ShareTrace: Proactive Contact Tracing with Asynchronous Message Passing

Ryan Tatton

Case Western Reserve University

7 February 2025

Introduction: Types of Contact Tracing

- Digital contact tracing (DCT)
- Proximity tracing
- Decentralized DCT
  - Broadcast model
  - Message-oriented model

Introduction: Limitations of Other Approaches

- No DCT approach exists that incorporates both non-diagnostic information and indirect contacts to estimate infection risk.
- Accounting for indirect contact can substantially improve the efficacy of DCT [12].
- Cherini et al. [5] propose exchanging pseudonyms of indirect contacts, but restrict themselves to diagnostic testing.
- Gupta et al. [7] incorporate non-diagnostic information, but do not account for indirect contact.

Introduction: ShareTrace

- Accounts for both non-diagnostic information and indirect contact to estimate infection risk.
- Developed in collaboration with Dataswyft [2].
- Ayday, Yoo, and Halimi [1] designed ShareTrace to use proximity tracing for contact discovery.
  - In practice, this was infeasible, because Apple and Google's Exposure Notification API did not permit the user's ephemeral identifiers to be stored remotely in a Dataswyft Personal Data Store.

Introduction: Prior Designs and Implementations

- “Thinking like a vertex” with Apache Giraph
- Factor subgraph actors
- Driver-monitor-worker framework
- Projected subgraph actors [18]
- Contact search

Proposed Design: Definitions

- *Risk score*, $s_t \in [0, 1]$: a timestamped infection probability where $t \in \mathbb{N}$ is the time of its computation
- *Symptom score*: prior infection probability; accounts for an individual's demographics, symptoms, and diagnosis [4, 11]
- *Exposure score*: posterior infection probability; accounts for direct and indirect contact with others

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i
2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$
3: $r_i^{(n)} \leftarrow \infty$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

3: $r_i^{(n)} \leftarrow \infty$

4: **while** $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

$\qquad \mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \mathinner{..} n-1] \,\}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

$\quad \mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

$\quad \lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

3: $r_i^{(n)} \leftarrow \infty$

4: **while** $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

5: $\quad \mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

6: $\quad \lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

7: $\quad R_i^{(n)} \leftarrow$ top k of $\{\, \lambda_{ji}^{(n)} \mid f_{ij} \in N_i \,\}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i
$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$
$r_i^{(n)} \leftarrow \infty$
while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$
    $\mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$
    $\lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$
    $R_i^{(n)} \leftarrow$ top k of $\{\, \lambda_{ji}^{(n)} \mid f_{ij} \in N_i \,\}$
    $r_i^{(n)} \leftarrow \max R_i^{(n)}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

    $\mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \mathinner{..} n-1] \,\}$

    $\lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

    $R_i^{(n)} \leftarrow$ top k of $\{\, \lambda_{ji}^{(n)} \mid f_{ij} \in N_i \,\}$

    $r_i^{(n)} \leftarrow \max R_i^{(n)}$

return $\mathbf{r}^{(n)}$

Asynchronous Risk Propagation

CREATE-ACTOR

$a.contacts \leftarrow \emptyset$

$a.scores \leftarrow \emptyset$

$a.exposure \leftarrow$ NULL-RISK-SCORE

return a

NULL-RISK-SCORE

$s.value \leftarrow 0$

$s.time \leftarrow 0$

return s

Asynchronous Risk Propagation

$\textsc{Risk-Score-Ttl}(s)$

1: **return** $T_s - (\tau - s.time)$

$\textsc{Contact-Ttl}(c)$

1: **return** $T_c - (\tau - c.time)$

Asynchronous Risk Propagation

$\textsc{Handle-Risk-Score}(a, s)$

1: **if** $\textsc{Risk-Score-Ttl}(s) > 0$

Asynchronous Risk Propagation

$\textsc{Handle-Risk-Score}(a, s)$

if $\textsc{Risk-Score-Ttl}(s) > 0$

  $s.key \leftarrow [s.time, s.time + T_s)$

Asynchronous Risk Propagation

$\textsc{Handle-Risk-Score}(a, s)$

1: **if** $\textsc{Risk-Score-Ttl}(s) > 0$

2: $\quad s.key \leftarrow [s.time, s.time + T_s)$

3: $\quad \textsc{Merge}(a.scores, s)$

Asynchronous Risk Propagation

$\textsc{Handle-Risk-Score}(a, s)$

if $\textsc{Risk-Score-Ttl}(s) > 0$

  $s.key \leftarrow [s.time, s.time + T_s)$

  $\textsc{Merge}(a.scores, s)$

  $\textsc{Update-Exposure-Score}(a, s)$

Asynchronous Risk Propagation

Handle-Risk-Score(a, s)

if Risk-Score-Ttl$(s) > 0$

    $s.key \leftarrow [s.time, s.time + T_s)$

    Merge$(a.scores, s)$

    Update-Exposure-Score(a, s)

    for each $c \in a.contacts$

        Apply-Risk-Score(a, c, s)

Asynchronous Risk Propagation

UPDATE-EXPOSURE-SCORE(a, s)

1: **if** $a.exposure.value < s.value$

Asynchronous Risk Propagation

$\textsc{Update-Exposure-Score}(a, s)$

1: **if** $a.exposure.value < s.value$

2: $\quad a.exposure \leftarrow s$

Asynchronous Risk Propagation

$\textsc{Update-Exposure-Score}(a, s)$

1: **if** $a.exposure.value < s.value$

2: $\quad a.exposure \leftarrow s$

3: **else if** $\textsc{Risk-Score-Ttl}(a.exposure) \leq 0$

Asynchronous Risk Propagation

$\textsc{Update-Exposure-Score}(a, s)$

if $a.exposure.value < s.value$

    $a.exposure \leftarrow s$

else if $\textsc{Risk-Score-Ttl}(a.exposure) \leq 0$

    $a.exposure \leftarrow \textsc{Maximum}(a.scores)$

Asynchronous Risk Propagation

$\textsc{Apply-Risk-Score}(a, c, s)$

1: **if** $c.time + \beta > s.time$

Asynchronous Risk Propagation

Apply-Risk-Score(a, c, s)

if $c.time + \beta > s.time$

  $s'.value \leftarrow \alpha \cdot s.value$

Asynchronous Risk Propagation

$\textsc{Apply-Risk-Score}(a, c, s)$

if $c.time + \beta > s.time$

  $s'.value \leftarrow \alpha \cdot s.value$

  $\textsc{Send}(c.name, s')$

Asynchronous Risk Propagation

$\textsc{Set-Send-Threshold}(c, s)$

1: $s'.value \leftarrow \gamma \cdot s.value$

Asynchronous Risk Propagation

$\textsc{Set-Send-Threshold}(c, s)$

$s'.value \leftarrow \gamma \cdot s.value$

$c.threshold \leftarrow s'$

Asynchronous Risk Propagation

UPDATE-SEND-THRESHOLD(a, c)

if $c.threshold.value > 0$

Asynchronous Risk Propagation

$\textsc{Update-Send-Threshold}(a, c)$

1: **if** $c.threshold.value > 0$

2: **if** $\textsc{Risk-Score-Ttl}(c.threshold) \leq 0$

Asynchronous Risk Propagation

$\textsc{Update-Send-Threshold}(a, c)$

if $c.threshold.value > 0$

  if $\textsc{Risk-Score-Ttl}(c.threshold) \leq 0$

    $s \leftarrow \textsc{Maximum-Older-Than}(a.scores, c.time + \beta)$

Asynchronous Risk Propagation

$\textsc{Update-Send-Threshold}(a, c)$

if $c.threshold.value > 0$

    if $\textsc{Risk-Score-Ttl}(c.threshold) \leq 0$

        $s \leftarrow \textsc{Maximum-Older-Than}(a.scores, c.time + \beta)$

        $s'.value \leftarrow \alpha \cdot s.value$

Asynchronous Risk Propagation

$\textsc{Update-Send-Threshold}(a, c)$

if $c.threshold.value > 0$
    if $\textsc{Risk-Score-Ttl}(c.threshold) \leq 0$
        $s \leftarrow \textsc{Maximum-Older-Than}(a.scores, c.time + \beta)$
        $s'.value \leftarrow \alpha \cdot s.value$
        $\textsc{Set-Send-Threshold}(c, s')$

Asynchronous Risk Propagation

$\textsc{Apply-Risk-Score}(a, c, s)$

$\textsc{Update-Send-Threshold}(a, c)$

Asynchronous Risk Propagation

APPLY-RISK-SCORE(a, c, s)

1: UPDATE-SEND-THRESHOLD(a, c)

2: **if** $c.threshold.value < s.value$ **and** $c.time + \beta > s.time$

Asynchronous Risk Propagation

$\textsc{Apply-Risk-Score}(a, c, s)$

1: $\textsc{Update-Send-Threshold}(a, c)$

2: **if** $c.threshold.value < s.value$ **and** $c.time + \beta > s.time$

3: $\quad s'.value \leftarrow \alpha \cdot s.value$

Asynchronous Risk Propagation

$\textsc{Apply-Risk-Score}(a, c, s)$

$\textsc{Update-Send-Threshold}(a, c)$

if $c.threshold.value < s.value$ **and** $c.time + \beta > s.time$

  $s'.value \leftarrow \alpha \cdot s.value$

  $\textsc{Set-Send-Threshold}(c, s')$

Asynchronous Risk Propagation

$\textsc{Apply-Risk-Score}(a, c, s)$

$\textsc{Update-Send-Threshold}(a, c)$

if $c.threshold.value < s.value$ **and** $c.time + \beta > s.time$

  $s'.value \leftarrow \alpha \cdot s.value$

  $\textsc{Set-Send-Threshold}(c, s')$

  $c.buffered \leftarrow s'$

Asynchronous Risk Propagation

HANDLE-FLUSH-TIMEOUT(a)

1: **for each** $c \in a.contacts$

Asynchronous Risk Propagation

$\textsc{Handle-Flush-Timeout}(a)$

for each $c \in a.contacts$

    if $c.buffered \neq \textsc{nil}$

Asynchronous Risk Propagation

HANDLE-FLUSH-TIMEOUT(a)

for each $c \in a.contacts$

  if $c.buffered \neq$ NIL

    SEND($c.name, c.buffered$)

Asynchronous Risk Propagation

$\textsc{Handle-Flush-Timeout}(a)$

```
for each c ∈ a.contacts
    if c.buffered ≠ NIL
        SEND(c.name, c.buffered)
        c.buffered ← NIL
```

Asynchronous Risk Propagation

$\textsc{Handle-Flush-Timeout}(a)$

for each $c \in a.contacts$
    if $c.buffered \neq \textsc{nil}$
        $\textsc{Send}(c.name, c.buffered)$
        $c.buffered \leftarrow \textsc{nil}$
    if $\textsc{Contact-Ttl}(c) \leq 0$
        $\textsc{Delete}(a.contacts, c)$

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

1: **if** $\textsc{Contact-Ttl}(c) > 0$

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

if $\textsc{Contact-Ttl}(c) > 0$

$\quad c.threshold \leftarrow \textsc{Null-Risk-Score}$

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

```
if Contact-Ttl(c) > 0
    c.threshold ← Null-Risk-Score
    c.buffered ← nil
```

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

if $\textsc{Contact-Ttl}(c) > 0$

    $c.threshold \leftarrow \textsc{Null-Risk-Score}$

    $c.buffered \leftarrow \textsc{nil}$

    $c.key \leftarrow c.name$

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

if $\textsc{Contact-Ttl}(c) > 0$

$\quad c.\mathit{threshold} \leftarrow \textsc{Null-Risk-Score}$

$\quad c.\mathit{buffered} \leftarrow \textsc{nil}$

$\quad c.\mathit{key} \leftarrow c.\mathit{name}$

$\quad \textsc{Merge}(a.\mathit{contacts}, c)$

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

if $\textsc{Contact-Ttl}(c) > 0$

  $c.threshold \leftarrow \textsc{Null-Risk-Score}$

  $c.buffered \leftarrow \textsc{nil}$

  $c.key \leftarrow c.name$

  $\textsc{Merge}(a.contacts, c)$

  $s \leftarrow \textsc{Maximum-Older-Than}(a.scores, c.time + \beta)$

Asynchronous Risk Propagation

$\textsc{Handle-Contact}(a, c)$

if $\textsc{Contact-Ttl}(c) > 0$

    $c.\mathit{threshold} \leftarrow \textsc{Null-Risk-Score}$

    $c.\mathit{buffered} \leftarrow \textsc{nil}$

    $c.\mathit{key} \leftarrow c.\mathit{name}$

    $\textsc{Merge}(a.\mathit{contacts}, c)$

    $s \leftarrow \textsc{Maximum-Older-Than}(a.\mathit{scores}, c.\mathit{time} + \beta)$

    $\textsc{Apply-Risk-Score}(a, c, s)$

Experiment 1: Accuracy I

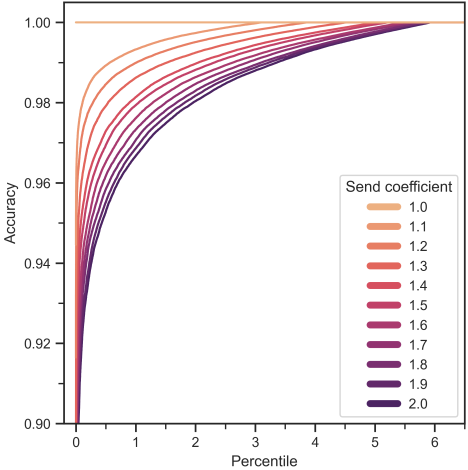

Figure: Cumulative accuracy distributions.

Experiment 1: Accuracy II

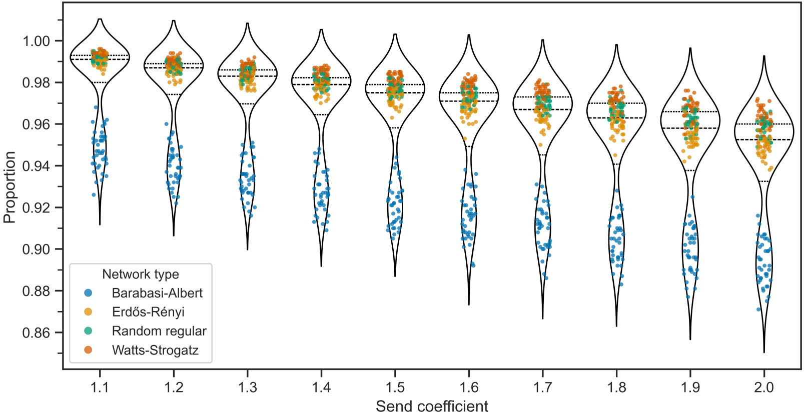

Figure: Send coefficient optimality distributions. The dashed line inside each violin marks the median. The upper and lower dotted lines inside each violin mark the upper and lower quartiles, respectively.

Experiment 1: Efficiency

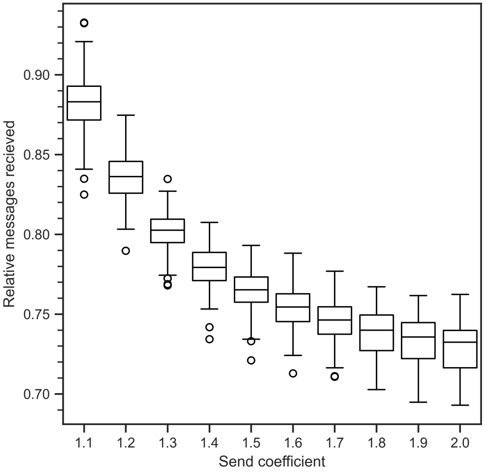

Figure: Message-passing efficiency. The send coefficient $\gamma = 1$ was used as a baseline for message-passing efficiency since it was found to be the maximum send coefficient that achieves perfect accuracy.

Experiment 1: Exploration I

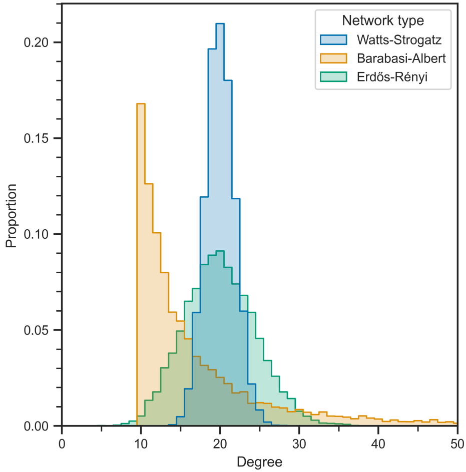

Figure: Contact network degree distributions. All vertices in random regular contact networks had a degree of 20, so the distribution was omitted to provide more visual space for the distributions of other contact networks.

Experiment 1: Exploration II

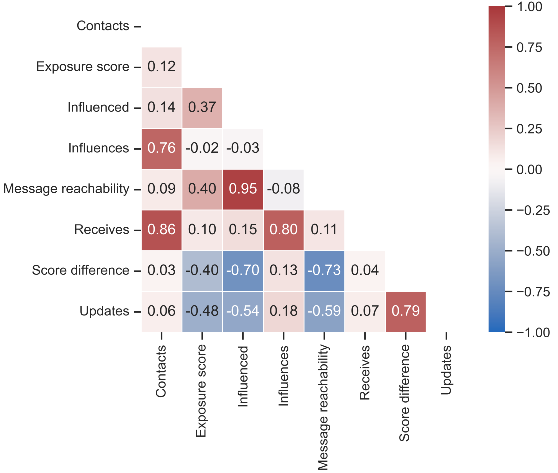

Figure: Correlation matrix of dataset attributes. Each cell is the Spearman rank partial correlation coefficient [17], controlling for the effect of the send coefficient. All coefficients are significant ($p < 0.01$), adjusting for multiple comparisons via the Holm–Bonferroni method [8].

Experiment 2: Benchmarking Hypothesis Testing

Experiment 3: Benchmarking I

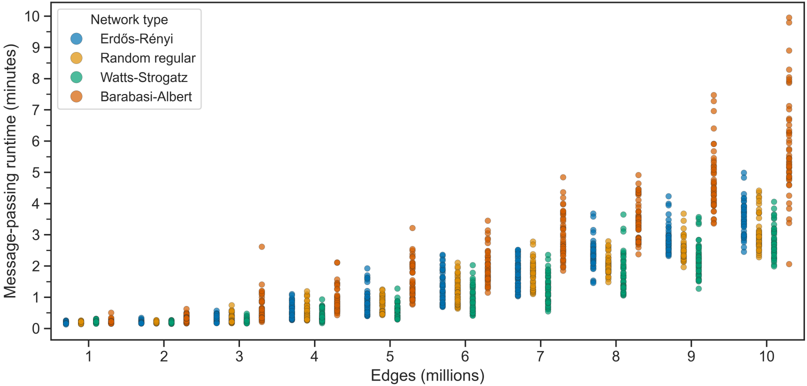

Figure: Message-passing runtimes.

Experiment 3: Benchmarking II

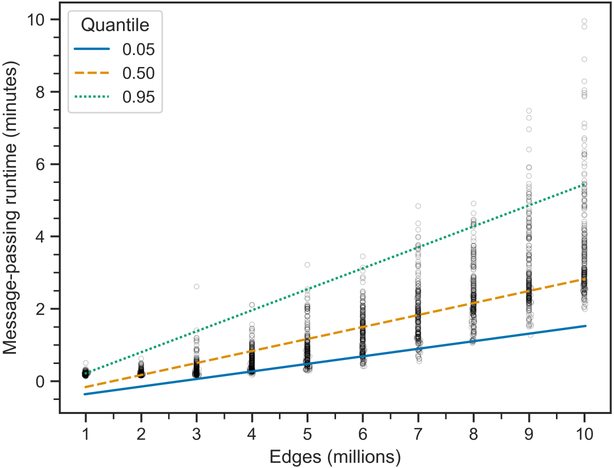

Figure: Message-passing runtimes with regression lines.

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [14].

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [14].
- Formally define the security and privacy characteristics of ShareTrace, using the framework proposed by Kuhn, Beck, and Strufe [10] to characterize the latter.

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [14].
- Formally define the security and privacy characteristics of ShareTrace, using the framework proposed by Kuhn, Beck, and Strufe [10] to characterize the latter.
- Conduct a simulation-based analysis of asynchronous risk propagation with COVI-AgentSim [6].

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [14].
- Formally define the security and privacy characteristics of ShareTrace, using the framework proposed by Kuhn, Beck, and Strufe [10] to characterize the latter.
- Conduct a simulation-based analysis of asynchronous risk propagation with COVI-AgentSim [6].
- Explore the utility and feasibility of integrating decentralized technologies [3, 9, 16, 19, 20] and self-soverign identity [13, 15] into the system design.

References I

[1] Erman Ayday, Youngjin Yoo, and Anisa Halimi. "ShareTrace: An iterative message passing algorithm for efficient and effective disease risk assessment on an interaction graph". In: *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2021. DOI: 10.1145/3459930.3469553.

[2] Erman Ayday et al. *ShareTrace: A smart privacy-preserving contact tracing solution by architectural design during an epidemic*. White paper. Case Western Reserve University, 2020.

[3] Juan Benet. *IPFS - content addressed, versioned, P2P file system*. 2014. arXiv: 1407.3561 [cs.NI].

[4] Mark Briers, Marcos Charalambides, and Chris Holmes. *Risk scoring calculation for the current NHSx contact tracing app*. 2020. arXiv: 2005.11057 [cs.CY].

[5] Renato Cherini et al. "Toward deep digital contact tracing: Opportunities and challenges". In: *IEEE Pervasive Computing* 22.4 (2023), pp. 15–25. DOI: 10.1109/mprv.2023.3320987.

[6] Prateek Gupta et al. *COVI-AgentSim: An agent-based model for evaluating methods of digital contact tracing*. 2020. arXiv: 2010.16004 [cs.CY].

[7] Prateek Gupta et al. "Proactive contact tracing". In: *PLOS Digital Health* 2.3 (2023), pp. 1–19. DOI: 10.1371/journal.pdig.0000199.

References II

[8] Sture Holm. "A simple sequentially rejective multiple test procedure". In: *Scandinavian Journal of Statistics* 6.2 (1979), pp. 65–70. URL: https://www.jstor.org/stable/4615733.

[9] Navin Keizer et al. "A survey on content retrieval on the decentralised web". In: *ACM Computing Surveys* 56.8 (2024). DOI: 10.1145/3649132.

[10] Christiane Kuhn, Martin Beck, and Thorsten Strufe. "Covid notions: Towards formal definitions—and documented understanding—of privacy goals and claimed protection in proximity-tracing services". In: *Online Social Networks and Media* 22 (2021). DOI: 10.1016/j.osnem.2021.100125.

[11] Cristina Menni et al. "Real-time tracking of self-reported symptoms to predict potential COVID-19". In: *Nature Medicine* 26.7 (2020). DOI: 10.1038/s41591-020-0916-2.

[12] Francisco Pozo-Martin et al. "Comparative effectiveness of contact tracing interventions in the context of the COVID-19 pandemic: A systematic review". In: *European Journal of Epidemiology* 38.3 (2023), pp. 243–266. DOI: 10.1007/s10654-023-00963-z.

[13] Alex Preukschat and Drummond Reed. *Self-sovereign identity: Decentralized digital identity and verifiable credentials*. Manning, 2021.

References III

[14] Rob Romijnders et al. "Protect your score: Contact-tracing with differential privacy guarantees". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 38.13 (2024), pp. 14829–14837. DOI: 10.1609/aaai.v38i13.29402.

[15] Frederico Schardong and Ricardo Custódio. "Self-sovereign identity: A systematic review, mapping and taxonomy". In: *Sensors* 22.15 (2022). DOI: 10.3390/s22155641.

[16] Ruizhe Shi et al. "A closer look into IPFS: Accessibility, content, and performance". In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8.2 (2024). DOI: 10.1145/3656015.

[17] Charles Spearman. "The proof and measurement of association between two things". In: *The American Journal of Psychology* 15.1 (1904), pp. 72–101. DOI: 10.2307/1412159.

[18] Ryan Tatton et al. "ShareTrace: Contact tracing with the actor model". In: *2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*. ©2022 IEEE. 2022, pp. 13–18. DOI: 10.1109/healthcom54947.2022.9982762.

[19] Dennis Trautwein et al. "Design and evaluation of IPFS: A storage layer for the decentralized web". In: *Proceedings of the ACM SIGCOMM 2022 Conference*. 2022, pp. 739–752. DOI: 10.1145/3544216.3544232.

References IV

[20] Carmela Troncoso et al. "Systematizing decentralization and privacy: Lessons from 15 years of research and deployments". In: *Proceedings on Privacy Enhancing Technologies* 2017.4 (2017), pp. 307–329. DOI: 10.1515/popets-2017-0056.