WavePurifier: Purifying Audio Adversarial Examples via Hierarchical Diffusion Models

Hanqing Guo[1,2], Guangjing Wang, Bocheng Chen, Yuanda Wang, Xiao Zhang, Xun Chen[3], Qiben Yan[1], Li Xiao[1]

MICHIGAN STATE UNIVERSITY

UNIVERSITY of HAWAI'I MĀNOA

SAMSUNG RESEARCH AMERICA

Dr. Hanqing Guo

Assistant Professor
Department of Electrical and Computer Engineering
Department of Information and Computer Sciences
University of Hawai'i at Mānoa
Office: Room 437, Holmes Hall
Email: guohanqi at hawaii dot edu

Welcome

I am an assistant professor at the University of Hawai'i at Mānoa, appointed by Electrical & Computer Engineering Department and Department of Information and Computer Sciences. I received my Ph.D. in Computer Science & Engineering at Michigan State University, US, in 2024, an M.S. in Computer Science from Ball State University (fully funded with Research Assistant Scholarship), US, in 2019, and a B.S. in Communications Engineering from Chongqing University of Posts of Telecommunications, China, in 2015. Prior to joining the University of Hawai'i, I was a research scientist at Samsung Research America (Knox security team) and Amazon (ML security team based in San Diego). My research interest includes trustworthy AI, mobile system & sensing. In particular, I am interested in research real-world AI applications, such as voice ID systems, smart speakers, mobile applications, watermarks, and wireless communication systems. I try to find the Vulnerabilities and potential fix solutions.

Openning: I am actively seeking highly-motivated students for Ph.D. or Research Intern positions. If you are interested, please send me your CV, transcripts, and brief descriptions your research interest

News & Updates

- 03/2025 Paper accepted: "ClearMask: Noise-Free and Naturalness-Preserving Protection Against Voice Deepfake Attacks". Accepted in **AsiaCCS 2025**.
- 04/2025 I was invited to serve a Program Committee for **MobiSys 2025 Artifact Evaluation**
- 03/2025 Service: I was excited to be invited to serve a Program Committee for **NDSS 2026**.
- 03/2025 Paper accepted: "Demo: Disrupting In-Car mmWave Sensing Through IRS Manipulation". Accepted in **S&P Workshop SecureTrans 2025**. See you on S&P 2025 in San Francisco!
- 03/2025 Service: I was invited to serve a Technical Program Committee for Human-Centered Sensing, Modeling, and Intelligent Systems **HumanSys 2025**.
- 02/2025 Service: I was invited to serve a Program Committee for **UbiComp/IMWUT 2025**.
- 01/2025 Service: I was invited to serve a Demos Chair for **EAI SmartSP 2025**. Check out the Conference Website and Welcome to submit Demos!
- 1/2025 Paper accepted: "AUDIO WATERMARK: Dynamic and Harmless Watermark for Black-box Voice Dataset Copyright Protection". Accepted in **USENIX Security 2025**. Check demo and code [here!]
- 11/2024 Paper accepted: "Secure-IRS: Defending Against Adversarial Physical-Layer Sensing in ISAC System". Accepted in ICNC 2025. Congrats to Ziyu for his first research paper! Check demo [here!]
- 11/2024 NSF Workshop: I have been accpeted to attend the **NeTS Early Career Investigator Workshop** on Jan. 15-16, 2025 at NSF Headquarters.
- 10/2024 Invited Talk: I was invited to give a talk about "Trustworthy AI and It's Applications" in **George Mason University**
- 10/2024 Paper accepted: "PiezoBud: A Piezo-Aided Secure Earbud with Practical Speaker Authentication". Accepted in **ACM SenSys 2024**.
- 10/2022 New award: OVPRS Faculty Research Travel Fund 🏆. I am excited to receive **UH OVPRS Faculty Research Travel Fund**! Can't wait to attend the Mobicom 2024 in DC!
- 10/2024 Invited Talk: I was invited to give a talk about "Towared the Secured Speech AI Services" in **University of Delaware**
- 10/2024 Paper accepted: "WavePurifier: Purifying Audio Adversarial Examples via Hierarchical Diffusion Models". Accepted in **Mobicom 2024**.
- 10/2024 Paper accepted: "Protecting Activity Sensing Data Privacy Using Hierarchical Information Dissociation". Accepted in **CNS 2024**.

Assistant Professor

BEH 311 | 813-396-0629

Email | **Google Scholar**| **LinkedIn**

Biography

Dr. Guangjing Wang joins the Department of Computer Science and Engineering (CSE) as a tenure-track assistant professor in the Fall of 2024. He earned his Ph.D. in Computer Science from Michigan State University in 2024. Prior to that, he obtained his master's degree in computer science from the University of Science and Technology of China in 2020, and his bachelor's degree in computer science from Southwest University in 2017. Dr. Wang's primary research work lies in data-centric AI, with interdisciplinary studies involving mobile sensing and IoT data management. His research work has been published in top-tier conferences such as SIGMOD, ICDE, MobiCom, MobiSys, UbiComp, and INFOCOM.

Research Interests

LLM Agents: Exploring large language models for various applications.

Sensing and its data management: Leveraging various devices to collect, analyze and manage data from the physical world.

Follow

Manage Profile

Bocheng Chen

Affiliation

Department of Computer Science
Michigan State University
East Lansing, MI, USA

Publication Topics

Accuracy Of Model, Accuracy Of The Final Model, Action Recognition, Actuator, Behavioral Model, Behavioral Rules, Call Graph, Capabilities Of Devices, Chain Coordination, Chain Interactions, Client–side, Concrete Examples

Show More

Biography

Bocheng Chen received the BS degree in electronic science and technology from Shanghai Jiaotong University, Shanghai, China, in 2020. He is currently working toward the PhD degree with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI. His research interests include mobile security, AI security, and IoT security. *(Based on document published on 31 May 2022).*

Yuanda Wang

Hello all, I am Yuanda Wang, a final-year PhD student from Michigan State University.

San Jose, California, USA

Email

Resume

LinkedIn

Google Scholar

ORCID

About me

I am a final year PhD candidate from SEIT lab, department of Computer Science and Engineering, Michigan State University. My research is focusing on machine learning & security and privacy, especially threats and safety topics related to speech AI (CCS'22, WiSec'23, RAID'23, MobiCom'24, AsiaCCS'25, USENIX Security'25), side-channel attacks (NDSS'22), LLM safety(RAID'23, Arxiv'24), and IoT (MobiCom'20, CNS'24). I am advised by Dr. Qiben Yan. Currently, I am an AI Security Research Scientist Intern at ByteDance, San Jose, where I work on leveraging GenAI/LLM to detect threats and enhance security in cloud services and firewalls.

Before my PhD career in MSU, I obtained my BEng from Xi'an Jiaotong University in 2016 and MEng from North China Electric Power Univesity (Beijing) in 2019.

I'm also exploring new opportunities in Software Engineer, Machine Learning Engineer, and Applied Scientist roles within the U.S. market. If you'd like to connect professionally, feel free to reach out via LinkedIn or contact me directly through email.

Industry Experience

ByteDance Inc., AI Security Research Scientist Intern, San Jose, Feb 2025 - Now.

Samsung Research America, Research Scientist Intern, Mountain View, Sep 2022 - Dec 2022.

Publications

My full paper list is here.

Awards

Dissertation Completion Fellowship (DCF) 2024, Michigan State University

Best Paper Honorable Mention Award, CCS 2022.

Student Travel Grant, CNS 2020.

Home

News

Research

Teaching

Students

Awards

Services

Xiao Zhang

Office: 212 CIS Bld. | Lab: 108 CIS Bld. 4901 Evergreen Road, Dearborn, MI 48128

I am an Assistant Professor in the Department of Computer and Information Science at the University of Michigan–Dearborn, where I lead the Trustworthy AI–boosted IoT Lab (**TAI Lab**). Before joining UM–Dearborn, I was a Postdoctoral Associate at Duke University. I obtained my Ph.D. degree in the Computer Science and Engineering at Michigan State University, my M.S. degree at Northwestern Polytechnical University and B.E. degree with honor at Taiyuan University of Technology. **I am looking for highly self–motivated students (Ph.D., master, undergraduate, visiting, intern, and high school) to join my group**. If you are passionate about related research topics, please email me with your CV, transcripts, and several sentences about your research interests.

My current research interests are ***mobile computing***, ***AIoT***, ***Cyber Physical Systems***, ***AI–assisted sensing/localization***, **and** ***HCI***. My previous work focused on exploring spatial–temporal diversities in Optical Wireless Communication (OWC) for its improved communication performance and enabled sensing/localization. My vision is that light can provide the secure and location–aware communication for Internet–of–Things and human–centered mobile computing. With explored OWC's spatial–temporal diversities and machine learning, we can use light as promising medium for next–generation wireless networks with broad applications (e.g., LiFi, V2X networks, underwater navigation, digital health, smart city, HCI, and AR/VR).

Umich Email / Google Scholar / Michigan Experts

Qiben Yan

Associate Professor

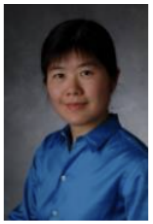
Keep you posted! We create "Security Posts" for you, researchers!

Computer Science and Engineering
Michigan State University
Email: qyan AT msu DOT edu
Office: 3115 Engineering Building
Phone: +1(517)432-3529
Public Key
SEIT Lab - Secure and Intelligent Things Lab
Google Scholar

We are always looking for PhD students to join exciting research projects on cyber security!
Please email your CV, Transcripts, Publications (if any) to me for opportunities.
Visiting/exchange scholars and students are also welcome to contact us.

Qiben Yan is an Associate Professor in the Department of Computer Science and Engineering at Michigan State University, where he founds and directs the SEIT (Secure and Intelligent Things) Lab. Dr. Yan is a recipient of the NSF CISE Career Research Initiation Initiative (CRII) Award. He currently serves as Associate Editor, IEEE Transactions on Information Forensics and Security (TIFS). Prior to joining academia, he worked in a cybersecurity startup company, Shape Security, in the silicon valley, where he participated in building the first "botwall". He received a Ph.D. in Computer Science from Virginia Tech, an M.S. and a B.S. degree in Electrical and Computer Engineering from Fudan University in Shanghai, China.

His research uses AI, acoustics, blockchain technologies for enhancing security of connected devices, networks and systems. His primary research interests are in IoT security, mobile security and privacy, wireless system security, AI security and privacy, and botnet and malware detection. A key thread of his research is the design of secure network infrastructure for Connected Things - including system/device security analysis, anomaly detection system, traffic monitoring and analysis system, attack resilient communication system - to provide security enhancement and protection for the massive IoT systems/networks under cyber attacks. He has been awarded the Best Paper Honorable Mention Award (Best Paper Runner-up) in ACM CCS 2022, Best Paper Award in IEEE SECON 2021 and Best Paper Award in ACM SenSys 2021. He has published papers in top-tier conferences such as CCS, NDSS, USENIX Security, MobiCom, SenSys, INFOCOM, etc. His research has been reported in various high-impact media outlets, including the BBC Radio, Scientific American, Science Daily, Forbes, Popular Mechanics, Gizmodo, The Register, etc. He is a Senior Member of the IEEE and a Member of the ACM.



Li Xiao
Professor
Department of Computer Science and Engineering
3115 Engineering Building
Michigan State University
East Lansing, MI 48824

Phone: 517-353-4386
Fax: 517-432-1061
lxiao@cse.msu.edu

Research Interests

Distributed and networking systems, wireless and mobile computing, overlay systems and applications, system resource management, and design and implementations of experimental algorithms.

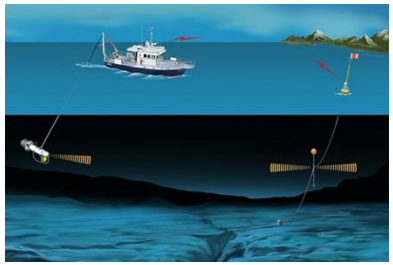
Research Lab

Experimental Laboratory for Advanced Networks and Systems (eLANS)

Selected Publications

- Y. Hu, M. Chen, H. Yan, C. Cheng, G. Tu, C. Li, T. Xie, C. Peng, L. Xiao, J. Tang, "Uncovering Problematic Designs Hindering Ubiquitous Cellular Emergency Services Access," *MobiCom 2024*.
- H. Guo, G. Wang, B. Chen, Y. Wang, X. Zhang, X. Chen, Q. Yan, L. Xiao, "WavePurifier: Purifying Audio Adversarial Examples via Hierarchical Diffusion Models," *MobiCom 2024*.
- X. Zhang, L. Xiao, M. Mutka, "HoloCube: 3D Opitcal IoT Connections via Software Defined Pepper's Ghost," *ICNP 2024*.
- J. Shi, T. Xie, G. Tu, C. Peng, C. Li, A. Hou, S. Wang, Y. Hu, X. Lei, M. Chen, L. Xiao, X. Liu, "When Good Turns Evil: Encrypted 5G/4G Voice Calls Can Leak Your Identities," *CNS 2023*.
- H. Guo, X. Chen, J. Guo, L. Xiao, Q. Yan, "MASTERKEY: Practical Backdoor Attack Against Speaker Verification Systems," *MobiCom 2023*.
- X. Zhang, G. Klevering, J. Wang, L. Xiao, T. Li, "RoFin: 3D Hand Pose Reconstructing via 2D Rolling Fingertips," *MobiSys 2023*.
- H. Guo, G. Wang, Y. Wang, B. Chen, Q. Yan, L. Xiao, "PhantomSound: Black-Box, Query-Efficient Audio Adversarial Attack via Split-Second Phoneme Injection," *RAID 2023*.
- H. Guo, Y. Wang, N. Ivanov, L. Xiao, Q. Yan, "SPECPATCH: Human-In-The-Loop Adversarial Audio Spectrogram Patch Attack on Speech Recognition," Best Paper Honorable Mention Award. *CCS 2022*.
- Y. Hu, M. Chen, G. Tu, C. Li, S. Wang, J. Shi, T. Xie, L. Xiao, C. Peng, Z. Tan, S. Lu, "Uncovering Insecure Designs of Cellular Emergency Services (911)," Best Community Paper Award Runner-Up, AT&T Security Award. *MobiCom 2022*.
- X. Zhang, J. Mariani, L. Xiao, M. Matt, "LiFOD: Lighting Extra Data via Fine-grained OWC Dimming," *SECON 2022*.
- J. Mariani, Y. Han, L. Xiao, "Co-Cache: Inertial-Driven Infrastructure-less Collaborative Approximate Caching," *SECON 2022*.
- H. Guo, C. Li, L. Li, Z. Cao, Q. Yan, L. Xiao, "NEC: Speaker Selective Cancellation via Neural Enhanced Ultrasound Shadowing," *DSN 2022*.
- X. Zhang, H. Guo, J. Mariani, L. Xiao, "U-Star: An Underwater Navigation System based on Passive 3D Optical Identification Tags," MOBICOM 2022.
- C. Li, Z. Cao, L. Xiao, "CurveALOHA: Non-linear Chirps Enabled High Throughput Random Channel Access for LoRa", *INFOCOM 2022*.
- S. Wang, G. Tu, X. Lei, T. Xie, C. Li, P. Chou, F. Hsieh, Y. Hu, L. Xiao, C. Peng, "Insecurity of Operational Cellular IoT Service: New Vulnerabilities, Attacks, and Countermeasures," *MobiCom 2021*.
- C. Li, H. Guo, S. Tong, X. Zeng, Z. Cao, M. Zhang, Q. Yan, L. Xiao, J. Wang, Y. Liu, "NELoRa: Towards Ultra-low SNR LoRa Communication with Neural Enhanced Demodulation," Best Paper Award. *Sensys 2021*.
- Y. Hu, S. Wang, G. Tu, L. Xiao, T. Xie, X. Lei, C. Li, "Security Threats from BitcoinWallet Smartphone Applications: Vulnerabilities, Attacks, and Countermeasures," *CODASPY 2021*.
- X. Zhang, L. Xiao, "Effective Subcarrier Pairing for Hybrid Delivery in Relay Networks," *MASS 2020*.
- M. Zarifneshat, L. Xiao, "Out-of-Band Multiple Path Discovery Protocol for Robust In-Band Millimeter Wave Links," MASS *2020*.
- M. Zarifneshat, L. Xiao, J. Tang, "Learning-based Blockage Prediction for Robust Links in Dynamic Millimeter Wave Networks," *SECON 2019*.
- M. Zarifneshat, R. Proteek , L. Xiao, "Multi-Objective Approach to Improve Load Balance and Blockage in Millimeter Wave Cellular Networks," *DySPAN 2019*.
- C. Liu and L. Xiao, "Interference and Blockage Prediction in mmWave-Enabled HetNets," *MASCOTS 2018*.

Acoustic + AI Applications

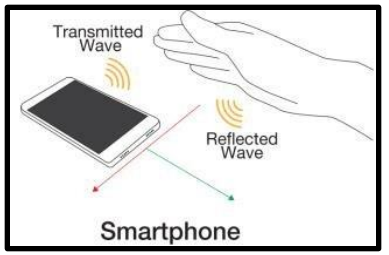


Underwater acoustic communication

Voice control in-car

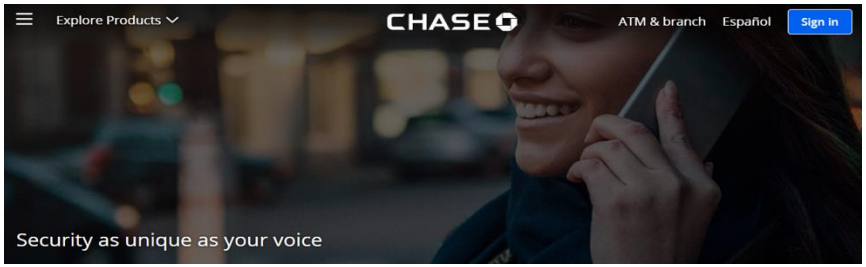
voice control in home

Voice ID



Gesture sensing

Acoustic monitoring



Voice authentication for customer service

Acoustic AI systems are vulnerable

Adversary

Adversarial Example

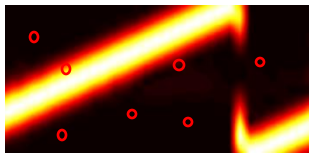
Speech Recognition

"Music"

↓

"**open the door**"

Adversary



Perturbations

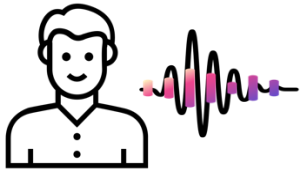
Signal Detection [1]

"0x40"

↓

"**0x45**"

Adversary



Attack

Speaker Authentication

"Bob"

↓

"**Evil**"

[1] SenSys 2021 NELoRa: Towards ultra-low SNR LoRa communication with neural-enhanced demodulation.

Existing Countermeasures

Input-Level Defense

Process the input to make perturbation invalid

- Feature Squeezing
- Input Preprocessing
- Randomization
- Adversarial Detection

Vulnerable to adaptive attacker

Model-Level Defense

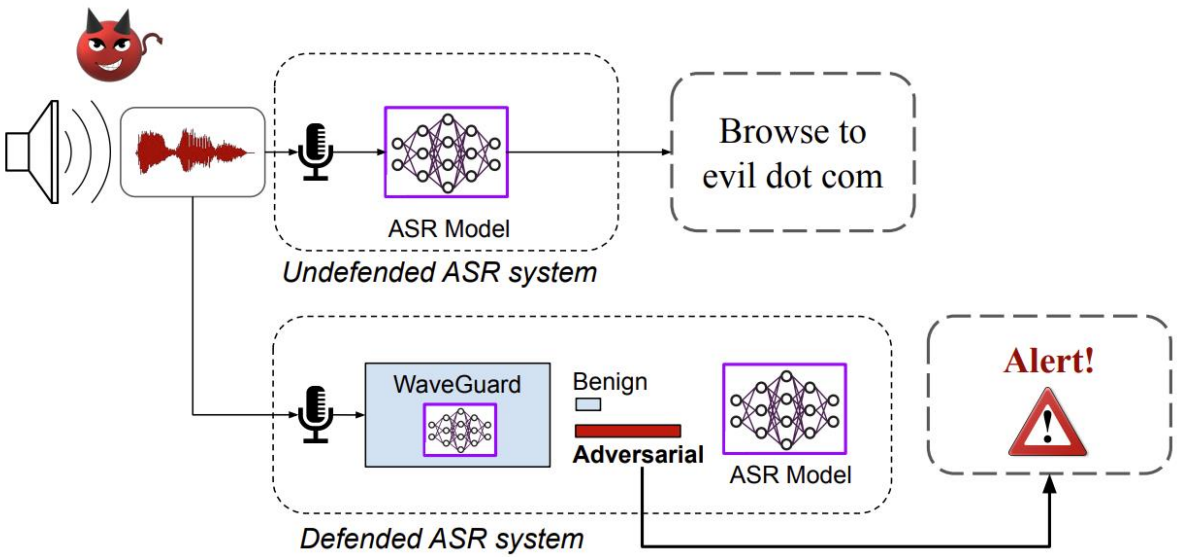
Make the model robust and hard to attack

- Adversarial Training
- Gradient Masking
- Defensive Distillation
- Ensemble Defense

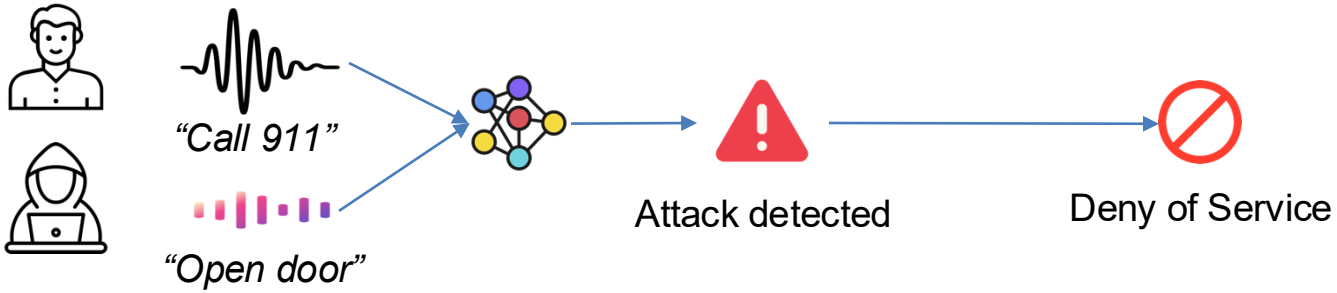
Computation intensive

Open attack interface

Prior Work – WaveGurad [2]

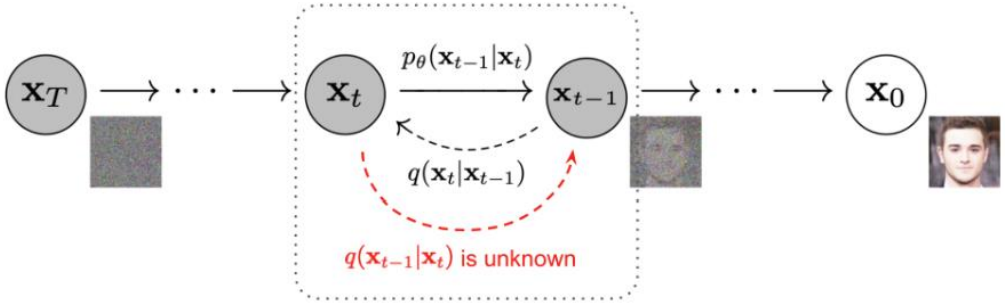


Vulnerable to adaptive attacker

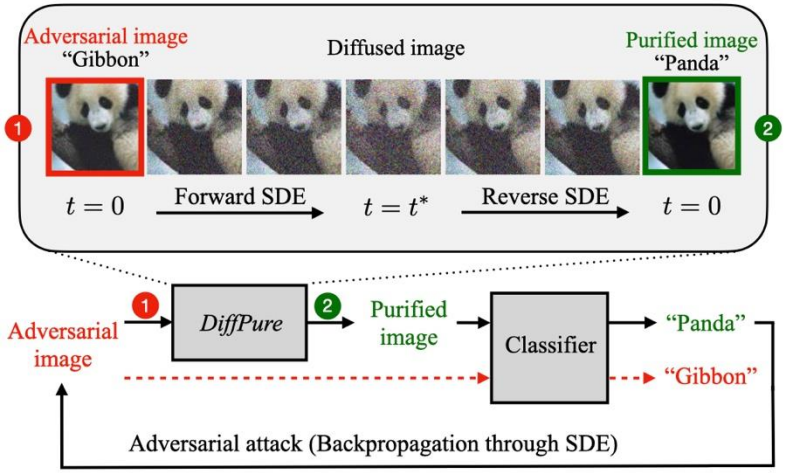


Block benign input

Inspiration



NeurIPS 2020 DDPM



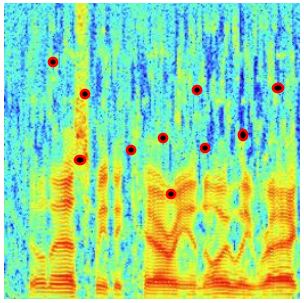
ICML 2022 Diffpure

Is it possible to purify the audio sample with diffusion model?

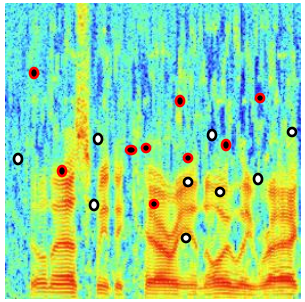
Challenge

1. How to determine the Diffusion Framework?
2. How to prove the existence of optimal purify step?
3. How to determine the optimal purification step?

Idea – Why diffusion model?



Forward Diffusion



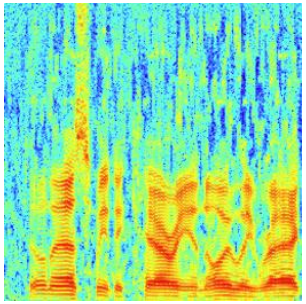
Adversarial Example

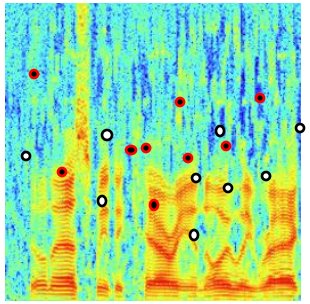
Benign sample

Add Noise to Disrupt Perturbation

Forward destroy perturbation

Backward Diffusion



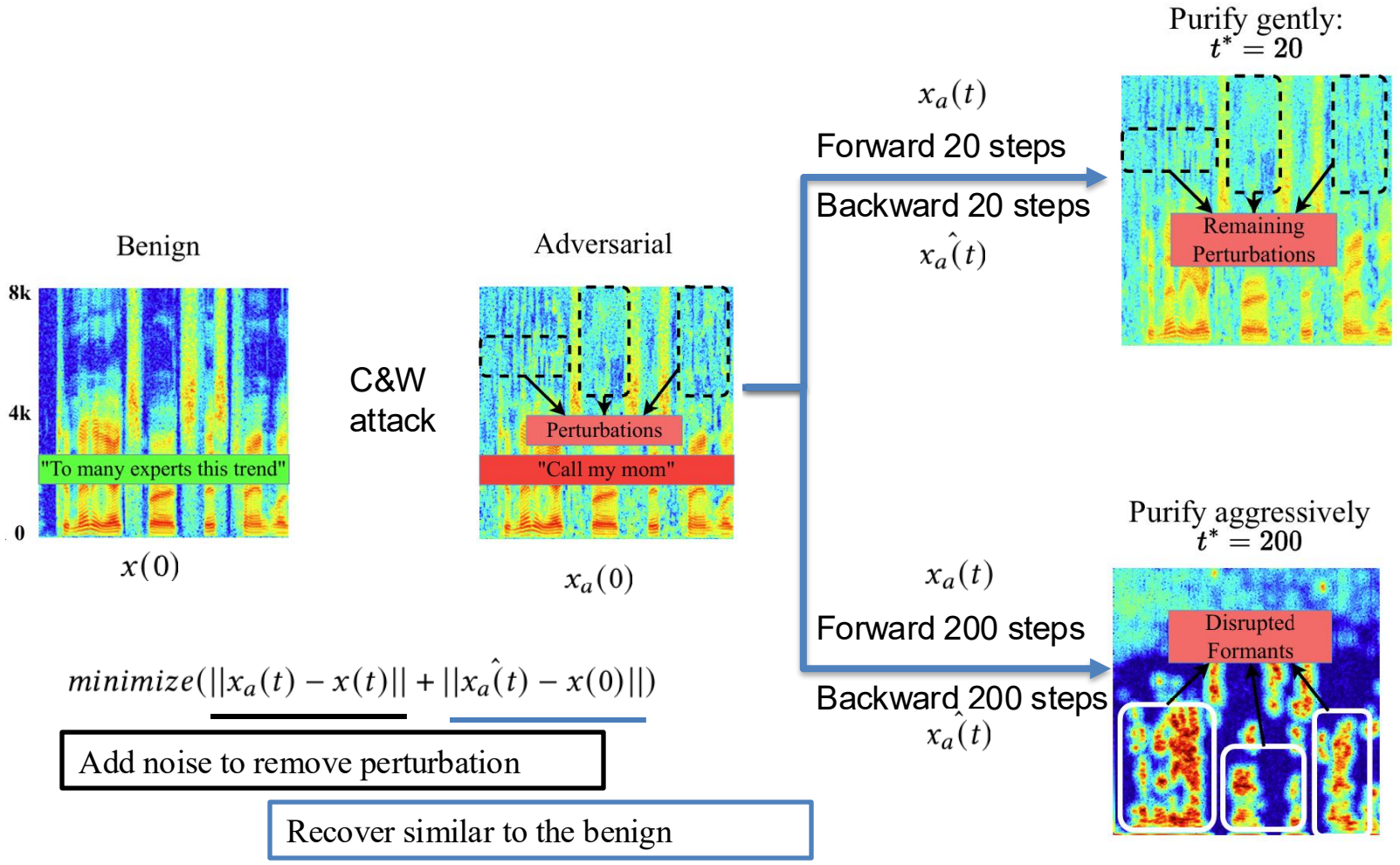


Benign Sample

Noisy Adversarial Example

Backward recover benign; Because the original diffusion model is trained on Clean data.

How to determine the optimal purify step?



$$minimize(||x_a(t) - x(t)|| + ||\hat{x_a}(t) - x(0)||)$$

Add noise to remove perturbation

Recover similar to the benign

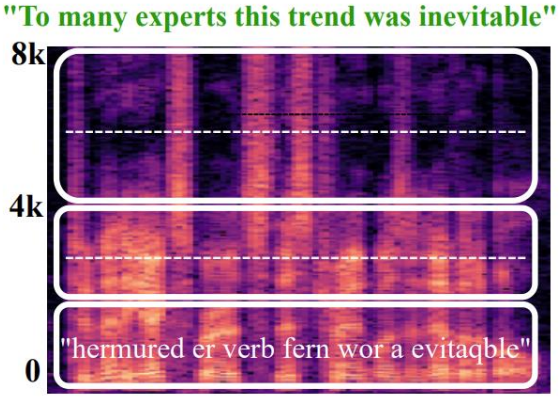
Image VS audio

Image classifier:

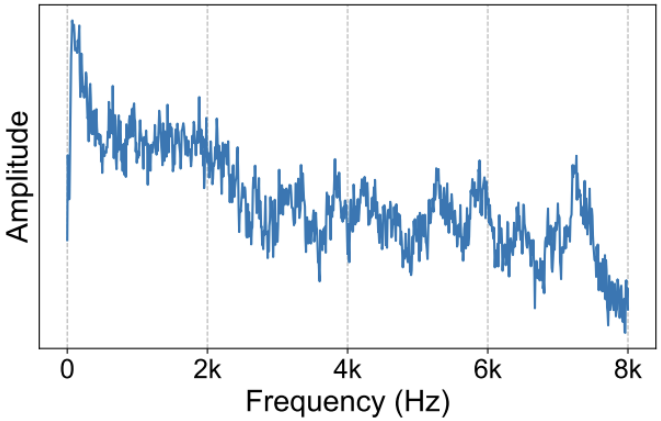
- Single input and single output
- Treats every pixel in the image equally
- Perturbation is evenly distributed

Speech recognition:

- Sequential model
- focus more on the low-frequency formants
- Perturbation is added on waveform, not even distributed on Spec

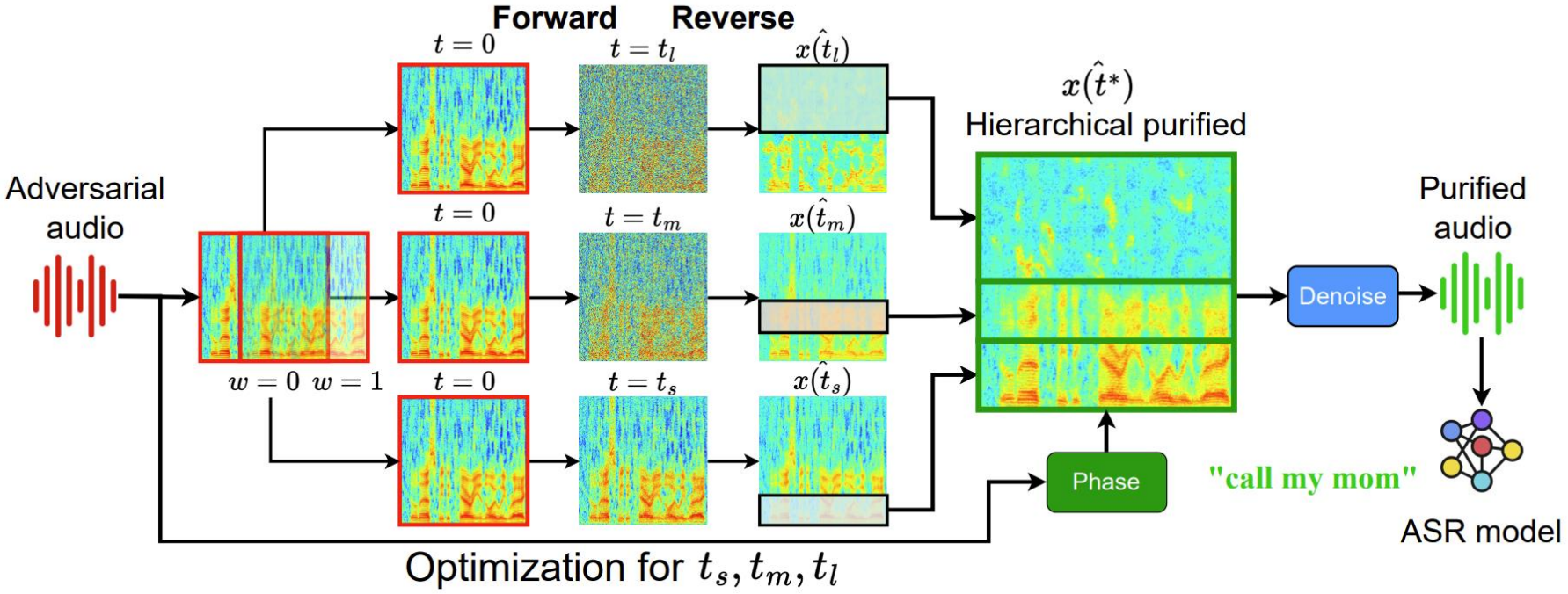


Different frequency components contribute differently to the transcription



Perturbation distribution

Our design: Hierarchical Diffusion Models



Our design: Optimization Goal

$$minimize(||x_a(t) - x(t)|| + ||x_a\hat{(t)} - x(0)||)$$

Does not reflect speech recognition goal: Get the correct Transcription

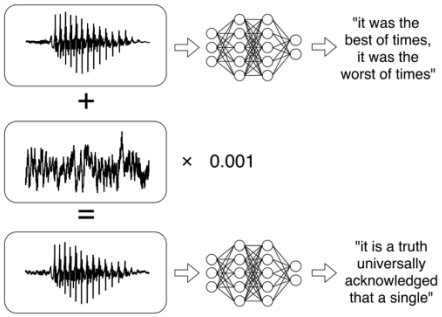
$$y = CER(f(x_a(\hat{t})), benign) - k * CER(f(x_a(t)), target)$$

Assure the transcription is recovered; Assure the target is not achieved.

$$x(\hat{t}^*) = x(\hat{t}_s)_{F<2k} || x(\hat{t}_m)_{2k<F<4k} || x(\hat{t}_l)_{F>4k}$$

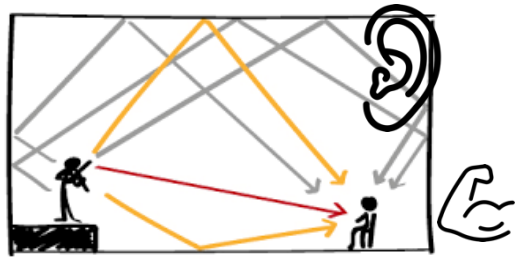
Instead of optimize single t, optimize the combinations.

Reproduce 3 Speech Recognition Attacks



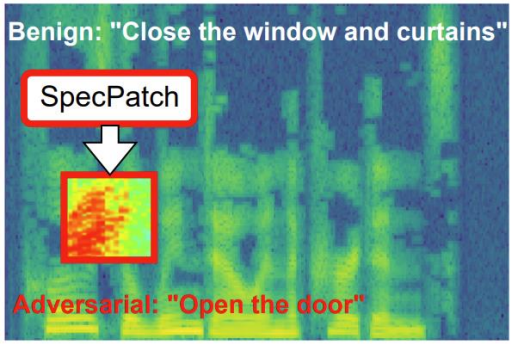
C&W Attack
SPW 2018 [3]

ASR: DeepSpeech



QIN-I Attack
PMLR 2019[4]

ASR: Lingvo



SpecPatch Attack
CCS 2022 [5]

ASR: DeepSpeech

[3] **SPW** Nicholas Carlini and David Wagner. 2018. Audio adversarial examples: Targeted attacks on speech-to-text. In 2018 IEEE Security and Privacy Workshops (SPW). IEEE, 1–7
[4] **PMLR** Yao Qin, Nicholas Carlini, Garrison Cottrell, Ian Goodfellow, and Colin Raffel. 2019. Imperceptible, robust, and targeted adversarial examples for automatic speech recognition.
[5] **CCS** Hanqing Guo, Yuanda Wang, Nikolay Ivanov, Li Xiao, and Qiben Yan. 2022. SpecPatch: Human-in-the-Loop Adversarial Audio Spectrogram Patch Attack

Defenses

WaveGuard [USENIX SECURITY 2021]:

- **Down-Up 2k:** Make a downsampling and upsampling to 2kHz to defend the AE.
- **LPC 10:** Use Linear Predictive Coding with 10th oreder filter to defend the AE.
- **Quant 8:** Quantilize the audio samples in 8 bits and then reconstruct back to defend the AE.

Denoise PLoS computational biology**:**

- **ANR:** Use adaptive noise reduction to defend the AE.
- **SNR:** Use stationary noise reduction to defend the AE.

Diffusion [ICLR 2023]:

- **DiffSpec:** Use Mel-spectrogram based diffusion model to purify AE.
- **DiffWave:** Use waveform based diffusion model to purify AE.

Metrics

CER - Character Error Rate:

Check if purified audio have the same transcription as original. **The lower, the better purification performance.**

WER – Word Error Rate:

Check if purified audio have the same transcription as original. **The lower, the better purification performance.**

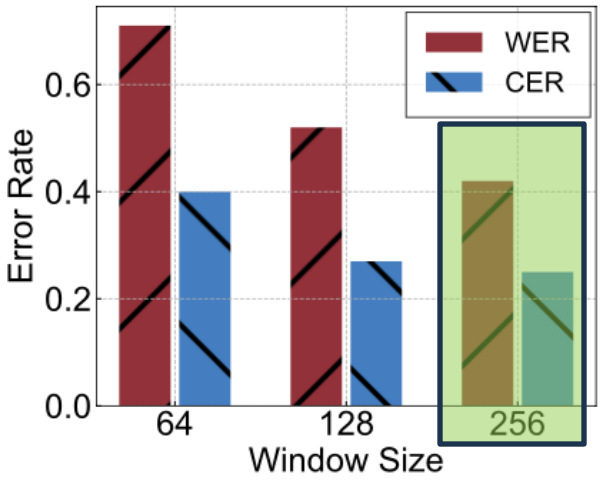
PSR - Purification Success Rate:

The rate at which purification is successful (When CER is lower than a threshold 0.2)

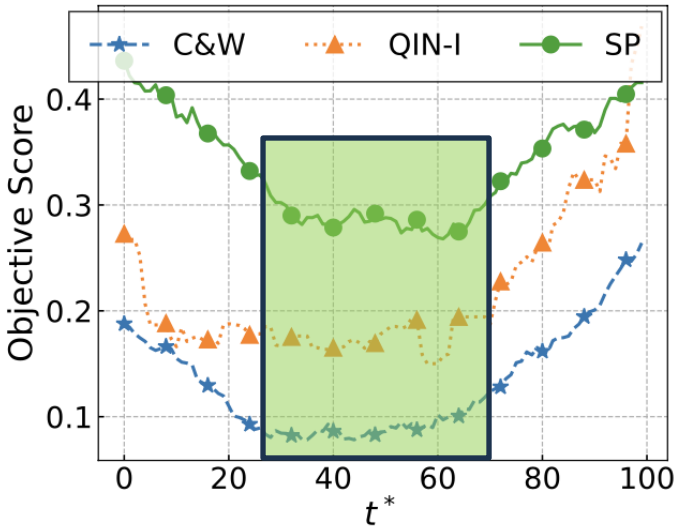
Typically, an effective model has a CER between 0.05 and 0.2 (e.g., OpenAI's Whisper)

Model Selection:

Guided diffusion from OpenAI
64 ×64, 128 ×128, 256 ×256

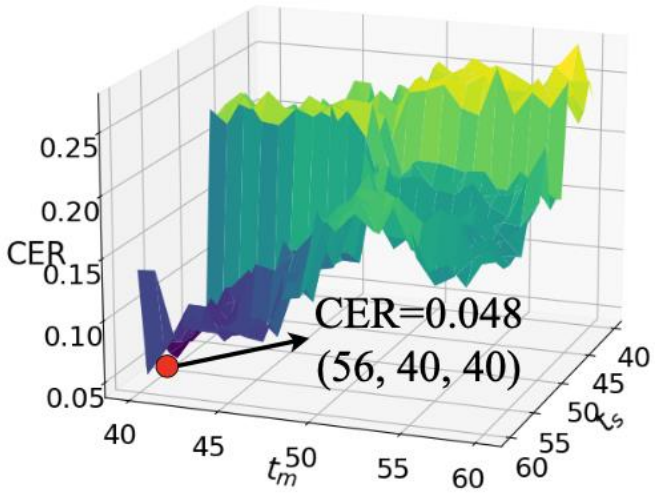


Global Purification Step
Searching start point for hierarchy optimal selection

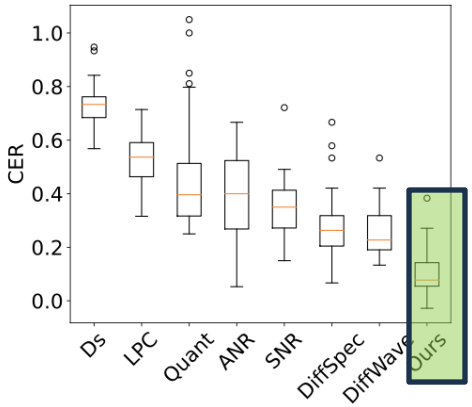


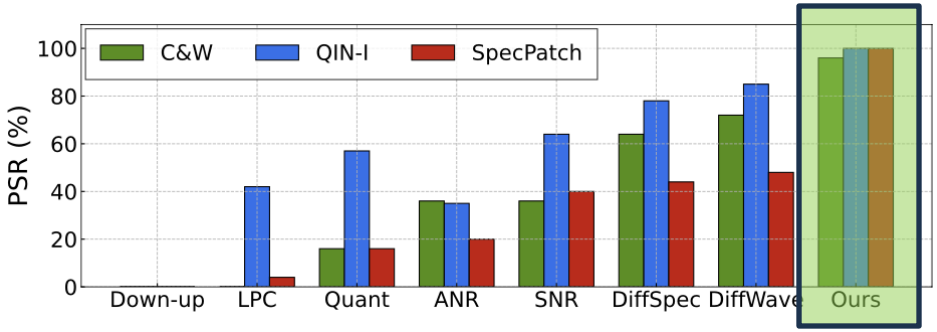
Result

Brutal force find the hierarchical optimal ts, tm, tl.



Achieve the best purification performance



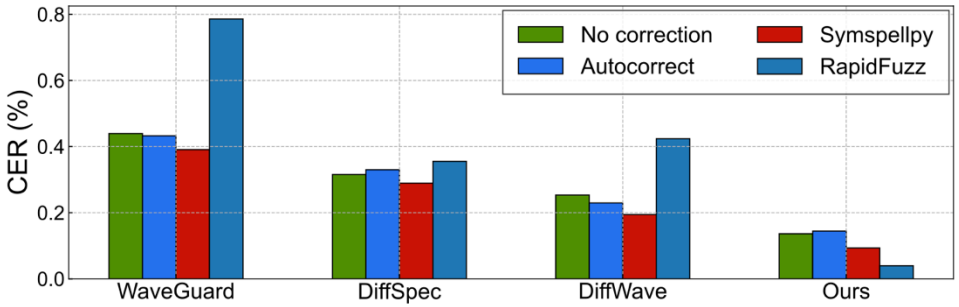


Discussion

Adaptive Attacker?

Out-of-memory;
According to DiffPure, Diffusion-based approach is robust to image adversary.

Post Auto-correction



Running time
Desktop VS Server

Process		Avg. Running Time (s)
ASR Models	DeepSpeech	9.04 / 0.56
	Lingvo	4.03 / 0.24
Purify Steps	Diffuse	8.77 / 0.44
	Reconstruct	0.11 / 0.12
	Denoise	0.02 / 0.02
Total Time of WavePurifier		8.9 / 0.58

Thank you!

Dr. Hanqing Guo

Assistant Professor
Department of Electrical and Computer Engineering
College of Engineering
University of Hawai'i at Mānoa
Office: Room 437, Holmes Hall
Email: guohanqi at hawaii dot edu

https://wavepurifier.github.io/
Code; Demo; PDF, slides

https://hanqingguo.github.io/
About me.
Team; Opening; Visiting;