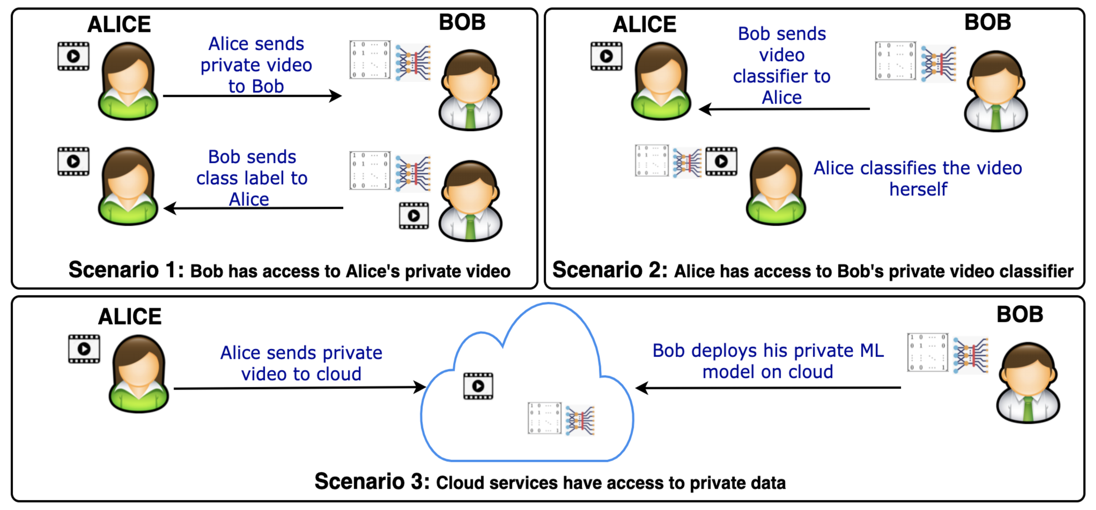
Privacy-Preserving Video Classification with Convolutional Neural Networks

Sikha Pentyala[1], Rafael Dowsley[2], Martine De Cock[1]
sikha@uw.edu, rafael@dowsley.net, mdecock@uw.edu
[1]School of Engineering and Technology, University of Washington Tacoma [2]Faculty of Information Technology, Monash University

Problem

Impactful applications of video classification
- Require **access to personal** videos
- **Invasive** in nature - security risk



Probable Scenarios

- *Alice* sends video to *Bob*
  - Privacy leak : Misuse of video
  - Resource limitation
- *Alice* classifies video herself
  - Privacy leak : Bob's proprietary model
  - Privacy leak : Bob's private training data
  - Evasion attacks
  - Resource limitation
- *Alice* and *Bob* outsource computation to cloud based infrastructure
  - No resource limitation
  - Privacy still a concern

Existing Solutions

- **No formal guarantee** of privacy
- Require **special secure hardware**

Contributions

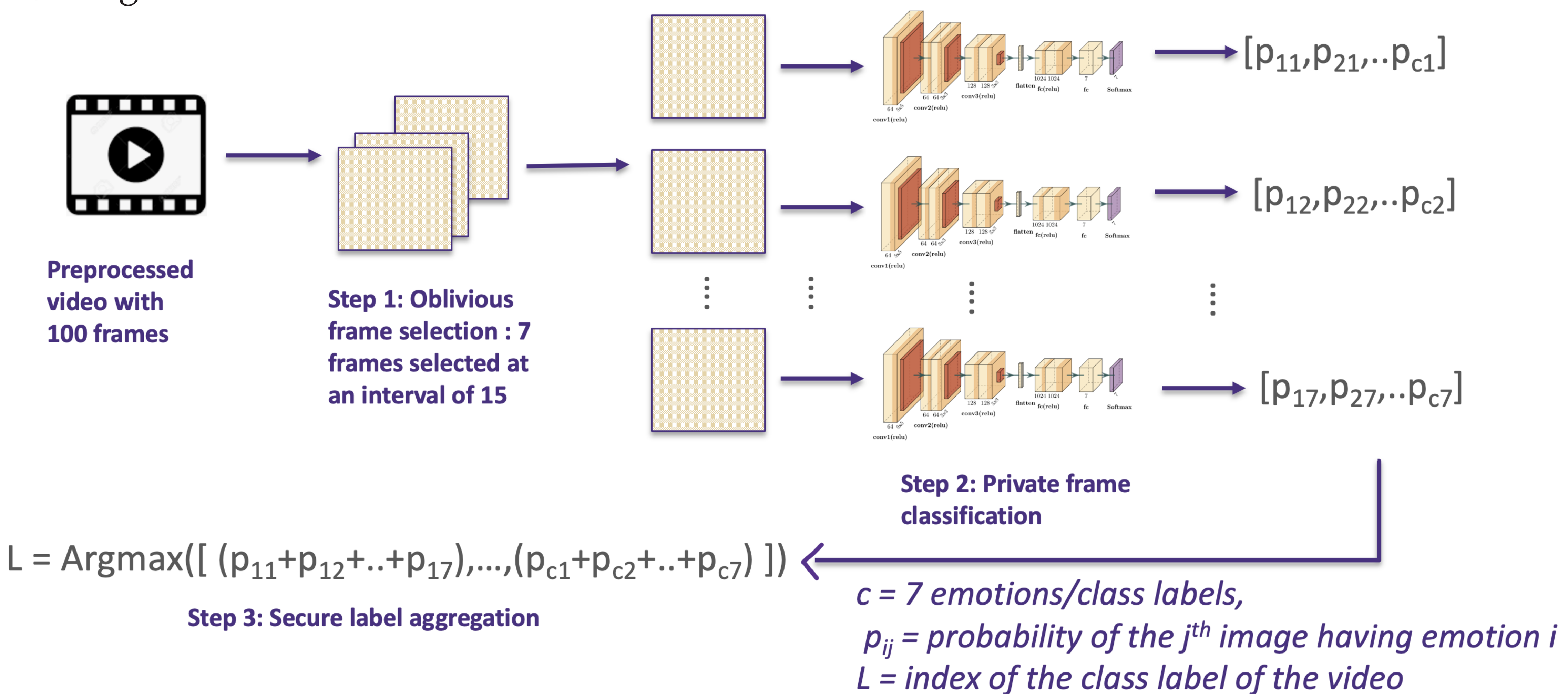
Our main novel contributions are:

1. An MPC protocol for selecting frames in an oblivious manner, such that the video owner remains unaware which frames from the video were selected for classification by the model owner.
2. An MPC protocol for privacy-preserving aggregation of the labels inferred for the individual frames.
3. An evaluation of our secure video classification pipeline in an application for human emotion detection from video on the RAVDESS dataset.

SCHOOL OF ENGINEERING & TECHNOLOGY UNIVERSITY of WASHINGTON TACOMA MONASH University

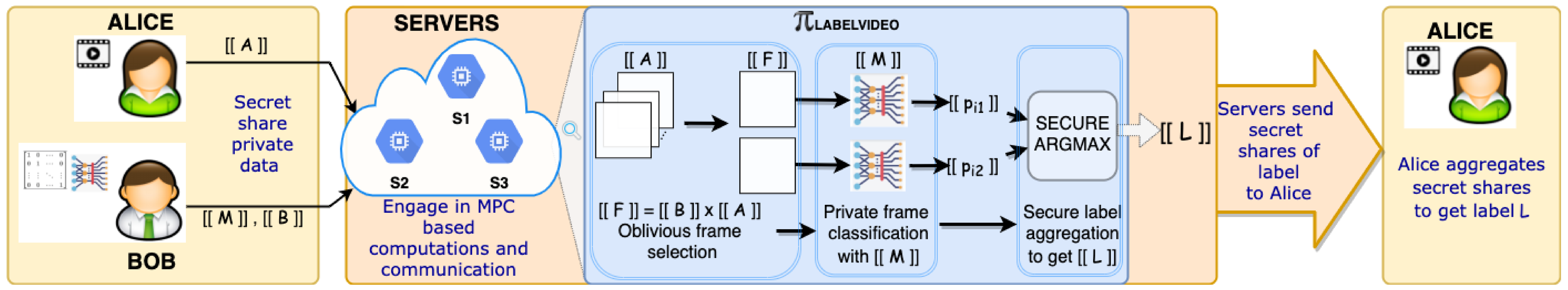
Method

- We classify a video based on the **single-frame method**, i.e. by aggregating predictions across single frames/images.



- We use a field of cryptology – **Secure Multiparty Computation (MPC)** – that allows two or more servers to jointly compute a specified output (the class label of the video) from their private information (the model and the video) in a distributed way, without revealing the private information to each other.
- Overview of steps to classify a video while preserving privacy:

1. *Alice* preprocesses the video on her end and generates a 4D tensor. *Bob* pretrains a 2D-CNN with 1.5 million parameters to classify 'frames'(images). *Bob* also generates the frame selection matrix with one-hot encoded entries of the frame numbers in the video he wants to select.
2. *Alice* **secret shares** her video as $[\![A]\!]$. *Bob* **secret shares** his model as $[\![M]\!]$ and frame selection matrix as $[\![B]\!]$. In this, $[[x]]$ represents the secret shares of private data ("secret") x.
3. The computations are carried out as per the privacy-preserving video classification pipeline [2] shown below.



- We evaluate our approach for **detecting emotions of a person in a video** - Preventing exposing emotions of a person, most private to oneself, and preventing compromising the security of video classification parameters.

The servers compute over data that they can not see.

Results

Dataset: 3-5 second videos of RAVDESS dataset with 120-150 frames, containing 7 emotions

Implementation: in MP-SPDZ [1] with mixed circuits and computations over integers modulo 64 with 16 threads.

Evaluation: Evaluated on F32s V2 Azure virtual machine - 32 cores, 64 GB RAM, and network bandwidth of upto 14 Gb/s. Evaluated the pipeline for different security settings.

		Time (sec)	Comm (GB)
Passive	2PC	302.24	374.28
	3PC	**8.69**	0.28
Active	2PC	6576.27	5492.38
	3PC	27.61	2.29
	4PC	**11.67**	0.57

Results show avg time to privately detect emotion computed over a set of 10 videos with 7-10 frames. Avg communication measured per party.

Accuracy of 56.8% on a held-out test set in line with state-of-the-art results.

Azure cloud credits donated by Microsoft

Conclusion

1. **First** baseline end-to-end privacy preserving video classification pipeline.
2. Feasible privacy preserving video classification with **state-of-the-art accuracy** for emotion detection in a RAVDESS video with **no privacy leakage** (mathematically provable!) and **no special hardware**.

Future Directions

1. Use of machine learning for intelligent frame selection.
2. Develop MPC protocols for state-of-the-art techniques in video classification in-the-clear.

References

[1] Keller, M. MP-SPDZ: A Versatile Framework for Multi-Party Computation
[2] Pentyala, S., De Cock, M., Dowsley, R. Privacy-Preserving Video Classification with CNNs, ICML2021