Consequences of Compromise: Characterizing Account Hijacking on Twitter

Frank Li
UC Berkeley

With: **Kurt Thomas** (UCB → Google),
Chris Grier (UCB/ICSI → Databricks), **Vern Paxson** (UCB/ICSI)

Accounts on Social Networks

- Accounts are valuable!
  - Precursor for abuse (spam, phishing, malware)
  - Twitter accounts are attractive

Accounts on Social Networks

- Accounts are valuable!
  - Precursor for abuse (spam, phishing, malware)
  - Twitter accounts are attractive
- Two ways for attackers to get accounts:
  - Fraudulent accounts
  - Compromised accounts

Prior Works

- Fraudulent accounts
  - Lots of prior work on detecting and preventing fake accounts
- Compromise accounts
  - COMPA (NDSS '13)
  - PCA-based Anomaly Detection (USENIX Security '14)

Compromise on Social Networks

- Is compromise occurring at large scales?
- What do miscreants do with compromised accounts?
- Who are being victimized?
- How do users react to compromise?
- What is causing compromise?

Detecting Compromise

- We take an external perspective of Twitter
- Looked at 8.7B tweets with URLs gathered from Jan – Oct 2013
  - 168M users in data set

Spam Tweets


Stephen M
@Dance_guy
Follow

Aweesomeee! I made $171.50 this week so far taking a couple of surveys. http://t.co/cwG67lh4

10:20 AM - 19 Nov 13

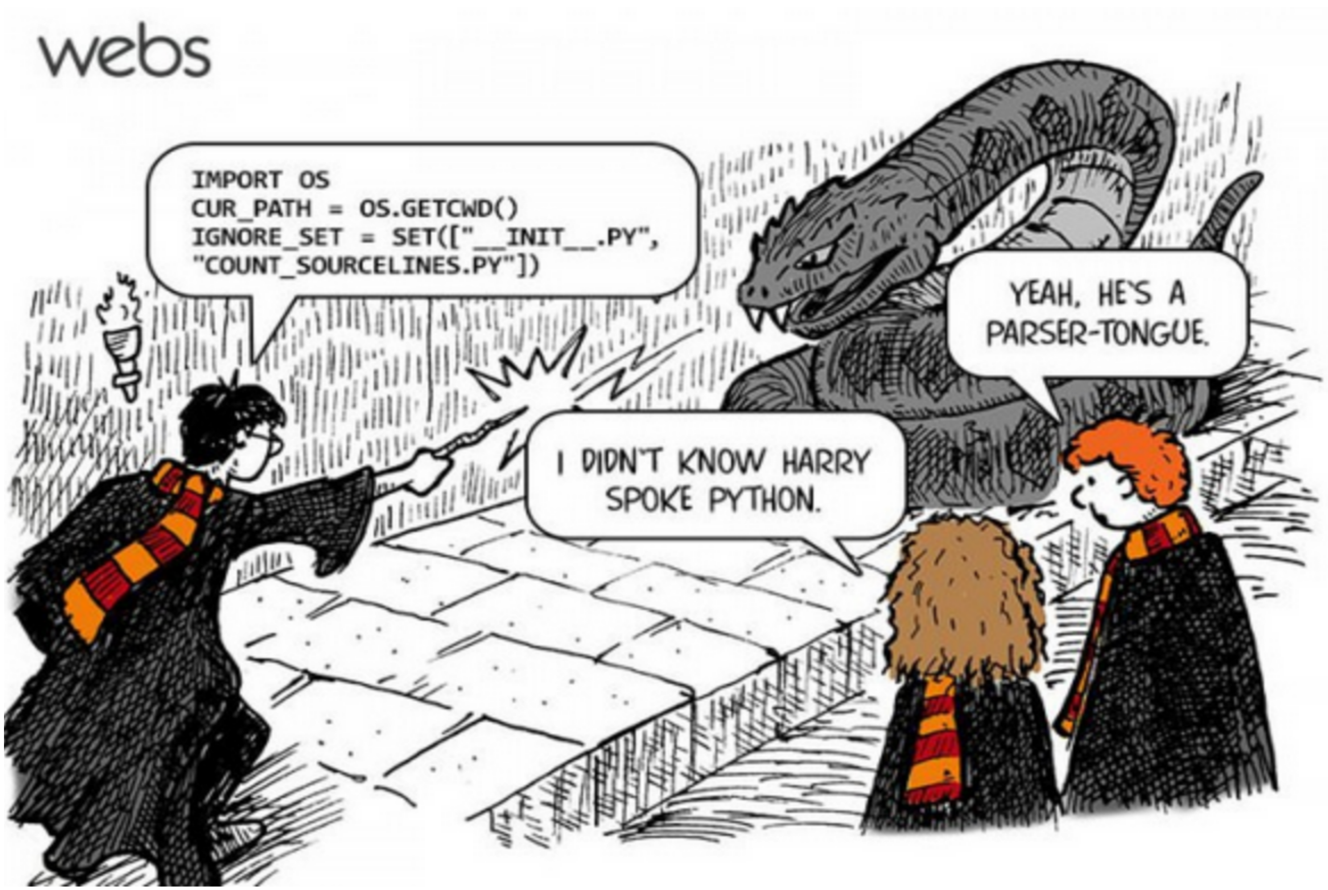

Nicole C.
@CheapCialisNow
Follow

Awesome! I made $106.03 this week so far just filling out a couple of surveys. http://t.co/PoHBayLz

Reply Retweet Favorite More

6:44 AM - 5 Dec 12

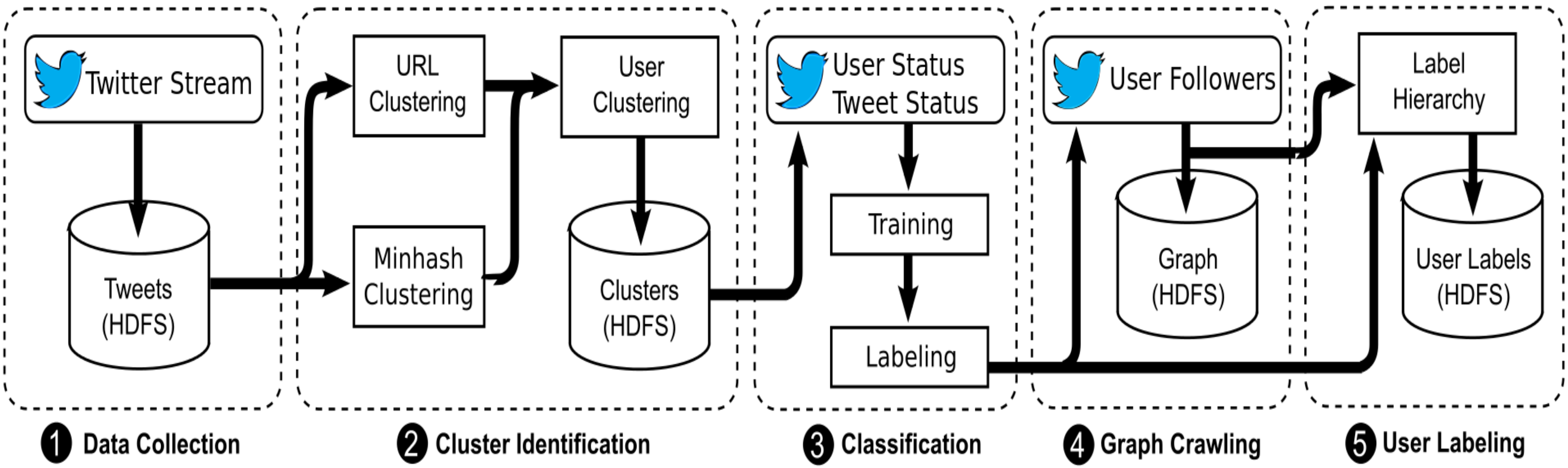
Meme Tweets



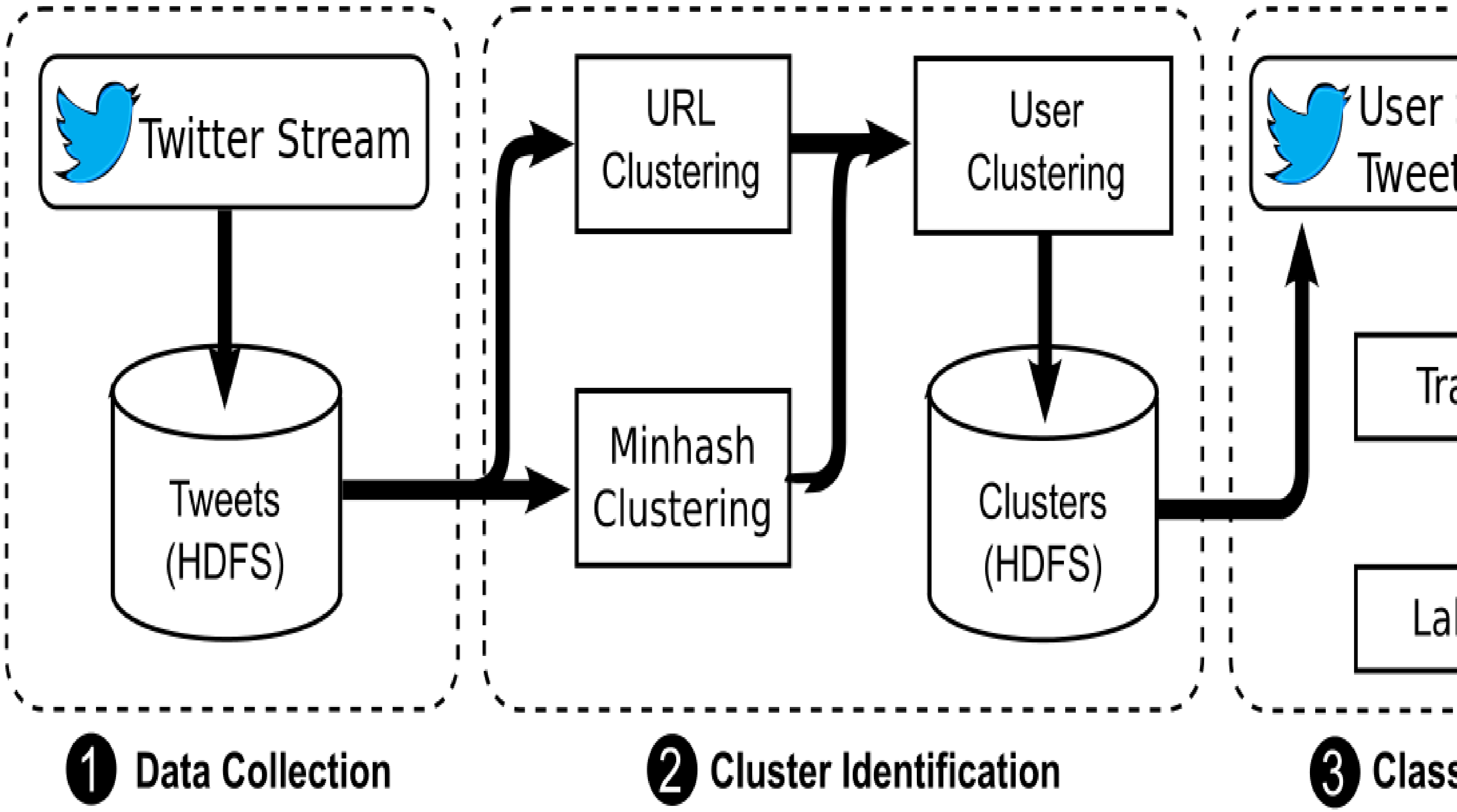
SEMrush @semrush · Oct 4
Hahaha! I didn't know Harry spoke Python :D #fun #itjokes #python

53 24

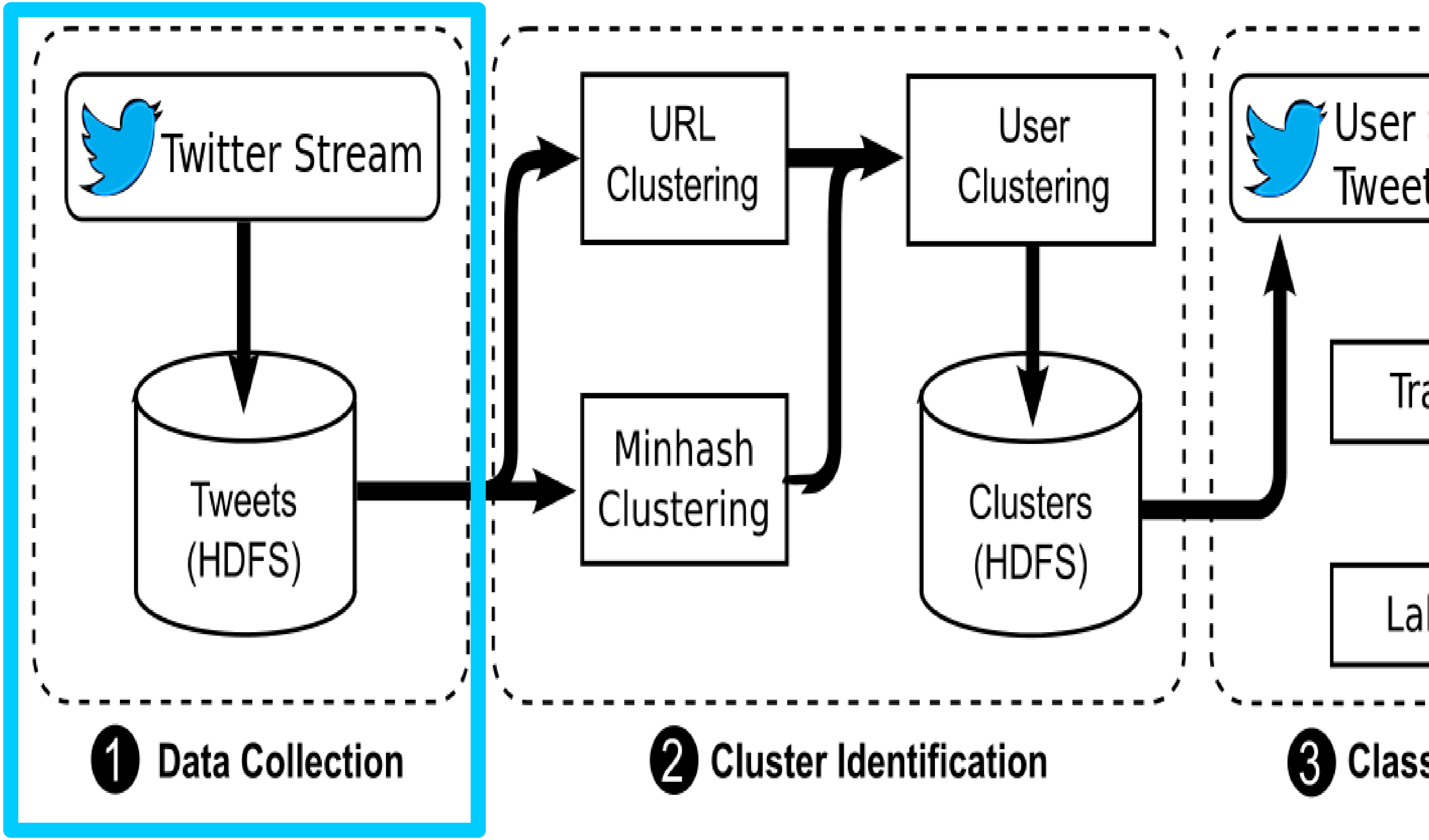
Analysis Pipeline



Identifying Compromised Users



Identifying Compromised Users



Twitter Stream Data

{"created_at":"Fri Oct 10 00:00:24 +0000 2014","id":520363179210072065,"id_str":"520363179210072065","text":"White people http:\/\/t.co\/gcOd6JqKKL","source":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter for iPhone\u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":
{"id":1912894320,"id_str":"1912894320","name":"Suck My Ass ","screen_name":"Janoskbiebs","location":"","url":null,"description":null,"protected":false,"verified":false,"followers_count":1136,"friends_count":1294,"listed_count":2,"favourites_count":2090,"statuses_count":5113,"created_at":"Sat Sep 28 03:00:09 +0000 2013","utc_offset":-
25200,"time_zone":"Arizona","geo_enabled":true,"lang":"en","contributors_enabled":false,"is_translator":false,"profile_background_color":"C0DEED","profile_background_image_url":"http:\/\/pbs.twimg.com\/profile_background_images\/479624239423553538\/9xMeKSoG.jpeg","profile_background_image_url_https":"https:\/\/pbs.twimg.com\/profile_background_images\/479624239423553538\/9xMeKSoG.jpeg","profile_background_tile":true,"profile_link_color":"31BF9C","profile
_sidebar_border_color":"000000","profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,"profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/520080781612290048\/A5pKzHGV_normal.jpeg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/520080781612290048\/A5pKzHGV_normal.jpeg","profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/1912894320\/1412831879","default_profile":false
,"default_profile_image":false,"following":null,"follow_request_sent":null,"notifications":null},"geo":null,"coordinates":null,"place":null,"contributors":null,"retweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"trends":[],"urls":[],"user_mentions":[],"symbols":[],"media":[{"id":489091648354549760,"id_str":"489091648354549760","indices":
[14,36],"media_url":"http:\/\/pbs.twimg.com\/media\/BsmaQ0rIgAA4fMN.jpg","media_url_https":"https:\/\/pbs.twimg.com\/media\/BsmaQ0rIgAA4fMN.jpg","url":"http:\/\/t.co\/gcOd6JqKKL","display_url":"pic.twitter.com\/gcOd6JqKKL","expanded_url":"http:\/\/twitter.com\/SteveMeans\/status\/489091648522301440\/photo\/1","type":"photo","sizes":{"medium":{"w":600,"h":552,"resize":"fit"},"small":{"w":340,"h":312,"resize":"fit"},"large":{"w":600,"h":552,"resize":"fit"},"thumb":
{"w":150,"h":150,"resize":"crop"}},"source_status_id":489091648522301440,"source_status_id_str":"489091648522301440"}]},"extended_entities":{"media":[{"id":489091648354549760,"id_str":"489091648354549760","indices":
[14,36],"media_url":"http:\/\/pbs.twimg.com\/media\/BsmaQ0rIgAA4fMN.jpg","media_url_https":"https:\/\/pbs.twimg.com\/media\/BsmaQ0rIgAA4fMN.jpg","url":"http:\/\/t.co\/gcOd6JqKKL","display_url":"pic.twitter.com\/gcOd6JqKKL","expanded_url":"http:\/\/twitter.com\/SteveMeans\/status\/489091648522301440\/photo\/1","type":"photo","sizes":{"medium":{"w":600,"h":552,"resize":"fit"},"small":{"w":340,"h":312,"resize":"fit"},"large":{"w":600,"h":552,"resize":"fit"},"thumb":
{"w":150,"h":150,"resize":"crop"}},"source_status_id":489091648522301440,"source_status_id_str":"489091648522301440"}]},"favorited":false,"retweeted":false,"possibly_sensitive":false,"filter_level":"medium","lang":"en","timestamp_ms":"1412899224466"}

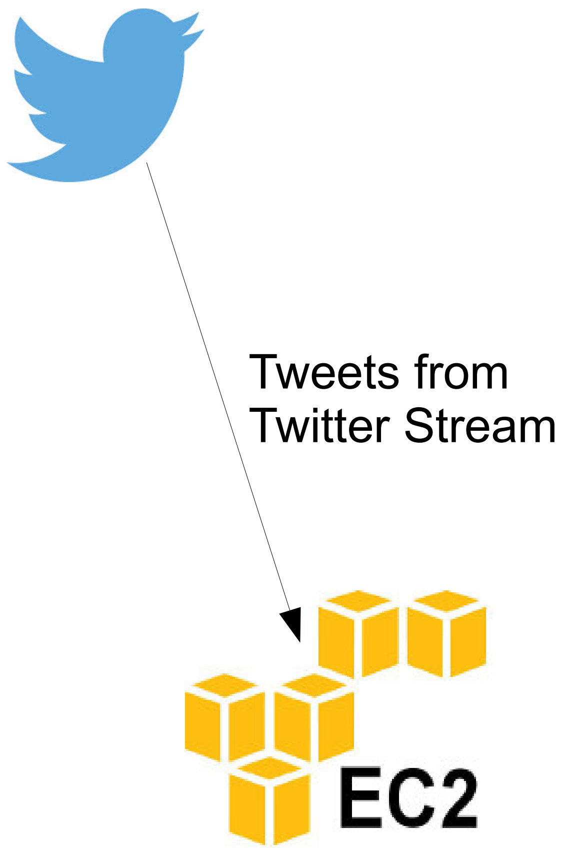
Twitter Stream Data

- **created_at** (UTC, seconds)
- **id** (>53 bits)
- **text** (UTF-8, <140 char)
- **source**
- **lang** (machine-detected, BPC-47)
- **in_reply_to_status_id**
- **in_reply_to_user_id**
- **in_reply_to_screen_name**
- **entities**
  - **hashtags**
  - **urls** (both URL and domain)
  - **user_mentions**

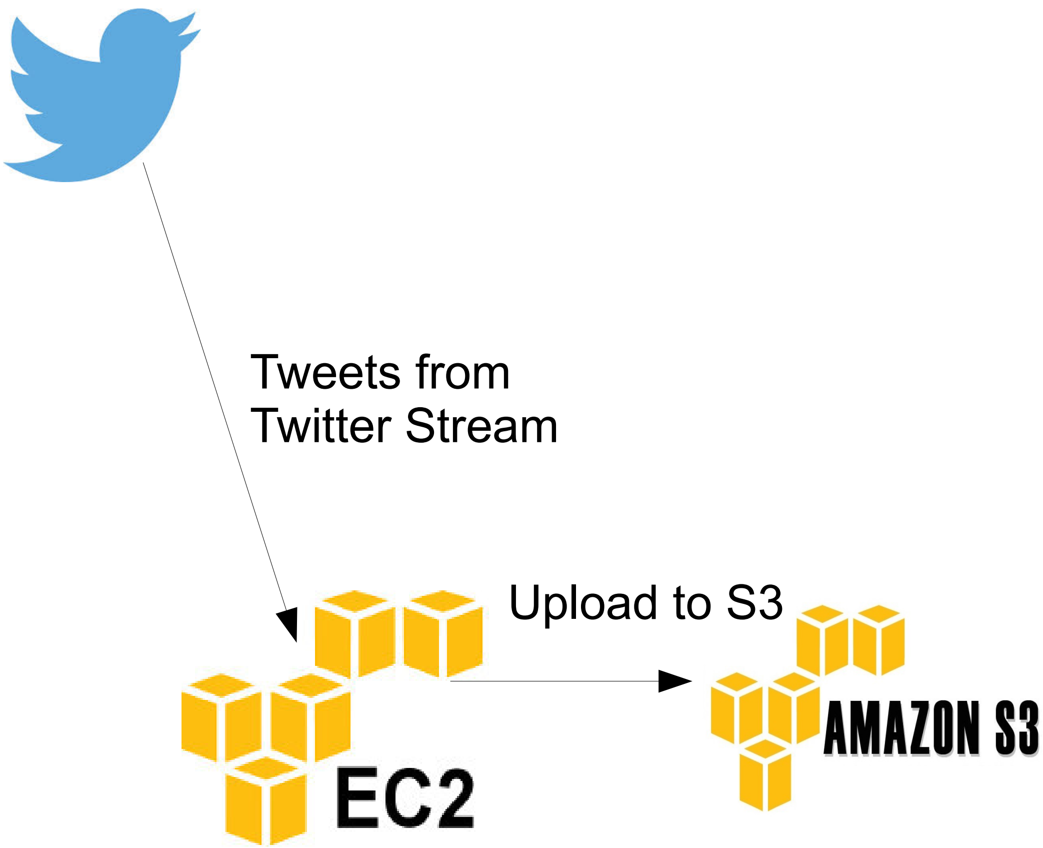
- **user**
  - **id** (>53 bits)
  - **name** (<=20 char)
  - **screen_name** (<=15 char)
  - **description** (<=160 char)
  - **protected**
  - **verified**
  - **followers_count**
  - **friends_count**
  - **statuses_count**
  - **created_at** (UTC, seconds)
  - **lang** (user self-declared, BPC-47)

Infrastructure

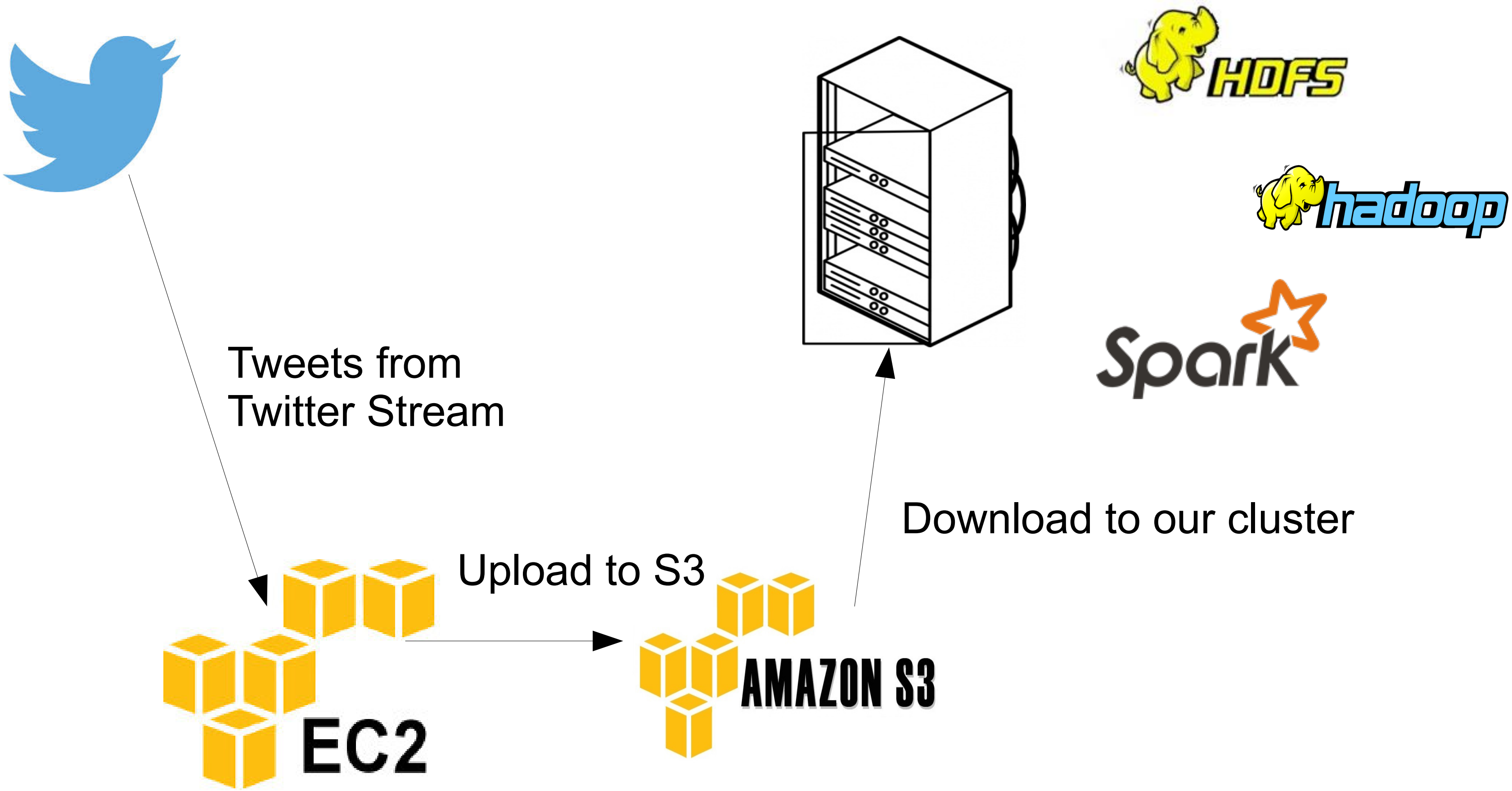
Infrastructure



Infrastructure



Infrastructure

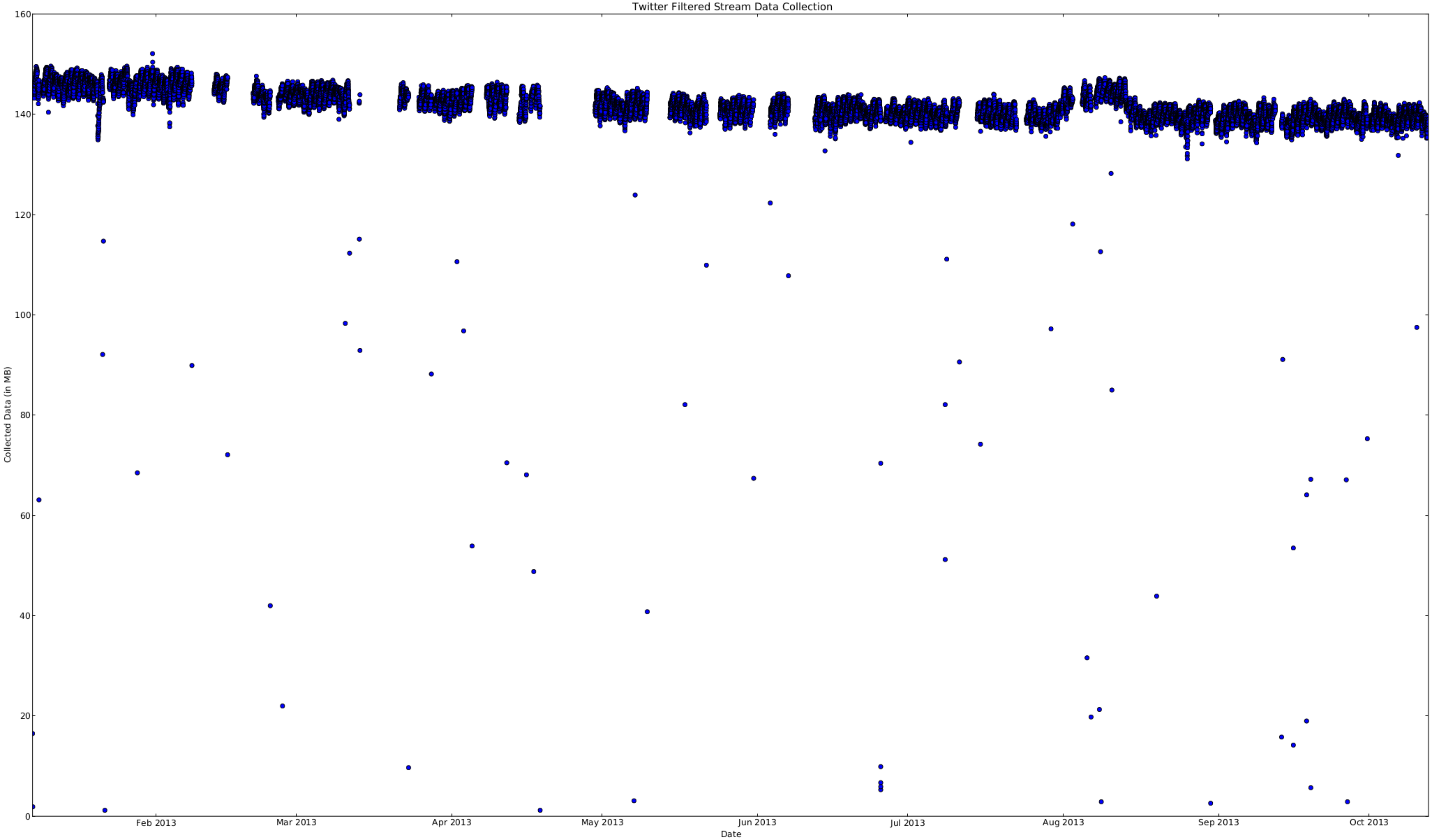


Filtered Stream

- Access to a filtered stream of URLs
- ~200 GB of data per day,

  compressed to ~20 GB per day
- In total, 4.1 TB of compressed data for 2013.

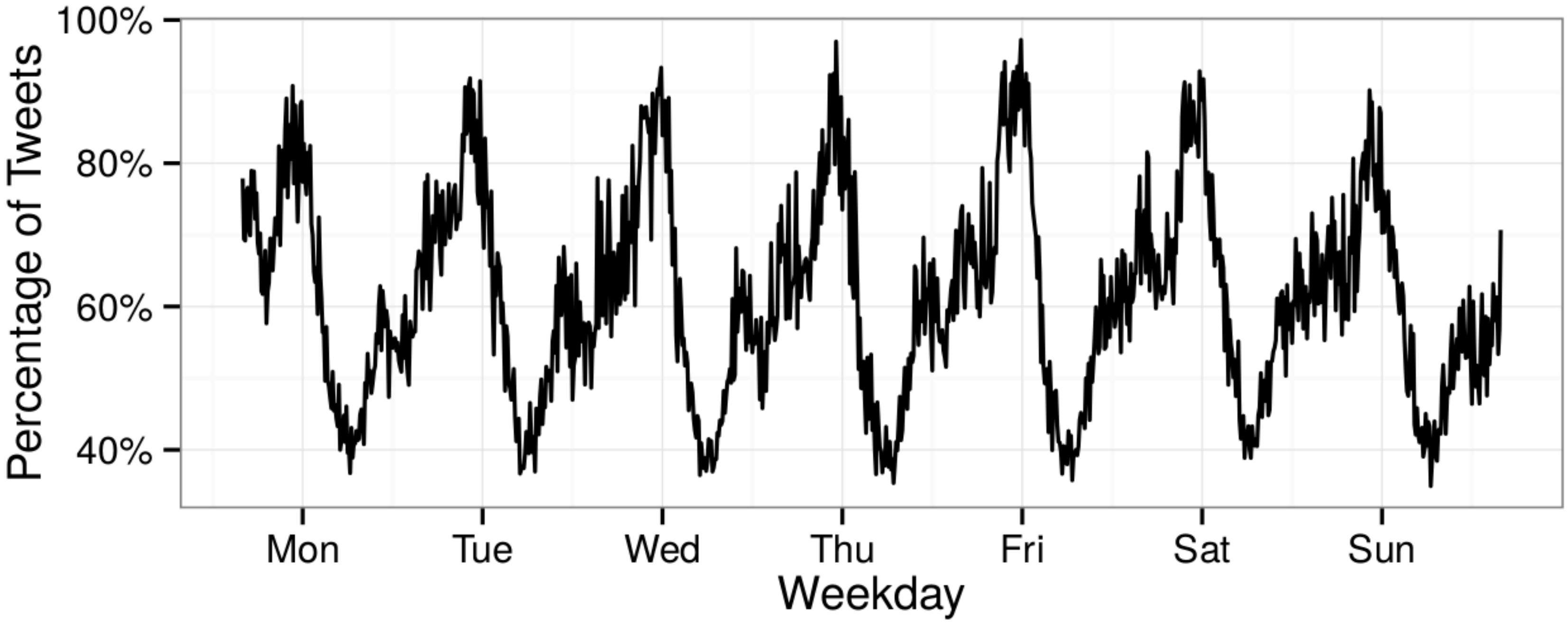
Data Collection



Infrastructure Issues

- Twitter feed outage
- EC2 reboot
- EC2 feed application crash
- Low disk space
- Disk failures
- Updates break things

Filtered Stream



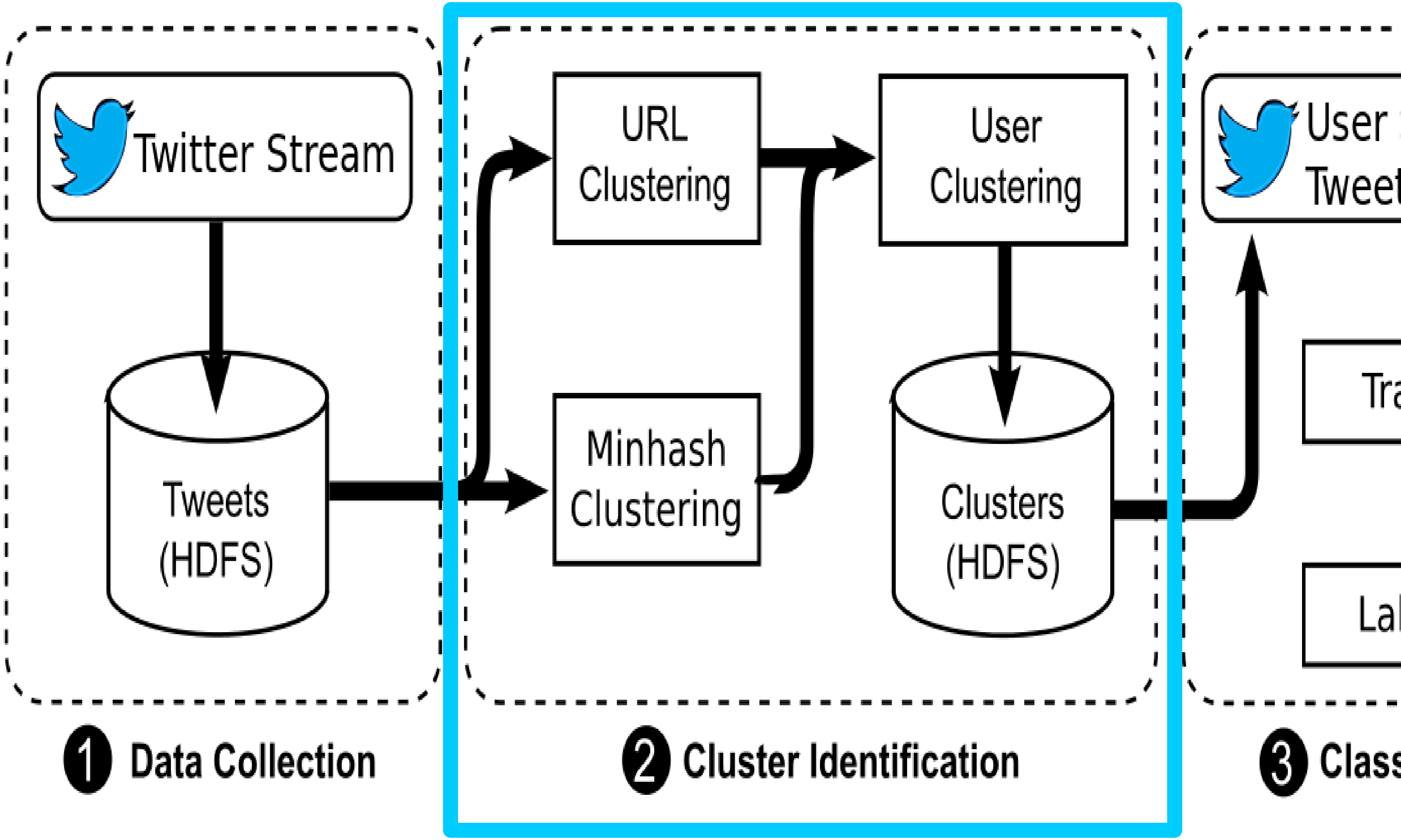
Roughly 61% of all Tweets with URLs

Sampling Error

- Under-estimate size of clusters

- Any graph analysis will under-represent social connectivity

Identifying Compromised Users



Similar Content Example

Near duplicate text
Different URL

Clustering Tweets

- Cluster on same URLs
- Cluster on similar content
  - Split text into n-grams
  - Want Jaccard similarity coefficient:

$$J(M_i, M_j) = \frac{|M_i \cap M_j|}{|M_i \cup M_j|}$$

  - To avoid O(n^2), where n = O(billion), use minhash estimation

Minhash Estimation

- Set A = {a1,...., aN}   Set B = {b1,..., bN}

Minhash Estimation

- Set A = {a1,...., aN}   Set B = {b1,..., bN}
- Hash all elements:

  A' = {h(a1),...,h(aN)}   B' = {h(b1),...,h(bN)}

Minhash Estimation

- Set A = {a1,...., aN}   Set B = {b1,..., bN}
- Hash all elements:

  A' = {h(a1),...,h(aN)}   B' = {h(b1),...,h(bN)}
- Sort hashes for each set:

  A'' = {h(a3), h(a7),...}   B'' = {h(b9}, h(b2),...}

Minhash Estimation

- Set A = {a1,...., aN}   Set B = {b1,..., bN}
- Hash all elements:

  A' = {h(a1),...,h(aN)}   B' = {h(b1),...,h(bN)}
- Sort hashes for each set:

  A'' = {h(a3), h(a7),...}   B'' = {h(b9}, h(b2),...}
- Key for each set is the k smallest hashes:

  Key_A = h(a3)||h(a7)   Key_B = h(b9)||h(b2)

Minhash Estimation

- Set A = {a1,...., aN}   Set B = {b1,..., bN}
- Hash all elements:

  A' = {h(a1),...,h(aN)}   B' = {h(b1),...,h(bN)}

- Sort hashes for each set:

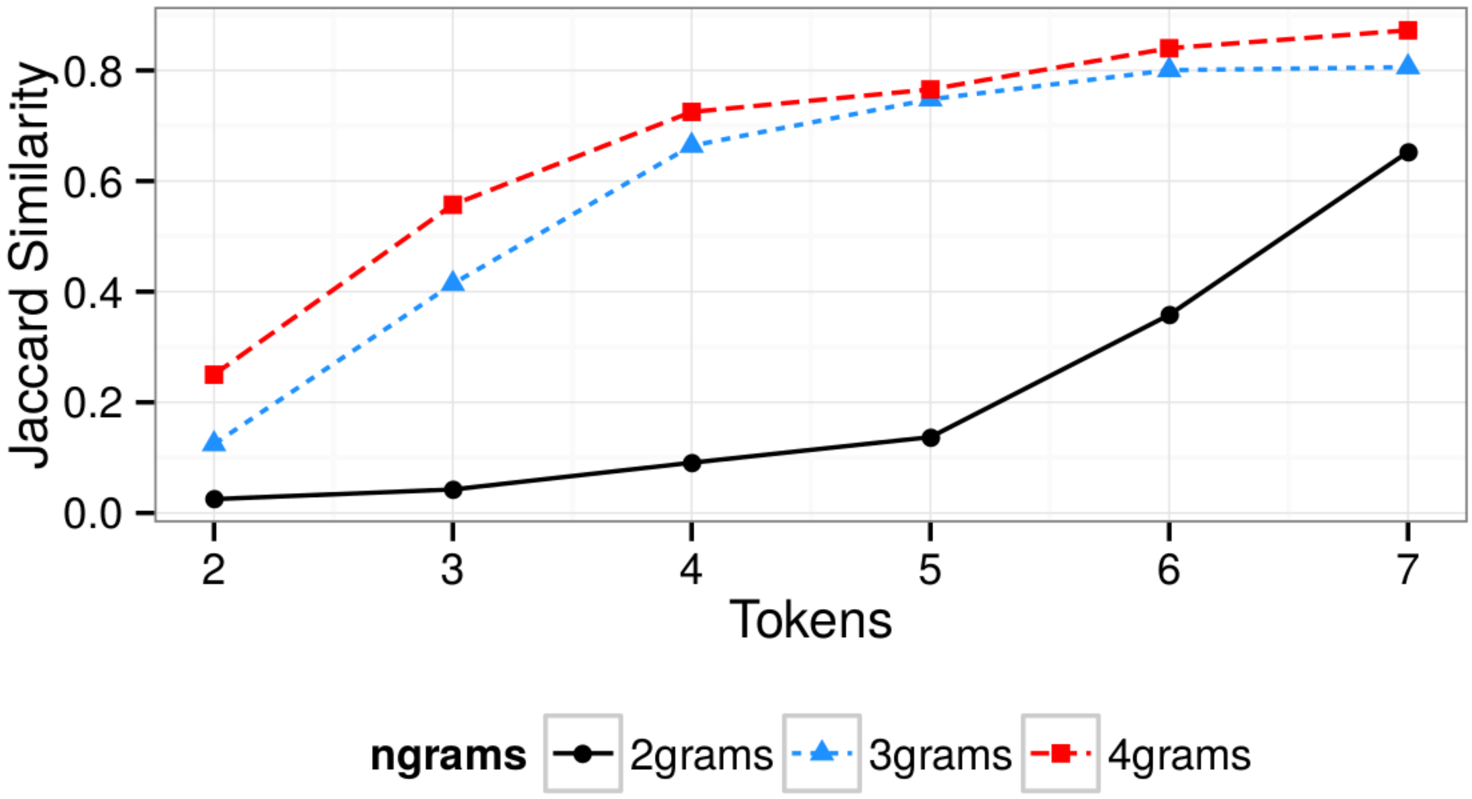
  A'' = {h(a3), h(a7),...}   B'' = {h(b9}, h(b2),...}

- Key for each set is the k smallest hashes:

  Key_A = h(a3)||h(a7)   Key_B = h(b9)||h(b2)

- The probability keys are equal for two sets is proportional to their Jaccard similiarity.

Minhash Parameters



Grid search on sample of 19 M tweets

Classifying a Group of Tweets

Classifying a Group of Tweets

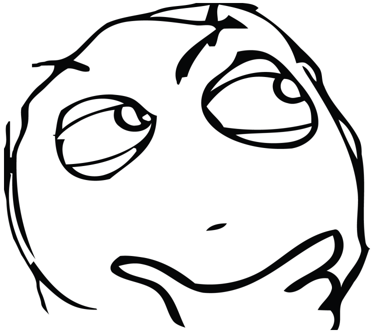
- Observation 1: Users delete tweets from compromise.

Classifying a Group of Tweets

- Observation 1: Users delete tweets from compromise.
- Observation 2: Twitter suspends fraudulent accounts.

Classifying a Group of Tweets

- Observation 1: Users delete tweets from compromise.
- Observation 2: Twitter suspends fraudulent accounts.

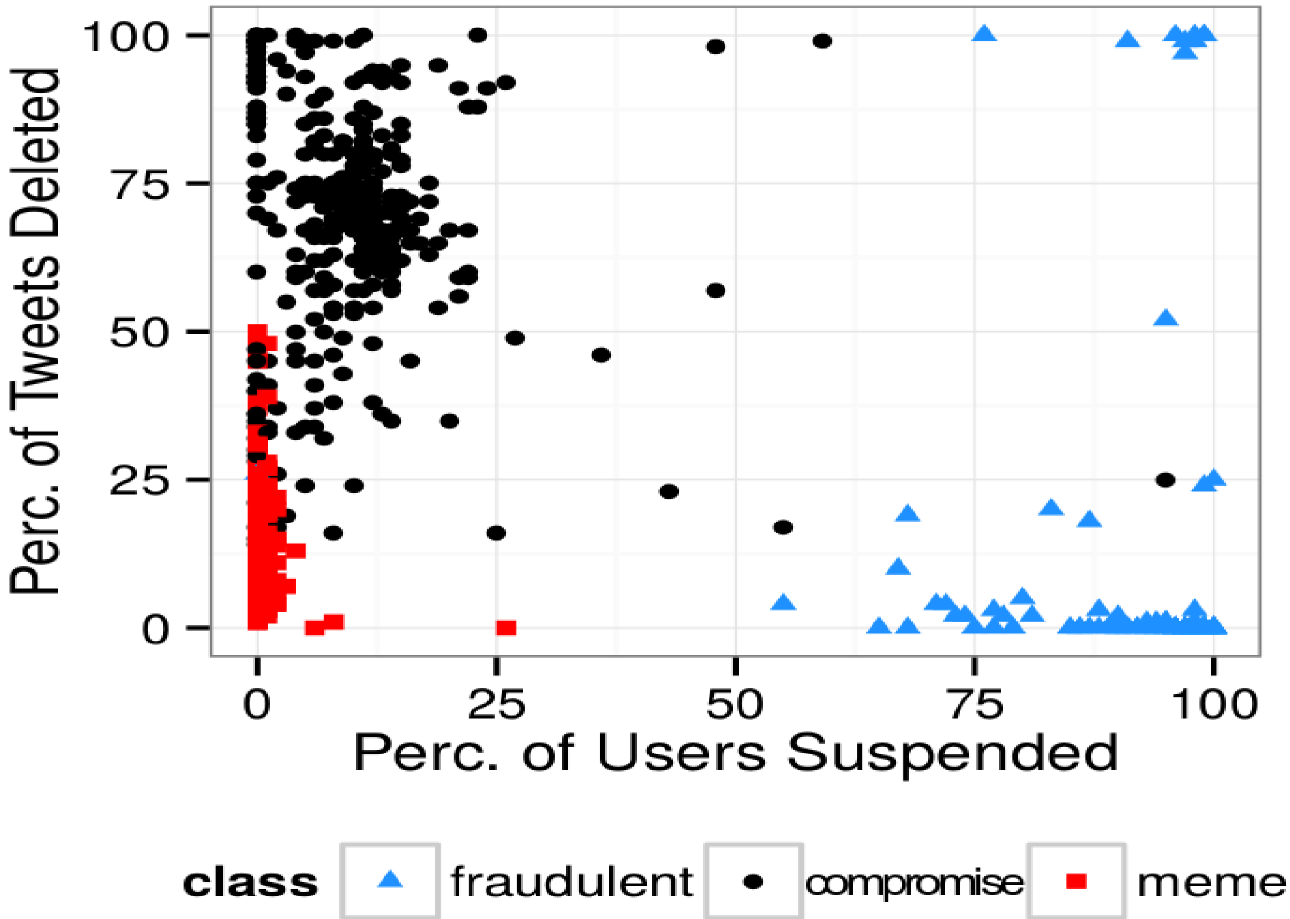


Deletions and Suspensions as Features

- Manually labeled 1700 random clusters

Deletions and Suspensions as Features

- Manually labeled 1700 random clusters



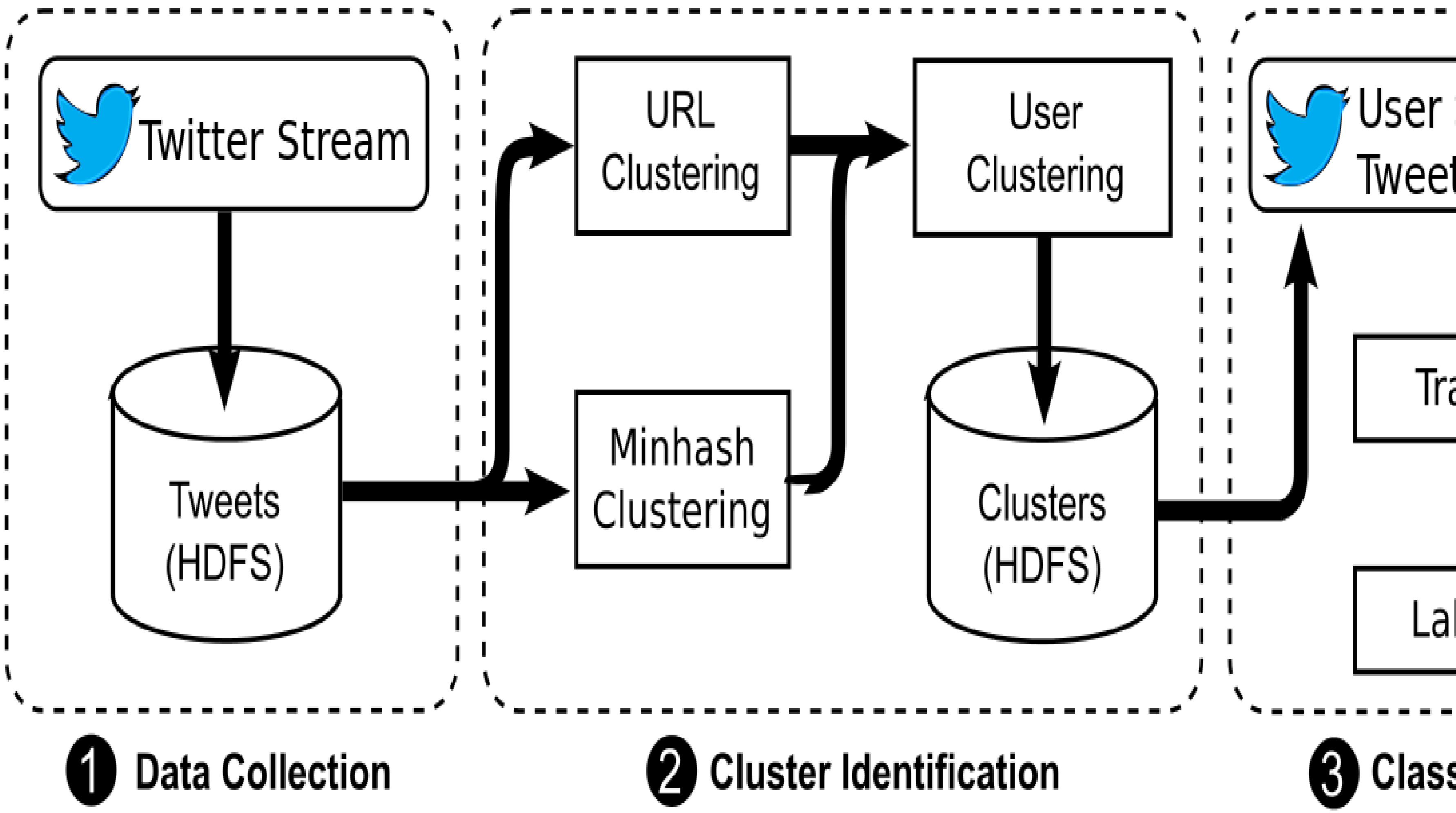
Other Features

- Fraction of tweets in a cluster that were retweets
- Average # of tweets per user in the cluster
- # of distinct tweet sources per cluster
- # of distinct languages per cluster

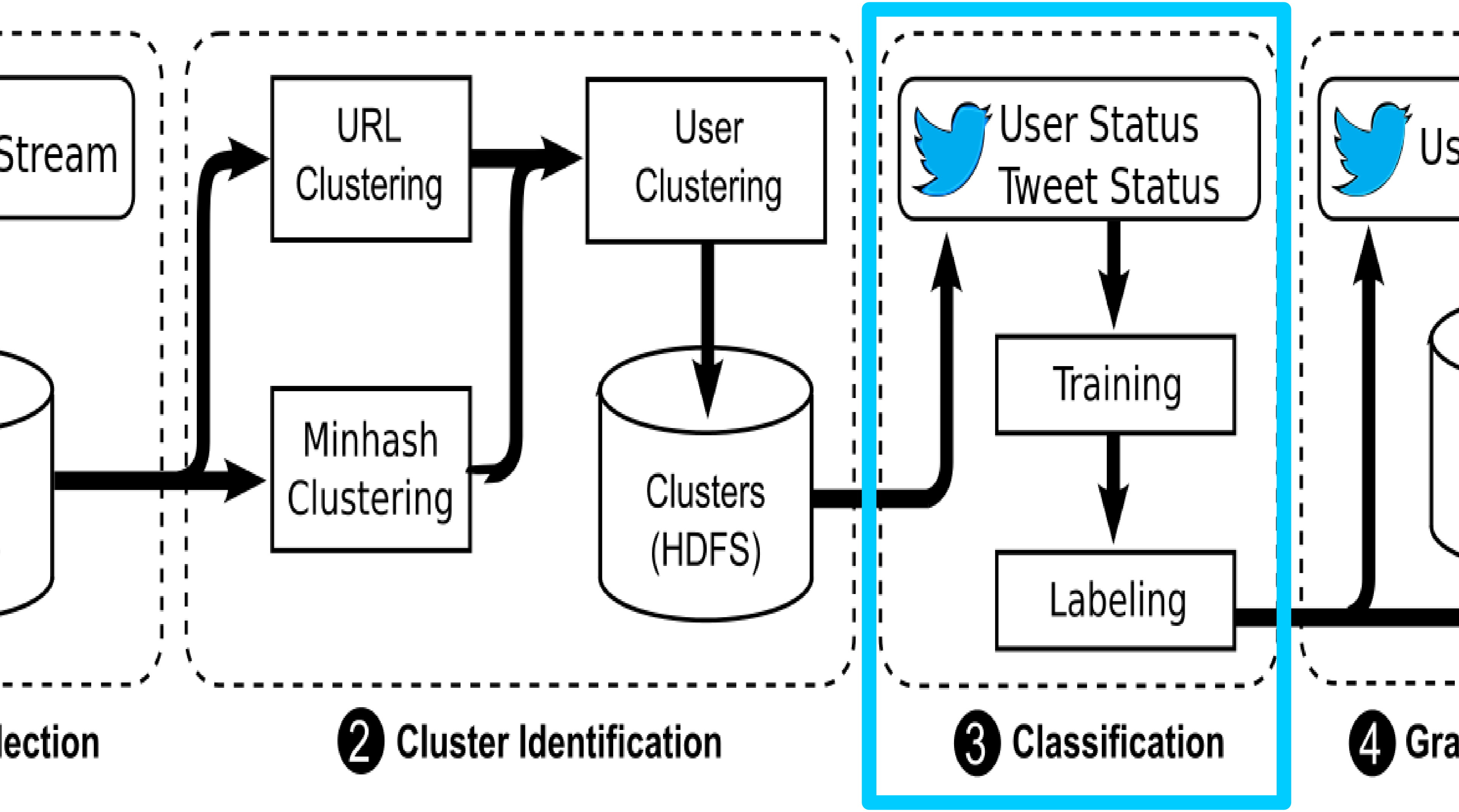
Classification

- Multi-class logistic regression
- 10-fold cross-validation: 99.4% accuracy
- Most important features:
  - Ratio of suspended users, ratio of deleted tweets, number of distinct languages

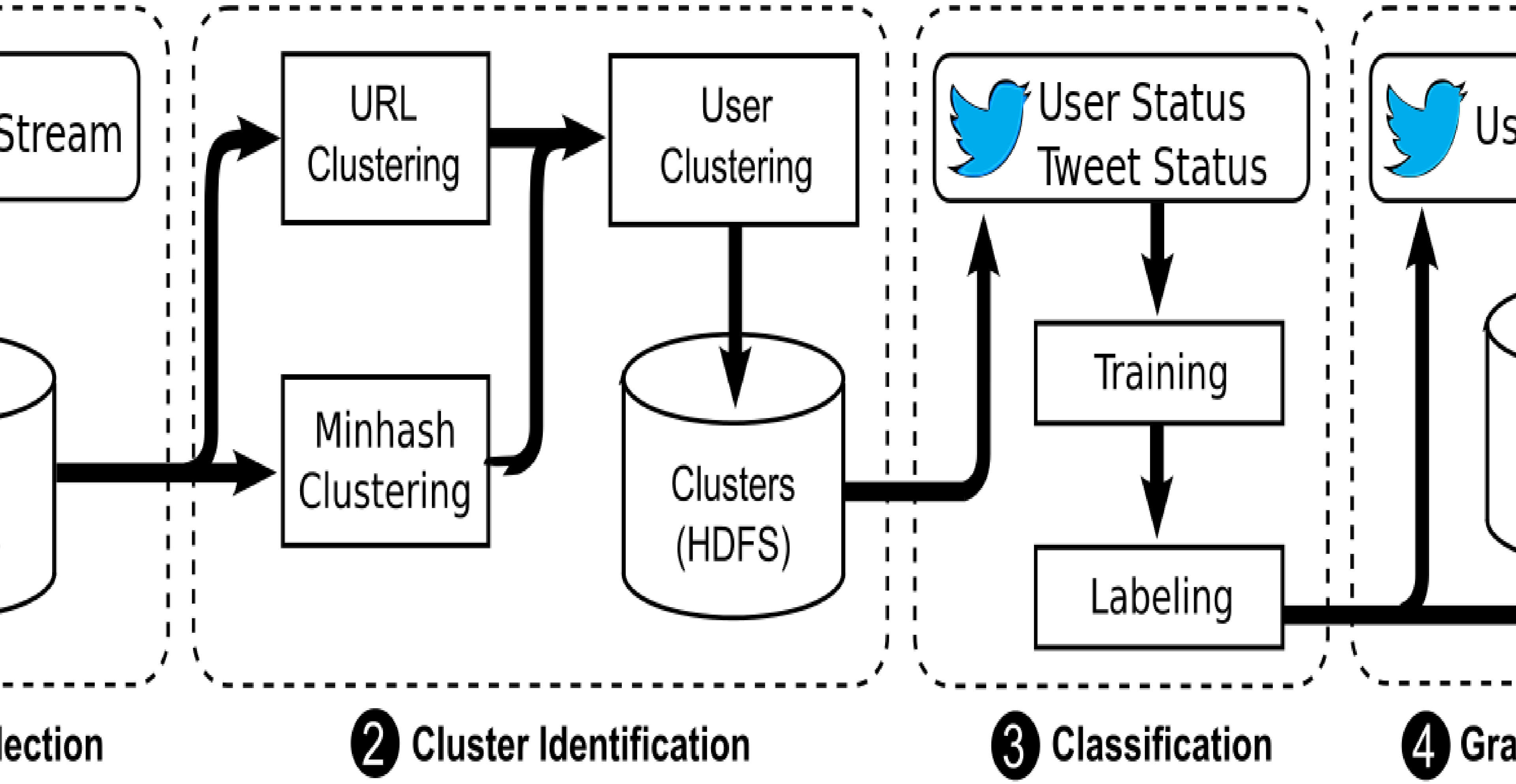
Identifying Compromised Users



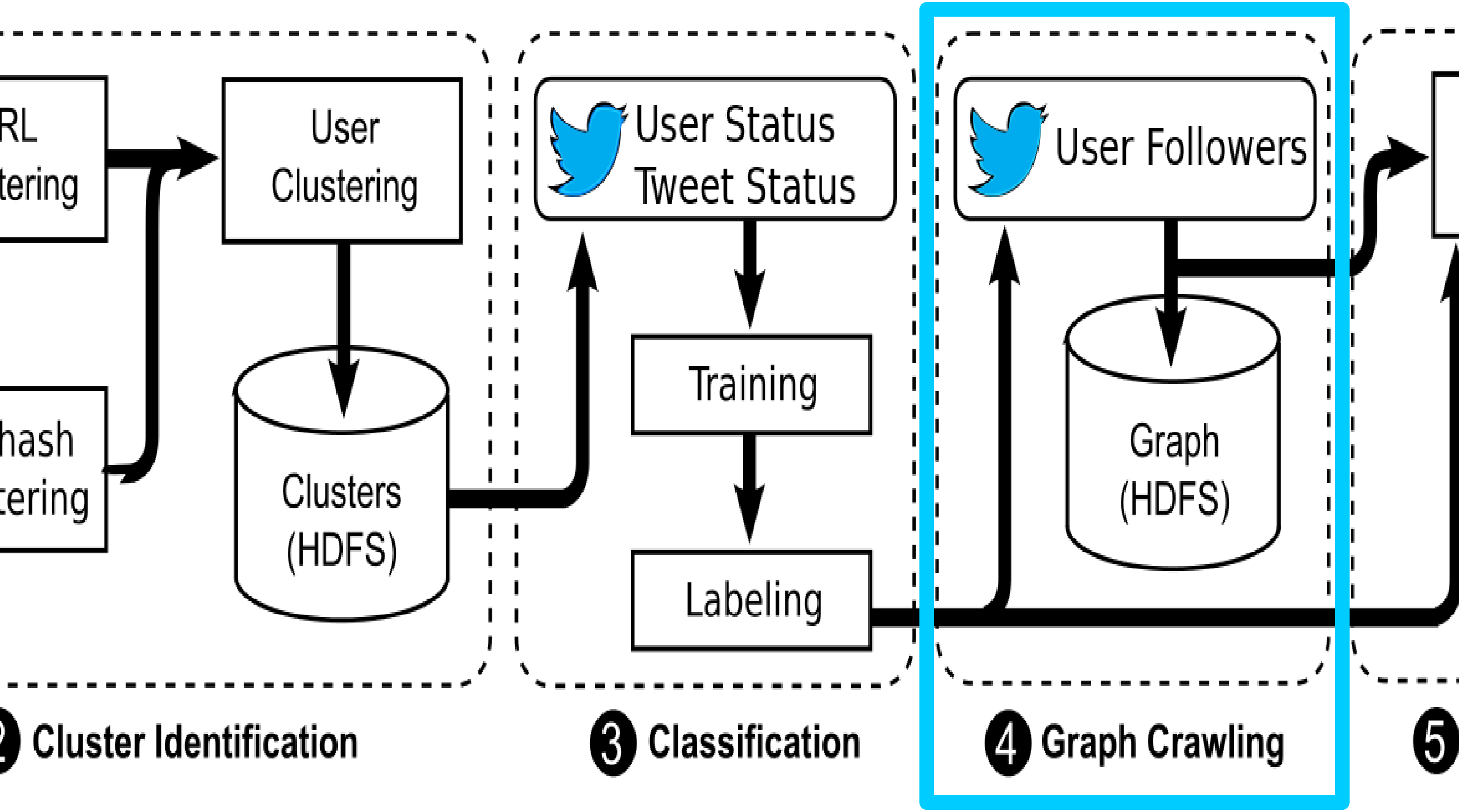
Identifying Compromised Users



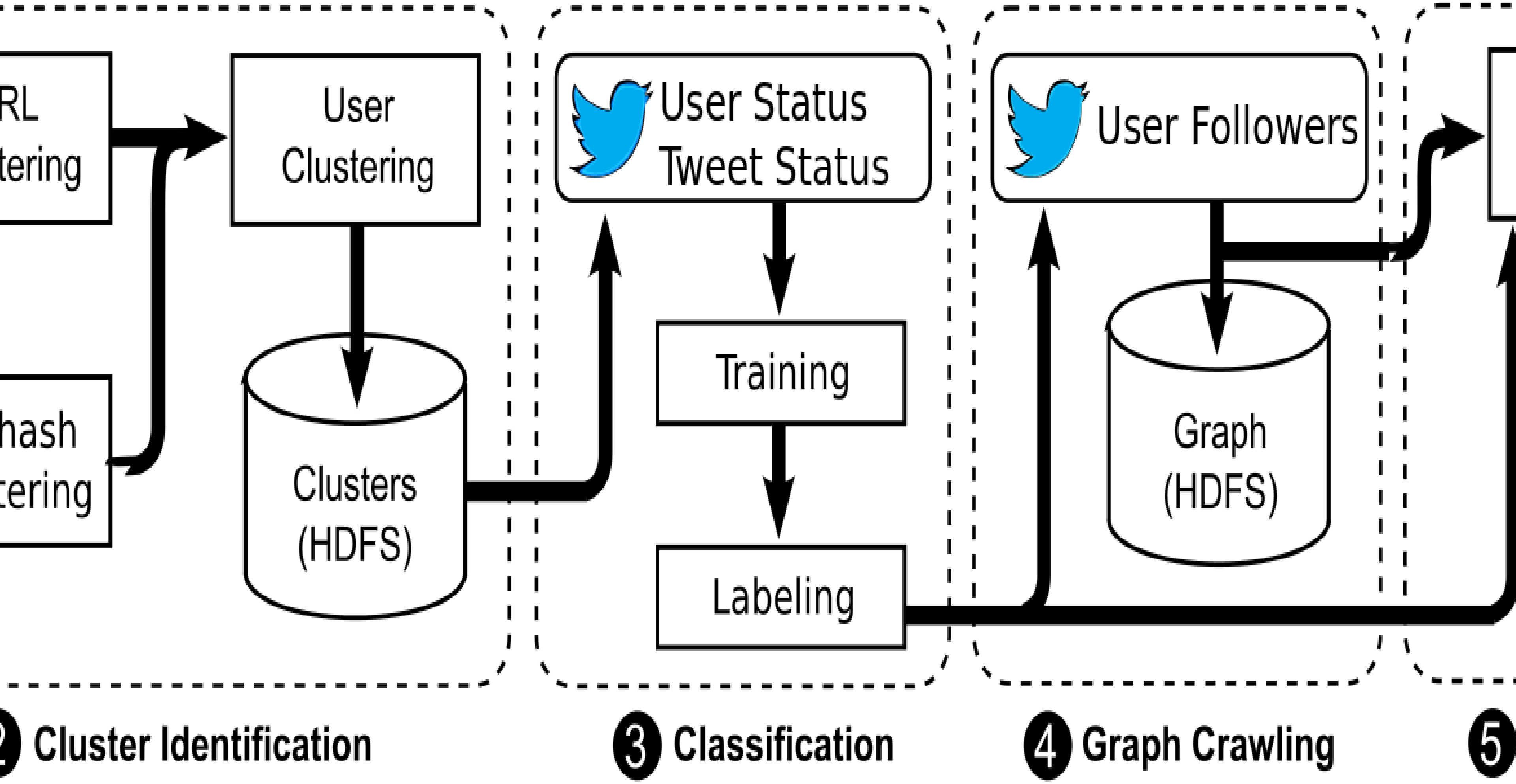
Analyzing Compromised Users



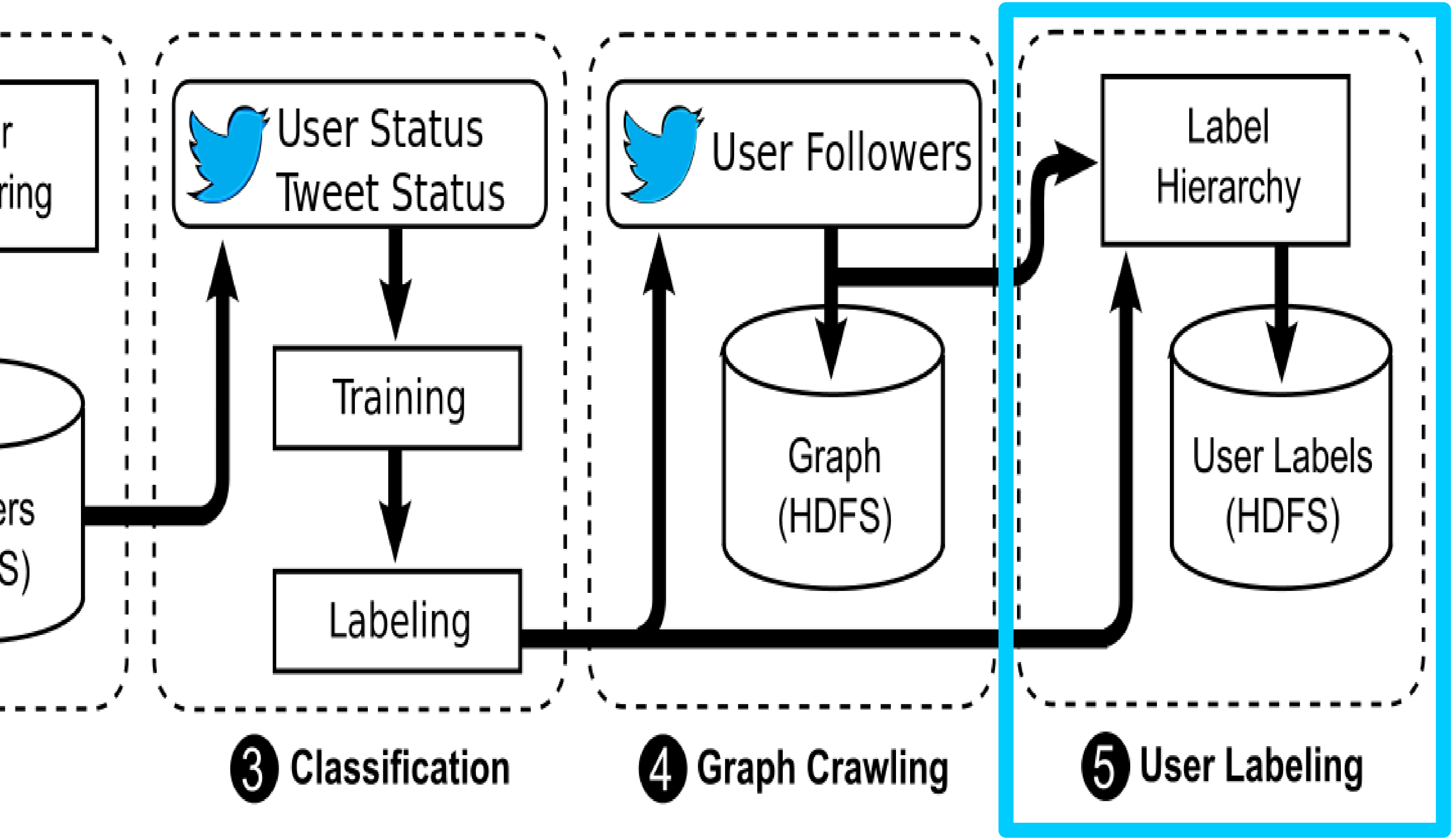
Analyzing Compromised Users



Analyzing Compromised Users



Analyzing Compromised Users



Scale of Compromise

Scale of Compromise

Measurement	Value
Meme clusters	10,792
Compromise clusters	**2,661**
Fraudulent account clusters	2,753
Meme participants	17.3 million
Compromised victims	**13.9 million**
Fraudulent accounts	4.7 million
Meme tweets	130 million
Spam tweets via compromised accounts	**81 million**
Spam tweets via fraudulent accounts	44 million

Monetizing Compromised Accounts

Monetizing Compromised Accounts

- Largest single campaign advertised Garcinia
  - 1.1M accounts
  - 70k distinct URLs
  - Lasted 23 days
- Nutraceutical campaigns were largest source
  - 4.7M accounts total (34% of all we detect)

sid bishop @sustainablesid · 7h

Dr. Oz **Garcinia** Cambogia Where To Buy Natural And Organic Food That Burns ... - Amersham People tinyurl.com/lq9wa5l

Expand

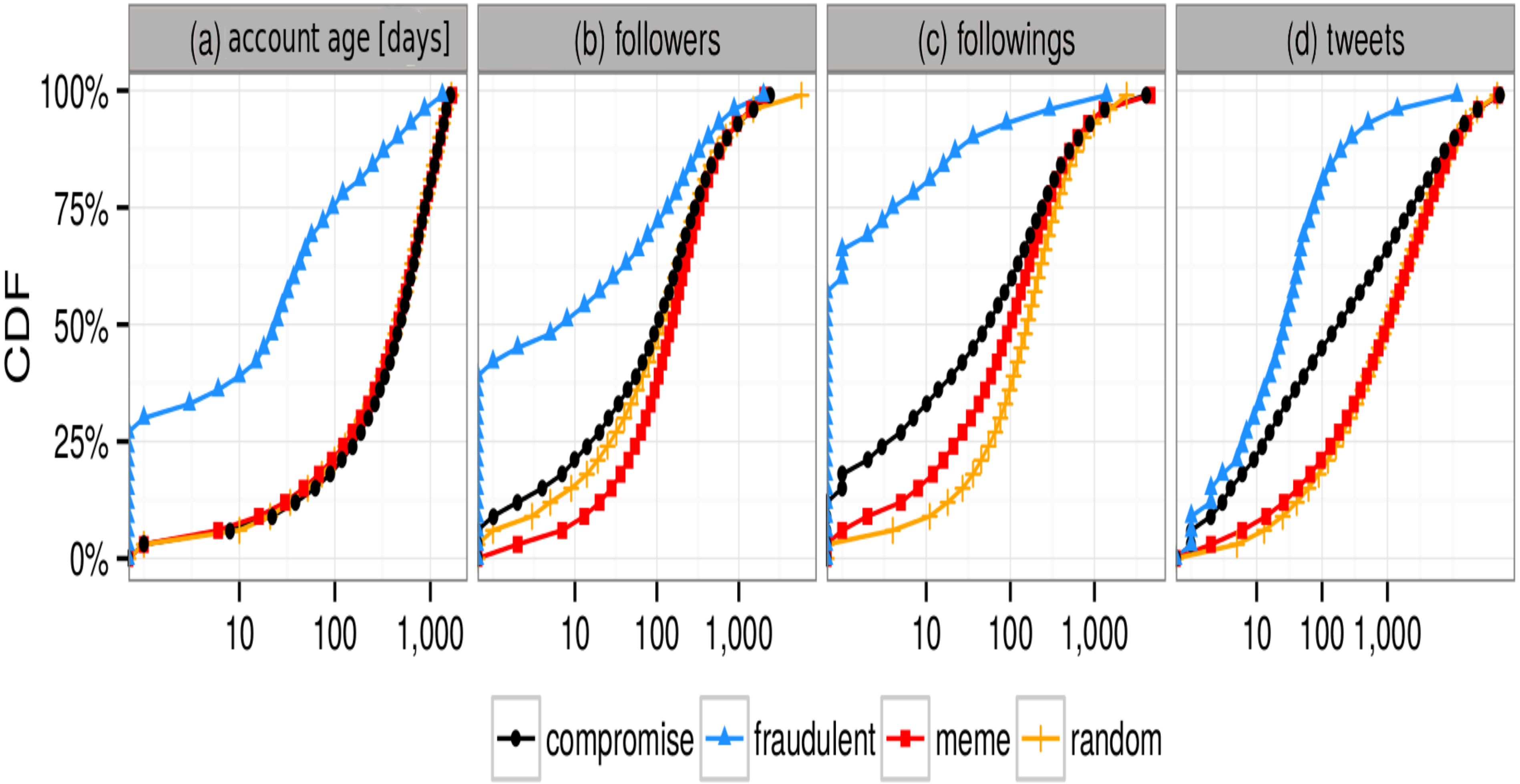
Reply Retweet Favorite More

Other Leading Monetization Vectors

- Gain followers and retweets
  - 3.7M users
  - 779 distinct clusters advertising free followers
- Generating Leads
  - 1M users, 1 cluster, lasting 31 days

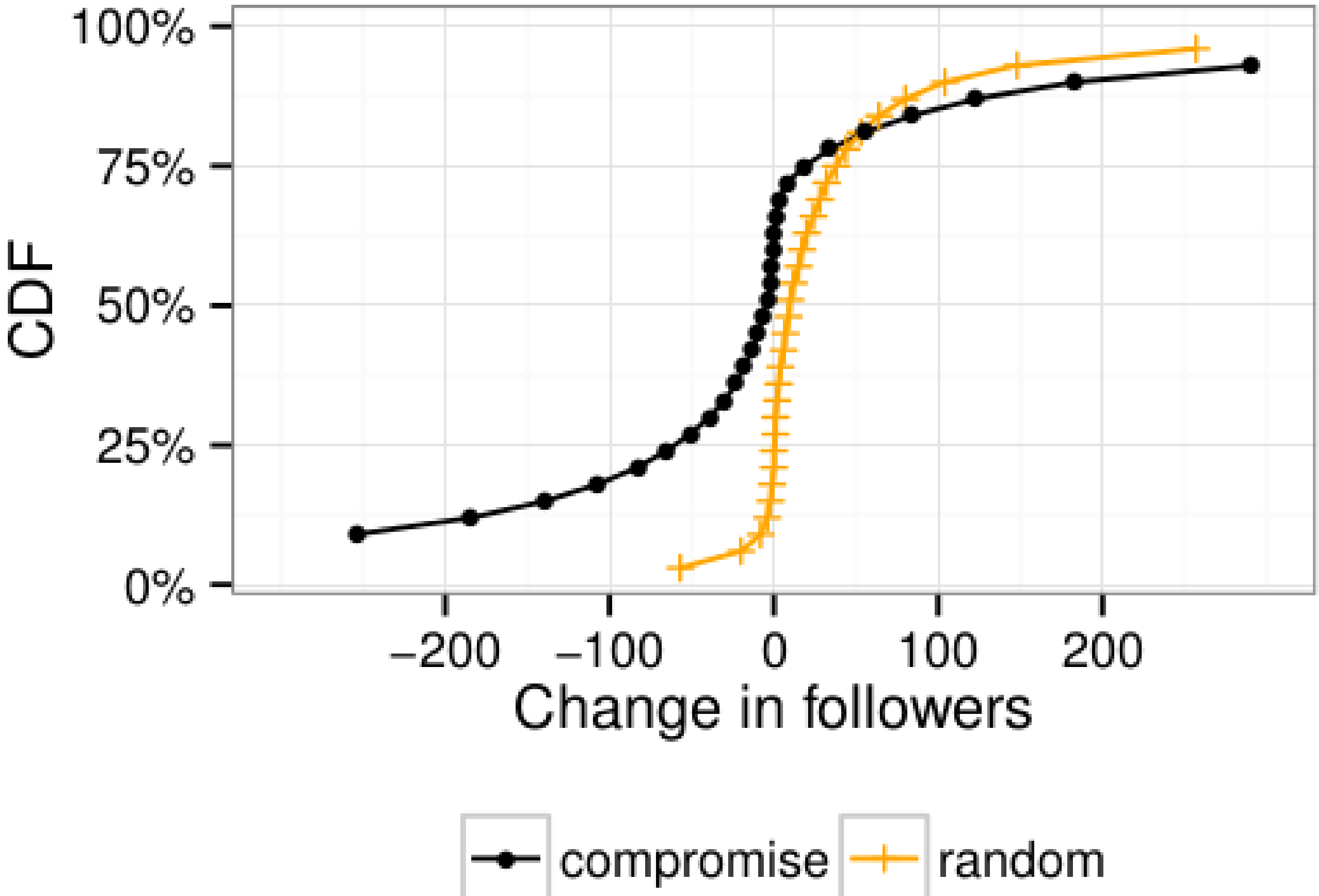
Compromise Demographics

Compromise Demographics

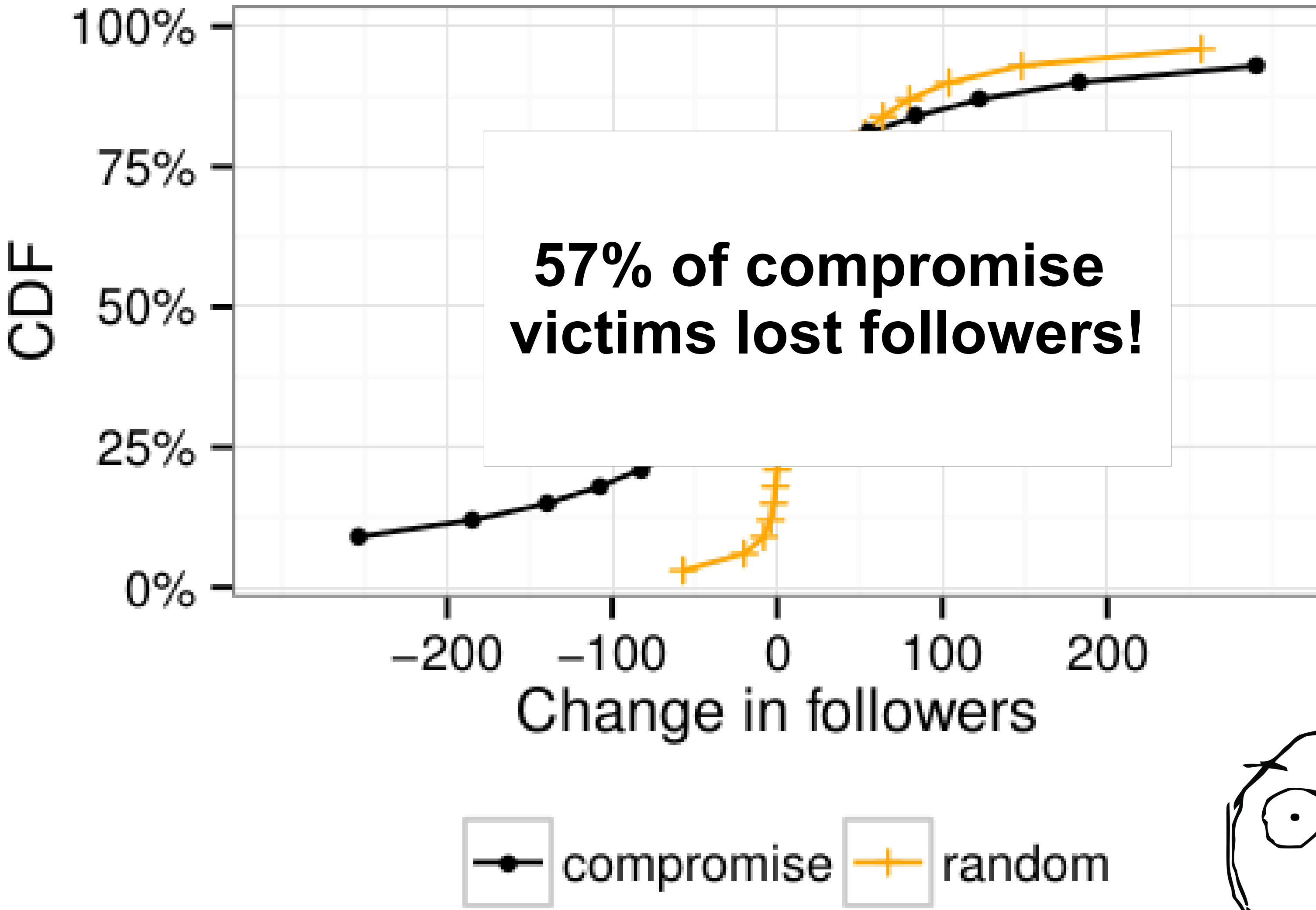


Accounts After Compromise

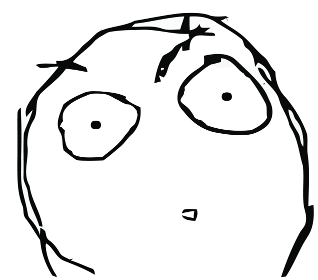
Accounts After Compromise



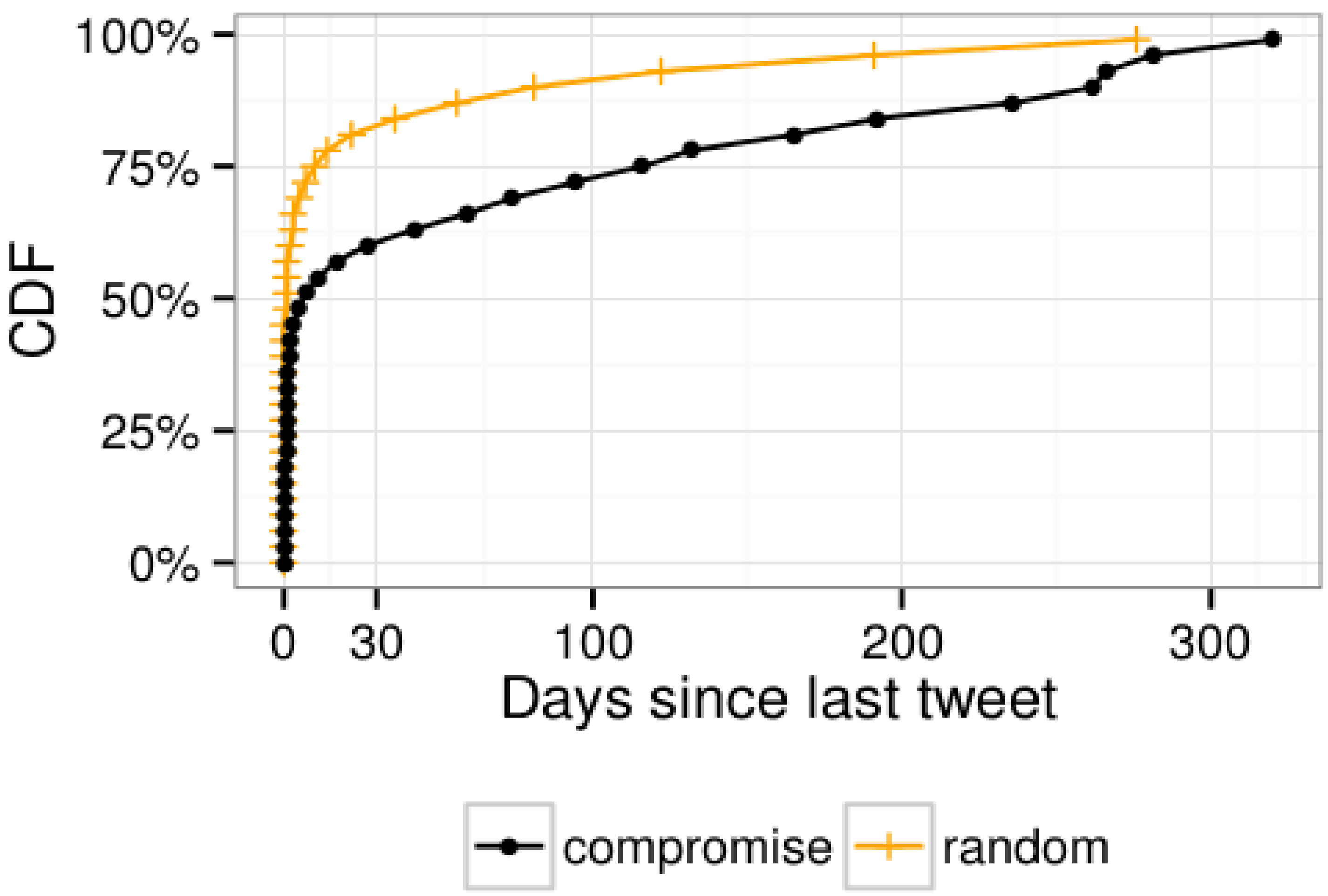
Accounts After Compromise



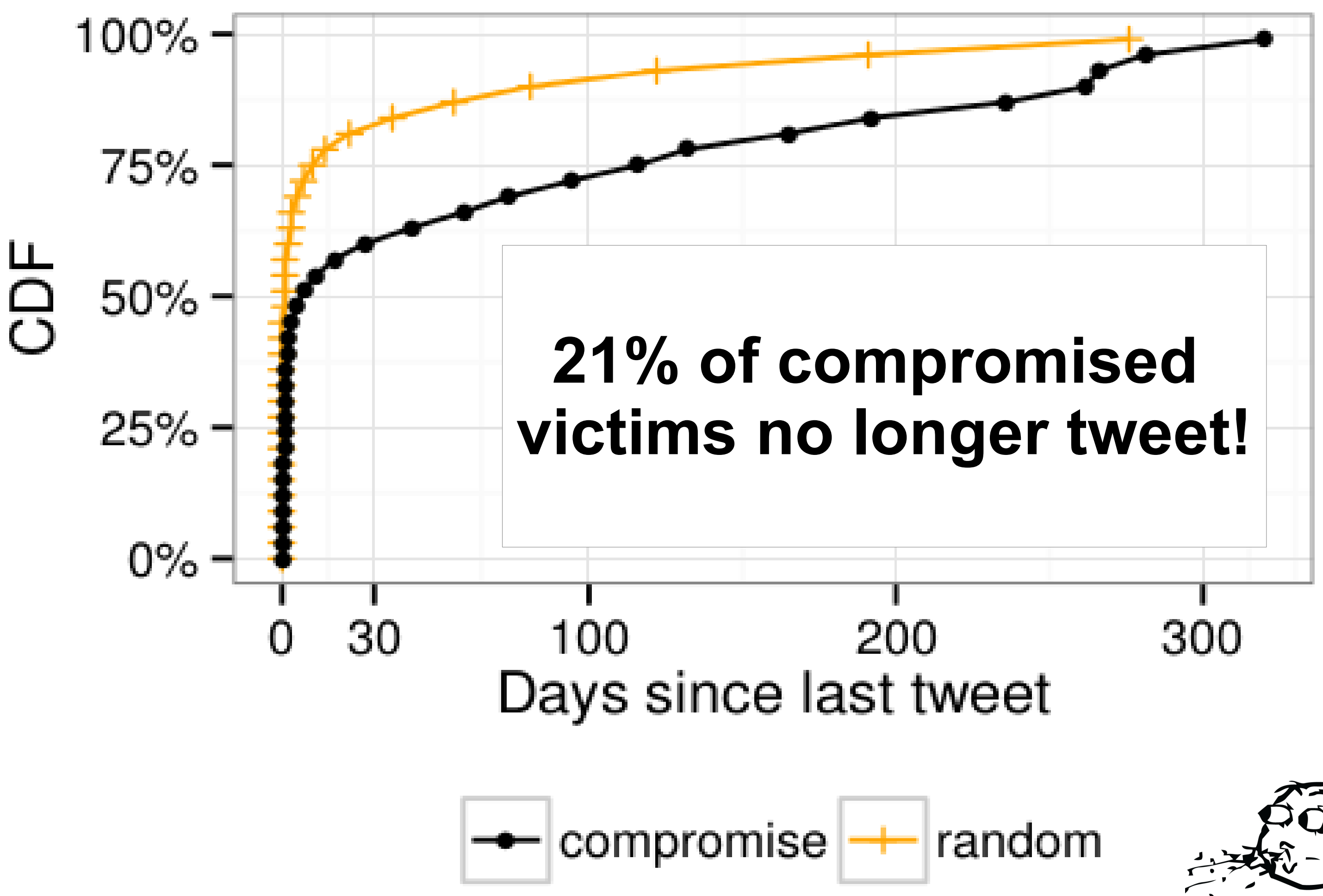
57% of compromise victims lost followers!

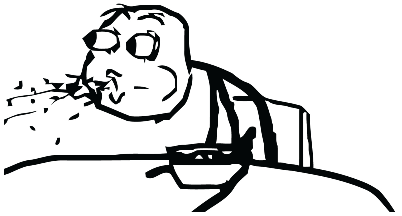


Accounts After Compromise



Accounts After Compromise

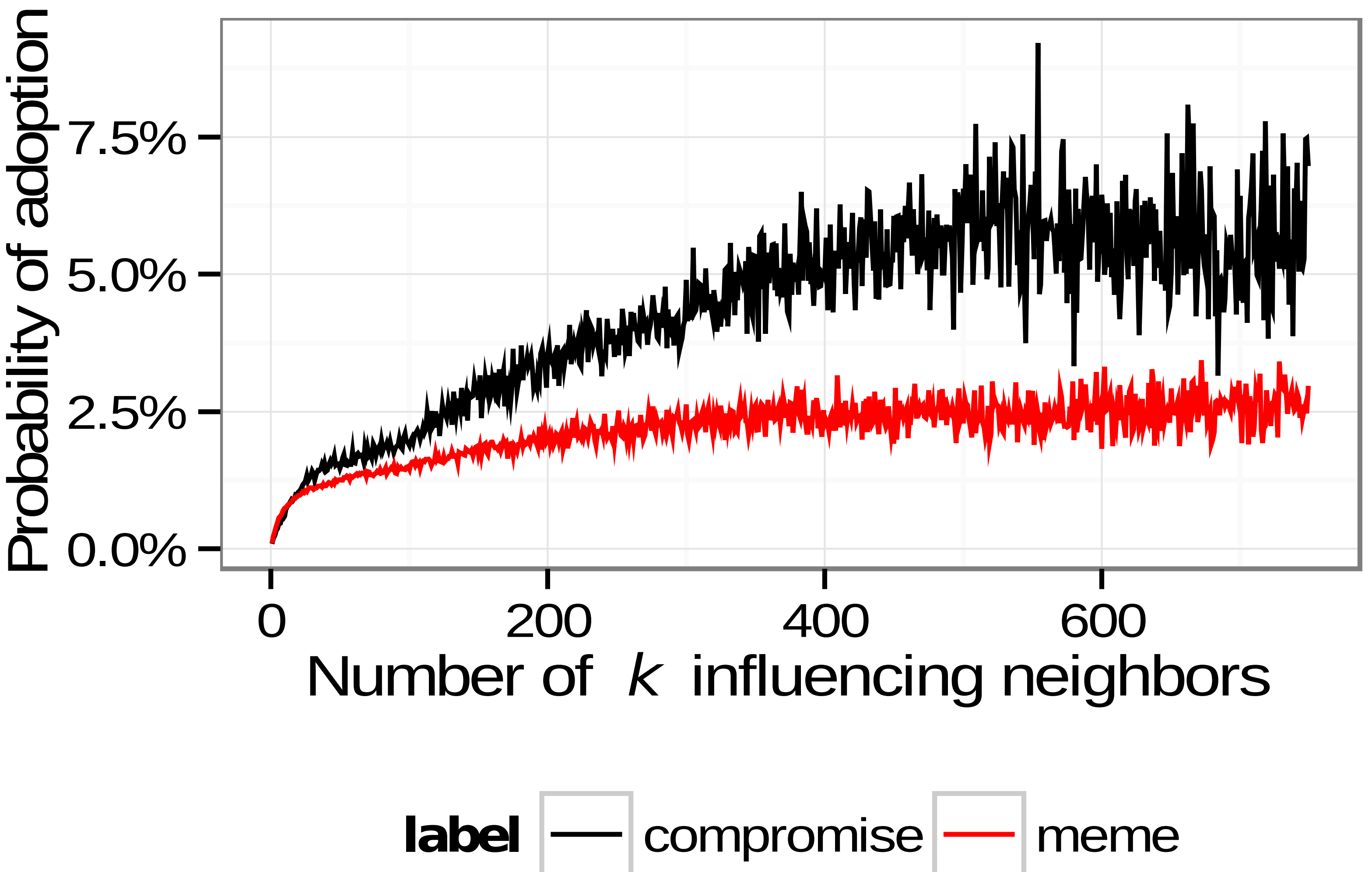




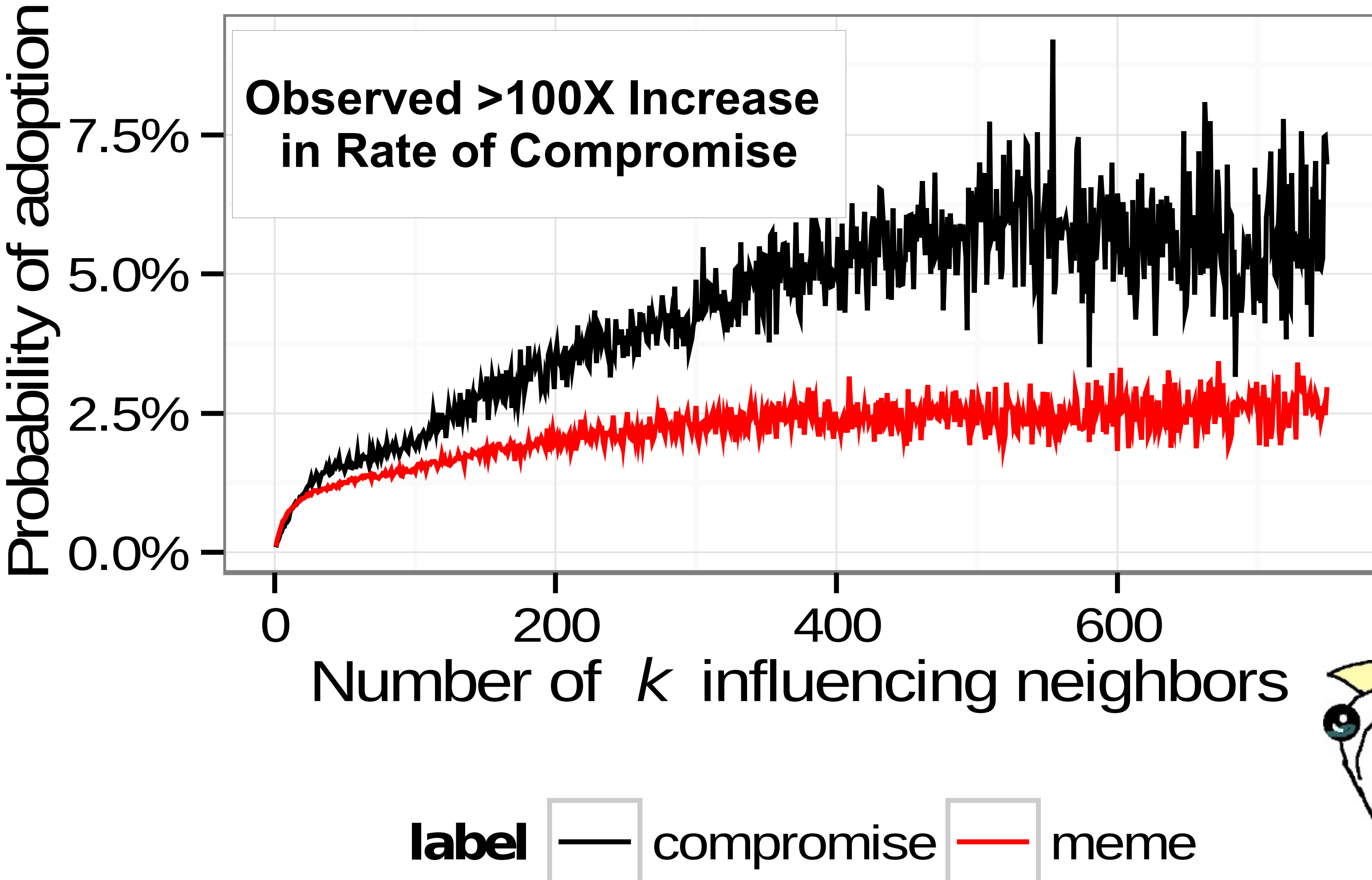
Sources of Compromise

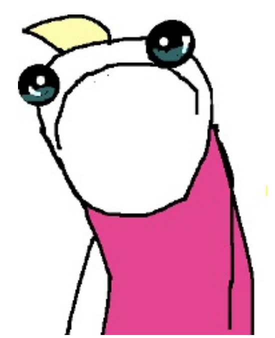
- Potential sources
  - Password brute-force
  - Database dumps
  - Social contagion (i.e. spread via your friends)
  - External contagion (i.e. driveby download site)

Compromise Can Spread



Compromise Can Spread





Sources of Compromise

- Potential sources
  - Password brute-force
  - Database dumps
  - Social contagion (i.e. spread via your friends)
  - External contagion (i.e. driveby download site)

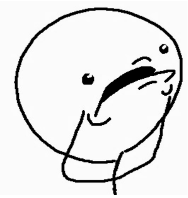
Sources of Compromise

- Potential sources
  - Password brute-force
  - Database dumps
  - Social contagion (i.e. spread via your friends)
  - External contagion (i.e. driveby download site)
- Defense: Early victims are indicators. If spread is on Twitter, quarantining can help.

Summary

Summary

- Is compromise occurring at a large scale?



YES! 14 million victims!

Summary

- Is compromise occurring at a large scale?

YES! 14 million victims!

- What do miscreants do with compromised accounts?

$$$ Profit! $$$

Summary

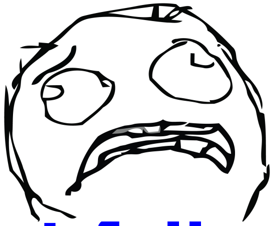
- Is compromise occurring at a large scale?

YES! 14 million victims!

- What do miscreants do with compromised accounts?

$$$ Profit! $$$

- How do users react to compromise?



Bad! 21% of victims quit, 57% lost followers

Summary

- Is compromise occurring at a large scale?

YES! 14 million victims!

- What do miscreants do with compromised accounts?

$$$ Profit! $$$

- How do users react to compromise?

Bad! 21% of victims quit, 57% lost followers

- How might compromise be occurring?

Highly potent social contagions

frankli@cs.berkeley.edu