
Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar

Hálózati Rendszerek és Szolgáltatások Tanszék
Mobil Kommunikáció és Kvantumtechnológiák Laboratórium (MCL) és
CrySyS Adat- és Rendszerbiztonság Laboratórium

# Privátszféra Védelem Strukturális De-anonimizációs Támadásokkal Szemben Közösségi Hálózatokban

Ph.D. tézisfüzet

**Gulyás Gábor György**

Konzulens:

Dr. Imre Sándor

2015

# 1 Bevezetés

A közösségi médiát naponta több száz millióan használják, azonban hasznos funkcióik mellett ezen szolgáltatások komoly problémát is okozhatnak, hiszen kiváló megfigyelési eszközök is egyben: legyen szó akár barátaik után leselkedő felhasználókról, adatbázisokat vásárló cégekről, vagy megfigyelést végző kormányzati szervekről [1]. Mindezeket figyelembe véve a közösségi hálózatok az egyik kulcs tényezői az információs társadalmakból kukkoló társadalmakat kialakító folyamatoknak [2].

A közösségi hálózatokban felmerülő problémákra [3, 4] számos privátszférát erősítő megoldási javaslat született már. Az egyik legnagyobb kihívást rejtegető probléma a kapcsolati rendszerből eredő azonosítási lehetőségek megnehezítése, vagy akár ellehetetlenítése. Egyes megoldási javaslatok a hagyományos szolgáltatások lecserélést szorgalmazzák elosztott közösségi hálózatokra, míg mások az alapvető funkcionalitást akarják valahogy megváltoztatni – azonban mind a kettőben közös, hogy a csatlakozni kívánó felhasználóknak végül el kell hagyniuk a bevált szolgáltatásokat a privátszférájuk megóvása érdekében (ilyen például a diaspora* [5]). Egy irányvonal szerint a közösségi szolgáltatóknak kellene alkalmazni anonimizálási eljárásokat, amely segítségével továbbra is kiadhatnák felhasználók adatait üzleti, vagy kutatási célra. Ide sorolhatóak például a differential privacy körébe tartozó eljárások [6].

Azonban a probléma megoldására a legalkalmasabbak a fokozatosan elterjeszthető megoldások, amelyek lehetővé teszik a privátszféra védelmét, és emellett a többi felhasználóval megmarad a kapcsolattartási lehetőség. Továbbá, mivel a legtöbb közösségi hálózatot rávehetik állami szereplők, hogy kiadják a felhasználóik adatait, illetve partnereikkel is megosztanak adatokat, ezért az adatok feletti irányítást nem szabad kivenni a felhasználók kezéből az irányítást. Még akkor sem, ha feltételezhetnénk egy-egy módszerről, hogy azt alkalmazva a közösségi hálózat üzemeltetői privát módon tudnák megosztani a felhasználók adatait.

Számos esetben nincs szükség explicit kapcsolatrendszerre a kapcsolati gráf felépítéséhez, ugyanis a meta-adatokból kiderül, ami már az azonosításhoz elégséges. Helyzetinformáció de-anonimizálására léteznek olyan támadások, amelyek először felépítenek a hely és idő információk alapján egy a potenciális ismertségeket reprezentáló gráfot, majd újraazonosítják az egyes csomópontokat egy közösségi hálózatból származó gráf segítségével [7–10]. Ezek, és az ehhez hasonló eseték miatt is szükséges, hogy a privátszférát erősítő megoldásoknál az irányítás a felhasználó kezében maradjon.

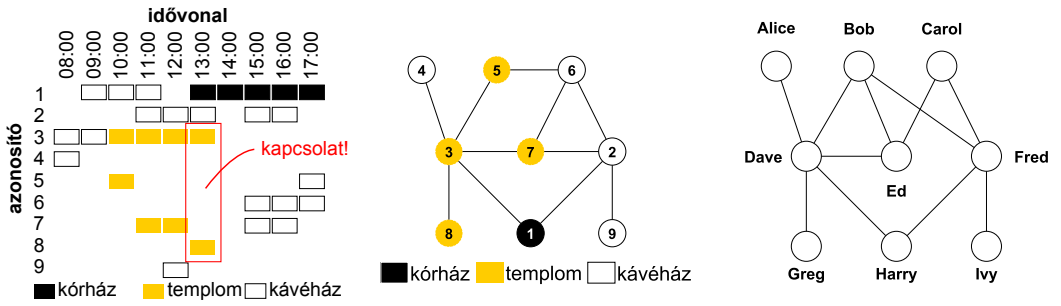
## 2 Motiváció

Mielőtt egy szolgáltató megoszt egy partnerével egy adathalmazt, általában naiv anonimizálási eljárással igyekszik védeni a felhasználók privátszféráját: eltávolítja az explicit azonosítókat (mint a nevek, azonosítók, email címek), és a gráf struktúrát kis mértékben módosítja (például néhány élet töröl, és néhány másikat felvesz). Ezek az eljárások azonban nem alkalmasak a megfelelő védelem biztosítására, az így védett adathalmazok résztvevőit nagy pontossággal újra lehet azonosítani [8, 10–19]. Ráadásul ezen támadások többsége a nagymértékű újraazonosításra is képes, akár több százezer vagy még nagyobb hálózatok esetén is.

A munkámban egy erős támadás típust vizsgálok, amely az újraazonosítást csupán a struktúra felhasználásával képes sikeresen elvégezni [8,10–15]. A következő példán keresztül szemléltetem ezen támadások alapelvét [10, 12]. A példa segítségével betekintést nyerhetünk abba is, hogy a helyzetinformáció megosztása milyen veszélyt hordoz a privátszférát illetően (mint például hívásinformációk helyadatait, vagy check-in jellegű bejelentkezések), ugyanis ahogy ezt korábban említettem, az efféle hely-idő adatokat gráf struktúrára lehet konvertálni [20].

Vegyünk egy támadót, aki az 1a. ábrán látható hely-idő adatokhoz hozzájut. Például megvásárolja egy kisváros nyilvános Wifi szolgáltatójától, aki szándékosan gyűjti az egyes hozzáférési pontoknál elhaladó eszközök MAC címét (például bekapcsolt Wifi hálózatú okostelefonokét). Ebből létrehozhat egy az 1b. ábrán látható kapcsolatrendszert az egyes eszközök egy időben és egy helyen tartózkodásának figyelembevételével. Üzleti szempontból az így kapott adathalmaz még értékesebb volna a támadó számára, ha ismerné az egyes csomópontok valódi neveit.

Ezért egy ismert közösségi portálról begyűjti a kisváros regisztrált felhasználóinak a kapcsolatrendszerét, majd ezt felhasználja a névtelen adathalmaz de-anonimizálásához. Ez az adathalmaz, a támadó háttértudása a 1c. ábrán látható. A támadás jellemzően két lépésben történik. Először a támadó globálisan kiugró csomópontokat keres, amelyek például a magas fokszámú csomópontok, majd ezeket összehasonlítja (az inicializálási fázis). A jelen példában ilyen pár lehet a két gráfban a $v_{Dave} \leftrightarrow v_3$ és a $v_{Fred} \leftrightarrow v_2$ összerendelések. Mivel több egyértelmű globális párosítási lehetőség nincs, a következő lépéstől a már ismert csomópontokhoz szomszédainak a vizsgálata lesz iteratív módon (ez a terjedési fázis). Például $v_{Harry}$-ről azt lehet tudni, hogy két szomszédja van (ami globálisan nem egyedi érték), és kapcsolatban áll a két már újraazonosított személlyel $(v_{Dave}, v_{Fred})$; végezetül ez elvezet a következő újraazonosítás lehetőséghez: $v_{Harry} \leftrightarrow v_1$. Ehhez hasonlóan több lépésben a hálózat jelentős része újraazonosítható.




(a) Anonimizált adathalmaz (b) (a) alapján generált hálózat (c) Publikus közösségi hálózat

1. àbra: Példa egy támadás lehetséges adathalmazaira. A támadó a hely-idő információból kapcsolati hálózatot építhet, amelyet de-anonimizálhat egy közösségi hálózat segítségével.

A támadás sikeres végrehajtása után a rosszindulatú fél már tudja, hogy Harry több óráig is tartózkodott a kórházban. Ezt a tényt használhatja célzott hirdetésekhez, vagy akár zsarolhatja is vele, hogy ezt az információt megosztja barátaival vagy főnökével.

A disszertációmban egy felhasználó-központú technikát vizsgálok, mint potenciális megoldást, az ún. identitás szeparációt. Ez a technika a már meglévő szolgáltatásokban is alkalmazható fokozatosan, anélkül, hogy a szolgáltatásoknak alkalmazkodniuk kellene, vagy hozzájárulásukat adnák, és megmaradhatna a lehetőség a védekezést választó és többi felhasználó között. Az identitás szeparáció annak az elvén alapul, ahogy részleges identitásainkat használjuk a mindennapokban: más és más információt osztunk meg különböző helyzetekben és különböző kapcsolatainkban [21]. Ezt az elvet a közösségi hálózatokban is lehet alkalmazni, ahol például csoportokat hozhatunk létre az információmegosztás finomítása céljából.

Visszatérve az előbbi példára, az identitás szeparációt akár a MAC címek kezelésénél is lehet alkalmazni: Harry például megváltoztathatta volna a MAC címét amikor megérkezett a kórházba, vagy ki is kapcsolhatta volna a hálózati hozzáférést, hogy megelőzze ezen információ kiszivárgását.

## 3 Kutatási célok

A struktúrális újraazonosítási támadások a közösségi hálózatok privátszférát érintő kérdéskörének viszonylag új és aktívan kutatott területét jelentik. Az első és azóta is a legkorszerűbbnek tekinthető támadást, amely lehetővé tette a kapcsolatok alapján a nagymértékű de-anonimizálást, Narayanan és Shmatikov készítette 2009-ben [12] (munkámban Nar09 néven hivatkozom a támadási algoritmusukra). Eredményeik egy új

kutatási irányvonalat nyitottak meg. A disszertációmban három problémakörrel foglalkoztam, amelyek mindegyikre ezekhez a támadásokhoz kapcsolódik.

**1. problémakör.** *Újraazonosítási algoritmusok vizsgálata.* Mivel ezeknek a támadásoknak a területe viszonylag újnak számít, számos nyitott kutatási lehetőséget kínál. A disszertációmban két kapcsolódó kérdéskört vizsgáltam: hogyan mérhető ezek esetén az anonimitás, illetve a támadások inicializálását. Mivel ezen támadások iteratívan működnek, nem lehet egyszerűen anonimitási halmazokat felállítani és eszerint mérni az anonimitás fokát – ezért volt szükség egy itt használható módszer kidolgozására. A második esetben az volt a célom, hogy feltárjam, hogy az inicializálás hogyan hat a támadás összteljesítményére. Ugyanis az irodalomban lévő munkák többféle eljárást is használtak az inicializáláshoz, de nem volt olyan elemzés, amely ezeket összehasonlította volna, megállapítva az egyes módszerek kapcsolatát, előnyét-hátrányát.

**2. problémakör.** *Az identitás szeparáció elemzése a de-anonimizálás megállítása szempontjából.* Az identitás szeparáció alkalmas privátszférát erősítő eljárásnak tűnik a jelenlegi kontextusban. Azonban modell szinten is meg kell vizsgálni a hatékonyságát a de-anonimizálással szemben. Ebben a problémakörben két összetettebb kérdés megválaszolása volt a célom. Először is, hogy milyen feltételekkel lehetséges a támadás megállítása, és így a hálózat szintjén a privátszféra megőrzése? Majd pedig azt vizsgáltam, hogy a különféle esetekben a résztvevő felhasználók privátszférája milyen mértékben sérül, függetlenül a támadás megállításának sikerességétől.

**3. problémakör.** *Egyéni, az információ szivárgást minimalizáló stratégiák elemzése.* Még ha lehetséges is a támadás megállítása, egy felhasználó dönthet úgy, hogy csak a saját érdekében kíván védekezési eljárást alkalmazni. Ehhez kapcsolódóan több felmerülő kérdést vizsgáltam a problémakörben. Vannak olyan stratégiák az identitás szeparáció alkalmazására, amelyek akkor is elérik egyéni szinten az információ szivárgás minimalizálását, ha mindössze csak néhányan alkalmazzák a technikát? Ha igen, képes lehet-e egy támadó az identitás szeparáció visszafordítására? Ezen kérdések megválaszolása mellett olyan stratégiákat is kerestem, amelyek elvi korlátokat adnak a támadó lehetőségeit illetően.

# 4 Módszertan

Mivel a szimuláció a tipikus eszköze a közösségi hálózatok privátszféra védelmének a kutatásának, én is sok esetben ezt a módszert alkalmaztam a disszertációmban. Azonban analitikus módszert is alkalmaztam néhány probléma esetén. Mivel nem léteznek olyan adathalmazok, amelyek segítségével az identitás szeparáció életszerűen modellezhető

volna (és ennek létrehozása jelentősen túlmutat a jelenlegi munka keretein), készítettem egy modellt amely mind a szimulációs, mind az analitikus esetekben alkalmazható az identitás szeparációt alkalmazó felhasználók viselkedésének modellezésére. A kísérleteim megismételhetősége okán megadtam az ehhez szükséges paramétereket, amelyeket úgy választottam meg, hogy a lehetséges befolyásoló tényezőket kizárjam: ilyen lehet például a hálózati struktúra vagy a kísérletek száma. Szimulációt mind a három **problémakörben** alkalmaztam. A **2. problémakörben** valószínűségszámítást alkalmaztam a támadó hibavalószínűségének levezetéséhez, illetve numerikus analízist ennek vizsgálatához. A **3. problémakörben** a k-anonymity modell [22] egy adoptációját készítettem el és vizsgáltam meg, illetve valószínűségszámítás és játékelmélet segítségével vizsgáltam privátszférát erősítő stratégiákat.

# 5 Új eredmények

A szimulációk során az eredményeket két különböző mértékkel mértem, melyek a támadás sikerét tükrözik különböző aspektusokból. Az ún. *helyes találati arány* a támadás kiterjedtségét tükrözi (recall rate a disszertációban), azaz az összes potenciálisan elérhető csomópontból mennyit sikerült a támadónak megtalálnia és helyesen azonosítania. Ez a támadó szempontjából határozza meg a siker mértékét, és mivel a hibás találatok aránya igen alacsony volt, elegendő volt ezt vizsgálni. A védekező felhasználók szempontjából a támadás sikerét a rájuk vonatkozó kiszivárgott információk mennyisége határozza meg, amit az ún. *felfedési arány* adott meg a méréseim során (disclosure rate a disszertációban).

## 5.1 Strukturális újraazonosítási algoritmusok vizsgálata

Megvizsgáltam a strukturális újraazonosítási támadásokat több szempontból. Javasoltam egy anonimitási mérték családot, amelyeket ezeknél a támadásoknál lehet használni. Megvizsgáltam több, ebbe a csoportba tartozó mértéket a Nar09 támadás segítségével. Ezek a mértékek megmutatják, hogy a hálózatban melyek a támadás szempontjából fontos csomópontok, amelyeket a támadó nagyobb valószínűséggel fog tudni sikeresen deanonimizálni. Ezen túlmenően megmutattam a Nar09 támadás inicializálásának jelentőségét, amely a támadás sikerességét alapvetően meghatározza.

---

**1. téziscsoport.** *Strukturális újraazonosítási támadások vizsgálata. Javasoltam egy anonimitási mérték családot, amelyet Local Topological Anonymity (LTA) néven neveztem el, és megmutattam, hogy az ebbe tartozó egyik LTA mérték és a csomópont fokszám is jól megmutatja, hogy mely csomópontokat tud a legkorszerűbb támadás valószínűleg újraazonosítani. Ugyanezen támadásnál megmutattam az inicializálási fázis fontosságát, és hogy különböző módszerek miként befolyásolják a támadás összteljesítményét.*

---

**1.1. tézis.** *Készítettem egy Local Topological Anonymity (LTA) néven elnevezett anonimitási mérték családot, amelyek lehetővé teszik az újraazonosítás kockázatának relatív meghatározását. Megmutattam, hogy az ebbe a családba tartozó $LTA_A$ mérték által adott értékek jelentős rangkorrelációt mutatnak a legkorszerűbb, illetve a Grasshopper támadás által adott csomópont újraazonosítási értékekkel.*
*Kapcsolódó publikációk: [C3, J2, J3]*

A nagymértékű újraazonosítási támadások második fázisa, az ún. terjedési fázis jellemzően iteratív működési elvet követ, és az aktuálisan vizsgált a csomópontot a tőle $d = 2$ távolságra lévőekkel veti össze (az ún. barátainak barátaival). Ezért minél hasonlóbb egy csomópont ezekhez, annál kisebb eséllyel fogja tudni egy támadó újraazonosítani. Ezt a tulajdonságot kell az anonimitási mértékeknek megragadni.

**1. definíció.** *A Local Topological Anonymity (LTA) mértékek $LTA(\cdot)$ módon jelölt függvények, amely megadják egy csomópont rejtőzködési képességét a strukturális deanonimizációs támadásokkal szemben, amelyek legfeljebb $d$ távolságra keresnek egymáshoz hasonló csomópontokat*[1].

A csomópontok hasonlóságát többféleképpen mérhetjük. A Nar09 támadás a háttértudásában szereplő csomópontok ($v_{src} \in G_{src}$) koszinusz hasonlóságát méri az anonimizált adathalmazban lévő csomópontokéval ($v_{tar} \in G_{tar}$). Mivel ez támadásonként eltérhet, ezért az egyes LTA mértékeknek ehhez adoptálhatónak kell lennie az ilyen esetekhez:

**2. definíció.** *Egy LTA mérték egy függvény, amelyet $LTA_\alpha(\cdot)$ módon jelölünk, és az a $f_\alpha(\cdot)$ csomópont ujjlenyomat függvényre épül, megadja egy csomópont strukturális újraazonosíthatóságát az adott közösségi hálózaton belül.*

[1] A munkámban $d = 2$ értékkel dolgoztam, amelynél nagyobb értékek a hálózatok tipikusan kicsi átmérői alapján már túlzott erőforrás igényű választásnak tűntek.

A disszertációmban három mértéket javasoltam, amelyek a koszinusz hasonlóságra épültek, amit $CosSim(\cdot)$ módon jelöltem. Ez a hasonlósági mérték lecserélhető, az aktuálisan vizsgált támadás működési elvének megfelelően. Ezek közül az $\mathrm{LTA}_A$ megadja egy csomópont átlagos hasonlóságát a $d = 2$ távolságú környezetében lévőkhöz képest:

$$\mathrm{LTA}_A(v_i) = \sum_{\forall v_k \in V_i^2} \frac{CosSim(v_i, v_k)}{|V_i^2|}, \tag{1}$$

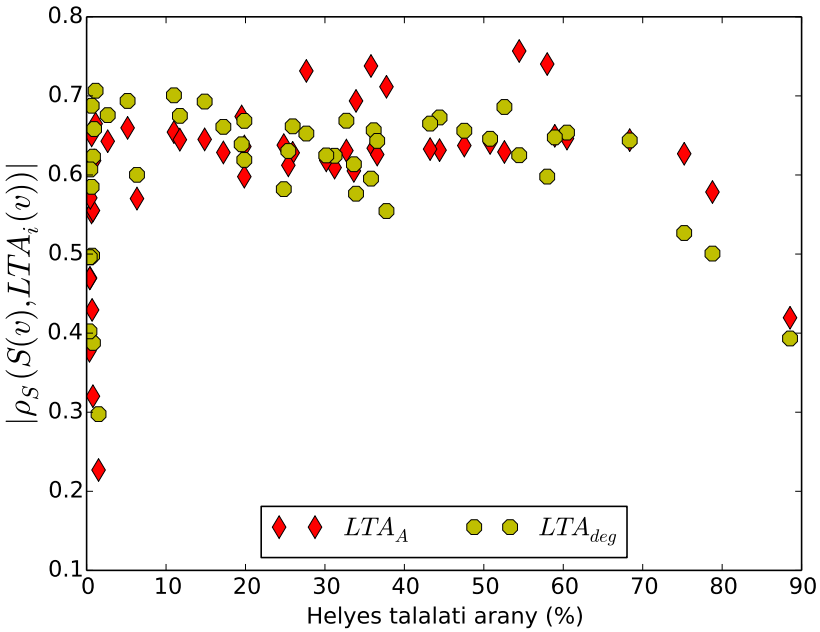
Az $\mathrm{LTA}_B$ esetén a fokszámmal normalizáltam a hasonlósági értékeket (de legalább kettővel), amely a magas fokszámú csomópontokat bünteti, ugyanis azok jellemzően könnyebben de-anonimizálhatóak. Az $\mathrm{LTA}_C$ szorosan az $\mathrm{LTA}_A$ mértékre épül, de azt tovább osztja a $v_i$ és $\forall v_j \in V_i^2$ ($d = 2$ távolságban lévő csomópontok) közötti fokszám eltérések szórásával, így figyelembe véve a környezet változékonyságát is. Ezeknek a képleteit a disszertációmban közöltem.

Összehasonlítottam a legkorszerűbb algoritmus újraazonosítási eredményeit az LTA mértékekkel több forrásból származó, és különféle paraméterek mentén előállított adathalmazon. Az összehasonlítást a Spearman rangkorreláció segítségével tettem meg (jelölése: $\rho_S$). Egy LTA mérték akkor tekinthető jónak, ha a korreláció abszolút értékben közel esik egyhez. Végül a három javasolt mérték közül az $\mathrm{LTA}_A$ adta a legjobb eredményeket, melyet jól szemléltet a 2. ábra. A disszertációmban az eredményekről és a vizsgálatról további részleteket közöltem.

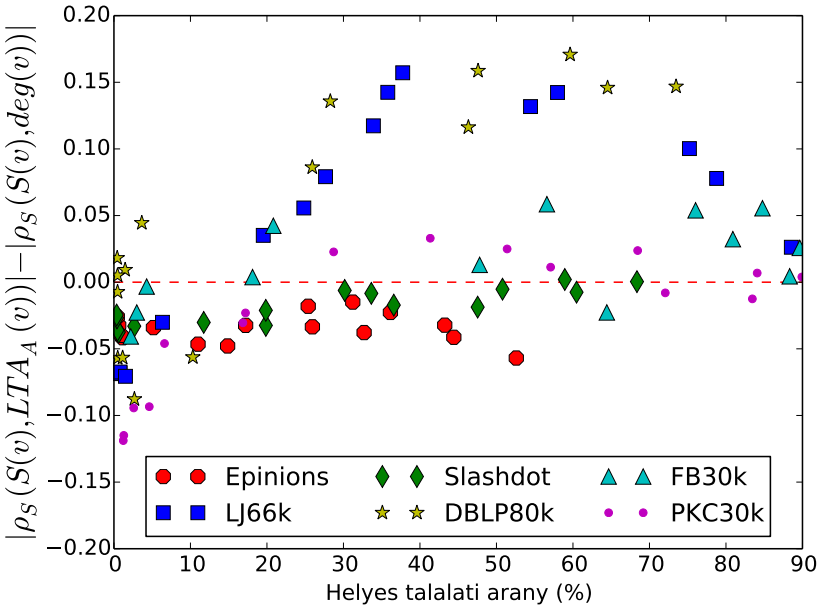
**1.2. tézis.** *Megmutattam, hogy a fokszám ($LTA_{deg}$) hatékony, gyorsan kiszámítható alternatívája az $LTA_A$ mértéknek. Megállapítottam, hogy az egyes hálózatok fokszám eloszlása hogyan határozza meg, hogy melyik mértéket érdemes ezek közül alkalmazni a legkorszerűbb támadás esetén: az $LTA_{deg}$ magasabb korrelációt ért el olyan hálózatokban, ahol a kis fokszámú csomópontok aránya relatíve nagyobb volt, míg az $LTA_A$ mérték ért el magasabb korrelációt a többi hálózat esetén.*
*Kapcsolódó publikációk: [J2, J3]*

A fokszám egy jelentős tulajdonság az újraazonosítás sikeressége szempontjából, és a méréseim szerint a Nar09 támadás erősen elfogult a magasabb fokszámú csomópontok irányába. Például egy mérésben a kis fokszámú csomópontoknak ($\deg(v) \leq 3$) mindössze 20%-át sikerült de-anonimizálni, addig a magasabb fokszámú csomópontok esetén ez 80% körül volt. Ezért a disszertációmban javasoltam a fokszámot is az előbbiekhez hasonló vizsgálatra ($\mathrm{LTA}_{deg}$ jelöléssel), és megmutattam, hogy ez is igen kedvező rangkorrelációs értéket képes felmutatni, ahogyan azt a 2. ábra is szemlélteti.

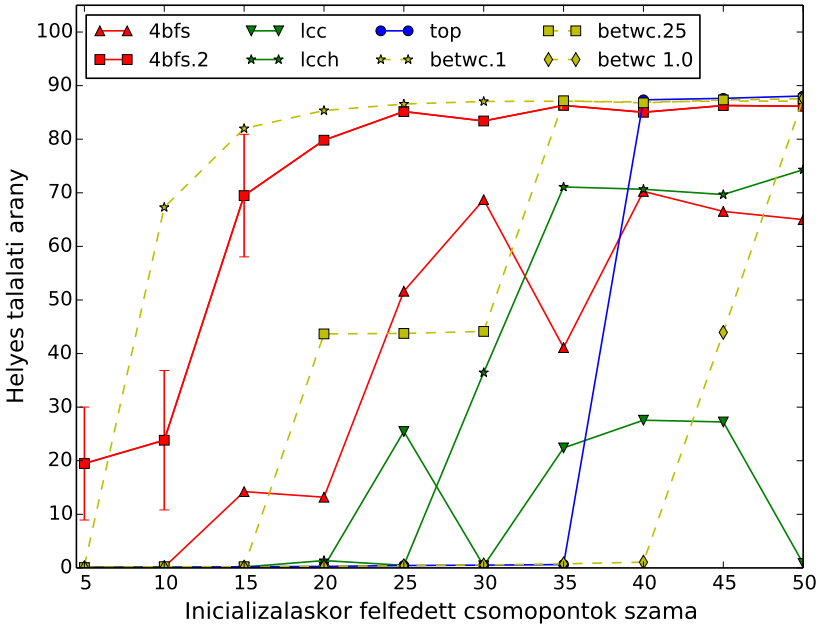
A vizsgálataim során az is kiderült, hogy az $\mathrm{LTA}_{deg}$ és az $\mathrm{LTA}_A$ által kiemelt csomópon-




2. àbra: LTA mértékek rangkorrelációjának összehasonlítása különböző perturbációs eljárások esetén a helyes újraazonosítások számával. Az ábrán szereplő $\text{LTA}_A$ és $\text{LTA}_{deg}$ adta a legjobb eredményeket a vizsgáltak közül.




3. àbra: Jelentős eltérés van az $\text{LTA}_A$ és az $\text{LTA}_{deg}$ mértékekhez kapcsolódó korrelációs értékekben a hálózati struktúrától függően. Csak az Epinions és Slashdot hálózatokban adott jobb értéket az $\text{LTA}_{deg}$, az összes többiben az $\text{LTA}_A$ használata tűnt célravezetőnek.




4. àbra: A különböző inicializálási módszerek igencsak eltérhetnek tulajdonságaikban, mint például az elérhető maximális hozam.

tok között jelentős átfedés van (kb. 80%), noha ennek ellenére nem ugyanúgy teljesít a két mérték ugyanazokban a hálózatokban. Megmutattam, hogy a korrelációs értékekben az eltérést jelentősen befolyásolja a hálózatok fokszám eloszlása: ahol jelentősebb mennyiségű kis fokszámú csomópont található, ott az $\mathrm{LTA}_{deg}$ ad magasabb korreláció értéket, míg az $\mathrm{LTA}_A$ a többi esetben. Ezt szemlélteti a 3. ábra. További részleteket adok a vizsgálatról a disszertációmban.

**1.3. tézis.** *A legkorszerűbb algoritmusnál megmutattam az inicializálási fázis jelentőségét. Megmutattam, hogy a maximum újraazonosított csomópontok száma hogyan függ különféle inicializálási módszerektől és annak paramétereitől, illetve hogy az inicializálás során kiválasztott minimum csomópontszám hogyan függ a hálózat tulajdonságaitól és a kiválasztási módszertől. Valamint megmutattam az instabil inicializálás jelentőségét, hogy akár igen kevés csomópont kezdeti felfedése esetén is sikeresen elérheti a hálózat nagymértékű de-anonimizálását a támadó.*
*Kapcsolódó publikációk: [C1]*

A Nar09 és a hasonló elven működő algoritmusok kétfázisúak. Az első, az inicializálási fázisban az algoritmus csupán néhány csomópontot azonosít be globális tulajdonságaik alapján. Már a kezdeti munkájában is felhívta Narayanan és Shmatikov a figyelmet, hogy

fázisátmenet tulajdonságot [23] tapasztaltak az algoritmusnál az inicializálás során felfedett csomópontok függvényében [12]: ahogy nőtt az inicializálás során felfedett csomópontok száma, egy ponton ugrásszerűen megnőtt az algoritmus második részének teljesítménye, szinte egyből elérve a maximumot. Részletek közlése nélkül azt is megjegyezték, hogy a fázisátmenet függ a hálózat struktúrájától és az inicializálási módszertől. A nagymértékű újraazonosítás valószínűségét munkájukban mint az inicializálási csomópontok számának függvényeként említik. Azonban ezek az említések nélkülözik a részleteket (például az egyes módszerek összehasonlítását), és az [12] után következő munkák egyike sem indokolja, hogy miért épp az aktuálisan használt módszert választja a többi helyett.

Ezeket a tulajdonságokat több módszer esetén is megvizsgáltam és összehasonlítottam a disszertációmban. A fázisátmenetet és további számos tulajdonságot szemléltet a 4. ábra többféle inicializálási módszer esetén. A disszertációmban megmutattam, hogy az inicializálásnál kiválasztott csomópontok globális jellemzői (például a magas fokszám vagy a magas betweenness érték), egymás közötti kapcsolata (például klikkekről van-e szó vagy szomszédos csomópontokról) hogyan határozza meg a támadás egészének sikerét. Arra is adtam példát, hogy vannak olyan módszerek, amelyek esetén csak alacsonyabb találati értéket lehet elérni, ahogy például ezt a 4. ábra szemlélteti az `lcc` módszerre.

## 5.2 Az identitás szeparáció technika elemzése

Az identitás szeparációt használatát közösségi hálózatokban először az irodalomban a [C7, C8] munkákban javasoltuk. Az alábbi vizsgálatokban pedig az [J4] cikkben publikált modellre építettem, amely négy különböző viselkedési típust foglalt magába. Ezek a modellek részletesen le vannak írva a disszertációmban, azonban madártávlatból szemlélve a következőképpen működnek: a $v_n$ felhasználó létrehoz $Y = y$ darab új identitást (amelyek egy külső szemlélő számára összeköthetetlenek), és az eredeti kapcsolatokat az új identitások között szétosztja. A javasolt modellek abban térnek el, hogy lehetséges-e egyes kapcsolatokat több helyre besorolni vagy nem, illetve hogy lehet-e egyes kapcsolatok törölni vagy sem.

Ez tehát négy viselkedési típust tesznek lehetővé (ezeket a disszertációmban almodelleknek hívom). Az ún. alapvető modell a legegyszerűbb mind közül (basic model disszertációmban), és nem teszi lehetővé sem a kapcsolatok törlését, sem másolását. Ezzel szemben az ún. valósághű modell (realistic model a disszertációmban) mindkettő funkciót megengedi. Élesben valószínűleg a felhasználók többsége így használná az identitás szeparáció technikáját. Ezen túlmenően megkülönböztetem még az ún. legrosszabb és legjobb modell lehetőségeket is (worst és best model a disszertációmban), amelyeket a

felhasználói aspektus alapján neveztem el. A legrosszabb modell nem engedi a kapcsolatok törlését, csak másolását, a legjobb modell pedig ennek az ellenkezőjét teszi lehetővé.

Ebben a téziscsoportban első sorban arra a kérdésre kerestem a választ, hogy az identitás szeparáció segítségével a közösségi hálózat képes-e védekezni az újraazonosítási támadásokkal szemben. Először megvizsgáltam, hogy az inicializálási fázissal szemben milyen védelmet nyújt ez a technika, majd pedig megnéztem az iterációs fázisra is ezt. Ez utóbbi esetén többféle lehetséges stratégiát megvizsgáltam, amelyek között voltak nem-kooperatív és kooperatív változatok is. Az előbbi esetén a felhasználók önállóan alkalmazzák az identitás szeparációt, míg kooperáció esetén vagy szomszédos csomópontok együttműködését vizsgáltam, vagy pedig valamilyen globális elv szerint kerültek kiválasztásra a védekező csomópontok.

---

**2. téziscsoport.** *A privátszféra védelmének hatékonyságát vizsgáltam az identitás szeparáció technika segítségével. Az általam javasolt modellek alapján a támadó inicializálási hibavalószínűségét vizsgáltam, illetve szimulációk segítségével több nem-kooperatív és kooperatív stratégiát vizsgáltam meg, és megmutattam hogy ezek közül melyek képesek hatékonyan csökkenteni a támadó által újraazonosított csomópontok számát, illetve a védekező felhasználókról kiszivárgott információ mértékét minimalizálni.*

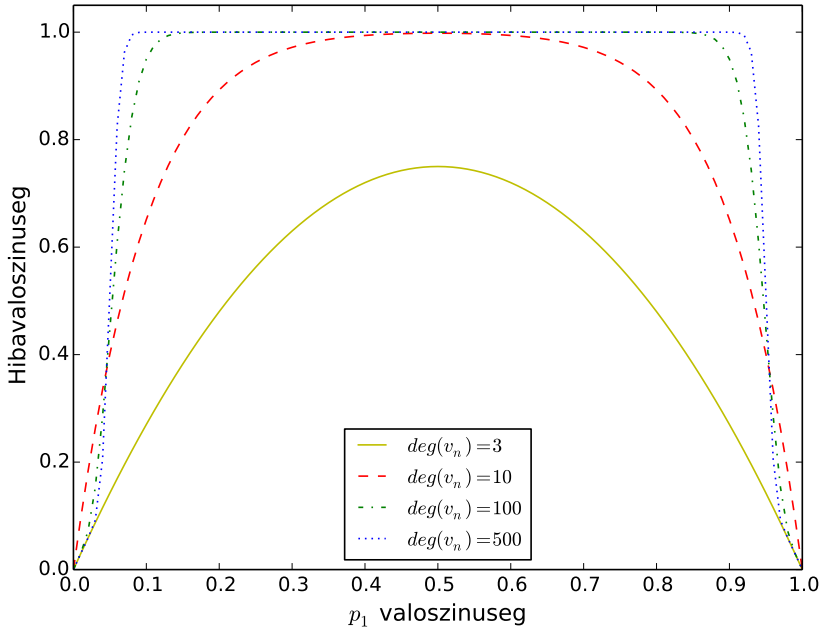
---

**2.1. tézis.** *Megadtam az inicializálási fázishoz a támadó hibavalószínűségét az identitás szeparáció esetén. Ezen képletből levezetve adtam egy alsó becslést a hibavalószínűségre a klikk alapú, illetve a legmagasabb fokszámú csomópontok kiválasztásán alapuló inicializálási módszerekhez. Numerikus analízis segítségével megmutattam, hogy számos olyan beállítás létezik, amely által a felhasználó meg tudja magát védeni a támadóval szemben.*
*Kapcsolódó publikációk: [J4]*

A $v_n$ csomópont számára a hibavalószínűséget a teljes valószínűség tételét alkalmazva lehet leírni a következőképpen:

$$P(\text{"hiba"}) = P(Y = 0) + \sum_{y=1}^{\deg(v_n)} P(\text{"hiba"}|Y = y) \cdot P(Y = y). \tag{2}$$

Ezt a képletet tovább ki lehet bontani attól függően, hogy milyen felhasználói viselkedési típust választunk. A [12] cikkben az inicializáláshoz 4-klikkeket használtak. Ezért a disszertációmban a klikkalapú inicializálási módszerekhez (ahol a klikk méretet $k$ paraméterként kezeltem) megadtam a támadó hibavalószínűségét az alapvető és valósághű modellek esetén. Megmutattam, hogy számos olyan identitás szeparációs beállítás létezik, amely esetén



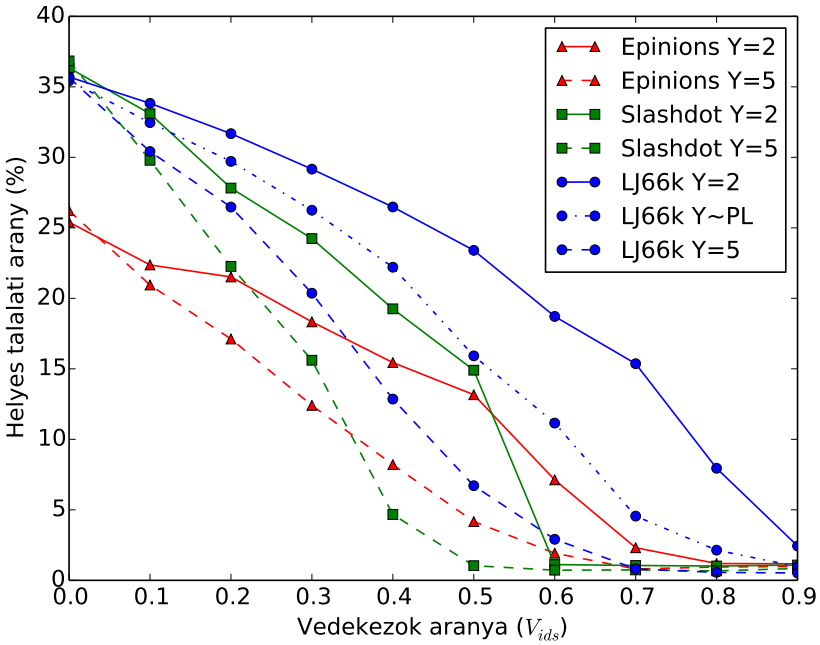

5. àbra: Az alapvető modell paramétereinek vizsgálata. A $P^B_{clique}$("failure"$|Y=2$) hibavalószínűség értékei a $p_1$ függvényében, rögzített $k=4$ és $\epsilon=0.05$ paraméterekre, többféle fokszám esetén.

a támadó hibavalószínűsége lényegében az 1.0 értéket eléri, még akkor is, ha a védekező felhasználó mindössze két új identitást hoz létre (ezt $Y=2$ módon jelöltem). Az 5. ábra megmutatja, hogy különböző fokszámok esetén hogyan alakulnak a hibavalószínűségek. Ez esetben a vizsgálat alapjául a következő képlet szolgált, amelyet (2)-ből vezettem le:

$$P^B_{clique}(\text{"hiba"}|Y=y) = 1+\sum_{\forall i\in[0,\ldots,y]} p_i^{k-1}\cdot\left(\sum_{x_1''+\cdots+x_y''=n-k+1}\left(\frac{(n-k+1)!}{x_1''!\cdot\ldots\cdot x_y''!}\cdot p_1^{x_1''}\cdot\ldots\cdot p_y^{x_y''}\cdot e(k-1,x_i'')\right)-1\right) \tag{3}$$

ahol $p_i$ annak a valószínűsége, hogy egy kapcsolatot az identitás szeparációt végrehajtó felhasználó az $i$. identitáshoz sorol be. Emiatt az alapvető modell esetén multinomiális eloszlással dolgoztam ($\sum p_i=1$).

A (2) képlet segítségével arra az esetre is vizsgáltam a hibavalószínűséget, amikor a támadó a legmagasabb fokszámú csomópontokat veti össze két hálózatban az inicializáláshoz. Egy esetben például megmutattam, hogy az 1000 legmagasabb fokszámú csomópont esetén azok többsége nem lesz újraazonosítható (80.4%), ha csupán két új identitást is használ-




6. ábra: Az alapvető modell működését bemutató szimulációs eredmények különböző hálózatokban több beállításban, amelyekben a felhasználók 0-90%-a alkalmazza nem-kooperatív módon az identitás szeparációt.

nak. Az általam bemutatott eljárás és modellek segítségével ez a vizsgálat más inicializálási módszerekre is elvégezhető.

**2.2. tézis.** *Vizsgáltam a legkorszerűbb támadás terjedési fázisának érzékenységét az identitás szeparáció technikára, és megmutattam, hogy a támadás igen robusztus: nagy számú nem-kooperatív felhasználó kell ahhoz, hogy a támadó által újraazonosított csomópontok számát kellően alacsonyra lehessen csökkenteni.*
*Kapcsolódó publikációk: [C2, J1, J3]*

Annak érdekében, hogy érdemben vizsgálhassam az identitás szeparáció technika legerősebb stratégiáit, először megmértem, hogy a különböző részeire, mint az élek törlése vagy a csomópontok "szétrobbantása" mennyire zavarja a támadó algoritmust. A kezdeti várakozásokkal szemben kiderült, hogy az alapvető modell esetén nem lehet hatékonyan megállítani a támadást, mivel több, mint a közösségi hálózat felének részt kell vennie a védekezésben ehhez (ez teljesül $Y = 5$ új identitás esetére is). Ezt szemlélteti a 6. ábra.

Megmértem más modellekkel azt az esetet, amikor az élek törlése is lehetséges volt, és ebben az esetben az újraazonosított csomópontok aránya hasonló volt az előbbi esethez, csupán kissé volt kedvezőbbnek mondható. Ezek az eredmények megmutatták, hogy

ezeknél az alapvető megközelítéseknél bonyolultabb stratégiára van szükség a közösségi hálózat privátszférájának megőrzéséhez. Szerencsére a felfedési arány eredményei kedvezőbbek voltak. A disszertációm a mérésekről további részleteket tartalmaz.

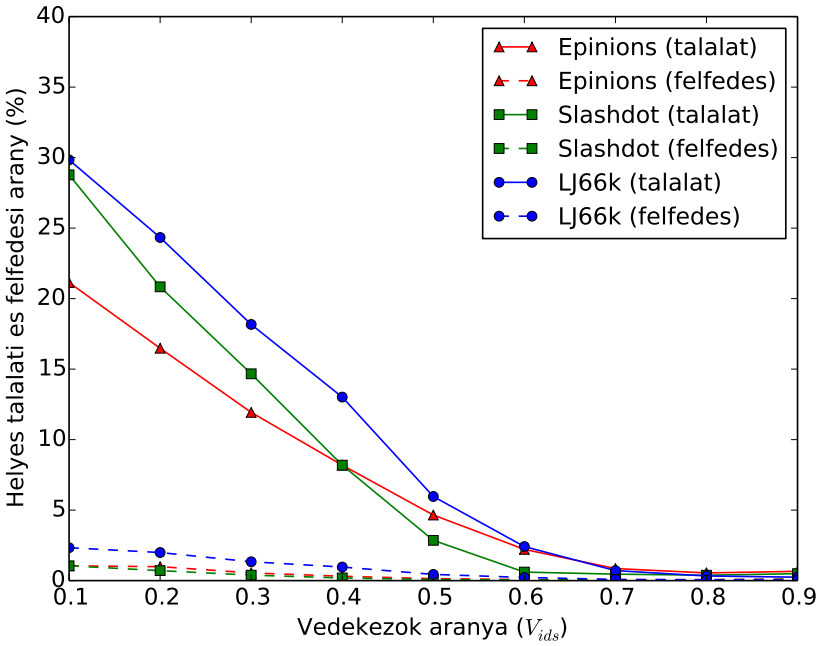
**2.3. tézis.** *Megmutattam különböző tulajdonságait a nem-kooperatív identitás szeparációnak. Megmutattam, hogy ha a támadó hiába kísérletezik az inicializációs eljárás cseréjével vagy az ott felfedezett csomópontok számának növelésével, nem tud jelentősen javítani az algoritmus összeredményén.*
*Kapcsolódó publikációk: [C2, J2, J3]*

Az előző eredményeket figyelembe véve új identitás szeparációs stratégiákat vizsgáltam. Az egyik legérdekesebb eredmény szerint nem érdemes az alapvető modellt két új identitással alkalmazni: ez esetben megmutattam, hogy egy kézenfekvőbb inicializálási módszer alkalmazása esetén a védekező felhasználók körében magasabb lesz a helyes találati arány, mint a hálózat összességében (részletek a disszertációban, ahol a jelenség okát is megadom). Ezen túlmenően megmutattam, hogy még a legjobb modell és $Y = 5$ alkalmazása (ld. 7. ábra) is csak akkor képes a támadás megállítására, ha a hálózat jelentős része alkalmazza az identitás szeparációt. Kiderült viszont, hogy ez esetén a résztvevő felhasználók minimalizálni tudják a róluk kiszivárgó információ mértékét már kisebb arányú részvétel esetén is: a felfedési arány értékek jellemzően 3% alatt maradtak.

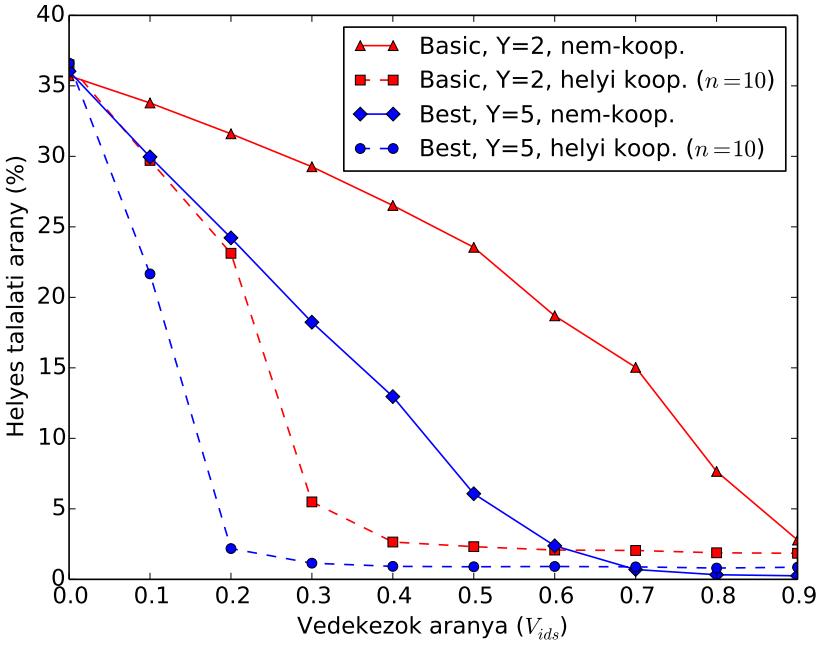
Összhangban az 1.3. tézishez kapcsolódó eredményekkel, több inicializálási módot is megnéztem, hogy azok segítségével a támadó képes-e javítani eredményén: a kísérletek azt igazolták, hogy legfeljebb csak minimális mértékben. Azonban mivel egyes inicializálási eljárások robusztusabbak voltak a többinél, ezek általában véve jobb választásnak bizonyulnak: ha kevés csomópont azonosítható újra a támadás előtt, vagy ha sokan alkalmazzák az identitás szeparációt. Ennek következménye lehet, hogy egy támadó eleve csak kevés felhasználót de-anonimizál inicializálásként, és ezzel addig kísérletezik, míg a nagymértékű újraazonosítás be nem indul. Erről, és a további vizsgálatokról a disszertációban találhatóak további részletek.

**2.4. tézis.** *Megmutattam, hogy egy egyszerűbb lokális kooperáció esetén is már jelentősen kevesebb résztvevő kell, hogy ugyanolyan eredményeket lehessen elérni mint a nem-kooperatív esetben.*
*Kapcsolódó publikációk: [J1]*

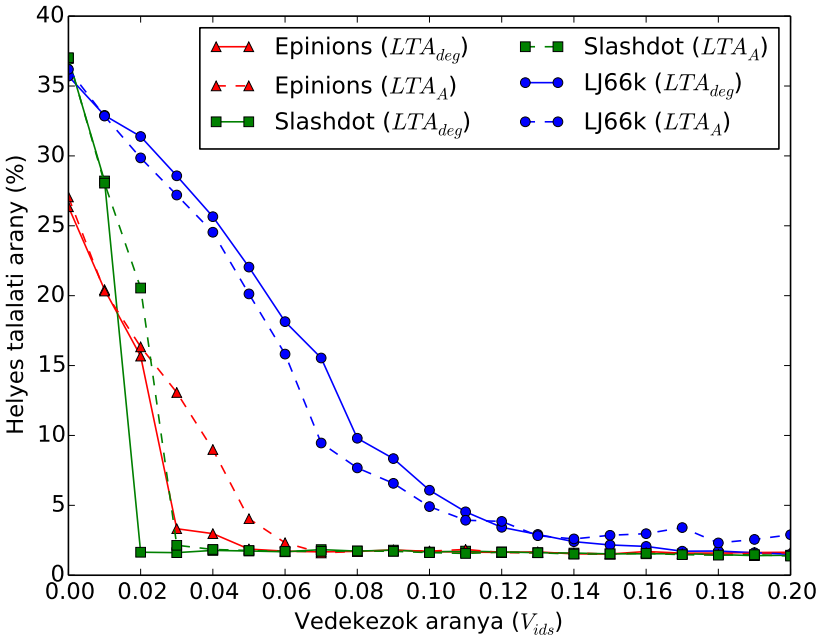
A korábbi mérések során kiderült, hogy a nem-kooperációs identitás szeparáció nem tudja hatékonyan megállítani a támadást a hálózat szintjén. Ezért megvizsgáltam egy egyszerűbb helyi kooperációra építő stratégiát, amelyben egy csomópont a szomszédai




7. àbra: Ha megengedjük az élet törlését, és öt új identitás alkalmazását ($Y = 5$), akkor bár a közösségi hálózat szintjén a támadás továbbra is nehezen lesz megállítható, de az védekezők egyéni szinten jól járnak.




8. àbra: A helyi kooperáció hatása összevetve a nem-kooperatív identitás szeparációval: jelentős javulás érhető el.

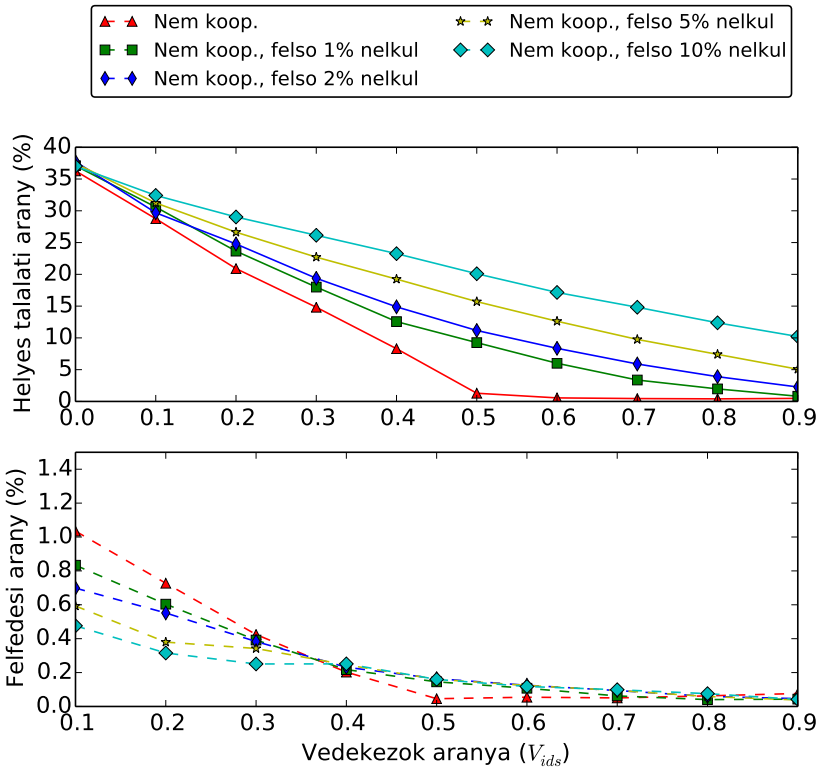



9. ábra: A globális kooperáció eredményeinek összehasonlítása több hálózatban, amikor a $\text{LTA}_A$, valamint a $\text{LTA}_{deg}$ mérték alapján választottam ki a releváns csomópontokat. A kísérletekben a legjobb modell beállításait használtam és az új identitások száma $Y = 5$ volt.

közül kiválaszt $n - 1$-et, és ők közösen alkalmazzák az identitás szeparációt. Mivel a támadás elég robusztus, arra számítottam, hogy nem lesz hatékony ez a védekezési forma, de a szimulációkból kiderült, hogy mégis az. Több paraméter értékre megvizsgáltam ezt a stratégiát, mint $n \in \{5, 10, 25\}$, az alapvető modell és $Y = 2$, valamint a legjobb modell és $Y = 5$ esetére. Kiderült, hogy már akár $n = 10$ esetén is jelentős javulás érhető el, amint az a 8. ábrán látszik (nagyobb $n$ esetén pedig a javulás mértéke tovább nő).

**2.5. tézis.** *Megmutattam, hogy ha az $LTA_A$ és $LTA_{deg}$ mértékek segítségével választunk ki globálisan a hálózatból a releváns csomópontokat az identitás szeparáció alkalmazására, akkor a sikeres védekezéshez szükséges minimum részvételi arány töredékére csökken a nem-kooperatív esetének. Megmutattam továbbá, hogy ez esetben is hiába változtatna a támadó inicializálási eljárást, vagy növelné a kezdetben újraazonosított csomópontok számát, nem tud jelentősen javítani az eredményein.*
*Kapcsolódó publikációk: [J1–J3]*

Szimulációval vizsgáltam a globális kooperáció esetét. A korábbi mérésekben igazolt anonimitási metrikák segítségéve azonosítottam be a fontos csomópontokat, hiszen ha egy csomópontnak alacsony szintű az anonimitása, akkor az érdekes lehet a támadó szempon-




10. ábra: Helyes találati és felfedési arány értékek a Slashdot hálózatban a nem-kooperatív beállításnál (legjobb modell, $Y = 5$). Ha a nagy fokszámú felhasználók nem vesznek részt, a védekezési eredmények jelentősen romlanak, azonban a védekező felhasználók ez esetekben is minimalizálni tudják a róluk kiszivárgó információkat.

tjából is. Kiderült, hogy a $\text{LTA}_A$ és $\text{LTA}_{deg}$ mértékek által kiválasztott alacsony anonimitású csomópontokra alkalmazott identitás szeparáció hatékony védekezés a legkorszerűbb támadással szemben. Mindkét mérték esetén a legalacsonyabb minimális részvételi arányt mértem az eddigi összes eredményekhez képest, ahogy azt a 9. ábra is szemlélteti. Ahogy az ábrán is látszik, bár eltért a különböző hálózatokban az egyes kiválasztási módszerek hatékonysága, de ez a különbség egyrészt nem volt jelentős, illetve konzisztens sem a hálózatok felépítését illetően. További részletek ezekről a vizsgálatokról a disszertációban olvashatóak.

**2.6. tézis.** *Megmutattam, hogy mind a nem-kooperatív és globálisan kooperatív identitás szeparáció esetén elengedhetetlen a legmagasabb fokszámú csomópontok részvétele, és nélkülük a védekezés hatékonysága jelentősen leromlik.*
*Kapcsolódó publikációk: [C2, J1, J3]*

Az előzőleg vizsgált esetekben alapértelmezettnek vettem, hogy minden érintett fel-

használó részt vesz a védekezésben. Azonban a valóságban könnyen előfordulhat, hogy a magas fokszámú, általában híresebb, ismertebb felhasználók nehezebben fognak erre hajlani. Ők valószínűleg kevésbé feltűnő megoldásokat fognak alkalmazni, például álca-profilnak meghagyják az ismert identitásukat, és álnéven, a következő problémakör részeként tárgyalt stratégiákat fogják használni egyéb, rejteni kívánt tevékenységükre.

A disszertációmban több szimulációval vizsgáltam, hogy milyen hatással van a hatékonyságra ha az ilyen ismertebb felhasználók nem vesznek részt a védekezésben. Kiderült, hogy akár ha csak a felső 1%-ról is van szó, akik nem hajlandóak együttműködni a többiekkel, akkor jelentősen több résztvevőre van szükség, hogy a támadó helyes találati arány szintjét akár csak az 1%-kal elérhető szinten tartsuk. Ezt a 10. ábra is alátámasztja, bemutatva mind a helyes találati és felfedési arány értékeket a Slashdot hálózatban (legjobb modell, $Y = 5$). Bár a hálózati szintű védekezés nem működik, egyéni szinten továbbra is; a legjobb modell, $Y = 5$ beállítás segítségével a résztvevő felhasználók minimalizálni tudták a kiszivárgó információ mértékét. További részletek a disszertációban találhatóak.

## 5.3 Egyéni stratégiák vizsgálata

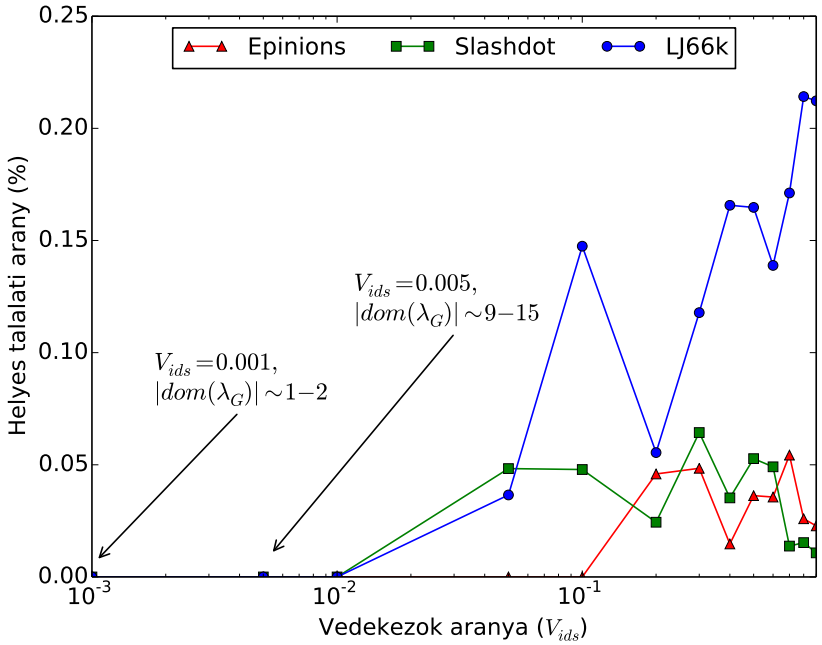
A 2. téziscsoportban megmutattam, hogy nehéz a hálózat szintjén megállítani a legkorszerűbb támadást, bár egyéni szinten van esély a privátszféra védelmére. Ezért a disszertációm utolsó részében megvizsgáltam, milyen egyéni stratégiák segíthetik a felhasználókat.

---

**3. téziscsoport.** *Megmutattam, hogy akkor is érdemes alkalmazni az identitás szeparációt, ha mindössze csak pár felhasználó teszi azt, és adtam egy módszert a részidentitások újraazonosításának a valószínűségének alsó becslésére. A k-anonymity modell alapján kidolgoztam annak egy identitás szeparációra építő változatát, és megmutattam, hogy ez a technika a jelen kontextusban nem alkalmazható hatékonyan. Ezért bevezettem egy új, y-identity néven hívott modellt, amely lehetővé teszi az információ rejtést akár a jelenleginél erősebb támadók esetén is.*

---

**3.1. tézis.** *Megmutattam, hogy ha csak néhány felhasználó is alkalmazza az identitás szeparáció technikát, a támadó által elért eredmények a technika nagyobb léptékű alkalmazása esetén elért eredményeivel lesznek arányosak. Javasoltam egy álca-profilt használó megoldást, amely az irányított információrejtést tette lehetővé, és megmutattam, hogy a legkorszerűbb támadás ezen eljárás alkalmazása esetén csak kevés esetben találja meg az elrejteni szándékozott információt.*
*Kapcsolódó publikációk: [C2, J3]*




11. àbra: A védekező csomópontokra vonatkoztatott találati arány az álca-profil eljárás alkalmazása esetén.

Az egyéni védekezésnél elsőként az a kérdés merül fel, hogy az eddig tárgyal technikákat ha csak néhány felhasználó alkalmazza, akkor hatékonyak maradnak-e. Ezért megvizsgáltam, hogy ha csak pár ezrelék (vagy akár egy) felhasználó védekezik, akkor a támadó mennyi információt fedhet fel róluk. A szimulációs eredményekből kiderült, hogy ez esetekben is ugyanolyan marad a felfedési arány, mint a korábbi kísérletekben, és ezért például jelen esetben is a legalkalmasabb beállítás a legjobb modell, $Y = 5$.

Azonban ezekben az esetekben nem a felhasználó dönti el, melyik identitását fogja felfedezni a támadó. Ezért javasoltam egy álca-profilra építő eljárást, ahol a kapcsolatok 90%-a egy publikus ál-identitásé lesz, valamint a maradék 10% egy rejtett identitásé (plusz 10% átfedésben a kettő között). A védekező felhasználók a 11. ábrán látható eredményeket érték el, amely túlmutat az eddigi legjobb stratégiák eredményein is (vö. 7. ábra). Az álca-profilokra vonatkozó találati arány jellemzően alacsony volt, tipikusan 0.25% és az alatt. Az eljárásnak a sikeressége, illetve az a tény, hogy egy jövőbeni támadó adoptálhatja eljárást ehhez a módszerhez, ösztönzött k-anonymity módszer vizsgálatára.

**3.2. tézis.** *Bevezettem egy módszert a részidentitások újraazonosításának a valószínűségének alsó becslésének mérésére, amelyet a legkorszerűbb támadó algoritmus módosításával hoztam létre. Megmutattam, hogy ezen eljárás segítségével csupán az egy felhasználóhoz tartozó részidentitások töredékét lehet összekötni és újra egyesíteni.*
*Kapcsolódó publikációk: [J1]*

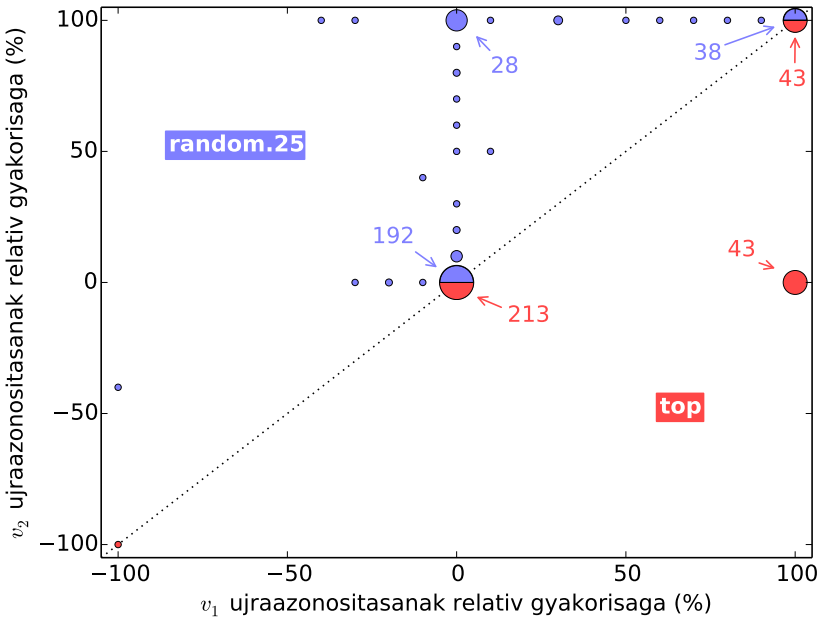
A Nar09 eljárást két tulajdonsága megakadályozta, hogy segítségével a részidentitások közül többet meg lehessen találni, vagy azok megtalálhatóságára becslést adni szimuláció segítségével. Az egyik, hogy a támadás csak egy-az-egyhez párosításokat képes kiadni, másrészt az algoritmus igen determinisztikus eredményt ad. Ezt úgy küszöböltem ki, hogy ha több részidentitást vizsgáltam, akkor az aktuálisan mért részidentitáson kívül valamennyi, ugyanahhoz a felhasználóhoz tartozó részidentitást töröltem. Majd pedig kicseréltem a mérések után a részidentitást egy másikra, és így tovább. A módosítás érdekessége, hogy ezt az eljárást akár egy támadó is alkalmazhatná az identitás szeparáció visszafordítására is.

A vizsgálat során azt tapasztaltam, hogy az alapvető modell ($Y = 2$) esetén az identitás szeparáció kb. 15% felhasználó esetén fordítható vissza, ami magasnak számító érték. Azonban a legjobb modell ($Y = 5$) esetén ugyanez az érték 2.83% volt, és csak az összes eset 1.72%-ban lehetett volna az identitás szeparációt teljesen visszafordítani (néhány törölt él veszteséggel). A 12. ábrán látható néhány eredmény, illetve további részletek a disszertációmban olvashatóak.

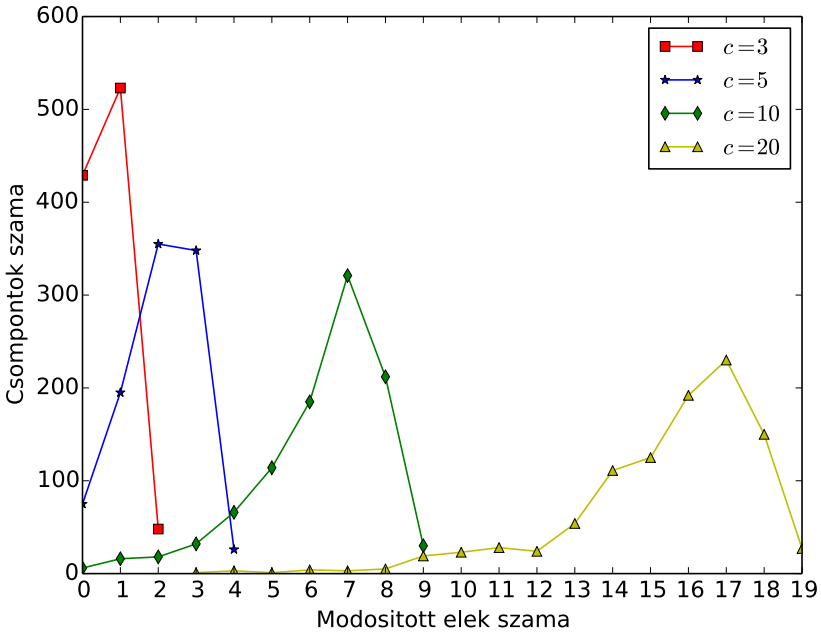
**3.3. tézis.** *Bevezettem a k-anonymity modell elvére épülő $(k, 2)$-anonymity modellt. Elkészítettem a* `K-AnonymizeNode` *algoritmust, amely egy adott csomópontra megmondja, hogy a $(k, 2)$-anonymity modell hogyan lenne alkalmazható, és segítségével megmutattam, hogy a k-anonymity koncepciója a jelenleg tárgyalt támadásokkal szemben nem alkalmazható hatékonyan.*
*Kapcsolódó publikációk: [J1]*

A k-anonymity a kvázi-azonosító fogalmára épül, amely az önmagukban nem azonosító erejű jellemzők együttes alkalmazása által nyert azonosítót jelenti.

**3. definíció.** *k-anonymity modell. Egy adathalmaz akkor teljesíti a k-anonymity feltételét, ha minden benne szereplő entitáshoz található legalább $k - 1$ másik, amelyik ugyanazzal a kvázi azonosítóval rendelkezik [22].*




12. ábra: Az alapvető modell ($Y = 2$) esetén az egyes részidentitások újraazonosítási gyakoriságát szemlélteti az ábra két különböző inicializálási mód esetén (`random.25` és `top`). Az esetek kevesebb mint 15%-ban lehetne visszafordítani az identitás szeparáció folyamatát.




13. ábra: Az Epinions hálózat eredményei a `K-AnonymizeNode` algoritmussal $k = 2$ esetén. Sajnos már kis $c$ értékek esetén is csak relatíve nagyszámú él felvételével sikerült a feltételt teljesíteni.

Ezt az elvet alkalmaztam egyéni felhasználók esetére:

**4. definíció.** $(k,2)$*-anonymity modell. Egy* $v_n \in G$ *felhasználó akkor teljesíti a* $(k,2)$*-anonymity feltételét, ha van legalább* $k$*-1 olyan másik nem szomszédos felhasználó, akiknek pontosan ugyanaz a szomszédságuk, azaz:*

$$\exists A_k = \{v_i : \forall v_i \in V_n^2, V_i = V_n\} \rightarrow |A| = k,$$

*ahol* $V_i$ *jelöli a* $v_i$ *szomszédait, és* $V_i^2$ *pedig a* $v_i$ *szomszédainak szomszédait.*

Ennek megfelelően elkészítettem a `K-AnonymizeNode` algoritmust, amely egy csomóponthoz ellenőrzi, hogyan lehetne megfelelnie a $(k,2)$-anonymity feltételének az identitás szeparáció alkalmazásával. A $k$ paraméteren kívül az algoritmus kér egy $c$ paramétert, amely a rejteni kívánt identitás kapcsolatainak számát adja meg. Ha ez nem lehetséges, az algoritmus minimális számú új kapcsolat felvételével próbálja teljesíteni a modell feltételeit.

A `K-AnonymizeNode` algoritmussal megnéztem a $(k,2)$-anonymity teljesíthetőségét 1000 csomópontra több hálózatban is. Az eredményeket jól szemlélteti a 13. ábra, és demonstrálja a technika hátulütőjét is: már egész kis $c$ értékekre is sok új él felvételével lehetne csak megfelelő privátszféra védelmet biztosítani. Hasonló eredményeket kaptam nagyobb hálózatokban is, és magasabb $k$ értékre is – a részletek a disszertációban megtalálhatóak.

**3.4. tézis.** *Bevezettem az y-identity modellt, amely egy új alternatív megoldás k-anonymity modell helyett. Bebizonyítottam, hogy különböző stratégiák a legalkalmasabbak erős és gyenge támadókkal szemben, valamint azt is, hogy az erős támadó esetén a játékelméleti levezetés által kapott egyensúlyi kevert stratégiát érdemes használni, ha nem ismert a támadó erőssége, mert ennek esetén felülről korlátos a felhasználó privátszféra sérülésből származó várható vesztesége.*
*Kapcsolódó publikációk: [J1]*

Az y-identity modellben a felhasználó készít $y$ új identitást, és véletlenszerűen hozzárendeli valamelyikhez a privátszféra szempontjából érzékenynek tekinthető adatot. A modell feltételezi, hogy a felhasználó racionálisan viselkedik, és a legjobb eredményeket szeretné elérni. Fontos, hogy a rejtendő információ mellé hihető alternatívákat lehessen odatenni, különben maga a rejtés nem sokat ér.

**5. definíció.** *y-identity modell. Egy felhasználó akkor teljesíti a modell feltételeit, ha létrehoz y új identitást (akár egy vagy több hálózatban), majd az egyikhez véletlenszerűen hozzárendeli a privátszféra szempontjából fontos információt, egy adott eloszlás szerint.*

A támadást így a következőképp képzelhetjük el: a támadó először felfedi a felhasználó részidentitásainak egy halmazát (akár az összeset), és ez után vagy kiválasztja az egyiket ezek közül, mint a hiteles információ hordozója, vagy egyiket sem. Ezt egy játékként is leírhatjuk, bár a disszertációmban nem mindegyik támadó esetét vizsgáltam így. A $\mathcal{P}$ jelöli a játékosok halmazát (felhasználó és támadó), a $\mathcal{S}$ jelöli a stratégiák halmazát (az $i$ darab részidentitás közül valamelyik kiválasztása), és az $\mathcal{U}$ a kifizetéseket. Egyes esetekben a támadó csak a felhasználó lehetséges lépéseinek egy részhalmazát látja: $\mathcal{S}' \subset \mathcal{S}$. A felhasználó döntését a $P(R = i) = r_i$ ($\sum_{\forall i} r_i = 1$) eloszlással, a támadóét pedig a $P(Q = i) = q_i$ ($\sum_{\forall i} q_i \leq 1$) eloszlással modelleztem.

Erre építve pedig a következő két támadó típus fogalmazható meg:

1. *Erős támadók*, akik képesek mind az $y$ identitást felfedezni, és a támadó ezzel tisztában is van. Mivel mindkét fél ismeri a lehetséges stratégiák terét, ezért ez az eset játékelméleti megközelítéssel egyszerűen kezelhető.

2. *Gyenge támadók*, akik valamennyi részidentitást képesek felfedezni (akár az összeset), de nem tudják, hogy van-e még a felfedezetteken kívül tovább részidentitás. Formálisabban a támadó csupán $\mathcal{S}' \subseteq \mathcal{S}$ ismerettel rendelkezik, és nem tudja eldönteni, hogy $\mathcal{S}' = \mathcal{S}$. Az egyszerűség kedvéért ebben az esetben a támadó döntését egy eloszlással modelleztem (ami független a felhasználó döntésétől, hiszen nem is ismeri azt), és a felhasználó stratégiáját ezen döntésre való optimalizálásként határoztam meg.

Az *erős támadó* vizsgálatához az *identitás szeparációs játékot* használtam, amely egy egykörös játék, amelyben egyik fél sem ismeri a másik lépését. A játék Nash-egyensúlya [24] egy olyan stratégia pár, amikor egyik játékos sem tud önállóan úgy lépni, hogy a saját kifizetését növelje. Szerencsére John Nash bebizonyította, hogy a véges játékokra mindig létezik kevert stratégia [25], és az alábbi tétel bizonyításával megmutattam a pontos kevert stratégia értékeket az identitás szeparációs játékhoz.

**1. tétel.** *Létezik egy kevert stratégia együttes, amelyik Nash-egyensúly is egyben az identitás szeparációs játékban (y részidentitás mellett), ahol a stratégia valószínűségei* $q_i = \frac{1}{y}, r_i = \frac{1}{y}$ *(*$\forall i$*).*

A bizonyítás a disszertációban megtalálható.

A *gyenge támadók* esetén feltételeztem, hogy a felhasználó meg tudja becsülni részidentitások megtalálási valószínűségét amelyet $P_i$ jelöl (például ahogy megmutattam az 3.2. tézisben). Ha ez nem lehetséges, az ismeretlen támadó esetén javasolt stratégia jó választás

lehet a felhasználó számára. A támadó első lépésben a felhasználó bizonyos identitásait megtalálja, amit jelöljünk a $\mathbf{m}$ vektorral, ahol $m_i \in \mathbf{m}$ jelöli, hogy az $i$. részidentitást sikerült-e felfedeznie ($m_i \in [0,1]$, $m_i = 1$ jelöli a sikert). Ebben az esetben $\mathbf{q_m}$ jelöli a támadó döntési eloszlását, ahol $q_i^{\mathbf{m}} \in \mathbf{q_m}$ jelöli annak a valószínűségét, hogy a támadó az $i$. részidentitás értékét fogja elfogadni.

Ezeknek a segítségével a felhasználó privátszféra sérülésből származó várható veszteségét a következő képlettel lehet leírni:

$$E_w[u_n] = \sum_{\forall \mathbf{m}} \left( \left( \prod_{\forall j} ((1 - m_j) + (-1)^{(1-m_j)} \cdot P_j) \right) \cdot \left( \sum_{\forall i} r_i \cdot q_i^{\mathbf{m}} \right) \right) \cdot u_n^- \tag{4}$$

ahol $i, j \in [1, y]$. A képlet kifejtési módja a disszertációban megtalálható. Azonban ez a képlet érdekes következtetésre vezet el bennünket.

**2. tétel.** *Egy gyenge támadó esetére, ahol ismertek a $\boldsymbol{q_m}$ vektorok ($\forall \boldsymbol{m}$), létezik egy tiszta stratégia halmaz $\mathcal{S}' \subseteq \mathcal{S}$, amelyet használva $E_w[u_n]$ a minimum értékét fogja felvenni. A $\mathcal{S}'$ halmazban szereplő tiszta stratégiák akár kevert stratégiákként is alkalmazhatóak.*

A részletes bizonyítást a disszertáció tartalmazza. A 2. tételből tehát következik, hogy egy gyenge támadóval szemben inkább tiszta stratégiák használata célszerű.

Mivel a két támadó típus esetén nem ugyanaz a stratégia javasolt, a következő életszerű kérdés felmerül: akkor *ismeretlen támadóval* szemben hogy érdemes eljárni? A választandó stratégiával szemben jogosan elvárhatjuk, hogy legalább a k-anonymity szintű védelmet nyújtsa, ezért pedig azt javaslom a disszertációmban, hogy ilyenkor a felhasználó használjon $r_i = \frac{1}{y}$ eloszlást a döntése meghozatalakor. A következő tétel kimondja, hogy ez esetben a felhasználó várható vesztesége legfeljebb annyi lesz, mint a k-anonymity esetén.

**3. tétel.** *Ha a támadó modell adott, de nem ismerjük a támadó pontos típusát, és a felhasználó döntése során az $r_i = \frac{1}{y}$ ($\forall i$) kevert stratégia szerint jár el, akkor privátszféra sérülésből származó vesztesége az alábbi szerint korlátos:*

$$E[u_n] \leq \frac{u_n^-}{y}.$$

A részletes bizonyítás a disszertációban található. Ez a tétel megmutatja, hogy bár a két támadó típus ellen alkalmazandó stratégiák ellentmondásban vannak egymással, mégis az egyensúlyi stratégiát érdemes követni, mivel akkor elvi korlát van a várható veszteségre.

## 6 Eredmények hasznosítása

Az 1. téziscsoportban bemutatott eredményeket új támadási algoritmusok és védekezési módszerek vizsgálatánál lehet alkalmazni. A többi eredmény (2. és 3. téziscsoportok) az identitás szeparáció technika vizsgálatával kapcsolatosak, és egy erre a technikára épülő kliensoldali privátszféra erősítő eszköz létrehozásánál szolgálhatnak hasznos útmutatóul. A disszertációban bemutatott eredményeim megmutatják, hogy hogyan lehetséges hálózati és egyéni szinten megőrizni a privátszférát. Bár ez nem minden esetben tűnik könnyű feladatnak, a munkámban javasoltam olyan egyéni stratégiákat, amelyek ma ismertebb legkorszerűbb algoritmussal szemben képesek a védekezést hatékonnyá tenni, illetve olyanokat is, amelyek akár erősebb, jövőbeni támadók esetén is képesek a privátszféra védelmét megvalósítani.

## Köszönetnyilvánítás

Szeretném megköszönni Dr. Imre Sándor témavezetőmnek a doktoranduszi pályafutásom során nyújtott támogatását. Valamint szeretném megköszönni Dr. Buttyán Leventének a támogatását, hogy a nemzetközi hírű CrySyS Labor munkatársaként dolgozhattam a disszertációm elkészítésének utolsó éveiben. Szeretném megemlíteni, hogy kutatásomhoz támogatást nyújtott a BME-n működő Nagysebességű hálózatok laboratóriuma (HSN Lab), a Mobil Innovációs Központ, valamint a BME-Infokom Innovátor Nonprofit Kft.

## 7 Hivatkozások

[1] "What nsa's prism means for social media users." `http://www.techrepublic.com/blog/tech-decision-maker/what-nsas-prism-means-for-social-media-users/`. Accessed: 2014-05-26.

[2] I. Szekely, "Building our future glass homes–an essay about influencing the future through regulation," *Computer Law & Security Review*, vol. 29, no. 5, pp. 540–553, 2013.

[3] A. Acquisti, B. V. Alsenoy, E. Balsa, B. Berendt, D. Clarke, C. Diaz, B. Gao, S. Gürses, A. Kuczerawy, J. Pierson, F. Piessens, R. Sayaf, T. Schellens, F. Stutzman, E. Vanderhoven, and R. D. Wolf, "D2.1 state of the art," tech. rep., SPION Project.

[4] S. Gurses and C. Diaz, "Two tales of privacy in online social networks," *Security & Privacy, IEEE*, vol. 11, no. 3, pp. 29–37, 2013.

[5] "diaspora*." `https://diasporafoundation.org`. Accessed: 2014-10-31.

[6] A. Sala, X. Zhao, C. Wilson, H. Zheng, and B. Y. Zhao, "Sharing graphs using differentially private graph models," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, (New York, NY, USA), pp. 81–98, ACM, 2011.

[7] C. Y. Ma, D. K. Yau, N. K. Yip, and N. S. Rao, "Privacy vulnerability of published anonymous mobility traces," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, MobiCom '10, (New York, NY, USA), pp. 185–196, ACM, 2010.

[8] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: using social network as a side-channel," in *Proceedings of the 2012 ACM conference on Computer and communications security*, CCS '12, (New York, NY, USA), pp. 628–637, ACM, 2012.

[9] G. Danezis and C. Troncoso, "You cannot hide for long: De-anonymization of real-world dynamic behaviour," in *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, WPES '13, (New York, NY, USA), pp. 49–60, ACM, 2013.

[10] S. Ji, W. Li, J. He, M. Srivatsa, and R. Beyah, "Poster: Optimization based data de-anonymization," 2014. Poster presented at the 35th IEEE Symposium on Security and Privacy, May 18–21, San Jose, USA.

[11] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography," in *Proceedings of the 16th international conference on World Wide Web*, WWW '07, (New York, NY, USA), pp. 181–190, ACM, 2007.

[12] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *Security and Privacy, 2009 30th IEEE Symposium on*, pp. 173–187, 2009.

[13] A. Narayanan, E. Shi, and B. I. P. Rubinstein, "Link prediction by de-anonymization: How we won the kaggle social network challenge," in *The 2011 International Joint Conference on Neural Networks*, pp. 1825–1834, 2011.

[14] W. Peng, F. Li, X. Zou, and J. Wu, "Seed and grow: An attack against anonymized social networks," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012 9th Annual IEEE Communications Society Conference on*, pp. 587–595, 2012.

[15] P. Pedarsani, D. R. Figueiredo, and M. Grossglauser, "A bayesian method for matching two similar graphs without seeds," in *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pp. 1598–1607, Oct 2013.

[16] S. Bartunov, A. Korshunov, S.-T. Park, W. Ryu, and H. Lee, "Joint link-attribute user identity resolution in online social networks," in *Proceedings of the sixth Workshop on Social Network Mining and Analysis*, 2012.

[17] D. Chen, B. Hu, and S. Xie, "De-anonymizing social networks," 2012.

[18] P. Jain, P. Kumaraguru, and A. Joshi, "@i seek 'fb.me': identifying users across multiple online social networks," in *Proceedings of the 22nd international conference on World Wide Web companion*, WWW '13 Companion, (Republic and Canton of Geneva, Switzerland), pp. 1259–1268, International World Wide Web Conferences Steering Committee, 2013.

[19] O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira, "Exploiting innocuous activity for correlating users across sites," in *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, (Republic and Canton of Geneva, Switzerland), pp. 447–458, International World Wide Web Conferences Steering Committee, 2013.

[20] H. Pham, C. Shahabi, and Y. Liu, "Ebm: an entropy-based model to infer social strength from spatiotemporal data," in *Proceedings of the 2013 international conference on Management of data*, pp. 265–276, ACM, 2013.

[21] S. Clauß, D. Kesdogan, and T. Kölsch, "Privacy enhancing identity management: protection against re-identification and profiling," in *Proceedings of the 2005 workshop on Digital identity management*, DIM '05, (New York, NY, USA), pp. 84–93, ACM, 2005.

[22] L. Sweeney, "K-anonymity: A model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 10, pp. 557–570, Oct. 2002.

[23] E. W. Weisstein, "Phase transition." `http://mathworld.wolfram.com/PhaseTransition.html`. Accessed: 2014-11-03.

[24] M. J. Osborne and A. Rubinstein, *A course in game theory.* MIT press, 1994.

[25] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.

# 8 Publikációk

A kiemelt publikációk erősen kötődnek a disszertációmhoz.

## 8.1 Könyvfejezet

[B1] K. Boda, A. M. Földes, G. G. Gulyás, and S. Imre, *Research and Development in E-Business through Service-Oriented Solutions*, ch. Tracking and Fingerprinting in E-Business: New Storageless Technologies and Countermeasures, pp. 134–166. IGI Global, 2013.

[B2] G. G. Gulyás, R. Schulcz, and S. Imre, *Digital Identity and Access Management: Technologies and Frameworks*, ch. Separating Private and Business Identities, pp. 114–132. IGI Global, 2012.

[B3] A. Kóbor, R. Schulcz, and G. G. Gulyás, *Szabad adatok, védett adatok 2.*, ch. Current threats of email - and what we can do against it (in Hungarian), pp. 315–340. INFOTA, 2008.

[B4] G. G. Gulyás, *Szabad adatok, védett adatok 2.*, ch. Using privacy-enhancing identity management in instant messaging services. (in Hungarian), pp. 285–314. INFOTA, 2008.

[B5] G. G. Gulyás, *Studies on information and knowledge processes 13., Alma Mater Series*, ch. Next generation of anonymous web browsers: a bit closer to democracy?, pp. 91–102. INFOTA, 2008.

[B6] G. G. Gulyás, *Tanulmányok az információ- és tudásfolyamatokról 11., Alma Mater Series*, ch. Analaysis of anonymity and privacy in instant messaging services (in Hungarian), pp. 137–158. BME GTK ITM, 2006.

[B7] G. G. Gulyás, *Alma Mater sorozat az információ- és tudásfolyamatokról 10.*, ch. Are anonymous web browsers anonymous? Analysis of solutions and services. (in Hungarian), pp. 9–30. BME GTK ITM, 2006.

## 8.2 Nemzetközi és hazai folyóiratcikkek

[J1] G. G. Gulyás and S. Imre, "Hiding information against structural re-identification," *Telecommunication Systems*, September 2014. (under review).

[J2] B. Simon, G. G. Gulyás, and S. Imre, "Analysis of grasshopper, a novel social network de-anonymization algorithm," *Periodica Polytechnica Electrical Engineering and Computer Science*, January 2015. (accepted for publication).

[J3] G. G. Gulyás and S. Imre, "Using identity separation against de-anonymization of social networks," *Transactions on Data Privacy*, January 2015. (accepted for publication).

[J4] G. G. Gulyás and S. Imre, "Analysis of identity separation against a passive clique-based de-anonymization attack," *Infocommunications Journal*, vol. 4, pp. 11–20, December 2011.

[J5] G. G. Gulyás, R. Schulcz, and S. Imre, "New generation anonymous web browsers (in hungarian)," *Híradástechnika (National Journal)*, vol. 62, no. 8, pp. 24–27, 2007.

## 8.3 Nemzetközi konferencia és workshop cikkek

[C1] G. G. Gulyás and S. Imre, "Measuring importance of seeding for structural de-anonymization attacks in social networks," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, 2014.

[C2] G. G. Gulyás and S. Imre, "Hiding information in social networks from de-anonymization attacks by using identity separation," in *Communications and Multimedia Security* (B. Decker, J. Dittmann, C. Kraetzer, and C. Vielhauer, eds.), vol. 8099 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013.

[C3] G. G. Gulyás and S. Imre, "Measuring local topological anonymity in social networks," in *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pp. 563–570, 2012.

[C4] K. Boda, A. M. Földes, G. G. Gulyás, and S. Imre, "User tracking on the web via cross-browser fingerprinting," in *Information Security Technology for Applications* (P. Laud, ed.), vol. 7161 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2012.

[C5] T. Besenyei, A. M. Földes, G. G. Gulyás, and S. Imre, "Stegoweb: Towards the ideal private web content publishing tool," in *Fifth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2011)* (M. Takesue and R. Falk, eds.), pp. 109–114, August 2011.

[C6] T. Paulik, A. M. Földes, and G. G. Gulyás, "Blogcrypt: Private content publishing on the web," in *Fourth International Conference on Emerging Security Information Systems and Technologies (SECURWARE 2010)*, pp. 123–128, July 2010.

[C7] G. G. Gulyás, R. Schulcz, and S. Imre, "Modeling role-based privacy in social networking services," in *Third International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2009*, pp. 173–178, June 2009.

[C8] G. G. Gulyás, "Design of an anonymous instant messaging service," in *Proceedings of PET Convention 2009.1* (S. Köpsell and K. Loesing, eds.), pp. 34–40, Fakultät Informatik, TU Dresden, March 2009.

[C9] G. G. Gulyás, R. Schulcz, and S. Imre, "Comprehensive analysis of web privacy and anonymous web browsers: are next generation services based on collaborative filtering?," in *Proceedings of the Joint SPACE and TIME Workshops 2008* (L. Capra, I. Wakeman, and M. S. Foukia, Noria, eds.), CEUR Workshop Proceedings, June 2008.

### 8.4 Egyéb

[T1] T. Paulik, A. M. Földes, and G. G. Gulyás, "Publishing private data to the web (in hungarian)," tech. rep., Budapest University of Technology and Economics, 2010.

[T2] S. Dargó and G. G. Gulyás, "Using privacy-enhancing identity management in anonymous web browsers (in hungarian)," tech. rep., Budapest University of Technology and Economics, 2010.