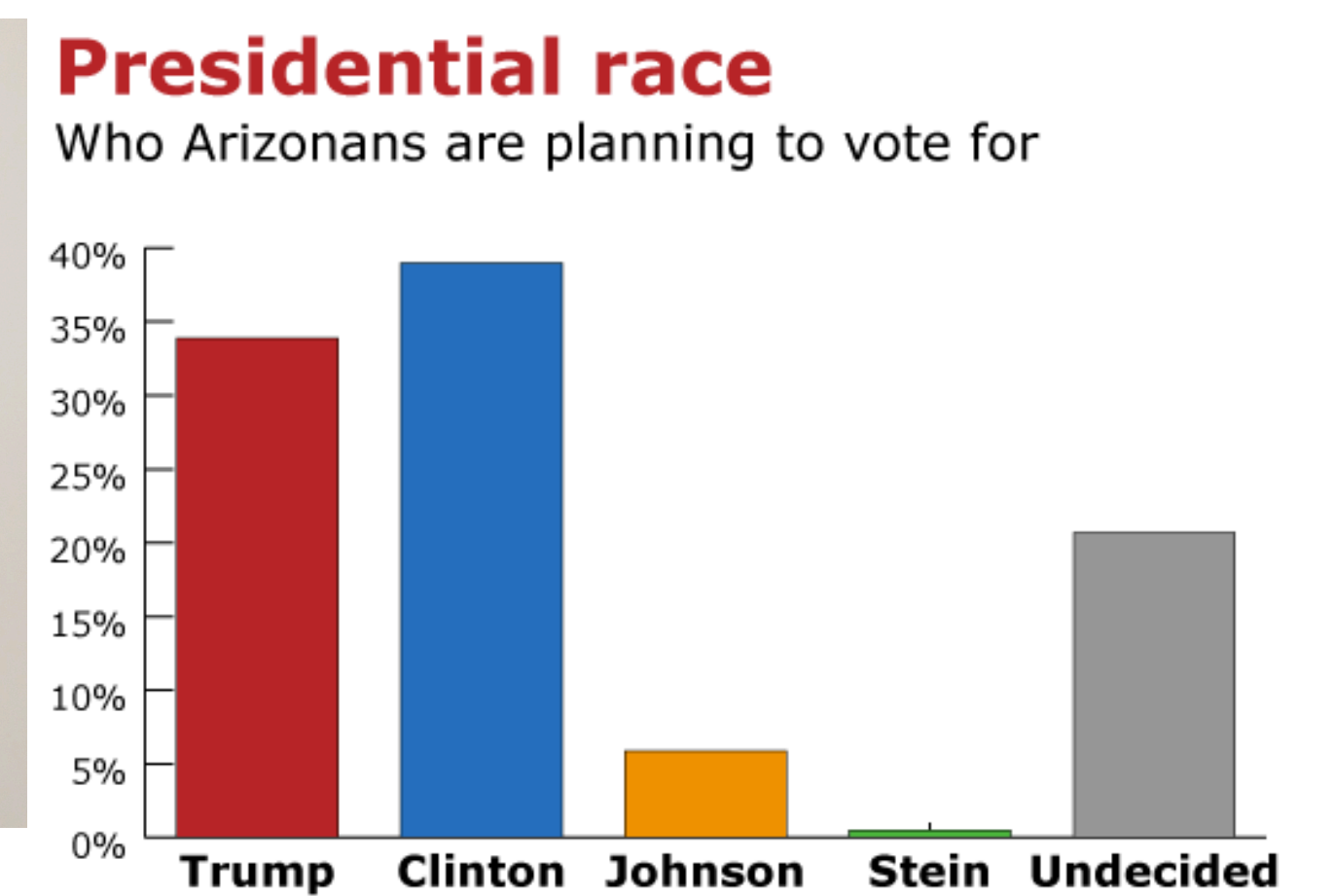


Differentially Private Heavy Hitter Problem

1. An embarrassing question
2. Toss a coin
3. With many people
4. Good estimation



⊖ true answer
 👍 false answer



Data from Arizona Republic/Morrison/Cronkite News poll from Oct. 10-15, 2016.

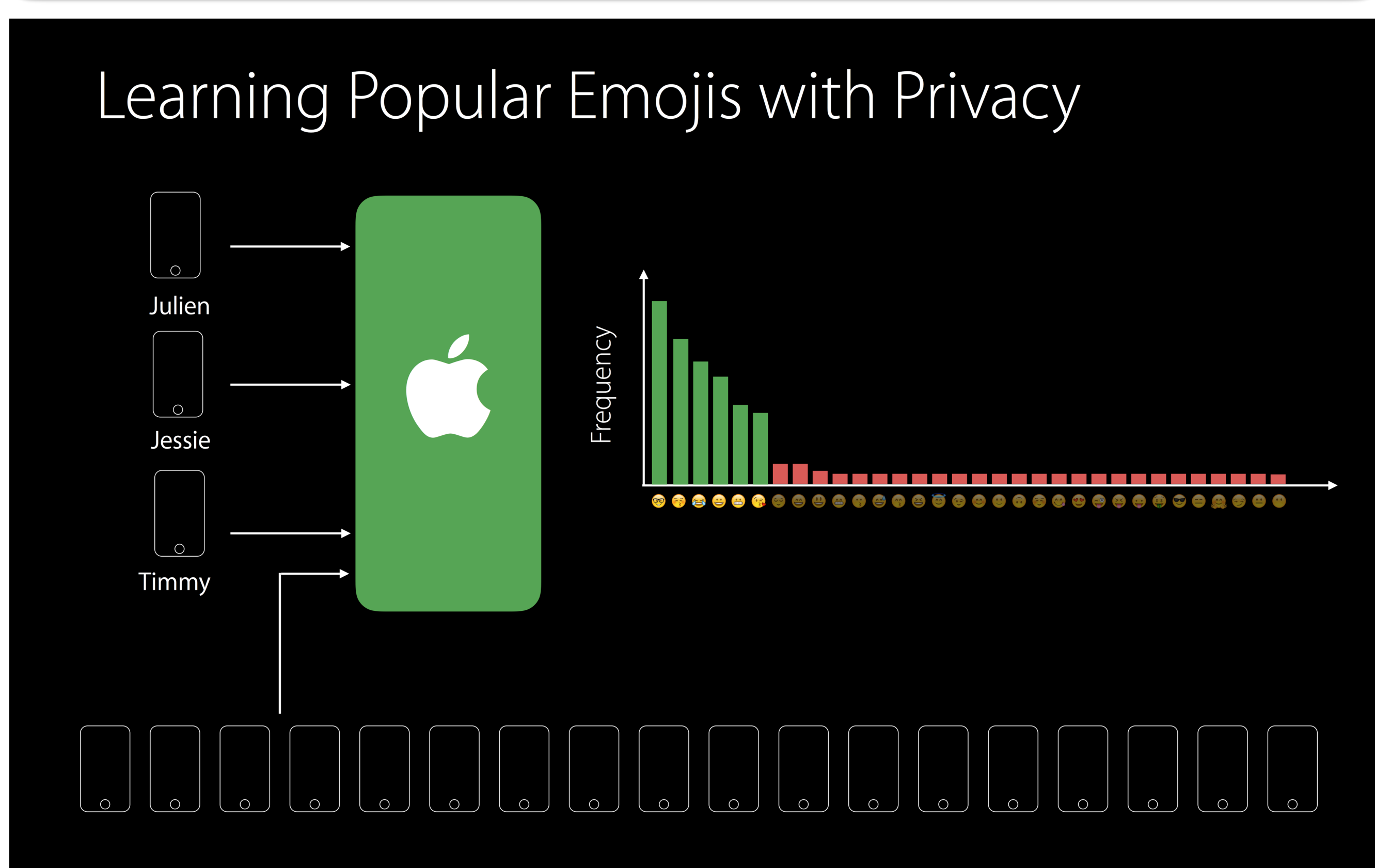
III. Proposed Method

I. Differential Privacy in the Local setting

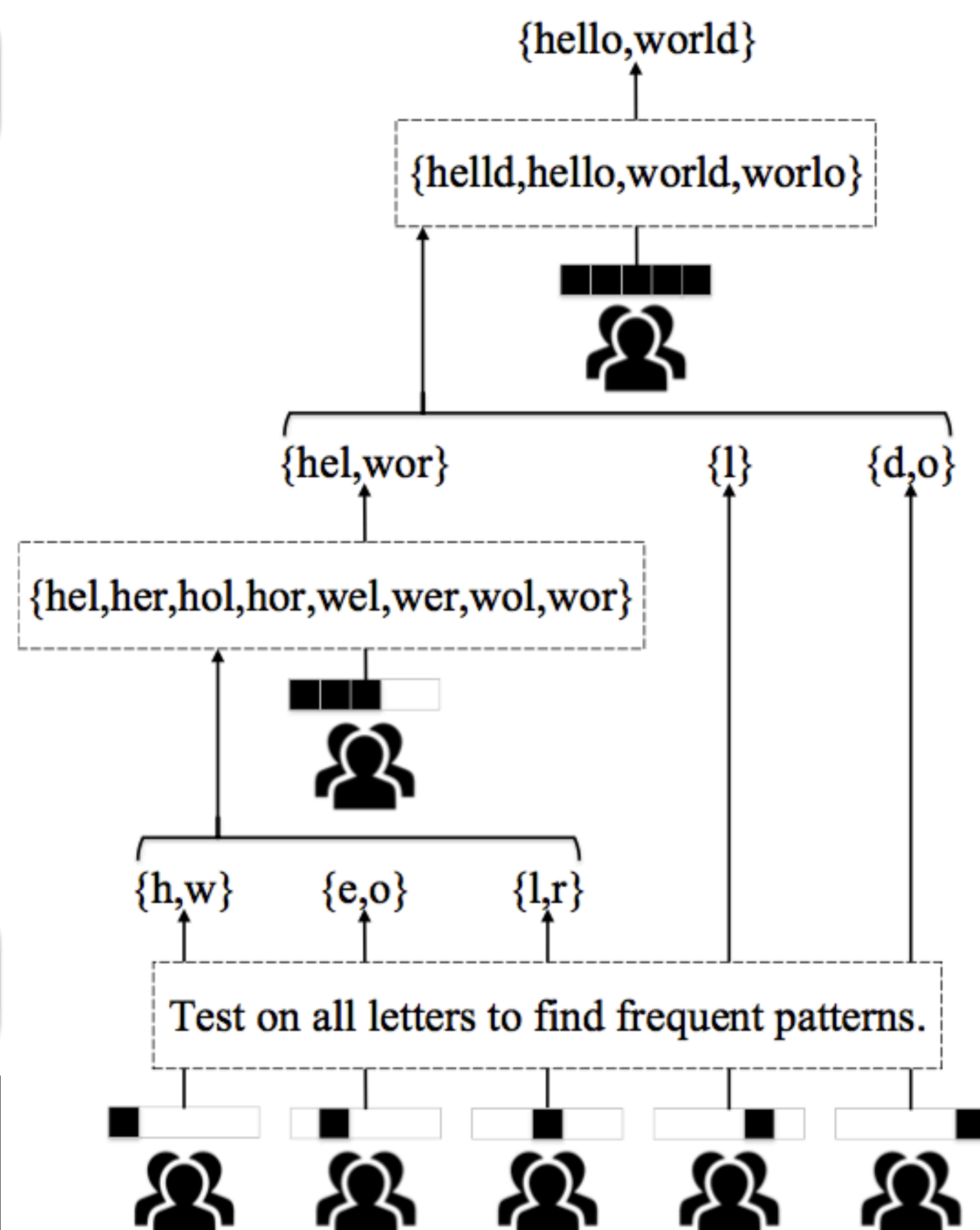
- Each user perturbs her own data before sending it to the aggregator



II. Frequency Estimator under LDP



- Good for small domain
- Inefficient and inaccurate for large domain -> focus of this research



- Combine the rest characters
- Combine three characters
- Analyze each character first

IV. Results

| ϵ | method | $n = 10^6$ | |
|------------|----------|----------------------|----------------------|
| | | $m = 64$ | $m = 256$ |
| 2 | HieProd | 19.8(1.2) ; 2.9(0.1) | 1.7(0.9) ; 0.4(0.2) |
| | SPM | 0.3(0.6) ; 0.1(0.1) | 0.0(0.0) ; 0.0(0.0) |
| | IterProd | 20.1(1.1) ; 2.9(0.1) | 2.1(0.7) ; 0.5(0.2) |
| | HASH | 0.0(0.0) ; 0.0(0.0) | 0.0(0.0) ; 0.0(0.0) |
| | SimHash | 16.0(1.2) ; 2.6(0.1) | 0.6(0.5) ; 0.1(0.1) |
| 4 | HieProd | 63.4(2.6) ; 5.1(0.1) | 24.9(0.7) ; 3.4(0.1) |
| | SPM | 11.8(1.0) ; 2.1(0.1) | 0.0(0.0) ; 0.0(0.0) |
| | IterProd | 62.6(1.6) ; 5.1(0.1) | 24.5(1.3) ; 3.3(0.1) |
| | HASH | 0.0(0.0) ; 0.0(0.0) | 0.0(0.0) ; 0.0(0.0) |
| | SimHash | 51.5(2.4) ; 4.7(0.1) | 21.8(1.3) ; 3.1(0.1) |