DOI: 10.35598/mcfpga.2025.024

A Comprehensive Approach to Improving Security Against Unauthorized Audio Recording with Soc Support

Oleksandr Vorgul
Scientific adviser
ORCID 0000-0002-7659-8796
dept.Microprocessor Technologies and System
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
oleksandr.vorgul@nure.ua

Oleksiy Bilotserkovets
dept.Microprocessor Technologies and System
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
oleksii.bilotserkivets@nure.ua

***Abstract*— The problem of unauthorized speech recording remains acute due to the wide availability of miniature recording devices and the potentially high damage from leaks of confidential information. Existing solutions, such as "jammers" or passive security sensors, have significant drawbacks: legal restrictions, impact on legitimate devices and people, lack of "intelligence" and high cost. This work proposes a concept of an integrated hardware and software protection system implemented on the basis of a System-on-Chip (SoC). The key idea is the intelligent detection of recording devices using AI/ML algorithms that analyze acoustic and electromagnetic signatures, their precise localization and subsequent selective active counteraction (acoustic and/or EM) only to confirmed threats. This approach on a single SoC platform provides compactness, energy efficiency, high detection accuracy and minimal impact on the environment, representing a promising alternative to traditional methods**

Keywords*— *Voice Protection, Unauthorized Recording, System-On-Chip (SOC), Threat* *Detection, Selective Suppression, Intelligent Sound Analysis, Hardware-Accelerated AI, Source Localization.

I. INTRODUCTION

Air traffic control (ATC), airspace monitoring (ASM), The urgency of the problem is difficult to overestimate. Miniaturization of recording devices – from smartphones to specialized “bugs” – makes them easy to covertly place in conference rooms, offices or during private negotiations [1, 4]. Leakage of speech information can cause significant reputational and financial damage. At the same time, the shortcomings of existing solutions are obvious: broadband acoustic or EM “jammers” create legal problems and interfere with the operation of legitimate devices [5]; passive security systems are often unable to detect modern hidden voice recorders; specialized control systems are cumbersome, energy-consuming and expensive; all of them lack the “intelligence” to accurately identify the threat. This creates a need for a qualitatively new approach.

The purpose of this work is to develop and substantiate the concept of a comprehensive hardware and software solution for protection against unauthorized speech recording, built on the basis of a System-on-Crystal (SoC). This system should integrate the functions of continuous monitoring of the acoustic and electromagnetic environment, intelligent data analysis for detection and localization of recording devices, and selective active counteraction to only identified threats.

To achieve this goal, it is necessary to solve a number of problems : analyze modern methods of covert recording and methods for detecting them; develop an SoC architecture capable of integrating the necessary modules; research and select effective AI/ML algorithms for sound analysis and device detection [2, 6]; develop or adapt methods of directional acoustic and point EM suppression; design a system for managing the interaction of modules on the SoC ; evaluate the efficiency and resource intensity of the proposed approach in comparison with traditional ones .

The scientific novelty of the concept lies in the creation of a single integrated platform based on SoC , which for the first time combines intelligent threat detection based on modern AI/ML algorithms with the ability to provide highly accurate, targeted and selective active counteraction. The practical significance is seen in the creation of a prototype device that can potentially be significantly more compact, energy efficient and affordable than existing solutions, while providing a higher level of protection due to "intelligence" and selectivity.

II. ANALYSIS OF THE PROBLEM AND EXISTING APPROACHES

The threat of unauthorized speech recording is realized through various devices: from ordinary voice recorders and smartphones to specially designed miniature "bugs". Their covert placement and operation pose a serious challenge to security systems. Physical principles of detection of such devices are based on the analysis of their side emissions. Acoustic methods focus on identifying uncharacteristic background noise or parasitic acoustic signals, such as ultrasound from quartz generators [1, 5]. Electromagnetic methods are aimed at detecting radio emissions (Bluetooth, Wi-Fi, GSM transmitters in voice recorders or smartphones) or parasitic interference in the power grid [6]. Optical or vibration methods can be used in specific cases, but have limited application for this task.

Traditional methods of active counteraction are also physical in nature. Acoustic "jammers" generate masking noise (white noise, noise-like speech) or use ultrasonic pulses to affect microphones [5]. Their main problems are low selectivity (they affect the entire area and people), legal restrictions on the use of powerful sound pressure, and potential harm to hearing. Electromagnetic jammers aim to jam radio frequency data transmission channels. However, their use is often illegal, they interfere with legitimate communications and consume significant energy. In general, traditional systems suffer from the disparity of components,1. high power consumption, large dimensions, cost and, most importantly, the lack of intelligent processing for accurate detection and selective response. This emphasizes the need for an integrated approach combining smart detection with pinpoint impact on a single platform.

III. Architecture of a complex system based on SoC 2.

The core of the proposed solution is the System-on-Chip (SoC) . The choice of SoC as a basis is dictated by key advantages: the possibility of high integration of heterogeneous components (CPU, DSP, FPGA/ASIC blocks, memory, ADC/DAC, radio interfaces) on a single chip; outstanding energy efficiency ; compactness; high performance required for real-time signal processing and execution of AI algorithms; and, critically, the possibility of3. hardware implementation of resource-intensive algorithms (for example, via FPGA or specialized NPU) to achieve the required performance [3].

The hardware platform of the system includes both components inside the SoC and the peripherals controlled by it. The key modules inside the SoC are:

- Multifunctional high-speed ADC: Provides digitalization of signals from all sensors - microphone arrays and wideband EM receivers (SDR)

- Central Processing Unit (CPU): Performs functions of operating system management (e.g. RTOS), coordinating the work of all modules, implementing high-level logic and less time-critical algorithms.

- Digital Signal Processing (DSP) and/or FPGA (ASIC) units: The heart of the real-time system. This is where signal pre-processing (filtering, FFT), hardware-accelerated execution of optimized AI models for threat detection and classification [2], and generation of complex signals for selective suppression are performed.

- Multifunctional DAC: Converts digital suppression signals into analog form to drive acoustic radiators and EM transmitters.

- Memory Blocks (RAM/ROM): Stores data, software, trained and optimized AI models.

- Communication interfaces (Ethernet, Wi-Fi, Bluetooth): Provide remote control, administrator notification and software updates. 1.

Peripheral components controlled by the SoC include:

- Sensors: Microphone arrays (necessary for localizing the sound source/threat using triangulation), wideband EM receivers (SDR) for scanning the radio airwaves , optionally – vibration sensors.

Active elements for counteraction: Directional acoustic emitters (loudspeakers, ultrasonic transducers) and low-2. power EM transmitters for point impact.

User Interface: Simple display or indicators to show status, light/sound alarm, control buttons.

IV. Intelligent threat detection and analysis

The detection process begins with preprocessing of signals coming from microphone arrays and EM receivers. At this stage, noise is removed, useful features are identified, and the data is prepared for deep analysis.

The key role is given to AI/ML-based detection algorithms running on SoC:

Anomaly detection: Algorithms analyze the acoustic and EM background of the room in a "quiet" state. Any appearance of signatures characteristic of the operation of recording devices (e.g. specific ultrasonic harmonics of generators [1, 5], unique patterns of EM radiation of a smartphone in recording mode [4, 6]) is identified as an anomaly candidate for a threat.

Signal classification: Optimized neural networks (such as CNN for spectrograms or RNN for time series) running on DSP or hardware accelerated in FPGA/NPU classify the signal. The model is trained to distinguish between "background noise", "human speech", "recording device noise", "voice recorder EM radiation", "smartphone EM radiation in recording mode", etc. [2, 6]. This allows for accurate detection of the threat type.

Source localization: Using data from the microphone array and signal processing algorithms (e.g. MUSIC, SRP-PHAT) implemented on the DSP/FPGA, the system calculates the direction and distance to the source of the suspicious acoustic signal. EM localization is more complex, but determining the approximate direction based on the signal level on different antennas is also possible [6].

Correlation of data from different sensors (for example, the simultaneous appearance of suspicious acoustic noise and EM radiation at a frequency typical for voice recorders) dramatically increases the reliability of detection and reduces the likelihood of false alarms.

A special feature of the implementation on SoC is the need for careful optimization of AI models (weight quantization, pruning) for operation under conditions of limited computing resources and memory [2]. Speed-critical tasks (pre-processing, neural network inference) are transferred to hardware accelerators (DSP, FPGA, NPU), ensuring operation in real time.

V. Selective active counteraction

The fundamental difference of the proposed system is the selectivity of the impact. Counteraction is activated only upon confident detection and localization of the threat and is aimed strictly at it, minimizing the impact on the environment and people.

Acoustic suppression is implemented in two main ways:

Targeted masking signal generation: Based on localization information, the system uses DSP/FPGA to generate a complex signal and a phased array of emitters to direct a narrow beam of masking noise (e.g. adaptive to current speech) precisely towards the detected voice recorder or smartphone. This dramatically increases the signal-to-noise ratio for the attack microphone only .

Focused Ultrasound: Short, powerful ultrasonic pulses focused on a threat area [5] can temporarily disable a

sensitive microphone (due to nonlinear effects) or create overwhelming interference to recordings while remaining undetectable to the human ear in other parts of the room.

Electromagnetic suppression is also implemented pointwise:

- "Smart" RF jamming: Instead of suppressing the entire range, the system, using EM analysis data , generates narrowband interference strictly on those specific frequencies (e.g. Bluetooth MAC address of a specific smartphone, Wi-Fi channel) that the detected device uses for data transfer or synchronization. This requires precise control of the transmitter parameters, implemented on the DSP/FPGA SoC .
- Injection of interference into the power grid: If the detected device is powered by the grid, it is possible to generate specific interference introduced into the grid through the DAC and related circuits to disrupt its stable operation.

The hardware implementation of complex signal generators for suppression and control of emitters is effectively implemented in programmable logic (FPGA) and signal (DSP) blocks of the SoC .

VI. Management system and integration

The control of this complex system is provided by specialized software running on top of a microkernel or real-time operating system (RTOS) . This is necessary to strictly adhere to time constraints, especially at the signal processing and threat response stages [3].

Software architecture includes several levels:

- Drivers: Provide low-level interaction with SoC hardware modules (ADC/DAC, memory, interfaces) and external peripherals (sensors, emitters).
- Signal processing services: Implement preprocessing, classification (AI), and localization algorithms. Performance-critical parts are executed on the DSP or hardware-accelerated in FPGA/NPU.
- Threat Manager: Is the "brain" of the system. Receives data from processing services (classification results, threat coordinates), makes a decision on the presence and type of threat, selects and launches the appropriate active counteraction module, initiates an alert.
- Active Countermeasure Modules: Control the operation of acoustic emitters and EM transmitters in accordance with threat manager commands and required jamming algorithms.
- User and network interfaces: Provide information about the system status, settings, channels for remote control and monitoring.

The flow of data and decision making is a continuous cycle: data collection from sensors -> pre-processing -> intelligent analysis (AI) -> decision making by the threat manager -> activation of selective countermeasures -> monitoring of effectiveness and threat status -> feedback for corrective actions.

VII. Evaluation of effectiveness and ways of implementation

Evaluation of the feasibility and benefits of the proposed concept requires the definition of clear criteria:

- Detection accuracy: Evaluated by standard metrics for AI classification models : Precision , Recall , F1-score (harmonic mean) on test datasets containing signatures of real devices [2, 6].
- Reaction speed: The time from the moment a threat appears in the action zone to the beginning of effective counteraction is measured. Should be minimal (fractions of a second).
- Suppression effectiveness: Evaluated objectively using speech quality metrics such as PESQ (Perceptual Evaluation of Speech Quality) and STOI (Short-Time Objective Intelligibility) applied to a recording made by an attacked device under the influence of the system.
- Selectivity: The number of false positives is measured (False Positive Rate and the degree of impact of the system on legitimate devices and people in the room (should be close to zero).
- Resource intensity: The load of key SoC blocks (CPU, DSP, FPGA, NPU) and the overall power consumption of the system in monitoring and active counteraction modes are assessed.
- Cost: Estimated cost of components and final device.

The testing methodology will require the creation of a stand simulating the protected premises, with the placement of various recording devices (dictaphones, smartphones of different models, specialized "bugs"). For objective measurements, it is necessary to use spectrum analyzers, high-precision microphones for recording "from the point of view of the attacking device", equipment for measuring EM fields.

To develop a prototype, it will be necessary to select a specific SoC platform that meets the requirements for signal processing performance, the presence of hardware AI accelerators (NPU or FPGA of sufficient size), interfaces for connecting sensors and emitters. Promising candidates are platforms such as Xilinx Zynq UltraScale+ MPSoC (ARM CPU + FPGA), Intel Agilex (ARM CPU + FPGA) or SoCs with built-in powerful NPUs (e.g. some NVIDIA Jetson, Qualcomm solutions, or dedicated audio analytics chips).

Expected results from implementing such a system on SoC compared to traditional solutions include: higher detection accuracy due to the use of AI; lower power consumption due to high integration and optimization; significant compactness of the final device; a fundamentally new possibility of selective action minimizing side effects; and potentially better price/performance per unit of protected space.

VIII. Conclusions

In this paper, we propose and substantiate the concept of a comprehensive system for protection against unauthorized speech recording , built on the basis of a System-on-Chip (SoC) . The key elements of the concept are: an integrated platform for monitoring the acoustic and electromagnetic environment; the use of artificial intelligence and machine learning algorithms for intelligent detection and classification of threats by their acoustic and EM signatures; precise localization of the threat source; and the use of selective active counteraction (acoustic and/or EM) strictly aimed at the detected device. The SoC architecture allows for the efficient implementation of all these functions, providing high performance, energy efficiency and compactness.

1. The main conclusions of the work confirm the prospects and advantages of the proposed approach:
2. Comprehensive integration of detection, analysis and countermeasure functions on a single SoC platform is a technically feasible and effective strategy.
3. The use of optimized AI/ML algorithms running with hardware acceleration on SoC allows achieving high accuracy and detection speed of modern recording devices.
4. The principle of selectivity in active counteraction is key to minimizing legal risks and impacts on the environment and legitimate devices.
5. The use of SoCs opens the way to creating compact, power-efficient and potentially more affordable security devices compared to traditional disparate systems.
6. Prospects for further research and development include:
7. Developing and training more complex, robust and accurate AI models for detection, possibly using deep learning methods on raw audio and RF signals [2, 6].
8. Research and implementation of new physical principles and algorithms for covert and highly effective suppression, further minimizing detectability of system impact.
9. Further minimization of the size, cost and power consumption of the final device by optimizing algorithms, selecting more advanced SoCs and peripheral components.
10. Study of the system's resistance to possible countermeasures by intruders (for example, the use of devices with shielding or non-standard recording/transmission methods).
11. A detailed legal analysis of the use of active suppression methods (especially EM) in various jurisdictions to ensure the legality of the system's use.
12. Development of a fully functional prototype on the selected SoC platform and conducting comprehensive full-scale tests under conditions as close to real ones as possible to validate all declared characteristics and advantages.

REFERENCES

[1] Kune , D. F., Kim , Y., & Backes , J. (2017). *Surprisingly Weakening Cryptographic Hardware : Acoustic Cryptanalysis Revisited* . IEEE Security & Privacy..

[2] Shi , Y., Ferdowsi , H., & Grosse , K. (2022). *DeepAudioML : Efficient Deep Learning Audio Classification on Edge Devices using Model Compression and Hardware Acceleration* . ACM Transactions on Embedded Computing Systems (TECS).

[3] Chen , L., Wang , Z., Liu , Y., & Ren , S. (2021). *A Survey of System-on-Chip Security : Threats , Countermeasures , and Design Challenges* . ACM Computing Surveys

[4] Roy , N., Shen , S., Hassanieh , H., & Choudhury , R. R. (2020). *Inaudible Audio as a Covert Channel for Mobile Device Tracking and Identification* . Proceedings on Privacy Enhancing Technologies (PETS).

[5] Zhang , G., Yan , C., Ji , X., Zhang , T., Zhang , T., & Xu , W. (2021). *DolphinAttack 2.0: Inaudible Voice Commands and the Effectiveness of Ultrasonic Interventions* . IEEE Symposium on Security and Privacy (SP).

[6] Kumar , A., Li , T., & Koutsonikolas , D. (2023). *PhyAuth : Physical-Layer Authentication of IoT Devices via Hybrid RF-Fingerprinting and Deep Learning* . IEEE Transactions on Information Forensics and Security (TIFS).