Abusing Social Networks for Automated User Profiling

Presented by
Niraj 2009CS50249
Sidhartha Gupta 2009CS50490

Recent Advances in Intrusion

Marco Balduzzi, Christian Platzer, Thorsten Holz
Engin Kirda, Davide Balzarotti and Christopher Kruegel

Motivations

- Social Networks have experienced a huge surge in popularity

  - The amount of personal information they store requires appropriate security precautions

  - Many users tend to be overly revealing when publishing personal information

  - People are not aware of all the possible way in which these info can be abused

- A simple problem can result in serious consequences for thousands of Social Networks users

The Finding-Friends functionality

Social Networks usually provide an email-based functionality to search for existing friends
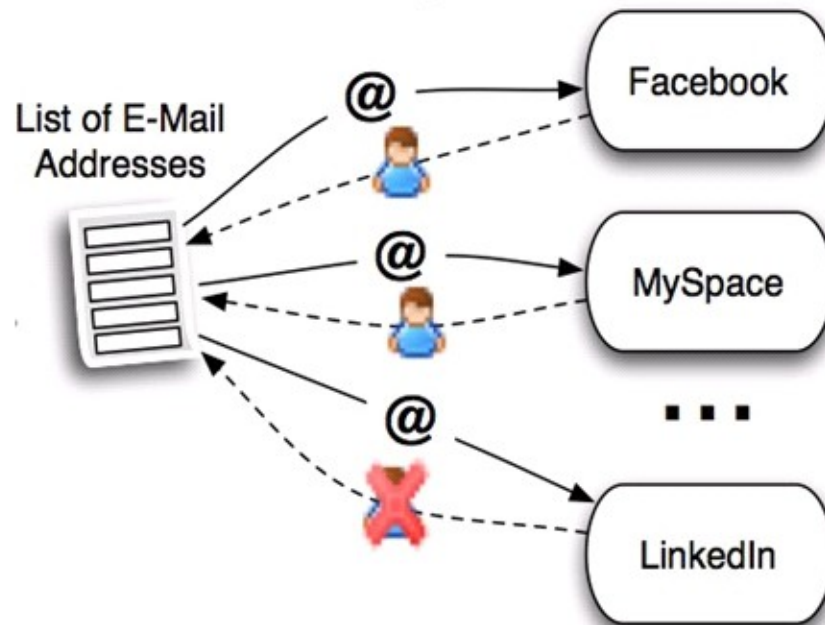
What is the Problem?

- Problems of the Finding-Friends functionality:

  - Map a profile to an email
  (normally considered a private information)

  - Bulk queries of thousands of emails (10,000 on FB)

  - Recursive queries via email fuzzing on user friends

How to build the Identity of a person?

- Different profiles with the same e-mail address belong to the same person

- The e-mail address can be used to build the identity of a person extracting info from multiple social networks
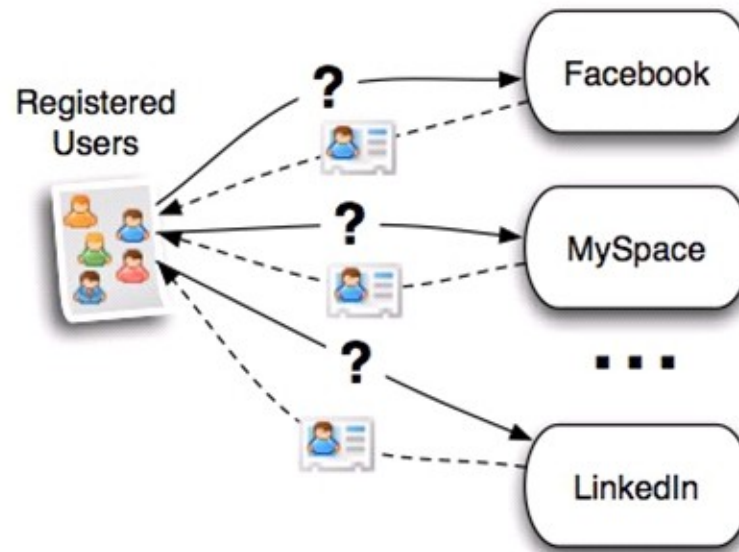
Impact

- Validation of e-mail addresses on large scale for massive spam attacks

  - Fast and automated

- Enrich e-mail addresses with private user information for targeted attacks

  - Phishing, social engineering

  - Information gathering

- Detection of inconsistent profile informations
- Discovering of "hidden" identities

System Overview
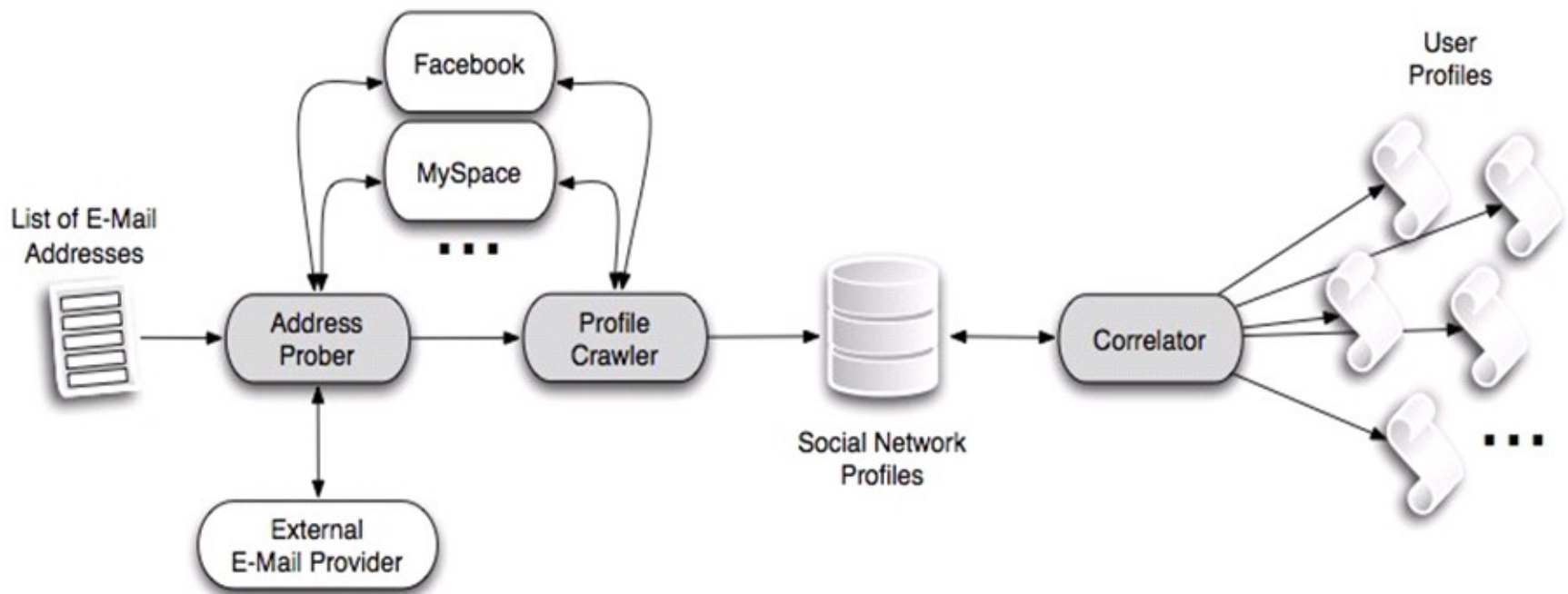
Experiments

- Identified 8 popular Social Networks providers that support the   Finding-Friends   functionality:
  - Facebook, MySpace, Twitter, LinkedIN, Friendster, Badoo, Netlog, XING

- Input data: 10 million e-mail addresses

- Two phases:

  1. Discover user profiles by e-mail querying

  2. Profile user identities by e-mail correlation

Discovered profiles

	Network	Query method method	E-mail list length size efficiency	# queried e-mails speed efficiency	# identified accounts	Percentage
1	Facebook	Direct	5000	10M/day	517,747	4.96%
2	MySpace	GMail	1000	500K/day	209,627	2.01%
3	Twitter	GMail	1000	500K/day	124,398	1.19%
4	LinkedIn	Direct	5000	9M/day	246,093	2.36%
5	Friendster	GMail	1000	400K/day	42,236	0.41%
6	Badoo	Direct	1000	5M/day	12,689	0.12%
7	Netlog	GMail	1000	800K/day	69,971	0.67%
8	XING	Direct	500	3.5M/day	5,883	0.06%
				Total of	1,228,644	

Discovered profiles

	Network	Query method	E-mail list length	# queried e-mails	# identified	Percentage
		method	*size efficiency*	*speed efficiency*	accounts	
1	Facebook	Direct	5000	10M/day	517,747	4.96%
2	MySpace	GMail	1000	500K/day	209,627	2.01%
3	Twitter	GMail	1000	500K/day	124,398	1.19%
4	LinkedIn	Direct	5000	9M/day	246,093	2.36%
5	Friendster	GMail	1000	400K/day	42,236	0.41%
6	Badoo	Direct	1000	5M/day	12,689	0.12%
7	Netlog	GMail	1000	800K/day	69,971	0.67%
8	XING	Direct	500	3.5M/day	5,883	0.06%
				Total of	1,228,644	

Discovered profiles

	Network	Query method method	E-mail list length *size efficiency*	# queried e-mails *speed efficiency*	# identified accounts	Percentage
1	Facebook	Direct	5000	10M/day	517,747	4.96%
2	MySpace	GMail	1000	500K/day	209,627	2.01%
3	Twitter	GMail	1000	500K/day	124,398	1.19%
4	LinkedIn	Direct	5000	9M/day	246,093	2.36%
5	Friendster	GMail	1000	400K/day	42,236	0.41%
6	Badoo	Direct	1000	5M/day	12,689	0.12%
7	Netlog	GMail	1000	800K/day	69,971	0.67%
8	XING	Direct	500	3.5M/day	5,883	0.06%
				Total of	1,228,644	

Profiling of the user identities

# SN	# Profiles
1	608,989
2	199,161
3	55,660
4	11,483
5	1,478
6	159
7	11
8	0
Total unique	876,941

Combination	Occurrences
Facebook - MySpace	57,696
Facebook - LinkedIn	49,613
Facebook - Twitter	25,759
Facebook - MySpace - Twitter	13,754
Facebook - LinkedIn - Twitter	13,733
Facebook - NetLOG	12,600
Badoo - FriendSter	11,299
Facebook - MySpace - LinkedIn	9,720
LinkedIn - Twitter	8,802
MySpace - Twitter	7,593

Parsing sensitive information

	Age	Sex	Spoken language	Job	Education	Current relation	Searched relation	Sexual preference
Facebook	0.35	0.50	n/a	0.23	0.23	0.44	0.31	0.22
MySpace	82.20	64.87	n/a	3.08	2.72	8.41	4.20	4.07
Twitter	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LinkedIn	n/a	n/a	n/a	96.79	60.68	0.00	n/a	n/a
Friendster	82.97	87.45	n/a	30.88	2.72	64.59	77.76	n/a
Badoo	98.61	98.61	47.81	17.06	19.92	22.48	n/a	22.80
Netlog	97.66	99.99	44.56	43.40	1.64	25.73	23.14	29.30
XING	n/a	n/a	84.54	99.87	49.21	n/a	n/a	n/a

Information Mismatch

Information	Value	% Total mismatches	% of mismatched values		
			2	3	4+
Name	*string*	72.65	62.70	35.37	17.66
Location	*city*	53.27	51.74	16.24	3.72
Age	$0 < n < 100$	34.49	33.58	17.84	30.56
Sex	*male, female*	12.18	12.18		
Sexual preference	*straight, homosexual, bisexual*	7.63	7.63		
Current relationship	*single, in a relationship, married, complicated*	35.54	35.42	5.13	

Information Mismatch

Information	Value	% Total mismatches	% of mismatched values 2	3	4+
Name	*string*	72.65	62.70	35.37	17.66
Location	*city*	53.27	51.74	16.24	3.72
Age	$0 < n < 100$	34.49	33.58	17.84	30.56
Sex	*male, female*	12.18	12.18		
Sexual preference	*straight, homosexual, bisexual*	7.63	7.63		
Current relationship	*single, in a relationship, married, complicated*	35.54	35.42	5.13	

Information Mismatch

Information	Value	% Total mismatches	% of mismatched values		
			2	3	4+
Name	*string*	72.65	62.70	35.37	17.66
Location	*city*	53.27	51.74	16.24	3.72
Age	$0 < n < 100$	34.49	33.58	17.84	30.56
Sex	*male, female*	12.18	12.18		
Sexual preference	*straight, homosexual, bisexual*	7.63	7.63		
Current relationship	*single, in a relationship, married, complicated*	35.54	35.42	5.13	

Mismatches in Provided Age Information

- 1/3 of the profiles have a mismatch in the age (of at least two years)

Range	#	%
2 - 10	4,163	60.17
11 - 30	1,790	25.87
31 +	966	13.96
Profiles with Age	20,085	
Total mismatched	6,919	

- Underage users claim to be > 18 to register on dating sites (Badoo)

Countermeasures

1. Do not provide a direct map between e-mail and user
(e.g. returning a list of registered accounts in random order)

2. CAPTCHA

3. Require contextual information to acknowledge the data

4. Raising awareness (e.g. use a different e-mail for sites with personal information)

Conclusions

- "Finding-Friends" can be abused to:
  1. Map thousands of profiles to their personal   e-mails
  2. Correlate profiles across multiple networks to build user identities
- Implemented an Automated System that shows
  the real impact of the attack
- Run experiments on 8 popular SN providers,
  most of which acknowledge/fixed the problem