Privacy-Preserving Speech Processing via STPC and TEEs

1. Motivation & Contribution

Current Situation: Voice-based Interfaces are Becoming Omnipresent
- > 28 smartphone users (Amazon Alexa, Apple Siri, Google Assistant, Microsoft Cortana)
- Increasing number of smart-home devices (Amazon Echo, Apple Home-Pod, Google Home)
- Automatic Speaker Verification (ASV) for authentication over the phone (e.g., for banking)

“How is the weather today?”

Risks: Voice Data Contains Sensitive Biometric Information as well as Spoken Words
- Impersonation attacks, extracting context, inferring sensitive data (health, ethnicity, etc.)

Problem Statement: Naive Solution of Performing Speech Processing on Client-side fails
- Shipping the model parameters to the client contradicts the business interests of vendors

Contributions: Secure and Private Speech Processing Architectures
- VoiceGuard [1]: secure & private speech processing via Intel SGX
- Offline Model Guard (OMG) [2]: secure & private speech processing on mobile devices
- Private ASV [3] via outsourced secure two-party computation (STPC)

2. Related Work

Privacy-Preserving Machine Learning
- Via Secure Multi-Party Computation
  - Orders of magnitude higher communication
  - Impractical for (large-scale) on-the-fly processing due to repeated initialization costs
- Via Homomorphic Encryption
  - Orders of magnitude higher computation time
  - Impractical for (large-scale) on-the-fly processing due to high latency

Privacy-Preserving Speech Processing (Nautsch et al., Computer Speech and Language’19)
- Speech processing via cryptographic means is still very inefficient
- Example: >3h to encrypt 1s of audio & recognize a word out of a 10 word vocabulary
- Currently far from suitable for speech recognition in real time due to high overhead
- Certain tasks like ASV can be performed via HE in reasonable time

PLDA ASV on i-vectors with dimension 250 takes >6m (Nautsch et al., Odyssey’18)
- Private ASV protocols by Rahulamathavan et al. (CybecSA79, TASEP17) found to be highly insecure (Treiber & Schneider, TPDS’19)
- Often, intermediate values, some model parameters, or even entire voice models of individuals are leaked (e.g., Portoles et al., EUSPIC’14)

 VoiceGuard & OMG: Novel Architectures for Privacy-Preserving Speech Processing

3. VoiceGuard [1]

Drawbacks of Previous HE-based ASV
- High computational demand for client device & inefficient transaction times
- Threshold value, comparison score, and intermediate fixed-point exponents leaked
- No security against malicious users

Our Construction based on Outsourced STPC
- Reference embedding stored by two non-colluding servers in secret shared form
- Probe embedding secret shared during verification step, servers then compute verification in STPC
- Hides the score, intermediate results, and is secure against malicious users
- Satisfies international standards on Biometric Information Protection (ISO/IEC IS 24745)

Evaluation
- Implementation based on ABY (Demmler et al., NDSS’15) evaluated on NIST i-vector ML challenge
- Run-time: 0.5s (improvement up to 4,000x)
- Fixed-point, but retains 100% accuracy

4. Offline Model Guard (OMG) [2]

Drawbacks of Previous SMC-based ASV
- Often impractical for (large-scale) on-the-fly processing due to high latency
- Practicality and security are still a trade-off

Our Construction based on Outsourced STPC
- Reference embedding stored by two non-colluding servers in secret shared form
- Probe embedding secret shared during verification step, servers then compute verification in STPC
- Hides the score, intermediate results, and is secure against malicious users
- Satisfies international standards on Biometric Information Protection (ISO/IEC IS 24745)

Evaluation: Offline Keyword Recognition
- Based on TensorFlow Lite for Microcontrollers
- Uses spectral fingerprints & CNN architecture
- Implementation on ARM HMP 960 development board (octo-core ARMv8 SoC, 3GB RAM)

5. Private Automatic Speaker Verification (ASV) [3]

Drawbacks of Previous HE-based ASV
- High computational demand for client device & inefficient transaction times
- Threshold value, comparison score, and intermediate fixed-point exponents leaked
- No security against malicious users

Our Construction based on Outsourced STPC
- Reference embedding stored by two non-colluding servers in secret shared form
- Probe embedding secret shared during verification step, servers then compute verification in STPC
- Hides the score, intermediate results, and is secure against malicious users
- Satisfies international standards on Biometric Information Protection (ISO/IEC IS 24745)

Evaluation
- Implementation based on ABY (Demmler et al., NDSS’15) evaluated on NIST i-vector ML challenge
- Run-time: 0.5s (improvement up to 4,000x)
- Fixed-point, but retains 100% accuracy

6. Conclusion

VoiceGuard & OMG: Novel Architectures for Privacy-Preserving Speech Processing
- Protects the user’s sensitive voice data & the vendor’s IP (i.e., model parameters)
- Support user-specific models, such as feature transformations (e.g., iMLLR), i-vectors, or model transformations (e.g., custom output layers)
- Deployment either in the cloud or on-premises (VoiceGuard)
- Prototype implementations demonstrate applicability for speech recognition in real time
- Generic & easy to work for related tasks (speaker verification or voice biometrics, including emotion recognition and medical speech processing)

Private ASV: PLDA-based Speaker Verification
- Outperforms HE-based solutions both in terms of efficiency and security
- Open-source implementation (https://encrypto.cs.tu-darmstadt.de/)
- LLR threshold precision limited to three decimal points (but sufficient for ASV)

7. References