Research Statement

Yili Ren (ren@cs.fsu.edu)
Department of Computer Science
Florida State University

1. Introduction

My research interests span two major areas: (1) Mobile Computing and (2) Cybersecurity. Specifically, my mobile computing research focuses on integrating the sensing capabilities of wireless and mobile devices to build new classes of Internet of Things (IoT) and smart home applications, as well as enable novel technologies for human-computer interaction (HCI). My cybersecurity research is concerned with mobile authentication leveraging mobile and wearable devices, as well as smart home and Internet of Things security with an emphasis on discovering novel human biometrics and inventing biometric sensing techniques. I enjoy sensing and understanding the physical world with multi-modal signals (e.g., radio frequency, light, acoustic, magnetic, etc.) and machine learning techniques, which I believe will be an important direction for my future research.

While engaging in research on mobile computing and cybersecurity, I have been actively involved in collaboration with multiple research groups and universities (WINLAB at Rutgers University, DAISY Lab at Rutgers University, and Trinity University). My work during my Ph.D. studies at Florida State University has resulted in multiple publications in prestigious conferences and journals such as USENIX Security, ACM SenSys, ACM UbiComp, ACM ACSAC, and IEEE IoT Journal. I have published 21 publications with 5 full papers in top-tier csrankings.org conferences as the first author or co-first author. Moreover, I received the Graduate Student Research Award (1 out of 85 Ph.D. students) in the Department of Computer Science at Florida State University. I also received Merit Scholarships when I was a master's student at Beihang University. For the rest of this statement, I will introduce the background of my research, highlight the topics I have worked on or still be working on, and present my research vision for possible future work.

2. Research Direction I: Mobile Computing

2.1 Smart Home Applications

Smart home enables the interconnections of ubiquitous devices planted in home appliances with artificial intelligence and the Internet of Things (IoT) for home automation. WiFi is one of the key technologies that enable connectivity for smart home services. Apart from its primary use for communication, the WiFi signal has now been widely leveraged for various sensing tasks due to its sensitivity to human and environmental dynamics. Building smart home applications based on WiFi sensing could achieve cost-effective, non-invasive, and massive deployment.

Intelligent Surveillance. Nowadays, the intelligent surveillance system has played an essential role in maintaining the safety and stability of society. An intelligent surveillance system can provide a wide range of security applications including person detection, person tracking, etc. Among these applications, person re-identification (Re-ID) is a fundamental one that identifies individuals across times and locations. Traditional methods for person Re-ID mainly rely on computer vision approaches and they achieved great progress using deep learning. However, vision-based systems have several limitations such as non-line-of-sight scenarios and user privacy concerns. Recently, WiFi devices have become more and more ubiquitous in public and private places. We also note that the number of WiFi devices is much larger than that of surveillance cameras. More importantly, as WiFi signals and visible lights are both electromagnetic waves, we ask the question that whether we can leverage the more pervasive WiFi signals to illuminate the human body and analyze the reflections for person Re-ID or to augment traditional camera-based surveillance systems. Therefore, we design a WiFi vision-based approach to "see" the person and thus achieve person Re-ID.

We leverage the advancement of WiFi technology and deep learning to help WiFi devices "see", identify, and recognize persons as we humans do. Our system helps WiFi devices "see" a person by leveraging the multiple spatially distributed antennas on WiFi devices, and the two-dimensional angle of arrival (2D AoA) estimation of the WiFi signal reflections. With spatially distributed antennas at the WiFi receiver, the signal reflections from the different directions could be separated with signal processing techniques, providing the theoretical foundation to derive spatial information of the physical space. Second, we leverage the 2D AoA of the signal reflections to visualize a person in the physical space. Once the WiFi devices "see" a human, we propose to leverage the deep learning models to digitize 2D AoA images of the human into a 3D human body representation. We then extract the intrinsic features of a person including both the static body shape and dynamic walking patterns for person Re-ID. We experimentally evaluate our system with twenty-eight people in various indoor environments including the laboratory, classroom, and home. Experimental results show that our Re-ID system is highly accurate across time and space.

Liquid Level Sensing. The popularity of IoT offers unprecedented opportunities to enable a variety of emerging applications in the context of smart homes. We propose to reuse home WiFi to estimate the liquid levels in containers, which provides critical information for building a smart home and mobile healthcare applications such as daily calory ingest tracking, household inventory checking, water consumption measurement, etc. The rationale behind our work is that the resonance frequencies of a container vary with its liquid level change. Moreover, when a container vibrates at its resonant frequencies, the vibrating level is the largest and can be sensed by WiFi signals. Based on this knowledge, we propose LiquidSense, a liquid level sensing system with commodity WiFi. Specifically, we mounted a low-cost transducer to the surface of the container and made the transducer emit a well-designed chirp signal. When the chirp signal sweeps to the frequency of resonance, the strenuous vibrations of the container cause phase changes in the WiFi signals. Therefore, we could decide the resonant frequency of a container by monitoring WiFi signals' phase changes and use that as the feature for liquid level decisions. Based on a few training data, we designed prediction models for both continuous and discrete predictions using curve fitting and SVM respectively. According to the experimental results, LiquidSense is highly accurate for sensing various types of liquids' levels in different materials of containers. Moreover, benefiting from the penetration of the WiFi signal, LiquidSense can work in non-line-of-sight scenarios.

2.2 Human-Computer Interaction Application

3D human body (i.e., pose and mesh) tracking/reconstruction has recently attracted increasing attention due to its widespread applications in a great variety of areas, such as human-computer interaction (HCI), robot, human motion analysis, surveillance, autonomous driving, etc. Inferring the 3D human body from optical images is a classic task in computer vision, but we explore the possibility of estimating the 3D human body with WiFi signals for several reasons. First, both visible lights and WiFi signals are electromagnetic waves and they have similar physical properties. Second, WiFi sensing can work in non-line-of-sight (NLoS) and poor lighting conditions. Third, using WiFi to sense humans involves fewer privacy concerns. At last, WiFi sensing can be easily integrated into IoT and smart home environments without additional hardware. We also aim to push the limit of WiFi sensing capability. Therefore, we accomplish 3D human body tracking and reconstruction using WiFi sensing approaches with an increasing level of difficulty: 3D human pose tracking for free-form activities, 3D human pose estimation for moving subjects, and 3D human mesh estimation.

3D Human Pose Tracking. While proliferating WiFi networks are intended for wireless communication, they have the potential to capture the human body and activity. A variety of WiFi-based systems have been proposed including large-scale activity sensing (e.g., walking and exercise), small-scale movement sensing (e.g., finger gesture), indoor localization, etc. More recent work has evolved towards constructing a 3D human pose that consists of a set of joints of the body at an unprecedented level of granularity. However, existing WiFi-based 3D human pose tracking is limited to only a set of predefined activities as it relies on the pre-trained model of known activities. Thus, we propose Winect, a 3D human pose tracking system for free-form activity using commodity WiFi devices. The basic idea of our system is to combine signal separation and joint movement decomposition to enable 3D free-form activity tracking. In particular, we separate the multi-limb signals based on blind source separation (BSS). Next, we can derive the position of each limb over time and infer the trajectory in 3D space with multiple transmitter-receiver pairs by leveraging phase changes of separated signals. Finally, we leverage deep learning to decompose the trajectory of each limb (e.g., the arm or leg) to the fine-grained trajectories of the joints (e.g., the wrist and elbow, or the ankle and knee) for 3D pose tracking. The evaluation results show that our system can track 3D human poses with centimeter-level accuracy for various free-form activities under challenging environments.

The existing WiFi-based 3D human pose estimation systems require the user to perform activities at a fixed position as a moving user can lead to complex signal reflections and environmental changes. However, many activities in real life involve the change of users' locations. Therefore, we propose GoPose, a novel 3D skeleton-based human pose estimation system for moving subjects using WiFi devices. Our system works for unseen activities even when the user is moving around. Specifically, our system extracts the two-dimensional (2D) angle of arrival (AoA) of the incident signals, which can represent the spatial information of different body parts or joints regardless of activities or user positions. Thus, we can leverage the spatial characteristic of the 2D AoA spectrum to separate the signals bounced off the human body from the ones reflected off the static environments. We then utilize the CNN-LSTM model to transfer the 2D AoA spatial information to the 3D joint locations in physical space. We conduct experiments under non-line-of-sight scenarios, different distances between WiFi devices, etc. Results show that our system is highly accurate in constructing 3D human poses for moving users even for unseen activities.

3D Human Mesh Reconstruction. 3D human mesh parameterizes the 3D surface of the human body, which represents how individuals vary in height, weight, somatotype, body proportions, and how the 3D surface deforms with articulation. Unlike the 3D human pose estimation only predicts a dozen joints, the 3D human mesh describes the fine-grained 3D human body shape with thousands of vertices, which is several orders of magnitude larger than the number of body joints defined in the 3D pose. In this work, we demonstrate that the commodity WiFi can be utilized to construct 3D human mesh and we propose a WiFi vision-based approach to 3D human mesh estimation. Our system utilizes WiFi devices to visualize a person by leveraging the fairly large number of antennas on the next-generation WiFi devices. This is done by estimating the 2D AoA image of the WiFi signal reflections off the human body. To increase the spatial resolution of the 2D AoA estimation, we further leverage the frequency diversity of the OFDM subcarriers, the spatial diversity of antennas at the transmitter, and the time diversity of the WiFi packet. Also, we design deep learning models to extract both the spatial body shape and temporal body deformation features from 2D AoA images. Then, both features are fitted into the SMPL model to obtain the 3D mesh representation of the human body. Extensive evaluations show that our system is highly accurate and robust across different environments and for unseen people.

3. Research Direction II: Cybersecurity

3.1 Mobile and Wearable Security

Advances in mobile and wearable hardware and communication technology overturn human information transmission and storage modes from physical interactions to the anytime-anywhere service of mobile internet, which connects users and all kinds of personal information, including privacy and credentials, via wireless devices. Meanwhile, wearables are rapidly emerging as a new platform encompassing a broad range of personal applications due to their rich human body sensing capability. To secure mobile devices and accessible information, many protection mechanisms, such as PINs, passwords, and biometrics, have been adopted. However, directly adapting to traditional authentication can be challenging. This is simply because wearables lack a suitable input interface to support rapid and reliable entry of passwords or most of the traditional biometrics. Voice-based authentication is convenient for wearables but has also been proven vulnerable to voice spoofing attacks. Despite the issue, wearables also provide novel opportunities to improve or redesign approaches to authentication due to their rich human body sensing capability.

Hence, we developed and designed two security enhancement solutions for user authentication on mobile devices. Particularly, we take advantage of the sensing capabilities of ear wearables and explored the physiological and behavioral characteristics of human speech production, and discovered two new biometrics: acoustic tooth gestures and ear canal deformations. These biometrics could be applied for user authentication, identification, or transparent liveness detection for user authentication to enhance wearable and mobile security. Moreover, we invented novel sensing methods to collect these biometrics with existing hardware on average wearable devices. Compared with existing work, our solutions are built on highly secure new biometrics, which is hard to be observed, stolen, or forged. Moreover, our solutions require only the wearable with no extra devices or cumbersome user operations.

Earable Authentication via Acoustic Toothprint. The ear wearables (i.e., earables) enable rich around-the-head sensing opportunities. Here we utilize these sensing advantages and discover an anti-spoofing, widely accepted, and implicit biometric authentication. We design a secure authentication system that leverages the toothprint-induced sonic effect produced by a user performing teeth gestures for earable authentication. In particular, when teeth slide or strike against each other, part of their mechanical energy is transformed into a sonic wave. The harmonics of the friction- and collision-excited sonic waves are dependent on the teeth composition, the dental geometry, and the surface characteristics of each tooth. The key insight is that the sonic waves produce from a teeth gesture (either slide or tap) carry the information of the toothprint. As every individual has a unique toothprint just like our fingerprint, two users performing the same teeth gesture will result in distinct toothprint-induced sonic waves, which could be sensed by the earables for user authentication. We validate our system through both experiments and user surveys and prove that toothprint could is biometric and differentiate different users.

Ear Canal Deformation-based User Authentication. We are also working on a security enhancement solution for emerging applications such as Virtual Reality (VR) and Audio Augmented Reality (AAR), where the users normally wear headsets and earphones to obtain immersive experiences. In these cases, the traditional authentication method relying on PIN input with a virtual keyboard is vulnerable to recording attacks. Thus, we propose ear canal deformation-based user authentication to provide a secure, continuous, and transparent security enhancement solution. The basic idea of this system is to measure the ear canal deformations caused by human articulation. Specifically, when the user speaks, different articulators' movements lead to characteristic ear canal deformations. Therefore, our system keeps sending a well-designed inaudible beep signal to the ear canal via the built-in microphone on the earbud and collecting the reflected signals for ear canal deformation analysis. Our preliminary study proves that ear canal deformation is biometric, which could differentiate different users and can be applied for user authentication.

3.2 Internet of Things Security

The rise of the IoT era has been drawing exponential popularity to voice user interface (VUI) capable devices, which are widely adopted to connect and control extensive physical IoT devices, systems, and objects via voice commands. For example, in a smart home environment, the user could control the house security systems, appliances and manage personal businesses via a smart speaker. In a smart vehicle environment, the user may require navigation help or even hands-free driving service via the in-car voice assistant system. A typical VUI-capable device's workflow composes of two main phases: the activation phase and the speech recognition phase. Some devices will be activated if the user could pass voice authentication, whereas others only require the user to speak the right wake words. An activated device may execute any commands from any user or only execute commands from specific users who possess enrolled voice biometrics. Unfortunately, since voice authentication is vulnerable to replay attacks, an attacker could spoof the VUI-capable device at both phases. After hijacking the VUI-capable devices, the attacker may access confidential information, privacy, and even mislead other IoT subjects to cause severe hazards like house security system failure and smart vehicle misleading. Therefore, to protect the VUI-capable device and other IoT devices, a continuous liveness detection that could protect the whole communication session is imperative.

Although the aforementioned user authentication solutions are well suited for mobile devices when the users keep the device close to themselves, these solutions could not be extended to universal VUI-capable devices in IoT environments like the smart home or the smart vehicle. In these use cases, the users could be half a meter or even meters away from the VUI. Moreover, the first two solutions require user enrollment and do not support continuous liveness detection. Additionally, the third solution requires the user to wear earbuds all the time, which could be inconvenient or even illegal in many scenarios. Therefore, we design an adaptive liveness detection system to support continuous and long-range liveness detection for secure VUI-capable devices in various use cases.

Bone-conducted Vibrations-based Liveness Detection. When humans articulate, both air-conducted voices and bone-conducted vibrations are generated. Although sharing the same sound source of vocal cords, these signals are different as they pass through different media and modulations. By contrast, when a loudspeaker replays a recording, the rigid body vibrations and the replayed sounds are always exactly the same as the input signal. Therefore, we proposed to compare the vibrations with the audio signal spoken or replayed by a live user or a loudspeaker for liveness detection. To catch such subtle vibrations, we emitted a probe signal via the built-in microphone to the vibrating human head/loudspeaker and received the modulated reflections with the built-in loudspeaker on the VUI-capable device. In particular, when the probe signal travels from the microphone to the human head/loudspeaker, the propagation distances vary due to the displacements of the human head/loudspeaker, and therefore the power of the probe signal changes accordingly. We considered this procedure as amplitude modulation, where the probe signal is the carrier wave, and it is modulated by the vibration signal. Based on this theory, we extracted the vibration signals from the reflections successfully by calculating the power changes of every few samples in the time domain. Compared with the previous solutions, this solution does not require user enrollment and supports continuous liveness detection since we could always collect the air-conducted voices and the corresponding vibrations for comparison, regardless of the content of the voices. Moreover, this system could adapt to different scenarios including both short-range applications on smartphones and long-range services in smart homes or smart vehicles.

4. Future Research Outlook

Most of my current projects focus on developing mobile computing and wireless sensing techniques to construct emerging IoT applications and improve the security of mobile or IoT devices. I intend to keep exploring these directions while absorbing novel technologies to solve sensing and security issues in everyday life.

4.1 Radio Frequency Vision

Radio frequency (RF) vision is an interesting and challenging problem in the evolving community of wireless sensing. Pervasive person and object sensing would be enabled in rich contexts (e.g., location, silhouette, pose of person and object) by RF vision. Thus, it could support new applications such as exercise assessment, interactive gaming, and human recognition in a privacy-preserving and cost-effective way. Although our work has obtained preliminary results on human visualization using WiFi signs, wireless vision is worth exploring and improving. To better visualize the target in complex environments, I propose to explore multi-band wireless fusion approaches to image the target as multi-band signals have complementary characteristics. For example, mmWave signals can achieve better accuracy but WiFi signals have a larger sensing range. Such a fusion-based approach could improve the robustness and accuracy of the system. Moreover, I plan to extend 2D AoA information-based wireless vision to higher dimensional information by jointly estimating the time of flight (ToF) and Doppler frequency shift (DFS). Therefore, the extended 4D information will consist of three dimensions of static information (i.e., azimuth, elevation, and distance) and one dimension of dynamic information (i.e., frequency). With such comprehensive information, we may extend the wireless sensing capacity even beyond vision.

4.2 Semantic Sensing Using WiFi Signals

In recent years, a number of WiFi sensing systems have achieved detection, localization, and recognition for human or human activities. However, semantic sensing is more promising to support a wider range of applications. In particular, we define the term semantic sensing as the task to analyze a scene by considering the semantic context of its contents and the intrinsic relationships between them. In other words, the semantic sensing system is capable of understanding and describing the scene in a human-understandable form. For example, Let's consider the scenario of a family gathering with multiple family members spread through different rooms and performing various activities (e.g., playing, cooking, and talking). I aim to propose semantic sensing that can sense both the activities of the children and their body language, thus we can detect potential bullying occurrences. Moreover, through mouth motion and posture sensing, it is possible to perceive if the person in the conversation is agitating or not. By leveraging object sensing and the motion of the user, the system can guide the person who is trying to cook a meal where the specific items are located in the kitchen, and if a restock of groceries or condiments is needed. Although semantic sensing is challenging due to the complex interactions between objects, complex environments, different viewpoints, and the inherent ambiguity in the limited information provided by a given scene, incorporating semantic sensing abilities into the WiFi sensing system can be a long-standing goal and desirable research topic in the future.

4.3 Multi-modal Sensor Fusion and Sensing for IoT Security Enhancement

Humans have great perceptions of the surroundings due to our ability to cross-modal perception with the five senses. Thus, by combining the wearable, acoustic, IoT devices, and WiFi sensing approaches, we could come up with novel and practical applications with multi-modal signals to effectively respond to complex environments. Inspired by this observation, I propose to improve the perception of the IoT environment with multi-modal sensing for IoT security enhancement. By utilizing the multi-modal sensor in the house, we could detect the owner/intruder of the house. When someone knocks on the door, we could capture the biometric information such as the arm propagation sound by the wearables of the person. We could also estimate the person’s height based on the propagation time of the knocking sound to each VUI-capable device. The WiFi signals could also capture general information about the person’s body. It could also detect if the front door is open, and then we could deploy the VUI-capable devices to sense the moving track of this person and analyze if his/her movements follow any family members’ routine. Last, our system could make the decision based on these three dimensions of information. A person could be an intruder, if her/his height is different from all the family members, she/he opens the door after knocking, and after entering the room, his/her moving trajectories do not match with any family members' routines. Specifically, I am interested in user authentication and liveness detection under multi-modal sensing in IoT, healthcare based on multi-modal IoT sensing, and multi-modal sensor data fusion.

I believe that my broad background will enable me to continue making contributions to research fields not limited to mobile computing and cybersecurity. I look forward to collaborating with researchers not only from other computer science areas such as AI, robotics, and embedded systems but also from areas other than computer science such as healthcare, education, and statistics. I hope to blend my knowledge with others and provide better solutions to everyday problems.

References.

[1] **Yili Ren**, Yichao Wang, Sheng Tan, Yingying Chen, Jie Yang. “Person Re-identification in 3D Space: A WiFi Vision-based Approach”, The 32nd USENIX Security Symposium **(USENIX Security 2023)**, full paper, 18 pages.

[2] Yichao Wang*, **Yili Ren***, Yingying Chen, Jie Yang. “Wi-Mesh: A WiFi Vision-based Approach for 3D Human Mesh Construction”, The 20th ACM Conference on Embedded Networked Sensor Systems **(SenSys 2022)**, full paper, 15 pages. (*co-first authors)

[3] **Yili Ren**, Zi Wang, Yichao Wang, Sheng Tan, Yingying Chen, Jie Yang. “GoPose: 3D Human Pose Estimation Using WiFi”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **(UbiComp 2022)**, full paper, 25 pages.

[4] **Yili Ren**, Zi Wang, Sheng Tan, Yingying Chen, Jie Yang. “Winect: 3D Human Pose Tracking for Free-form Activity Using Commodity WiFi”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **(UbiComp 2021)**, full paper, 29 pages.

[5] **Yili Ren**, Sheng Tan, Linghan Zhang, Zi Wang, Zhi Wang, Jie Yang. “Liquid Level Sensing Using Commodity WiFi in a Smart Home Environment”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **(UbiComp 2020)**, full paper, 30 pages.

[6] Zi Wang, **Yili Ren**, Yingying Chen, Jie Yang. “ToothSonic: Earable Authentication via Acoustic Toothprint”, Proceedings of the ACM on Interactive, Mobile, Wearable, and Ubiquitous Computing **(UbiComp 2022)**, full paper, 24 pages.

[7] Zi Wang, Sheng Tan, Linghan Zhang, **Yili Ren**, Zhi Wang, Jie Yang. “EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables”, Proceedings of the ACM on Interactive, Mobile, Wearable, and Ubiquitous Computing **(UbiComp 2021)**, full paper, 27 pages.

[8] Sheng Tan, **Yili Ren**, Yingying Chen, Jie Yang. “Commodity WiFi Sensing in 10 Years: Status, Challenges, and Opportunities”, **IEEE Internet of Things Journal**, 2022, full paper, 12 pages.

[9] Linghan Zhang, Sheng Tan, Zi Wang, **Yili Ren**, Zhi Wang, Jie Yang. “VibLive: A Continuous Liveness Detection for Secure Voice User Interface in IoT Environment”, 36th Annual Computer Security Applications Conference **(ACSAC 2020)**, full paper, 13 pages.