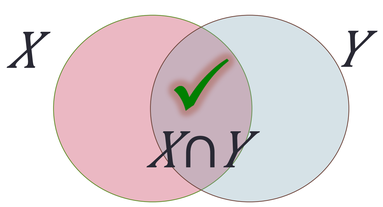
# Private Set Intersection

**Peter Rindal**, Mike Rosulek -- Hao Chen, Kim Laine
Oregon State University Microsoft Research

## ABSTRACT

Private set intersection (PSI) allows two parties (computers), who each hold a set of items $X,Y$, to compute the intersection of those sets without revealing anything about other items. Recent advances in PSI have significantly improved its performance, making PSI a practical alternative to insecure methods for computing intersections.

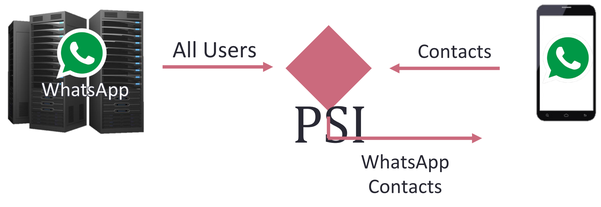



## OBJECTIVES

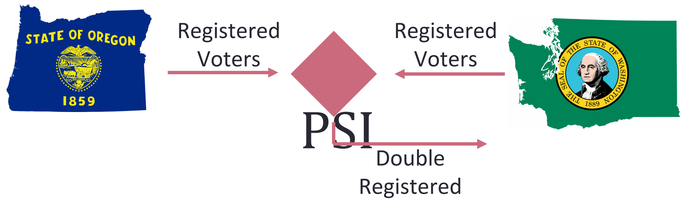
- Learn which items are common to both parties
- Hide items unique to one party
- Efficient to run on a laptop or cellphone
  - Computation
  - Communication
- Security:
  - Other party can not learn any additional information
  - Cheating behavior (hacking) does not help.

## APPLICATIONS

**Contact Discovery**




**Voter Registration**

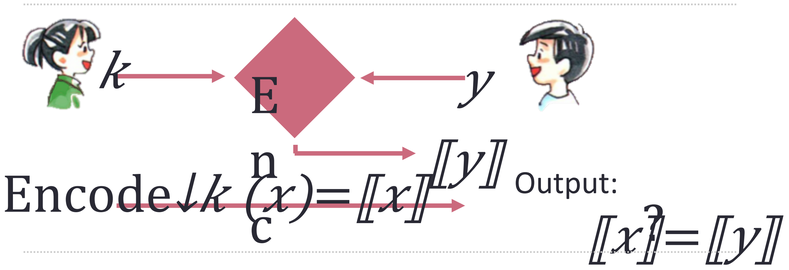



## METHODS

***Comparing two items:*** If Alice has $x$ and Bob has $y$, we need to check if

$$x=y \quad ?$$

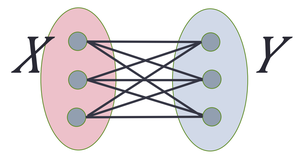
without revealing $x$ or $y$. First Alice picks a secret encoding key $k$.



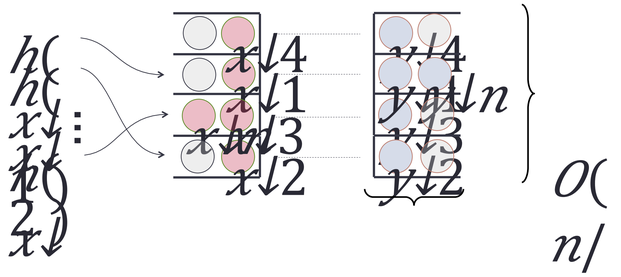

The encoding of $x$ must reveal nothing about $x$.

***Comparing Sets $X,Y$:*** When sets are of equal size, $n=|X|=|Y|$, we desire the work and communication between parties to be proportional to $n$.

- **Method 1:** Compare all pairs will result in $n^2$ comparisons. In practice we may have $n=1{,}000{,}000$ This would result in 1 trillion comparison!



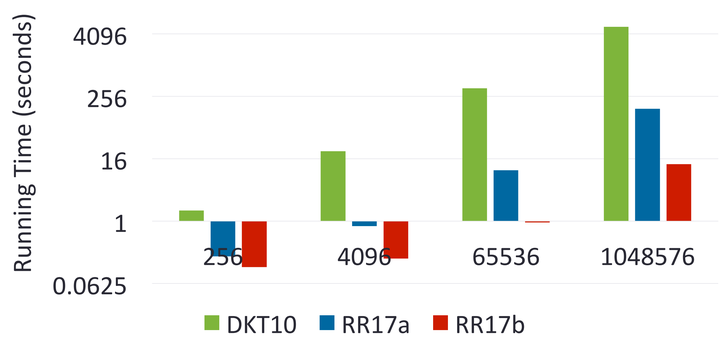

- **Method 2:** Use a hash table to reduce comparison.



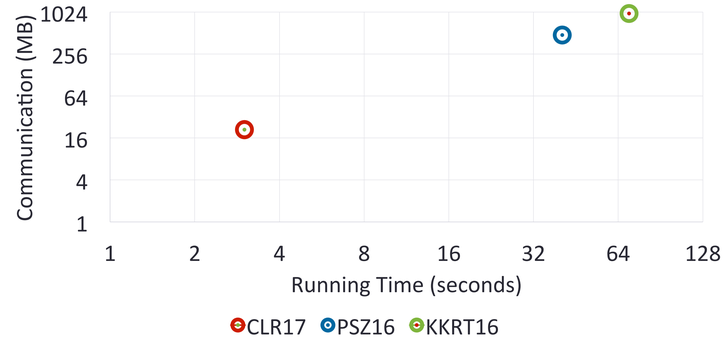

Compare each bin. Overall, $n \log n$ comparison need to be made. Can be further reduced to $n$ comparisons.

## RESULTS

***Balanced Set sizes:*** Rindal & Rosulek (RR17a,RR17b) improved on the state-of-the-art (DKT10) running time by 40x and later improved it further by 450x. **12 seconds to compare two sets of 1 million items.**




***Unbalanced set sizes:*** In many cases, one set is much larger than the other, e.g. Contact Discovery. Here Chen, Laine & Rindal (CLR17) improved the running time by 40x and communication by 25x. **Requiring 3 seconds and 20MB to compare 5000 items with 16 million items.**




## CONCLUSIONS

Recent advances in private set intersection has resulted in very efficient techniques for comparing sets. Combining this with a wide variety of applications ranging from secure messaging apps, voter registration, ad revenue tracking, and many more, it is expected that this technology will soon start impacting millions of people.