Privacy-preserving voice anti-spoofing using secure multi-party computation

Oubaïda Chouchane[1], *Baptiste Brossier*[1], *Jorge Esteban Gamboa Gamboa*[1], *Thomas Lardy*[1], *Hemlata Tak*[2], *Orhan Ermis*[1], *Madhu Kamble*[1], *Jose Patino*[2], *Nicholas Evans*[2], *Melek Önen*[1], *Massimiliano Todisco*[1]

[1]Security protocols and applied cryptography, EURECOM, France
[2]Audio, security and privacy, EURECOM, France
{name.surname}@eurecom.fr

Abstract

In recent years the automatic speaker verification (ASV) community has grappled with vulnerabilities to spoofing attacks whereby fraudsters masquerade as enrolled subjects to provoke illegitimate accepts. Countermeasures have hence been developed to protect ASV systems from such attacks. Given that recordings of speech contain potentially sensitive information, any system operating upon them, including spoofing countermeasures, must have provisions for privacy preservation. While privacy enhancing technologies such as Homomorphic Encryption or Secure Multi-Party Computation (MPC) are effective in preserving privacy, these tend to impact upon computational capacity and computational precision, while no available spoofing countermeasures preserve privacy. This paper reports the first solution based upon the combination of shallow neural networks with secure MPC. Experiments performed using the ASVspoof 2019 logical access database show that the proposed solution is not only computationally efficient, but that it also improves upon the performance of the ASVspoof baseline countermeasure, all while preserving privacy.

Index Terms: Secure multi-party computation, privacy-preservation, automatic speaker verification, spoofing

1. Introduction

Automatic speaker verification (ASV) technology provides a low-cost and reliable biometric solution to person authentication [1]. While deployed in a growing range of practical applications such as for online banking and call centers, ASV systems are susceptible to spoofing attacks [2, 3], referred to as presentation attacks according to the ISO/IEC 30107-1:2016 standard[1]. Spoofing attacks are launched by fraudsters who masquerade as other enrolled subjects to gain illegitimate access to resources protected by biometric recognition.

Spoofing countermeasures have been developed to help protect ASV systems from such attacks. These take the form of auxiliary systems which operate upon speech signals in a similar fashion to ASV systems in order detect to deflect attacks. In many use case scenarios, both ASV and spoofing countermeasure systems are implemented as remote, cloud-based services, requiring the transmission of speech data, potentially involving insecure infrastructure. Since speech signals can contain a wealth of sensitive information, such as the speaker's age, gender, ethnicity, emotional and health condition [4], privacy preservation solutions are needed to prevent eavesdroppers from intercepting and exploiting speech data for nefarious purposes. Adequate safeguards are now mandated by privacy law worldwide, e.g. the well-known General Data Protection Regulation (GDPR) in Europe which contains specific clauses relevant to biometric data such as speech [5].

From the user perspective, privacy concerns can be averted by processing speech data locally so that speech data need never be shared with an online service provider. This solution, though, requires the service provider's models to be distributed to the user device. Models potentially stem from the collection of massive quantities of proprietary data and result from tremendous research and development effort. With service providers being reluctant to put their intellectual property in jeopardy, this solution is impractical. The solution is then some form of privacy-enhancing technology, involving advanced cryptographic primitives such as Homomorphic Encryption or secure Multi-Party Computation (MPC). These techniques support computation upon encrypted data without the need for a user to share their speech data or a service provider to share their model. Both HE and MPC have been applied successfully to support privacy-preserving computation upon speech data, e.g. for privacy-preserving speaker verification [6, 7]. The literature also shows some efforts to bring privacy preservation to anti-spoofing for other biometric characteristics [8, 9].

In this paper, we present the key challenges in privacy-preserving voice biometric anti-spoofing and the first solution. We propose PRIVASP, an efficient shallow neural network architecture combined with secure MPC and consider two different scenarios. Since our focus is upon privacy preservation for anti-spoofing and not ASV, for simplicity we assume that the latter is performed on the user device. Nonetheless, our solution is applicable to scenarios in which ASV is also performed remotely. In the first scenario, anti-spoofing is performed by an independent service provider. The anti-spoofing service provider outsources its model in clear text form to a cloud service provider. In the second scenario, only a *protected* model is shared with the cloud service provider. Speech data is protected in both cases. The difference between these two scenarios is hence the level of trust granted to the cloud service provider and hence the implied level of privacy preservation with regards to the model.

We have evaluated our system on the ASVspoof 2019 Logical Access database. Experiments show that the new approach to privacy-preserving anti-spoofing even performs better than the ASVspoof 2019 baseline system and is also computationally efficient. To the best of our knowledge, it is the first reported solution to privacy-preserving voice biometric anti-spoofing.

2. Secure computation

In this section, we introduce secure MPC and further overview existing privacy-preserving solutions for neural networks in general and ASV, in particular.

[1]https://www.iso.org/standard/53227.html

2.1. Secure multi-party computation

Secure Multi-Party Computation (MPC) was firstly proposed by Yao in 1982 as a two-party computation (2PC) solution to the Millionaires' problem [10].

MPC protocols enable multiple parties to jointly and securely compute a function f over their inputs and reveal nothing but the output of the function. Existing MPC solutions are based on Yao's garbled circuits [11] or secret sharing (additive or Boolean) [12, 13]. In our work, we use a 2PC protocol that makes use of additive secret sharing. More specifically, private input data (secret s) is split into two pieces (two secret shares), namely $\langle s \rangle_1$ and $\langle s \rangle_2$ and distributed among two non-colluding parties (P_0 and P_1) (see Figure 1). One of the shares is chosen randomly, e.g. $\langle s \rangle_1$, and the second share is: $\langle s \rangle_2 = s - \langle s \rangle_1$. The secret shares on their own reveal no information about the original value s but together they recover it perfectly ($s = \langle s \rangle_1 + \langle s \rangle_2$). Each party later performs computations over their shares in order to finally obtain their output. While the addition operation is directly translated to the addition of the shares, the multiplication operation requires the storage of additional coefficients and the interaction among the two parties (see [12]).

MPC protocols usually consider two types of adversaries who can corrupt a subset of the contributing parties [14]: (i) passive (or semi-honest/honest-but-curious) adversaries strictly follow the protocol specification and are curious to extract information of the other parties' inputs; (ii) active adversaries who can arbitrarily deviate from the protocol.

2.2. Privacy preserving Neural Network solutions

There exist a large number of solutions proposed to build privacy-preserving Neural Networks (see [15] for a survey). Most of these solutions are client-server based and consider a scenario whereby the client protects its input from the server (i.e., model provider) and the model provider protects its model parameters from the client. These approaches can be classified into four main categories: (i) MPC-based solutions such as [16]; (ii) Homomorphic Encryption (HE)-based techniques, like [17, 18]; (iii) Hybrid solutions such as [19] and [17] that combine the use of MPC and HE; and, (iv) Trusted Execution Environment (TEE)-based solutions like in [20] that perform the operations in an isolated environment [21], e.g. Intel Software Guard Extensions (SGX).

More recent studies focused on the design of privacy-preserving speech processing solutions based on Neural Networks. The authors in [22] propose an approach based on 2PC and TEEs. On the other hand, Nautsch et al. [23, 24] use HE and MPC to protect the input biometric templates for speaker recognition. While these solutions aim to guarantee user privacy, they only consider the problem of speaker/speech recognition and do not focus on their security against spoofing. In this work, we propose the first privacy-preserving voice anti-spoofing scheme that is based on a customized shallow Neural Network that is compatible with MPC. This solution is presented in the next section.

3. PRIVASP

In this section, we provide a description of PRIVASP by first introducing the actual environment and further introducing two versions of PRIVASP. The main difference between the two versions is the level of trust anti-spoofing services gives to the cloud servers with respect to access to the model's parameters.

3.1. Environment

We consider a scenario whereby an individual/user tries to authenticate himself/herself to an ASV system, e.g. smart speaker such as Google Home, using a voice command. During the verification process of the identity of the user, the ASV system needs to check if the input data, i.e. the user's voice, is spoofed or not. With this aim, the latter exploits an anti-spoofing service to detect whether the system is under spoofing attacks or not. This anti-spoofing service owns an anti-spoofing Neural Network model that it outsources to some cloud service providers. In this work, the cloud service provider is represented by two servers.

In this scenario, since the ASV system is using an external anti-spoofing service to detect spoofing attacks, the input to this service should remain confidential against the service and the cloud servers. On the other hand, the anti-spoofing service does not wish to reveal the anti-spoofing model to its customers (the ASV system here) as this can be considered as an asset and hence should remain confidential. Additionally, the cloud servers may also be considered as potential adversaries against the privacy of the model. We propose to use multi-party computation to address these privacy requirements and consider all parties involved in computation as honest-but-curious. The actual solution is described in the next section.

3.2. Proposed solution

In this section, we describe our new solution PRIVASP that is based on the use of 2PC.

Current ASV systems implementing spoofing countermeasures are highly accurate and efficient. Yet, directly applying MPC to these systems is challenging due to the complexity of their underlying building blocks. For example, solutions in [25, 26] consider deep neural networks with a non-negligible number of non-linear operations which would result in a significant increase in the inference time with MPC. Additionally, these systems work with real numbers whereas MPC only supports integers; The conversion of real numbers into integers causes an information loss problem and hence decreases the accuracy of the system.

Therefore, we propose to design a spoofing countermeasure from scratch in order for it to be compatible with MPC. The new solution PRIVASP builds a shallow neural network with one hidden linear layer and one ReLU activation layer and truncates the parameters. Thanks to this new architecture, MPC can easily be integrated. In section 4.2, we show that the accuracy of PRIVASP remains efficient.

In the following two sections we explain how MPC is used in PRIVASP and suggest two versions of it as illustrated in Figure 1. In both versions, the input is secretly shared among the two cloud servers. On the other hand, while in the first version, the anti-spoofing service fully trusts the cloud and sends the model parameters in clear, in the second version, in addition to the input, the model parameters are also secretly shared. The first version is represented in red in Figure 1, where the same model is shared between the servers, and the second version is represented in blue in which the model is secretly shared.

3.2.1. Scenario 1: PRIVASP with Model privacy against the client

The client generates two additive secret shares of the input, i.e. the user's voice matrix, and sends them to the two non-colluding servers. Both servers have the same model M and the additive secret shares and perform the classification task, i.e. spoofing

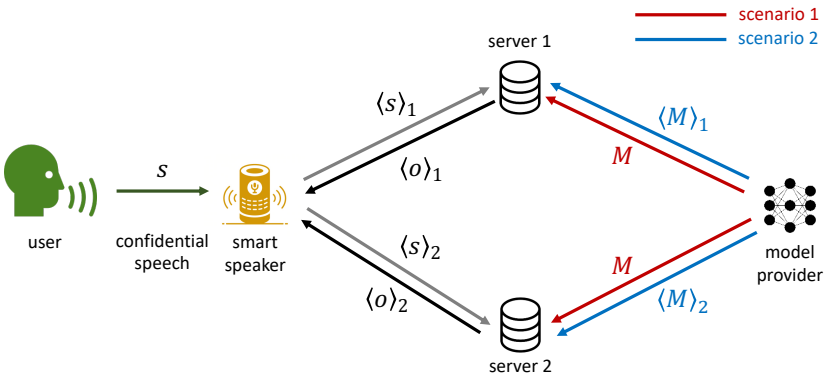

Figure 1: *Scenario 1: PRIVASP with Model privacy against the client (red arrows). Scenario 2: PRIVASP with model privacy against both the client and the cloud servers (blue arrows).*

detection over the shares. The trained anti-spoofing model is kept hidden from the client and never leaves the servers' side.

As previously mentioned, in this first version of PRIVASP, the model is revealed to the cloud servers and this may suffer from intellectual property problems. Therefore, in the next section, we propose a second version of PRIVASP which ensures model privacy against the cloud servers.

3.2.2. Scenario 2: PRIVASP with model privacy against both the client and the cloud servers

The basic idea behind this second version is to use the concept of 2PC, where each server receives one secret share of the client's data and the model provider's model and invoke a 2PC protocol to securely compute the spoofing detection without gaining any knowledge about the inputs. The trained model is kept hidden from the two servers and the client, and the model's input and output are only revealed to the client.

4. Experimental setup

This section describes the ASVspoof 2019 LA database, evaluation metrics, baselines of the challenge and competing, state-of-the-art countermeasure and PRIVASP implementation details. Note that all the countermeasure systems concerned in the experiments are single systems, which means that they are not the result of the fusion of many individual systems.

4.1. ASVspoof 2019 LA database, baselines and evaluation metrics

Experiments were performed using the publicly available ASVspoof 2019 logical access (LA) database [27]. The LA subset is a collection of bona fide utterances and 6 different spoofing attacks for the training and development partitions and 13 unseen spoofing attacks for the evaluation partition. The attacks were generated using state-of-the-art text-to-speech (TTS) and voice conversion (VC) techniques. There were two different evaluation metrics for the ASVspoof 2019 Challenge. The primary metric is the *tandem detection cost function* (t-DCF) [28] and Equal Error Rate (EER) [29] is kept as the secondary metric.

Two different baseline systems were provided for the ASVspoof 2019 Challenge edition. Baseline B01 consists of constant Q cepstral coefficients (CQCCs) feature set along with 512 Gaussian Mixture Model (GMM) as a back-end classifier. [30, 31]. The number of bins per octave is kept to 96 and the re-sampling period is set to 16 within the bandwidth range of 15 Hz to 8 kHz. The 30-dimensional static coefficients (includes zeroth coefficients) along with their delta and delta-delta feature vector are extracted resulting in 90-dimensional features. Baseline B02 consists of linear frequency cepstral coefficients (LFCCs) feature set along with 512 GMM as a back-end classifier [32]. The bandwidth used here is from 30 Hz to 8 kHz. LFCCs are extracted using a 512-point discrete Fourier transform applied to windows of 20 ms with 50% overlap. 60-dimensional features are computed that include 20-dimensional static coefficients (includes zeroth coefficients) along with their delta and delta-delta feature vector. Three post-evaluation systems were also used for comparison purposes. They are the high-spectral resolution linear frequency cepstral coefficient system with a conventional Gaussian mixture model classifier (LFCC-GMM) [33], the RawNet2 [26] system and the ResNet18-SP [25] system. Note that the last two systems, well known in the literature, are complex deep neural networks implying millions of parameters.

4.2. Implementation details of PRIVASP

PRIVASP uses LFCC features as the front-end. The processing to extract the audio features is as follows. The speech waveform is frame-blocked using a sliding window of 30 ms with a 15 ms shift. We extract 30 linear frequency cepstral coefficients from the first 1500 ms of each utterance (note that if the length of the utterance is less than 1500 ms, we merely repeat it). The obtained matrix is then vectorized into a column vector of 2970 elements. A shallow neural network with one hidden layer and ReLU activation function has been used as back-end. Two variants of PRIVASP in terms of numbers of neurons have been evaluated: PRIVASP-1024 and PRIVASP-512, which use 1024 and 512 nodes, respectively. The network is trained with the Adaptive Moment Estimation (Adam) algorithm with the binary cross-entropy loss. The model chosen is the one with the lowest loss value during the training phase. In the experiments, a PC with a 6-core Intel i5-9400F processor at 2.9GHz, a GPU GeForce GTX 1050 with 4GB memory, and 64GB RAM was employed. We used PyTorch framework [2] to build the NN and to implement PRIVASP we utilized the Pysyft [3] library [34] that supports MPC within Pytorch.

[2] https://pych.org/

[3] https://github.com/OpenMined/PySyft

Table 1: *Average inference time in ms per utterance.*

system / type	**PRIVASP-1024**	**PRIVASP-512**	B01	B02	LFCC-GMM	RawNet2	ResNet18-SP
plaintext	2.8	2.7	339.9	89.9	100.6	12.0	2.8
scenario 1	95.8	59.9	-	-	-	-	-
scenario 2	349.6	208.1	-	-	-	-	-

5. Experimental results

The evaluation follows a threefold objective: i) analyzing the performance of countermeasures, ii) assess privacy-preserving algorithms, and iii) evaluate the computational costs.

Table 2 and Table 3 show experimental results in terms of pooled EER and min t-DCF for the two baselines, B01 and B02, the high-spectral-resolution LFCC, RawNet2, ResNet18-SP, and our proposed PRIVASP-1024 and PRIVASP-512 systems. In order to assess detection performance without compromising privacy, two sets of experiments were carried out for the two PRIVASP systems for the ciphertext scenarios 1 and 2. In the development partition, Table 2, all the systems perform quite well. PRIVASP-1024 and PRIVASP-512 have a perfect performance even in the two PRIVASP scenarios. Regarding the evaluation partition, Table 3, as expected, PRIVASP-1024 performs slightly better than PRIVASP-512 in plaintext, 7.03%/0.1485 EER/min-tDCF. Of note, is that there is no drop in performance in the two PRIVASP scenarios. A comparison with other systems also demonstrates the effectiveness of PRIVASP, which performs better than RawNet2 and the two baselines B01 and B02.

In terms of efficiency, Table 1 shows the average time in ms to infer whether an utterance is bona fide or spoof. Due to its reduced number of neurons, PRIVASP-512 is more efficient than PRIVASP-1024 in both plaintext and ciphertext scenarios. In particular, in scenario 2, an utterance is detected as bona fide or spoof in 350ms and 208ms for PRIVASP-1024 and PRIVASP-512, respectively. This time is close to B01 (plaintext) and however acceptable in real-time applications. Scenario 1, in which the privacy of models is not taken into account, an utterance is detected as bona fide or spoof in 95ms and 60ms for PRIVASP-1024 and PRIVASP-512, respectively, which is better than the systems B01, B02, and LFCC-GMM, all in the plaintext scenario. Finally, the plaintext efficiency of PRIVASP systems is comparable to the two deeper networks RawNet2 and ResNet18-SP. Although these measurements may seem odd, they are justified by the fact that we did not use GPUs to evaluate PRIVASP.

Table 2: *Performance for the ASVspoof 2019 LA development partition in terms of pooled EER and min t-DCF for the two baselines, B01 and B02, the high-spectral-resolution LFCC, RawNet2, ResNet18-SP and our proposed PRIVASP-1024 and PRIVASP-512 systems. PRIVASP systems are also evaluated in privacy-preserving scenario 1 and 2.*

system	type	EER [%]	min-tDCF
B01	plaintext	0.43	0.0123
B02	plaintext	2.71	0.0663
LFCC-GMM	plaintext	0.00	0.0000
RawNet2	plaintext	1.09	0.0362
ResNet18-SP	plaintext	0.07	0.0018
PRIVASP-1024	plaintext	0.00	0.0000
	scenario 1	0.00	0.0000
	scenario 2	0.00	0.0000
PRIVASP-512	plaintext	0.00	0.0000
	scenario 1	0.00	0.0000
	scenario 2	0.00	0.0000

Table 3: *Same as in Table 2 for the evaluation partition.*

system	type	EER [%]	min-tDCF
B01	plaintext	9.57	0.2366
B02	plaintext	8.09	0.2116
LFCC-GMM	plaintext	3.50	0.0904
RawNet2	plaintext	5.54	0.1547
ResNet18-SP	plaintext	6.82	0.1140
PRIVASP-1024	plaintext	7.03	0.1485
	scenario 1	7.02	0.1481
	scenario 2	7.02	0.1481
PRIVASP-512	plaintext	7.10	0.1549
	scenario 1	7.13	0.1550
	scenario 2	7.13	0.1550

6. Conclusions

In this work, we propose PRIVASP, the first privacy-preserving solution to voice anti-spoofing. Rather than considering *privacy as an add-on*, we adopt a *privacy by-design* approach. The spoofing countermeasure is designed from scratch so that it is compatible with the computational capacity of secure multiparty computation, thereby ensuring efficient privacy preservation and reliable spoofing detection. The approach is based upon a shallow neural network with only one layer and a ReLU activation function. Experiments were performed on the recent ASVspoof 2019 Logical Access database. Two scenarios were considered, depending on whether or not the spoofing countermeasure service providers reveal their model to the cloud service provider. Results confirm the computational efficiency of the system, as well as its ability to discriminate bona fide from spoofed utterances despite the use of a shallow underlying network. PRIVASP can be used for real-time applications while protecting both the user's private speech data as well as the intellectual property of the service provider's model. In our future work, we intend to explore the potential of our approach to bring privacy preservation to state-of-the-art anti-spoofing systems that provides enhanced performance, and alternative cryptographic primitives whose integration has no impact on computational capacity or accuracy.

7. Acknowledgements

This work is supported by the TReSPAsS-ETN project funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860813. It is also supported by the ANR-DFG RESPECT project.

8. References

[1] A. Nagrani, J. S. Chung, W. Xie, and A. Zisserman, "Voxceleb: Large-scale speaker verification in the wild," *Computer Speech & Language*, vol. 60, p. 101027, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230819302712

[2] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification: A survey," *Speech Communication*, vol. 66, pp. 130–153, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639314000788

[3] M. Sahidullah, H. Delgado, M. Todisco, T. Kinnunen, N. Evans, J. Yamagishi, and K. Lee, "Introduction to voice presentation attack detection and recent advances," in *Handbook of Biometric Anti-Spoofing - Presentation Attack Detection, Second Edition*, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2019, pp. 321–361.

[4] A. Nautsch, A. Jiménez, A. Treiber, J. Kolberg, C. Jasserand, E. Kindt, H. Delgado, M. Todisco, M. A. Hmani, A. Mtibaa *et al.*, "Preserving privacy in speaker and speech characterisation," *Computer Speech & Language*, vol. 58, pp. 441–480, 2019.

[5] A. Nautsch, C. Jasserand, E. Kindt, M. Todisco, I. Trancoso, and N. Evans, "The gdpr & speech data: Reflections of legal and technology communities, first steps towards a common understanding," *arXiv preprint arXiv:1907.03458*, 2019.

[6] A. Nautsch, J. Patino, A. Treiber, T. Stafylakis, P. Mizera, M. Todisco, T. Schneider, and N. Evans, "Privacy-preserving speaker recognition with cohort score normalisation," 2019.

[7] M. A. Pathak and B. Raj, "Privacy-preserving speaker verification and identification using gaussian mixture models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 397–406, 2013.

[8] A. Gumaei, R. Sammouda, A. M. S. Al-Salman, and A. Alsanad, "Anti-spoofing cloud-based multi-spectral biometric identification system for enterprise security and privacy-preservation," *Journal of Parallel and Distributed Computing*, vol. 124, pp. 27–40, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731518307391

[9] R. Shao, P. Perera, P. C. Yuen, and V. M. Patel, "Federated face presentation attack detection," 2020.

[10] A. C. Yao, "Protocols for secure computations," in *Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society, 1982, pp. 160–164.

[11] A. Yao, "How to generate and exchange secrets (extended abstract)," in *FOCS*, 1986.

[12] D. Beaver, "Efficient multiparty protocols using circuit randomization," in *Advances in Cryptology — CRYPTO '91*, J. Feigenbaum, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 420–432.

[13] O. Goldreich, S. Micali, and A. Wigderson, "How to play any mental game," in *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, ser. STOC '87. New York, NY, USA: Association for Computing Machinery, 1987, p. 218–229. [Online]. Available: https://doi.org/10.1145/28395.28420

[14] Y. Lindell, "Secure multiparty computation (mpc)." *IACR Cryptol. ePrint Arch.*, vol. 2020, p. 300, 2020.

[15] M. Azraoui, M. Bahram, B. Bozdemir, S. Canard, E. Ciceri, O. Ermis, R. Masalha, M. Mosconi, M. Önen, M. Paindavoine *et al.*, "Sok: Cryptography for neural networks," in *IFIP International Summer School on Privacy and Identity Management*. Springer, 2019, pp. 63–81.

[16] S. Wagh, S. Tople, F. Benhamouda, E. Kushilevitz, P. Mittal, and T. Rabin, "Falcon: Honest-majority maliciously secure framework for private deep learning," 2020.

[17] B. Bozdemir, O. Ermis, and M. Önen, "ProteiNN: Privacy-preserving one-to-many Neural Network classifications," in *SECRYPT 2020, 17th International Joint Conference on Security and Cryptography*, Lieusaint (on line), France, Jul. 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03151068

[18] A. Boudguiga., O. Stan., A. Fazzat., H. Labiod., and P. Clet., "Privacy preserving services for intelligent transportation systems with homomorphic encryption," in *Proceedings of the 7th International Conference on Information Systems Security and Privacy - Volume 1: ICISSP,*, INSTICC. SciTePress, 2021, pp. 684–693.

[19] C. Juvekar, V. Vaikuntanathan, and A. Chandrakasan, "Gazelle: A low latency framework for secure neural network inference," *CoRR*, vol. abs/1801.05507, 2018. [Online]. Available: http://arxiv.org/abs/1801.05507

[20] A. Gangal, M. Ye, and S. Wei, "Hybridtee: Secure mobile dnn execution using hybrid trusted execution environment," in *2020 Asian Hardware Oriented Security and Trust Symposium (AsianHOST)*, 2020, pp. 1–6.

[21] M. Sabt, M. Achemlal, and A. Bouabdallah, "Trusted execution environment: What it is, and what it is not," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1, 2015, pp. 57–64.

[22] S. P. Bayerl, F. Brasser, C. Busch, T. Frassetto, P. Jauernig, J. Kolberg, A. Nautsch, K. Riedhammer, A.-R. Sadeghi, T. Schneider, E. Stapf, A. Treiber, and C. Weinert, "Privacy-preserving speech processing via stpc and tees," in *PPML 2019, Privacy Preserving Machine Learning Workshop, CCS 2019 Workshop, November 15, 2019, London, UK*, London, 2019.

[23] A. Nautsch, S. Isadskiy, J. Kolberg, M. Gomez-Barrero, and C. Busch, "Homomorphic encryption for speaker recognition: Protection of biometric templates and vendor model parameters," *Odyssey 2018 The Speaker and Language Recognition Workshop*, Jun 2018. [Online]. Available: http://dx.doi.org/10.21437/Odyssey.2018-3

[24] A. Treiber, A. Nautsch, J. Kolberg, T. Schneider, and C. Busch, "Privacy-preserving plda speaker verification using outsourced secure computation," *Speech Communication*, vol. 114, pp. 60–71, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639318304369

[25] T. Chen, A. Kumar, P. Nagarsheth, G. Sivaraman, and E. Khoury, "Generalization of audio deepfake detection," in *Proc. Speaker Odyssey*, 2020, pp. 1–5.

[26] H. Tak, J. Patino, M. Todisco, A. Nautsch, N. Evans, and A. Larcher, "End-to-end anti-spoofing with rawnet2," in *Proc. ICASSP*, 2021.

[27] M. Todisco, X. Wang, V. Vestman, M. Sahidullah, H. Delgado, A. Nautsch, J. Yamagishi, N. Evans, T. H. Kinnunen, and K. A. Lee, "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection," in *INTERSPEECH*, Graz, Austria, 2019, pp. 1008–1012.

[28] T. Kinnunen, H. Delgado, N. Evans, K. A. Lee, V. Vestman, A. Nautsch, M. Todisco, X. Wang, M. Sahidullah, J. Yamagishi *et al.*, "Tandem assessment of spoofing countermeasures and automatic speaker verification: Fundamentals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2195–2210, 2020.

[29] N. Brümmer and E. de Villiers, "The bosaris toolkit: Theory, algorithms and code for surviving the new dcf," 2013.

[30] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in *Speaker Odyssey Workshop*, vol. 25, Bilbao, Spain, 2016, pp. 249–252.

[31] ——, "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," *Computer Speech and Language*, vol. 45, pp. 516–535, 2017.

[32] M. Sahidullah, T. Kinnunen, and C. Hanilçi, "A comparison of features for synthetic speech detection," in *INTERSPEECH*, Dresden, Germany, 2015, pp. 2087–2091.

[33] H. Tak, J. Patino, A. Nautsch *et al.*, "Spoofing Attack Detection using the Non-linear Fusion of Sub-band Classifiers," in *Proc. INTERSPEECH*, 2020, pp. 1106–1110.

[34] T. Ryffel, A. Trask, M. Dahl, B. Wagner, J. Mancuso, D. Rueckert, and J. Passerat-Palmbach, "A generic framework for privacy preserving deep learning," 2018.