Privacy and Security in Online Social Media
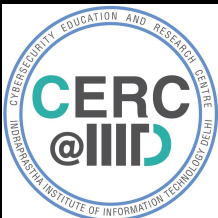
Course on NPTEL

NOC21-CS28

Week 8.2

Ponnurangam Kumaraguru ("PK")
Full Professor
ACM Distinguished Speaker
fb/ponnurangam.kumaraguru, @ponguru

Anonymous Networks

- 4chan

- Whisper

- Secret

- Yik Yak

- Wickr

Why use Anonymous Networks?

- Increasing awareness of privacy
- Snowden disclosures
- PRISM Surveillance program
- Bal Thackeray incident
- Many other incidents around the world

What is Whisper?

https://whisper.sh

What is Whisper?

https://www.youtube.com/watch?v=pX9I9kR2tTc

Hearts / Chat

Terminology / Claims

- Whispers
- Replies
- Anonymous names
- Does not associate any personal information with user ID
- Does not archive any user history
- Does not support persistent social links between users
- "Heart" a message anonymously
- Private messages

Whisper

Goals

● How do whisper users interact in an anonymous environment?

● Do users form communities similar to those in traditional social networks?

● Does whisper's lack of identities eliminate strong ties between users?

● Does it eliminate stickiness critical to long term engagement as in traditional SN?

Data collection

- Feb 6th – May 1st 2014
- Collected "Latest" list by scrapping
- Data include
  - WhisperID
  - Timestamp
  - Plain text of the whisper
  - Author's nickname
  - A location tag
  - # of replies (marked with the whisper)
  - Likes

Data collection

- 9,343,590 whispers

- 15,268,964 replies

- 1,038,364 GUIDs
    - Global Universal Identifier
    - Makes it possible to track user, but was removed in June 2014

- Interacted with Whisper team

Data
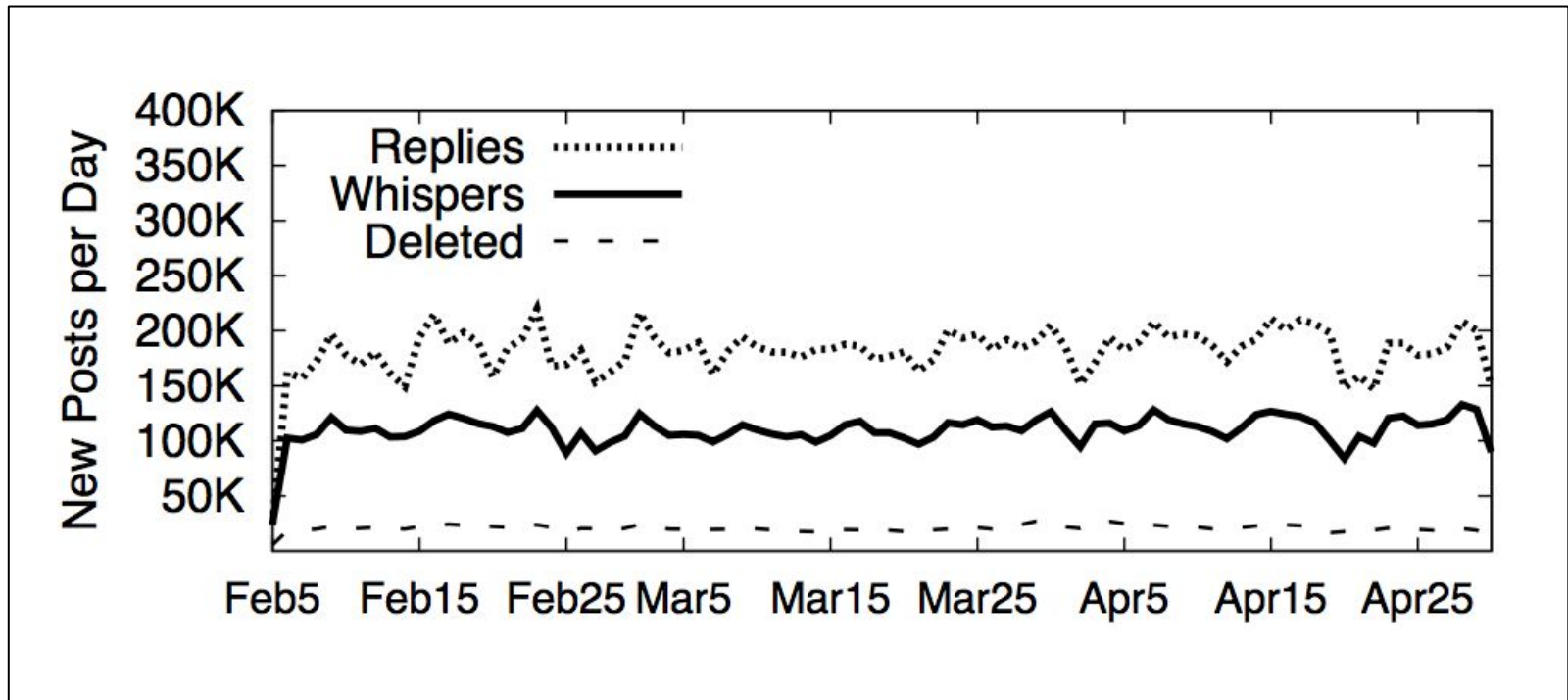
- 55% of whispers receives no replies
- 25% have a chain of at least 2 replies

Time between original & reply
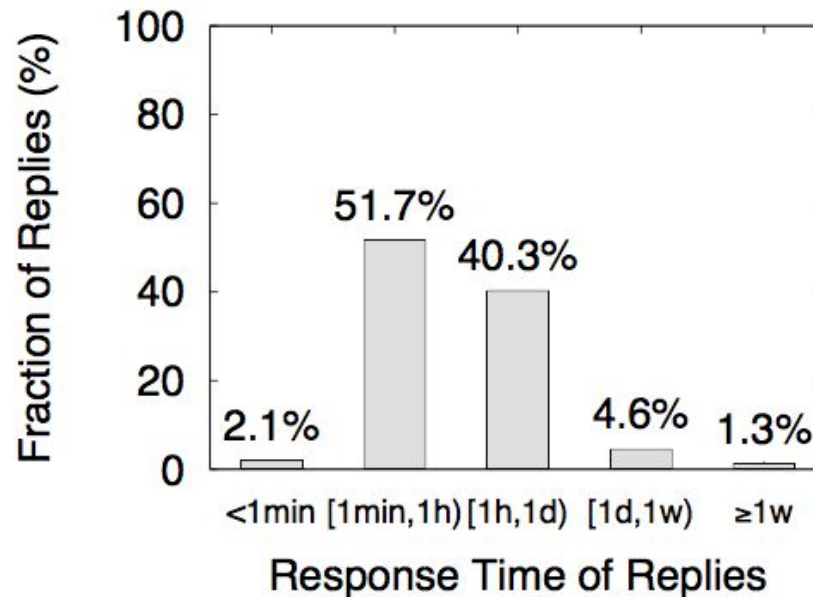
- 54% of replies arrive within an hour of the original whisper
- 94% of replies arrives in one day
- 1.3% of replies arrive a week or more
- "If a whisper does not get attention shortly after posting, it is unlikely to get attention later."

Posts per user

- 80% users post less than 10 total whispers or replies

- 15% of users only post replies but no original whispers

- 30% of users only post whispers but no replies

Network analysis

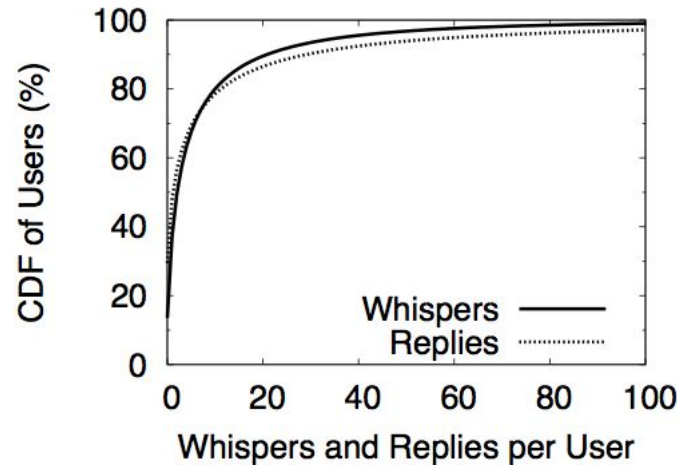| Graph	# of Nodes	# of Edges	Avg. Degree	Clustering Coef.	Avg. Path Length	Assortativity Coef.
Whisper	690K	6,531K	9.47	0.033	4.28	-0.011
Facebook	707K	1,260K	1.78	0.059	10.13	0.116
Twitter	4,317K	16,972K	3.93	0.048	5.52	-0.025

● High average Degree. Users interact with large sample of other users.

● Whisper users are likely to interact with complete strangers who are highly unlikely to interact with each other (low clustering coef.)

● 100 random nodes. Avg. path length calculated. Shortest average path among 3.

● Above 3 used to infer that graph is random than "small world"

Network analysis

Graph	# of Nodes	# of Edges	Avg. Degree	Clustering Coef.	Avg. Path Length	Assortativity Coef.
Whisper	690K	6,531K	9.47	0.033	4.28	-0.011
Facebook	707K	1,260K	1.78	0.059	10.13	0.116
Twitter	4,317K	16,972K	3.93	0.048	5.52	-0.025

- Assortativity measures the probability for nodes in a graph to link to other nodes of similar degrees.

- Close to zero → random graph

Content moderation

- 1.7 million whispers have been deleted in 3 months

- 18% of content deleted compared to 4% in Twitter

Content moderation: Process

- Extracted keywords from all whispers

- Removed common stop-words

- Removed words that appear in less than 0.05% of whispers

- Compute *deletion ratio* for each word
  - # of deleted whispers with this word / all whispers with this keyword

- Rank the words with deletion ratio

- Top and bottom keywords

Content moderation: Process

● Run on all 9 million original whispers

● 1.7 M were deleted

● 2,324 keywords ranked by deletion ratio

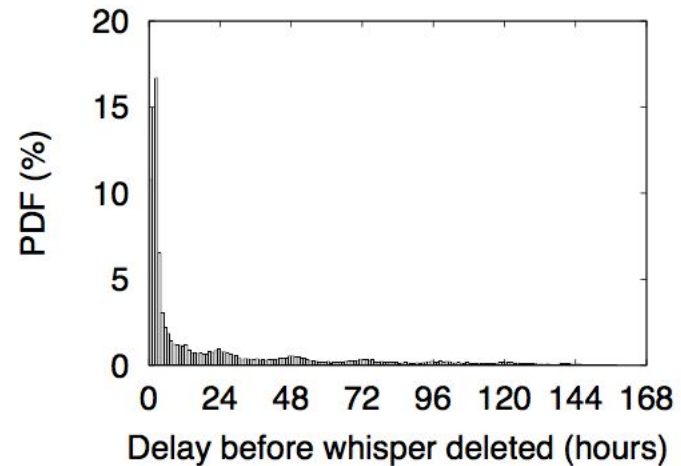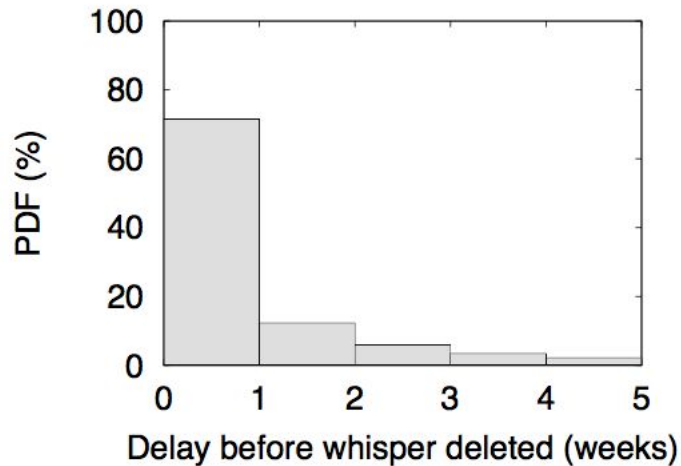● Manually put them in categories

Content moderation

Topic	Top 50 Keywords Most Related to Deleted Whispers
Sexting (36)	sext, wood, naughty, kinky, sexting, bj, threesome, dirty, role, fwb, panties, vibrator, bi, inches, lesbians, hookup, hairy, nipples, freaky, boobs, fantasy, fantasies, dare, trade, oral, takers, sugar, strings, experiment, curious, daddy, eaten, tease, entertain, athletic
Selfie (7)	rate, selfie, selfies, send, inbox, sends, pic
Chat (7)	f, dm, pm, chat, ladys, message, m

| Topic	Top 50 Keywords Least Related to Deleted Whispers
Emotion (17)	panic, emotions, argument, meds, hardest, fear, tears, sober, frozen, argue, failure, unfortunately, understands, anxiety, understood, aware, strength
Religion (10)	beliefs, path, faith, christians, atheist, bible, create, religion, praying, helped
Entertain. (8)	episode, series, season, anime, books, knowledge, restaurant, character
Life story (6)	memories, moments, escape, raised, thank, thanks
Work (5)	interview, ability, genius, research, process
Politics (1)	government
Others (3)	exactly, beginning, example

Deletion delay

- 70% of deleted whispers are "deleted" within one week after posting

- 2% of the whispers stay for more than a month
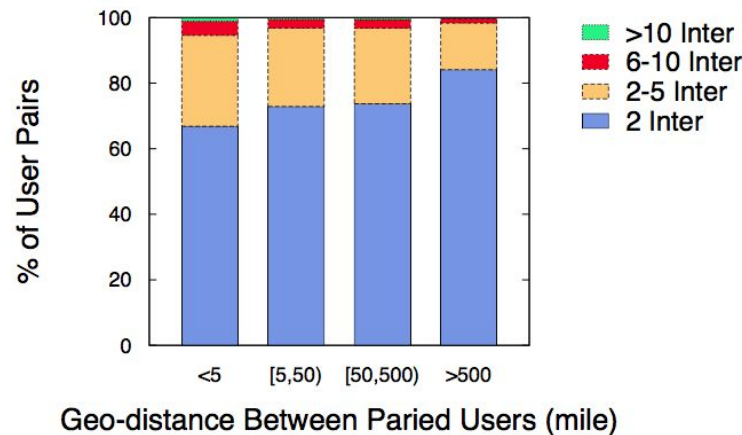 - Done by moderators

Deletion delay
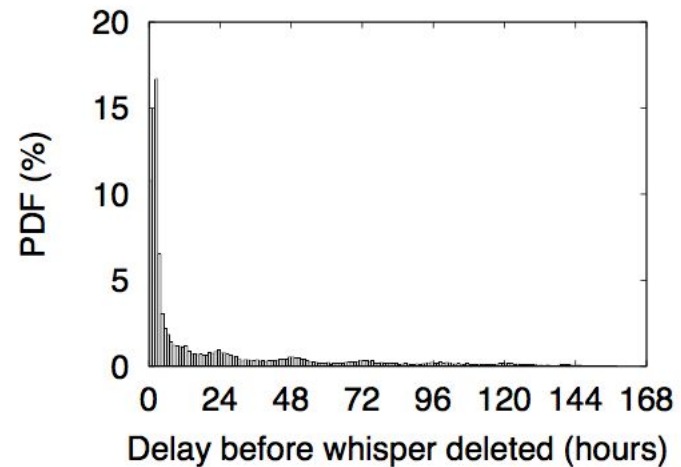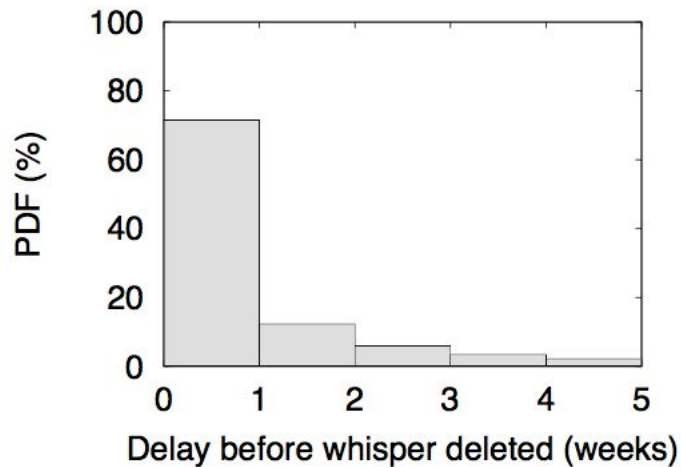
- Fine grained analysis
- Recrawled for 200K latest whispers
- 32,153 was deleted
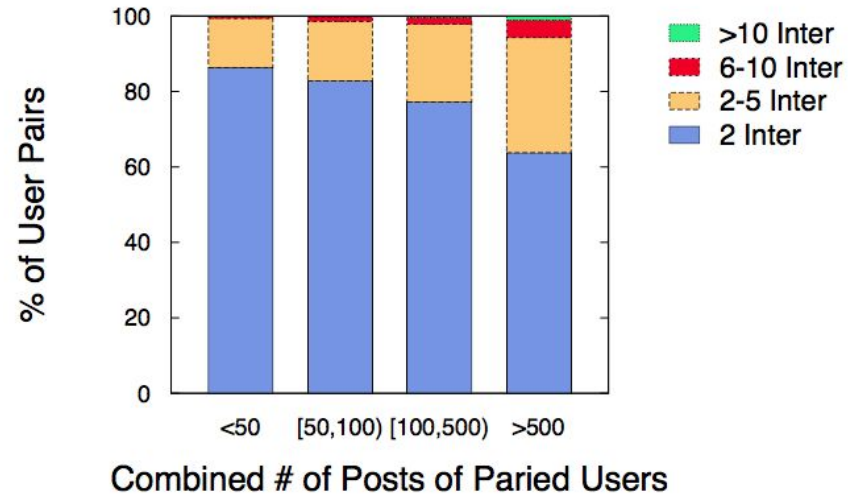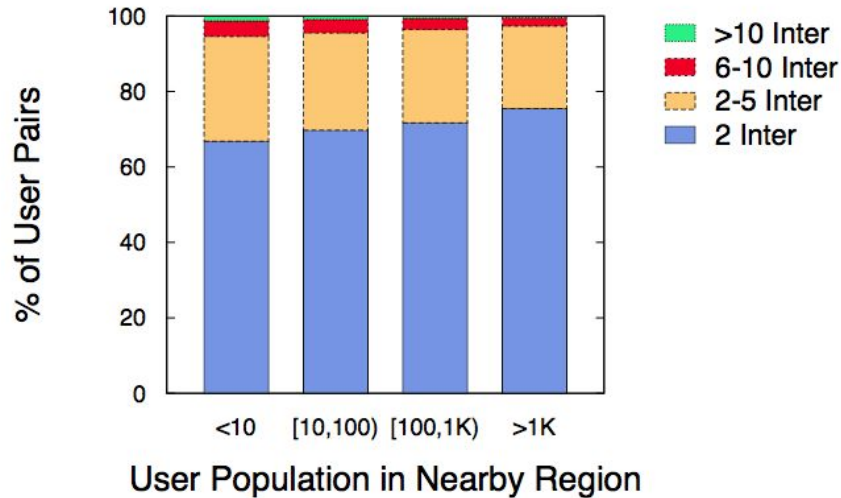- Peak deletion 3 – 9 hrs
- Majority deletes within 24 hrs

User interactions

- 503K user pairs
- 90% of the two users co-located in the same "State"
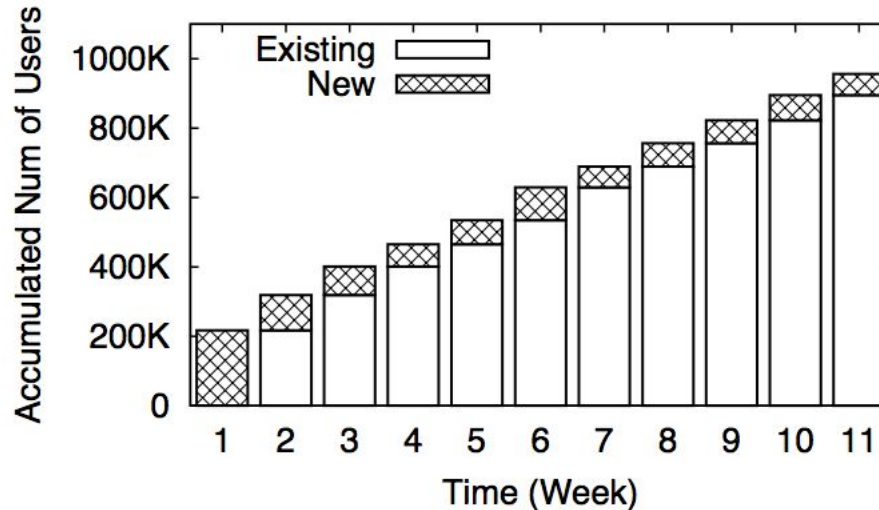- 75% have their distance < 40 miles

User interactions

- Smaller user population in same nearby area, higher chance of encounter
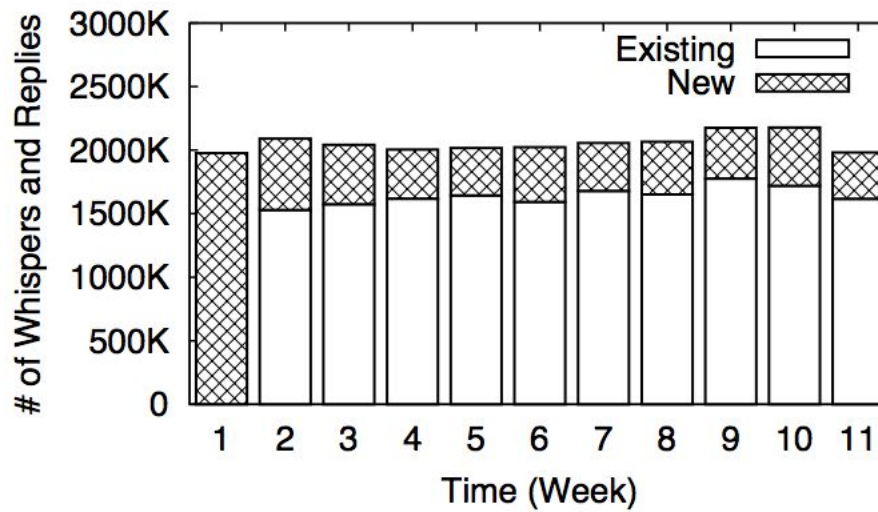- More whispers 2 users post, more likely they encounter each other

User engagement

- Roughly 80K users per week
- Daily new posts in the entire network remain stable, despite new users (earlier conclusion)
- Shows users "disengage"

User engagement

- # of whispers and replies by both new and old users
- New users make 20% contribution in the content
- Content by new users does not grow significantly

Conclusions

- Clearly different from traditional social networks

- Without strong user identities or persistent social links, users interact with strangers

- Moderation is necessary

References

- http://www.cs.ucsb.edu/~ravenben/publications/pdf/whisper-imc14.pdf

Thank you

pk@iiitd.ac.in

precog.iiitd.edu.in

fb/ponnurangam.kumaraguru