korai projektektől napjainkig

A hazai egyetemeken ill. kutatóhelyeken a 90-es években szintén megkezdődött a párhuzamos és elosztott rendszerekkel kapcsolatos kutatás, ill. kutatásfejlesztés, ami nagyban hozzájárult ahhoz, hogy manapság is lépést tudjunk tartani e tématerület fejlődésével. E kutatói közösség számára sok segítséget és támogatást jelentettek az IKTA (Információs és Kommunikációs Technológiai Alkalmazások) program keretében meghirdetett pályázatok. Ezek közül számunkra a legjelentősebb talán a Virtuális Szuperszámítógép Szolgáltatás Kialakítása az Akadémiai Hálózat felhasználásával (ViSzSzKi) elnevezésű projekt volt, mely lényegében megalapozta azt a tudást és ismeretet, melyet ma számos hazai és nemzetközi projektben kamatoztathatunk.

Annak ellenére, hogy kiemelt állami támogatást a Grid kutatások Magyarországon nem kaptak, több sikeres IKTA, NKFP (Nemzeti Kutatási és Fejlesztési Programok), GVOP (Gazdasági Versenyképesség Operatív Program) és egyéb olyan államilag támogatott kutatásfejlesztési projekt indulhatott el az utóbbi 15 évben, amely valamilyen módon kapcsolódik a Grid technológiához. (*ViSzSzKI, DemoGrid, SuperGrid, JGrid, Chemistry Grid, ClusterGrid, Hungarian Grid, MEGA*)

2003-ban az 5 legaktívabb Grid kutatói közösség létrehozta a Magyar Grid Kompetencia Központot (MGKK) [10] azzal a céllal, hogy a központ a kutatási tevékenységeket, pályázatokat koordinálja, az elért eredményeket pedig minél szélesebb körben publikálja. Az eltelt 3 évben az alapító tagok (BME, ELTE, KFKI-RMKI, MTA-SZTAKI, NIIF) számos közös hazai és nemzetközi projektben vettek ill. vesznek részt. A központhoz később csatlakozott a Veszprémi Egyetem is, ahol szintén jelentős Grid kutatások folynak. Az MGKK két regionális Grid központhoz is csatlakozik, a Közép-Európai Grid Konzorciumhoz (CEGC) és a Dél-Kelet-Európai Grid Konzorciumhoz.

2003 nem csak a hazai erők összpontosításának éve volt, mivel szintén 2003-ban megalakult a Közép-Európai Grid Konzorcium (CEGC) is Ausztria, Csehország, Lengyelország, Magyarország, Szlovákia és Szlovénia részvételével.

A hazai és nemzetközi összefogás nagyban hozzájárult ahhoz, hogy a közép-európai régió országai, így hazánk is rész vehessen több nemzetközi projektben (*EDG, GridLab, APART-2, COST, D23 action, UK e-science OGSA test-bed, EGEE, SEEGRID, CoreGrid, GridCoord, ICEAGE*).

4.2 EGEE (Enabling Grids for E-science)

Az EGEE projekt [4,5] az Európai Unió által támogatott egyik legjelentősebb Grid projekt. Célja a Grid technológia legújabb eredményeinek felhasználásával létrehozni egy olyan szolgáltatást, amely napi 24 órában rendelkezésre áll a kutatás-fejlesztés, később az ipari szolgáltatási alkalmazások számára is.

A CERN (European Organization for Nuclear Research) által vezetett projekt első fázisa 2006 márciusával ért véget, és 27 ország 70 kutató intézetének munkáját fogta össze. Az első fázis sikerének köszönhetően 2006 áprilisától újabb 2 éves kutatás indulhatott el lényegében az előző fázisban megkezdett kutatás folytatására és a szolgáltatás széles körű kiterjesztésére, a gLite Grid köztesréteg folyamatos fejlesztésére, hogy a felhasználók számára megbízható szolgáltatást nyújtson. A projekt mindkét fázisának megvalósításában több magyar intézmény is részt vett ill. részt vesz. Ezek a következők: BME, ELTE, KFKI-RMKI, NIIF, SZTAKI.

A projekt eredményeinek alkalmazása jelenleg két tudományterületen jelentős:

- Nagyenergiájú fizikai kutatások: a CERN-ben 2007-ben induló nagyenergiájú részecskék ütközését vizsgáló kísérletsorozat során évenként több mint 10 petabájt adat fog keletkezni. A Grid projekt feladata, hogy ezt az információmennyiséget letárolja, analízálhatóvá és elérhetővé tegye több ezer fizikus számára világszerte.
- Biomedikai problémák megoldása: a több különböző számításgépes applikáció közül a legjelentősebb a WISDOM (Wide In Silico Docking On Malaria) gyógyszerkutató projekt, mely során 46 millió vegyületet vizsgáltak meg a malária ellenszerét keresve.

A projekt nemzetközi együttműködéssel egy olyan szervizorientált köztesréteg kidolgozására és alkalmazására vállalkozott, amely egyesíti a jelenleg meglévő köztesrétegek előnyeit, és nem kötődik szorosan egy tudományterülethez sem. A gLite elnevezésű köztesréteg fejlesztése a projekt második fázisában is továbbfolytatódik, és jellemzően szolgáltatásorientált architektúrának (Service Oriented Architecture – SOA) megfelelően épül fel.

A projektben jelenleg elérhető számítási és tárolási kapacitás nagysága lenyűgöző: jelenleg (2006. május) **184** helyszínen összesen **33938** processzor működik [11]. A rendelkezésre álló tárolókapacitás **10180 TB**. A helyszínek a föld legkülönbözőbb pontjain találhatók (2. ábra).



2. ábra: EGEE projekt erőforrásait adó helyszínek

4.3 Globus

A Globus Alliance szervezetek és egyének egy közössége, akik részt vesznek a Grid eszközök fejlesztésében, használatában. A Globus szoftvere a Globus Toolkit [2], eszközök és programkönyvtárak gyűjteménye, melyek az elosztott rendszerek fejlesztésekor felbukkanó leggyakoribb problémákra kínálnak megoldásokat.

A Globus Toolkit (GT) nagyban meghatározta ill. meghatározza a különböző fejlesztések irányát. A programcsomagot a 90-es évek óta fejlesztik. A tapasztalatok alapján, a különböző Grid rendszerek specializált volta ellenére sok közös alapfeladat megvalósítása szükséges minden elosztott rendszerben. A szoftver célja, hogy olyan fő Grid feladatok megvalósítását tegye lehetővé, mint például az erőforrás-kezelés, biztonság, erőforrás-hozzáférés, adatmozgatás, erőforrás-felfedezés, és így egy általánosan használható köztes réteget kínáljon a fejlesztők számára. A Toolkit nem jelent kész Grid szoftvert, szükség van a Grid rendszer specifikus területeinek önálló fejlesztésére, az egyes építőelemek összeintegrálására.

4.4 NorduGrid

A NorduGrid projekt [7] 2001 májusában indult azzal a céllal, hogy északi országok (Dánia, Norvégia, Svédország és Finnország) különböző akadémiai intézményeinek összefogásával létrehozson egy Grid infrastruktúrát. A projekt célja, hogy megállapítsák, képesek-e a Grid rendszerek kielégíteni a mai tudományos kutatások megnövekedett számítási és tároló kapacitás igényét. A cél egy saját Grid infrastruktúra megvalósítása volt egészen az alapoktól. A gyors fejlesztéshez felhasználták a European Data Grid (EDG) korábbi eredményeit és a Globus Toolkit 2-t (GT2), így egy év alatt elkészült az első infrastruktúra. A NorduGrid

jelenleg több mint 1500 CPU-t tartalmaz, melyek napi 24 órában állnak a tudósok rendelkezésére.

4.5 ClusterGrid

A NIIF ClusterGrid infrastruktúra [8] létrehozása az Oktatási Minisztérium 2002 júliusában meghirdetett számítógéplaborok létrehozására vonatkozó pályázatára vezethető vissza. A cél különböző magyar egyetemeken elhelyezett közel kétezer asztali PC egységes rendszerbe történő összefogása volt. A gépek nappal szokásos egyetemi feladatokat látnak el, míg éjjel és hétvégén egységes rendszerbe szerveződve számításigényes tudományos feladatokat végeznek, ezzel megvalósul a gépek 24 órás hatékony kihasználása.

A rendszer erősen támaszkodik az akadémiai gerinchálózat (HBONE) által nyújtott szolgáltatásokra, amelyen keresztül egy virtuális magánhálózatot (VPN) alakít ki, ezáltal különválasztva a nyilvános Internetet és a Grid-en belüli hálózati forgalmat.

Operációs rendszer tekintetében időbeli és térbeli elválasztásról is beszélhetünk. Az egyetemi, irodai jellegű feladatok elvégzésére Windows környezet áll rendelkezésre, míg a tudományos munkák számára egy UNIX környezet nyújt több lehetőséget.

4.6 MEGA

A BME IK által vezetett, Magyar informatikai erőforráshálózat (Grid) alapjai (Grid közmű) projekt [6] elsődleges célja a magyar Grid infrastruktúrális kutatások kiléptetése az akadémiai és kutatási szférából, ipari alkalmazások és felhasználási területek keresése, kidolgozása és tesztelése.

A projekt alapvetően abból a tényből indul ki, hogy akadémiai viszonylatban növekszik a Grid technológia ismertsége. A Grid technológia hazai iparban történő felhasználása, ismertsége azonban elmarad a kívánatostól. Ennek oka részben technológiai, mivel még nem beszélhetünk teljesen kiforrott technológiáról, részben pedig a társadalmi ismertség, támogatottság hiánya. A projekt célja ezért olyan Grid rendszer létrehozása és bemutatása, amely az ipar és a szolgáltató környezet számára is vonzó, általános célú „közműszolgáltatásként” jelenik meg.

Az NKFP támogatású projektben projektben olyan alkalmazásokat dolgozunk ki, melyek távol esnek az informatikától, ugyanakkor a technológia igen jól alkalmazható. Ilyen pl. a hídgerendák méretezése, ill. a talajeróziós számítások is.

5. Összefoglalás

Büszkén állíthatjuk, hogy a hazai Grid kutatások színvonala kiállja a nemzetközi megmérettetést is. Ennek legjobb bizonyítéka, hogy a magyar kutatóintézetek számos nemzetközi projektben is részt vesznek, eredményeiket elismerik. A hazai kutatóintézetek eredményei közül külön érdemes kiemelni az NIIF ClusterGrid projektjét, a BME MEGA projektjét, a SZTAKI P-Grade portálját, a Veszprémi Egyetem JGrid projektjét és számos Grid-hez kapcsolódó kutatást mint pl. a KFKI-RMKI-ban folyó részecskefizikai kutatásokat.

A kutatások egyre inkább áttevődnek a alkalmazások irányába, habár számos nyitott kérdés van az infrastruktúra kialakításának terén is. Megítélésem szerint a legfontosabbak ezek közül a bizalomépítő, biztonsági megoldások, illetve a mára letisztulni látszó 9-10 különböző köztesréteg együttműködési kérdéseinek, megoldása. Ez a törekvés leginkább a

szabványosítási fórumok munkáján látszik, melyben magyar szakemberek is részt vesznek. Ezek közül kiemelkedő a Global Grid Forum (GGF) keretein belül készülő (Open Grid Service Architecture – OGSA) architektúra, melynek célja, hogy lehetővé tegye elosztott rendszerek integrációját, virtualizációját, és menedzsmentjét egy széles körben elfogadott keretrendszeren keresztül.

Köszönetnyilvánítás

Jelen cikk közvetlenül felhasználta az NKFP OM-00262/2004 és 2/009/04 projekt, a Pázmány Péter RET-06/2006 program, valamint az EGEE-II INFSO-031688 program támogatásával elért eredményeket.

Irodalomjegyzék

- [1] Ian Foster, Carl Kesselman: *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 1999.
- [2] Ian Foster, „Globus Toolkit Version 4: Software for Service-Oriented Systems”, Argonne, University of Chicago, 2005
- [3] Stefan, P. and Szalai, F. and Vitez, G., „*Toward the Implementation of Production Grid Systems: The Hungarian ClusterGrid Infrastructure Project*”, Office for National Information and Infrastructure Development, 2003
- [4] EGEE project, *EGEE project portal*, <http://www.eu-egee.org>.
- [5] EGEE projekt: <http://egee.ik.bme.hu>
- [6] MEGA projekt, *MEGA projekt portal*, <http://mega.ik.bme.hu>
- [7] NorduGrid project, *The NorduGrid portal*, <http://www.nordugrid.org>, The NorduGrid Collaboration
- [8] NIIF ClusterGrid projekt, ClusterGrid projekt honlap, <http://www.clustergrid.iif.hu>., Nemzeti Információs Infrastruktúra Fejlesztési Program
- [9] GridCafé: <http://gridcafe.web.cern.ch>
- [10] MGKK honlap: <http://www.mgkk.hu>
- [11] EGEE erőforrásmonitorok, <http://goc.grid-support.ac.uk/gridsite/monitoring/>