Security Issues in Situational Awareness: Adversarial Threats and Mitigation Techniques

Arslan Munir, *Senior Member, IEEE,* Erik Blasch, *Fellow, IEEE,* Alexander Aved, *Senior Member, IEEE,* Edward Paul Ratazzi, *Senior Member, IEEE,* and Joonho Kong, *Member, IEEE*

Abstract—Situational awareness (SAW) is of paramount significance for successful operations in many domains, such as surveillance, humanitarian, and search and rescue missions. SAW, however, is susceptible to adversarial attacks. This article contemplates adversarial threats and attacks on SAW systems and discusses various mitigation approaches.

1 Introduction

Situational awareness (SAW) is defined as "keeping track of prioritized significant events and conditions of one's environment", which includes the perception of entities in the environment, comprehension of their meaning, and the projection of their status in the near future [1], [2]. From a tactical perspective, SAW complements situation assessment (SA), which refers to the capability to understand the current and future dispositions of entities and threats within a volume of space. Space SAW is defined in the U.S. Department of Defense (DoD) dictionary of military and associated terms as: "The requisite foundational, current, and predictive knowledge and characterization of space objects and the operational environment upon which space operations depend" [3].

SAW is of paramount significance for military and the Air Force and is regarded as the decisive factor in military and air combat engagements. For example, survival in combat dogfight is heavily dependent on SAW as it relies on perceiving the enemy's aircraft current movement and predicting its future action fractions of a second before the enemy perceives his/her aircraft's movement himself/herself. The combat example was used to motivate SAW, which is regarded as tantamount to the "observe" and "orient" stages of the observe-orient-decide-act (OODA) loop, described by the United States Air Force (USAF) war theorist Colonel John Boyd [4]. SAW is also an important part of military command and control (C2), which can be construed as consisting of SAW, planning, tasking, and control. SAW is also indispensable for dismounted operators. Dismounted operators rely on SAW to perceive, comprehend, and project the entities in the environment in order to adapt their actions for efficient engagement with the red forces. Similarly, pilots need to be equipped with an avant-garde SAW system to better engage with red aircraft and cope with other strenuous situations such as higher levels of aviation traffic, harsh weather (e.g., storms, fog), and presence of an increasingly large number of unmanned aerial vehicles (UAVs) in the airspace.

Although SAW is indispensable for piloted devices (e.g., planes, cars, tanks), SAW devices and systems[1] providing information to humans are susceptible to adversarial threats and attacks that can compromise the security and trust of SAW systems. These attacks include both passive and active attacks that target sensors, communication links, electronics, and artificial intelligence (AI) of SA devices, equipment, and systems. To safeguard SA devices and systems against these adversarial attacks and to institute trust in these SAW systems, approaches need to be adopted to mitigate these adversarial attacks on SA and SAW. This article examines the security and trust issues in SA/SAW and contemplates different approaches for alleviating the adversarial attacks on SAW systems towards enhancing SAW trust.

2 Security and Trust of Situational Awareness

Security and trust of SAW is imperative for the commanders, operators, and pilots relying on SA for perception, comprehension, and prediction of entities in the surroundings based on which decisions are taken regarding engagement and/or appropriate response to the situation. SAW is often portrayed to commanders in terms of a common operating picture (COP), while an operator engaged in SAW is supported with a user-defined operating picture (UDOP). The UDOP is an evolution of COP that enables an operator to dynamically assemble his/her own view of information. The U.S. DoD dictionary of military and associated terms defines COP as: "A single identical display of relevant information shared by more than one command that facilitates collaborative planning and assists all echelons to achieve situational awareness" [3]. The COP is typically a single display shared by more than one command teams where each command team is in charge of their relevant aspect of the operating picture. For example, the air picture would be the responsibility of the air command. The information displayed in COP is obtained from various information sources. However, different aspects of SAW are susceptible to security attacks as depicted in Figure 1 and the integrity of COP can

[1] SAW systems are typically an aggregate of individual SAW devices networked together in a wired and/or wireless manner.

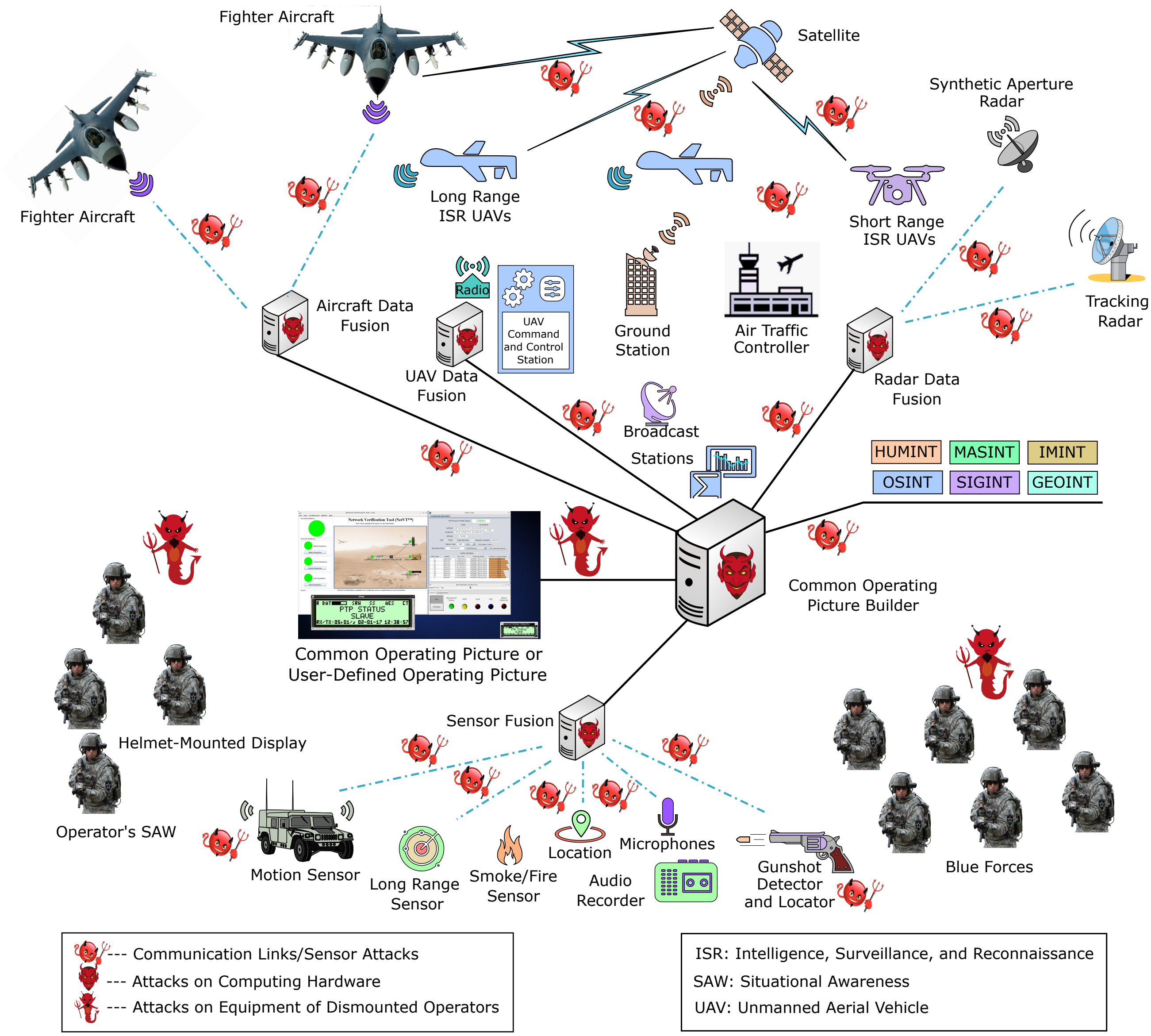

Fig. 1: Security issues in situational awareness.

be negatively affected by faults (hardware and software), security attacks, and/or misperception or misprojection of information. Hence, there is a need for UDOPs to enable each separate group to tailor their interpretation of the situation to mitigate a single misinterpretation propagated to the single COP.

Figure 1 depicts security attacks on different SA components that can affect the integrity and trust of the COP. In many cases, distributed UDOPs feeding a COP can provide enhanced trust through human-machine teaming. The cybersecurity attacks on SAW can be mainly categorized for four main components of SAW: attacks on communication links, attacks on sensors, attacks on computing hardware and software of SA infrastructure, and attacks on equipment of dismounted operators and pilots. Communication links, both wired and wireless are vulnerable to security attacks. As shown in Figure 1, an attacker can compromise the communication links between various entities, such as edge sensors, satellites, UAVs, aircraft, and ground station, etc. An attacker can also compromise computing hardware and software of SA infrastructure whether it be sensor data fusion center, vehicle data fusion center, information fusion center, and command and control station. Finally, sensors and equipment, such as heads-up displays (HUDs) and helmet-mounted displays (HMDs), carried by dismounted operators can be compromised by an attacker.

Considering that different aspects of SAW can be attacked, there is a possibility that the situation reports compiled by an operator and the SAW acquired by the COP/UDOP builder are subjective, imprecise, and compromised. Consequently, the integrity of resultant COP may become questionable. For COP to have integrity, a level of consistency is needed across the UDOPs in terms of measurements, methods, and values within the COP. The impact of security attacks on SA devices and SAW UDOP can be demonstrated through an example of a compromised link. During an operation, information is displayed on COP from multiple systems, that is, geographical information system, blue force tracking, aircraft position and flight path monitoring system, etc., are collated into a single air picture (or COP). If one of the links to the air picture, say aircraft

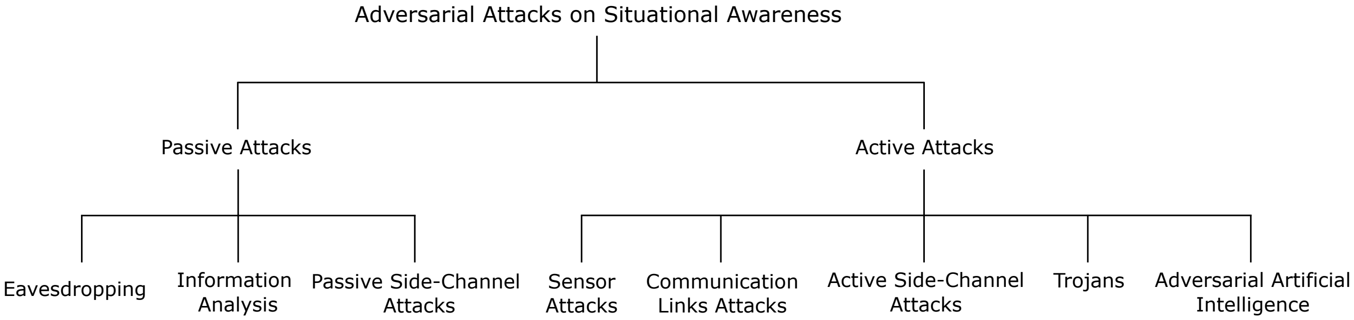
Adversarial Attacks on Situational Awareness
Passive Attacks
Active Attacks
Eavesdropping
Information Analysis
Passive Side-Channel Attacks
Sensor Attacks
Communication Links Attacks
Active Side-Channel Attacks
Trojans
Adversarial Artificial Intelligence

Fig. 2: Classification of adversarial attacks on situational awareness.

positions and flight paths, goes down or is compromised, then it can impact the COP and the air traffic controller's (ATC's) SAW as follows [5], [6]:

1) If the *COP continues to display the last available aircraft positions and flight paths*, the COP will no longer be accurate. If the ATC discerns the link failure, he/she will know that there are aircraft that can affect the operation. In case the ATC fails to recognize that the link is down, the commander will continue to trust the corrupted COP as real-time information.
2) If the *COP deletes the last available aircraft positions and flight paths*, then the COP will no longer be accurate. The ATC in this case will not be certain whether there are actually no aircraft flying through the area of interest or no aircraft are displayed in the COP because the link is down.

If COP cannot be trusted, the commander or the ATC may not make appropriate decisions, which can impact the outcome of the situation and the safety of the troops and the first responders. It has been observed that a commander or ATC either opts to trust the COP (overtrust) or totally ignores it (distrust) [5]. For large COPs, it is likely that a single element of the COP is not 100% accurate because of some compromised elements (e.g., sensors, equipment), but since this impacts the integrity of the COP, the commander or ATC may disregard the entire COP as untrustworthy. This leads to accurate inputs to the COP being unnecessarily neglected, which adversely impacts SAW.

3 Adversarial Threats and Attacks on SAW

Adversarial attacks compromise the security and trust of SAW systems. Adversarial conditions and threats for SAW can be broadly classified into two categories: passive and active attacks. The passive and active attacks on SAW can be further classified into different categories as depicted in Figure 2. We clarify that this article is not a comprehensive survey of all potential attacks on SAW and Figure 2 is not an exhaustive taxonomy of attacks on SAW, though, the article covers salient threats and attacks on SAW.

3.1 Passive Threats and Attacks

Passive attacks are those where attackers employ non-disruptive and covert methods to avoid detection. Many security data breaches (e.g., breaches of sensitive data, plans or locations) are caused by passive attacks. We discuss common passive threats and attacks from both malicious (bad actor) and adversarial (opponent actor) players for the commander (good actor) in the following.

3.1.1 Eavesdropping

An adversary infiltrating the communication channels of sources providing SAW can carry out passive eavesdropping (e.g., sniffing and storing all the traffic for SAW). Thus, the adversary becomes privileged to the SA information that only a blue force commander or the legitimate authorities are supposed to receive. This eavesdropping compromises the SA process and hence the commander loses the strategic advantage in the combat or situation as he/she will not likely have a better SAW than the opponent and thus will not be able to get in the opponent's OODA loop.

3.1.2 Information Analysis

An adversary eavesdropping on the communication channels of sources providing SA can further capitalize on the eavesdropped information by performing traffic and information analysis, thus obtaining critical information about the strategies of opposing forces. For instance, from the information analysis of the eavesdropped information, the adversary can locate the sources of information and thus distribution of information assets of opposing forces, which paves the way for further active manipulation attacks.

3.1.3 Passive Side-Channel Attacks

To process the data securely by SA computing systems, handheld operators' equipment, Internet of battlefield things (IoBT) devices, and other sensitive systems; cryptography algorithms are often employed. Side-channel attacks circumvent the theoretical strength of cryptographic algorithms by exploiting weaknesses in the hardware implementation of a cryptographic system via nonprimary, side-channel inputs and outputs [7]. Commonly utilized side-channel outputs include power consumption, light, timing, electromagnetic (EM) emissions, and sounds.

3.2 Active Threats and Attacks

Active attacks are those in which an attacker endeavors to make changes to the data of a target system or the data en route to the target. In context of SA (Figure 1), an intruder can attack computer servers (e.g., data fusion servers, COP builder servers), sensors, operators' equipment, and any of the communication links. Common methods of active attacks include masquerade attacks, replay attacks, message modification attacks, message injection attacks, denial of service (DoS) attacks, and distributed DoD (DDoS) attacks. In *masquerade attacks*, an intruder masquerades or pretends as a particular user of a system to gain unauthorized access or greater privileges. In *message modification* attacks, an attacker alters a message header to direct it to a different destination or modifies the message content. An attacker injects malicious messages in the system or network in *message injection* attacks. In *DoS attacks*, an attacker deprives a user of a system or network by injecting more data or traffic that the system/network can handle. In *DDoS attacks*, a large number of compromised devices or systems, often referred to as botnet, conduct a large-scale attack on a single target. Active attacks are often overt in nature and victim becomes aware of the attacks as they occur. Both active and passive attacks can be used in combination to disrupt or gain unauthorized access to a system, network, or data. The sensitive SAW systems are also vulnerable to active and passive attacks as demonstrated by Hack the Air Force 4.0 challenge—a hacker-powered challenge to examine the security of the Air Force assets [8]. In Hack the Air Force 4.0 challenge, a team of 60 hackers managed to hack the Air Force systems and found 460 vulnerabilities in the pool of cloud-based servers and systems known as the U.S. Air Force Virtual Data Center. In the following, we discuss some of the active threats and attacks that can be utilized to compromise the SA of a system.

3.2.1 Sensor Attacks

A sensor attack, also known as *transduction attack*, exploits vulnerabilities in the physics of a sensor to manipulate its output or induce errors [9]. For example, malicious acoustic interference can affect the output of sensors in a variety of systems ranging from autonomous vehicles to handheld electronic equipment to medical devices. Researchers have shown that sound waves can alter the output of sensors, for instance, accelerometers [9]. Since many handheld electronic equipment carried by dismounted operators (e.g., radio, smart phones, etc.) include both a speaker and other sensors (e.g., accelerometer), an adversary can carry out transduction attacks without any special equipment.

3.2.2 Communication Link Attacks

A communication attack seeks to compromise the confidentiality, integrity, or availability of the signals. Communication links are most susceptible to adversarial attacks as compared to other components of an SAW system. Both the wired and wireless communication links can be attacked, although attacks on wireless links are relatively easier. The wired links can either be broken or tapped by an adversary. The broken wired links cause DoS from a particular component or subsystem of an SAW system. An adversary can not only eavesdrop (passive attack) through the tapped link but can also inject or alter messages in the communication link, thus facilitating active attacks including replay attacks, substitution attacks, and injection of viruses, worms, and malware. For wireless channels, adversaries can utilize jamming to thwart the information flow from sensors and other sources to SA receivers and computing systems.

3.2.3 Active Side-Channel Attacks

SAW systems and devices are also susceptible to active side-channel attacks. An active side-channel attack exploits side-channel inputs, such as supply voltage, temperature, sound, light, or environmental conditions to tamper, modify, or influence the targeted device or system in a way that bypasses security mechanisms either directly or leads to malfunctioning that in turn enables attacks. *Fault injection attacks* are a type of active side-channel attacks that applies glitches, which are fast changes in the signal (typically power supply or clock) supplied to a device, to affect its normal operation. Applying a clock glitch (a clock pulse much shorter than the normal) or a power glitch (a swift transient in supply voltage) affects only some transistors in the chip and leads to a few flip-flops adopting a wrong state, which can be exploited to make the processor execute a number of wrong instructions. *Electromagnetic fault injection* is another type of active side-channel attack that utilizes short, high-energy electromagnetic pulses to alter the state of memory cells, causing erroneous calculations. *Laser/Optical fault attacks* utilize a focused laser beam to change the state of a transistor in an integrated circuit (IC), causing bit flips in memory cells.

3.2.4 Software and Hardware Trojans

SAW devices and systems are susceptible to both software and hardware Trojans. Trojans can change the functionality, leak information from, and disable the operation of SAW devices and systems. A software Trojan is a malicious code or software that appears legitimate but can damage, disrupt, steal, or inflict some harm on a computing device, memory, data, or network. It is named as such because it looks like a bona fide application, and thus tricks a user into loading and executing the application containing Trojan. A hardware Trojan is a malicious addition or modification of the components in an IC that can change the functionality, decrease the reliability, disable or modify operation, and leak information. Hardware Trojans can severely impact the security and trust of SAW systems. Trojans can change the functionality or disable an SAW device and affect mission-critical components.

3.2.5 Adversarial Artificial Intelligence

There is a continuous evolution from information warfare to intelligent warfare. The result of future military conflicts will not be determined by who controls the information but rather who applies AI to the information, monitors it, harnesses it, and degrades it to attain potential goals [10]. Recent advances in AI have made AI an integral part of SA systems. AI assists operators, pilots, and commanders in developing SAW by perceiving the situation and then making projections about the future actions of entities in the environment. However, AI is also susceptible to adversarial attacks. Research has shown that several machine learning models, including deep neural networks (DNNs), are vulnerable to *adversarial examples*, which are carefully modified inputs, also known as perturbations, crafted to manipulate the system into generating a particular output. Figure 3 depicts an adversarial attack on the AI of an SA device/system, where the attacker perturbs the input data with carefully crafted data that appears like noise but results in misclassification of outputs by the DNN of the SA device/system. To generate such adversarial examples, several algorithms have been proposed, such as the fast gradient sign method (FGSM) and the Jacobian saliency map algorithm (JSMA) approach. The attack surface of AI ranges from adversaries attempting to manipulate the collection and processing of data, corrupt the model, or alter the outputs. AI of SAW systems is susceptible to both black-box (access to inputs and outputs) and white-box (access to model) attacks at training and inference phases.

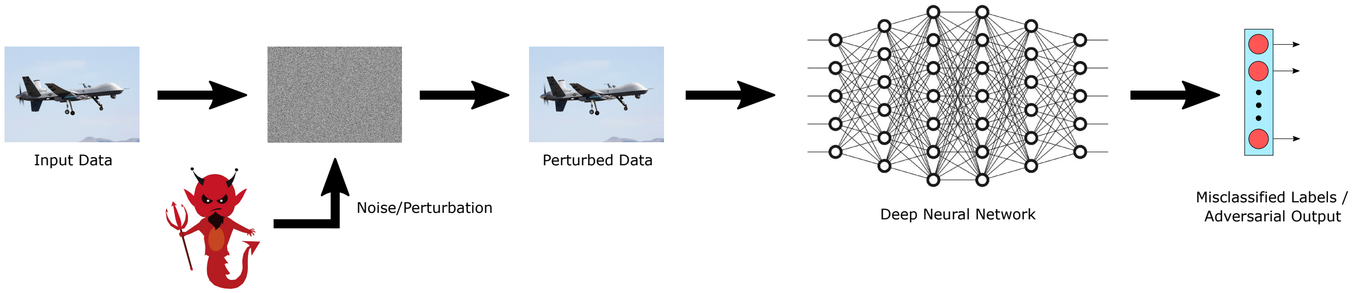

Fig. 3: Adversarial attacks on artificial intelligence of situational awareness systems.

Adversarial attacks on AI of SA assets can affect the integrity and availability of the system. The attacks attempting to manipulate the output are at the heart of integrity attacks, introducing false positives or false negatives in SA asset decisions. For example, an integrity attack can cause AI-based SA assets to misclassify resources, which can have detrimental ramifications for SAW and decision-making, and can change the outcome of a situation. Adversarial examples can also be exploited for misclassification of equipment, people, and behaviors. The attacks on availability aspire to reduce the quality (e.g., confidence or consistency), performance, or access (e.g., denial of service). For example, an adversary attacking an AI-based SA asset may cause the system to behave erratically or can cause the asset to stop working resulting in DoS. If vehicles, aircraft, and/or actuators are beguiled into taking or not taking actions based on adversarial examples, then catastrophic consequences would ensue. In short, the connotations of adversarial AI are significant leading to challenges in trust, interpretability, and explainability in decision-making.

4 Approaches for Mitigating Adversarial Attacks on SAW

Since security incursions on SAW can have excruciating ramifications, defending against these security attacks is of paramount significance to maintain the trust and integrity of SAW. This section discusses some approaches for mitigating the effects of adversarial attacks on hardware SA and human SAW, although we point out that this article is not a comprehensive survey of all approaches for mitigating attacks on SAW.

4.1 Robustness Against Sensor Attacks

Since sensors lie at the heart of SA systems, trustworthiness and robustness of sensors against adversarial attacks is of paramount significance for obtaining trustworthy SA. Safeguarding against transduction or sensor attacks is challenging because the manifestations appear as software or environmental symptoms (e.g., noise) whereas the problem lies in the physics of hardware. Standard security practices such as static analysis, fuzz testing, and signed software updates are inadequate to provide protection against transduction attacks. Three techniques for mitigation against transduction attacks have been recommended [9]: (i) transition from component-centric security to system-centric security and tolerance of untrustworthy components, (ii) continuously checking the output of sensor hardware for adversarial attacks, and (iii) manufacturing circuits to diminish the effects of resonance.

4.2 Anti-Jamming

Jamming attacks on SA information sources are a subset of DoS attacks that aim at blocking the legitimate communication by creating intentional interference in the communication networks. To address jamming issues, mechanisms are needed for jamming detection, localization, and countermeasures. Various techniques that can be utilized for anti-jamming include spatial retreat, consistency checks, and channel or frequency hopping [11]. For defense against jamming attacks by fast-following jammers, direct-sequence spread spectrum (DSSS) and frequency-hopping spread spectrum (FHSS) are often utilized. DSSS uses a wide bandwidth for signal transmission while FHSS hops through different frequency channels to avoid interference.

4.3 Symmetric and Asymmetric Cryptography

Symmetric and asymmetric (or public key) cryptography techniques can be utilized by SA systems to provide various security services including confidentiality, integrity, message authentication, non-repudiation, access control, and security auditing. Symmetric cryptography is widely used for data encryption and integrity checking of messages. Symmetric cryptography relies on a secret (symmetric) key that is in possession of legitimate senders and receivers to provide security services. Asymmetric cryptography is typically utilized for generating and distributing the secret key (a mechanism known as key establishment) for symmetric cryptography as well as additional security services, such as authentication, and digital signatures.

4.4 Hardware-based Security

Although traditional symmetric and asymmetric cryptography can provide many security services to integrate security and trust in SA systems, traditional cryptography suffers from various shortcomings, the most noteworthy being secure storage of secret keys and the distribution and handling of certificates by a trusted third party to potentially billion of devices. To address the shortcomings of traditional cryptography, hardware-based security techniques, such as those based on physically unclonable functions (PUFs), can be utilized by SA systems and devices. In particular, lightweight PUF-based authentication and key establishment protocols can be employed by SA devices and systems [12]. Hardware-based security techniques can alleviate the need for storing the secret keys in non-volatile memories of SA devices and can provide authentication and secret key establishment services at run-time. Some other services provided by hardware security primitives include IC metering and hardware watermarking. IC metering is a set of security protocols that enables a design house to attain post-fabrication control over their designs by uniquely tagging each chip to facilitate tracing the chips. Hardware watermarking embeds a tag in the design so that each chip produced by that design would carry the same watermark/tag.

4.5 Robustness Against Side-Channel Attacks

To mitigate side-channel attacks on SA systems or components, SA systems or components need to incorporate remedies against side-channel attacks. Different countermeasures have been developed for side-channel attacks which can be categorized into hiding, masking/blinding, design partitioning, and physical security. Side-channel attacks aim to recover a signal from the side-channel which typically has noise, and thus these attacks aim at boosting the signal-to-noise ratio (SNR) of the side-channel information as much as possible. Hence, one way of *hiding* the side-channel information is to increase the noise, which decreases the SNR of the side-channel information. Consequently, many researchers have proposed noise generator circuits to secure ICs [7]. Alternatively, logic gates have been designed whose side-channel emissions are independent of the data being processed. Low-power design techniques and physical shielding provide other techniques for hiding side-channel information. *Masking* or *blinding* techniques aim at removing the correlation between input data and the side-channel emissions. *Design partitioning* targets separating regions of the chip that operate on plaintext from the regions that operate on ciphertext. *Physical security* techniques aim at denying an attacker physical access to a sensitive system/device. The goal of *anti-tamper* techniques is to prevent an attacker from applying invasive techniques (e.g., decapsulation, reverse engineering) for extracting side-channel information.

4.6 Robustness Against Trojans

Software Trojans in SAW devices and systems can be detected and removed through anti-malware programs. Many of these malware programs help mitigate additional infection by disabling the communication between a Trojan and any backend server. Other good practices for preventing software Trojans in SAW devices include: executing periodic diagnostic scans, automatic updates for operating system (OS) software, installing latest security updates, regularly updating the applications, and patching the security vulnerabilities. The good practices for preventing software Trojans in SAW systems include the above-mentioned practices for SA devices as well as avoiding suspicious websites, being cautious of unverified attachments and links in unknown email, and installing a firewall [13].

Providing robustness against hardware Trojans in SAW devices and systems is more challenging as compared to software Trojans. To provide resilience against hardware Trojans in SA ICs, hardware Trojans need to be detected. Hardware Trojan detection approaches can be broadly classified into *destructive* and *non-destructive* approaches. The destructive techniques perform systematic delayering through chemical mechanical polishing and imaging of ICs via scanning electron microscope to detect the presence of additional circuitry in an IC. However, destructive approaches are extremely expensive and time consuming and do not scale well with increasing density of transistors on modern ICs. The non-destructive approaches can be classified into: (a) run-time monitoring approaches, and (b) test-time approaches [14]. Run-time monitoring approaches exploit redundancy in the circuit (built at the design time in design-for-security paradigm) to bypass an infected part of the circuit. Furthermore, SA devices and chips can be encapsulated in a self-destructive packaging which can either be externally triggered by an operator or internally triggered by a run-time Trojan monitor on detection of a malfunction. Test-time approaches can be further categorized into: (i) *logic-testing* based approaches, and (ii) *side-channel analysis* based approaches. Logic-testing approaches aim at test vector generation and application for activating a Trojan circuit and observing its pernicious effect at primary outputs of an IC. Logic-testing has its limitations for detecting sequential Trojans and for ICs comprising of a large number of tranistors/gates. Side-channel analysis approaches detect the presence of Trojans by observing side-channel parameters, such as leakage current, quiescent supply current, path-delay characteristics, electromagnetic radiation, etc., because insertion of additional gates of a Trojan circuit will increase the side-channel emissions from an IC. However, side-channel approaches can give erroneous indications due to process and environment noise.

4.7 Robustness Against Adversarial AI

Although various research studies have been conducted for crafting adversarial example attacks, the studies on defense techniques are still in infancy. We contemplate a **two-layer defense approach** for attacks against the AI of SAs. The two defense layers are: (1) Attack mitigation: design of defense techniques against known attacks; and (2) Model security: enhances ML model robustness against (unknown) attacks. For attack mitigation defense layer, redundancy and sparse approximation-based approaches (training and inference phase) are some of the examples. For model security layer, ensemble based approaches and explainable AI are some of the examples.

4.7.1 Redundancy and Sparse Approximation based Approach

This approach is based on the observation that the vulnerability of ML models including DNNs to adversarial examples primarily arises from the existence of rarely-explored subspaces [15]. In this approach, N_r *redundant* ML/DNN *models*, which we refer to as redundant modules for AI security (RMAIS), are activated along with the *main* ML *model* performing the inference where the N_r value can be selected based on the safety-criticality of the SA function. In the training phase, each ML model in RMAIS characterizes the explored subspace by learning the probability density function (PDF) of typical data points and marks the complementary regions as unexplored/risky [15]. In the inference phase, N_r redundant ML models evaluate the input sample in parallel with the main ML model and raise alarm flags if the input sample lies in any risky region. To mitigate the risk of an adversary adding structured noise to a legitimate sample such that the input sample is moved from one cluster center (corresponding to a class i) to some other cluster center (corresponding to some class j, $j \neq i$, thus resulting in misclassification), *dictionary learning* can be utilized to determine the peak signal-to-noise ratio (PSNR) of each incoming data and filter out atypical samples in the input space.

4.7.2 Ensemble based Approach

It has been observed in ML that given enough data, a more complex hypothesis class (e.g., non-linear classifier as opposed to a linear classifier) provides better prediction. Based on this observation, an ensemble method variation which combines N different ML classifiers to form a complex hypothesis can provide resilience against adversarial attacks. In this ensemble method variation, the input is supplied concurrently to an ensemble of ML models, and the output is majority voted. Each of the ML classifiers in the ensemble is trained on similar but not exactly the same data set. The intuition behind this approach is that not all the models in the ensemble will be vulnerable to the same adversarial example. The N value for this ensemble method can be selected based on the safety-criticality of SA system.

4.7.3 Explainable AI

Explainable AI can be exploited to provide resilience against adversarial attacks on the AI of SA assets. The autonomous or intelligent functions in SA assets (e.g., UAVs, IoBTs) can be made explainable so that the assets will explain the decisions taken in different scenarios. The explainable AI will give the operator insights into autonomous decisions made by the assets, thus enabling the commander to take control of the assets in case of wrong logic, malfunctioning, or attack scenarios.

5 Conclusions

This article discusses security and trust issues in SA and the impact of these issues on perceived SAW and decision-making. Design of secure SA/SAW systems needs understanding of how each layer of computation, from sensors to human behavior, can fail when subjected to adversarial attacks. This article contemplates various passive and active adversarial threats and attacks on SA systems. The article then discusses different approaches for mitigating adversarial attacks on SAW, such as robustness against transduction attacks, symmetric and asymmetric cryptography, hardware-based security, and AI security.

Although, we have presented some approaches for mitigating adversarial attacks on SA systems, there exist many challenges in developing defenses against adversarial attacks on SAW systems. For instance, comprehensive detection of hardware Trojan circuits of arbitrary sizes in a multi-million gate design remains an intractable problem. Both destructive and non-destructive approaches for hardware Trojan detection have their limitations and there is no silver bullet solution. Adversarial attacks on AI of SAW systems present another avenue where further research is needed. Finally, tradeoffs exist in balancing security with constraints on performance, area, and cost of SA devices and systems.

Acknowledgments

This research was supported in part by the Air Force Research Laboratory (AFRL) Information Directorate (RI), through the Air Force Office of Scientific Research (AFOSR) Summer Faculty Fellowship Program®, Contract Numbers FA8750-15-3-6003, FA9550-15-0001 and FA9550-20-F-0005. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the AFRL and AFOSR.

References

[1] M. R. Endsley, "Situation Awareness in Aviation Systems," in *Handbook of Aviation Human Factors*, D. J. Garland, J. A. Wise, and V. D. Hopkin, Eds. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers, 1999, pp. 257–276.

[2] E. Blasch, E. Bossé, and D. A. Lambert, Eds., *High-Level Information Fusion Management and Systems Design*. Artech House, 2012.

[3] DOD, "DOD Dictionary of Military and Associated Terms," https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf, July 2019, Last visited on August 16, 2019.

[4] B. McKay and K. McKay, "The Tao of Boyd: How to Master the OODA Loop," https://www.artofmanliness.com/articles/ooda-loop/, May 2019, Last visited on August 14, 2019.

[5] J. Robertson, "Integrity of a Common Operating Picture in Military Situational Awareness," in *Proc. of Information Security for South Africa (ISSA)*, Johannesburg, South Africa, August 2014.

[6] E. Blasch, R. Sabatini, A. Roy, K. A. Kramer, G. Andrew, G. T. Schmidt, C. C. Insaurralde, and G. Fasano, "Cyber Awareness Trends in Avionics," in *IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*. San Diego, California: IEEE, September 2019.

[7] K. Mai, "Side Channel Attacks and Countermeasures," in *Introduction to Hardware Security and Trust*, M. Tehranipoor and C. Wang, Eds. Springer, 2012, pp. 175–194.

[8] D. Winder, "U.S. Air Force Successfully Hacked by 'Battalion' of 60 Hackers," https://www.forbes.com/sites/daveywinder/2020/04/16/us-air-force-successfully-hacked-by-battalion-of-60-hackers/#1179ab8e39f9, April 2020, Last visited on July 15, 2020.

[9] K. Fu and W. Xu, "Risks of trusting the physics of sensors," *Communications of the ACM*, vol. 61, no. 2, pp. 20–23, February 2018.

[10] J. Burton and S. R. Soare, "Understanding the Strategic Implications of the Weaponization of Artificial Intelligence," in *Proc. of International Conference on Cyber Conflict (CyCon)*, Tallinn, Estonia, May 2019.

[11] K. Grover, A. Lim, and Q. Yang, "Jamming and Anti-jamming Techniques in Wireless Networks: A Survey," *International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC)*, vol. 17, no. 4, pp. 197–215, 2014.

[12] M. A. Qureshi and A. Munir, "PUF-RAKE: A PUF-based Robust and Lightweight Authentication and Key Establishment Protocol," *IEEE Transactions on Dependable and Secure Computing (TDSC)*, 2021.

[13] Malwarebytes, "Trojan," https://www.malwarebytes.com/trojan/, July 2020, Last visited on July 27, 2020.

[14] S. Narasimhan and S. Bhunia, "Hardware Trojan Detection," in *Introduction to Hardware Security and Trust*, M. Tehranipoor and C. Wang, Eds. Springer, 2012, pp. 339–364.

[15] B. D. Rouhani, M. Samragh, M. Javaheripi, T. Javidi, and F. Koushanfar, "DeepFense: Online Accelerated Defense Against Adversarial Deep Learning," in *Proc. of the International Conference on Computer-Aided Design (ICCAD)*, San Diego, California, November 2018.

Arslan Munir is currently an Associate Professor in the Department of Computer Science at Kansas State University. His current research interests include embedded and cyber-physical systems, secure and trustworthy systems, and artificial intelligence. Contact him at amunir@ksu.edu.

Erik Blasch is a program officer at the United States (US) Air Force Research Laboratory (AFRL)—Air Force Office of Scientific Research (AFOSR) in Arlington, VA. He is an AIAA Associate Fellow, SPIE Fellow, and IEEE Fellow. Contact him at erik.blasch.1@us.af.mil.

Alexander Aved is currently a technical advisor at the Air Force Research Laboratory Information Directorate in Rome, NY. Alex's research interests include multimedia databases, stream processing, and dynamically executing models with feedback loops. Contact him at Alexander.Aved@us.af.mil.

Edward Paul Ratazzi is currently a technical advisor to the Air Force Research Laboratory's Information Exploitation and Operations Division in Rome, NY. In this position, he is the Senior Advisor to the Division Chief, where he provides oversight and direction to the Division's portfolio spanning interests in cyber agile and resilient architectures and systems, automation of cyber operations, cyber vulnerability analysis, and cyberspace/SIGINT integration. Contact him at edward.ratazzi@us.af.mil.

Joonho Kong is currently an Associate Professor in the School of Electronics Engineering at Kyungpook National University, South Korea. His research interests include computer architecture, embedded system, and hardware/software co-design. Contact him at joonho.kong@knu.ac.kr.