Third Paper Selection: Comparison of Candidates

November 24, 2025

1 Selected Papers for Detailed Study

1.1 Paper Links

Paper Title	Link	Shortform
Data Poisoning Attacks to Local Differential Privacy Protocols	USENIX PDF	Cao2020
Locally Differentially Private Protocols for Frequency Estimation	USENIX PDF	Wang2017

1.2 Publication Details

Shortform	Year	Venue	Rank	Google Scholar Category
Cao2020	2020	USENIX Security	#1	Computer Security & Cryptography
Wang2017	2017	USENIX Security	#1	Computer Security & Cryptography

2 Candidate Papers

2.1 Paper Links

Paper Title	Link	Shortform
Locally Differentially Private Heavy Hitter Identification	Semantic Scholar	Wang2021
Discrete Distribution Estimation under Local Privacy	Semantic Scholar	Kairouz2016
Further Study on Frequency Estimation under Local Differential Privacy	USENIX	Fang2025
Mitigating Data Poisoning Attacks to Local Differential Privacy	ACM DL	Li2025Mitig

2.2 Publication Details

Shortform	Year	Venue	Rank	Google Scholar Category
Wang2021	2021	IEEE TDSC	#6	Computer Security & Cryptography
Kairouz2016	2016	ICML/PMLR	#3	Artificial Intelligence
Fang2025	2025	USENIX Security	#1	Computer Security & Cryptography
Li2025Mitig	2025	ACM CCS	#5	Computer Security & Cryptography

3 Unreviewed Candidate Papers

3.1 Paper Links

Paper Title	Link	Shortform
RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response	Scopus	RAPPOR2014
Locally Differentially Private Frequency Estimation with Consistency	NDSS 2020 PDF	Wang2020

3.2 Publication Details

Shortform	Year	Venue	Rank	Google Scholar Category
RAPPOR2014	2014	ACM CCS	#5	Computer Security & Cryptography
Wang2020	2020	NDSS	#7	Computer Security & Cryptography

4 Analysis

4.1 Wang2021

4.1.1 Relevance to Cao2020

Introduces the PEM, which is mentioned in Cao2020 in the following sections:

Cao2020 Section	PEM Discussion
Section 2.2	Introduces PEM as state-of-the-art heavy hitter protocol with iterative prefix-based mechanism using OLH
Section 4.2	Data poisoning attacks (RPA, RIA, MGA) manipulate bits in each iteration to push attacker-chosen items into top-k

Continued on next page

Table 7 – *Continued from previous page*

Cao2020 Section	**PEM Discussion**
Section 5.3	MGA achieves 100% attack success with ∼5% fake users on multiple datasets
Section 6.2	Fake user detection via frequent itemset mining at each PEM iteration

4.2 Fang2025

4.2.1 Usefulness

- It's a very recent paper and introduces a latest LDP protocol called RWS
- To be filled
- To be filled

4.2.2 Relevance to Wang2017

Shows that OUE and OLH are not optimally parameterized, introducing improvements with RUE and RLH. The paper discusses OUE and OLH in the following sections:

Fang2025 Section	**Protocol**	**Discussion**
Section 3.2	OUE	Main definition section for Optimized Unary Encoding
Section 3.3	OLH	Main definition section for Optimized Local Hashing
Section 3.5	OUE & OLH	Summary comparing protocols; states OUE and OLH only achieve optimal MSE for large d
Section 4	OUE & OLH	Explains that OUE and OLH were optimized using approximate equations that need improvement
Section 4.1.1	OUE → RUE	Introduces Re-optimized Unary Encoding (RUE) built from OUE
Section 4.1.2	OLH → RLH	Introduces Re-optimized Local Hashing (RLH) built from OLH
Section 4.1.3	OUE vs RUE	Parameter discussion comparing OUE and RUE optimization approaches
Section 4.2	OLH → RLH	Addresses OLH's slow server-side computation and how RLH solves it

4.3 Kairouz2016

This paper is **not ideal for selection** as it is published in an Artificial Intelligence venue rather than a cryptography one

4.4 Li2025Mitig

4.4.1 Usefulness

- This paper directly references Cao2020
- To be filled
- To be filled

4.4.2 Relevance to Cao2020

Builds on top of MGA (the strongest attack) introduced in Cao2020. Li2025Mitig discusses MGA and proposes defenses in the following sections:

Li2025Mitig Section	**Discussion**
Section 3.2	Main formal description of MGA extended to different LDP protocols (OUE, OLH-User/Server, HST-User/Server, GRR)
Section 4.1	Theoretical analysis of fake user detection including expected squared error under MGA and chi-square error statistics
Section 5	Detection evaluation comparing methods such as Diffstats and FIAD against MGA attacks
Figures 6 & 8	Experimental results showing utility recovery and detection performance under MGA and MGA-A for various protocols