Poirot: Private Contact Summary Aggregation

Yanping Zhang[1]*, Chenghong Wang[1]*, David Pujol[1]*, Johes Bater[1], Matthew Lentz[1,2], Ashwin Machanavajjhala[1], Kartik Nayak[1], Lavanya Vasudevan[3,4], Jun Yang[1]

Department of Computer Science, Duke University[1], VMware Research[2], Department of Family Medicine and Community Health, Duke University[3], Duke Global Health Institute[4]

Duke

Poirot: In a Nutshell

Physical distancing between individuals is key to preventing the spread of contagious diseases.

Poirot: Measures physical distancing while ensuring user privacy.

- Measures physical distancing through contact events between individuals (independent of diagnosis status).
- Protects user privacy by collecting and releasing aggregate statistics that can't be linked back to every individual.

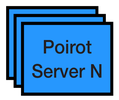
Poirot complements traditional contact tracing techniques by enabling more proactive measures.

Poirot collects and shares aggregate statistics without linkability.

Threat Model

Semi-honest: Honestly participate in the protocol but may attempt to learn sensitive information.

Learn their own number of contacts with locations and times.



Learn which user is using the system and metadata. e.g., how many records are updated each day, times of updates, and whether a user has sufficient number of tokens.

Learn the set of users participating in the system.

Untrusted

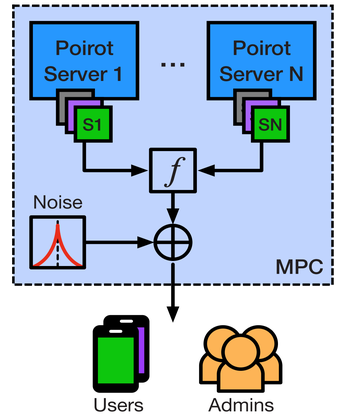
Administrators and potentially everyone learn differentially-private aggregate statistics.

Design: Data Processing

Use multiparty computation (MPC) and differential privacy (DP) to compute and release noisy aggregates.

What is the average number of contacts for <location, time> pair?

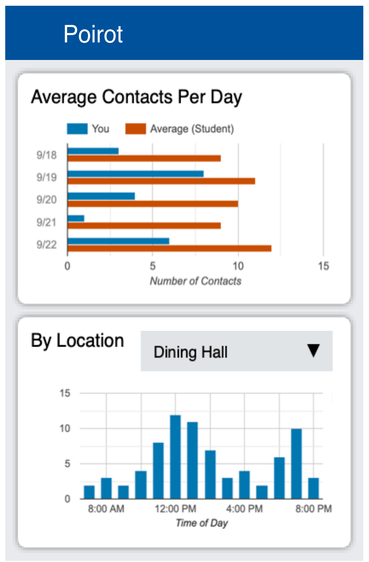
e.g., query f = SELECT * FROM contact_summary_table GROUPBY location and time.



MPC allows computation on secret-shared data. DP ensures statistics do not reveal individual's data.

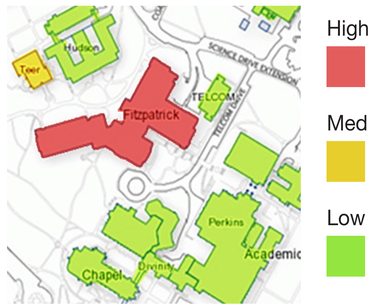
How would Poirot be used?

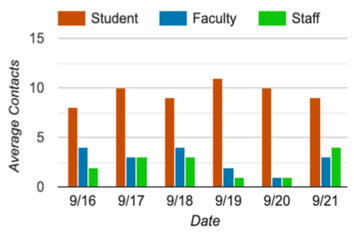
Individuals



How many contacts do I have on a daily basis? When is it safest for me to visit a give building?

Administrators



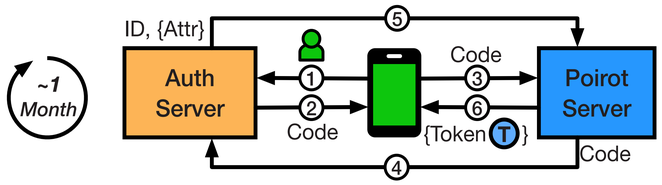


Which buildings require policy changes? Are certain groups at higher risk?

Poirot provides actionable information to both individual users and decision makers.

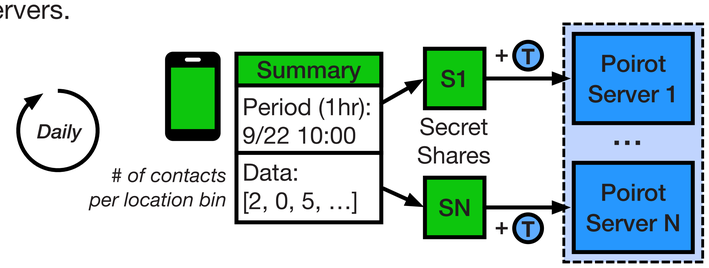
Design: Data Collection

Permission: Access control for the set of participating users. Blind signatures decouple identity from uploaded summaries.



Discover contacts: Broadcast pseudorandom ephemeral identifiers ensures long term unlinkability but allows detecting contacts.

Upload summary: Upload secret-shared summaries to compute servers.

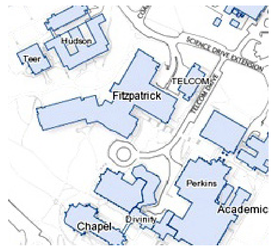


Servers only learn metadata about contact summaries.

Evaluation

How scalable is the server-side data processing?

Duke University:
- 20K members
- 256 buildings
- Daily and hourly contacts



North Carolina:
- 10M residents
- 100 counties
- Daily contacts

Performance for computing aggregate statistics:

	Number of Bins			Execution Time	
Case	Location	Time	Users	App (ms)	Server (s)
Duke	256	1	20K	15.2 ± 4.5	3.9 ± 0.0
Duke	256	24	20K	366.1± 8.9	94.3 ± 0.4
NC	100	1	10M	6.0 ±4.4	776.1 ± 1.7

Poirot scales effectively even for large (e.g., state-wide) deployment scenarios.

Authors: * denotes equal contribution.

This work was partially supported by the NSF Award 2029853.