

AZ INFORMATIKA SZÁMÍTÁSTUDOMÁNYI ÉS MATEMATIKAI PROBLÉMÁIRÓL¹

ARATÓ MÁTYÁS

Budapest

ARCHIMEDES szerint egy, a földünkön kívüli fix pont elegendő lenne a föld kimozdításához. Ugyanilyen alaptétel — mely azonban kimondatlan —, hogy az irányítás elvégezhető lenne, ha megfelelő időpontban rendelkezésre állnának a szükséges adatok. Archimedes óta a világgép megváltozott, az általa kimondott tétel lényegében igaz maradt, a megvalósítás érdekében természetesen azóta sem történt egyetlen lépés sem, azonban jelentősen fejlődtek világunk változtatásának lehetőségei. Az irányításhoz szükséges adatok biztosításával kapcsolatban alaptételünk alapvetően nem változott, csak az igények nőnek beláthatatlanul. Példaként említhető, hogy tudományos publikációkban abból az alapvető feltevésekből indulnak ki, hogy ha adva lennének a népgazdasági tervezéshez szükséges éves adatok, vagy egy vállalaton belül az irányításhoz szükséges műszakonként kész adatok, vagy a folyamatirányításban a másodpercnyi kiértékeléshez szükséges adatok, akkor a népgazdasági optimális terv kidolgozása, a vállalati optimális irányítás vagy termelésirányítás feladata, a folyamatirányítás feladata egyértelműen és jól lenne megoldható. A megoldás alapvetően helytelen feltételezésből indul ki, így tulajdonképpen mind a népgazdasági optimális tervezésben, mind az optimális irányításban addig nincsenek megoldásaink, amíg az adatok megfelelő szintű biztosítása meg nem történik.

Az optimális irányítással, az optimális tervek kidolgozásával kapcsolatos matematikai vizsgálatok gyakran olyan feltételezéseken alapulnak, amelyek majdnem ekvivalensek a földön kívüli fix pont biztosításával. Nincsen legtöbbször kimondva és megfogalmazva, hogy milyen típusú adatok, milyen gyorsan és milyen feltételek mellett biztosíthatók az optimális irányítás megoldásához. Emellett nem kevés gondot okoz, hogy az adatállomány legkisebb változása alapvetően megváltoztatja a tervet is. Erre vonatkozóan rengeteg példa, illusztráció látott napvilágot, emlékeztetünk itt a dinamikus programozás, a nem korrekt kitzésű feladatok megoldásaira, vagy a sztochasztikus kitzésű feladatok megoldásaira, vagy a sztochasztikus folyamatok átmeneti függvényeiben történő apró változtatásokra, melyek kétségessé teszik megbízható matematikai megoldások biztosítását.

A felsorolt negatívumok arra mutatnak, hogy nagyobb kutatói és tudományos igényességgel kell foglalkozni az adatbiztosítás, a megbízhatóság, a stabilis megoldások kérdéseivel, melyek alapvetőek a számítástechnikai megoldások esetén. Az adatok és adatkezelés feladatai az informatikában új módon vetődnek fel, olyan újfajta feladatmegfogalmazások kerültek előtérbe, amelyek korábbi vizsgálatainkban nem szerepeltek.

1. A hiba megjelenése

Annak illusztrálását, hogy a számítástudomány új feladatok elé állította a kutatókat, csak egyetlen példán kívánom elvégezni. A hibás feldolgozási problémákkal, a hiba megjelenési formáival kívánok foglalkozni. A hiba megjelenését matematikailag először a mérések kiértékelésében, majd a híradástechnikában kezelték és kezeljük ma is. A számítástechnikában, a számítástechnikai alkalmazásokban ma már

¹ Elhangzott az MTA Matematikai és Fizikai Tudományok Osztályának a *Magyar Tudományos Akadémia* 1978. évi közgyűlése keretében 1978. május 11-én megtartott ülésén.

megszokottá vált a sajátos hibakezelés, a megfelelő matematikai modell azonban még hiányzik. A hibák időbeli megjelenésének leírása, a különböző hibák osztályozása és felismerése, a működő rendszerekre kifejtett hatások vizsgálata úgy érezzük érdekes és új matematikai modelleket igényel.

Azoknak a matematikai modelleknek a kialakítása, amelyek az ún. egyenes feladatokkal foglalkoznak, ahol adott feltételek mellett, adott adatok alapján optimális megoldásokat keresünk, ma is jelentős helyet foglal el a kutatásokban. Az ellenítés, vagy a matematika nyelvén fogalmazva inverz feladat az, amikor is azt vizsgáljuk, hogy hogyan biztosíthatók az adatok a modell működéséhez, hogyan biztosíthatók a modell feltételei és hogyan látható be bizonyos körülmények között (pl. időben) változó adatok mellett a megoldások megbízhatósága. A számítástechnikában a megbízhatóság és stabilitás fogalma, ami azt jelenti, hogy az adatokban bekövetkező kis változtatások esetén a megoldásnak az eredetihez közeli megoldást kell biztosítani, alapvető kategória. Jelentősége nem kisebb, mint az optimális üzemeltetésé, vagy az optimális működtetésé. Az optimális üzemeltetés, az optimális működtetés minden esetben a megbízható megoldás biztosítása után vizsgáljuk. A valószínűségelmélet feladatmegfogalmazásával szemben a matematikai statisztika feladatmegfogalmazásai jelentik az inverz feladatot. A valószínűségelméletben abból indulunk ki, hogy adott eloszlások esetén esemény vagy eseménysorozat tulajdonságait határozzuk meg. Ezzel szemben a statisztikában a megfigyelési adatok alapján kell meghatározni az eloszlások tulajdonságait, majd ennek alapján következtetéseket levonni. Az adatok összegyűjtése, a feltételek megvizsgálása jóval több munkát, a gyakorlattal való szoros kapcsolatot jelenti. Ehhez járul bizonyos események valószínűségeinek vizsgálata, és azok megbízhatóságának figyelembe vétele is.

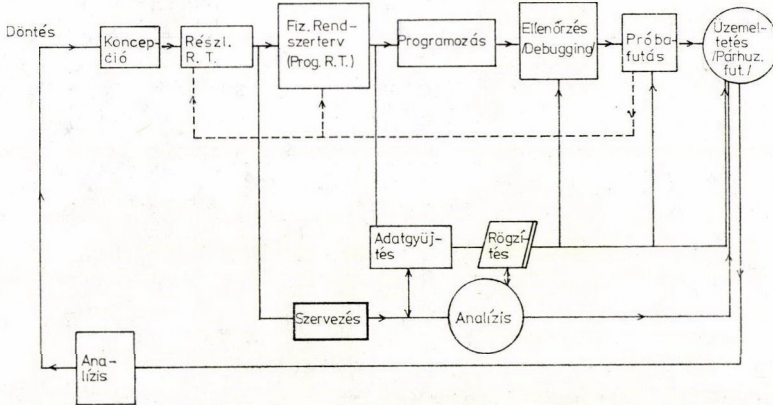
Másik példaként a jóval egyszerűbben megfogalmazható, lineáris egyenletrendszer megoldhatósága feltételének vizsgálata említhető. A matematikai megoldással szemben, a determináns nulla értékének a meghatározása a gyakorlat számára nem ad használható megoldást. A számítástechnikai megoldások az ilyen és hasonló kérdések elvi tisztázását igénylik.

Nagyon sok rendszer, „elvileg” jól felépített számítógépes rendszer bukásának szemtanúi lehettünk amiatt, hogy a rendszer készítői nem számoltak az adathibákkal, vagy műszaki hibákkal, vagy a programrendszer készítői által elkövetett hibákkal. A hibák analízise éppúgy hozzá tartozik a rendszer készítéséhez, mint a rendszer logikai struktúrájának felépítése.

A továbbiakban a rendszer szót általában olyan információs rendszerekre fogom használni, amelyekbe a számítástechnikai eszközök alapvetően beépültek, amikor az információ-feldolgozás rendjében részt vesznek az adatrögzítő és feldolgozó berendezések, a kiértékelő berendezések, software-rendszerek is. Ilyen rendszer pl. népgazdasági szinten egy tervezési rendszer, vállalatoknál egy termelés-irányítási, pénzgazdálkodási, állóeszköz-gazdálkodási, vagy bér- és munkaügyi rendszer stb. Ezeknek a rendszereknek a legfontosabb jellemzője, hogy a számítástechnika megjelenése előtt is működő rendszerek voltak, az információs szolgáltatás akkori rendje alapján is történtek döntések és folyt termelésirányítás is. Éppen ezért a számítástechnikai eszközök bevonulása a rendszerek jelentős mértékű megváltozását, a rendszerek jobbá tételét kell hogy szolgálja. Ennek a jobbá tételnek a hatékonyság szempontjából egyik legfontosabb feltétele, hogy a rendszer bevezetésének átfutási

ideje megfelelő módon alakuljon. Éppen ennek az átfutási időnek a megvizsgálását tekintjük most alapvető feladatunknak.

Vizsgáljuk meg az 1. ábrát, amely egy számítógépes irányítási rendszer időbeni kialakítását mutatja be. A rendszert visszacsatolósónak képzeljük, mert ez



1. ábra

felel meg leginkább a gyakorlatnak. Időben először a már működő rendszer működésének vizsgálata szerepel, ami után a számítógépes rendszer működésére, üzemeltetésére vonatkozó alapkoncepció elkészítésére kerül sor. A régi rendszer kiváltása újabb és jobb, esetleg más típusú szolgáltatásokkal együtt történik. Az alapkoncepció elkészülte után rendszerterv, majd részletes rendszerterv készül. Közben elkezdődik a szervezési munka a régi rendszer átalakítására. Ez a szervezési munka kiterjed az oktatásra, a felkészülésre, az új rendszer üzemeltetésére és így tovább. A részletes rendszerterv megvalósítása után szokott elkezdődni az adatelőkészítés, a feladat megvalósításához szükséges adatok biztosítása, mely a későbbiekben a folyamatos adatszolgáltatássá alakul. Fizikai rendszerterv vagy programrendszerterv készül, amely után a programrendszer készítésére kerül sor. Az adatelőkészítés és a szervezési munka alapján adatbiztosítás is történik a programrendszer próbájához. Ekkor lényeges visszacsatolás történik, egyrészt a programrendszer javítására, másrészt a rendszertervek helyességének ellenőrzésére. A programrendszer és a próbadatok biztosítása után kerülhet sor próbaüzemre, majd a párhuzamos üzemeltetésére és az új rendszer bevezetésére, értékelésre. Jellemzője a rendszerek készítésének, hogy az üzemeltetés tapasztalatai alapján, idővel a rendszer koncepciójának megvizsgálására, a rendszer hatékonyságának értékelésére újból vissza kell térni, és esetleg a rendszert újra kell alakítani. Felvetődik a kérdés, hogy a hiba megjelenés, a hiba javítás mennyiben változtatja a rendszer működését és mennyiben módosítja a rendszer kialakítását. A részletes rendszertervben, előzőleg a rendszerkoncepcióban is, a számítástechnikai eszközök lehetőségeire és a rendelkezésre álló erőforrásokra bizonyos becslések szerepelnek. Erre épül a részletes rendszerterv logikai felépítése. A több kötet, néha több száz oldalas rendszerterv logikai felépítésében adódhatnak hibák, téves elképzelések, amelyek pl. a csoportmunka nehézségeiből keletkeznek;

vagy abból, hogy az egyes készítőik nem értik a feladat megoldását teljes mértékben. Előfordulnak olyan hibák is, amelyek elírásból vagy nem világos fogalmazásból adódnak.

A fizikai rendszerterv (programrendszerterv) elkészítése a részletes rendszerterv félreértéseiből adódó hibákat és a programtervezésből adódó hibákat is tartalmazhat. A programrendszer-készítés speciális hibáit a számítástechnikában már régen észrevették, itt alakult ki először a hibák kezelése. A „debugging” a programozásban, a programban jelenlevő hibák keresését, javítását jelenti. A bevezetendő rendszerek megbízható működésével és átfutási idejének csökkentésével kapcsolatban a debugging tanulmányozása bizonyult a leghasznosabbnak. Ezt mutatja a debugging idejének jelentős csökkentése. Ismertek software-eszközök a debugging támogatására és software-eszközök a debugging idő lecsökkentésére. Az interaktív eszközök kialakítása az egyik legszebb és legjobb példája a praktikus alkalmazkodásnak az elméleti kutatásokhoz és egyben a gyakorlathoz. A debugging idő dinamikus vizsgálata természetes módon függ a rendelkezésre álló hardware- és software-eszközök milyenségétől.

Az adatbiztosítás hasonló módon hibákkal jár együtt. Az adatbiztosítást a programrendszer debugging ideje szempontjából vizsgálva azt találjuk, hogy amennyiben az adatok nem bizonyulnak megbízhatónak, a programrendszer javítása többszöröse is kiterjedhet annak az időnek, mint ami a megbízható adatok biztosítása esetén adódik. Az adatbiztosítás megfelelő szintjét a programozók és a rendszer készítői természetesnek tekintik, de nem fogalmazzák meg. A programrendszer működőképes állapotba kerülése, ami nem ekvivalens azzal a megfogalmazással, hogy a programrendszerben nincs hiba, azt jelenti, hogy az adott adatszolgáltatás mellett a programrendszer jól működik. Az összes lehetőség teljes kipróbálására egy programon belül általában nincs mód, mivel ez meghaladja a jelenlegi számítógépes rendszerek lehetőségeit.

A próbaüzemeltetés lehetővé teszi a rendszer működésének elemzését, mind a hatékonyság, mind a megfelelő feldolgozási idők, a szervezéshez való kapcsolódás szempontjából is. Itt kapcsolódnak a szervezés eredményei, ekkor derül ki, hogy valójában a rendszer működőképes állapotba hozható-e, hiszen a megfelelő időben történő szolgáltatások biztosítása elsősorban a szervezés milyenségétől függ.

A párhuzamos üzemeltetés alatt a régi rendszer értékelése ismételtelen megtörténik, és az új rendszer lehetőségei kiderülnek. Mind a próbaüzemeltetés, mind a párhuzamos üzemeltetés alatt kiderülnek további adathibák, programrendszerbeli hibák, előfordulnak a számítógépes rendszer üzemeltetésében is hibák. A programrendszer és a szervezés kapcsolatában előforduló ellentmondások visszacsatolása a rendszertervbe nem mindig megoldható. Az 1. ábrán látható rendszer kialakításával kapcsolatos problémáknak szavakban megadott leírását a gyakorlati rendszerek bevezetésében járatos ember ismeri, a rendszer ilyen leegyszerűsített vázát természetes módon kezeli, a jelenségekben fellelhető módszeres és véletlen hibák megjelenését tovább pontosítja. A mintegy 2-3 éves átfutási időben kialakuló rendszernek a megfelelő matematikai modelljét megadni ma még nem tudjuk, mert nem állnak pl. rendelkezésre a becslésekhez szükséges empirikus adatok. További nehézség, hogy a programrendszerbeli hibák lényegesen különböznek az adathibáktól, nemcsak jellegükben, hanem a rendszerre kifejtett hatásuknál fogva is. Egy programrendszeren belül részprogramok kapcsolódási hibái, az input-output félreértelmezéséből, vagy

pontatlanságaiból adódó hibák lényegesen más jellegűek, mint azok a hibák, amelyek logikaiak, vagy az olyan hibák, amelyek csak bizonyos kezdeti feltételek teljesülése esetén jelennek meg.

Az 1. ábrához még két gyakorlati megjegyzést helyes fűznünk. Az első megjegyzésben két alapvető hibát szeretnénk megemlíteni, amely rendszerek bukásához vezethet. Az első hiba az, amikor a részletes rendszerterv hiányosságai csak az üzemi próbáknál derülnek ki. A javítás olyan több éves késést okozhat — hiszen a részletes rendszertervtől a próbaüzemig 1-2 év telt el —, hogy ennek a periódusnak a megismétlésére már nincs lehetőség. Másik hiba: feltételezve, hogy a rendszer készítői képesek a feladatok megoldására, előfordulhat, hogy az adatelőkészítés és az adatellenőrzés között olyan hosszú idő telik el, hogy közben a változó adatok száma nagyságrendben túlhaladja az előkészített alapadatokat. Ez a jelenség különösen a nagynyilvántartási rendszereknél léphet fel.

Másik gyakorlati megjegyzés az, hogy a rendszer kialakításának, átfutási idejének már említett becslését kell megadnunk. Az 1-2 vagy 3-4 éves terminusok megállapítása a rendszer egyes elemeinek becsléséből és az elemek összeadásából történik. Megkérdezve a programrendszer készítőit debugging időben, hogy mennyi idő alatt tudják a programrendszert működőképes állapotba hozni, legtöbbször 2-3-szorosát mondják annak a becslésnek, mint amit a rendszer készítésekor adtak. Mielőtt egy statisztikai modell kialakítását elkezdenénk, vizsgáljuk meg, hogy a hibák megjelenésére és javítására, szűrésére vonatkozóan milyen korábbi vizsgálatokra támaszkodhatunk.

2. A hibák és a hibás feldolgozás matematikai modelljeiről

A híradástechnikában, a távközlésben és a mérésekben legtöbbször a hasznos jelre egy időben és nagyságban véletlen additív zaj rakódik. A vett jel ζ_t a t időpillanatban $\zeta_t = \zeta_t + \varepsilon_t$ alakú, ahol ε_t jelenti a zajt. Mind a ζ_t időben lejátszódó folyamat természetére vonatkozóan, mind az ε_t zajfolyamat természetére vonatkozóan különböző statisztikai feltevésekkel szokás élni és ennek alapján szokás megoldani a filtrációt, az előrejelzés változatos feladatait attól függően, hogy ismereteink milyenek.

A híradástechnikában és távközléseméletben mind a ζ_t , mind az ε_t folyamat természetének vizsgálatára vonatkozóan kidolgozódtak azok a módszerek, amelyek lehetővé teszik a paraméterek meghatározását, a paraméterek becslését, azok megbízhatóságának vizsgálatát is.

Az informatikai rendszerek kialakításánál az előbbi típusú zajok ritkán fordulnak elő, még akkor is, hogyha az additivitás mellett a multiplikatív hibák megjelenését is tekintetbe vesszük.

A hibás elemekből álló, de mégis megbízhatóan működő rendszerek működésének kérdésében az első érdekes számítások SHANNONNÁL szerepelnek, aki nagy valószínűséggel hibásan működő logikai kapcsoló elemekből felépített, megbízhatóan működő rendszereket vizsgált. Hasonló típusú kérdéskör megoldásával állunk szemben az informatikai rendszerek felépítésénél is. A hibák megjelenése, illetve a meghibásodás vizsgálata azonban más formában vetődik fel. A SHANNON általi vizsgált rendszerekben ugyanis csak egyszerű *Boole-függvényekkel* leírható elemek hibás működéséről, duplázásáról, illetve párhuzamosításáról volt szó. Az informa-

tikában nem két- vagy többértékű logikai függvények kapcsolását, hanem időben és állapotban is változó rendszerek felépítését kell vizsgálnunk.

Tudva azt, hogy a rendszerterv, a szervezés hibái lényegesen befolyásolják a rendszer megbízható működését, ezt egyelőre figyelmen kívül hagyjuk, és csak a számítástechnikai rendszer hibáival, majd az adatok, az adatrendszerek hibáival foglalkozunk.

A programrendszerek zavaraira vonatkozóan legjellemzőbb példaként az operációs rendszerekben tapasztalható hibákat vizsgáljuk. Egy már több éve működő operációs rendszer működésében tapasztalható hibákról rendszerint a gyártó vállalatok hibajegyzéket és hibajavításokat nyújtanak. Ezek a hibajavító rendszerek kötelet szokta kitenni. Általában a hosszabb idő után felgyülemlett hibák ismeretében egy, a lehetőségekben is új programrendszer vagy operációs rendszer bevezetésére, ún. új „release” bevezetésére kerül sor. A hibák javítása azonban egy nagy hiányossággal rendelkezik, mégpedig az addig hibás rendszerben működő alkalmazói programrendszerek az új „release”-ben meghibásodhatnak. Ez a „rendezés” esetleg működő rendszereket tehet tönkre, tehát egy új „release” bevezetése, annak ellenére, hogy jobb és nagyobb lehetőségekkel rendelkezik, mégis nagy gazdasági károkat okozhat. Miből adódik ez a zavar? A programozók rendszerint az új „release” előtt észreveszik, hogy az operációs rendszer nem azt csinálja, mint ami a leírásban szerepel. Ezt kikerülendő, a valódi működő rendszert használják programjaik megírásánál, és ily módon a programok alkalmazási szempontból jól működnek. A megváltoztatott operációs rendszer kijavítja az eredeti, valóban hibás megoldást és a leírásnak megfelelővé teszi a rendszert, de egyúttal a korábban kialakult alkalmazói rendszereket tönkre teszi. Itt most nem soroljuk fel az operációs rendszerekben, a hibalistákon szereplő hibákat. Adattömegeket gyakran feldolgozó programrendszerek működésével kapcsolatban „érdekes” hibajelenség az, amelyik a ritkán változó, vagy ritkán az adatbázisba bekerülő adatok esetén még „emberi szemmel” sem észrevehető eltérést okoz a programban. A programrendszer hibás működésének észrevételére csak a rendszer-feldolgozás után néhány hónappal, esetleg 1-2 évvel kerül sor, amikor a hibás adatokban történő összesítés egy meghatározott szint fölé kerül. A zavar oka általában az, hogy a programrendszer egy olyan ága kerül felhasználásra, amelyik csak az esetek tíz- vagy százszázad részében ad hibás eredményt.

Jelentős zavart okoz nagy adattömegben való konvertálás közben az adatokban létrehozott olyan módszeres hiba, amelyik nem állítható vissza. Például egy adatcsoport olyan más adatszoportba történő konvertálása, amely más módon is előállítható. A hibás tétel előállítása egyértelmű, de a visszaállítása lehetetlen, csak az eredeti adatok felhasználásával és helyes újrakonvertálásával állítható vissza az eredeti állapot. Ha egy ilyen hiba felfedezésére csak néhány hónap múlva, a rendszeres feldolgozások elvégzése után kerül sor, kritikus helyzet alakulhat ki. A zavar kiküszöbölése ugyanis vagy az egész adatfeldolgozási folyamat megismétlését jelentené, ami majdnem megvalósíthatatlan, vagy a hiba kiszűrését egy nagy biztonságú, különböző szintű logikai, esetleg nyelvi próbákra alapuló javító rendszer beépítésével lehet megvalósítani.

Nagy adatrendszerek feldolgozásánál gyakran használt gyorsító eljárás a szótárak bevezetése. Segítségükkel a keresés, a visszakeresés folyamata felgyorsítható és a feldolgozás folyamata jóval rövidebbé tehető. A szótárak bevezetésének azonban

nagy hátránya, hogy a bennük előforduló hibák vagy pontatlanságok az adatrendszer jó adatainak hibáit jelentik. A szótárak állhatnak családnevekből, utcanevekből, utónevekből, vagy a termelésben a különböző fajta eszközök elnevezéseiből. A szótárakba olyan beírások vagy leírások is kerülhetnek, amelyek nem felelnek meg a magyar nyelvtan szabályainak. Ezeknek a hibáknak a kihatása alapvető az adatfeldolgozásban.

Nagy adatrendszerek feldolgozásában jelentős problémát okoz a következő típusú zavar. Többlépcsős feldolgozás második lépcsőjében ellenőrzött adatok feldolgozását tételezi fel a programrendszer (maga már nem ellenőríz). Az első lépcsőben véletlenszerűen fellépő adathiba a második lépcső feldolgozási programrendszerét teszi tönkre.

Az adatrendszerek előállításával, a feldolgozás színvonalával kapcsolatban tett megjegyzések azt jelentik, hogy az adatok rögzítése, az adatok ellenőrzése nagy mértékben függ a feldolgozók emberi tulajdonságaitól. A szubjektív elemekkel, azok statisztikai értékelésével az adatfeldolgozásnál ugyanúgy kell foglalkozni, mint ahogyan azt a mindennapi munkában, a termelésben tesszük.

Az adatforgalom szempontjából ma Magyarországon három rendszert szoktunk megkülönböztetni:

— A nagy, de elég statikusan viselkedő rendszereket, ami azt jelenti, hogy mintegy 10^6 — 10^7 nagyságrendű rekordtömeg kezelése folyik gépen. Erre a rekordtömegre és annak file-jaira vonatkozó kérdések általában elég stabilak, illetve a kérdések rendszere könnyen kialakítható;

— A kisebb, 10^4 — 10^5 nagyságrendű rekord, illetve adattömeg, amely viszonylag gyorsan változó és összetételében is bonyolult lekérdezési rendszernek van alávetve;

— A mérés- és adatgyűjtéssel, vagy a hiradástechnikában ismeretes rendszerrel kapcsolatos a harmadik típus, ahol az x_1, \dots, x_n , (x_i vektort jelent), alakú leírásokkal kapcsolatos, ahol az index azt jelenti, hogy az időben, esetleg térben, vagy más paraméterben változó sorozat kezeléséről van szó. Ebben az esetben azonban a vektorok dimenziója nem nagyon nagy.

A tudományos kutatások középpontjában elsősorban a harmadik típusú rendszerek vizsgálata szerepelt, hiszen itt álltak rendelkezésre a fizikai és műszaki alkalmazások szempontjából fontos matematikai modellek. Ez a rendszer nagyobb figyelmet érdemel a számítástechnikai feldolgozás szempontjából.

Az első két típus vizsgálata mindaddig empirikus úton történt, a továbblépés módszeres tanulmányozást és modellalkotást igényel. Ehhez figyelembe kell venni, hogy az adatok gépbe kerülése emberi munkához kapcsolódik és ez befolyásolja az adatok milyenségét, és ellenőrzésük automatizálhatóságát. Nagyon hosszú a feldolgozás átfutási ideje és hosszadalmas a bekerülő adatok ellenőrzése. Az adatok az információs rendszerbe legtöbbször bizonylatokról, formátumokról kerülnek, de nem automatikus módon. Az ellenőrzött adatoknak a számítógépbe kerülése felgyorsítása az egyik legfontosabb feladat. Az egy-két hetes, sőt néha hónapos előzetes ellenőrzési, karbantartási folyamatnak 1—2 órára történő lecsökkentése alapvetően változtathatja meg a működő alkalmazási rendszerek hatékonyságát. Különösen a vállalati alkalmazások terén van erre szükség, ahol a csoportos adatrögzítés kérdése nem megoldott.

Az adatok, adatrendszerek biztonságos kezelésével, tehát alkalmazási működő rendszerek biztonságos kezelésével kapcsolatban igen sok értékes eredmény és matematikai modell került már kidolgozásra, megemlíthető pl. GELENBE, CHANDI neve, de magyar kutatók nevéhez is fűződnek sikeres vizsgálatok. Tekintsük a következő példát, amikor az adatbázis állapotát az ún. „módosító tranzakciók egymásutánja” váltja ki. A tranzakciók beérkezése között véletlen idő telik el, a tranzakció végrehajtásához szükséges idő is általában véletlen. A módosítást kiváltó információk egy meghatározott ideig tárolóban maradnak azért, hogy az esetlegesen bekövetkező hiba esetén megismételhetők legyenek. Ezek a hibák adódhatnak az adatok helytelen kezeléséből, a hardware- és software-hibákból. Az adatbázis meghatározott időpontokban kimentésre kerül és a következő kimentésig tárolódik. A kimentő eljárást szokás kontrollpont-képzésnek nevezni. A különböző kontrollpont-képzési időpontok közötti távolság általában valószínűségi mennyiség, véletlen változó. A kimentés időtartamát tekinthetjük fixnek, vagy ugyancsak véletlen mennyiségnek. Hiba bekövetkezése esetén a legutolsó kontrollpont-képzésnél kimentett adatbázis-állapotból a tárolt módosítások segítségével újrafeldolgozás kezdődik. Az egyes hibák előfordulása közötti intervallum ugyancsak véletlen mennyiség. A visszaállítási időt hibajavítási periódusnak nevezzük. Abban az esetben, hogyha a hiba észrevétele 1 valószínűséggel történik, a rendszer működésével kapcsolatban a következő feltevésekkel szokás élni. Kontrollpontképzés, a tranzakció-sorozat tárolása, valamint a hibajavítás költsége mérhető. A hibák bekövetkezése a kontrollpont-képzés időszakában nem lehetséges, a tárgyalás egyszerűsítése céljából sokszor azt is feltételezik, hogy a hibajavítás periódusa is hibamentes. A különböző valószínűségi változókra vonatkozóan exponenciális eloszlású feltevessel, illetve determinisztikus feltevessel szokás élni.

A kérdéskör vizsgálatában alapvető feladat a fent leírt tranzakciókat kiszolgáló rendszernek mint sztochasztikus rendszernek a vizsgálata, hogy mennyiben biztosítható a hibamentes működés a meghatározott időkeretek között. Ebbe a feladatkörbe tartozik a megbízhatóság vizsgálatának kérdése és a különböző állapotokban (mint a kiszolgálás, visszaállítás, kontrollpont-képzés állapotában) való tartózkodás stacionér valószínűségeinek meghatározása.

Annak vizsgálatához, hogy milyen ingadozások történhetnek a feldolgozásban — meglepő módon —, nem független valószínűségi változók sorozataira, illetve sztochasztikus folyamatokra vonatkozó centrális határeloszlás tétel felhasználása szükséges. A különböző közelítéseknel és megbízható számításoknál — ma már nem meglepő módon — az időléptékek megválasztása, éppen a felmerülő nagy költségek miatt, igen jelentős gyakorlati felhasználással is bír. Mint érdekes és lényeges következményt említem meg, hogy az időléptékek megválasztása útján a feldolgozási folyamatoknál a diffúziós folyamatok, amelyek a fizikában és a híradástechnikában természetes módon vetődnek fel, az informatikai feldolgozásban is előtérbe kerülnek.

Az adatbázis-kezelő rendszerek biztonságos kezelésének fenti példája mutatja, hogy az informatikai rendszerben a matematikai eszközök, a matematikai módszerek felhasználása terén lényeges előrehaladás történt.

Visszatérve eredeti feladatunkra, ha formális valószínűségszámítási megfontolásokat használnánk, azt tapasztalnánk, hogy az 1. ábrán leírt rendszer hibátlan működésének valószínűségére egyszerű feltételezések alapján a következőt kapjuk:

— A rendszerterv hibátlanságának valószínűsége 10^{-3} — 10^{-4} nagyságrendű (egy 5-600 oldalas rendszerterv elkészítése esetén).

— A programrendszerterv hibátlanságának valószínűsége egy több programból, tízezer utasításból álló rendszer esetén ugyancsak 10^{-3} — 10^{-4} körüli érték.

— A programrendszer hibátlansága, mivel utasításokról van szó, mintegy 10^{-4} — 10^{-5} nagyságrendű és 10^5 nagyságrendű adatrendszer esetén az adatok hibátlanságának valószínűsége ugyancsak 10^{-4} — 10^{-5} nagyságrendűnek tételezhető fel.

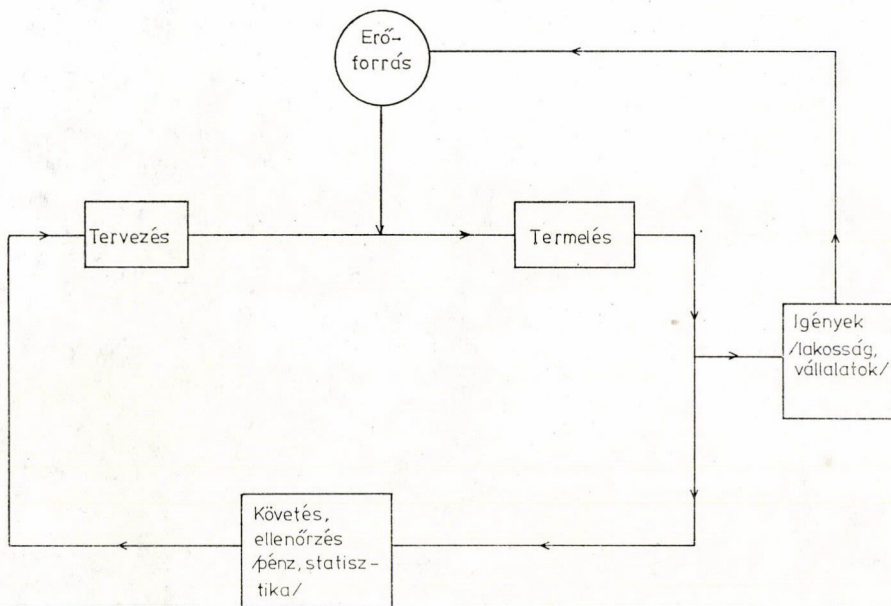
Feltételezve, hogy a különböző rendszerek egymástól függetlenül valósulnak meg, az egész rendszer hibátlan megvalósításának valószínűsége 10^{-13} nagyságrendű. Ez pedig azt jelenti, hogy gyakorlatilag hibátlan rendszer nem alakítható ki. Ennek ellenére pár hónapos együttes hibakereséssel, a rendszer javításával, ha nem is hibátlan, de működő rendszerek alakulnak ki. Egy ilyen jól (stabilisan) működő rendszer kialakulásának (ami nem a hibátlansággal azonos) vizsgálata az, ami matematikai szempontból modellezhető és alapvető fontossággal bír. A modell kialakítása csak a hibák megjelenésére vonatkozó feltételezések alapján képzelhető el. Ma még az egész rendszerre vonatkozóan ez nem várható, azonban egyes részekre, mint ahogyan az adatbázis-kezelő rendszerekkel kapcsolatban a fentiekben szerepelt, eredmények várhatók.

3. A körülmények változásai

A rendszerek működésével kapcsolatban a következőt kell még figyelembe venni. Mivel a rendszerek kialakítása több éves munka eredménye, a kialakítás folyamata alatt (pl. olyan rendszereknél, mint a bér- és munkaerő-gazdálkodás) megjelennek vállalaton belüli és országos rendelkezések, változások. Ezek a változások az egész rendszer működése szempontjából mint perturbáló tényezők jelennek meg. A rendszer kialakításakor legtöbbször lehetőség van a rendszer változtatására és a rendszer átalakítására. Amikor bevezetett rendszerrel van szó, ezeknek a perturbáló tényezőknek a megjelenése a követés és karbantartás problémáját érinti. Tapasztalataink szerint még az elnevezések, az ún. szótárak rögzítésében jelentkező új rendelkezések is évente több lényeges változtatást eredményeznek. A most említett perturbáló hibatényezőknek a figyelembevétele a rendszereknél úgy értendő, hogy azoknak rugalmasaknak kell lenniük a külső behatásokkal szemben.

Egy vállalati, illetve népgazdasági szintű irányítási rendszer leegyszerűsített modelljének a vizsgálata, amint az a 2. ábrából látható, a rendszert időben változóknak tekinti. Kiindulva a tervezési rendszerből, előbb figyelembe vesszük a lehetséges erőforrásokat, majd a végrehajtás, szervezés egy döntési rendszerbe jut. A munka eredménye a lakosság és más intézmények rendeléseinek, igényeinek kielégítése.

Mind a végrehajtásról, mind a termelés folyamatáról, mind az igényekről, az igényeknek az erőforrásokra való visszahatásáról elszámolási rendszerek készülnek, amelyek lehetnek statisztikai, pénzügyi, munkaügyi, beruházási típusúak. A visszacsatolás a tervezési rendszerhez rendszeres időközökben megtörténik, ahol a tervek módosítására kerül sor. Így alakul ki egy időben dinamikusan változó rendszer, amelyben a tervezés végrehajtása az elszámolási rendszerek, a teljesítési rendszerek figyelembevételével történik. Vállalatoknál a visszacsatolást a termékgyártás vagy termékkövetés, a pénzügyi, statisztikai elszámolási rendszerek szerves beépülése jelenti. A visszacsatolási periódus — ami a híradástechnikában másodpercek, esetleg percek alatt történik — népgazdasági, vagy vállalati szinten havi, éves, öt éves szín-



2. ábra

ten valósul meg. Amennyiben az adatszolgáltatás felgyorsulása lehetővé teszi a sűrűbb visszacsatolást, operatív irányítás megvalósítása válik lehetségessé. A vállalati elszámolási és irányítási rendszerekben a napi, illetve a műszakra vonatkozó operativitás jelenti a legfontosabb megoldandó feladatot ma *Magyarországon*.

A népgazdasági és vállalati irányítási rendszereknek egy ilyen leegyszerűsített modellje feltételezi, hogy a tervezési rendszer és irányítási rendszer részére az adatok biztosítása megfelelő időben történik. Ebben a rendszerben nem engedhető meg nagy elcsúszás, hiszen 1 hónapos késés az adatok biztosításában a rendszer stabilis működését akadályozza meg. A hónapra, vagy naprakész adatok biztosítása jelenti az informatika egyik alapvető feladatát. Vállalati irányítási rendszereinkben jelenleg még csak a hónapos, illetve negyedéves követési, elszámolási, gazdálkodási rendszerekben gondolkozunk.

Összehasonlításként említem, hogy a számítógépek lehetőségei az aritmetikai, műszaki számítások, a közgazdasági számítások végzésében jelentősen megnöttek; 10^8 nagyságrendű másodpercenkénti műveletszám elvégzésére is alkalmasak. Ez azonban az adatfeldolgozás, az informatika szempontjából még keveset nyújt, annyit, hogy bizonyos leszámolásokat gyorsabban tudunk végezni. Az informatikai feldolgozásokban a különböző sorrendek meghatározása fontos és egy 100 elemből álló halmaz különböző lehetséges leszámolásainak, különböző sorrendben történő előállításainak száma mintegy $3 \cdot 10^{100}$ nagyságrendű műveletet igényelne másodpercenként. Ezeket a nagyságrendeket még a legújabb gépek sem tudják produkálni.

Befejezésül egy gyakorlatból vett példát mutatok számításaink, becsléseink megbízhatóságára. Egy automatizált bér- és munkaerő-gazdálkodási rendszert vizsgáltunk meg, melyet több éves munkával alakítottak ki. Egy 5—10 ezer embert foglalkoztató

üzem bérszámfejtésében dolgozók száma mintegy 20 fő, havi bérük 50—60 ezer Ft. A feldolgozási gépidő mintegy 20—24 óra havonta, melynek ára 200—240 ezer Ft. Ugyanez a gépóra egy másik országban 5 ezer pénzegységbe kerül, és 2 bérszámfejtő havi illetménye ugyanott 4500—4800 pénzegység. Magyarországon, úgy tűnik, még 20 bérszámfejtő esetén sem gazdaságos a gépi igénybevétel, viszont a kézi feldolgozásra emberi kapacitást már ma sem lehet biztosítani, így a gépi feldolgozás mindenképpen előremutató és hasznosnak ítélandó.

A rendszer mintegy másfél-kétéves üzemeltetésének tapasztalatai alapján kb. 2-3 hónapos ráfordítással sikerült a 20—24 órás feldolgozási időt a felére csökkenteni. Ez a példa mutatja, hogy egyetlen rendszer eredményeit az országra kivetíteni, statisztikai adatként kezelni nem szabad. Másrészt felvetődik a kérdés, hogyan kezdjük az ilyen rendszerek bevezetését? Előre lefektetett 8—10 órás célkitűzéssel, vagy egy lassabban működő, esetleg 20—24 órát is felhasználó rendszer bevezetésének kipróbálásával? Mindenképpen az utóbbi megoldást kell választani, hiszen a rendszer megbízható üzemeltetése még 20—24 órában is elképzelhető, és a későbbiekben kell betervezni a rendszer működési idejének a lerövidítését.

Alapvető tanulság, hogy a jelenlegi periódusban az információs folyamatok milyenségéről, bonyolultsági fokáról, mechanizmusáról, automatizálható elemeiről csak működő rendszerek mélyreható vizsgálata alapján győződhetünk meg, ekkor ismerhetjük meg a fellépő hibákat is. Az informatika számítástudományi és matematikai kérdéseinek egyetlen, konkrét oldalát vizsgáltam meg. Egy előadás keretében nincs lehetőség annak megmutatására, hogy milyen matematikai módszerek használhatók általában, mindezt speciális szakmai munkák kell, hogy elvégezzék. A felvetődő feladatok mély analitikus eredményeket, logikai-algebrai megfontolásokat, számelméleti vizsgálatokat és nem utolsósorban statisztikai szemléletet igényelnek. Az eredmények sokszor hardware és software megoldásokként jelentkeznek, így segítve az információs folyamatok vitelében az egyre javuló számítástechnikai szolgáltatásokat.

(Beérkezett: 1978. május 11.)

ARATÓ MÁTYÁS
SZÁMÍTÓGÉPALKALMAZÁSI KUTATÓ INTÉZET
1536 BUDAPEST, POSTAFIÓK 227.

ON COMPUTATIONAL AND MATHEMATICAL PROBLEMS OF INFORMATICS

M. ARATÓ

This is a survey paper. It was presented on the general meeting of the *Hungarian Academy of Sciences* at 11th May, 1978.