Ben-Gurion University
of the Negev

Hunting Organization-Targeted Socialbots

Abigail Paradise, Asaf Shabtai, Rami Puzis

Information Systems Engineering
Ben-Gurion University of the Negev
Beer-Sheva, Israel

Abundance of

- People
- Personal information
- News
- Opinions
- Impressions
- Offensive content
- Distress signals

- Fewer barriers, limited awareness to privacy

Offenders and defenders often have similar

- Exploit personal information
- Hide within the crowd
- Divert public opinion

Problems
Tools
Algorithms

- Detect misusers
- Identify terrorists
- Protect national secrets

Fake LinkedIn profile gathering info for targeted attacks

Posted on 04 November 2013.

Social networks are great sources of information for cyber criminals and a great way to enter the potential victims' circle of trust.

An ongoing social engineering campaign targeting LinkedIn users has been using the "professional" social network to popularise a specific dating site but, acco[rding to Websense researchers, the final aim of the campaign is] likely mo[netization]

The attac[ker] Jessica F[ox] both to vi[sit] aforemen[tioned]

"Clandestine Fox" Attackers Target Energy Firms via Social Media: FireEye

By Eduard Kovacs on June 11, 2014

An advanced persistent threat (APT) group whose activities have been monitored by FireEye has started using social networks to trick the employees of targeted organizations into installing malware, the company said Tuesday.

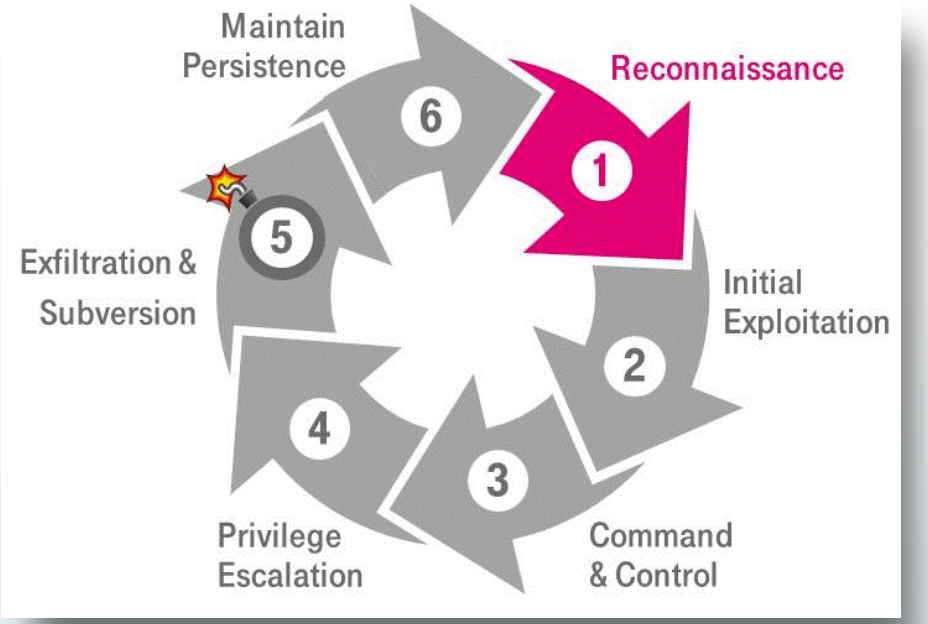
The activities of the actors involved in what FireEye calls "Operation Clandestine Fox" were first brought to light by the security firm back in April when the group leveraged an Internet Explorer zero-day exploit in targeted attacks. In May, researchers spotted a new version of the **attack** specifically targeting Windows XP machines running Internet Explorer 8.

Iranian [hackers use] Facebook to spy on Israelis

Over 2,000 executives and officials fall victims to brazen, complex 3-year 'Newscaster' scam run by social network 'friends,' report says

BY DAVID SHAMAH | May 29, 2014, 9:56 am | 2

APT attack



- APT is a group of sophisticated, determined and coordinated attackers that have been systematically target organizations, government and commercial networks.
- Detecting reconnaissance activities is very difficult since it is performed **outside of the organization's premises** and without direct interaction with the organizational resources.

Extracting information from social networks

- Attacker use online social networks in order to extract useful information about the target organization.
- Information extracted from SNs may include organizational structure, positions and roles, contact information, and other information that may not appear on the official organizational website.

Fire, M., & Puzis, R. (2012). Organization mining using online social networks. *Networks and Spatial Economics*, 1-34.

SocialBots

- Attackers use socialbots or human operated fake accounts to connect to real members of an organization.
- A fake profile in a SN tries to connect to real profiles to gain social capital and harm the organization.
- Attackers employ variety strategies to connect members, ranging from random friend-requests, to sophisticated approaches that **maximize the chance of their targets to accept the request.**

Initial penetration

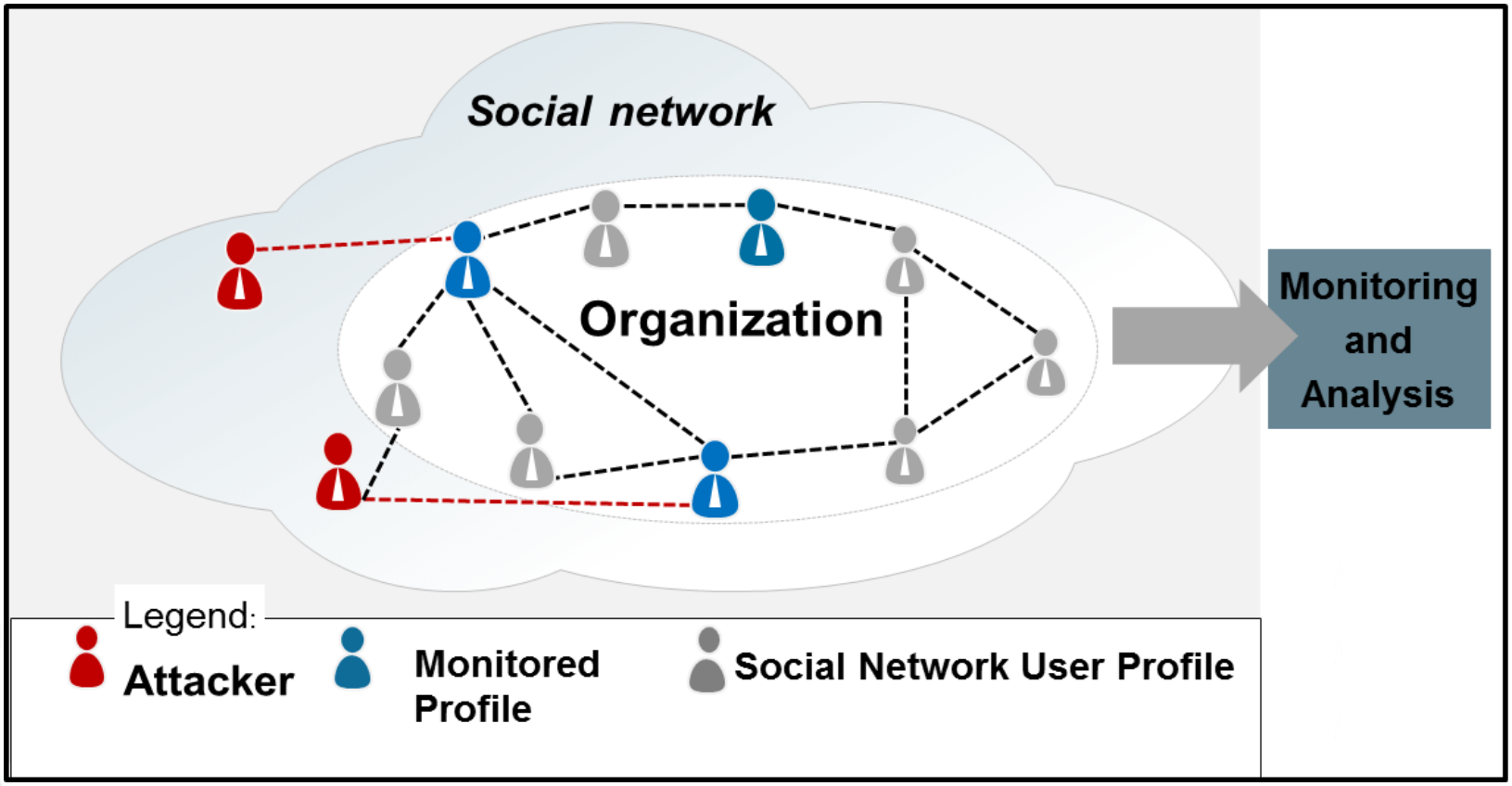
Malware can be injected through:

- Direct messages.
- Group messages.
- Job Postings.
- Status update.
- Emails with malicious attachment.
- **USB drive with conference proceedings**

Social Network Monitoring

- Allow organizations identifying attackers during the early, reconnaissance, phase.
- Option 1: Intelligently select organization member profiles and monitor their activity.
- Option 2: Deploy social network honeypots that mimic vulnerable employees holding key positions.

Profile Monitoring



Profile Monitoring

- Assuming that fake profiles can be identified by careful inspection.
- Assuming that the organization is legally allowed to inspect the activities of the monitored employees in the SN.
- Inspect friend requests received by the selected profiles.
  - Such inspection may include investigation of profiles behind each friend request; for example, validating the indicated place of work, educational background, contact information.

Profile Monitoring - cost

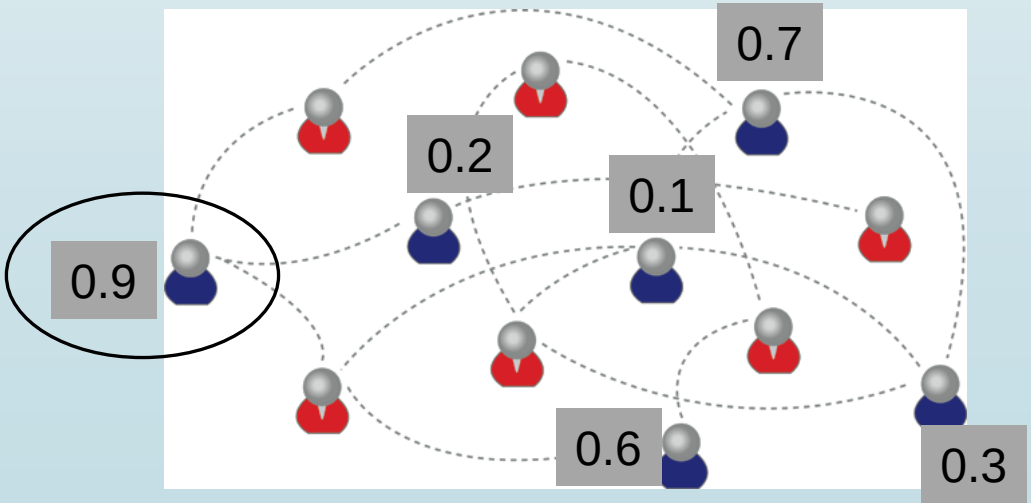
- The main **challenge** is to accurately select a small number of profiles with a minimal monitoring cost.
- In general, the more friends a profile has, it will have a larger number of incoming friend requests, wall posts, and comments that need to be analyzed.
- We assume that the **cost** of monitoring a profile is **proportional to its number of friends**.

Monitoring strategies

- **Random** (***def_rnd***) - The monitored profiles are chosen randomly among the employees of the organization.
- **Preferential attachment** (***def_mc***) - Profiles that have more friends in the organization are more likely to be chosen.
- **Eigenvector centrality** (***def_e***) - Assigns each profile a score that is proportional to the sum of the scores of its neighbors.
- **PageRank** (**def_pr**) - Measuring the importance of each profile.

Profile selection for monitoring

- For each profile in the organization's SN we calculate the measures based on the monitoring strategy currently employed.
- The probability of employee x to be selected is p_x and is proportional to the scores of the other employees.



Attack strategies

- No knowledge attacker
- Partial knowledge attacker
- Full knowledge attacker

Attack strategies

We consider and evaluated attack strategies with different knowledge about the employed defenses:

- **No knowledge attacker** - This attacker is unaware of the use of any monitoring strategy.
- **Partial knowledge attacker -** The attacker is aware of the probability of each organization member being selected to be monitored (p_x).
- **Full knowledge attacker** - This attacker has full knowledge about the monitoring strategy and avoids monitored profiles.

No knowledge attacker

- ***attk_rnd*** - A baseline approach in which the attacker randomly sprays friend-requests.
- ***attk_opt -*** An attacker can estimate the probability of a SN user accepting a friend request given the total number of friends the user has and the number of common friends with the attacker. $\underset{x \in O_s}{\operatorname{argmax}}(P_{accept}(x))$

Boshmaf, Y., et al. (2011). The socialbot network: when bots socialize for fame and money. 27th Annual Computer Security Applications Conference (93-102).

Partial knowledge attacker

- ***attk_adv1-*** Sends a friend request to the employee with the lowest chance of being selected for monitoring $\underset{x \in O_S}{\operatorname{argmin}}(P_x)$.
- ***attk_adv2***- This attack strategy considers the acceptance probability and chance of a profile to be selected for monitoring $\underset{x \in O_S}{\operatorname{argmax}} \left(\frac{P_{accept}(x)}{P_x(x)} \right)$

Full knowledge attacker

- ***attk_rnd_known*** - This attacker focuses only on profiles that are not monitored de-facto, while using the same attack strategy as *attk_rnd*.

- ***attk_opt_known*** - This attacker focuses only on profiles that are not monitored de-facto, while using the same priorities as the strategy *attk_opt*.

Evaluation

- We conducted a set of extensive simulations based on real data topologies.
- Research questions:
  - Which method for selecting the monitored profiles **increases** the chances of detecting socialbots while **minimizing** the overall monitoring cost?
  - What are the best monitoring strategies against each attack strategy?

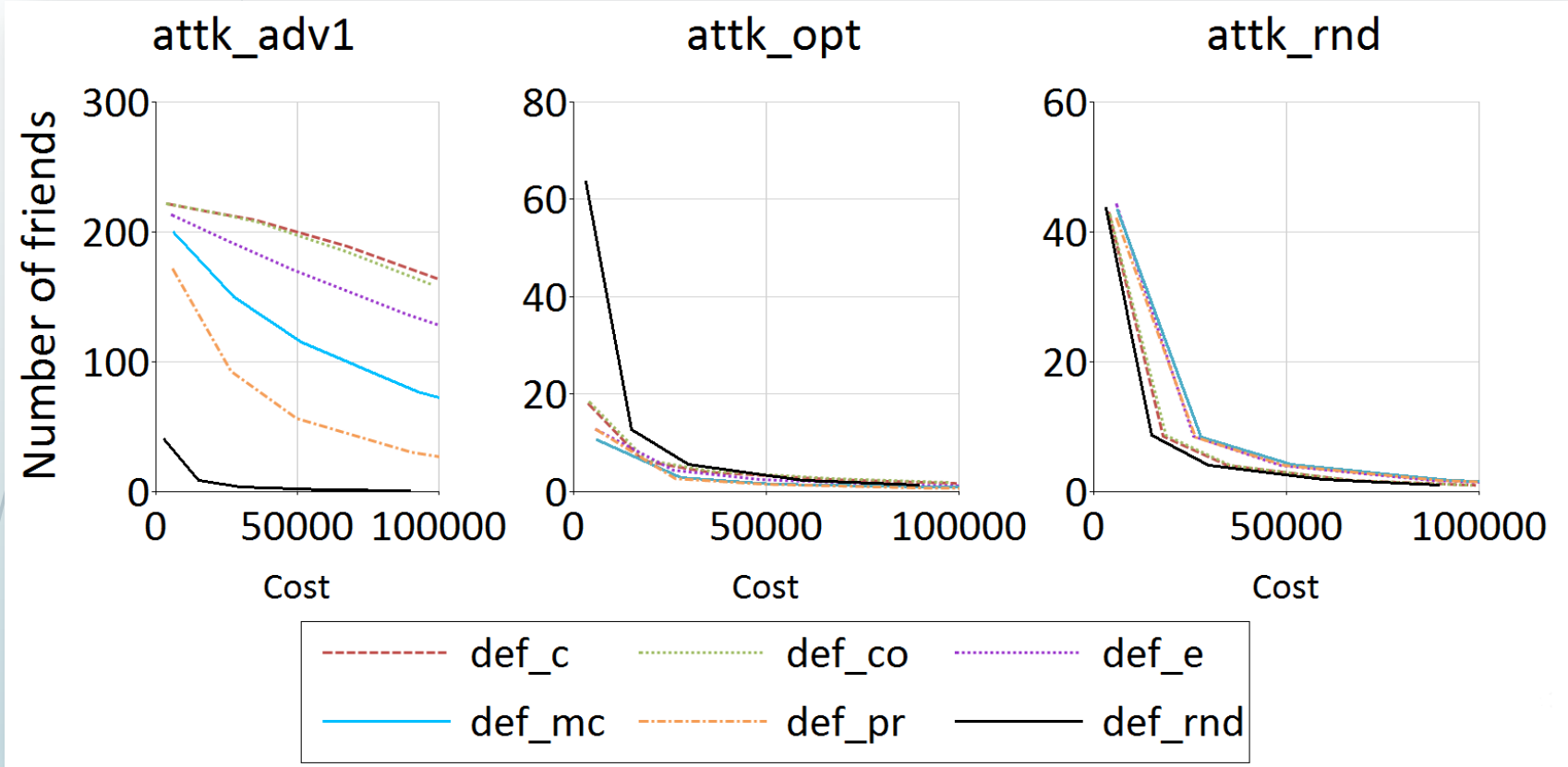
Evaluation - parameters

- **Social network**: Orkut, LiveJournal, Friendster
- **Organization**: 50 organizations selected from each SN
- **Monitoring strategy**: *def_rnd, def_c, def_co, def_pr, def_e, def_mc*
- **Number of monitored profiles**: 1%, 5%, 10%, 20%, 30% of organization size and 40%, 50%, 60%, 70%, 80%, 90% tested only for the full knowledge attacks.
- **Attack method:**
  - No knowledge: *attk_rnd, attk_opt*
  - Partial knowledge: *attk_adv1, attk_adv2*
  - Full knowledge: *attk_rnd_known, attk_opt_known*

Evaluation - measures

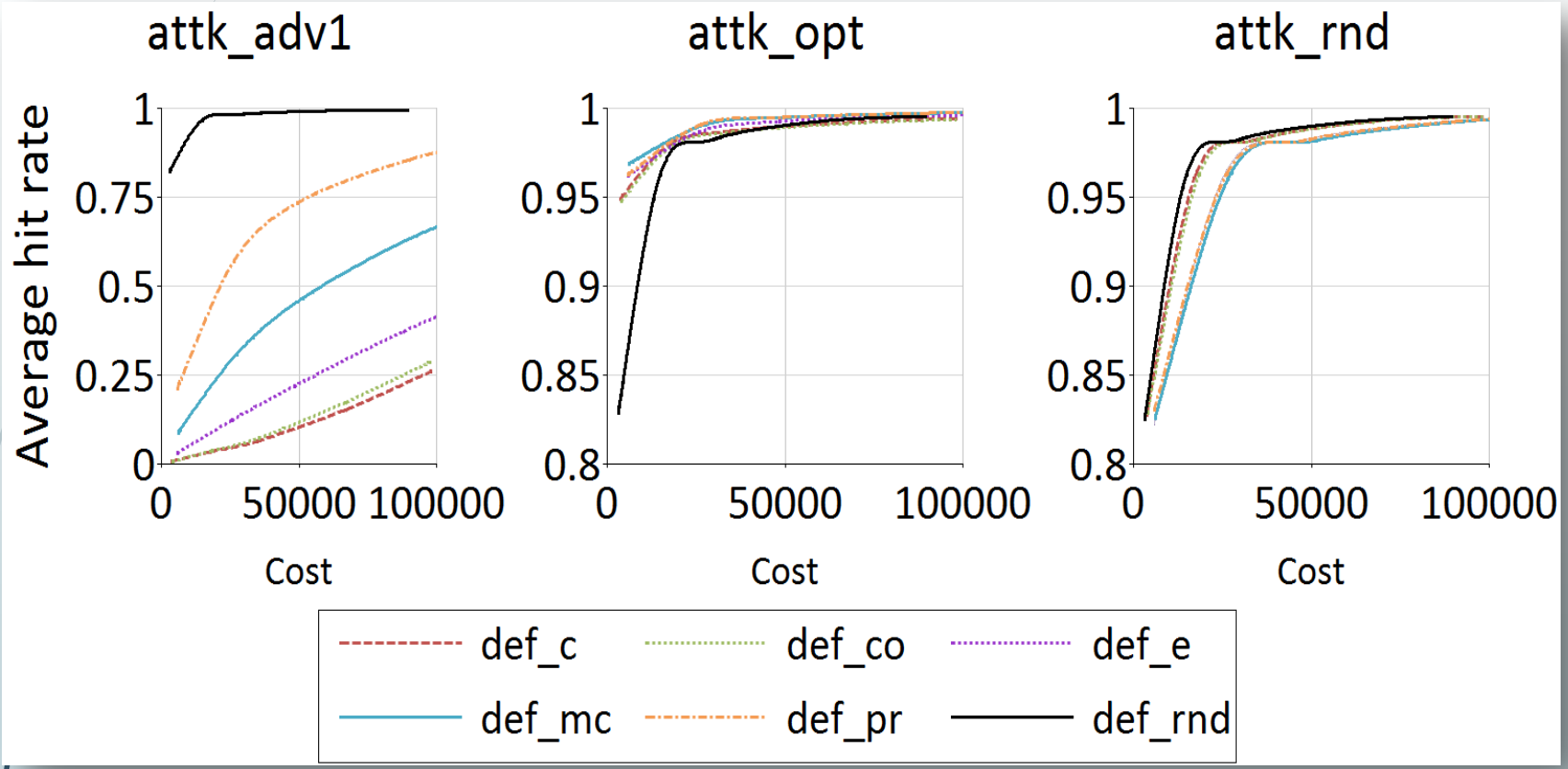
- **Acceptance rate** measures the effectiveness of the attack strategies by calculating the fraction of accepted friend requests out of the total friend request sent by the attacker.
- **Hit rate** the cumulative probability of contacting a monitored profile by at least one of the requests sent by the attacker. Hit rate was calculated for each friend request sent.
- **Average hit rate -** The average hit rate across all friend requests.
- **Number of friends before hit** is the number of friends obtained before hitting a monitored profile for the first time.
- **Monitoring cost** is the sum of the connectivity degrees of the set of monitored profiles.

Results – social capital



- ***def_mc*** **and** ***def_pr*** **are optimal against** ***attk_opt*****.**
- **def_rnd is the most effective against partial knowledge attackers and the random attacker**

Results – hunting



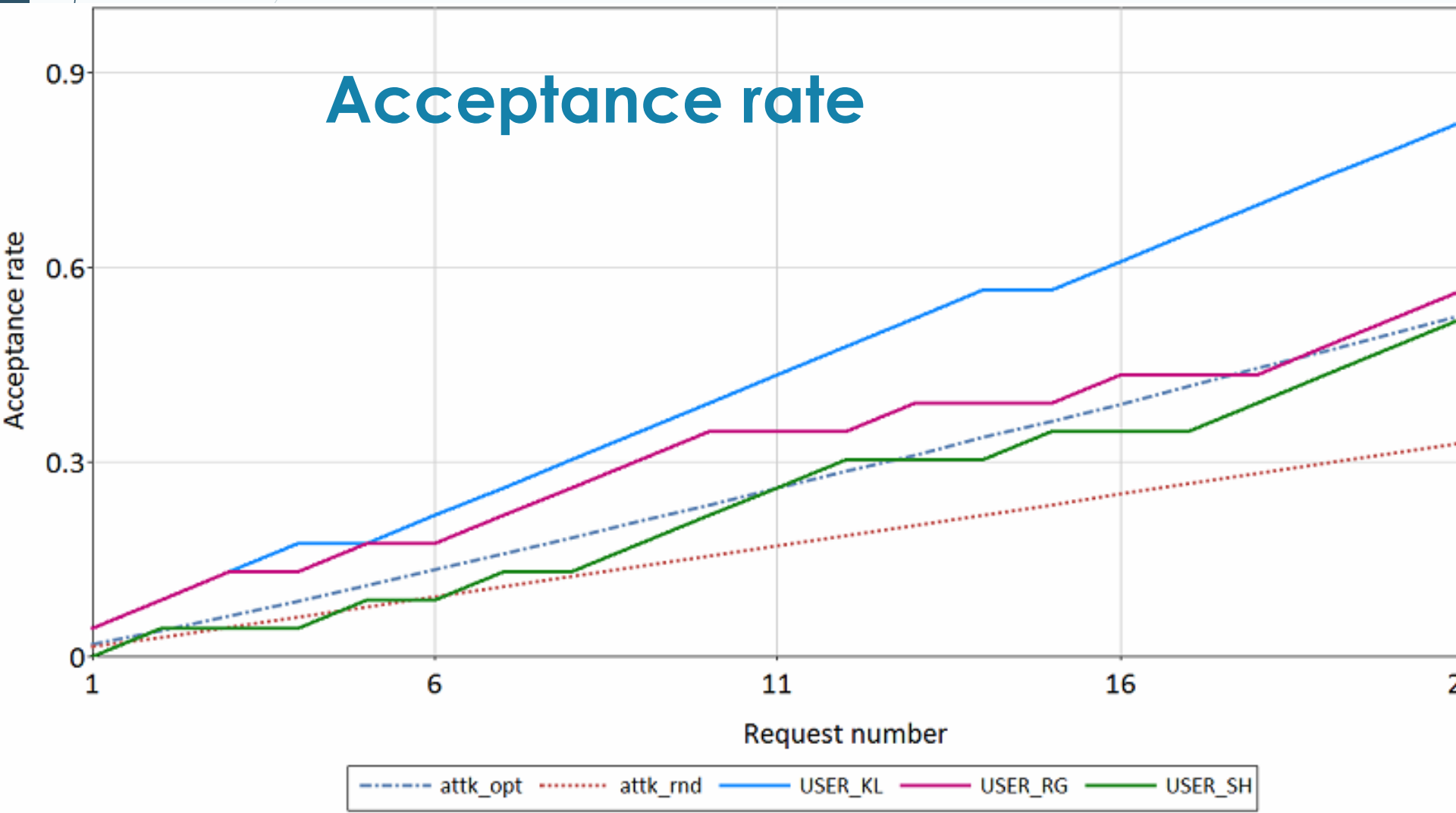
- ***def_mc*** **and** ***def_pr*** **are preferable against** ***attk_opt.***
- **def_rnd is the most effective against partial knowledge attackers and the random attacker**

Fresh results

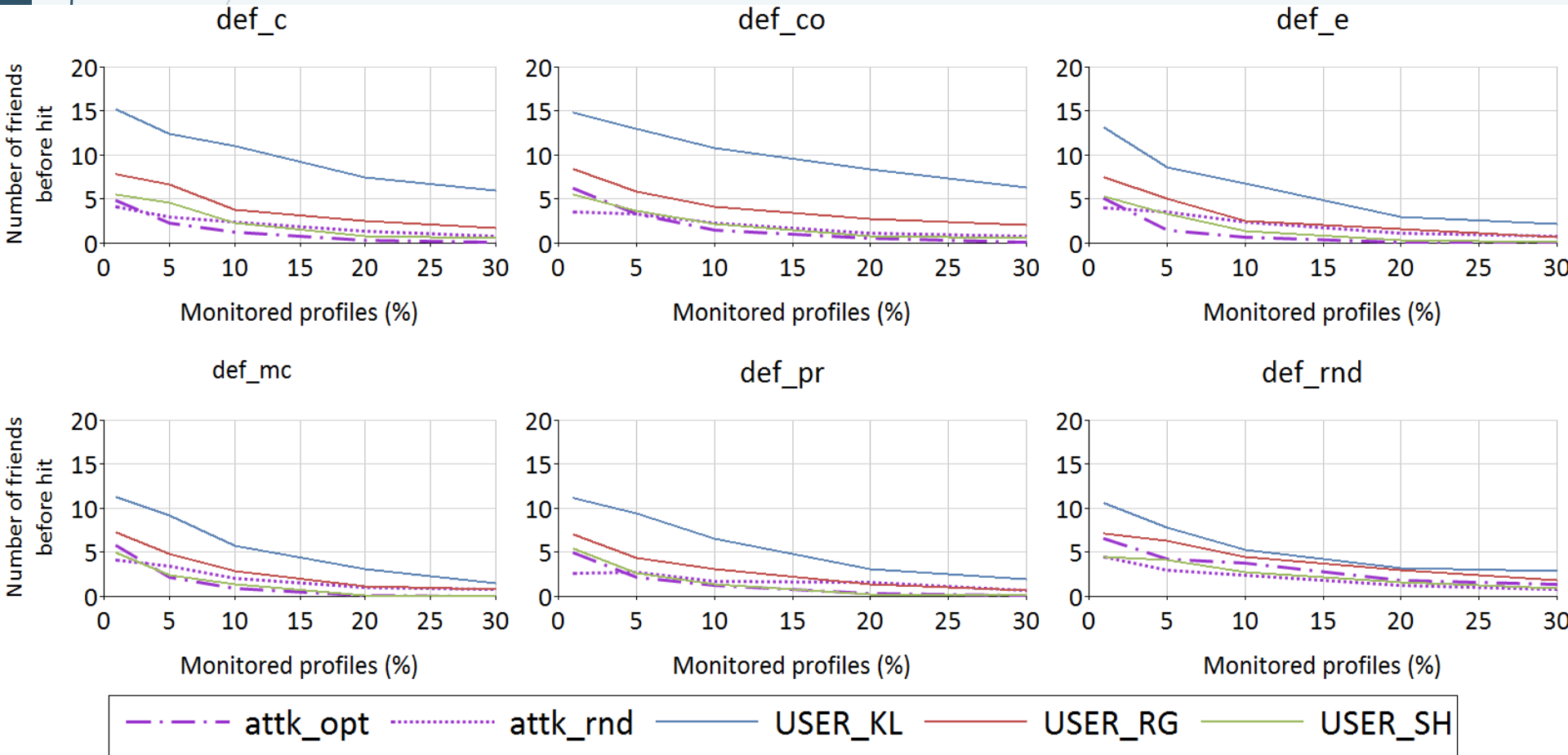
- In future, we plan to conduct an evaluation of the proposed monitoring strategies using real data of an intrusion via artificial profiles.
  - Real scenario
  - Using real company social network
  - No need to simulate the attack strategies!

Case study:
Real organization – fake profiles

Acceptance rate



Case study:
Real organization – fake profiles
Social capital



Questions