Naval Cybersecurity in the Age of AI: deceptive ISAR Images Generation with GANs

Giulio Meucci ¹, Bertan Karahoda ¹, Amir Hossein Oveis ¹, Francesco Mancuso ¹, Edmond Jajaga ¹, and Alessandro Cantelli-Forti ²,²

¹Affiliation not available
²CNIT

October 31, 2023

Abstract

Navigational systems, the heart of maritime operations, face escalating cybersecurity risks due to their system of systems nature and reliance on diverse suppliers. Amid concerns about supply chain attacks and insider threats, the potential for malicious radar image injections, including Inverse Synthetic Aperture Radar (ISAR) images, is emerging. Such manipulations can critically undermine navigational integrity through the generation of decoy targets, the strategic relocation of existing ones, and the effective concealment of additional targets. We have identified where an Advanced Persistent Threat (APT) could be concealed within the processing chain of a modern radar system. This study demonstrates the potential of APTs to exploit Generative Adversarial Networks (GANs) for the creation of deceptive ISAR images, thereby spotlighting this previously unexplored threat vector. Our findings offer novel insights into bolstering maritime cybersecurity in an increasingly AI-dominated landscape.
Naval Cybersecurity in the Age of AI: deceptive ISAR Images Generation with GANs

Giulio Meucci†‡, Graduate Student Member, IEEE; Bertan Karahoda§; Amir Hosein Oveis‡; Francesco Mancuso†‡, Graduate Student Member, IEEE; Edmond Jajaga§; Alessandro Cantelli-Forti†‡, Member, IEEE;
†Department of Information Engineering, University of Pisa, Pisa, Italy
‡Radar and Surveillance Systems (RaSS) National Laboratory - CNIT, Pisa, Italy
§Department of Computer Science, University for Business and Technology, Prishtina, Kosovo

alessandro.cantelli.forti@cnit.it

Abstract—Navigational systems, the heart of maritime operations, face escalating cybersecurity risks due to their system of systems nature and reliance on diverse suppliers. Amid concerns about supply chain attacks and insider threats, the potential for malicious radar image injections, including Inverse Synthetic Aperture Radar (ISAR) images, is emerging. Such manipulations can critically undermine navigational integrity through the generation of decoy targets, the strategic relocation of existing ones, and the effective concealment of additional targets. We have identified where an Advanced Persistent Threat (APT) could be concealed within the processing chain of a modern radar system. This study demonstrates the potential of APTs to exploit Generative Adversarial Networks (GANs) for the creation of deceptive ISAR images, thereby spotlighting this previously unexplored threat vector. Our findings offer novel insights into bolstering maritime cybersecurity in an increasingly AI-dominated landscape.

Index Terms—GAN, ISAR, Radar, Maritime Surveillance, Threat, Cybersecurity, Supply Chain Attacks

I. INTRODUCTION

The critical role of radar systems in ship navigation underscores the complexity and interconnectivity inherent within maritime systems of systems paradigm. The amalgamation of diverse subsystems sourced from various suppliers, including corporate entities and research institutions, enhances this complexity. Often, these subsystems are designed based on overarching but not well-defined ship specifications, resulting in shared standards across multiple shipyards. Consequently, beyond the well-recognized challenge of supply chain attacks, it becomes an arduous task to secure the ship’s sensor network with a holistic and well-structured approach. Additional threats may arise from disloyal crew members or infiltrators who have the capacity to compromise the system from within. A skilled adversary could potentially inject deceptive radar images at different points within the ship’s processing chain, including Inverse Synthetic Aperture Radar (ISAR) images. The ISAR is a radar technique used for generating high-resolution images of non-cooperative targets by exploiting the target’s relative motion with respect to the radar. In Synthetic Aperture Radar (SAR) systems, the radar platform or antenna moves to create a synthetic aperture, which allows for high-resolution imaging. However, in ISAR, the target itself is in motion, such as an aircraft or a ship. This relative motion between the radar and the target is utilized to create the synthetic aperture effect. The digital subsystems of a modern ship are software-defined and installed on specific industrial computers, certainly general-purpose with the computational capabilities of modern architectures. An APT (Advanced Persistent Threat) is a sophisticated, long-term cyber attack, often state-sponsored, aimed at stealing data or compromising systems over an extended period. We have therefore formalized the effects of an APT that has sufficient computational resources to generate synthetic ISAR images. As ISAR images afford detailed target and obstacle classification, their manipulation poses a substantial risk. This paper embarks on a journey to explore the threat modeling of the naval system, specifically investigating the feasibility of generating credible ISAR images using Generative Adversarial Networks (GANs). Our findings shed light on new dimensions of cybersecurity in the maritime domain, offering novel insights for enhancing navigational integrity and safety.

In Section II, we do a literature review about both the cybersecurity risk and the radar image generation with GANs. Section III delves deeper into the cyber threat, while Section IV presents the ISAR imaging and the framework that we propose for the fake images synthesis. In Section V, the real data used for the experiment and the obtained results are shown. Finally, in Section VI, conclusion will be drawn.
II. RELATED WORKS

The driving force of this work is the potential for an attacker to inoculate an APT within a naval radar system. While previous research of Longo et al., Kessler and Shepard [3], [4] has already explored this possibility, and Cherminod, Cantelli-Forti et al. [5], [6] proposed solutions to industrial networking attacks via Software Defined Networks, our study takes a step further. We demonstrate that a compromised radar system can exploit GANs to produce deceptive ISAR images inside the processing unit, a threat not yet considered. A Synthetic Aperture Radar (SAR) is a remote sensing technology that uses radar to create high-resolution images of targets by measuring the time delay and amplitude of the radar signal reflected from the target area. There are two primary classifications of Synthetic Aperture Radar (SAR) simulators: raw signal simulators, which provide SAR raw data for processing, and image simulators, which directly offer focused images [7]. Note that the creation of precise computer-aided design (CAD) models for each target in SAR scenes is not practical due to the extensive scene swaths and the presence of noncooperative targets. To overcome this challenge, Guo et al. [8] proposed a novel SAR simulation method using GANs. This approach directly generates SAR images based on real samples without relying on any prior assumptions about the targets. Another significant development in SAR simulation is the work of Zhou et al. [9]. They created an ISAR simulation imagery dataset using the bidirectional analytic ray tracing (BART) [10] electromagnetic scattering calculation tool. To address the challenges posed by limited training samples and rapid angular variations in SAR target recognition, Song et al. [11] introduced an adversarial autoencoder method for SAR image generation. They conducted experiments using the moving and stationary target recognition (MSTAR) dataset [12], leveraging the adversarial autoencoder to generate SAR images with diverse aspect angles. In an effort to enhance training stability, Zheng et al. [13] proposed a model called Multidiscriminator GAN (MDGAN). The MDGAN tackles this issue by employing multiple discriminators trained in parallel. Furthermore, Liu et al. [14] utilized CycleGAN to refine simulated images generated by tools like RaySAR. Their objective was to minimize the disparities between these simulated images and real images.

III. THE CYBER THREAT

A. Cyber Resilience: A Cultural Challenge in Defense

Advanced radar technologies, provided by defense sector companies often in collaboration with research institutions and universities, require a pervasive “cyber resilience” culture from the outset of projects. This culture should span horizontally across specialized internal teams, rather than being confined to isolated tasks. Our practical experience shows that the absence of this culture can lead to cyber resilience being treated as a time-limited task, sometimes even before the subsystems needing protection are identified. This lack of understanding can result in logical aberrations, such as the premature enumeration of hardware components for a secure, software-defined radar system. The intricate hardware-software architecture, often based on Field-Programmable Gate Array (FPGA) systems, should be developed by specialized researchers unrestricted in their design choices.

Further complicating matters, the corporation in charge of shipbuilding might be in competition with sensor suppliers in other industrial sectors such as radar systems, prompting them to withhold full project specifications. This results in situations where cyber resilience teams may lack knowledge about which systems share the same hardware, rack, ship area, or local network.

B. APT Hide and Seek: Unusual Residences for Advanced Persistent Threats

Fig. [1] illustrates the key components of a modern, software designed, radar system. These components include the antenna unit, signal processing unit, system manager, and associated components involved in processing, control, and networking. The diagram depicts the flow of signals and information within the radar system, achieving tasks such as beamforming, signal processing, and data visualization. The dashed square on the right encompasses the Processor Unit that is the potential entry point that cyber attackers could exploit by means of an APT. In the attack scenario considered, the installation of the APT should occur within the “signal processing” block. This assumes that the system is disconnected and unreachable from the outside. In this context, the APT analyzes the scenarios seen by the radar and alters them according to procedures established during the creation of the APT code. The signal processing of a modern radar partially takes place on FPGAs and partially on general-purpose CPUs. FPGAs are integrated circuits that can be configured post-manufacturing to perform complex functions and handle tasks such as interfacing with the Analog-to-Digital Converter, the Digital Down-Converter, or the Digital Beamforming, and a significant portion of the signal processing (adaptive filtering and range doppler map, part of the ISAR processing). The CPU takes care of tracking and is part of the ISAR processing. In the case of Supply Chain compromise, this flexible reconfigurable computing capability of FPGAs makes them a prime target for APTs seeking to manipulate radar systems. In fact, the presence of an FPGA shifts the “cyber perimeter”, eliminating the possibility of malicious code installation but also preventing the implementation of Host System and Network Security probes [3] so making the attack difficult to mitigate [6].
IV. METHODS AND MATERIALS

A. ISAR Imaging

The ISAR technique plays a decisive role in maritime surveillance [15], enabling high-resolution imaging of maritime targets for identification and threat assessment purposes, regardless of weather conditions or the time of day. As the targets are usually non-cooperative, a motion compensation scheme must be applied for removing the target motion prior to the imaging process. As illustrated in Fig. 2, the instantaneous location of a hypothetical target varies with respect to time $t_i$, and $R_0(t_i)$ denotes the distance between the target and receiver. The radar aperture is synthetically formed by the varying view angle $\theta(t_i)$, thus enabling high cross-range resolution for radar imaging. By combining that with the high range resolution achieved by wide radar signal bandwidth, the range-Doppler (RD) image of the target is formed (example of ISAR image provided in Fig. 3).

It is important to highlight the differences between ISAR and optical images. From a geometrical perspective, the motion-induced image in ISAR significantly differs from its optical counterpart. For example, to obtain the equivalent target side view illustrated in Fig. 4, the target must undergo a pitch motion in the direction of the radar observer, that is $90^\circ$ off the view angle of the optical case.

The utilization of AI-based image generation in ISAR images is influenced by this because it is not possible to know the appearance of the target in advance and thus it is very hard to identify features to characterize it. In optical images, objects typically exhibit identifiable visual characteristics that can be utilized for both generation and classification purposes.

B. GAN architecture

GANs are a class of machine learning models introduced by Goodfellow et al. [16]; they are designed to produce new data instances that resemble a given dataset, starting from a random input vector. GANs consist of two sub-networks: the Generator and the Discriminator, which are trained at the same time. The generator learns to generate images to fool the discriminator, whereas the discriminator tries to minimize the classification error between real and fake images. Through this adversarial process, two sub-networks reach an equilibrium point. The basic architecture of GANs is shown in Fig. 5.

C. The proposed framework

As mentioned earlier, GANs are adapted for generating ISAR images. It is important to note that our training dataset consists of 720 ISAR images, each containing $[512, 512]$ complex values. GANs are primarily designed for processing real image inputs rather than complex images, which is the case with ISAR images. Therefore, before training, the complex ISAR images are encoded and embedded into three-channel RGB images. The real part of the ISAR image is embedded in the R (Red) channel, the imaginary part in the G (Green) channel, and once again, the real part is embedded in the B (Blue) channel to introduce pixel variations in the Blue channel as well. Note that there are different ways to encode the complex image, and the approach mentioned above is a simple one that we have utilized. Training the GAN with image inputs of $[512, 512, 3]$ values allows the generator to learn fine details of the pixel variations. However, this can result in the generated images being different from the actual ISAR.
Fig. 4: How a profile view of a target is formed from ISAR and from the human eye.

images. To make the generated images more closely resemble the actual ISAR images, all images in the training set are downsampled to \([78, 78, 3]\) dimensions. After training, the generator output is then upscaled back to the original \([512, 512, 3]\) values.

Fig. 6 illustrates the architectures of the generator and discriminator networks utilized in this study. The generator network employs two transposed 2D convolutional layers to generate ISAR images from a random input vector of dimensions \([1, 1, 128]\). In order to capture distinctive features of true ISAR images, the generator network employs larger filter sizes \([20, 20]\) and fewer layers compared to the discriminator network. The discriminator network, on the other hand, utilizes a filter size of \([4, 4]\) in five convolutional layers, all with identical stride and padding values. GAN is trained using the training dataset, described in Section V-A, aiming to capture the specific features of each individual ISAR image. The generator network is trained to produce output with dimensions \([78, 78, 3]\) within the range of \([-1, 1]\) for each channel. The real and imaginary parts of the generated images are extracted from the first and second channels, respectively, and upscaled to \([512, 512]\). Subsequently, the real and imaginary parts are combined to obtain generated complex ISAR images. The results obtained from this process will be discussed in the results section.

Fig. 5: The basic architecture of GANs [17]

![Fig. 5: The basic architecture of GANs](image)

Fig. 6: The proposed architecture: (a) Generator (b) Discriminator

![Fig. 6: The proposed architecture](image)

Fig. 7: Google Earth view of the experimental setup of the NATO-SET-196 trials and image of the Astice A5379.

V. SIMULATION AND FURTHER ANALYSIS

A. Exploiting Real Data

This study utilized radar data acquired during the NATO SET-196 trials, which were carried out in 2014 at the Istituto Vallauri, Livorno, Italy. The results of these trials have been presented and analyzed in [18]. The experimental setup is depicted in Fig. 7. The data were collected using a multichannel FMCW [19] X-band radar known as PIRAD, with a central frequency of 10.7 GHz and a bandwidth of 300 MHz. One of the targets acquired during the trials was a training ship provided by the Italian Naval Academy, the Astice A5379 (Fig. 7). It measures 33.25 meters in length, 6.47 meters in width, 12 meters in height and it can reach a maximum speed of 12 knots. This specific target has been selected for the experiment presented in this work. All the ISAR images used for the training of the GAN were extracted from the raw data acquired during this measurement. A tracking algorithm and the information from the onboard IMU (Inertial Measurement Unit), were used to ensure that Astice was the only ship represented in the training set. The first step is to create a range-Doppler map from the raw temporal data. The range-Doppler (RD) map is a visualization technique used in radar
systems to display the spatial and frequency information of radar echoes. It provides a two-dimensional representation of the targets detected by the radar, showing their range and their doppler frequency (i.e., radial velocity), therefore it is a valuable tool for radar operators and analysts as it provides a visual representation of the targets’ locations and velocities. They are commonly used in applications such as target detection, tracking, and identification.

B. Generated images

Despite the visual similarities between the real and generated ISAR images when compared in absolute value, some differences could not be overcome even with fine-tuning the GAN. In the right side of Fig. 8(a) and 8(b), different patterns can be observed in the real component of the extracted ISAR images and the generated ones. In particular, to each positive peak in the extracted image, a corresponding negative peak is observed in its proximity, while in the generated images this characteristic cannot be found. The same difference is observable also in the imaginary component. After a thorough analysis, it became evident that the generated images lacked this sinusoidal pattern that was always noticeable in the extracted ones. To solve this issue, we propose the following processing steps to apply to all the generated images.

- A two-dimensional Fast Fourier Transform (FFT)
- A circular shift to match the shift (or modulation) of the extracted images
- An inverse two-dimensional FFT to obtain the ISAR image

The final result can be observed in Fig. 8(c).

C. Quality Assessment

To ensure a meaningful comparison between the generated fake ISAR images and the real ones, it is essential to employ an appropriate metric when assessing their quality. The structural similarity index (SSIM) [20] can be used for this context, as it measures the similarity between two ISAR images based on three parameters, namely, luminance, contrast, and structural characteristics, specific to radar imaging. By measuring these three parameters of the ISAR images, which take into account both radar signatures and also scattering properties, SSIM provides a scalar value between 0 and 1, where 1 represents a perfect match. This allows for a quantitative assessment of the similarity between the GAN-generated ISAR images and the real reference ISAR images, based on the unique characteristics of ISAR imaging. It should be noted that the original SSIM index is not applicable to complex-valued images. Therefore, in this context, we consider only the magnitude of the images. Furthermore, while the SSIM index is typically used to compare two images, we have two sets of images in this scenario: the real images from the input training dataset and the generated images from the GAN output.

We first calculate the SSIM indexes between the training images. Fig. 9(a) illustrates the histogram of the SSIM index computed between the training images. We can see that the most frequent SSIM indexes between the training images are in the range of [0.76, 0.78], close enough to 1, i.e., the full match. Next, we separately calculate the SSIM indexes between the three generated images and every single image in the training dataset. In Fig. 9 (b)-(d), three histograms are shown, which correspond to these three fake images. These histograms, exhibiting peak index values close to that of the training images and relatively large in size, indicate a high similarity between all three generated fake images and
the real images. This shows the capability of the proposed GAN to generate images that closely resemble the real image successfully.

It is worth noting that SSIM is not the only metric used in evaluating image quality; the peak-signal-to-noise ratio (PSNR) has been consistently considered as an alternative. PSNR value increases as the mean square error (MSE) approaches zero, indicating higher image quality.

D. Limitations

While our research findings demonstrate the feasibility of generating fake ISAR images, it is essential to consider several limitations. Firstly, the generation of AI-based images requires a significant number of ISAR images featuring different targets. However, due to limited access to real data, we were only able to utilize a single target in our study. To create a genuinely threatening system, it is crucial to incorporate diverse targets with varying shapes and sizes into a real-world application.

Furthermore, the ISAR imaging mechanism poses a challenge as the appearance of the same target can vary significantly depending on the geometry and dynamics of the radar-target system. Therefore, multiple distinct views of the same target would be required to accurately capture its characteristics. To overcome these limitations, it is necessary to prepare a larger and hence more diverse dataset of ISAR images. This implies gathering data from multiple targets with different characteristics, as well as considering different radar-target system configurations.

VI. Conclusions

The objective of this study was to assess the feasibility of generating fake complex-valued ISAR images using GANs for potential injection into a naval control system. The implications of such a threat in a maritime scenario are significant, as it could prompt a ship to alter its course or even engage in missile firing under certain circumstances. To some extent, in the realm of modern warfare, we are witnessing a new form of attack that digitally modifies the number and quality of military targets, reminiscent of the “Terracotta Army” strategy of the first Chinese empire, but executed in the cyber domain.

A. Future works

To enhance this work, we can strive to generate more realistic fake images on one hand, and concurrently seek technology to mitigate the risk and detect them, in a typical cat-and-mouse chase characteristic of cyber-physical scenarios. Regarding the first objective: while understanding the amplitude information of ISAR images is relatively straightforward, both for humans and neural networks, the same cannot be said about the phase of these images. Therefore, it would be interesting to explore how the GAN uses the phase information. About the second goal, further attempts to develop techniques for recognizing decay images based on their complexity shall be pursued. It would also be beneficial to address civilian applications in which similar cybersecurity risks could impact non-military industries, such as transportation, infrastructure, or telecommunications.

References