Logelemzés
– avagy megfejthető-e emberi közreműködés nélkül az informatikai logokba kódolt intelligencia?

FABIÁNYI GÁBOR, FRÉSZ FERENC, SZABÓ LÁSZLÓ, ZSILINSZKY SÁNDOR

KÜRT Zrt.
{gabor.fabianyi, ferenc.fresz, laszlo.szabo, sandor.zsilinszky}@kurt.hu

Kulcsszavak: informatika, korrelációs logelemzés, monitoring, forensics

Mottó: „A bölcsesség egyik titka, hogy tisztában lenni azzal, mit kell figyelmen kívül hagyni."

Az adatok gyűjtése és elemzése egyidős az emberi civilizációval. Napjainkra a számítógép forradalmasította ezt a tevékenységet, de ugyanakkor önmaga is komoly problémák forrásává vált. Ezek jelentős részének megoldásához szükséges az informatikai rendszerekben lezajló folyamatokat, eseményeket rögzítő naplóbejegyzések, logok mélyreható elemzése. A logelemzés a vállalatok menedzsmentje számára nagy jelentőséggel bíró, egyéb információkat is képes szolgáltatni, segítségével jövőbeli trendekre is lehet következtetni. Az IT rendszerek azonban óriási mennyiségű logot termelnek, melyek adekvát feldolgozása a normál üzemeltetés keretei között lehetetlen. Megjelentek tehát a piacon a különböző megoldások, melyek közül a legnagyobb hozzáadott értéket a humán intelligenciával támogatott logelemző szolgáltatás adja.

1. Bevezetés

Divatos kifejezés manapság a logelemzés. Az utóbbi években a szakmai körökön túl szélesebb rétegek is megismerkedhettek a fogalommal a különböző, informatikához kötődő közéleti botrányok nyomán. Tény azonban, hogy sok félreértelmezés és tévhit kapcsolódik e témához, így hát érdemes alaposabban körüljárni, mit érdemes tudni róla, mire használható pontosan és milyen előnyöket kínál.

A logelemzésről elmondható, hogy az informatika jelenének és jövőjének egyik legnagyobb jelentőségű eszközrendszere.

A különböző adatok gyűjtése és értékelése már akkor is kritikus része volt az élet számos területének, amikor az elemzést még nem támogatta modern technológia. A historikus adatok értékelése visszatekint egészen az ókori időkre, a sumérok korára, akik például összegyűjtötték a termésmennyiség-adatokat és a termés megfelelő elosztására használták fel azokat. Egy másik példa a dél-amerikai Inka Birodalom, ahol a növényi kultúrákat analizálták bizonyos gazdálkodási trendek és minták meghatározásához, melyhez adatrögzítő módszerként a quipu névre hallgató csomóírást alkalmazták. (Érdekesség, hogy ezt az írásrendszert szokás háromdimenziós kettes számrendszernek is hívni.)

Ha kicsit ugrunk az időben, érdekes példát találunk az 1880-as USA népszámlálás idején. A népszámlálás adatainak feldolgozása és szerkesztése 9 évig tartott, ebből 7 teljes évet vett igénybe a statisztikai analízis.

Az adatfeldolgozást és elemzést a számítógép 20. századi megjelenése forradalmasította. 1952-ben az első számítógépek egyikén, az UNIVAC-on készült az első számítógépes előrejelzés az USA elnökválasztás várható kimeneteléről. A számítógép a CBS előrejelzésével ellentétben Eisenhowert jósolta a választások győztesének és igaza is lett.

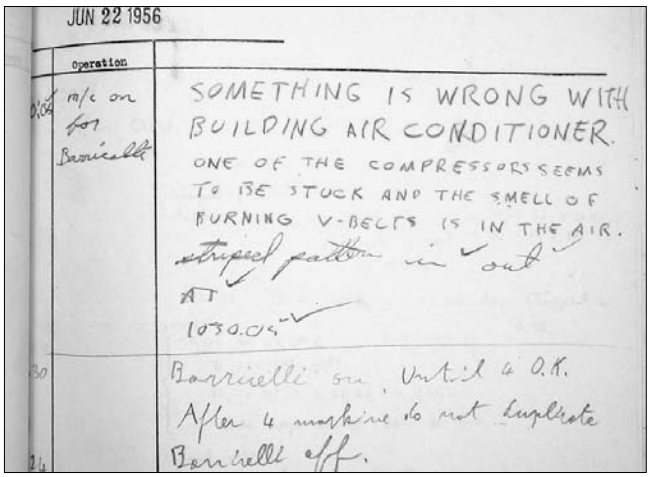
A tömeges alkalmazás elterjedésében az adatok számítógéppel végzett elemzésének üzleti célú alkalmazása, a döntéstámogatás segítése volt az igazi fegyvertény. 1989-ben nevezték el a módszert üzleti intelligenciának és írták le a koncepcióját, metodológiáját. Az üzleti intelligencia tökéletesítette a döntéshozatalt a tényalapú támogatási rendszernek köszönhetően. A prediktív analízis, vagyis az előrejelző vizsgálati módszertan 1999-ben debütált. Az üzleti életben nagy jelentősége van annak, hogy a meglévő adatokból következtetni tudjunk a jövő trendjeire, ez az egyik alapvető funkciója egy teljes értékű informatikai logelemzésnek is, mint azt a későbbiekben látni fogjuk.

A számítógép megjelenése nem csak azt eredményezte, hogy az adatfeldolgozás sebessége ugrásszerűen megnőtt, így az elemzések soha nem látott komplexitással és hatékonysággal lettek elvégezhetők, hanem egyben a számítógép maga is problémák forrásává vált, többek között a saját maga által generált, hatalmas mennyiségű adat miatt.

2. A log

A fenti példák egyértelműen igazolják, hogy az adatok összegyűjtése és elemzése minden korban rendkívül fontos szerepet játszott. Napjainkban az információs rendszerek megfelelő működésének és fejlődésének biztosításához a rendszerek naplóadatainak elemzése szolgáltat megfelelő alapot.

A naplózás alkalmas arra, hogy feltérképezzük a rendszerben felmerülő problémákat, időrendbe állítsuk azokat, megtaláljuk a lehetséges megoldásokat, kidolgozzuk az elhárítási terveket, s azok segítségével végül felülkerekedjünk a problémákon, mi több, az elemzett adatainkból következtetéseket vonjunk le a jövő fenyegetéseinek elkerülése érdekében.


JUN 22 1956

Operation

m/c on for Barnicelli

SOMETHING IS WRONG WITH BUILDING AIR CONDITIONER. ONE OF THE COMPRESSORS SEEMS TO BE STUCK AND THE SMELL OF BURNING V-BELTS IS IN THE AIR.

stripped pattern in out

AT 1030.04

Barnicelli on Until 4 O.K.

After 4 machine do not duplicate

Barnicelli off.

1. ábra Kézi naplózás az ötvenes évekből

A log a számítógép naplóbejegyzése. Amióta számítógép létezik, azóta létezik log is. Kicsit leegyszerűsítve a definíciót, a log nem más, mint a számítógép által generált naplóbejegyzés valamilyen, a számítógép működése közben megtörtént eseményről és annak fontosabb paramétereiről.

A számítógépek működése során rengeteg esemény következik be. Minden egyes esemény változást jelent az adott rendszer vagy eszköz állapotában. Számítógépes biztonsági szempontból az esemény egy tevékenység eredménye, mely egy adott cél elérésének érdekében történik.

A számítógép és a rajta futó szoftverek szimbiózisa meglehetősen bonyolult, komplex rendszer, sok beavatkozási ponttal, s persze rengeteg hibalehetőséggel. A rendszer az emberi felfogóképességhez képest rendkívül nagy sebességgel működik. Ennek köszönhetően az embernek (a számítógépkezelőnek) esélye sincs az eseményeket, esetleges hibákat, rossz működést vagy valamilyen, a működésre jellemző fontos paraméter változását valós időben, működés közben észlelni, nemhogy kezelni.

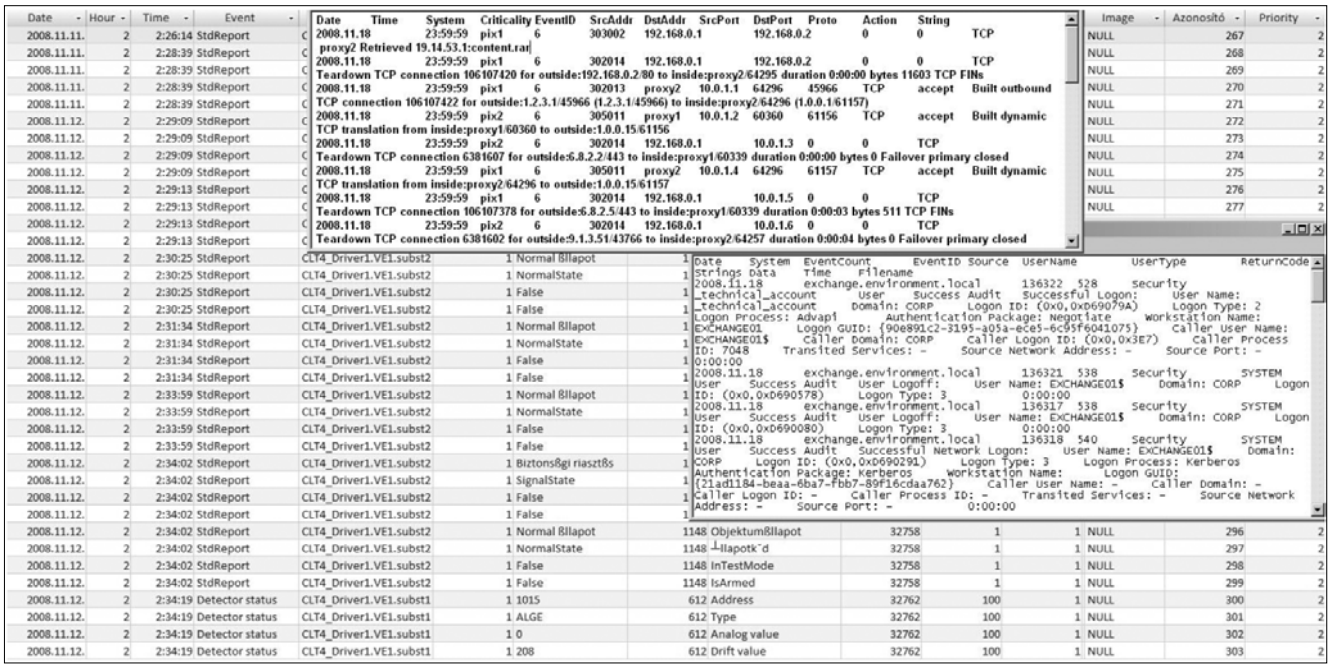
Ugyanakkor, ha ezekről az eseményekről, történésekről a számítógép vagy a rajta futó szoftverek készítenek egy-egy feljegyzést, akkor a feljegyzések alapján később visszakövethetővé, értelmezhetővé válik, hogy mi minden zajlott le a gépben, mi okozott hibát, leállást, biztonsági krízishelyzetet stb. A naplófájl a számítógépes események fontosabb paramétereit automatikusan rögzíti és olvashatóan megjeleníti, így lehetővé teszi azok utólagos ellenőrzését, elemzését.

Önmagában véve egyetlen számítógép is figyelemre méltó mennyiségű logot tud generálni, ám a számítógép-hálózatok minden korábbi várakozást felülmúló elterjedésével a szakembereknek ma már igen nagy méretű és bonyolultságú, komplex informatikai rendszerekkel kell megbirkózniuk. Ezek működése hihetetlenül összetett, így nagyságrendekkel több a hibalehetőség, a valamilyen reakciót, beavatkozást igénylő rendszerállapot-változás, amit kezelni kell. Elengedhetetlen tehát, hogy az eseményekről, hibákról, állapotváltozásokról valamiféle visszajelzést kapjunk naplóbejegyzések formájában.

Az informatikai rendszerek főbb alkotóelemei, az operációs rendszerek, alkalmazások, különféle szoftver és hardver komponensek működésük során mind, mind naplófájlokat generálnak, milliónyi naplóbejegyzéssel, loggal. Az információbiztonsági szempontok megkövetelik, hogy a felhasználók tevékenységéről, a rendszerek működéséről, a hozzáférési jogosultságok változásairól szóló bejegyzések folyamatosan követhetőek legyenek.

Az adatoknak azonban önmagukban nincs jelentésük. Az adatok az értelmezéstől, a feldolgozás módjától, alkalmazásuktól nyernek értelmet és válhatnak értékes információvá.

2. ábra „Gépi” naplózás fél évszázaddal később



Date	Hour	Time	Event	
2008.11.11.	2	2:26:14	StdReport	C
2008.11.11.	2	2:28:39	StdReport	C
2008.11.11.	2	2:28:39	StdReport	C
2008.11.11.	2	2:28:39	StdReport	C
2008.11.11.	2	2:28:39	StdReport	C
2008.11.12.	2	2:29:09	StdReport	C
2008.11.12.	2	2:29:09	StdReport	C
2008.11.12.	2	2:29:09	StdReport	C
2008.11.12.	2	2:29:09	StdReport	C
2008.11.12.	2	2:29:13	StdReport	C
2008.11.12.	2	2:29:13	StdReport	C
2008.11.12.	2	2:29:13	StdReport	C
2008.11.12.	2	2:30:25	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:30:25	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:30:25	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:30:25	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:31:34	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:31:34	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:31:34	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:31:34	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:33:59	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:33:59	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:33:59	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:33:59	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:02	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:02	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:02	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:02	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:02	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:02	StdReport	CLT4_Driver1.VE1.subst2
2008.11.12.	2	2:34:19	Detector status	CLT4_Driver1.VE1.subst1
2008.11.12.	2	2:34:19	Detector status	CLT4_Driver1.VE1.subst1
2008.11.12.	2	2:34:19	Detector status	CLT4_Driver1.VE1.subst1
2008.11.12.	2	2:34:19	Detector status	CLT4_Driver1.VE1.subst1

Image	Azonosító	Priority
NULL	267	2
NULL	268	2
NULL	269	2
NULL	270	2
NULL	271	2
NULL	272	2
NULL	273	2
NULL	274	2
NULL	275	2
NULL	276	2
NULL	277	2
NULL	296	2
NULL	297	2
NULL	298	2
NULL	299	2
NULL	300	2
NULL	301	2
NULL	302	2
NULL	303	2

A napjaink informatikai rendszereiben keletkező logok mennyisége messze meghaladja azt a mértéket, ami hagyományos rendszergazdai vagy üzemeltetői kompetenciát és erre fordítható időt feltételezve emberléptékűnek és feldolgozhatónak nevezhető, a hatékonyságot nem is említve. Pedig éppen a rendszerek komplex mivolta követelné meg még inkább a naplóadatok rendszeres, alapos elemzését. A naplófájlok vizsgálatával és értelmezésével a rendszer működése rekonstruálható, az abban előállt rendellenességek és egyéb nem várt események forrásai felkutathatóak, különös tekintettel például a rendszert érő külső és belső támadásokra. Később erről részletesebben is írunk.

Konkrét példát nézve, egy átlagos, közepesen terhelt tűzfal naponta kb. 2 és fél millió logot termel. Ebből gyakorlatilag néhányszor tíz, esetleg néhányszor száz log az, amely konkrét és releváns információt hordoz, és nagyjából egymillió a figyelmen kívül hagyható bejegyzés. Ahhoz viszont, hogy a konkrét jelzésekből trendekre, a körülményekre, az ok-okozati összefüggésekre következtetni lehessen, meg kell vizsgálni a maradék másfél millió logot is.

3. A logelemzés

A logelemzés segítségével feldolgozásra kerülnek a különböző formátumú, struktúrájú és forrású események, így lehetővé válik az informatikai rendszer állapotával kapcsolatos kép rekonstruálása és riportokba foglalása.

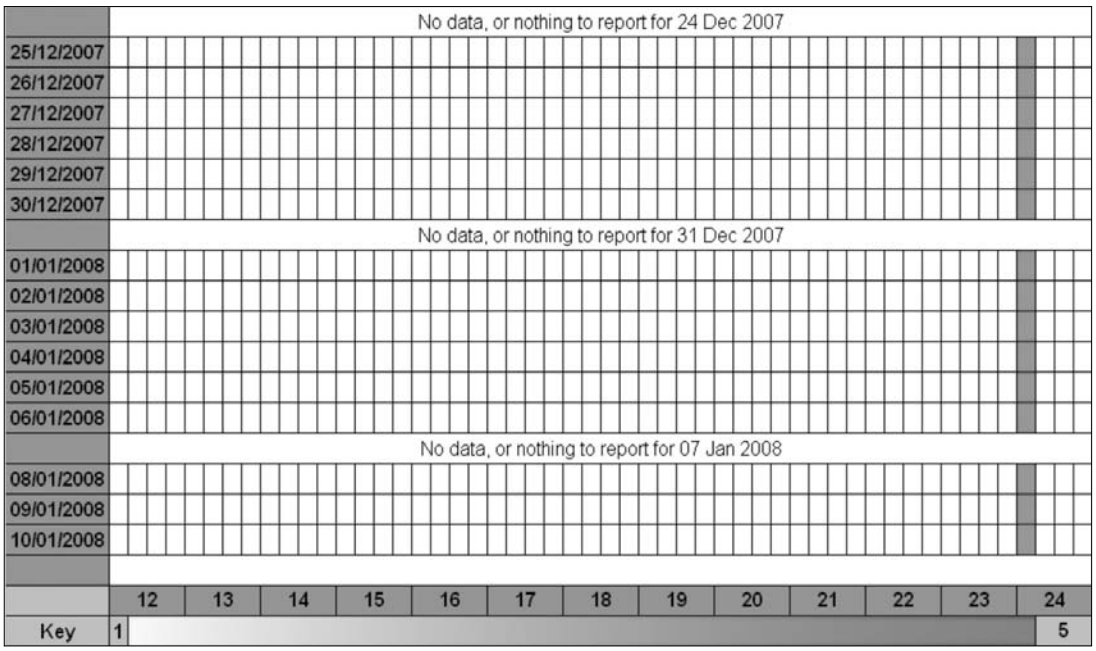
Mivel a rendszerek elképesztő mennyiségű logot generálnak, a rendszerüzemeltetők gyakran éreznek indíttatást a logok mennyiségének minimalizálása. Ezzel sok esetben halálra ítélik a feldolgozás lehetőségét, hiszen az események úgy csökkenthetőek a leghatékonyabban, ha a bejegyzések legnagyobb részét alkotó, informális üzeneteket kikapcsoljuk. Ezzel azonban elvágjuk magunkat attól a lehetőségtől, hogy egy vizsgálandó incidens esetén az azzal összefüggő, azt leíró eseményeket, körülményeket együttesen vizsgáljuk.

Az általunk incidensnek nevezett, nem kívánt események nem csak úgy „lesznek”. Nincs olyan esemény, amely egymagában képes romba dönteni jól felépített és féltve őrzött rendszereinket, hogy aztán az ismeretlenség jótékony homályába húzódva várakozzon a helyreállítási műveletek befejezéséig, hogy azt követően aztán újra lecsaphasson.

Az informatikai hálózat a komplex rendszerekre jellemző módon az elemek közötti összefüggésekkel írható le a legjobban. Ebből következően az incidenseket mindig megelőzi a rendszerkörnyezetben bekövetkező „viselkedésbeli” változás. Bizonyos események száma megnő, míg ezzel egyidőben másoké csökken. Ha csak a kritikus hibákra vagy figyelmeztető üzenetekre koncentrálunk, gyakran éppen a lényeget tévesztjük szem elől. A tettes felkutatásához kevés lesz pusztán a nyitva hagyott ablak és az üres ékszeres doboz közötti összefüggés vizsgálatára hagyatkoznunk.

Az elmúlt 5 év gyakorlati tapasztalatainak összegzéseképpen 2007-ben elkészítettük a logelemzés ökölszabályain alapuló ötlépcsős elemzési modellt, amelyben nagyvonalakban rögzítettük a logokkal kapcsolatos teendőket. Az öt lépcső: a gyűjtés, a normalizálás, az értelmezés, az elemzés és a riportolás.

Az első és legfontosabb a feldolgozandó információk összegyűjtése egy központba. Az adatokat ezután különböző tulajdonságaik alapján egységes, közös struktúrába rendezzük, ami biztosítja, hogy az adatbázis mezőiben homogén adatokat találjunk. (Ilyenek például a keletkezés dátuma, ideje, helye, típusa stb.) Az értelmezési fázisban kutatjuk a bejegyzések jelentését, melyek alapján emberi, de legalábbis elemzői fogyasztásra alkalmassá válnak. A negyedik, elemzési fázis a tényleges adatbányászat, az összefüggések és korrelációk felderítésének ideje, itt kerülnek a helyükre a kirakós elemei. Az utolsó munkafolyamat a riport elkészítése, ahol



3. ábra
Éjszakai mentés hibamintázata egy hőtérképes megjelenítőn

A hőtérkép kifejezés itt nem az eredeti jelentésével értendő, a színskála (az ábra legalján lévő csík) szerinti érték-megjelenítésre utal. A színskála színei 1-től 5-ig terjedő értékeket mutatnak, ami az adott időintervallumban bekövetkező események számát jelzi. Az ábrán a mentés hibajelzése látható, ami a 23h-ás mindennapi mentéskor megjelenik, vagyis amikor mentés történik, akkor az hibára fut.

a feltárt jelenségeket tényadatokkal és grafikonokkal alátámasztva dokumentációvá alakítjuk.

A logelemzést szakmai körökben információs hulladék-újrahasznosításnak is hívjuk, mivel a log az informatikai rendszerek működésének szükségszerű mellékterméke, ahogy a hulladék mellékterméke az emberi fogyasztásnak. Ezt a mellékterméket megfelelően kezelve értékes nyersanyagokhoz juthatunk. Az elemzési modell fázisai logikusan egymásra épülnek, pont úgy, ahogy a hulladékgyűjtő telepeken a beérkező anyagok kezelését, válogatását, továbbítását, csomagolását és elszállítását szabályozó munkafázisok.

4. Mivel elemezzünk?

Melyik a legjobb eszköz a logok elemzéséhez? Milyen szoftvert/hardvert válasszunk? Mint ahogyan az általában lenni szokott, itt sem létezik olyan univerzális termék, ami mindenben a legjobb. Vannak azonban jól bevált részmegoldások, melyeket megfelelően ötvözve hatékony gyűjtő és elemző rendszerek alakíthatóak ki.

A logelemzés egyik alapvető problémája, hogy szabványos logformátum mint olyan, egyáltalán nem létezik. Ahány rendszer, annyi féle felépítésű naplóbejegyzéssel és tárolási formátummal találkozhatunk. Vannak szövegfájlként, bináris adatfájlként és adatbázisrekordként tárolt bejegyzések, valamint teljes a skála az egyszerű egysoros jól strukturált bejegyzéstől a több sort kitöltő, dinamikusan változó tartalmúig. A naplóbejegyzések egyaránt tartalmazhatnak hagyományos alfanumerikus, vagy speciális írásjeleket, illetve a bináristól a hexadecimális skáláig terjedő numerikus értékeket is. Minél átfogóbb, teljesebb körű megoldást szeretnénk kialakítani, annál komolyabb, összetettebb problémákkal találjuk magunkat szemben.

A gyűjtés legelterjedtebb módja a syslog protokollon alapuló átirányítás, amely még a UNIX kezdeti időszakából származik és bizony mára jócskán eljárt felette az idő. Nagy előny viszont, hogy többé-kevésbé minden rendszer támogatja.

Változást jelent napjainkban, hogy egyre több gyártó és felhasználó szervezet kezdi felismerni, hogy a logok kezelése legalább annyira fontos (sőt, sok esetben fontosabb!), mint maguk a rendszerekben tárolt adatok. Emiatt a gyors, ámde nem garantált UDP protokoll helyett megjelentek a TCP protokollt és azon felül titkosítást is alkalmazó gyűjtő technológiák, de igen gyakori a fájlszintű másolás is.

Szakembereink, a komplex, elosztott terhelésű logadattárházak építését látják a leghatékonyabb megoldásnak. Ennek méretezésekor úgy kell kalkulálni, hogy a keletkező logok gyűjtésén és hosszú távú tárolásán felül az adatbányászatra is megfelelő lehetőség nyíljon. Az értelmezési feladatok és műveletek elvégzéséhez nélkülözhetetlen, de sokszor nem kis nehézséget jelentő normalizálás során a logok származási hely és formátum szerint azonos struktúrába kerülnek, így olvasás helyett inkább a számolási képességünkön van a hangsúly.

A logok bányászatához szükséges OLAP (On-Line Analytical Processing – valós idejű adatelemzés) adatbázisok alapjául, a kereskedelmi, pénzügyi elemző rendszereknél megszokotthoz hasonlóan, a produktív rendszerekből kinyert információk szolgálnak. Ezeket különböző szempontok szerint rendezve vizsgáljuk.

5. Logelemző intelligencia

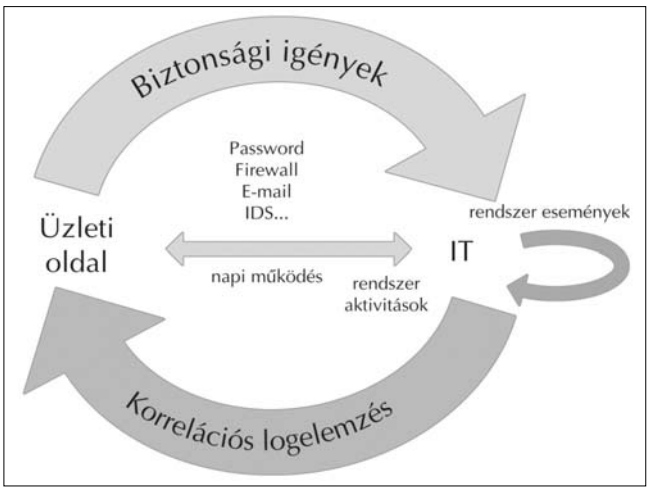
A piacon fellelhető logelemző termékek legfőbb ismérvei, hogy képesek az általánosan elterjedt rendszerek logjainak gyűjtésére, tárolására, archiválására, rendelkeznek néhány törvényi vagy jogszabályi megfelelőségi paraméterrel és ezekre optimalizált előre definiált riportokkal (GLBA, HIPAA, PCI, SOX stb.) Ezeken túl a riportok személyre szabhatóak a felhasználó szájíze szerint, tartalmaznak valamilyen riasztási funkciót és képesek az események közötti egyszerűbb korrelációk megvilágítására.

Ezek a logelemző termékek azonban nem oldják meg teljeskörűen a logelemzés problémáját, a szervezetek menedzsmentje és az informatikát üzemeltetők számára a logelemzésben rejlő széleskörű lehetőségeket és előnyöket csak kismértékben használják ki.

A gyártók fejlesztési tervei az általuk ismert univerzumból indulnak ki, mely azonban sok esetben hiányos. Hatékony terméket létrehozni egy ilyen kaotikus, szabványmentes környezetben nagyon nehéz, majdhogynem lehetetlen feladat. A túl sok logformátum támogatásával a rendszer egésze lassú lesz, a döntési mechanizmusok pedig rendkívül bonyolulttá válnak. Ha viszont kizárólag az elterjedt formátumok támogatására fókuszálnak, akkor a fejlesztések sablonos logmonitoring rendszerek létrehozásába fulladnak, amelyek csak igen korlátozott logelemző funkcionalitással bírnak.

Tapasztalataink alapján jelenleg nem létezik egyetlen olyan kész logelemző termék sem, amely egy szervezet logelemzési igényeit képes lenne maradéktalanul kielégíteni. A különböző területeken alkalmazott részmegoldások és az elemzéshez használt emberi intelligencia,

4. ábra
Üzleti oldal és logelemzés kapcsolata (forrás: KÜRT Zrt.)



illetve döntéshozatali mechanizmusok ötvözésével és segítségével azonban képesek vagyunk hatékony elemző rendszerek kialakítására és üzemeltetésére. Nem szabad azt sem elfelejteni, hogy az elemzői tevékenység a rendszerüzemeltetőkétől távolabbi, holisztikus nézőpontot igényel. Ennek a szemléletnek a hiánya lehet az oka annak is, hogy viszonylag magas számú, félresikerült termékbevezetésre van példa a piacon.

A logelemzés szükségességét egyre több szervezet ismeri fel, a megoldást azonban csak kevesen ismerik. A megoldás egy olyan termékfüggetlen logelemző szolgáltatás, ahol a rendszer felmérését követően optimális, folyamatos szolgáltatás megvalósítása a cél. Az alkalmazott szoftverek körét minden esetben az ügyfél rendszerének ismeretében alakítják ki a szakemberek és annak érdekében, hogy legkönnyebben juthassanak a megfelelő információk birtokába, saját fejlesztésű szoftver-komponensekkel egészítik ki ezt a kört. A gyakorlat eddig számos esetben igazolja, hogy a hatékony logelemző rendszer legfontosabb láncszeme a beégetett algoritmusoktól mentes, szabad döntések meghozatalára és intuícióra képes humán elemző.

A lehetőségeket jól kiaknázó, humán intelligenciával támogatott logelemző szolgáltatás az informatikai rendszer biztonsági szintjének fenntartásában betöltött nélkülözhetetlen szerepe mellett mérhetővé teszi az üzleti oldal számára az informatikai szolgáltatásokat, beruházásokat, fejlesztési igényeket és az informatikai rendszer sokszor átláthatatlannak tűnő működését. Erről lesz szó részletesebben a következő szakaszban.

6. Logelemzésre épülő szolgáltatások

Az informatikai rendszerek hőskorában, különösen a hálózatok kialakulásának kezdetén a logelemzés, mint informatikai „módszer”, önmagában nem létezett. A rendszerfejlesztők, alkalmazásfejlesztők különböző programrészeket „használtak” arra, hogy nyomon követhessék a programokban bekövetkező hibákat, állapotváltozásokat.

Az „áttörést” a hálózatos működéssel együtt megjelenő, többfelhasználós rendszerek megjelenése jelentette, mivel itt már azt is ki kellett mutatni, hogy mikor, ki használt egy-egy terminált, szolgáltatást. Az akkori informatikai rendszerek nem voltak annyira összetettek, mint a mai hálózatok, jellemzően célmegoldásokra használták azokat, a legtöbbször kutatók és programozók. A mai felhasználói réteg akkor még nem alakult ki, a nyomonkövethetőség igénye azonban már a kezdetek kezdetekor is létezett, mivel az első hálózatok és alkalmazások leginkább katonai célokra alakultak ki.

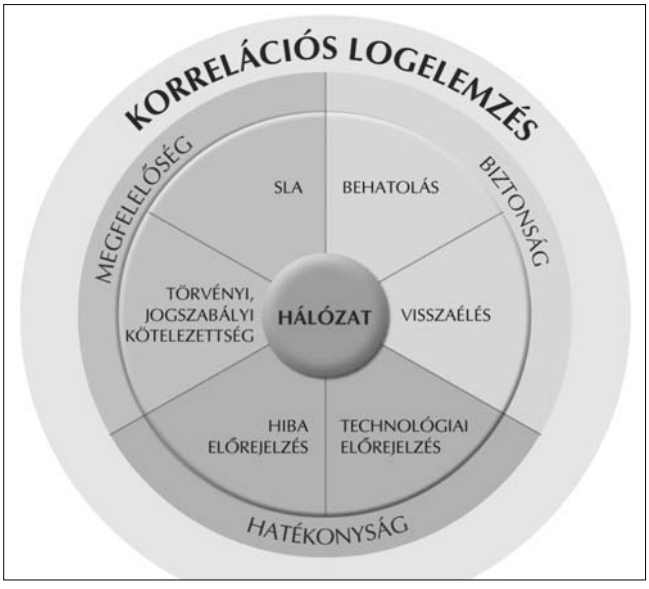
A nyomonkövethetőséget a nagy vállalati hálózatok és az internetes szolgáltatások megjelenése erősítette, de akkor ez kimerült a határvédelmi rendszerek és a szerverek erőforrásainak monitorozásában. A tűzfalak és szerverek eseményeinek monitorozását megoldani képes hálózat-monitoring, erőforrás-monitoring rendszerek elterjedésével a felhasználók valós idejű információkkal rendelkeztek az üzemeltetett hardver és szoftverpark pillanatnyi állapotáról, változásairól. A 90-es évek elejétől terjedő web exponenciálisan növelte meg a rendszerek működéséről kimutatható információéhséget, mivel itt bekapcsolódtak a vállalatok üzleti szintjei is.

A vállalati hálózatok növekedésével a bennük tárolt információ mennyisége elképesztő mértékben növekedett és ez a növekedés a mai napig tart. Az adatbázisokban egyre gyorsabban, egyre több adat tárolódott és megjelentek a vezetői információs igények is. Ennek kézbentartására aztán kialakultak a nagy ERP és HR rendszerek, a védelmi szinteken elindult az IDS-ek (Intrusion Detection Systems – behatolásjelző rendszerek) és a tűzfalmegoldások forradalma is.

A korábbi esemény-monitoring funkciókat rendre kiegészítették az adatokat historikusan kezelő modulok is, a múltbeli események „visszajátszására” képes programrészek. Az alapfelfogás mind a mai napig az, hogy a rendszereseményekből ki kell tudni válogatni a biztonságra, felhasználásra vonatkozókat és azokat „jelezni” kell tudni az üzemeltetők felé. Az üzleti alkalmazhatóság, a pénzügyi rendszerek és az on-line szolgáltatások elterjedése azonban a rendszeresemények „visszajátszhatóságát” is megköveteli. Így már nem elég csupán néhány kiválasztott esemény bekövetkezését jelezni, hanem a rendszerek eseményeit leíró adatokat, logállományokat el is kell tárolni. Kialakultak a loggyűjtést és tárolást megoldó központi szerverek. A hálózatok sebességnövekedése egy időn túl lehetővé tette a rendszer összes eseményének egy központi helyre való továbbítását is.

Az így kialakult megoldások számos lehetőséget biztosítanak a rendszerüzemeltetők, vállalatok számára. Az események statisztikai elemzése leginkább a vezetői információs rendszerekben jelenik meg, az üzemeltetés területén pedig az események mielőbbi jelzése a prioritás. Az on-line marketig térhódítása a rendszerfelhasználás, látogatottság-mérés, a pénzügyi-tranzakciók „videomagnószerű” visszajátszhatóságát, mérését célzó

5. ábra
A korrelációs logelemzés helye a szervezet információbiztonsági rendszerében (forrás: KÜRT Zrt.)



megoldások fejlődését indukálják, míg a megnövekedett biztonsági igények az internetes támadások, vírustámadások jelezhetőségét erősítik. A kialakult helyzetben ez a két fő irány szervesen elválasztásra került, így a pénzügyi, vállalatirányítási rendszerek a naplóállományok elemzésére, míg az üzemeltetői rendszerek az események jelzésére helyezik a hangsúlyt.

Az általunk kidolgozott megoldás e két területet együtt célozza meg, így a logok összegyűjtését követően, azok elemzésével mindkét terület számára képes olyan információkat kinyerni, amelyek korábban nem voltak elérhetők. A tűzfalak, IDS-ek, szerverek eseményeit nem önmagukban, hanem a vállalatok informatikai rendszereinek összes alkotóelemével együtt vizsgálják. Ezzel a megoldással lehetőség nyílik arra, hogy az informatikai vezetők, vállalatvezetők naprakész információkhoz jussanak az informatikai rendszer mindenkori állapotáról, annak viselkedéséről.

A napi logelemzés elsődlegesen az üzemeltetők számára biztosít információkat a rendszerhibák, anomáliák elhárításához. Ennek segítségével az üzemeltetőknek nem kell a logokban keresgélniük a problémák után, hiszen azt az elemzés elvégzi, ráadásul javaslatokat tesz a hibák elhárítására, pénzt és időt megtakarítva ezzel az amúgy is kislétszámú informatikai területeknek.

A szolgáltatás alapvető funkciója továbbá a rendszerek biztonsági szempontú elemzése. A rendszerek biztonságát fenyegető események kiszűrésével és elemzésével elkerülhetőek a tömeges hibás riasztások és rendkívül hatékony incidens-kezelés alakítható ki.

Az összegyűjtött és központilag tárolt naplóállományok segítségével a rendszerösszeomlások, csalások, visszaélések informatikai nyomai is feltárhatóak, így támogatva a forensics, azaz az események okait utólag felderíteni szándékozó, nyomozati jellegű vizsgálatokat.

A logállományok napi feldolgozásával az informatikai rendszerek rendelkezésreállás-mérése is kiválóan megoldható. A napjainkban elterjedt outsource megoldások, valamint a vállalatvezetők stratégiai rendelkezésreállás követelményei elengedhetetlenné teszik a komoly SLA-k (Service Level Agreement-ek) „bevállalását”, a rendelkezésreállás folyamatos biztosítását.

Leginkább a pénzügyi területeken, on-line szolgáltatások esetén van igény az úgynevezett fraud-management megoldásokra, ahol a logelemzés az esetleges visszaélések kivizsgálását képes támogatni.

Az informatikai vezetők folyamatosan harcolnak a megfelelő anyagi erőforrások biztosításáért, alapvetően az üzemeltetők információira támaszkodva. A vállalatok újabb és újabb szolgáltatások bevezetésével szeretnék bevételeiket növelni, amelynek következtében az informatikai rendszerek folyamatosan változnak. Időnként elavulnak, az új rendszerelemek fejlesztése komoly változásokat eredményez, a folyamatos változás pedig megnöveli a biztonsági kockázatokat. A logelemző szolgáltatás napi információkat képes nyújtani az elavult rendszerkomponensekről, a fejlesztés alatt álló alkalmazások, szegmensek hibáiról, valamint a komplex hálózatok pillanatnyi sérülékenységeiről is.

7. Jövőkép

A logelemzés fejlődése a jelenlegi tendenciák alapján a real-time logelemzés és a konvergencia irányába halad.

Folyamatosan változó világunkban a döntéshozók egyre pontosabb és egyre gyorsabb eredményeket követelnek, ezért az adatok elemzése, így a logelemzés is a valós idejű monitoring és elemzés megvalósítására törekszik. Ez egyben a két funkció egymáshoz való közelítését is jelenti, vagyis a monitoring és az elemzés, amelyek ma még elkülönülnek, egyre inkább összeolvadnak majd. Ez persze egyenesen következik abból, hogy a monitoring eleve valósidejű funkció és a tendencia az, hogy az elemzés is ebbe az irányba halad.

A logelemzésnél jelenleg elsősorban technológiai akadályai vannak annak, hogy real-time, valós időben folyhasson az elemzés. Ezek elsősorban a logok, naplófájlok azonnali hozzáférhetőségével, feldolgozhatóságával, értelmezhetővé tételével, normalizálásával vannak kapcsolatban. A technológiák fejlődésének iránya ugyanakkor egyértelműen azt mutatja, hogy a jövőben a real-time logelemzés kap majd egyre nagyobb hangsúlyt.

A konvergenciáról, ugyancsak elmondható, hogy a technológia számos területén varázsszónak számít. A logelemzés vonatkozásában a cél a fizikai és logikai biztonság területéről származó információk és logok közös platformon, együtt történő kezelése. Ezzel a módszerrel lényegesen hatékonyabbá és gyorsabbá válhat az informatika számtalan területéről, s a különböző biztonsági (beléptető, tűzvédelmi, behatolásjelző, zártláncú video stb.) rendszerekből érkező logok és jelzések értékelése. A KÜRT-nél ezzel a témával kapcsolatban konkrét fejlesztések folynak, ilyen értelemben a konvergencia már nem is annyira jövőbeli tendenciája a logelemzésnek.

A szerzőkről

FABIÁNYI GÁBOR 1987-ben végzett a BME Villamosmérnöki Karán. 12 éve dolgozik a KÜRT Zrt-nél, 10 éve marketingmenedzserként. Részt vett a cég információbiztonsági portfoliójának kialakításában, két évig szerkesztette a KÜRT Informatikai Biztonság című szakmai hírlevelét.

FRÉSZ FERENC Budapesten született, 2003-ban diplomázott, tanítói szakon. Újságíróként dolgozott, majd az informatikában indított oktatási, tanácsadói vállalkozást a 90-es évek közepén. Biztonsági szakértőként, majd a Budapest Airport informatikai vezetőjeként szerzett tapasztalatokat a vállalati rendszerek sérülékenységeiről, viselkedéséről. Számos vállalat IT vezetőjeként folytatta pályáját, amelynek keretében több mint 500 biztonsági projektet irányított. Az így szerzett tapasztalatai alapján dolgozta ki a KÜRT logelemzési és legális hackelési módszertanait. Jelenleg a KÜRT Zrt. Biztonsági Intelligencia Központjának vezetője.

SZABÓ LÁSZLÓ a KÜRT Információmenedzsment Megoldások Üzletág Logelemzés csoportjának szakértője. Több hazai vállalat logelemzés projektjében vett és vesz részt. Munkája során loggyűjtő és -elemző, illetve IDS rendszerek finomhangolásával, trendelemzéssel, valamint biztonsági események detektálásával és mintaelemzésével foglalkozik. Tevékenységi területe kiterjed a fizikai biztonsági terület és az informatikai rendszerek eseményeinek vizsgálatára is.

ZSILINSZKY SÁNDOR Budapesten született, itt végezte középiskolai tanulmányait, majd 1984-ben kapott diplomát a Budapesti Műszaki Egyetem Villamosmérnöki karán, Híradástechnika szakon. Az egyetem elvégzése után a hazai informatikai iparban helyezkedett el, mint hardverszakértő. Később a KÜRT-nél számos informatikai biztonsági fejlesztésben vett részt, mint például a cég Informatikai Biztonsági Technológiájának kifejlesztése (IBiT). Jelenleg üzletágvezetőként dolgozik a KÜRT Zrt-ben és aktív részese a logelemzési technológia továbbfejlesztésének, melyről több cikke és konferencia-anyaga is megjelent.