Voiceprint-based Access Control for Wireless Insulin Pump Systems

Presenter: **Xiaojiang Du**

Bin Hao[†], Xiali Hei[†], Yazhou Tu[†], Xiaojiang Du[‡], and Jie Wu[‡]

[†]School of Computing and Informatics, University of Louisiana at Lafayette, LA, USA

[‡]Department of Computer and Information Sciences, Temple University, PA, USA

IEEE MASS 2018

Chengdu, China

OCT. 9-12, 2018

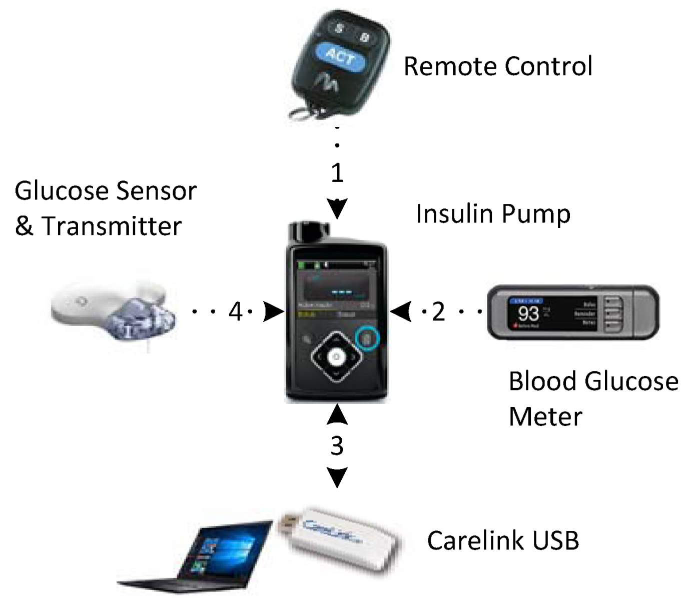
Insulin Pump System

- As of 2015, there were an estimated 30.3 million people of all ages in the U.S. suffering from diabetes
- People with type 1 diabetes (about 5% of diabetics) need insulin pumps
- Insulin pump systems adopt wireless channels with few cryptographic mechanisms
  - Vulnerable to many attacks (eavesdropping, remote dosage setting, etc.)
  - Threatening the privacy and safety of the users

Existed Attacks and Countermeasures

Attacks Countermeasures

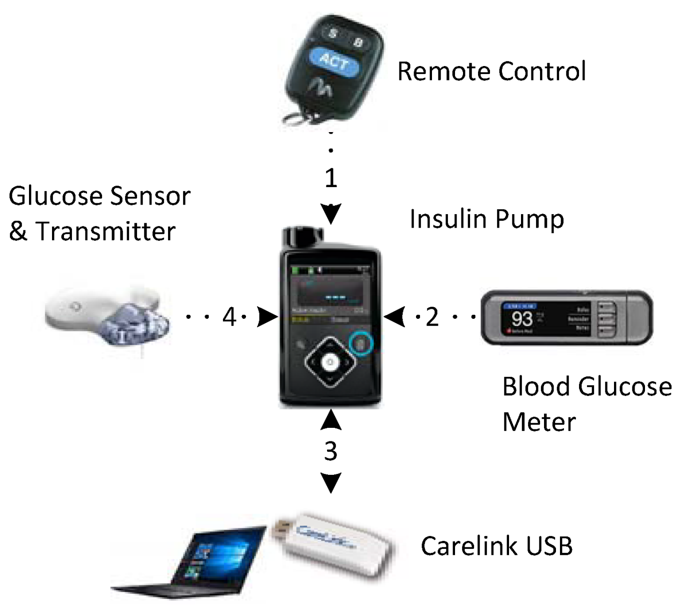
- Radcliffe, 2011
  - intercepted glucose data in link 4, caused wrong readings displaying
- Jack, 2011
  - captured data transmitted from computer (link 3), made the pump deliver fatal doses
- Li et al., 2011 and Marin et al., 2016
  - fully reverse-engineered the wireless communication protocol (link 1-4)



A real time close-loop insulin pump system

Existed Attacks and Countermeasures

Attacks



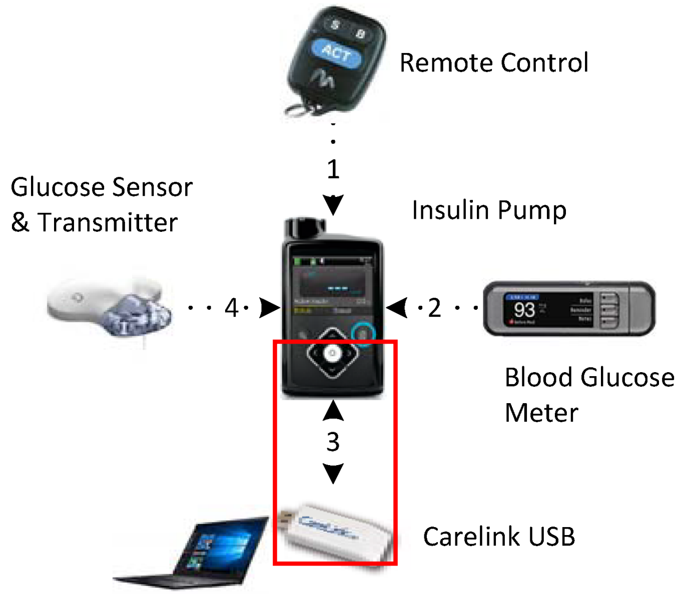
A real time close-loop insulin pump system

Countermeasures

- AES-MAC-based cryptographic solution (Marin et al., 2016)
  - focuses on link 1, applicable to link 2/3/4
  - needs sharing of symmetric keys
- Patient infusion pattern based access control (PIPAC, Hei et al., 2013)
  - focuses on link 3
  - assumes the patient's parameters can only be changed manually, not suitable in a closed-loop control system

Our Motivation

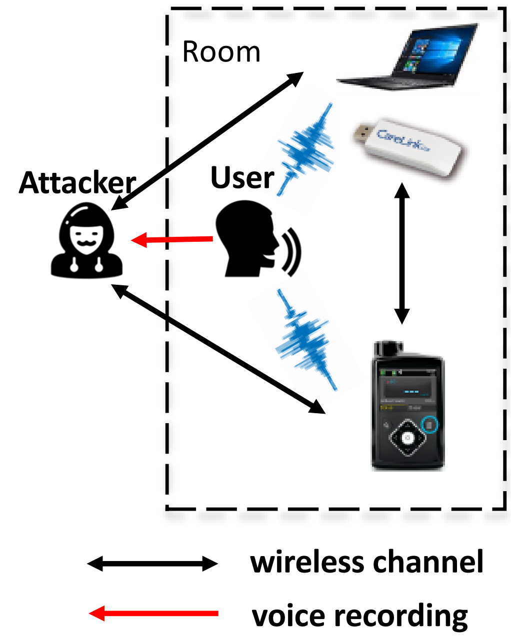
- We focus on the wireless channel between the Carelink USB and insulin pump (link 3) in a close-loop insulin pump system
- Attacks over link 3
  - Eavesdropping (Privacy)
  - Remote dosage setting (Safety)



How to establish a secure channel between unacquainted devices in a close-loop system?

Basic Idea

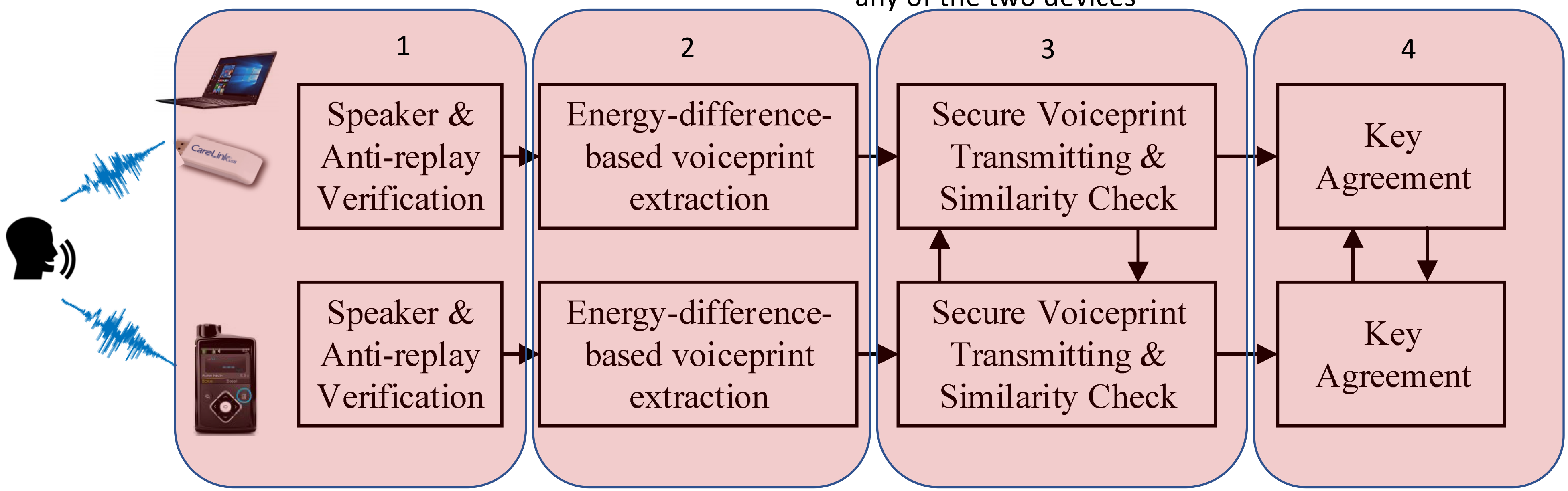
- Cascaded fusion of speaker verification and anti-replay countermeasure
  - to ensure the insulin pump is accessed by the Carelink USB only after the legitimate user passes the identity/speaker verification
- Key Agreement based on energy-difference-based voiceprint extraction [Schürmann et al., 2013 and Haitsma et al., 2002] and secure multi-party computing (SMC)
  - to generate a common cryptographic key between the two unacquainted devices only when the user and the devices are in close proximity



Our Solution: Voiceprint-based Access Control

Phase 1: Speaker & Anti-replay Verification: accept the legitimate user, reject the replay impostor

Phase 3: Secure Voiceprint Transmitting & Similarity Check: abort if the voiceprint similarity check fails in any of the two devices



Phase 2: Energy-difference-based voiceprint extraction

Phase 4: Key agreement to establish a secure channel between the two devices

System Model

- Considered Scenario: CareLink USB wants to acquire access to an insulin pump to request data or remotely modify the therapy settings
- Access Control Process
  - First, CareLink USB sends request to the pump to activate the access privilege
  - Then, the pump starts the speaker verification and the user says random passphrase
  - After successful verification, the pump then bootstraps a key agreement with the Carelink USB

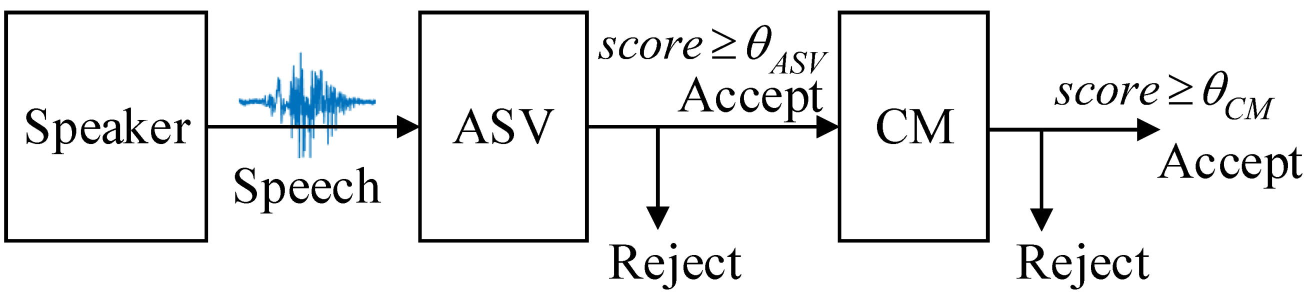
Authentication can be achieved if the user passes the speaker verification and Carelink USB is in close proximity to the pump and the user.

Attacker Model

- Scenario A (Remote impersonation)
  - The attacker not in close proximity tries to pass speaker verification and perform key agreement with the pump by remotely receiving the user's voice or just using the voice previously recorded.
- Scenario B (Passive eavesdropping)
  - The attacker eavesdrops on the messages transmitted over the wireless channel and records the voice of the legitimate user.
- Scenario C (Man-in-the-middle, MITM)
  - The attacker tries to actively participate in the authentication process to establish a secure channel with the insulin pump.

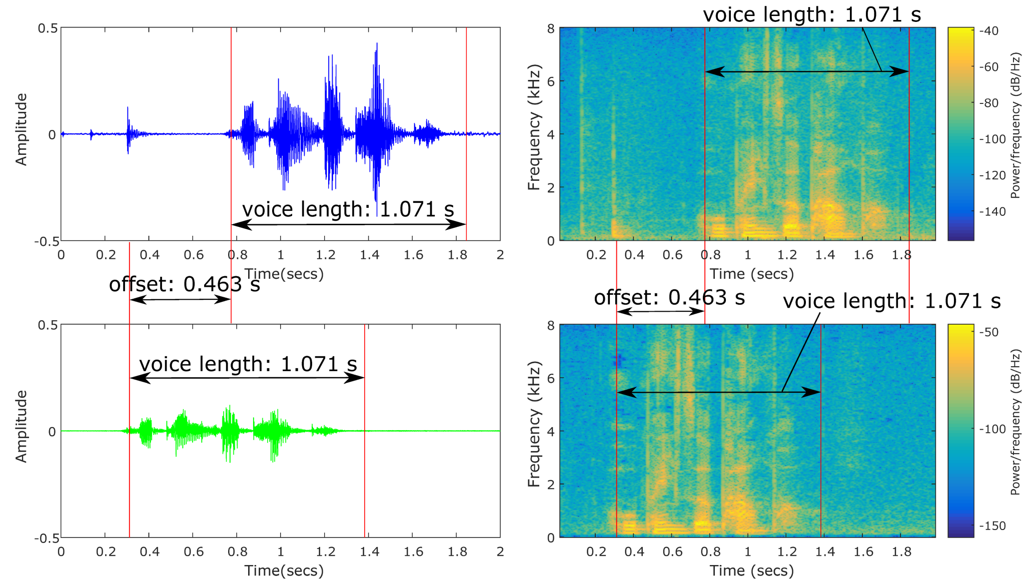
Voiceprint-based Access Control Scheme

- **Phase 1**: Speaker & Anti-replay Verification
  - Speaker-dependent: only the legitimate user can pass the verification.
  - Text-independent: the user can use any passphrases.
  - Lightweight Speaker Model: only one speaker (the pump user) in each system.
  - Cascaded Fusion of ASV (Automatic Speaker Verification) and Anti-replay Countermeasure (CM)
    - ASV confirms that the voice comes from the target user (genuine or replayed)
    - CM confirms that the voice comes from a real person, not a replay device (e.g., loudspeaker)



Voiceprint-based Access Control Scheme

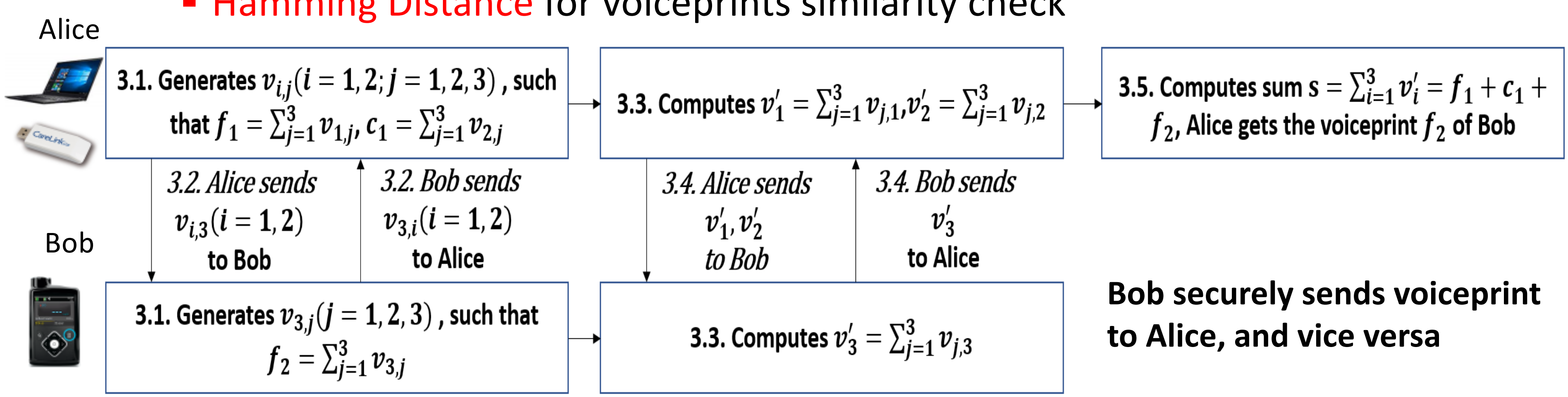
- **Phase 2**: energy-difference-based voiceprint extraction
  - The pump and Carelink USB record same passphrase simultaneously
  - Each device extracts a binary sequence (voiceprint) with the length of N*M bits using energy-difference-based scheme (M frequency bands of each of N frames)
  - Cross-correlation is used to align the two recorded voices



Amplitude and Frequency spectrum of passphrase "Open the pump" recorded by iPhone 5S (top) and Samsung Galaxy S5 (bottom). The similarity of the two extracted voiceprints is 85.49%.

Voiceprint-based Access Control Scheme

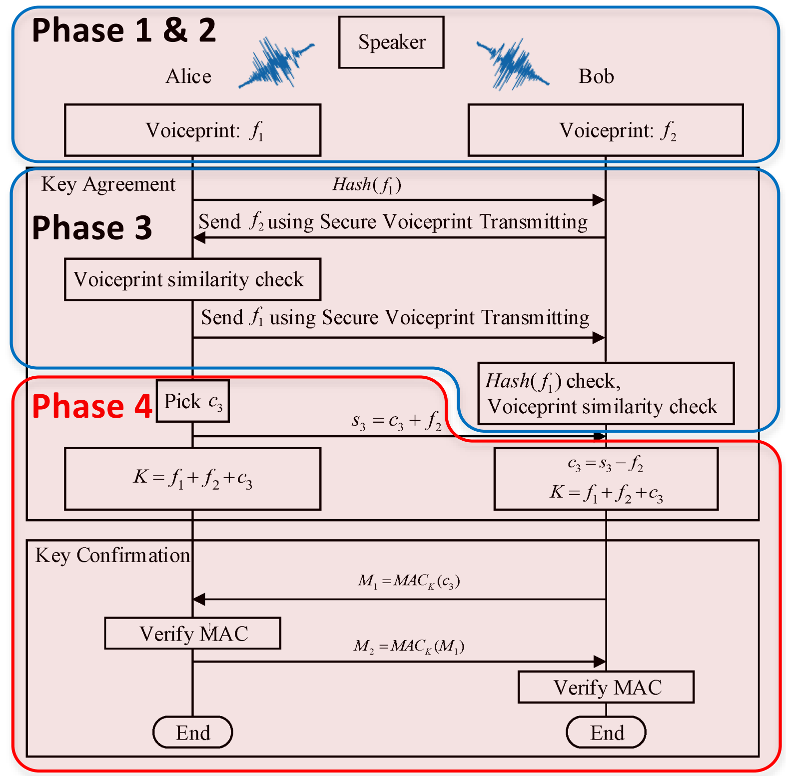
- **Phase 3**: Secure Voiceprint Transmitting & Voiceprint Similarity Check
  - Voiceprints cannot be directly used as a key: similar but not identical
  - Secure Voiceprint Transmitting (SVT) Protocol to securely exchange voiceprints
  - Hamming Distance for voiceprints similarity check



Bob securely sends voiceprint to Alice, and vice versa

Voiceprint-based Access Control Scheme

- **Phase 4**: Key agreement to establish a secure channel between the two devices
  - Voiceprints as seed
  - Secure Voiceprint Transmitting Protocol as basic unit
  - $Key = f_1 + f_2 + c_3$
  - Key Confirmation using MAC



Evaluation (I)

- Feature Selection
  - Short-term power spectrum features (MFCC, IMFCC, etc.)
  - Constant-Q Cepstral Coefficients (CQCC)
- Speaker Model
  - ASV : Gaussian mixture model with universal background model (GMM-UBM)
  - Countermeasure (CM): Gaussian mixture model (GMM)
- Datasets:
  - ASVspoof 2017 (T. Kinnunen et al.)

Subset	**# Speakers**	**# Utterances**	
		Genuine	***Spoof***
Training	10	1507	1507
Development	8	760	950
Evaluation	24	1294	11988
Total	42	3561	14445

Evaluation (II)

- Influence of VAD (voice activity detector)
  - 30 coefficients for CQCC
  - 20ms frame length and 40 filter banks for other features
  - VAD is critical: without VAD, there is no successfully trained classifier except CQCC.
  - MFCC, LPCC, and CQCC as candidates to train ASV
    - MFCC and LPCC achieve better performance
    - CQCC not sensitive to VAD

Features	***Training set (VAD)***	***Training set (No VAD)***
CQCC	0.66	**0.44**
MFCC	**0.54**	50.89
IMFCC	0.88	50.89
LPCC	**0.44**	55.56
LFCC	0.66	50.89
RFCC	0.57	50.89
SCFC	1.62	50.89
SCMC	0.88	50.89
SSFC	1.20	55.56

Standalone ASV feature performance (Equal Error Rare, % EER) with and without VAD

Evaluation (III)

- Standalone ASV performance of zero-effort and replay impostors
  - Zero-effort impostors: impersonate the genuine target speaker using their own sounds
  - Replay impostors: impersonate target speaker using recordings of target speaker

The higher the EER, the lower the performance:

$$EER_{replay} > EER_{zero-effort}$$

Speakers	Zero-effort Impostors			Replay Impostors		
	MFCC	*CQCC*	*LPCC*	*MFCC*	*CQCC*	*LPCC*
M0001	0.00	1.45	4.20	0.05	1.19	2.56
M0002	0.00	0.13	1.30	0.20	0.25	2.03
M0003	3.55	2.37	**0.66**	14.54	11.11	**10.40**
M0004	1.32	0.00	3.70	4.78	2.94	3.70
M0005	0.39	2.63	0.13	0.56	1.56	0.44
M0006	1.97	4.21	3.68	16.16	**8.89**	12.58
M0007	0.00	0.26	0.26	0.22	0.22	0.11
M0008	10.39	11.67	**1.67**	8.69	6.67	**2.46**
M0009	1.75	0.26	1.18	1.75	0.42	1.75
M0010	0.39	0.53	0.92	0.22	0.22	1.88

Evaluation (IV)

- Standalone CM Performance of Replay Impostors
  - Trained a 2-class GMM model using the Development (Dev) subset as enrollment and Evaluation (Eval) subset as prediction (column 2), and vice versa (column 3)
  - IMFCC feature achieves the best performance when Eval (larger than Dev) as enrollment set and Dev as prediction set

Features	**Enrollment/Prediction dataset**	
	Dev/Eval set	***Eval/Dev set***
CQCC	27.58	8.94
MFCC	38.78	8.00
IMFCC	34.67	**6.57**
LPCC	30.90	8.42
LFCC	37.06	7.23
RFCC	36.14	8.04
SCFC	**25.11**	29.05
SCMC	34.97	8.14
SSFC	33.94	8.03

Evaluation (V)

- ASV & CM Fusion Performance of Replay Impostors
  - In most cases, the fusion of MFCC/LPCC ASV and IMFCC CM gets a lower EER than a standalone ASV or CM
  - The fusion of LPCC ASV and IMFCC CM achieves the best performance: the maximal EER value for all evaluated speakers is 4.02%.

System	Speakers									
	M0001	***M0002***	***M0003***	***M0004***	***M0005***	***M0006***	***M0007***	***M0008***	***M0009***	***M0010***
ASV1 (MFCC)	0.20	0.00	11.11	3.70	0.78	4.78	0.16	6.67	1.75	0.17
CM1 (IMFCC)	7.11	6.06	6.96	6.71	6.67	6.31	7.59	6.61	6.11	6.81
Fusion1 (MFCC+IMFCC)	1.19	0.00	4.60	0.83	0.72	2.22	0.16	8.33	1.75	0.06
ASV2 (LPCC)	3.17	1.30	**0.80**	1.44	0.50	2.22	0.22	1.67	1.63	0.50
CM2 (IMFCC)	7.11	6.06	6.96	6.71	6.67	6.31	7.59	6.61	6.11	6.81
Fusion2 (LPCC+IMFCC)	**4.02**	2.60	0.52	0.50	0.22	1.77	0.11	**3.33**	1.63	0.33
ASV3 (CQCC)	2.31	0.20	4.94	0.56	0.89	4.44	0.05	6.67	0.00	0.22
CM3 (IMFCC)	7.11	6.06	6.96	6.71	6.67	6.31	7.59	6.61	6.11	6.81
Fusion3 (CQCC+IMFCC)	7.14	11.69	1.38	0.00	3.70	6.67	0.05	9.84	1.75	0.17

Evaluation (VI)

- Feasibility of Energy-difference-based Voiceprint Extraction
  - Two devices: iPhone 5S and Honor 10 (H10)
  - 270 passphrases, i.e., 135 for each device
  - 27 different distance settings relative to the voice source (speaker)
  - In each distance setting, the speaker spoke 5 sentences (each contains either 4 or 5 English words or numbers)
  - Voiceprint extraction: 16 kHz sampling frequency, 63 ms frame length, and 17 frequency filter banks.

(1) Average voiceprint similarity (AVS) is larger than ~80% when the two devices are positioned within distance <= 30cm to speaker

(2) AVS drops down to ~75% when one device is 300cm away, ~62% when one device is outside

Distance (cm)	**Average voiceprints similarity (%)**					
	5S 20	***5S 30***	***5S 50***	***5S 150***	***5S 300***	***5S outside***
H10 20	81.55	80.22	80.78	78.66	74.97	64.36
H10 30	81.35	80.37	77.69	78.34	75.89	62.27
H10 50	80.73	80.50	78.45	78.32	77.00	61.11
H10 150	75.29	76.17	74.17	-	-	-
H10 300	75.06	75.79	72.55	-	-	-
H10 outside	60.39	60.90	61.22	-	-	-

Security Analysis

- Remote Impersonation
  - The attacker MUST pass the anti-replay speaker verification (4.02% success rate when using recorded audio)
  - MUST pass the voiceprint similarity check by
    - generating a random bit (the voiceprints have high entropy), or
    - extracting from recorded voice (the user use random passphrases; the attacker cannot get a voiceprint having high similarity to the one in the pump in another context (similarity drops down to <75% according to the evaluation results))
- Eavesdropping
  - The attacker can pass the speaker verification using recorded audio in 4.02% success rate(EER = 4.02%)
  - Usage of long length of voiceprint (e.g., >=512 bits) and random passphrases (sequences of words) to resist voiceprint brute-force attack
  - Secure Voiceprint Transmitting: no information leakage of voiceprint
- Man-in-the-middle (MITM): Hash check, Key confirmation in Key Agreement

Discussion

- Storage overhead
  - Needs to store classifier models in the pump for only 1 user
  - ASV: 1 GMM-UBM model, 1 GMM user model (genuine), 1 GMM background users model (spoof)
  - CM: 1 GMM Genuine model, 1 GMM Spoof model
  - Total permanent storage: < 1 MB
- Computation complexity
  - Evaluated in Raspberry Pi 1 Model B+ with 700 MHz Broadcom BCM2835 CPU
  - 16 bits voiceprint for each voice frame; passphrases length: 19-42 frames; voiceprint length: 304-672 bits
  - The whole access control duration: ~1 s (after voice recording)
- Communication complexity
  - Total received and transmitted data: < 10 Kbits
  - can be exchanged within 1 s using the RF channel (Pump to Carelink USB Frequency: 961.5 MHz, Bandwidth: 185 kHz)

Conclusion

- Proposed a voiceprint-based access control scheme comprising
  - anti-replay speaker verification
    - The insulin pump can be accessed by the Carelink USB after the legitimate user passed the identity verification
  - voiceprint-based key agreement
    - The pump established a secure channel with the device in its close proximity
- Evaluated the performance of cascade fusion of ASV and CM
- Demonstrated the feasibility of energy-difference-based voiceprint extraction and secure multi-party computing-based key agreement scheme

Thank you!