AGREE: Efficient and Privacy-preserving Speaker Recognition for Internet of Things

Rongxing Lu*

Contact Email: rlu1@unb.ca

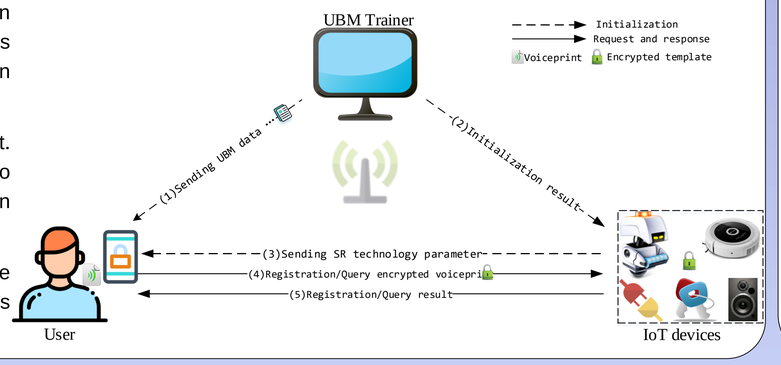
Canadian Institute for Cybersecurity (CIC), Faculty of Computer Science, University of New Brunswick (UNB)

ABSTRACT

With the booming development of intelligent terminals, the applications of speaker recognition have been seen increasingly rapid advances in the past two decades. However, the flourish of speaker recognition technology still faces many challenges in IoT (Internet of Things) scenarios, especially preventing the disclosure of voiceprint. In this , an efficient and privacy-preserving speaker recognition scheme, named AGREE, is proposed for IoT. With AGREE, the speaker's identity can be recognized in different security levels without the data leakage of voiceprint. To be specific, based on random matrix theory, a voiceprint encryption algorithm and the corresponding privacy-preserving similarity computation over ciphertext algorithm are proposed to achieve the efficient and accurate speaker recognition scheme by computing the match score of voiceprint over the encrypted i-vector data. Detailed analysis shows that AGREE can resist various known security threats. Moreover, AGREE is implemented with a real speaker voiceprint dataset, and extensive simulation results demonstrate that our proposed scheme is highly efficient and can be implemented effectively.

System Model of AGREE

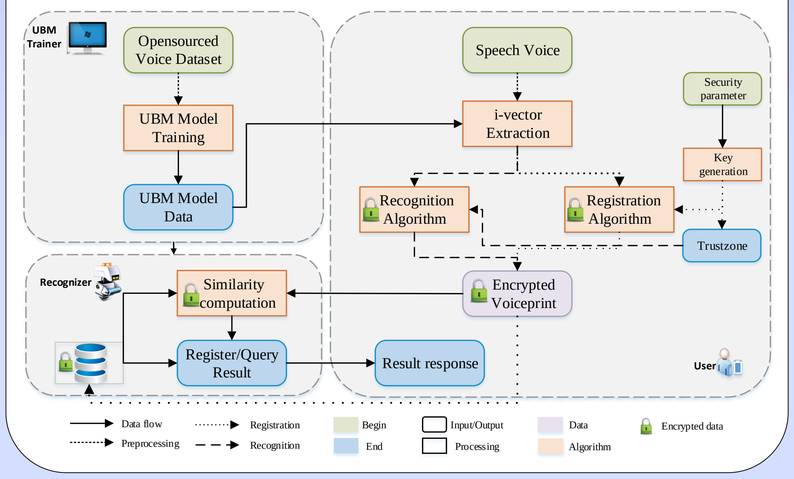
- **IoT devices** are waiting for the user's recognition query and then choose a suitable speaker recognition technology to use, which holds encrypted voiceprints of users. During the recognition phrase, IoT devices compute the similarity between encrypted registering voiceprint and encrypted recognizing voiceprint, then send the result of recognition to user. By the way, it can be extended to some smart devices in practice.
- **User** is a participant of privacy-preserving speaker recognition, it can be a smart phone or a PC client. User receives the UBM parameter data from UBM trainer. User sends encrypted voiceprint data to recognition during the registration and recognition. Last but not least, there is a trust zone in user which can store the key and other applications cannot read from it.
- **UBM trainer** is the UBM model trainer and the generator of UBM parameters. It can use some open source speech voice datasets to train the model and generate parameters. After generating, UBM trainer sends the parameters to user.



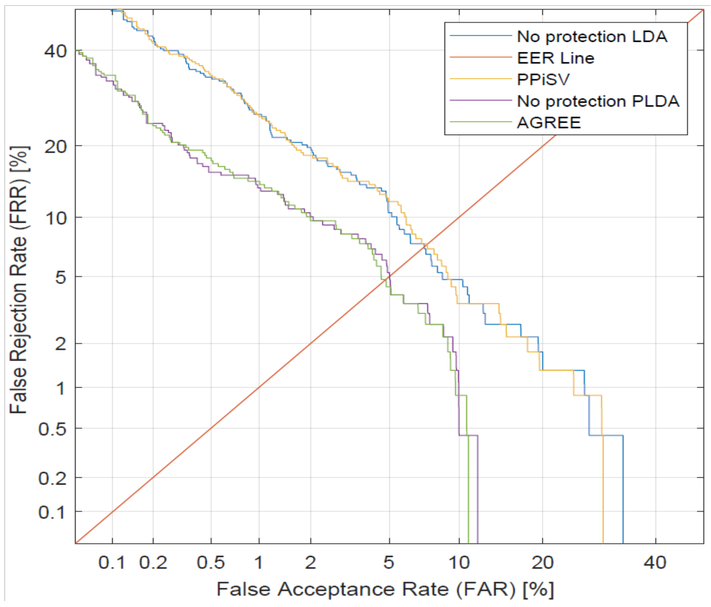
Security Model

- **Level-I: Security model under ciphertext-only attack.** The model is secure under this kind of adversary attack which can obtain the encrypted voiceprint and the encrypted queries.
- **Level-II: Security model under known-sample attack.** With Level-I security, we assume that the adversary has access to some plaintext samples in the database but do not know the corresponding ciphertexts. The model is secure under this kind of adversary attack.
- **Level-III: Security model under chosen-plaintext attack.** With Level-I and Level-II security model, we assume that the adversary can get a set of samples in the database and know the corresponding ciphertexts of these samples. Moreover, the adversaries are assumed to enable to generate queries of their interest arbitrarily. The model is secure under this kind of adversary attack.

Overview of AGREE



Proposed AGREE vs other schemes



- This work is still under revision with the collaborators at Xi'dian University.

i-vector Encryption Algorithm

Algorithm 2 i-vector Encryption Algorithm in Registration

Input: the i-vector w_e; the LDA matrix P; the UBM matrices $U_k, \Lambda, \hat{Q}$; $(n+\gamma)\times(n+\gamma)$ random invertible matrices M_1, M_2; a random $(n+\gamma)$-dimension vector H;

Output: Encrypted ciphertext C_i, C_e;

1: Channel compensation: $T_e = Pw_e^T$;
2: $\eta_e = U_k^T T_e$;
3: $C_e = \eta_e^T \hat{Q} \eta_e$;
4: Extending η_e to a $(n+\gamma)$-dimension vector as $\hat{\eta}_e$, where the $(n+1), (n+2), \cdots, (n+\gamma)$th element are all set as 1;
5: Vector diagonalization: $D = diag(\hat{\eta}_e)$;
6: $W_D = D \times A$, where A is a random $(n+\gamma)\times(n+\gamma)$ matrix, and $A_i \times H^T = 1, A = [A_1, A_2, \cdots, A_{(n+\gamma)}]^T$;
7: Encryption: $C_i = M_1 \times W_D \times M_2$;
8: **return** C_i, C_e.

Algorithm 3 i-vector Encryption Algorithm in Recognition

Input: the i-vector w_a; the LDA matrix P; $(n+\gamma)\times(n+\gamma)$ random invertible matrix M_1, M_2; the PLDA matrices $U_k, \Lambda, \hat{Q}$; a random γ-dimension vector r; a random $(n+\gamma)$-dimension vector H;

Output: Encrypted ciphertext C_k, C_M;

1: Channel compensation: $T_a = Pw_A^T$;
2: $\eta_a = U_K^T T_a$;
3: Norm computation: $C_a = \eta_a^T \hat{Q} \eta_a$;
4: $\widetilde{\eta}_a = \eta_a^T \Lambda$;
5: Extending $\widetilde{\eta}_a$ to $\hat{\eta}_a$: $\hat{\eta}_a =< \widetilde{\eta}_a, r_1, \cdots, r_\gamma >$;
6: $C_H = M_2^{-1} \times H^T$;
7: Encryption: $C_M = 2\hat{\eta}_a^T M_1^{-1}$;
8: **return** C_H, C_M, C_a.

Privacy-preserving similarity computation

During the similarity computation, the IoT device performs the following steps to recognition:

1) The IoT device computes the score:

$$V_i = C_e + C_a + C_M \times C_i \times C_H. \quad (4)$$

2) The IoT device selects the one with the largest score in the result set as the matching result. After computing all V_i for all the encryption i-vector C_i, the IoT device can find out the $\{I_i, w_i\}$ which has the maximum similarity to w_a.

Performance Evaluation

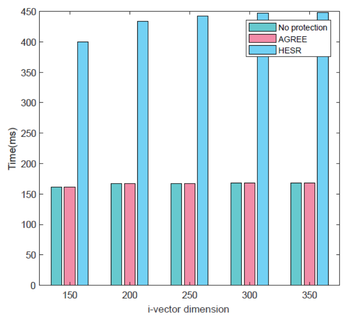
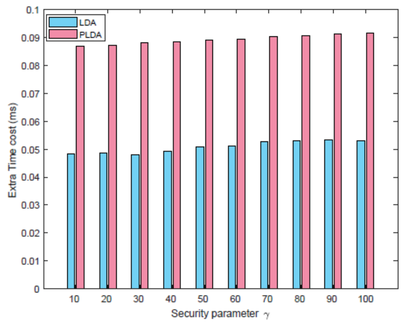
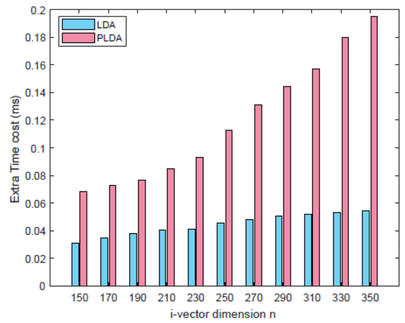


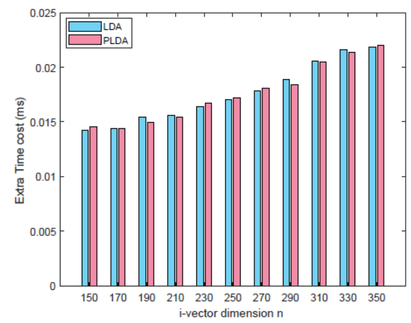
Fig. 7: Query time in no protection scheme, AGREE and HESR($\tau = PL$).



(a) Extra computation cost of different technology in AGREE with different security parameter γ



(b) Extra computation cost of different techniques for user in AGREE with different i-vector projection dimensions n



(c) Computation cost of different techniques for IoT device in AGREE with different i-vector projection dimensions n

Conclusion

In this work, we propose an efficient and flexible scheme for privacy-preserving speaker recognition, named AGREE. With our proposed scheme, the voiceprint of users would not be revealed to server and users can set different security parameters according to their own security requirements. AGREE can provide flexible recognition approaches with encrypted voiceprint data for privacy-preserving speaker recognition in devices of IoT by setting speaker recognition parameters, and guarantee accurate verification results.

Future work

Detailed security analysis shows that the proposed scheme is privacy-preserving, i.e., no one can read each user's voiceprint data. In addition, extensive experiments in this paper were conducted to demonstrate its precision and efficiency. In future work, we may take other techniques into consideration and combine these techniques with AGREE to achieve a more efficient scheme for privacy-preserving speaker recognition with deep learning.