Protect your score: Contact-Tracing with Differential Privacy Guarantees

Rob Romijnders[1], Christos Louizos[2], Yuki M. Asano[1], Max Welling[1]

[1]University of Amsterdam, [2]Qualcomm AI research

Concerns about privacy are the main reason for low adoption of contact-tracing algorithms, even though they have been shown to be effective [1,2,3]. We present a novel decentralized algorithm that satisfies **differential privacy (DP)** against the following attack:

The adversary installs a contact tracing app and makes contact with a victim. The next day, the adversary observes a change in its covidscore, which is due to the victim, and the attacker can reconstruct the covidscore of the victim.

DP definition: for every $\varepsilon > 0$, $\delta \in [0, 1]$, an algorithm $f(\cdot)$ satisfies the following constraint for any outcome Φ, and any two adjacent data sets D, D' that differ in at most one element [6]:

$$p(f(D) \in \Phi) \leq e^{\varepsilon} p(f(D') \in \Phi) + \delta$$

Two data sets $D, D\prime$ are adjacent when the covidscore of one contact differs between the data sets, i.e. $d(D, D') = 1$. Then upper bound

$$\Delta = \max_{\mu_1, \mu_1'} |f(\{(\mu_1, t_1)\} \cup D) - f(\{(\mu_1', t_1)\} \cup D)| \; \forall \; D.$$

For our novel DPFN, we uncover a composite structure in the update function for statistical contact tracing. Each contact, c, sends a message $\omega_{c,t} = 1 - p_1 \phi_{c,t}$ to user u, and this user calculates covidscore $\phi_{u,t}$:

$$\phi_{u,t} = F_1(\phi_{1,t-5}, \phi_{2,t-5}, \cdots, \phi_{C_1,t-5}, \phi_{1,t-3}, \phi_{2,t-3}, \cdots, \phi_{C_2,t-3})$$

The product structure: when the release of each product term is DP, the update function is DP by the post-processing property [6].

$$\phi_{u,t} = F_2(\prod_{i=1}^{C_5} \omega_{i,t-5}, \quad \prod_{i=1}^{C_3} \omega_{i,t-3})$$

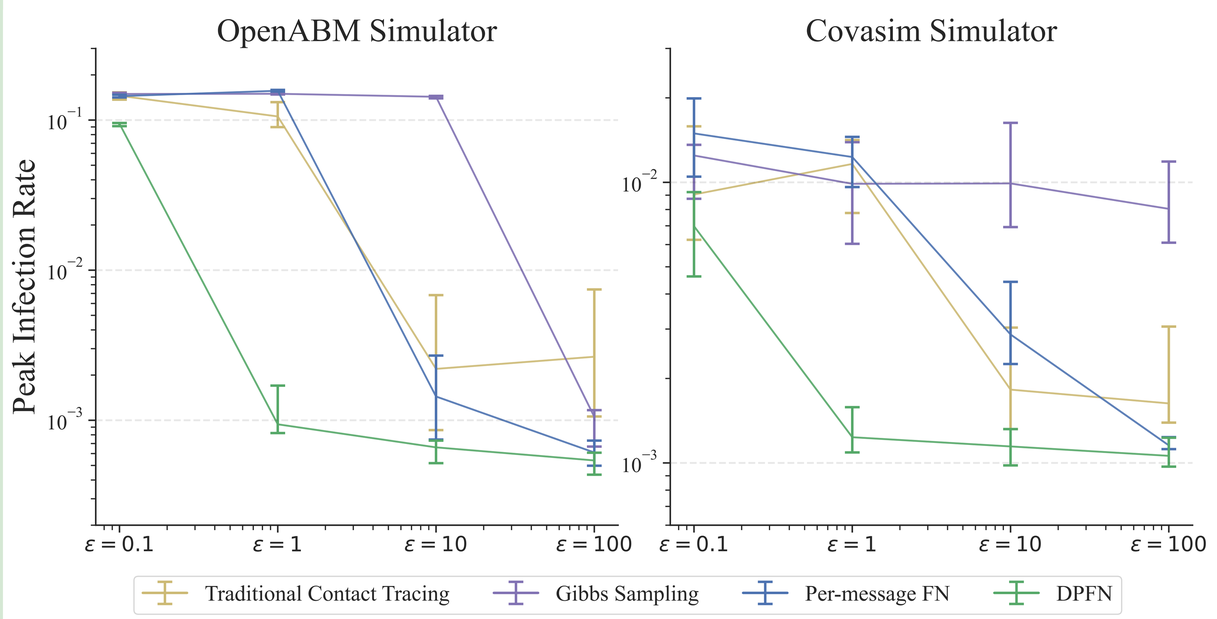
We achieve Renyi DP with the log-normal distribution, which is closed under multiplication and has a closed-form for its divergence:

$$D_a(p_u | p_v) = \frac{a}{2C\sigma^2}(\mu_u - \mu_v)^2$$

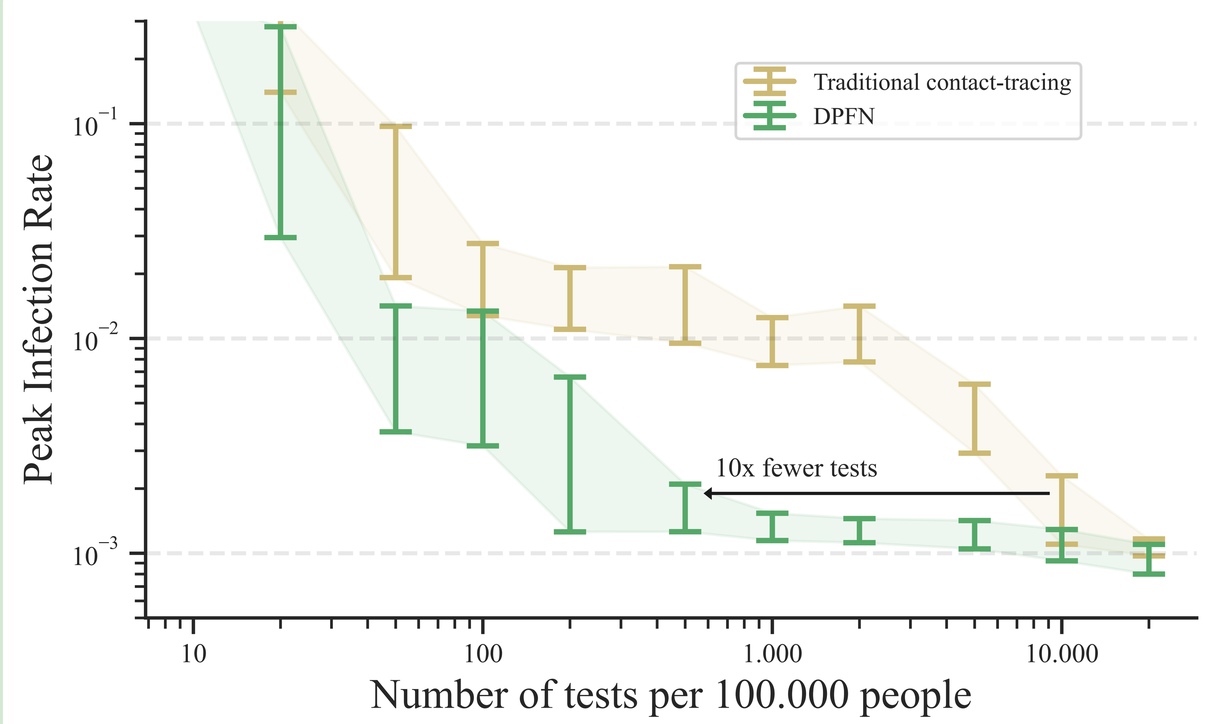
After finding the largest deviation in the means, clipping messages in the interval $[\gamma_l, \gamma_u]$, the noise parameter for the log-normal distribution is:

$$\sigma^2 \geq \frac{a}{2C\rho}\left(\log(1 - \gamma_u p_1) - \log(1 - \gamma_l p_1)\right)^2.$$

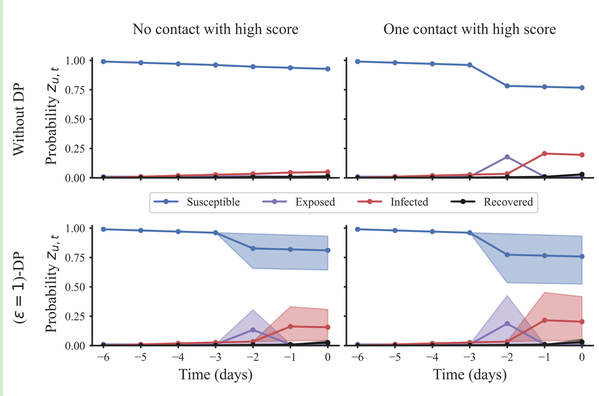
This equation shows that achieving DP requires less noise a) with more contacts C or b) with tighter clipping bounds.
In the appendix, we present a closed-form solution for converting the parameters of Renyi DP to convert to (ε, δ)-DP.



Trade-off between Peak Infection Rate (PIR) and privacy-level. At ε=1, DPFN achieves a lower PIR than all other methods. In general, it is pareto optimal. The simulators OpenABM and Covasim are the two most widely used simulators for COVID19 [4,5].



Trade-off between PIR and the number of available tests. The noise from DP can counteracted with more testing. Compared to traditional contact tracing, for a low PIR of about 2-3 per thousand users, DPFN needs more than ten times fewer tests.



The effect of differential privacy on inference. A user has two contacts with a low score, left column, while in the right column, one contact has a high score. The median prediction for infectiousness is higher in the right column (red solid line), but noisy predictions overlap, which gives plausible deniability and privacy.

Test setup	No privacy	DPFN	DPFN+
(fpr 0.0%; fnr 0.0%)	0.5 [0.5,0.6]	1.1 [0.7,1.4]	0.6 [0.5,0.7]
(fpr 1%; fnr 0.1%)	0.5 [0.4,0.6]	1.1 [0.9,1.7]	0.6 [0.5,0.6]
(fpr 10%; fnr 1%)	0.6 [0.5,0.8]	17.6 [11.4,20.4]	0.9 [0.7,1.0]
(fpr 25%; fnr 3%)	0.6 [0.5,0.8]	46.6 [40.4,48.0]	0.7 [0.6,0.8]
No testing	200 [190,212]		

An important message to policy makers. DP reduces robustness against noisier tests, but this can be counteracted with more available tests (DPFN+ has more tests than the results in the DPFN column). Units in the table are 1 daily infection per thousand users.
fpr = false positive rate, fnr = false negative rate.

r.romijnders@uva.nl, romijndersrob@gmail.com
github.com/RobRomijnders/dpfn_aaai

[1] "Epidemic mitigation by statistical inference from contact tracing data," Baker et al. PNAS 2020
[2] "CRISP: A Probabilistic Model for Individual-Level COVID-19 Infection Risk Estimation Based on Contact Data," Herbrich et al. 2020
[3] "No time to waste: Practical statistical contact tracing with few low-bit messages," Romijnders et al. AISTATS 2023
[4] "OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19... ," Hinch et al. PLOS 2021
[5] "Covasim: an agent-based model of COVID-19 dynamics and interventions," Kerr et al. PLOS Computational Biology 2021
[6] "The algorithmic foundations of differential privacy," Dwork et al. Foundations and Trends in Theoretical Computer Science 2014

University of Amsterdam Faculty of Science
QUVA Deep Vision Lab
Qualcomm