

Satellite-Assisted Continuous Roving Unmanned IoT/RF Sensor Enhanced Robot (SATCRUISER) for AI Data Generation

Michael Jenkins, Calvin Leather, and Sean Kelly

Knowmadics, Herndon, VA 20171, USA

ABSTRACT

The proliferation of Internet of Things (IoT) devices has revolutionized the landscape of facility and supply chain security, offering unprecedented opportunities for real-time monitoring, proactive threat detection, and enhanced operational resilience. However, the increasing adoption of IoT devices in the context of facility and supply chain security has resulted in novel implications for safeguarding assets and mitigating risks. The dynamic nature of modern facilities and complex supply chains demands robust security measures to safeguard assets, prevent disruptions, and ensure operational continuity. Traditionally, security systems have relied on static, fragmented approaches, often lacking comprehensive visibility and real-time insights. The advent of IoT devices presented a paradigm shift, enabling the creation of interconnected ecosystems that monitor and secure critical assets throughout the facility and supply chain. IoT devices offer an array of benefits for facility and supply chain security. These devices can be deployed as sensors, actuators, and monitoring devices, collecting and transmitting data on various parameters such as access control, environmental conditions, inventory levels, and equipment performance. While the proliferation of IoT devices presents potential, it also brings forth certain challenges. The interconnectedness of devices increases the attack surface, raising concerns about cybersecurity vulnerabilities and potential breaches. Ensuring robust security measures, including secure device authentication, encryption, and regular firmware updates, is crucial to safeguard against unauthorized access and potential data compromises. Artificial Intelligence (AI) is a rapidly maturing and evolving technology domain with potential to bolster the security of this class of IOT-based security solutions. Specifically, AI holds potential to provide more robust and responsive capabilities to ensure IOT device and endpoint security (e.g., anomaly detection, predictive maintenance, threat intelligence, automated security response, behavior monitoring, etc.). AI-driven security measures have the potential to provide a robust defense against emerging threats that can continuously learn and adapt to detect and enable mitigation of emerging threat vectors. However, high-quality data to train AI models is vital. The accuracy and performance of AI algorithms heavily depend on the quality, diversity, and quantity of the training data. To train AI algorithms effectively, a vast amount of diverse and labeled data is required. However, acquiring such data can be challenging, as it necessitates extensive and continuous data collection from various sources. Traditional methods of data gathering often fall short due to limitations in coverage, scalability, and real-time data availability. This paper provides an overview and initial findings from a rapid prototyping and hackathon effort to develop a Satellite-Assisted Continuous Roving Unmanned IoT/RF Sensor Enhanced Robot (SATCRUISER) that enables continuous collection, geolocalization, and backhaul of a facility's IOT endpoints and wireless activity. The intent of SATCRUISER is to enable the continuous collection of quality IOT wireless data that can be used for initial AI model training and, eventually, ongoing model learning and facility monitoring. We present the SATCRUISER system architecture and initial findings from pilot collection periods to source a baseline corpus of data to train novel AI IOT security models.

Keywords: IoT, IIoT, Artificial intelligence (AI), RF communications, GIS, Data analysis, Data generation, Supply chain security, Cybersecurity

INTRODUCTION

The landscape of facility and supply chain security has been transformed by the proliferation of Internet of Things (IoT) devices, offering unprecedented opportunities for real-time monitoring, proactive threat detection, and enhanced operational resilience. However, the increasing adoption of IoT devices in this context has introduced new implications for safeguarding assets and mitigating risks. The dynamic nature of modern facilities and complex supply chains necessitates robust security measures to safeguard assets, prevent disruptions, and ensure operational continuity. Traditional security systems often rely on static, fragmented approaches that lack comprehensive visibility and real-time insights. The emergence of IoT devices has brought about a paradigm shift, enabling the creation of interconnected ecosystems that monitor and secure critical assets throughout facilities and supply chains.

In the realm of facility and supply chain security, IoT devices provide a wide range of benefits. These devices can be deployed as sensors, actuators, and monitoring devices, collecting and transmitting data on various parameters such as access control, environmental conditions, inventory levels, and equipment performance. By leveraging IoT devices, organizations can achieve real-time visibility into their operations, enabling proactive decision-making and rapid response to security threats or anomalies. The ability to monitor and manage critical assets in real-time enhances security measures and improves overall operational efficiency.

The proliferation of IoT devices also presents challenges. The interconnectedness of these devices increases the attack surface, raising concerns about cybersecurity vulnerabilities and potential breaches. Safeguarding against unauthorized access and data compromises requires robust security measures, including secure device authentication, encryption, and regular firmware updates. Ensuring the integrity and confidentiality of IoT device communication is paramount to maintaining the security and resilience of facility and supply chain operations.

Artificial Intelligence (AI) is a rapidly maturing and evolving technology domain that has the potential to bolster the security of IoT-based security solutions. AI offers robust and responsive capabilities to enhance IoT device and endpoint security. AI-driven security measures, such as anomaly detection, predictive maintenance, threat intelligence, automated security response, and behavior monitoring, can provide a strong defense against emerging threats. By continuously learning and adapting, AI algorithms can detect and mitigate emerging threat vectors, contributing to a proactive and dynamic security posture.

However, the effectiveness of AI algorithms relies heavily on the quality, diversity, and quantity of training data. High-quality data is vital to train AI models effectively. Acquiring such data can be challenging, as it requires extensive and continuous data collection from various sources. Traditional methods of radio frequency (RF) data gathering relevant to IoT often fall short due to limitations in coverage, scalability, and real-time data availability.

In this research paper, we propose a novel approach to address the challenges associated with acquiring high-quality data for training AI algorithms in the context of facility and supply chain security. We introduce SATCRUISER (Satellite-enabled Continuous RF Data Survey and Localization), a prototype system designed to continuously survey and localize RF data over a predefined geographic area. SATCRUISER can be affixed to a mobile platform, such as a UGV, and collect RF data in real-time. This system combines the capabilities of RF data surveying and localization with the potential of AI algorithms for analyzing and identifying transmitting devices or anomalies in patterns of life.

SATCRUISER's ability to collect and backhaul RF data to a central repository using ad hoc Wi-Fi or periodic satellite communications aligns with the needs of acquiring high-quality data for training AI algorithms. The continuous and comprehensive coverage provided by SATCRUISER enhances the diversity and quantity of the collected RF data, contributing to a robust training dataset. This dataset, combined with advanced AI/ML models, empowers the development of intelligent solutions for identifying transmitting devices and detecting anomalies in patterns of life.

In the subsequent sections of this research paper, we delve into the technical details of SATCRUISER, including its hardware components, data collection, and processing methodologies, and its initial evaluations. We also discuss the potential applications and implications of SATCRUISER for facility and supply chain security, highlighting its role in bolstering the security of IoT-based security solutions. By bridging the gap between RF data surveying, localization, and AI-driven analysis, SATCRUISER represents a promising solution for enhancing the security of facility and supply chain operations. Its innovative design and capabilities will enable real-time monitoring, proactive threat detection, and improved operational resilience by providing a continuous and autonomous RF data collection solution. Furthermore, the system's ability to collect high-quality RF data as a corpus for training AI models holds the potential to benefit the field of facility and supply chain security, ensuring robust defenses against emerging threats and safeguarding critical assets can be developed and tested.

SATCRUISER PROTOTYPE OVERVIEW

System Architecture

We designed and implemented a flexible data collection prototype before using this prototype to collect data for evaluation. The prototype's system architecture is shown in Figure 1. SATCRUISER is designed to enable continuous surveying and localization of RF data over a predefined geographic area. It consists of two main components: a small hardware payload and a data processing element. The hardware payload comprises a Software-Defined Radio (SDR) for Electronic Intelligence (ELINT), power management, local data storage, and communications capabilities. Additionally, it incorporates protocol-specific receiver ASICs/FPGAs for capturing Bluetooth, ZigBee, Wi-Fi, 3G, 4G, and 5G. The Data Processing element includes source localization algorithms, a user interface for data query and visualization, and data

export functions. The system employs an adaptive communications protocol to automate switching between satellite, Wi-Fi, LTE, or physical cable communications based on availability.

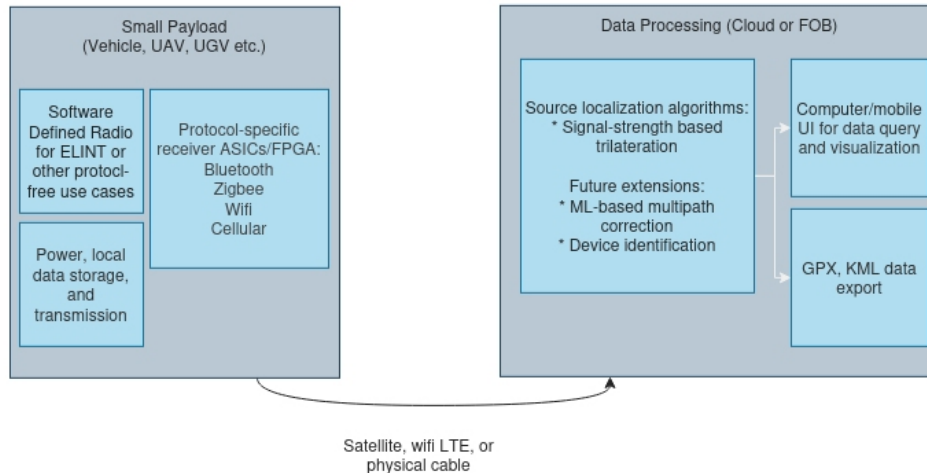


Figure 1: System architecture, consisting of a payload hardware device and cloud data processing and storage.

Hardware Payload

The hardware payload of SATCRUISER consists of various components that enable RF data collection and communication. It includes an SDR, which provides the capability to receive and analyze RF signals across a wide frequency range. The SDR is programmable, allowing flexibility in configuring the receiver for different spectra and protocols. Additionally, the hardware payload incorporates receiver ASICs/FPGAs specific to various protocols (e.g., nRF 52840 for ZigBee communication), resulting in spectrum coverage for Bluetooth, ZigBee, Wi-Fi, 3G, 4G, and 5G RF signals.

The payload includes a small Amlogic SoC-based microcontroller with integrated storage, a battery and optional solar panel, and a GPS receiver for location data. Local data storage within the hardware payload facilitates immediate data processing and temporary storage before transmitting it to the central repository. This local storage allows for efficient data handling and analysis without relying solely on real-time communication with the data processing element.

Communications capabilities within the hardware payload are essential for transmitting data to the central repository. SATCRUISER employs an adaptive communications protocol that can switch between different communication modes based on availability. This adaptive approach allows the system to utilize satellite communications (currently configured for StarLink's SWARM constellation), Wi-Fi, LTE cellular, or physical cable connections (e.g., when docked for charging), depending on the availability of each option. This flexibility ensures reliable and continuous data transmission,

regardless of the specific communication infrastructure within the deployment area. An example deployment of the full hardware payload, removed from its enclosure for visibility, is shown below in Figure 2.

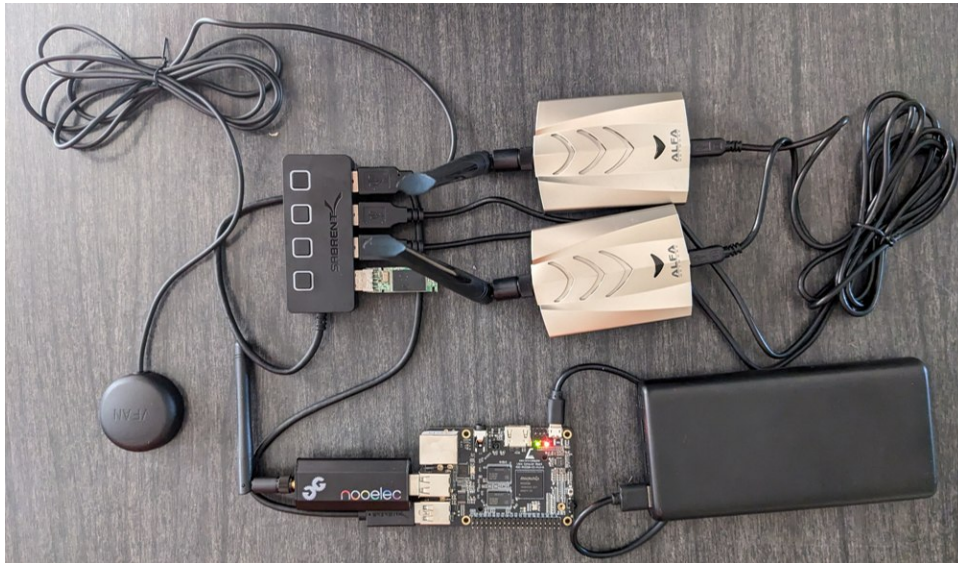


Figure 2: Hardware prototype example configuration, removed from case.

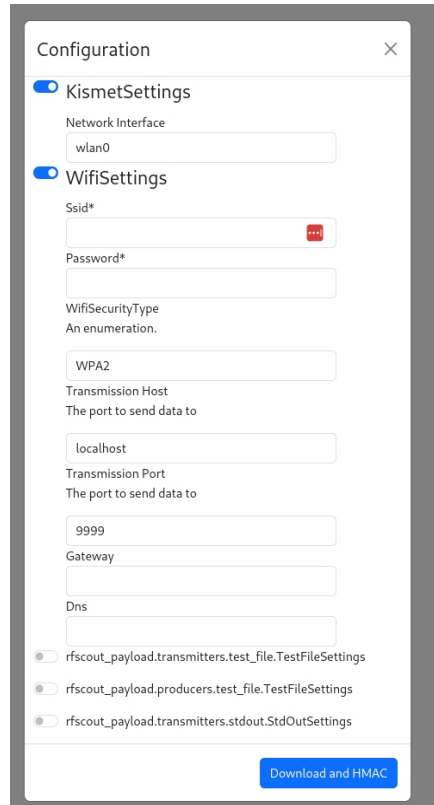
Data Processing Element

The data processing element of SATCRUISER is responsible for analyzing, visualizing, and managing the collected RF data. It includes source localization algorithms and data export functions. Users interact with the system by using a website to generate a configuration file (see Figure 3), which is then transferred to the payload (via a USB data drive that is left inserted during operation). This configuration file contains information controlling how data is collected by each component of the payload in addition to how the payload is to transmit the data back to the cloud processing environment.

To ensure security, this configuration file is validated using a private key, and the resulting hash-based message authentication code (HMAC) is transmitted with the configuration file. The payload will only load the configuration if the HMAC is valid. In this way, we ensure that a malicious third party cannot supply configuration files to, for example, exfiltrate data to a malicious endpoint.

On the payload itself, data from the various sensor components is pre-processed to prepare it for transmission. In the case of protocol-specific processing, this includes formatting packet header fields and computing signal strength estimates. Currently, source localization is not performed on the payload itself but after data is transmitted to the central repository. We currently perform signal-strength weighted trilateration to identify source locations. Note that as the focus of this device is to collect training datasets

for ML applications, including source location, source localization capabilities were left simple. In future efforts, this processing could be improved and performed on the payload itself if warranted by a use case.



The image shows a web-based configuration interface titled "Configuration". It features several sections and input fields:

- KismetSettings**: A toggle switch is turned on. Below it is a "Network Interface" field with the value "wlan0".
- WifiSettings**: A toggle switch is turned on. Below it are fields for "Ssid*" (with a red eye icon for visibility), "Password*", "WifiSecurityType" (with a subtext "An enumeration." and the value "WPA2"), "Transmission Host" (with the value "localhost" and subtext "The port to send data to"), "Transmission Port" (with the value "9999" and subtext "The port to send data to"), "Gateway", and "Dns".
- At the bottom, there are three disabled toggle switches for "rfscout_payload.transmitters.test_file.TestFileSettings", "rfscout_payload.producers.test_file.TestFileSettings", and "rfscout_payload.transmitters.stdout.StdOutSettings".
- A blue button labeled "Download and HMAC" is located at the bottom right.

Figure 3: Configuration interface.

The data produced by the hardware prototype fundamentally consists of a series of datapoints containing an identified device’s protocol-specific ID (e.g., MAC address; in the case of protocol-free SDR data, this ID is just the wavelength measured), an estimate of device-specific signal strength, and the device’s location when that measurement was taken. We also collect and transmit some device metadata for protocol-specific measurements (e.g., security protocols identified from message headers). This data is transmitted via a flexible interface over HTTPS (secured via TLS) to the data processing backend, where it is processed and stored for visualization. In our initial prototyping, we performed minimal processing, as our goal is to support data collection for next-generation AI-based algorithms. As such, the main function of the cloud backend for this prototyping was to store the received data in blob storage and provide basic visualization and data export.

Data export functions facilitate the transfer of processed RF data from the central repository to industry standard formats (e.g., GPX, KML) for geospatial visualization or analysis within third-party applications, security systems, or AI/ML models. This ensures the availability of comprehensive

data for training new AI/ML models and conducting in-depth analysis for security purposes.

PILOT COLLECTION DATASETS

Exemplar UGV Platform

To expedite the pathway to data collection for testing and evaluation of SATCRUISER, we adopted the Husqvarna Automower UGV. This commercially available platform was designed specifically for autonomous lawn mowing. However, while primarily used for residential and commercial lawn care, the Automower possesses several technical features that make it a potential UGV platform for various applications beyond its intended purpose.



Figure 4: SATCRUISER payload mounted to the automower UGV in the charging dock (left) and out of the dock (right).

Navigation System: The Automower utilizes a sophisticated navigation system to autonomously navigate and mow lawns. This system consists of multiple sensors and components that enable the UGV to perceive and interact with its environment. Onboard GPS and compass functions allow designated areas to be delineated as target zones (for mowing), transport paths (for moving between target zones), keep-out zones (areas to avoid), and maintenance points (specific points to check in on a cyclical basis). Real-time kinetic (RTK) navigation is then used to ensure the UGV adheres to the designated area layouts, and advanced UGV pathing algorithms allow for further pattern control within target zones (e.g., mowing a grid or diamond pattern into the lawn). This RTK navigation functionality provides highly accurate and precise positioning information in real-time at centimeter-level accuracy (e.g., the standard used for surveying, mapping, precision agriculture, construction, and more advanced autonomous vehicle navigation solutions).

Collision Sensors: The Automower is equipped with collision sensors that enable it to detect and avoid obstacles in its path. These sensors use various

technologies, such as ultrasonic sensors, bumper sensors, or infrared sensors, to detect objects and adjust the UGV's path accordingly.

Lift and Tilt Sensors: To ensure safety and prevent injury or damage, the Automower is equipped with lift and tilt sensors. These sensors can detect when the UGV is lifted or tilted, automatically stopping the cutting blades to prevent accidents.

Power and Charging: To ensure continuous operation, the Automower relies on a power and charging system. The main components of this system are the rechargeable battery and a fixed-position charging station that uses a dedicated frequency and visual beacon to enable the Automower to automatically return to the charging station when the battery power is low or when a charging cycle is scheduled.

Connectivity: For remote monitoring and control, the Automower may utilize Wi-Fi and Bluetooth wireless connectivity to facilitate communication with the UGV via a mobile application.

Weather Resistance: The UGV is designed to withstand various weather conditions, including rain, allowing it to operate safely and autonomously even in inclement weather.

The above collection of features made the Automower an ideal platform to serve as an exemplar UGV for data collection with SATCRUISER. Specifically, the ability to plan designated paths and areas and then rely on the Automower's scheduling pathing, charge management, automated docking (to serve as a potential data upload trigger), and safety features allowed for straightforward planning for repeatable, consistent, wide-area (up to 2 acres), and long-duration (when combined with the SATCRUISER battery and solar charging panel) collection cycles.

Collection Methods

We collected three pilot datasets while evaluating the efficacy of the prototype. These datasets consisted of:

1. Collection of RF data from the UGV platform around a residential property with a variety of RF-transmitting smart home devices, including those of neighboring houses.
2. Collection of RF data from a car around a suburban area (Mt. Pleasant, SC, USA) with a variety of Wi-Fi and cellular devices.
3. Collection of RF data from a car around an urban area with a variety of Wi-Fi and cellular devices.

Summary statistics of these datasets are presented below.

Table 1. Summary statistics of the pilot datasets.

Measures	Smart Home	Suburban	Urban
Collection Duration	168 hours	1 hour	30 minutes
Identified Devices	2673	6085	9825
Raw Dataset Size	424 Mb	31 Mb	30 Mb



Figure 5: Visualizations of the collected datasets (smart home obfuscated for privacy concerns).

RESULTS

Dataset Evaluations

To evaluate the usefulness of the generated data, we performed a basic trilateration-based localization of signal sources. Some signal sources had known locations, either because we placed the signal sources ourselves or because the signal sources were associated with places of business with known names and locations from OpenStreetMap (OSM; www.openstreetmap.org). An example of this can be seen in Figure 5, where one of the resulting buildings is aligned with the detected signal strength measurements and packet captures. Using this alignment method, we were able to identify 30 buildings from our Suburban (Mount Pleasant) dataset, consisting of a mixture of hotels, restaurants, and event venues, where we were able to align an SSID with the name of the business and OSM building data.

This dataset, especially with a larger recording volume and, as a result, more buildings, can be a useful mechanism to build improved methods for network security characterization and RF source localization. For example, using a trivial weighted-averaging-based localization algorithm (i.e., weighting the locations of measurements based on RSSI), we were able to compute localization errors for each of these resolved buildings (we found a final mean absolute error of around 200 feet within these resolved buildings). This could then be used as training data for more sophisticated machine learning models.

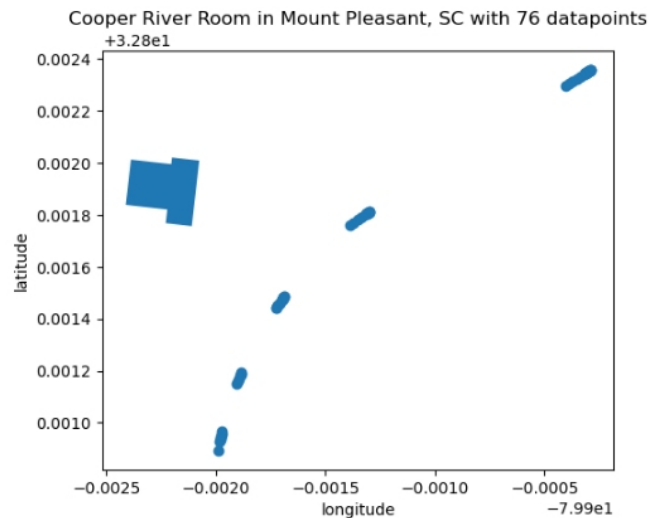


Figure 6: Example building data from open street map, along with locations of signal strength measurement from the business.

AI Data Corpus Generation

Depending on the hardware payload configuration, the resulting data contains signal frequency, estimated signal strength, (where applicable) device hardware ID, and many protocol-specific properties (e.g., for Wifi, the encryption protocol, temporal aspects such as packet rate) for Bluetooth, Wifi, ZigBee, cellular, and raw RF spectrum. These pieces of data may be correlated to develop AI models that can predict features (in the case that they are not available for a future arbitrary device), or the data may be collected in tailored scenarios (e.g., scenarios with known emitter locations or identities) to support supervised learning.

USABILITY LESSONS LEARNED

Through the course of prototyping, several usability lessons were learned. Strategies we prototyped that made the device easier to use include:

1. Use a schema-based configuration file, and then use a tool to automatically generate web GUIs with type validation to generate configuration

files. This simplifies usage for less technical operators, who may not be comfortable modifying XML or JSON documents.

2. Provide real-time feedback about errors via a screen or embedded LEDs. This makes it much faster for a user to diagnose issues with the system.

We initially built the cloud front end to visualize data. It quickly became apparent that many users already had preferred software (e.g., ArcGIS). As such, we added the ability to export data in several common formats (GPX, KML, geoJSON). This allowed users to leverage existing tools they were familiar with.

CONCLUSION

The successful preliminary testing of SATCRUISER has demonstrated its capability to continuously collect RF data and backhaul it for later analytics. This research paper has presented the technical overview of SATCRUISER, highlighting its innovative features, such as the hardware payload comprising a SDR, protocol-specific receiver ASICs/FPGAs, and adaptive communications protocols. The Data Processing element includes source localization algorithms, a user interface for data query and visualization, and data export functions. SATCRUISER's ability to collect and transmit RF data over a pre-defined geographic area in real-time holds great potential for applications in RF spectrum management and security.

This successful preliminary testing effort has validated the functionality and feasibility of SATCRUISER. The system has demonstrated its ability to continuously survey and localize RF data, ensuring comprehensive coverage of the predefined area. The integration of an adaptive communications protocol enables efficient data backhaul, utilizing available communication options such as satellite, Wi-Fi, LTE, or physical cable connections. This flexibility ensures reliable and continuous transmission of RF data, irrespective of the specific communication infrastructure in the deployment area.

The collected RF data serves as a valuable corpus for training new AI/ML models. The ability to identify transmitting devices and anomalies in patterns of life through advanced AI algorithms enhances the system's potential for accurate and efficient analysis of the RF spectrum. SATCRUISER's contribution to RF data analytics opens new avenues for spectrum management, interference detection, and security applications.

Based on the successful preliminary testing effort, several recommended future research directions and next steps have been identified for SATCRUISER:

Enhanced Localization Algorithms: Continuing research on source localization algorithms can improve the accuracy and reliability of SATCRUISER's RF data analysis. Exploring machine learning techniques and incorporating advanced signal processing algorithms may enhance the system's ability to accurately identify and localize transmitting devices.

Integration with AI/ML Models: Further research can explore the integration of advanced AI/ML models with SATCRUISER's data processing capabilities. By leveraging the collected RF data corpus, more sophisticated

AI algorithms can be developed to detect complex patterns, identify new wireless devices, and classify anomalous behaviors in the RF spectrum.

Field Testing and Validation: Conducting extensive field testing and validation is vital to assess SATCRUISER's performance in real-world scenarios. Collaborating with industry partners and relevant stakeholders will provide valuable insights and feedback for further improvements and optimization.

Application-Specific Adaptations: Exploring potential adaptations of SATCRUISER for specific application domains, such as telecommunications, defense, or public safety, can unlock new use cases and further refine the system's functionalities to cater to unique requirements.

In conclusion, the successful preliminary testing effort of SATCRUISER showcases its potential for continuous RF data collection and backhaul, providing a robust platform for RF spectrum management and security. By recommending future research and next steps, this paper highlights the directions for further enhancement of SATCRUISER's performance, scalability, and integration with AI/ML models, as well as addressing security considerations. Continued research and development efforts will ensure that SATCRUISER evolves into a powerful tool for effective RF data analysis, enabling improved spectrum management and security in diverse application domains.