The Doppelgänger Bot Attack: Exploring Identity Impersonation in Online Social Networks

Oana Goga (MPI-SWS), Giridhari Venkatadri (MPI-SWS),
Krishna P. Gummadi (MPI-SWS)

IMC 2015, October 28th

Max
Planck
Institute
for
Software Systems

Weak identities

Are unverified identities that do not require users to prove their online identities match their offline person.

Weak identities

Are unverified identities that do not require users to prove their online identities match their offline person.

✓ Lower sign-on barriers, provide anonymity

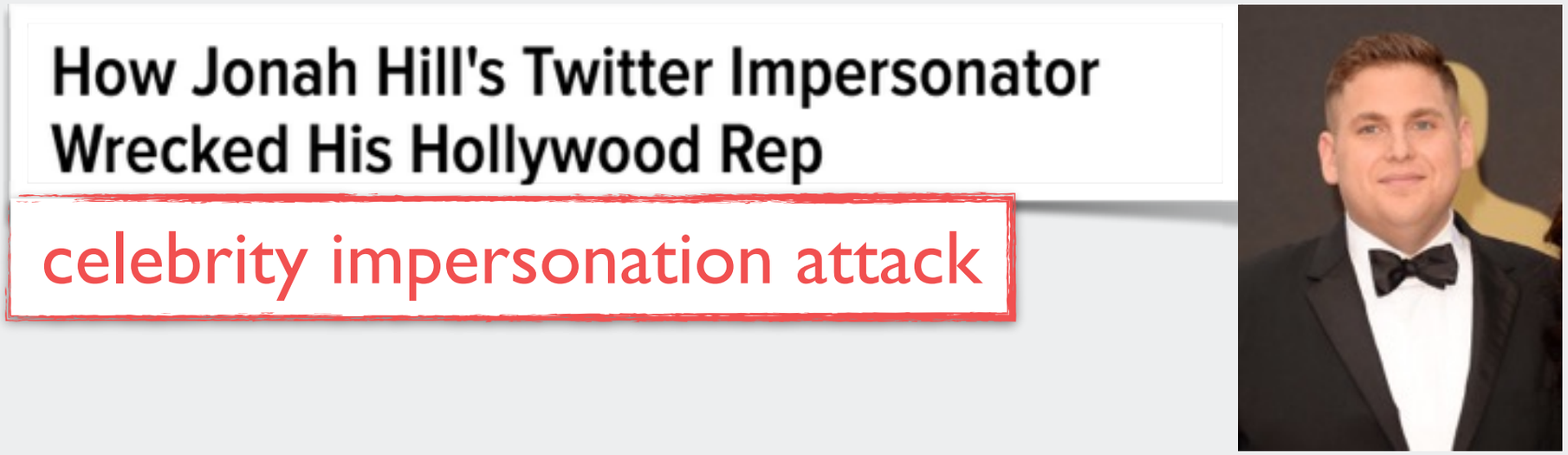
✗ Leave systems vulnerable to Sybil attacks (fake identities)

Identity impersonation attacks

Special class of fake identities attacks: the attacker spoofs the identity of another real-world user.

Identity impersonation attacks

Special class of fake identities attacks: the attacker spoofs the identity of another real-world user.



celebrity impersonation attack

Identity impersonation attacks

Special class of fake identities attacks: the attacker spoofs the identity of another real-world user.

Identity impersonation attacks

Special class of fake identities attacks: the attacker spoofs the identity of another real-world user.

- Damage the online image of victims & affect victims in the offline world!

Identity impersonation attacks

Special class of fake identities attacks: the attacker spoofs the identity of another real-world user.

- Damage the online image of victims & affect victims in the offline world!
- Impersonation attacks are increasingly easy to mount due to the availability of personal information online!

Current situation

- Lack of understanding of impersonation attacks online!
  - No large dataset about real-world impersonation attacks

Current situation

- Lack of understanding of impersonation attacks online!
  - No large dataset about real-world impersonation attacks
- Lack of frameworks to automatically detect impersonation attacks online
  - Detection relies on manual reports

Contributions

First extensive study of real-world impersonation attacks in online social networks.

1. Methodology to gather data about impersonation attacks
2. Characterization of impersonation attacks in Twitter
3. Automatic detection of impersonation attacks

Contributions

First extensive study of real-world impersonation attacks in online social networks.

1. Methodology to gather data about impersonation attacks
2. Characterization of impersonation attacks in Twitter
3. Automatic detection of impersonation attacks

Challenges in data gathering

People results for **nick feamster**

Nick Feamster @feamster
Associate Professor of Computer Science, Georgia Tech
Followed by Pablo Rodriguez and 2 others

Follow

Nicholas Feamster @ntfeamster

Follow

Nick Feamster @feamster_
Associate Professor of Computer Science, Georgia Tech

Follow

Challenges in data gathering

How to determine which identities try to portray the same user?

Nick Feamster @feamster
Associate Professor of Computer Science, Georgia Tech
Followed by Pablo Rodriguez and 2 others

Follow

Nicholas Feamster @ntfeamster

Follow

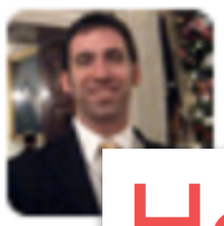
Nick Feamster @feamster_
Associate Professor of Computer Science, Georgia Tech

Follow

Challenges in data gathering

How to determine which identities try to portray the same user?

How similar the profiles of two identities should be to qualify as portraying the same user?



Nick Feamster @feamster
Associate Professor of Computer Science,

Follow

Nick Feamster @feamster_
Associate Professor of Computer Science,
Georgia Tech

Follow

Challenges in data gathering

Challenges in data gathering

How to determine if a doppelgänger pair is an impersonation attacks?

Challenges in data gathering

How to determine if a doppelgänger pair is an impersonation attacks?

victim-impersonator pair

Nick Feamster @feamster

Associate Professor of Computer Science,
Georgia Tech

Follow

Nicholas Feamster @ntfeamster

Follow

Nick Feamster @feamster_

Associate Professor of Computer Science,
Georgia Tech

Follow

Challenges in data gathering

How to determine if a doppelgänger pair is an impersonation attacks?

Nick Feamster @feamster

Associate Professor of Computer Science,
Georgia Tech

Follow

victim-impersonator pair

avatar-avatar pair

Nicholas Feamster @ntfeamster

Follow

Nick Feamster @feamster_

Associate Professor of Computer Science,
Georgia Tech

Follow

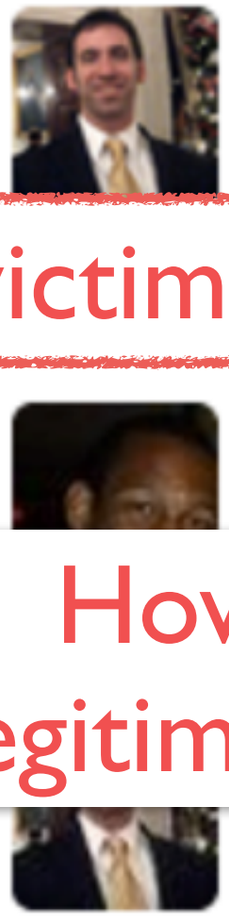
Challenges in data gathering

How to determine if a doppelgänger pair is an impersonation attacks?

victim-impersonator pair

avatar-avatar pair

How to determine which identity is legitimate and which is an impersonator?



Challenge 1:

Identifying doppelgänger pairs

- Identify pairs of identities that most humans believe they portray the same person
  - Every identity has a name, location, bio and photo
  - Automated rule-based matching scheme (trained on human-annotated data, determines when the profile attributes of two identities matches sufficiently)

Challenge 2 and 3

Challenge 2 and 3

Identify victim-impersonator pairs

- Exploit Twitter suspension signals: when Twitter suspends one but not both identities

Challenge 2 and 3

Identify victim-impersonator pairs

- Exploit Twitter suspension signals: when Twitter suspends one but not both identities

Identify avatar-avatar pairs

- Exploit interactions between identities: clear indication that one identity is aware of the other

Challenge 2 and 3

Solves challenge 3 as well!

impersonating identity = suspended identity

Identify victim-impersonator pairs

- Exploit Twitter suspension signals: when Twitter suspends one but not both identities

Identify avatar-avatar pairs

- Exploit interactions between identities: clear indication that one identity is aware of the other

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Contributions

First extensive study of real-world impersonation attacks in online social networks.

1. Methodology to gather data about impersonation attacks
2. Characterization of impersonation attacks in Twitter
3. Automatic detection of impersonation attacks

Contributions

First extensive study of real-world impersonation attacks in online social networks.

1. Methodology to gather data about impersonation attacks
2. Characterization of impersonation attacks in Twitter
3. Automatic detection of impersonation attacks

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Types of impersonation attacks

Types of impersonation attacks

- Celebrity impersonation attacks
  - Purpose: exploits or maligns the reputation of the victim
  - Detection: victim has more than 10,000 followers or is verified

Types of impersonation attacks

- Celebrity impersonation attacks ← 3% (in the random dataset)
  - Purpose: exploits or maligns the reputation of the victim
  - Detection: victim has more than 10,000 followers or is verified

Types of impersonation attacks

- Celebrity impersonation attacks ← 3% (in the random dataset)
  - Purpose: exploits or maligns the reputation of the victim
  - Detection: victim has more than 10,000 followers or is verified
- Social engineering attacks
  - Purpose: abuses victim's friends: reveal sensitive info, send money
  - Detection: attacker contacts victim's friends

Types of impersonation attacks

- Celebrity impersonation attacks ← 3% (in the random dataset)
  - Purpose: exploits or maligns the reputation of the victim
  - Detection: victim has more than 10,000 followers or is verified
- Social engineering attacks ← 2% (in the random dataset)
  - Purpose: abuses victim's friends: reveal sensitive info, send money
  - Detection: attacker contacts victim's friends

Types of impersonation attacks

- Ce taset)
  - ified
- Social engineering attacks ← 2% (in the random dataset)
  - Purpose: abuses victim's friends: reveal sensitive info, send money
  - Detection: attacker contacts victim's friends

Most impersonation attacks do not target celebrities or try to mount social engineering attacks!

Types of impersonation attacks

- Ce… …taset)
  - …
  - …ified
- Social engineering attacks ← 2% (in the random dataset)
  - …
  - Detection: attacker contacts victim's friends

Most impersonation attacks do not target celebrities or try to mount social engineering attacks!

What is possibly motivating the attackers?

Doppelgänger bot attacks hypothesis

H1: The attackers create these identities to abuse Twitter (and not the victims)

H2: The attackers attempt to create real-looking fake identities to evade the Twitter Sybil defense system

Doppelgänger bot attacks hypothesis

H1: The attackers create these identities to abuse Twitter (and not the victims)

H2: The attackers attempt to create real-looking fake identities to evade the Twitter Sybil defense system

doppelgänger bot attacks

Doppelgänger bot attacks hypothesis

H1: The attackers create these identities to abuse Twitter (and not the victims)

H2: The attackers attempt to create real-looking fake identities to evade the Twitter Sybil defense system

doppelgänger bot attacks

≠ doppelgänger pair!

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Twitter dataset

	RANDOM DATASET	BFS DATASET
initial accounts	1.4 million	142,000
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605

Doppelgänger bot attacks evidence for hypothesis I

H1: The attackers create these identities to abuse Twitter (and not the victims)

Doppelgänger bot attacks evidence for hypothesis I

H1: The attackers create these identities to abuse Twitter (and not the victims)

Evidence:

- Large number of impersonators follow the same users
- The users they follow are suspected of having bought fake followers (http://trulyfollowing.app-ns.mpi-sws.org/)

Doppelgänger bot attacks evidence for hypothesis I

H1: The attackers create these identities to abuse Twitter (and not the victims)

Evidence:

- Large number of impersonators follow the same users
- The users they follow are suspected of having bought fake followers (http://trulyfollowing.app-ns.mpi-sws.org/)

follower fraud

Doppelgänger bot attacks evidence for hypothesis 2

H2: Attackers create real-looking fake identities to evade the Twitter Sybil defense system

Doppelgänger bot attacks evidence for hypothesis 2

H2: Attackers create real-looking fake identities to evade the Twitter Sybil defense system

Evidence:

- Twitter took in median 278 days to suspend the impersonating identities
- Other traditional Sybil detection schemes perform badly

Doppelgänger bot attacks evidence for hypothesis 2

H2: Attackers create real-looking fake identities to evade the Twitter Sybil defense system

Can we do something to detect impersonating identities faster?

Eviden

- Twit rsonating identities
- Other traditional Sybil detection schemes perform badly

Contributions

First extensive study of real-world impersonation attacks in online social networks.

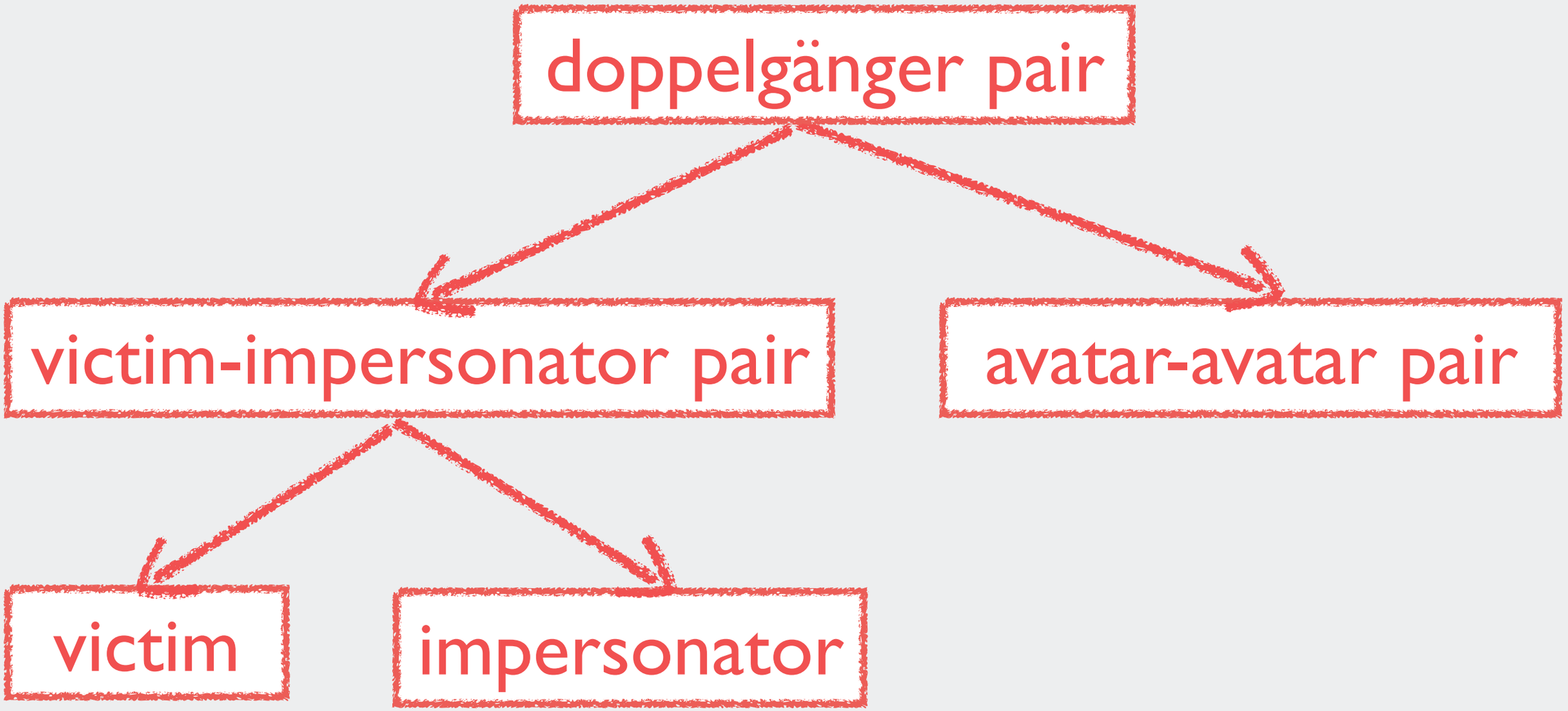
1. Methodology to gather data about impersonation attacks
2. Characterization of impersonation attacks in Twitter
3. Automatic detection of impersonation attacks

Contributions

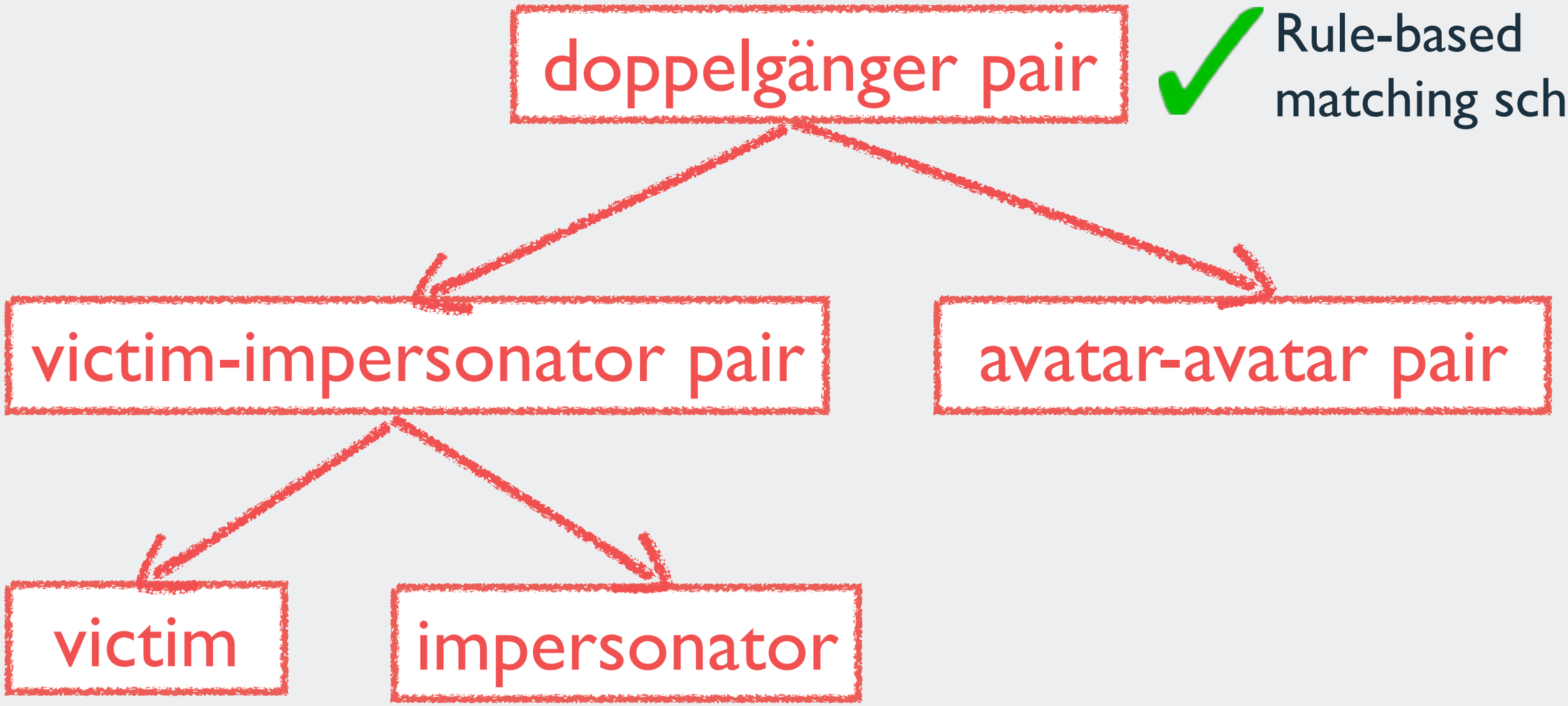
First extensive study of real-world impersonation attacks in online social networks.

1. Methodology to gather data about impersonation attacks
2. Characterization of impersonation attacks in Twitter
3. Automatic detection of impersonation attacks

Detection of impersonation attacks

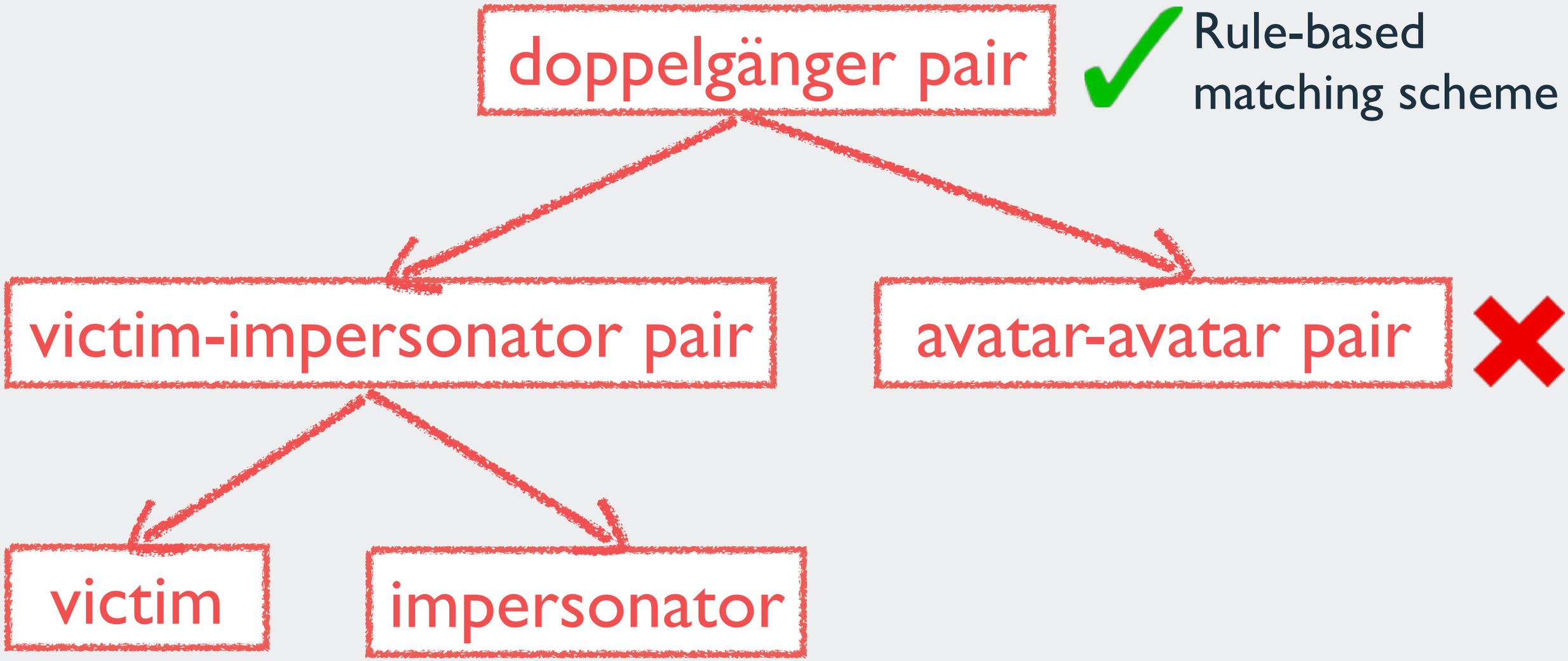


Detection of impersonation attacks

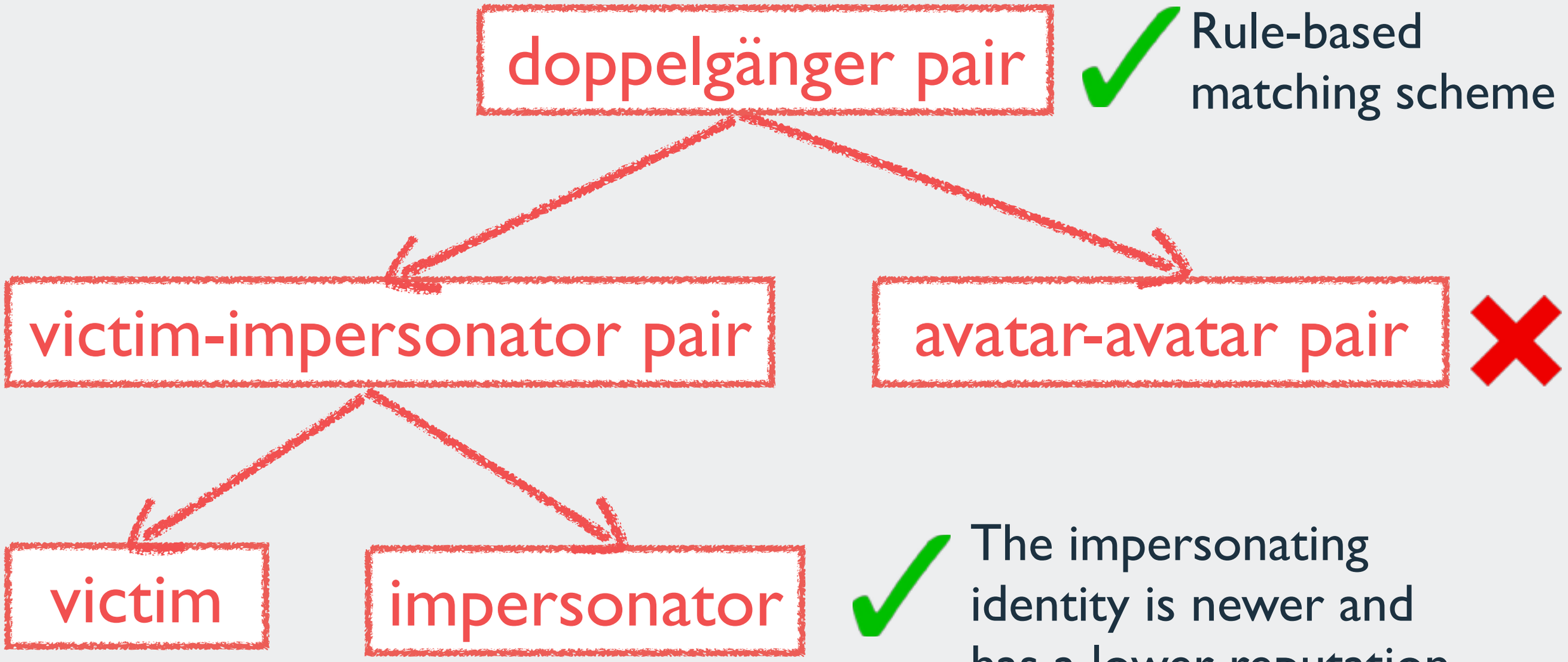


✓ Rule-based matching scheme

Detection of impersonation attacks



Detection of impersonation attacks



✓ Rule-based matching scheme

✓ The impersonating identity is newer and has a lower reputation

Detection of impersonation attacks

doppelgänger pair

✓ Rule-based matching scheme

victim-impersonator pair

avatar-avatar pair ✗

victim

impersonator

✓ The impersonating identity is newer and has a lower reputation

Automated detection of victim-impersonator pairs

SVM classifier to distinguish between victim-impersonator pairs and avatar-avatar pairs

- Training and testing:
  - labeled doppelgänger pairs from our dataset
- Features that characterize pairs of identities:
  - user-names, screen-names, location, profile photos, bios, interest *similarity*; *number of common* followers, followings, users mentioned, and retweeted; *time difference* between creation dates, first and last tweets, outdated account

Automated detection of victim-impersonator pairs

SVM classifier to distinguish between victim-impersonator pairs and avatar-avatar pairs

- Training
  - label
- Feature
  - user
    bios,
    inter
    wings,
    users mentioned, and retweeted; *time difference* between creation dates, first and last tweets, outdated account

detects

90% of victim-impersonator pairs

80% of avatar-avatar pairs

at less than 5% false positive rate

Classifying unlabeled doppelgänger pairs

	RANDOM DATASET	BFS DATASET
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605
detected victim-impersonator pairs	1,863	9,031
detected avatar-avatar pairs	4,390	4,964

Classifying unlabeled doppelgänger pairs

	RANDOM DATASET	BFS DATASET
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605
detected victim-impersonator pairs	1,863	9,031
detected avatar-avatar pairs	4,390	4,964

Classifying unlabeled doppelgänger pairs

	RANDOM DATASET	BFS DATASET
doppelgänger pairs	18,662	35,642
victim-impersonator pairs	166	16,408
avatar-avatar pairs	2,010	1,629
unlabeled pairs	16,489	17,605
detected victim-impersonator pairs	1,863	9,031
detected avatar-avatar pairs	4,390	4,964

one year later 50% were suspended!

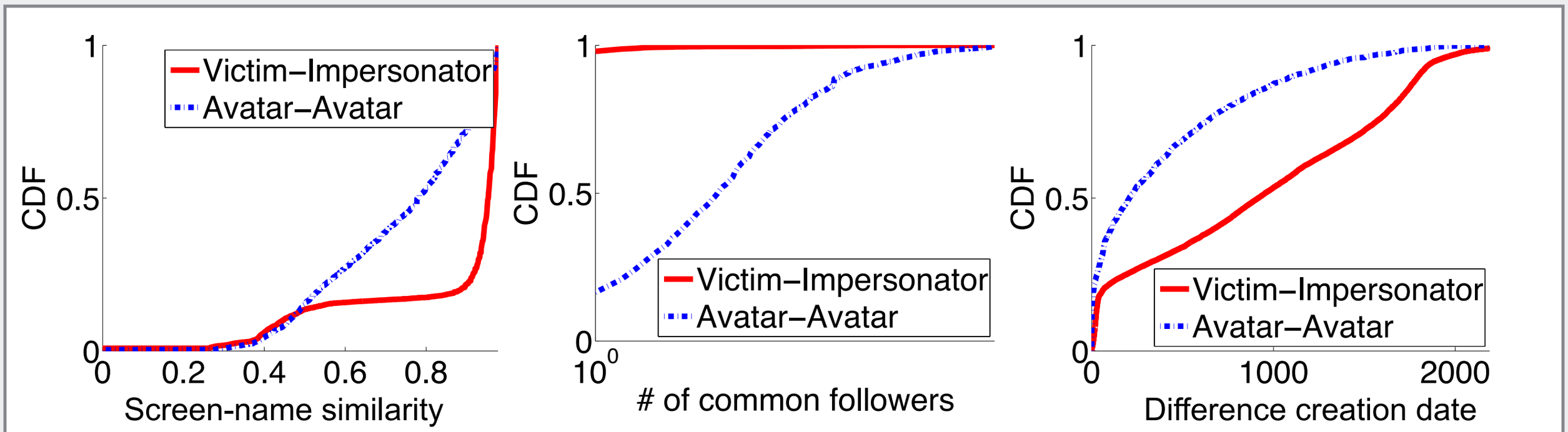
Summary

- First study to characterize and detect identity impersonation attacks online
- Method to gather real-world large-scale data about impersonation attacks
- Beside celebrity impersonators and social engineering attacks there are doppelgänger bot attacks
  - Attackers target a wide range of users, anyone can be a victim!
- Method to automatically detect impersonation attacks online

Questions?

Backup slides

Features



- Victim-impersonator pairs have more similar profile attributes
- Victim-impersonator pairs have no social neighborhood overlap
- Bigger time difference between accounts creation date in victim-impersonator pairs

Doppelgänger bot attacks: characterization

	Who are the victims?	Who are the attackers?
How popular?	73 followers	60 followers* *lower than victims, higher than random
How influential?	40% victims appear in lists	0% attacker appear in lists
How old?	October 2010	June 2013
How active?	181 tweets* *0 for random users, 20 for random users with one post	100 tweets* higher numbers of retweets, favorite and followings but not excessive