Distributed one-class learning

Ali Shahin Shamsabadi, Hamed Haddadi,
and Andrea Cavallaro

Published in: *IEEE International Conference on Image Processing (ICIP) 2018*

Centre for Intelligent Sensing
Queen Mary University of London

CIS centre for intelligent sensing

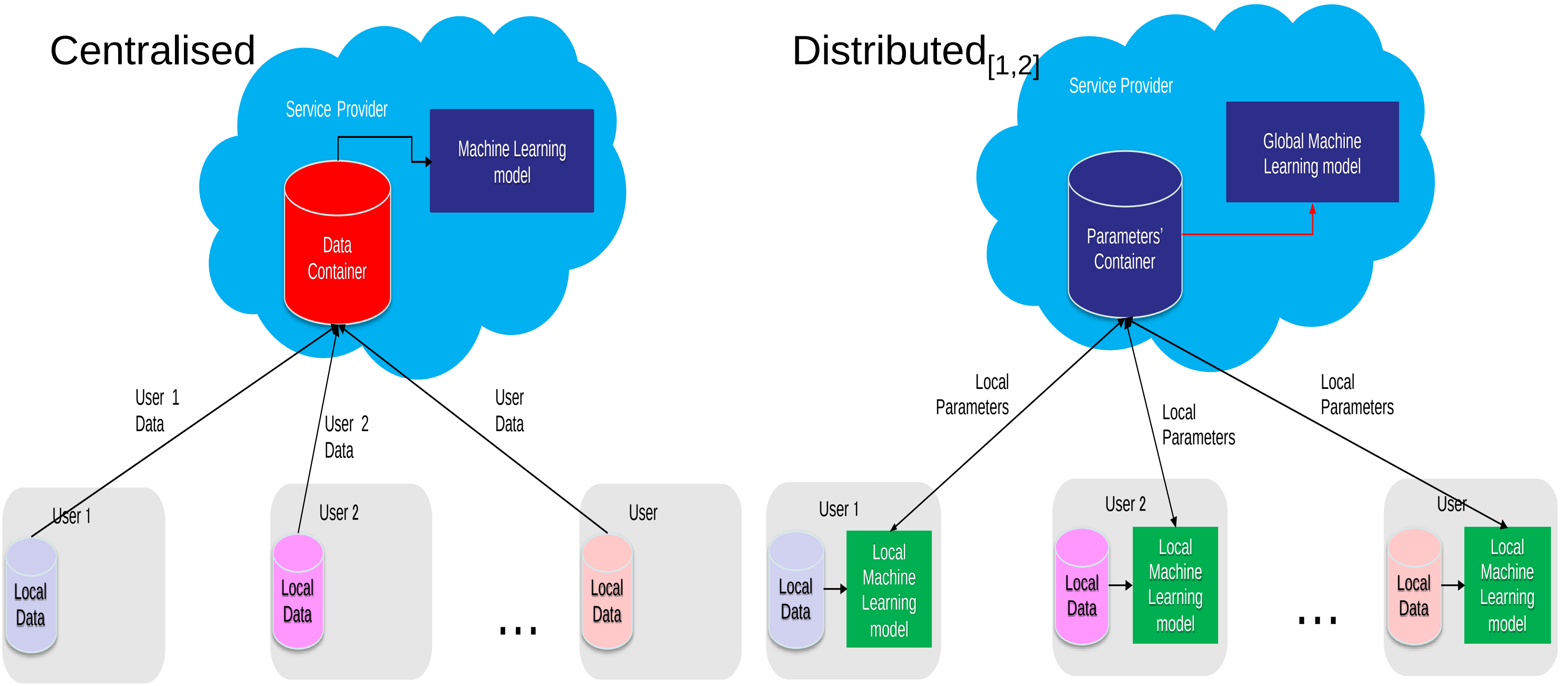
Queen Mary University of London

Outline

- Introduction to privacy in machine learning
- Centralised and distributed learning and their challenges
- Background in Autoencoder and One-Class classifier
- Machine learning solution for revenge pornography
- Proposed Distributed One-Class Learning
- Datasets, results and conclusion

Privacy in machine learning

- Training data, Parameters and Test data
- Training process with users' collaboration



CIS centre for intelligent sensing

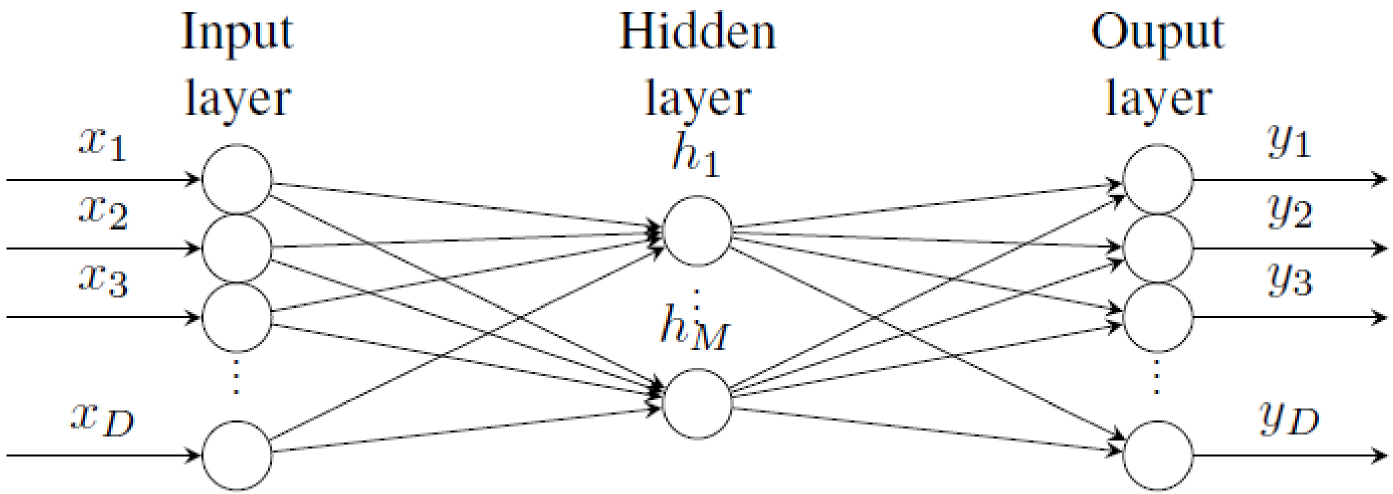
Queen Mary University of London

Challenges

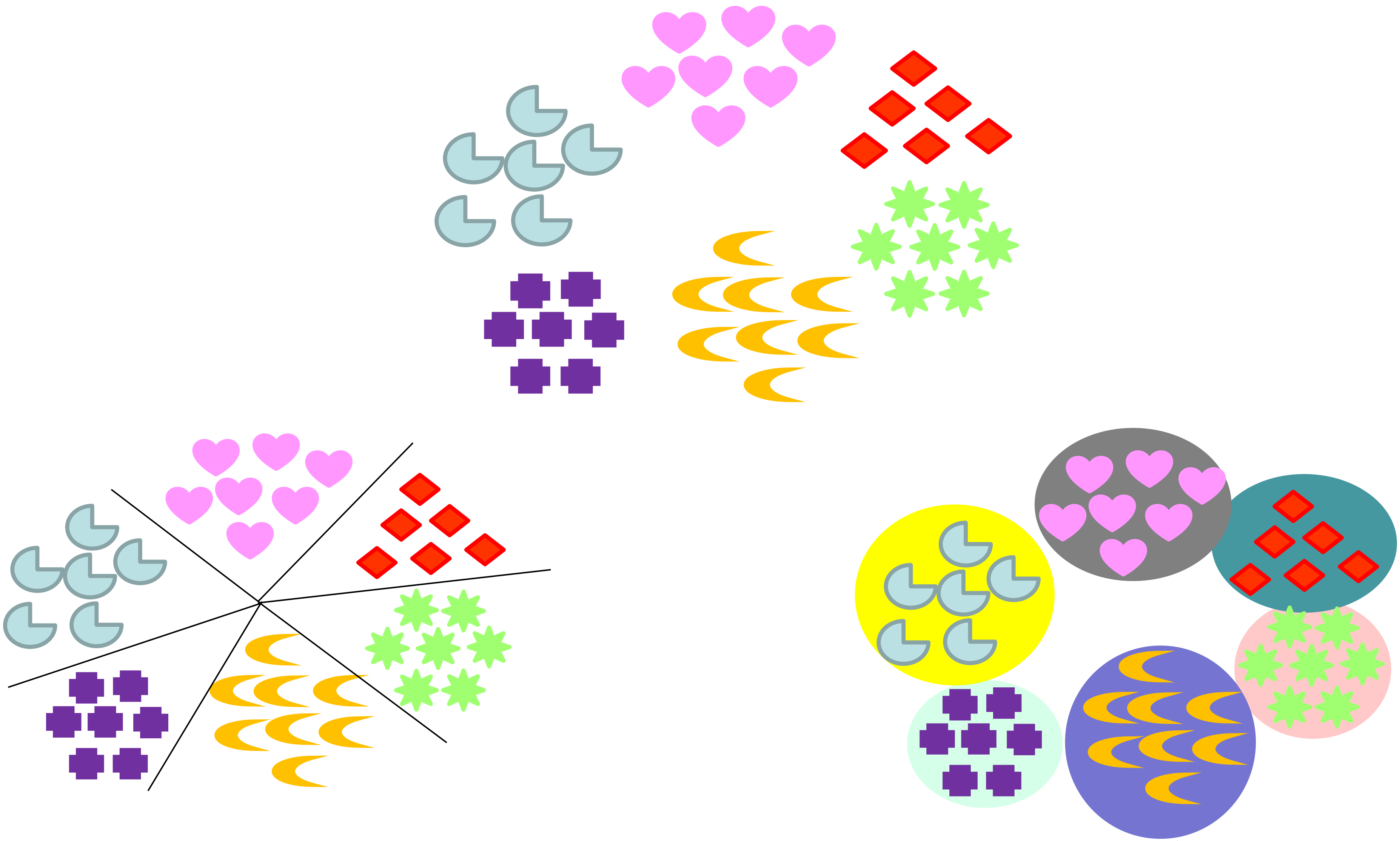
- Users and service provider share
  - Data
  - Parameters
- Training data of each user has different
  - Size
  - Number of classes
- Scalability
- Complex distribution of users' data (e.g. faces)

Background: Autoencoder[3]

- Encoder-Decoder neural network



Background: One-Class Classifier[4]



Online photo-sharing social media

- Revenge pornography[5]
  - Upload private images without consent

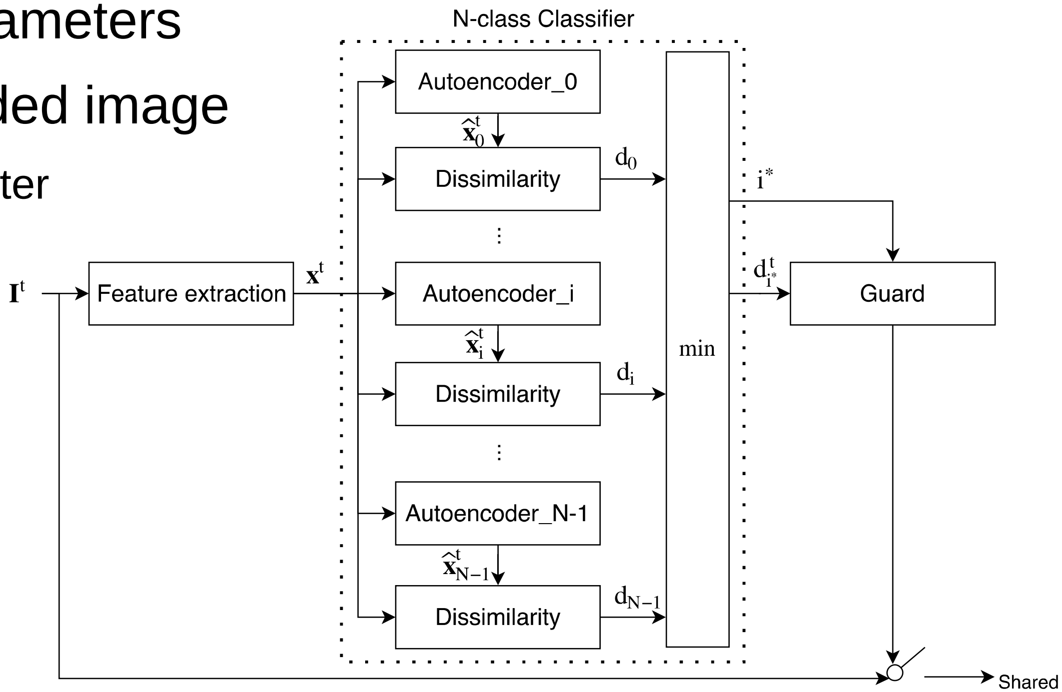
- ***How can prevent users from uploading privacy-sensitive images of other users?***

- Cloud-based Filter with users' collaboration
  - Share permission or block uploading images

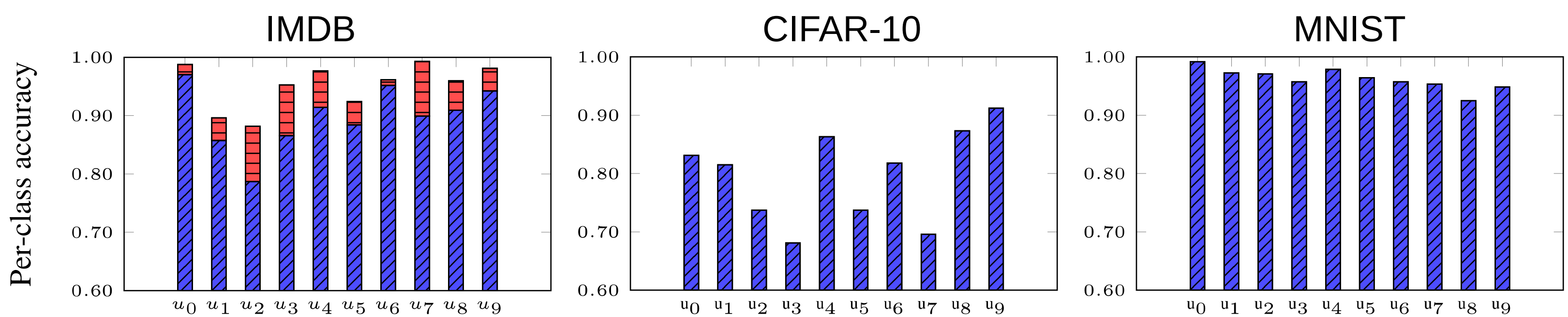
- Train blocking filter (N-class classifier)
  - Private-sensitive training data
    - → Not centralized learning
  - Parameters contain sensitive information, each user one class
    - → Not distributed learning

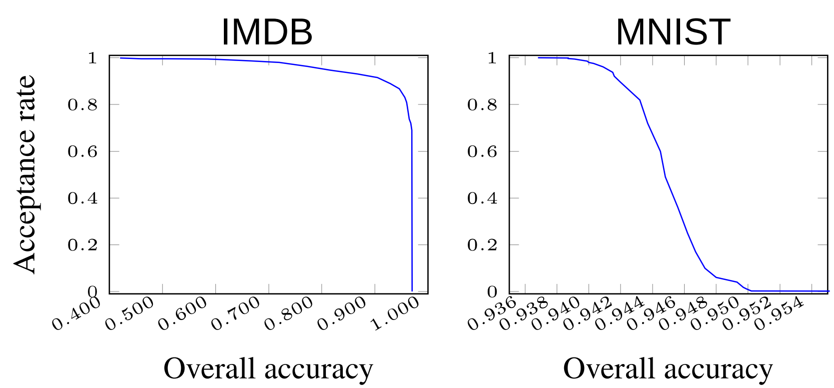
Distributed One-Class Learning

- N users with N private classes → N-class classifier
- Decompose N-class classifier to N **one-class classifier**
- Distribute N one-class classifiers (= Autoencoders)
- Train N one-class autoencoders locally by users **independently**
- Upload parameters
- New uploaded image
  - Feed to filter



Dataset & accuracy private/non-private images



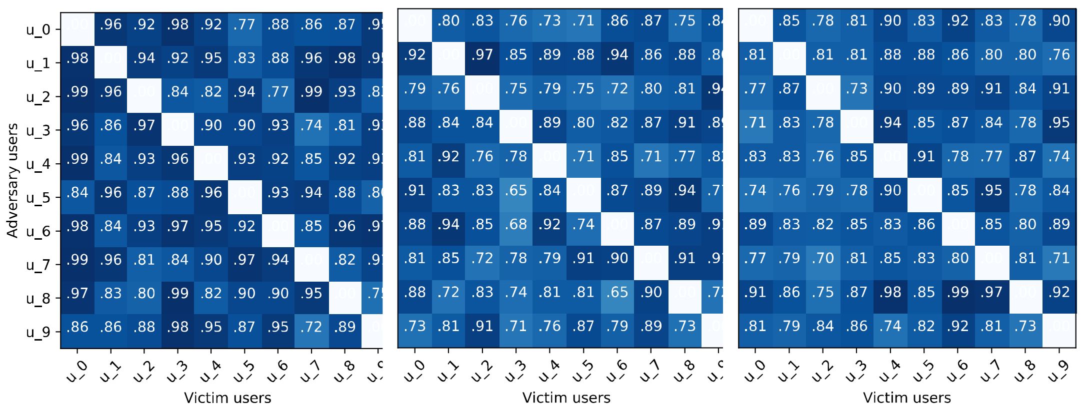


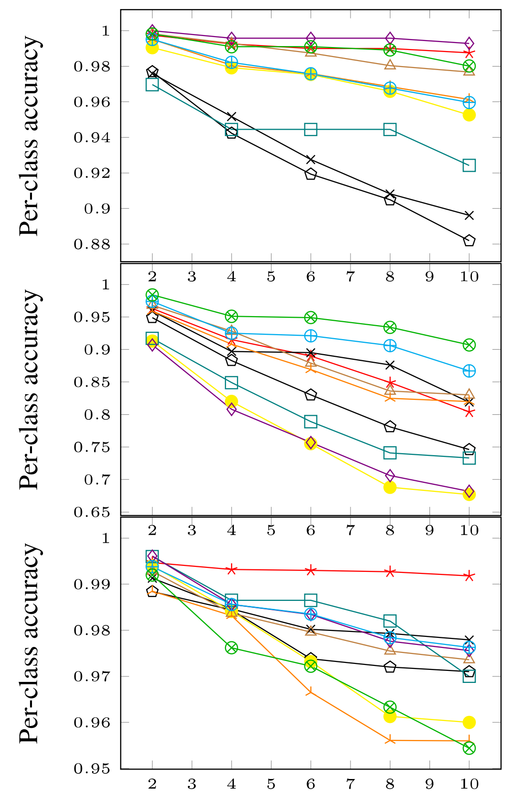
CIS centre for intelligent sensing

Queen Mary University of London

Threat & scalability

- Adversary user:
  - Access to data of victim user
  - Train one-class classifier with victim & adversary users data





- Scalability:
  - Impact of increasing number of user

Conclusion

DISTRIBUTED ONE-CLASS LEARNING

Ali Shahin Shamsabadi[⋆], *Hamed Haddadi*[†], *Andrea Cavallaro*[⋆]

[⋆]Queen Mary University of London,[†]Imperial College London

- Cloud-based Filter with users' collaboration
  - Each user capture property of their class independently
- Training phase
  - Not uploading users' data to cloud
  - Not sharing parameters among users
  - Each user data of one class
- Join new user at any time

Thank you!

CIS centre for intelligent sensing

Queen Mary University of London

References

- [1] R. Shokri and V. Shmatikov, “Privacy-preserving deep learning,” in Proceedings of the Computer and Communications Security (CCS), pp. 1310–1321, ACM, 2015.
- [2] H.B.McMahan,E.Moore,D.Ramage,S.Hampson,andB.A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proceedings of the International Conference on Articial Intelligence and Statistics (AISTATS), 2016.
- [3] A.S.Shamsabadi,M.Babaie-Zadeh,S.Z.Seyyedsalehi,H.R. Rabiee, and C. Jutten, “A new algorithm for training sparse au- toencoders,” in Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 2141–2145, IEEE, 2017.
- [4] D. M. J. Tax, One-class classification. PhD thesis, Delft University of Technology, 2001.
- [5] O. Solon, “Facebook asks users for nude photos in project to combat ’revenge porn’.” https: //www.theguardian.com/technology/2017/ nov/07/facebook-revenge-porn-nude-photos, 2017.
- [6] B. Hitaj, G. Ateniese, and F. Perez-Cruz, “Deep models under the gan: Information leakage from collaborative deep learning,” in Proceedings of the Computer and Communications Security (CCS), pp. 603–618, ACM, 2017.