future internet

MDPI

Article

Lightweight AI-Based Attack Detection for LED VLC in Multi-Channel Airborne Radar Systems

Vadim A. Nenashev [1], Vladimir P. Kuzmenko [2], Svetlana S. Dymkova [3,4,*] and Oleg V. Varlamov [3,4]

[1] Department of Design and Technology of Electronic and Laser Devices, Saint-Petersburg State University of Aerospace Instrumentation, 190000 St. Petersburg, Russia; nenashev.va@gmail.com
[2] Department of Electromechanics and Robotics, Saint-Petersburg State University of Aerospace Instrumentation, 190000 St. Petersburg, Russia; mr.konnny@gmail.com
[3] Scientific Research Department, Moscow Technical University of Communications and Informatics, 111024 Moscow, Russia; vov@mtuci.ru
[4] Institute of Radio and Information Systems (IRIS), 1010 Vienna, Austria
* Correspondence: ds@media-publisher.eu

Abstract

Compact multi-channel airborne radar stations increasingly rely on an LED-based visible light communication (VLC) service link under radio-frequency spectrum restrictions and strict end-to-end delay constraints. Despite the directional nature of optical links, the VLC channel remains vulnerable to active optical interference and signal injection; furthermore, when an AI-enabled integrity monitor is embedded into the receiver, the AI decision layer becomes a direct target of evasion and online poisoning. This paper proposes a lightweight, interpretable AI-based attack detection architecture in which a Poisson photon-counting observation model is used to form physically consistent features over the preamble and control-sequence interval, while the final decision is produced by an AI ensemble combining a monotonic logistic detector and a one-class detector. The considered threat profile includes sustained illumination and synchronized flashes (jamming/blinding), spoofing via false preambles, replay of recorded fragments, and online data poisoning during self-calibration. The adequacy of solutions is assessed using the detection probability P_D (ensemble: $P_D \geq 0.90$ for DC-jamming mean-count increment $\Delta\lambda_{DC} \approx 7.56$, pulsed-interference mean-count increment $\Delta\lambda_{pulse} \approx 12.89$, and spoofing signal-scaling factor $\alpha \approx 1.02$), the false-alarm probability $P_{FA} = 0.045$, and the per-packet end-to-end latency (bounded by the observation-window duration $L\Delta T = 20$ μs, where window length $L = 20$ and interval duration $\Delta T = 1$ μs), which confirms real-time CPU operation without GPU acceleration.

Keywords: secure artificial intelligence; lightweight anomaly detection; visible light communication (VLC); light-emitting diode (LED); multi-channel airborne radar

Academic Editors: Peter Kieseberg and Jungwoo Ryoo

Received: 28 January 2026
Revised: 25 February 2026
Accepted: 27 February 2026
Published: 28 February 2026

1. Introduction

Modern airborne radar stations for small carriers are increasingly implemented as multichannel (MIMO-type) systems with several parallel transceiver paths, digital beamforming, and spatio-temporal processing [1–4]. MIMO (Multiple-Input Multiple-Output) is a wireless technology that uses multiple transmitting and receiving antennas for simultaneous data transmission. MIMO architecture increases the bandwidth of the communication channel and improves connection stability through the use of spatial multiplexing.

This architecture improves angular resolution, enhances interference robustness, and enables reliable detection of low-contrast targets, including ground objects with a small effective scattering area (ESA), which is essential for civil search-and-monitoring missions. At the same time, multichannel operation substantially increases the volume of internal data exchange between receive modules, digitization units, beamforming/processing blocks, and the central computing module. In addition to radar data, the amount of service traffic grows (mode-control commands, synchronization markers, status telemetry, and aggregated primary-processing outputs). For small onboard radar stations (SORA), external connectivity to ground infrastructure also becomes more demanding, and bandwidth/latency constraints are frequently reported as a bottleneck when transferring complex radar modes to small platforms [5–8].

In response to these challenges, increasing attention has been devoted to optical wireless communication (OWC) links in the "SORA carrier–ground infrastructure" segment. Such links use laser or LED emitters as radiation sources and can provide high service data throughput without competing for the congested radio-frequency spectrum and without being directly affected by electromagnetic interference in the radar operating environment [9–11].

Within optical wireless technologies, LED-based visible light communication (VLC) occupies a distinct position. Compared with narrow-beam laser links, VLC enables a combination of illumination (or optical beaconing) and data transmission using the same emitter, while maintaining a practically stable communication link whenever line-of-sight visibility is available. Surveys of VLC solutions relevant to multichannel SORA indicate that the technology has matured in recent years for both indoor (enclosed) and outdoor (open-space) deployments, including scenarios where channel conditions vary due to geometry and ambient illumination [5,12–15].

For multichannel SORA platforms, this is particularly attractive because the carrier typically establishes a line-of-sight interval with the ground station during operation, and LED emitters can simultaneously serve as a service optical beacon and as a dedicated service data channel. Existing studies on OWC/VLC links for airborne or mobility-relevant settings demonstrate that stable transmission in the visible and near-infrared ranges can be achieved in the "SORA carrier–ground infrastructure" channel by applying appropriate compensation for atmospheric attenuation and geometry-induced distortions.

As a result, a natural two-link communication architecture emerges for SORA carriers. The radio-frequency (RF) channel supports the conventional exchange of telemetry and control commands and, when required, the transfer of radar products. The LED-based VLC channel provides a dedicated low-emission service path for transmitting operating parameters of the radar and the carrier, synchronization markers, results of primary processing, and other service/measurement data. In a multichannel SORA configuration, this separation allows VLC to carry time-critical service information under favorable optical conditions. The RF channel remains the universal link and a fallback path when VLC conditions degrade [16,17].

For a long time, VLC was considered more secure at the physical layer due to the geometrically limited reception area and the inability of light to propagate through opaque obstacles. However, recent VLC-security studies show that practical deployments remain exposed to a broad range of attacks. These include passive interception via reflected components or side-lobe radiation, as well as active illumination of the photodetector with high-intensity optical radiation that can push the receiving path into non-linear operation or saturation. Attacks may also involve injecting false optical packets into the receive path (e.g., by replacing the preamble or the payload) and other physical layer impacts that deliberately distort the power, spectrum, or temporal structure of the optical signal, thereby degrading

communication quality. A number of works propose physical layer security mechanisms (e.g., non-Lambertian patterns and controllable reflective surfaces), but these approaches are mainly developed for stationary rather than mobile SORA platforms [18–20].

Despite the often-cited physical layer security of VLC due to spatial confinement, recent studies show that practical VLC links remain vulnerable to active adversarial impacts. An attacker does not need to decode the payload to cause harm. It is sufficient to manipulate the optical front-end and, consequently, the photon-count statistics observed at the receiver. In open-space deployments, an external light source can inject continuous illumination or time-aligned pulses, which bias the receiver and distort preamble- and control-segment statistics. An adversary can also inject structured optical packets (spoofing) or retransmit previously recorded legitimate fragments (replay), thereby violating message freshness and logical consistency.

In the considered compact airborne radar architecture, the service VLC channel is used to deliver synchronization markers, telemetry, and mode-control information. Therefore, the dominant security objective is not confidentiality but the integrity and availability of the service link: even short disturbances or subtle manipulations can trigger incorrect mode transitions, desynchronize subsystems, or degrade coordination between sensing and control components, ultimately affecting radar functionality and platform safety.

Accordingly, the threat model is formulated in terms of attacks that perturb the receiver-side photon-counting process, either by shifting the background component (continuous illumination/blinding), introducing short, synchronized flashes (preamble/control distortion), or altering the structure and timing consistency of service messages (spoofing/replay). These mechanisms operate at the physical and protocol-adjacent layers by deliberately shaping the observed counting features, which directly motivates real-time integrity monitoring based on intra-packet statistics over the "preamble + control" segment.

In parallel with the development of physical layer security mechanisms, machine-learning techniques have been increasingly adopted in VLC to improve robustness to interference and to support adaptive link operation. Prior work shows that ML/DL models can assist with channel-state estimation and tracking, suppression of structured optical noise during decoding, and adaptive selection of modulation/coding and operating parameters under changing illumination and geometry conditions [21–23]. However, when ML is integrated into the receiver processing chain, the learned decision function itself becomes part of the attack surface. As emphasized in the broader literature on embedded and IoT-class networked devices, ML models are vulnerable to targeted evasion and online data poisoning: carefully crafted perturbations of the input can induce misclassification, while contamination of adaptation data can shift decision boundaries and degrade reliability over time [24–26].

These observations motivate the use of interpretable, resource-efficient anomaly detection architectures that can operate in real time and remain resilient to adversarial impacts on the observed data stream, which is particularly important for multichannel SORA platforms with strict onboard constraints [27,28]. Consequently, at the intersection of three converging trends—multi-channel SORA deployment on small carriers, the adoption of service VLC links in "carrier–ground infrastructure" architectures, and the growing reliance on ML within communication and cyber–physical pipelines—the problem of ensuring the resilience of the service VLC channel to intentional interference becomes practically relevant and insufficiently addressed.

At the same time, much of the existing security literature for OWC/VLC and the robustness literature for ML-based detectors predominantly targets stationary indoor VLC/LiFi installations or RF-centric protocols, and therefore provides limited guidance for

service VLC channels embedded into onboard avionics-grade sensor control loops on small, unmanned carriers. The key challenge is that the onboard AI subsystem that monitors the service VLC link becomes the primary target for evasion and poisoning attacks, while the protection mechanism must remain interpretable and executable in near–real- time on a CPU without graphical acceleration.

Accordingly, this study develops and evaluates an interpretable protection architecture for the AI subsystem that monitors the service VLC channel in a compact multichannel SORA system under constrained onboard resources. The work formalizes a threat model for the considered VLC link (including evasion, spoofing, replay, and online poisoning), proposes a lightweight detector ensemble driven by intra-packet features extracted from a Poisson photon-counting observation model, and quantitatively evaluates performance in terms of detection probability, false-alarm probability, and end-to-end response latency to assess suitability for near–real-time operation.

2. Service VLC Channel Data Processing and Anomaly Detection in the AI Subsystem of a Multichannel SORA System

The objective of the present study is to develop and evaluate service VLC channel data processing and AI-based anomaly detection for a multichannel SORA system installed on a small carrier and operating as a radar–computing complex. For external data exchange with ground infrastructure, the multichannel SORA system employs two logically independent communication channels: a radio-frequency channel and a service VLC channel based on light-emitting diodes. The specific type of carrier hosting the multichannel SORA system is considered in this formulation only through constraints on mass, dimensions, and power consumption, as well as through geometric constraints on the line-of-sight region.

Figure 1 presents the structural diagram of the multichannel SORA system and the placement of the artificial intelligence (AI) module within the onboard processing loop. Onboard the carrier, the SORA radar chain exchanges radar data and sounding commands with the SORA onboard computer. The onboard computer interfaces the RF modem (radio link) for service RF data exchange with the ground infrastructure. It also interfaces the VLC optical modem/Digital interfaces that implement the service VLC chain.

Throughout the manuscript, the term "AI subsystem" refers to the onboard *ML-based* decision layer for VLC integrity monitoring (logistic regression + one-class detection + fusion), whereas the Poisson photon-counting model is used only as a physics-consistent observation model for feature formation. The contribution of the present work is limited to developing and investigating the feature-formation mapping $\Phi(N, S)$ and the ML decision layer implemented as an interpretable detector ensemble $f(x; \theta)$ executable on a central processing unit without graphical acceleration, for protecting the AI subsystem under adversarial impacts on the service VLC channel.

In the proposed pipeline, the Poisson photon-counting model is used strictly as a physics-consistent observation model that stabilizes the formation of intra-packet features $\Phi(N, S)$. The ML component is the trainable decision layer $f(x; \theta)$ (logistic regression + one-class modeling with calibrated thresholds) and its ensemble fusion rule executed on a CPU. Hence, the protection problem addressed in this paper is the robustness of the ML decision layer against adversarial evasion and online poisoning, rather than proposing a new photon-counting model.

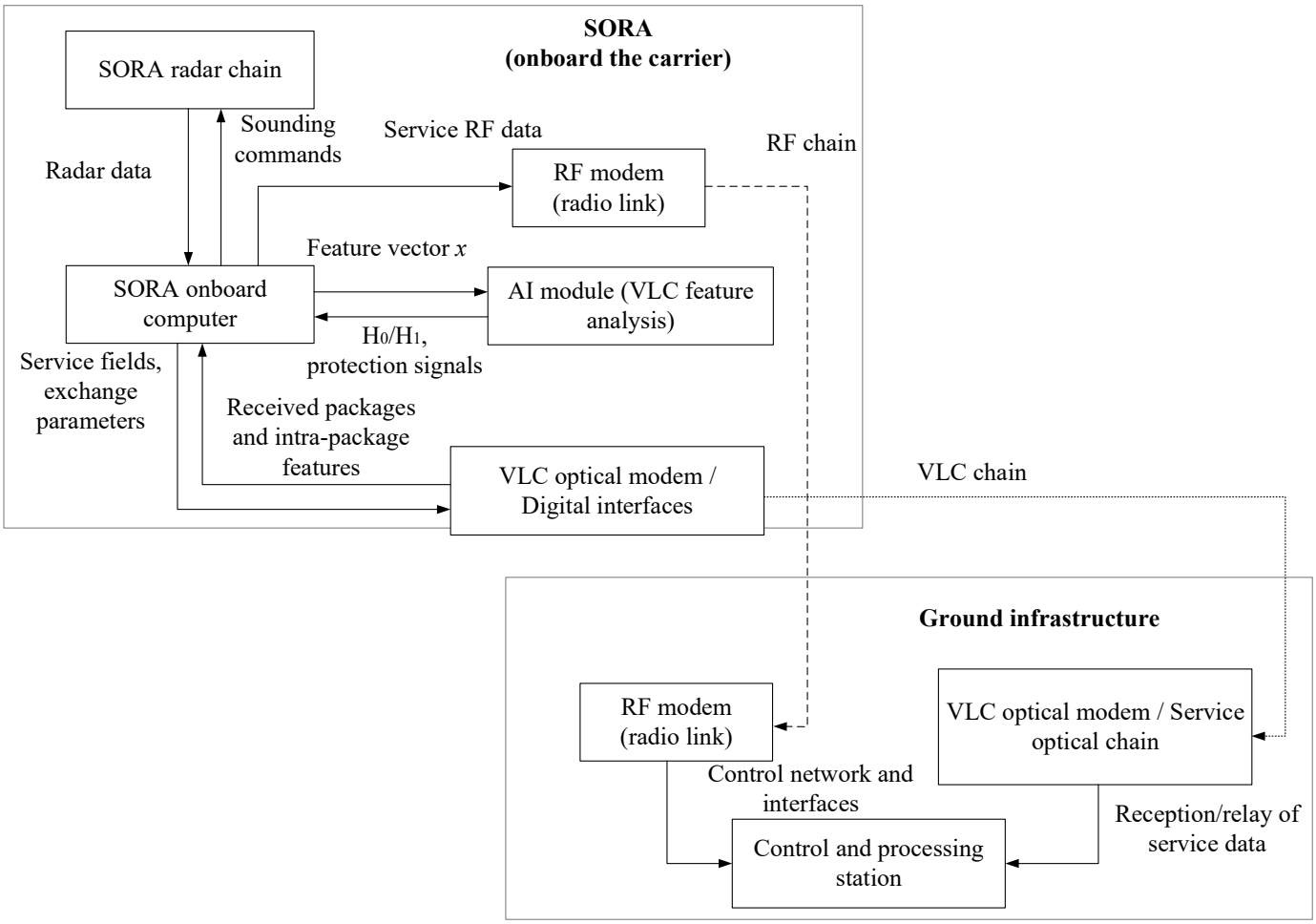

Figure 1. Structural diagram of a multichannel SORA system with radio-frequency and optical communication channels and an AI module.

The radio-frequency channel is used to transmit telemetry data on the state of the carrier and onboard systems, control commands, and, when required, aggregated radar products. The service visible light communication channel is employed as a low-emission dedicated link for transmitting operating-mode parameters of the SORA system, timing and synchronization markers, diagnostic messages, and a subset of service and measurement data.

The onboard computing module provides the software implementation of the AI subsystem, which is used to monitor the state of the service VLC channel. The AI software subsystem compares feature vectors corresponding to nominal and anomalous channel operation and, in a near–real-time mode, generates decisions regarding the channel state, issuing control signals to adjust system parameters based on the results of the feature-based state assessment.

Let the data transmission over the service VLC channel be represented as a sequence of service bit packets with a fixed structure. At the physical layer, each packet is realized as a sequence of optical radiation pulses of fixed duration, generated according to a pulse modulation scheme in accordance with the binary service data sequence of the transmitted packet. At the logical layer, each packet comprises three consecutively arranged segments of the optical signal: a preamble segment, a payload (service data) segment, and a control sequence of check bits.

The preamble is defined by a predefined binary bit sequence of the service packet, which, according to the employed modulation method, is mapped onto a corresponding sequence of optical radiation pulses. The preamble is used for acquisition and tracking of

time synchronization, as well as for estimating the current channel state based on deviations of the registered preamble realization from its known binary template. The control sequence of check bits is specified by a predefined binary sequence associated with the service packet and is used to verify the integrity of the payload segment and to assess the channel state based on deviations of the registered realization of the check sequence from its known binary template.

For a mathematical description of photon-counting observations at the photodetector output during the reception of service messages over the VLC link, we adopted a standard counting model for optical radiation detection. The combined "preamble + control bit sequence" segment of a single service message is partitioned into L identical time intervals of duration ΔT (s). In the k-th interval ($k = 1, 2, \ldots, L$), the number of registered events is described by (1):

$$N_k \in N_0, \tag{1}$$

where N_k (counts) is a random variable representing the number of photons registered within the k-th time interval of duration ΔT;

N_0 denotes the set of non-negative integers (dimensionless).

The vector $N = (N_1, N_2, \ldots, N_L)$ describes a single realization of a discrete-time counting process of photon detection of optical radiation over the combined segment "preamble + control bit sequence" of one service packet.

Here, N is an L-dimensional vector (dimensionless) and each component N_k is measured in counts. In the nominal reception mode of service bit packets over the VLC channel, a discrete-time Poisson model of photon registration is adopted, as given in (2):

$$N_k \sim Pois(\lambda_k), k = 1, 2, \ldots, L; \tag{2}$$

where Pois(λ_k) denotes a Poisson distribution on N_0 with parameter λ_k;

λ_k (counts per time interval ΔT) is the expected number of registered photons in the k-th time interval of duration ΔT.

Conditional independence of the values N_k for different intervals is assumed at fixed values of the parameters λ_k.

Here, the discrete-time Poisson model is used at the observation level to describe photon registration in the service VLC channel: each random variable N_k is the number of registered photons (counts) within the k-th time interval of duration ΔT, and the parameter λ_k (counts per time interval ΔT) is the expected number of registered photons in that interval. The intensity parameter λ_k is represented in the manuscript as the sum of the expected contribution from the LED transmitter during the transmission of a service bit packet and the expected background component; the background component explicitly includes background illumination and the components of the internal noise current of the receiver front-end. Consequently, more complex interference and noise patterns in practical scenarios are reflected in the realized intensity profile $\{\lambda_k\}$ of the service segment, i.e., in the expected useful and background photon counts across the L intervals, which directly determines the nominal distribution of the Poisson-counting feature vector under the hypothesis H_0 used for threshold setting (2).

The explicit probability mass function under the nominal hypothesis H_0 is provided in Appendix A.1.

The parameter λ_k is represented as the sum of the contributions of the useful optical signal radiation and the background component, including natural interference sources, as given in (3):

$$\lambda_k = \lambda_k^{(s)} + \lambda_k^{(b)}, \tag{3}$$

where $\lambda_k^{(s)}$ (counts per interval) is the expected number of photons registered in the k-th time interval of duration ΔT, caused by the useful optical signal emitted by the LED transmitter during the transmission of a service bit packet;

$\lambda_k^{(b)}$ (counts per interval) is the expected number of photons registered in the same k-th time interval of duration ΔT, caused by background illumination and internal noise sources of the photodetector receiving chain, including photodetector dark current noise, thermal noise of resistive elements, noise of amplification stages, and other components of the internal noise current of the receiver front-end.

The optional physical relationship between expected photon counts and received optical power is provided in Appendix A.1.

The ranges of variation in the expected number of registered photons due to the useful optical signal of the LED transmitter, and the expected number of registered photons caused by the ambient background illumination and internal noise of the optoelectronic receiver chain of the photodetector in the numerical simulation setup are specified as a regular finite grid of values over and within the operating ranges of the average optical power of the transmitting device and the background illumination levels of the "SORA carrier–ground infrastructure" link.

A formalized description of the characteristics of possible attacks (attack profile) on the considered system is defined as a set of parametric scenarios that modify the intensity of the useful and background optical radiation and, consequently, the statistical properties of the photon-counting registration process. These modifications are reflected in the parameters and of the counting-signal model and, ultimately, in the distribution of the observation vector N.

As studies on the security of VLC and OWC systems indicate, typical active threats at the physical layer and the optical channel protocol layer can be grouped into several persistent categories: illumination- and interference-based attacks (jamming, blinding), injection and modification of optical information (injection, spoofing), and replay attacks. Passive interception (eavesdropping) is typically treated separately as a threat to channel confidentiality [29–31]. For communication tasks in transport platforms and unmanned systems, it has been shown that jamming, spoofing, and replay attacks contribute most substantially to the risk of communication failure and loss of control stability [32,33]. Most existing active protection schemes for VLC channels primarily target these attack classes, highlighting their fundamental and critical importance [34].

To explicitly account for the random occurrence of the most critical active attacks, we introduce three Bernoulli event variables per received service packet (equivalently, per observation window). For Monte Carlo realization m, let $Z_{DC} \in \{0, 1\}$, $Z_{sp} \in \{0, 1\}$, and $Z_{rep} \in \{0, 1\}$ denote the presence of DC-jamming/blinding, spoofing, and replay, respectively. Their occurrence can be characterized by prior probabilities $\Pr[Z_{DC} = 1] = \pi_{DC}$, $\Pr[Z_{sp} = 1] = \pi_{sp}$, $\Pr[Z_{rep} = 1] = \pi_{rep}$, which depend on the operational environment and the adversary's capability. Under the nominal hypothesis H_0, no active attack is present, i.e., $(Z_{DC}, Z_{sp}, Z_{rep}) = (0, 0, 0)$. The alternative hypothesis is composite and corresponds to the presence of at least one active attack event, H_1:$(Z_{DC}, Z_{sp}, Z_{rep}) \neq (0, 0, 0)$. These event variables are used at the modeling level to select the realization-specific expected-value profile $\lambda^{(m)}(k)$ under H_1, i.e., to activate the corresponding parametric deviation of $\{\lambda k\}$ from the nominal reference $\lambda^{(0)}(k)$. This preserves the binary detection setup (H_0 versus composite H_1) while keeping the occurrence of the dominant attack mechanisms explicit in the generative description of the photon-count observations.

In the considered architecture, the VLC channel is used for unidirectional transmission of service and telemetry messages from the multichannel SORA system to the ground infrastructure; passive interception in this context does not affect the correct operation of

the SORA system itself, whereas impacts that compromise the availability and integrity of service messages are critical. Therefore, the threat profile includes four parametric classes of impacts: continuous illumination and synchronization-aligned light flashes (an analog of jamming/blinding), as well as spoofing and replay of service messages (spoofing/replay). This attack scenario is minimally sufficient for analyzing the sensitivity of the AI subsystem to energy-based and protocol-level attacks while maintaining a controlled dimensionality of the parameter space.

Based on the above classification and attack scenario, this study considers four basic types of active impacts on the service VLC channel.

The first type comprises attacks in the form of sustained (continuous) illumination of the optical receiver. In this attack scenario, there exist time intervals in which a non-zero background brightness is present, i.e., a persistent background impact produced by a directed light flux from an external radiation source. Let such attacks be modeled by adding a constant component to the intensity parameter of the background term, as given in (4):

$$\lambda_k^{(b)} \rightarrow \lambda_k^{(b)} + \Delta\lambda^{(DC)}, k \in K_{DC}; \quad (4)$$

where $\Delta\lambda^{(DC)}$ (counts per interval) is the increment of the expected number of registered photons of optical radiation caused by sustained directed illumination;

K_{DC} is the subset of indices of discrete time intervals $k \in \{1, 2, \ldots, L\}$, in which a persistent external optical impact is present, resulting in a shift in the expected number of registered photons.

In the case of attacks in the form of synchronized flashes reproduced in time intervals coinciding with elements of the preamble or the control sequence of the signal packet, short-duration optical pulses are generated by an external light source. These pulses lead to an increase in the number of photons registered at the receiver side, which can be described as in (5):

$$\lambda_k^{(b)} \rightarrow \lambda_k^{(b)} + \Delta\,\lambda_k^{(pulse)}, k \in K_{pulse}; \quad (5)$$

where $\Delta\,\lambda_k^{(pulse)} \geq 0$ (counts per interval) is the increment of the expected number of registered photons in the k-th time interval caused by an external optical flash;

K_{pulse} is a subset of indices of time intervals (dimensionless set) corresponding to flashes synchronized with known segments of the service packet structure.

In the case of spoofing attacks on service messages, an optical signal is generated that imitates the structure of the original service message, while the legitimate signal is suppressed, typically either geometrically (by removing the LED transmitter from the line-of-sight region), energetically (by dominance in optical power), or through a combined attack scheme. In this case, for the time intervals onto which the spoofed message is projected, a substitution of the useful signal component is realized as in (6):

$$\lambda_k^{(s)} \rightarrow \lambda_k^{(s,sp)}, k \in K_{sp}, \quad (6)$$

where $\lambda_k^{(s)}$ (counts per interval) is the expected number of registered photons of the useful signal in the nominal operating mode;

K_{sp} (dimensionless set) is the subset of indices of time intervals $k \in \{1, 2, \ldots, L\}$ over which the spoofed service packet is present;

$\lambda_k^{(s,sp)}$ (counts per interval) is the expected number of registered photons due to the useful signal generated by the attacker during the transmission of the spoofed service packet.

Depending on the attack scenario, $\lambda_k^{(s,sp)}$ may implement:

(i) A binary structure of the preamble and control sequence that differs from the nominal one;

(ii) Scaling of the legitimate structure in terms of optical power;
(iii) A combination of segments whose statistics are close to the nominal ones and segments with deliberately introduced distortions.

In all cases, spoofing is realized as a transition from the parameter vector $\lambda = (\lambda_1, \ldots, \lambda_L)$ to the vector $\lambda^{(sp)} = (\lambda_1{}^{(sp)}, \ldots, \lambda_L{}^{(sp)})$ which leads to a systematic shift in the distribution of the counting-observation vector N defined in (1) and (2) with the intensity decomposition given in (3), and consequently in the feature vector formed on the basis of the "preamble + control bit sequence" segment. The corresponding parametric modifications of vector λ for the considered impact classes are specified in (4)–(6).

The final type of attack considered in this study is a replay of service messages. In this case, the attack consists of recording a legitimate service message and subsequently re-emitting its optical equivalent at a different point in time. Unlike spoofing attacks on service messages, this scenario relies on a previously generated legitimate message; however, its temporal relevance and contextual binding to the current state of the multichannel SORA system are violated.

Then, at the time instant when there is a service bit packet with index i^*, and s is expected from the multichannel SORA system, under such an attack, instead of the expected set of useful-signal intensity parameters, a set corresponding to substituted data is realized at the receiver side, and a service bit packet with index j appears as given in (7):

$$\lambda_k^{(s)} \rightarrow \lambda_k^{(s,j)}, k \in K_{rep}, \tag{7}$$

where K_{rep} (dimensionless set) is the subset of indices of time intervals corresponding to the "preamble + control bit sequence" segment of the replayed service packet;

$j \neq i^*$ is the index of the original service bit packet used to construct the replay attack;

$\lambda_k^{(s,j)}$ (counts per interval) is the corresponding set of intensity parameters for the packet with index j, whose optical signal was previously recorded by the attacker and is reproduced during the replay attack.

Under a replay attack on a service message, for time intervals $k \in K_{\text{rep}}$, the intensity of the counting process λ_k is determined as the sum of the intensity of the false useful signal and the background component $\lambda_k^{(b)}$, as given in (8):

$$\lambda_k^{(rep)} = \lambda_k^{(s,j)} + \lambda_k^{(b)}, \tag{8}$$

In the presence of dynamically varying service fields (such as packet counters, timestamps, or pseudorandom preamble templates), for the expected service packet with index i^*, the nominal operating mode should exhibit statistics $\lambda_k^{(s,i*)}$ consistent with the current state of the multichannel SORA system. Replaying a service packet characterized by the parameter set $\lambda_k^{(s,j)}$ leads to the emergence of structural inconsistencies between the observed counting realizations and the predicted statistics corresponding to the current service message, while maintaining the energy characteristics within the nominal range. As a result, replay attacks constitute a typical class of impacts aimed at violating the logical consistency of service communication while introducing minimal distortion to low-level energy-based features.

Problem Formulation for Attack Detection in the Artificial Intelligence Subsystem

The considered attack types and scenarios can be described as controlled modifications of the vector of expected numbers of registered photons over the "preamble + control bit sequence" segment of a single service bit packet. For each time interval $k = 1, 2, \ldots, L$ λ_k

denotes the expected number of photons registered in that interval and can be written as the sum of the useful-signal term $\lambda_k^{(s)}$ and the background term $\lambda_k^{(b)}$.

Let these values be combined into the vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_L)$, where λ is the vector of expected numbers of registered photons over all L time intervals of the "preamble + control bit sequence" segment of a single service bit packet; each component λ_k is measured in counts per time interval of duration ΔT.

Let us assume that the operating range of the service VLC channel in the nominal mode is defined by the set (9):

$$\Lambda_0 \subset R_+^L, \tag{9}$$

where Λ_0 denotes the set of all vectors λ of expected numbers of registered photons over the "preamble + control bit sequence" segment of a single service bit packet that arise for admissible values of the average optical power of the LED transmitter and the background illumination level of the "SORA carrier–ground infrastructure" link in the absence of attacks.

Similarly, the set (10) is introduced:

$$\Lambda_{att} \subset R_+^L, \tag{10}$$

where Λ_{att} denotes the set of all vectors $\lambda_k^{(a)} = \left(\lambda_1^{(a)}, \lambda_2^{(a)} \ldots, \lambda_L^{(a)}\right)$ of expected numbers of registered photons over the "preamble + control bit sequence" segment of a single service bit packet in the presence of adversarial impacts, realized through the increments $\Delta\lambda^{(\text{DC})}$, $\Delta\ \lambda_k^{(pulse)}$ and through the index sets of time intervals K_{DC}, K_{pulse}, K_{sp}, K_{rep}, as well as through admissible structures of spoofed and replayed service bit packets.

For $\lambda \in \Lambda_0$, the realization of the counting-observation vector $N = (N_1, N_2, \ldots, N_L)$, previously introduced as a single experimental outcome of the discrete-time photon-counting process over the "preamble + control bit sequence" segment of a single service bit packet, belongs to the set (11):

$$N_0 = \{N : N_k \sim Pois(\lambda_k), k = 1, 2, \ldots, L, \lambda \in \Lambda_0\}. \tag{11}$$

For $\lambda^{(\text{a})} \in \Lambda_{\text{att}}$ the realization of the counting-observation vector under an attack belongs to the set (12):

$$N_{att} = \left\{N^{(a)} : N_k^{(a)} \sim Pois\left(\lambda_k^{(a)}\right), k = 1, 2, \ldots, L, \ \lambda^{(a)} \in \Lambda_{att}\right\}, \tag{12}$$

where N_{att} denotes the set of all realizations of the counting-observation vector in the presence of one of the considered attacks on the service VLC channel;

$N^{(a)} = \left(N_1^{(a)}, \ldots, N_L^{(a)}\right)$ is the counting-observation vector under attack.

The optical modem of the VLC channel and the onboard computer of the multichannel SORA system form a channel-state feature vector (whose deviations indicate anomalies) based on the vector of photon-counting observations registered in each time interval over the "preamble + control bit sequence" segment and on the known structure of the service bit packet, as given in (13):

$$x = \Phi(N, S), \tag{13}$$

where $x \in R^d$ is the channel-state feature vector (with anomalies corresponding to non-nominal deviations) in the AI subsystem of dimension d (each component of the vector x is a deterministic function of the component N_k and the set of structural parameters S of the service bit packet);

$\Phi(\cdot,\cdot)$ is the feature-formation mapping from the counting-observation vector over the "preamble + control bit sequence" segment and the structure of the service bit packet;

S is the set of structural parameters of the service bit packet (binary preamble template, binary template of the control bit sequence, indices of time intervals corresponding to these templates, and the parameters defining the partitioning of the segment into L intervals of duration ΔT);

d is the dimension of the feature space describing the nominal channel state of the service VLC channel; anomalies are represented as non-nominal feature vectors in the same space.

The mapping $\Phi(N, S)$ defines two classes of feature vectors corresponding to the nominal channel state and to anomalous (attack-induced) conditions in terms of interaction with the AI subsystem, as given in (14):

$$X_0 = \{x : x = \Phi(N, S), N \in N_0\}, \tag{14}$$

where X_0 denotes the set of feature vectors of the AI subsystem corresponding to the nominal operating mode of the service VLC channel, as given in (15):

$$X_{att} = \left\{x^{(a)} : x^{(a)} = \Phi(N, S), N^{(a)} \in N_{att}\right\}, \tag{15}$$

where X_{att} denotes the set of feature vectors of the AI subsystem corresponding to the presence of attacks on the service VLC channel.

The AI subsystem implements a binary decision rule, as given in (16):

$$f(x;\ \theta) \in \{0,\ 1\}, \tag{16}$$

where $f(x; \theta) = 0$ denotes the decision that the service VLC channel operates in the nominal mode (hypothesis H_0);

$f(x; \theta) = 1$ denotes the decision that the service VLC channel operates in an anomalous mode (hypothesis H_1), corresponding to the presence of at least one of the considered attack types (without attributing the decision to a specific attack class);

θ is the parameter vector of the detector ensemble of the AI subsystem (logistic-model coefficients, one-class detector parameters, and threshold values) implemented on the central processing unit of the onboard computer of the multichannel SORA system.

To avoid ambiguity between the decision output and the reference class information, two binary quantities are distinguished. The binary output of the AI subsystem is a decision produced from the currently observed feature vector and represents an estimate of the operating condition ("nominal" versus "attack-present"). In contrast, the variable y introduced in the dataset description is a reference class label assigned to simulated (labeled) realizations for supervised training and for performance evaluation; it indicates whether a given realization was generated under the nominal operating mode or under one of the considered attack scenarios. Consequently, y specifies which hypothesis holds for a labeled realization, whereas the AI subsystem output provides the corresponding estimated decision; false-alarm and missed-detection probabilities are evaluated by comparing these decisions to the reference labels over the nominal and attack-induced feature distributions.

The distribution of channel-state features in the nominal operating mode is denoted by P_0, while the distribution of channel-state features in the attacked (anomalous) mode is denoted by P_{att} as given in (17):

$$x \sim P_0 \Leftrightarrow x \in X_0, \tag{17}$$

where P_0 is the distribution of the AI subsystem feature vector over the set X_0, induced by the Poisson model of the photon-counting registration process for $\lambda \in \Lambda_0$ and a fixed structure S of the service bit packet; similarly, P_{att} is defined as in (18):

$$x^{(a)} \sim P_{att} \Leftrightarrow x^{(a)} \in X_{att}, \tag{18}$$

where P_{att} (dimensionless) is the aggregated distribution of the AI subsystem state (anomaly) feature vector over the set X_{att} induced by the Poisson model of the photon-counting registration process for $\lambda^{(a)} \in \Lambda_{\text{att}}$ across all considered attack scenarios.

For a fixed parameter vector θ, the false-alarm probability and the missed-detection probability of the AI subsystem are defined as in (19) and (20):

$$P_{FA}(\theta) = P_x \sim P_0(f(x;\theta) = 1), \tag{19}$$

where $P_{\text{FA}}(\theta)$ (dimensionless) is the probability that, under the nominal operating mode of the service VLC channel (feature distribution P_0), the AI subsystem erroneously produces a decision indicating an anomalous mode;

$$P_{MD}(\theta) = P_x^{(a)} \sim P_{att}\left(f\left(x^{(a)};\theta\right) = 0\right), \tag{20}$$

where $P_{\text{MD}}(\theta)$ is the probability that, in the presence of attacks on the service VLC channel (feature distribution P_{att}), the AI subsystem erroneously produces a decision indicating a nominal operating mode (dimensionless).

From the perspective of attacks on artificial intelligence systems, the scenarios of sustained illumination and synchronized flashes implement evasion attacks against the AI subsystem: the attacker selects a vector $\lambda^{(a)} \in \Lambda_{\text{att}}$ such that the corresponding features $x^{(a)} \in X_{\text{att}}$ lead, with non-zero probability, to the decision $f(x^{(a)}; \theta) = 0$ despite the actual presence of attacks.

The spoofing and replay scenarios of service bit packets, in addition to evasion attacks, also create the possibility of online data poisoning: when service traffic is used to adapt the parameter vector θ false (anomalous) channel-state feature vectors $x^{(a)} \in X_{\text{att}}$, may be incorporated into the training dataset. This results in a systematic bias in the estimates of nominal-mode statistics and a shift in the class decision boundary in the feature space.

The working hypothesis of this study is that, for the service VLC channel of a multichannel SORA system, it is possible to construct a feature-formation mapping $\Phi(N, S)$ and a detector ensemble $f(x; \theta)$, implementable on a central processing unit without graphical acceleration, such that, under physical constraints on the attack parameters $\lambda^{(a)} \in \Lambda_{\text{att}}$ there exists a parameter vector θ^* of the anomaly detector ensemble for the service VLC channel satisfying the inequalities in (21):

$$P_{FA}(\theta^*) \leq \overline{P}_{FA}, P_{MD}(\theta^*) \leq \overline{P}_{MD}, \tag{21}$$

where $\overline{P}_{FA}$ (dimensionless) is the prescribed admissible level of the false-alarm probability of the service VLC channel of the multichannel SORA system;

$\overline{P}_{MD}$ (dimensionless) is the prescribed admissible level of the missed-detection probability for attacks on the service VLC channel of the multichannel SORA system;

θ^* is the parameter vector of the detector ensemble of the artificial intelligence subsystem, optimal with respect to the selected attack detection performance criterion under the given constraints on $\overline{P}_{FA}$ and $\overline{P}_{MD}$.

The data-processing scheme in the service VLC channel and the operation of the AI subsystem are illustrated in Figure 2.

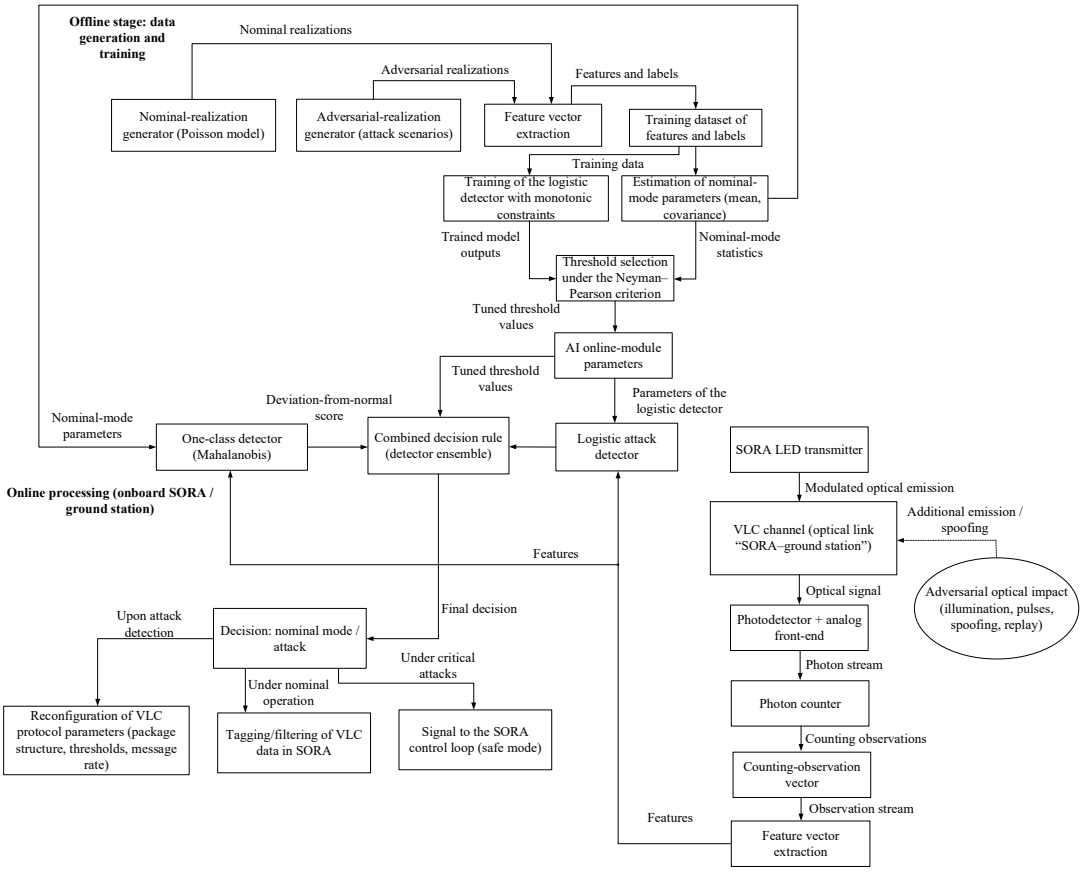

Figure 2. Data-processing scheme in the service VLC channel and the operation of the AI subsystem (ML-based anomaly detection).

The lower part of the diagram (Figure 2) shows the sequence of data processing in the service VLC channel during the carrier flight (online mode). The light-emitting diode transmitter generates an optical signal that propagates in an open environment, where it may be affected by natural background illumination and atmospheric effects, as well as by intentional adversarial impacts. After the optical signal is registered by the photodetector, time discretization is performed and photon-counting observations are formed within a fixed observation window divided into L intervals of duration ΔT.

The upper part of Figure 2 describes the formation of synthetic and semi-synthetic datasets of counting observations (offline stage), which are used to construct training datasets of channel-state features, to estimate nominal-mode statistics, and to train the parameters of channel anomaly detectors and the AI subsystem. At this stage, the threat profile (attack scenarios) is also formalized, incorporating the attack types described above.

3. Simulation Setup and Attack Emulation

The simulation model of the service VLC channel is constructed as a physico-mathematical interpretation of the photon-registration process at the receiver, based on specified parameters of the light-emitting diode transmitter, the propagation path, and the photodetector, in order to generate synthetic realizations of the counting-observation vector $N = (N_1, \ldots, N_L)$, which are used by the AI subsystem.

The relationship between the optical power incident on the photodetector aperture and the parameter of the counting process is defined by a quantum-optical relation, as given in (22):

$$\lambda_k = \frac{\eta}{h\nu} \int_{(k-1)\Delta T}^{k\Delta T} P_k^{(opt)}(t)dt + \lambda_k^{(b)}, \tag{22}$$

For reproducibility of the physical layer setup, we specify the effective received optical power levels at the photodetector aperture that correspond to the nominal mean photon-count levels used in the Monte Carlo study. The photon-counting (Poisson) observation model and the mapping between received optical power and mean photon counts are widely used in low-power optical wireless communication analyses [35].

For physical interpretability of the receiver-side parameters, we align the representative values used here with a standard silicon PIN photodiode datasheet (Hamamatsu Photonics, *"Si photodiode S9219 series (S9219-01)"*): the datasheet reports a peak sensitivity wavelength of 550 nm, photosensitivity of approximately 0.22 A/W at the peak, and microsecond-scale temporal response (typical rise time ≈0.5 μs under the stated test conditions). Accordingly, we use a representative visible wavelength λ_{opt} = 550 nm ($\nu = c/\lambda_{opt}$), adopt an integration time ΔT = 1 μs, and set the quantum efficiency to a conservative reproducible value η = 0.5 for the Monte Carlo setup. For these settings, the nominal mean counts $\{\lambda_{p,nom}, \lambda_{c,nom}, \lambda_{b,nom}\} = \{14, 7, 2\}$ per interval correspond to effective received optical powers $\{1.0 \times 10^{-11}, 5.1 \times 10^{-12}, 1.4 \times 10^{-12}\}$ W for the preamble "1"-intervals, the CRC/control "1"-intervals, and the background component, respectively. In this work, transmitter power, link geometry, receiver aperture, and pointing angles are not fixed to a single hardware configuration; instead, their combined impact is represented by the Poisson intensity profile λ_k via an effective channel transfer coefficient and an additive background term, consistent with the line-of-sight Lambertian assumptions stated below.

The line-of-sight optical channel transfer coefficient is modeled using the standard Lambertian DC-gain expression; the explicit formulation is provided in Appendix A.1 for completeness.

Here, it is assumed that within a single interval ΔT, the optical power $P_k^{(opt)}(t)$ is approximately constant.

To focus on the tasks of protecting the AI subsystem and on analyzing the impact of attacks on counting features, a number of assumptions are introduced in the simulation model:

- A line-of-sight (LOS) channel is only considered. Multiple reflections, diffuse scattering, and multipath effects are not explicitly modeled, and their overall contribution is assumed to be absorbed into variations in the effective channel transfer coefficient and the background optical power.
- Within the counting-observation time window, which includes a fixed number of equal-duration intervals, the propagation geometry and background conditions are assumed to be quasi-stationary; that is, for each realization of the simulation experiment, the distance between the transmitter and the receiver, the orientation of the optical axes of the emitter and the photodetector, the average background optical power at the photodetector aperture, and the effective level of internal noise of the receiver front-end are fixed and do not change within a single window.
- The photodetector front-end operates in the linear regime. Photodetector saturation and nonlinear effects of the amplification stages are not considered, and it is assumed that the optical power levels in the considered operating modes do not drive the system outside the linear range.

- Intersymbol interference is not modeled. The time intervals ΔT are chosen such that each interval corresponds to a single symbol position in the preamble or the control sequence, and the receiver response does not overlap across adjacent intervals.
- Timing synchronization errors and mismatches in selecting the counting-observation time window are not considered. It is assumed that the "preamble + control bit sequence" segment is correctly identified at the receiver and that synchronization is maintained by lower-layer protocol mechanisms.

The simulation model implements the data-processing scheme shown in Figure 2. The modeling process is divided into two loops:

- A synthetic data generation loop (offline stage), which includes forming reference intensity templates, modeling the effects of the channel and attacks, and generating counting observations and features for training and testing the detectors;
- An anomaly detection loop (online stage/online emulation), which includes applying the trained detector ensemble to the generated feature stream to estimate the false-alarm probability P_{FA} and the detection probability P_D.

The transmit chain is simulated under the assumption that the transmitter converts the logical structure of the service packet S into a sequence of optical pulses. A binary one (bit '1') is mapped to a rectangular optical pulse of duration ΔT with average optical power P_1. A bit '0' corresponds to the absence of radiation (power $P_0 \approx 0$). For the preamble, the transmitter optical power level P_p is used, and for the control sequence P_c, is used, with $P_p > P_c$. The ratio $P_p/P_c = 2$ is chosen to prioritize synchronization.

The optical channel model is simulated based on the assumption that the signal propagates through the "SORA carrier–ground infrastructure" link, which is characterized by:

- Geometric losses, calculated using a Lambertian radiation model for a given transmission distance, photodetector aperture area, and orientation angles;
- Atmospheric attenuation, which depends on the signal propagation range (distance between the transmitter and the receiver) and meteorological conditions;
- Background illumination, which creates a constant background irradiance at the photodetector.

For each realization, the true channel-state label $y \in \{0, 1\}$, is additionally fixed, where $y = 0$ corresponds to the nominal operating mode (hypothesis H_0), and $y = 1$ corresponds to the presence of an attack (hypothesis H_1). This enables the formation of simulation datasets $\{(x^{(m)}, y^{(m)})\}$ for training and testing the detector ensemble $f(x; \theta)$.

The proposed protection architecture is scalable in the sense that the online computation in the AI subsystem is $O(L)$ per received service packet: feature formation $\Phi(N, S)$ consists of linear-time aggregation over the index sets K_p, K_c, and K_0, and the one-class, energy, and logistic detectors operate on a fixed feature dimension $d = 5$.

Adaptability to different operational environments (including varying link geometry, background illumination, and meteorological conditions that affect atmospheric attenuation) is achieved by treating their combined impact as a realized Poisson-intensity profile λ_k (effective channel transfer coefficient plus an additive background term) and by calibrating nominal-mode feature statistics and detector thresholds on environment-specific nominal data. Under this interpretation, changing atmospheric conditions primarily manifest as increased inter-window variability of the useful and background components and, consequently, as heavier tails of the nominal feature distributions that control P_{FA} through threshold tuning. The applicability boundary under extreme environmental variability is defined by the modeling assumptions stated above: within-window quasi-stationarity, LOS dominance with multipath absorbed into effective parameters, and linear operation of the photodetector front-end without saturation.

Ensemble of Channel-State Feature Detectors for the Artificial Intelligence Subsystem

In the AI subsystem of the service VLC channel, an ensemble of lightweight anomaly detectors is used, implemented on the central processing unit of the multichannel SORA system without graphical acceleration. The ensemble operates in the space of channel-state features (with anomalies corresponding to non-nominal deviations in the same feature space) $x \in R^d$, formed from the counting-observation vector N and the set of structural parameters S of the service bit packet.

A total of $d = 5$ state features is considered for the protected channel and the AI subsystem.

The adopted feature dimension is fixed at five because the channel-state feature vector is defined as a five-component aggregation of counting observations over the service-packet structure. Specifically, the components quantify relative deviations over the preamble and control sequence, the contrast with respect to the background segment, an integral deviation measure over the service segment, and the variability of deviations over background intervals where the useful signal is absent. This compact representation preserves the observability of the main non-nominal effects considered in the manuscript while keeping the online detector computations lightweight.

To describe the service packet structure, three index sets are introduced:

$K_p \subset \{1, \ldots, L\}$ is the set of indices of time intervals corresponding to preamble elements;

$K_c \subset \{1, \ldots, L\}$ is the set of indices of time intervals corresponding to elements of the control sequence;

$K_0 \subset \{1, \ldots, L\}$ is the set of indices of time intervals corresponding to the background segment (absence of a useful signal).

The cardinalities $|K_p|$, $|K_c|$, $|K_0|$ (dimensionless quantities) are determined by the structure of the service bit packet and the parameter L.

If $\lambda^{(0)} = (\lambda_1^{(0)}, \ldots, \lambda_L^{(0)})$ is the reference vector of expected photon counts over the preamble-and control-segment under the nominal operating mode, then, for a single realization $N = (N_1, \ldots, N_L)$, consistent with the line-of-sight Lambertian assumptions stated below, channel-state feature vector $x = (x_1, \ldots, x_5)$ is defined as follows.

First, deviations of the observed photon counts in the preamble and in the control sequence from their reference values are quantified; these deviations are sensitive to changes in the useful-signal level and propagation conditions. Second, (i) the contrast relative to the ambient background illumination, (ii) the sum of squared deviations from the reference profile over the service segment, and (iii) the variance of deviations over background intervals where the useful signal is absent and the observations are dominated by ambient illumination and receiver front-end noise are computed.

Then, the mean relative deviation over the preamble is defined as in (23):

$$x_1 = \frac{1}{|K_p|} \sum_{k \in K_p} \frac{N_k - \lambda_k^{(0)}}{\lambda_k^{(0)}}, \tag{23}$$

where x_1 (dimensionless) is the mean relative deviation of the number of registered photons from the reference expected value, averaged over the preamble elements; mean relative deviation over the control sequence is defined as in (24):

$$x_2 = \frac{1}{|K_c|} \sum_{k \in K_c} \frac{N_k - \lambda_k^{(0)}}{\lambda_k^{(0)}}, \tag{24}$$

where x_2 (dimensionless) is the mean relative deviation averaged over the intervals of the control sequence; the remaining notations coincide with those in the formula for x_1.

The preamble contrast with respect to the background illumination is defined as the difference between the mean photon counts over the preamble and background segments, as given in (25):

$$x_3 = \frac{1}{|K_p|}\sum_{k\in K_p} N_k - \frac{1}{|K_0|}\sum_{k\in K_0} N_k, \tag{25}$$

where x_3 (counts) is the difference between the mean values of the number of registered photons in the preamble and in the background segment; the first sum yields the mean photon count over the preamble intervals, and the second sum yields the mean photon count over the background intervals.

The sum of squared deviations over the service segment (the "preamble + control bit sequence" segment) is defined as in (26):

$$x_4 = \sum_{k\in K_p\cup K_c} (N_k - \lambda_k^{(0)})^2, \tag{26}$$

where x_4 (counts^2) is the unnormalized sum of squared deviations of the observed values N_k from the reference values $\lambda_k^{(0)}$ over the "preamble + control bit sequence" segment, i.e., an integral measure of the overall deviation level across the entire service part of the packet.

The variance of the background noise component is defined as in (27):

$$x_5 = \frac{1}{|K_0|-1}\sum_{k\in K_0}\left[\left(N_k - \lambda_k^{(0)}\right) - \overline{\delta}_0\right]^2, \tag{27}$$

where $\overline{\delta}_0$ is given by (28):

$$\overline{\delta}_0 = \frac{1}{|K_0|}\sum_{k\in K_0}\left(N_k - \lambda_k^{(0)}\right), \tag{28}$$

where x_5 (counts^2) is the variance of deviations of the number of registered photons from the reference expected value $\lambda_k^{(0)}$ over those time intervals $k \in K_0$, in which the useful signal of the service packet is absent; this quantity characterizes the noise level due to background illumination and internal noise in the receiver front-end; (counts) is the sample mean deviation over the background segment.

$\overline{\delta}_0$ (counts) is the sample mean deviation over the background segment.

The ensemble includes three lightweight detectors implemented on the central processing unit and operating on the channel-state feature vector x defined in (13) (with components $x_1, \ldots, x_5$ given in (23)–(28)):

- A one-class (OC) detector models the feature distribution in the nominal operating mode. For each new feature vector, a deviation score with respect to the reference distribution is computed, and an anomaly decision is issued if the score exceeds a threshold.
- An energy detector computes the total energy of the feature vector (the sum of squares of its components) and compares it with a threshold. High energy values indicate large joint deviations across the features.
- A logistic-regression (LR) detector is trained on a labeled dataset containing both nominal-mode samples and attack examples. For each feature vector, an estimate of the attack probability is computed, and a positive decision is issued if it exceeds a threshold.

The explicit detector scoring functions and threshold calibration procedures for the OC detector, the energy detector, and the LR detector are provided in Appendix A.2.

Within the implementation of the one-class elliptical detector, the sample of channel-state features corresponding to the nominal operating mode $X_0 = \{x^{(m)}: y^{(m)} = 0\}$ is used to

obtain sample estimates of the mean feature vector in the nominal mode $\mu_0 \in R^5$ and the covariance matrix $\Sigma_0 \in R^{5\times5}$. These dimensions follow directly from the five-component definition of the channel-state feature vector.

The local decision rule of the one-class detector is given in (29):

$$f_{OC}(x;\theta_{OC}) = \begin{cases} 1, s_{OC}(x;\theta_{OC}) > \gamma_{OC}, \\ 0, s_{OC}(x;\theta_{OC}) \le \gamma_{OC}, \end{cases} \quad (29)$$

where the threshold γ_{OC} is determined using the simulation sample of nominal-mode features X_0.

For each feature vector $x^{(m)} \in X_0$, the corresponding scoring value is computed. Based on the set of obtained values, an empirical distribution consisting of $M_0 = |X_0|$ points is constructed. The threshold γ_{OC} is then selected as the level of this empirical distribution such that the fraction of nominal-mode realizations exceeding the threshold does not exceed the prescribed admissible false-alarm probability. In this way, γ_{OC} is uniquely determined from the sample X_0 and defines a controlled false-alarm level for the one-class detector.

The decision rule for the energy detector has the form (30):

$$f_{EN}(x;\theta_{EN}) = \begin{cases} 1, S_{EN}(x;\theta_{EN}) \ge \gamma_{OC}, \\ 0, S_{EN}(x;\theta_{EN}) < \gamma_{OC}, \end{cases} \quad (30)$$

where $f_{EN}(x;\theta_{EN}) \in \{0, 1\}$ is the binary decision of the energy detector; the value 1 is interpreted as the presence of an anomaly.

In what follows, the "energy" of the feature vector refers specifically to the quantity $S_{EN}(x;\theta_{EN}) = \sum_{i=1}^{5} x_i^2$, i.e., the sum of squares of its components, by analogy with the notion of the energy of a discrete-time signal in signal processing theory; this is not physical energy in the sense of joules, but a convenient scalar characteristic of the scale of feature deviations.

The local decision of the logistic detector is given by (31):

$$f_{LR}(x;\theta_{LR}) = \begin{cases} 1, \pi_{LR}(x;\theta_{LR}) \ge \tau_{LR}, \\ 0, \pi_{LR}(x;\theta_{LR}) < \tau_{LR}, \end{cases} \quad (31)$$

where $f_{LR}(x; \theta_{LR}) = 1$ corresponds to classifying the realization as an attack.

The parameters w and b are determined from the simulated labeled dataset $\{x^{(m)}, y^{(m)}\}$ as the solution of the problem of minimizing the mean logistic loss function: w and b, are chosen such that the values of $\pi_{LR}(x;\theta_{LR})$ are as close as possible to 0 for nominal-mode realizations ($y^{(m)} = 0$) and as close as possible to 1 for attack realizations ($y^{(m)} = 1$).

The complete detector ensemble in the AI subsystem implements the decision rule in the form of a logical expression, as given in (32):

$$f(x;\theta) = f_{OC}(x;\theta_{OC}), \vee f_{EN}(x;\theta_{EN}) \vee f_{LR}(x;\theta_{LR}), \quad (32)$$

where $f(x;\theta) \in \{0, 1\}$ is the ensemble decision;

$\theta = (\theta_{OC}, \theta_{EN}, \theta_{LR})$ is the joint parameter set of all three detectors, i.e., the parameter vector of the detector ensemble implemented on the central processing unit of the onboard computer of the multichannel SORA system.

This means that the ensemble declares the presence of an attack according to a "2-out-of-3" voting rule. The "2-out-of-3" rule operates on the three local binary decisions produced by the one-class detector, the energy detector, and the logistic detector for the same observation window and the same channel-state feature vector. An attack-present decision is issued only when at least two of the three detectors declare an anomaly/attack;

otherwise, the ensemble output corresponds to the nominal mode. This fusion rule reduces sensitivity to occasional excursions of a single detector while preserving the low-false-alarm operating point ensured by nominal-mode threshold calibration of the individual detectors.

Under extreme adversarial conditions, the performance of the ensemble (one-class detector, energy detector, and logistic-regression (LR) detector) combined by the "2-out-of-3" voting rule is limited by whether spoofing produces a detectable shift in the Poisson-counting feature vector $x = \Phi(N, S) = (x_1, \ldots, x_5)$ beyond the nominal H_0 variability at the fixed low-false-alarm thresholds (γ_{OC}, γ_{EN}, τ_{LR}). In the spoofing impact model considered in this manuscript, spoofing modifies the expected-value profile $\lambda(k)$ over the "preamble + CRC" intervals and therefore typically shifts the aggregated features x_1–x_5. The limiting extreme case corresponds to a high-capability spoofing process engineered to reproduce the nominal photon-count feature statistics of the service bit packet under H_0. This would require simultaneous matching of the "preamble + CRC" expected-value profile $\lambda(k)$ that governs the mean-deviation and energy-related features (x_1, x_2, x_4) and matching of the background level and background variability over the interval set K_0 that directly determines the contrast feature x_3 and the background-variance feature x_5.

Operationally, this implies fine synchronization with the receiver sampling grid ΔT, accurate knowledge of the instantaneous optical channel gain (Lambertian DC gain with atmospheric attenuation) and background illumination, and the ability to shape injected optical power across the service segment without inducing excess energy or excess variability in the background intervals. The quantitative sensitivity curves for spoofing and the most challenging close-to-H_0 case (replay) are reported in Section 4 (Figure 3g,h).

4. Results and Baseline Analysis

The objective of the numerical experiment is to obtain (i) the empirical false-alarm probability P_{FA} of the AI subsystem in the nominal operating mode of the service VLC channel (hypothesis H_0) and (ii) the sensitivity characteristics of the attack detection probability (hypothesis H_1) $P_D(\xi)$ as a function of the attack strength parameter ξ, applied to the communication channel or the AI subsystem for each impact type $a \in$ {DC-jamming, pulsed, spoofing, replay}. The detector thresholds are fixed using the nominal-mode sample and are not retuned when the attack strength ξ is varied; this ensures a physically consistent comparison of the sensitivity of the detectors and the ensemble based on the plots in Figure 3e–h.

The results in this section are reported both for the individual detectors and for their ensemble. For the individual-detector curves, the corresponding local binary decision is used. For the ensemble curves, the final decision is formed by the "2-out-of-3" voting rule applied to the three local decisions within the same observation window: an attack-present decision is issued only when at least two local decisions indicate an anomaly/attack; otherwise, a nominal decision is retained.

The numerical experiment is implemented using the Monte Carlo method within a discrete-time Poisson model of photon-counting over the "preamble + control bit sequence" segment of a single service packet. For each realization, a vector of expected numbers of registered photons over the observation-window intervals $\lambda^{(m)} = \left(\lambda_1^{(m)}, \ldots, \lambda_L^{(m)}\right)$ (counts per interval ΔT) is formed, after which a vector of photon-counting observations $N^{(m)} = \left(N_1^{(m)}, \ldots, N_L^{(m)}\right)$ is generated according to the procedure described in Section 2.

A reference (nominal) profile of the expected numbers of registered photons $\lambda^{(0)} = \left(\lambda_1^{(0)}, \ldots, \lambda_L^{(0)}\right)$ is then introduced.

By reference, the following is meant: it is a deterministic (non-random) vector $\lambda^{(0)}$ that describes the expected number of registered photons λ_k that should be realized in the nominal operating mode at each interval of the observation window, given nominal levels

of the useful and background radiation components and a known structure of the service packet. The reference profile $\lambda^{(0)}$ is used only as a baseline when computing the features $\Phi(N, S)$, which measure deviations of the observed values N_k from the nominal statistics of the normal operating mode.

The nominal statistical levels of the normal operating mode are specified as the nominal expected numbers of registered photons over the corresponding segments of the observation window. A graphical representation of $\lambda^{(0)}(k)$ is shown in Figure 3a.

In each Monte Carlo realization, a realization-specific profile $\lambda^{(m)}$ is used, which incorporates variations in channel conditions and, when applicable, adversarial impacts.

That is, $\lambda^{(0)}$ represents the reference nominal statistics of the normal operating mode, whereas $\lambda^{(m)}$—corresponds to the statistics of a specific realization according to which the photon counts $N_k{}^{(m)}$ are actually generated.

The initial prescribed parameters used for the simulation modeling, providing comparable data structures and comparable sensitivity curves, are summarized in Table 1.

Table 1. Parameters of the observation window and the structure of the service bit packet used in the simulation model.

Parameter	Symbol	Value	Unit	Description
Number of intervals in the observation window of the "preamble + control bit sequence (CRC) + background" segment	L	20	-	Dimension of the vectors λ and N
Duration of a single observation-window interval	ΔT	1×10^{-6}	s	1 μs per interval
Indices of preamble intervals in the observation window	K_p	0–4	-	5 intervals
Indices of the control-bit sequence intervals in the observation window	K_c	5–9	-	5 intervals
Indices of background-radiation intervals in the observation window	K_0	10–19	-	10 intervals
Binary preamble template of the service bit packet	S_{preamble}	[1, 1, 0, 1, 0]	-	For intervals K_p
Binary control-sequence (CRC) template of the service bit packet	S_{crc}	[1, 0, 1, 1, 0]	-	For intervals K_c

The number of observation-window intervals $L = 20$ is chosen as minimally sufficient to ensure that a single window simultaneously contains: (i) preamble intervals, from which deviation and contrast features are formed; (ii) control bit sequence (CRC) intervals, from which independent deviation features are formed for the second service segment; and (iii) a separate background segment, from which the level and variability of the background component are estimated. The presence of a dedicated background segment is essential for reproducible computation of the background-variance feature and for stable computation of the "service intervals–background intervals" contrast; therefore, 10 background intervals are used rather than 1–2.

The partitioning into five preamble intervals, five CRC intervals, and 10 background intervals is sufficient for two reasons. First, averaging over the preamble and CRC is based on at least five samples, which yields stable mean deviations (otherwise, a single random

Poisson outlier would begin to dominate the feature). Second, estimating the variance of the background component over 10 intervals ensures stability of the estimate; when variance is estimated from only 2–3 points, the estimation spread is excessively large and begins to "noise" at the detector level.

The interval duration $\Delta T = 1$ μs is chosen as a compromise between: (i) the requirement of the discrete-time photon-counting model that "one interval corresponds to one element of the service sequence"; (ii) neglecting intersymbol interference within the observation window; and (iii) ensuring a sufficient number of photons per interval for reproducible computation of statistical features. At the level of the numerical experiment setup, $\Delta T = 1$ μs defines a typical time scale of the elementary structure of the service segment and is consistent with the assumption of quasi-stationarity of conditions within a single observation window.

Varying the number of observation-window intervals L affects both the statistical stability of the Poisson-counting feature formation $\Phi(N, S)$ and the decision latency of the protection block. If L is reduced (and, consequently, fewer intervals are available in the "preamble + CRC" segment and in the background set K_0), the sample-mean and sample-variance estimates implicit in the features x_1–x_5 become noisier, which widens the nominal feature distribution under H_0 and tends to either increase the false-alarm probability for fixed thresholds or require more conservative thresholds that reduce detection sensitivity. If L is increased while keeping the same relative partitioning between preamble, CRC, and background, feature estimates become more stable, but the observation-window duration $L\Delta T$ increases the reaction time and may challenge the within-window quasi-stationarity assumption for the intensity profile $\lambda(k)$, which can again broaden the nominal H_0 statistics. For this reason, $L = 20$ is used as a minimally sufficient configuration that preserves stable estimation in all three segments while keeping latency low under the adopted ΔT.

The nominal expected numbers of registered photons in the normal operating mode are listed in Table 2.

Table 2. Nominal expected numbers of registered photons used in the simulation model.

Parameter	Symbol	Value	Unit	Description
Nominal expected number of registered photons in preamble "1"-intervals	$\lambda_{\mathrm{p,nom}}$	14	counts per interval ΔT	For $k \in K_p$ with $S_{\mathrm{preamble}} = 1$
Nominal expected number of registered photons in CRC "1"-intervals	$\lambda_{\mathrm{c,nom}}$	7	counts per interval ΔT	For $k \in K_c$ with $S_{\mathrm{crc}} = 1$
Nominal expected number of registered photons of the background component	$\lambda_{\mathrm{b,nom}}$	2	counts per interval ΔT	For $k \in K_0$ and for "0"-intervals

The values $\lambda_{\mathrm{p,nom}}$, $\lambda_{\mathrm{c,nom}}$, $\lambda_{\mathrm{b,nom}}$ define the nominal expected numbers of registered photons over the "preamble + control bit sequence" segment of a single service bit packet in the normal operating mode of the service VLC channel. Here, $\lambda_{\mathrm{p,nom}}$ applies to those observation-window time intervals in which, according to the binary preamble template, a bit '1' is transmitted; $\lambda_{c,\mathrm{nom}}$ applies to those observation-window time intervals in which, according to the binary control-sequence template, a bit '1' is transmitted; $\lambda_{b,nom}$ applies to those observation-window time intervals in which the useful signal is absent, and the counting is formed by background illumination and internal noise of the receiver front-end.

These nominal levels are selected such that the contrast between service intervals and background intervals of the observation window is pronounced at the level of the expected values λ_k, while preserving the Poisson variability of the counting process, which subsequently manifests itself in the realizations N_k.

To prevent the nominal statistics from becoming an "ideal template", inter-window variability of reception conditions is modeled in each Monte Carlo realization: each realization corresponds to one observation window (one received service bit packet over the "preamble + control bit sequence" segment), and the variability is introduced as variations in the expected values $\lambda_k^{(s)}$ and $\lambda_k^{(b)}$ between such observation windows.

Numerical values of the inter-realization variability parameters and the Monte Carlo sample sizes used for threshold tuning, testing, and sweeping over the attack strength are summarized in Table 3.

Table 3. Monte Carlo sample sizes and inter-realization variability parameters used in the numerical experiments.

Parameter	Symbol	Value	Unit	Description
Training sample size for the nominal mode	$M_{\text{train},0}$	80,000	-	Threshold tuning and estimation of nominal-mode statistics
Test sample size for the nominal mode	$M_{\text{test},0}$	20,000	-	Estimation of P_{FA}
Number of attack samples per type for training the logistic detector	$M_{\text{train},1}$	20,000	-	Training of the logistic detector (labeled data)
Number of realizations per sweep point	M_{sweep}	4000	-	Estimation of $P_{\text{D}}(\xi)$
Relative inter-window variation in the useful signal component	*signal_variation_rel*	0.15	-	Variations in the useful component between observation windows
Absolute inter-window variation in the background component	*background_var_abs*	0.4	counts per interval ΔT	Variations in the background component between observation windows
Additional fluctuation on service '1'-intervals	*noise_std_counts*	0.5	counts per interval ΔT	Residual uncertainty on service intervals

The sample sizes are chosen to ensure statistically stable estimates of trigger rates. The nominal-mode test sample size $M_{\text{test},0}$ = 20,000 provides sufficient capability to estimate the false-alarm probability at the level of a few 10^{-3}, while the number of realizations per sweep point M_{sweep} = 4000 yields sufficiently smooth detection-probability curves without noticeable "jitter" with respect to the attack-strength parameter applied to the channel and the AI subsystem. The nominal-mode training sample size $M_{\text{train},0}$ = 80,000 is required for stable threshold tuning based on the tails of the nominal feature distributions (without overfitting to random outliers), and 20,000 attack examples per attack type used to train the logistic detector ensure representativeness of within-scenario attack variations and stability of the learned weights under fixed thresholds.

The empirical false-alarm probability of the AI subsystem is estimated on an independent nominal-mode test sample as the fraction of realizations for which a decision indicating an attack is produced, as given in (33):

$$P_{FA} = \frac{1}{M_{test,0}} \sum_{m=1}^{M_{test,0}} I(f(x^m;\theta) = 1), \tag{33}$$

where I(·) is the event indicator function (dimensionless).

The sensitivity characteristic of the attack detection probability is constructed for a fixed impact type and a fixed attack strength as the fraction of attack realizations recognized as an attack, as given in (34):

$$P_D(\xi) = \frac{1}{M_{sweep}} \sum_{m=1}^{M_{sweep}} I(f(x_{a,\xi}^{(m)}; \theta) = 1), \quad (34)$$

where $x_{a,\xi}^{(m)}$ (dimensionless vector of the corresponding dimension) is the feature vector for the m-th realization under an impact of type a and attack strength ξ.

The numerical experiment is implemented in Python 3.12.2 (Anaconda distribution): Poisson counting realizations are generated using NumPy 1.26.4 and SciPy 1.12.0, and the AI subsystem detectors are implemented and trained using Scikit-learn 1.3.0. The simulation is organized as a sequence of fixed steps for each observation window of a single service packet.

First, a realization-specific profile of expected values over the observation window is formed: a background component with inter-window fluctuation is specified, and then, on the "1"-intervals of the preamble and the control bit sequence, a useful component is added with inter-window variation in the useful-signal level. Next, if an impact is present, the expected-value profile is modified according to the selected impact type. Based on the resulting expected-value profile, a counting-observation vector for the observation window is generated under the Poisson model, after which the AI subsystem feature vector is computed from the counting observations and the decisions of the one-class detector, the energy detector, the logistic detector, and the ensemble (using the "2-out-of-3" rule) are obtained.

That is, in the model, an impact modifies the expected photon-counting values over the observation window, while changes in the channel-state features and detector decisions are a consequence of changes in the counting observations generated under the modified statistics.

The results of the simulation modeling of the attack detection system for the VLC channel with the AI subsystem are presented in Figure 3a–h.

In Figure 3a, the reference nominal expected-value profile $\lambda^{(0)}(k)$ is shown and used as the H_0 baseline for computing $\Phi(N, S)$, i.e., deviations of photon counts are measured relative to a physically consistent nominal profile on both service and background intervals. Figure 3b clarifies the observability mechanisms of the considered impact types at the level of $\lambda(k)$: DC-jamming mainly shifts the background intervals upward, pulsed interference adds excess counts on a subset of service intervals, spoofing disrupts the correspondence between the service-interval structure and the nominal expected counts, and replay partially preserves the service structure while mixing recorded fragments. Figure 3c indicates that the Poisson variability is preserved for a fixed observation-window structure; consequently, decisions are formed statistically over the window rather than by deterministic template matching. The feature projection in Figure 3d provides an empirical consistency check that $\Phi(N, S)$ induces separation between H_0 and H_1, which is consistent with the high ranking performance of the monotonic logistic component (0.952).

Figure 3e–h quantify sensitivity margins for a fixed deployed operating point because all thresholds are calibrated on H_0 only and then kept unchanged during the sweep over the impact-strength parameter ξ. Therefore, the resulting $P_D(\xi)$ curves should be interpreted as *margins* of a fixed deployed configuration at a fixed H_0-calibrated operating point, rather than as performance after retuning to each ξ. On an independent H_0 test sample, the ensemble false-alarm probability is $P_{FA} = 0.045$, while the component trigger rates

are comparable (OC = 0.0468; Energy = 0.0506; LR = 0.0508), which confirms consistent H_0-based calibration across heterogeneous scoring rules and enables a fair comparison of $P_D(\xi)$ across impact types.

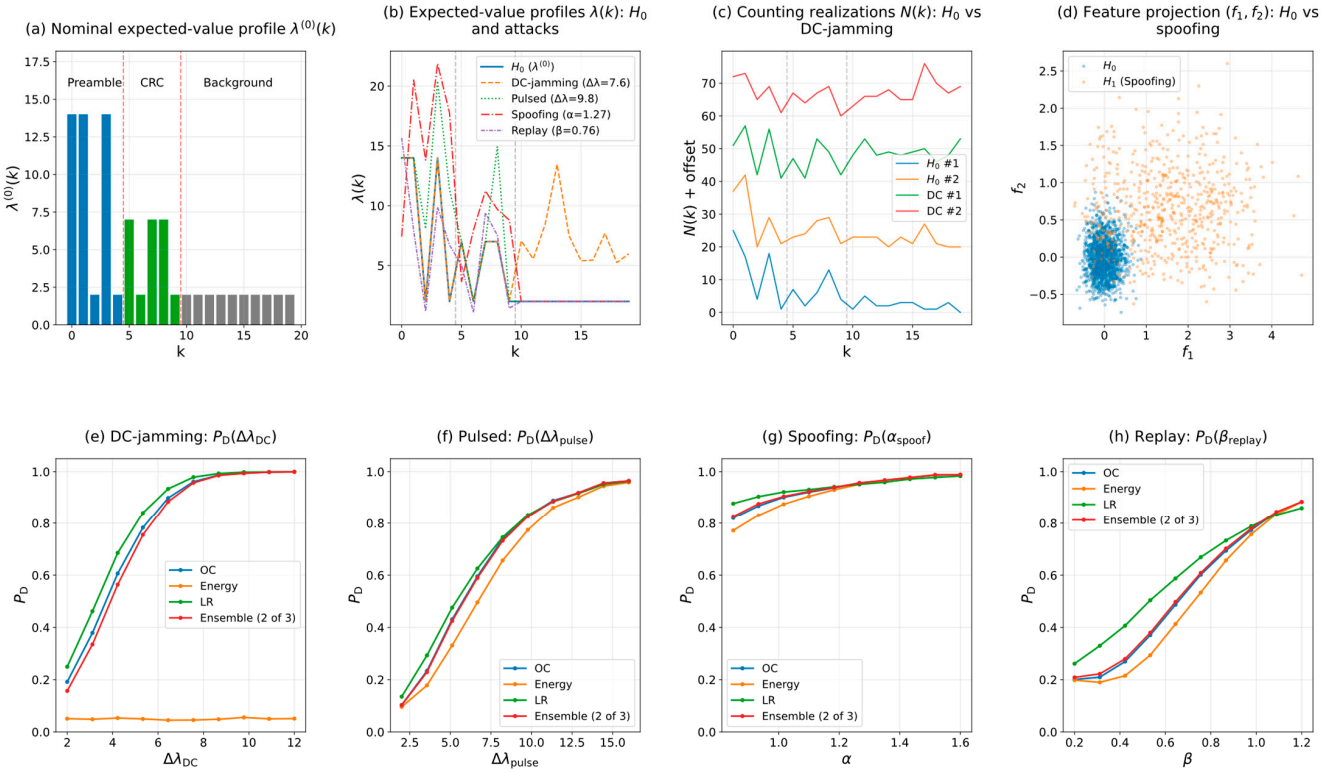

Figure 3. Results of the simulation experiment for evaluating false alarms and the sensitivity of the artificial intelligence subsystem under adversarial impacts on the service VLC channel and the AI subsystem of the onboard computer of a multichannel SORA system.

For the "2-out-of-3" ensemble, the minimum impact strength required to achieve $P_D \geq 0.90$ is $\Delta\lambda_{DC} \approx 7.56$ for DC-jamming (Figure 3e), $\Delta\lambda_{pulse} \approx 12.89$ for pulsed interference (Figure 3f), and $\alpha \approx 1.02$ for spoofing (Figure 3g). These thresholds have a direct interpretation in terms of the nominal expected-count levels used in the model: DC-jamming becomes reliably observable once the upward shift in background intervals produces a window-level deviation that dominates the nominal Poisson variability, whereas pulsed interference requires a larger $\Delta\lambda$ because it affects only a subset of service intervals and its effect is partially averaged in the aggregated features. Spoofing reaches the same detectability margin at a small parameter change because it is a structural mismatch impact relative to $\lambda^{(0)}(k)$ on the service segment, which produces persistent inconsistency in the service-interval aggregates used by $\Phi(N, S)$. In contrast, for replay, the target level $P_D \geq 0.90$ is not reached on the specified grid of β (Figure 3h). This identifies replay as the limiting case for the current feature set within the selected range: replay can preserve a substantial fraction of the service-segment structure and therefore induce smaller shifts in the same window-level Poisson-counting aggregates relative to H_0 than interference-type impacts and spoofing. From an operational viewpoint, achieving the same detectability margin for replay-like impacts with CPU-only processing requires either a more informative control structure within the service packet or an augmented feature map that is more sensitive to segment reuse while remaining compatible with the low-dimensional Poisson-counting formalism.

It is important to emphasize the practical link to the quality of the LED emitter, since it determines the stability and reproducibility of the photon-counting statistics on which the AI subsystem is trained and calibrated. Optical power instability, temperature drift, growth

of the noise component, and luminous-flux degradation (aging) lead to systematic shifts in the nominal statistics and an increase in inter-window variability, which directly affects P_{FA} (through heavier tails of the H_0 feature distributions) and $P_D(\xi)$ (through reduced separability of H_0 and H_1). Therefore, the curves in Figure 3 should be interpreted as sensitivity estimates under the specified source-quality level (stable nominal levels and limited variations). A natural extension of the model is to introduce a parameterization of LED instability/degradation [36].

Benchmarking Against AI-Based Methods and Poisson Baselines

A quantitative performance comparison is provided using (i) representative security-detection results reported in the literature with explicit numerical metrics, and (ii) an internal apples-to-apples benchmark of Poisson photon-counting baselines and lightweight ML detectors within the unified photon-counting control-segment model.

In this work, the AI component is the learned decision layer that classifies integrity states from photon-counting features. Logistic regression operates as a supervised ML classifier, while the final decision is produced by an AI-assisted fusion rule combining LR, a robust one-class anomaly detector, and a conventional energy statistic. This setting corresponds to AI-assisted security because the integrity decision is driven by a trainable ML component and decision fusion rather than by a single fixed-form statistical test.

All internal detectors are calibrated on nominal data to the same operating point $P_{FA} \approx 0.05$ and compared using the minimum attack strength required to achieve $P_D \geq 0.90$, or the maximum achieved P_D when the target is not reached on the tested grid. Under this criterion, LR yields the highest sensitivity to DC-jamming and spoofing, whereas EN (a representative Poisson-statistic baseline) can be insensitive to DC-jamming in the integrity-attack setting. Replay remains the most challenging case for all tested detectors within the examined β range.

For VLC-specific attack detection, a cooperative attack detection method for LED-based VLC reports 91% attack detection accuracy and a minimum detection rate of 84% in obstacle-rich environments [37].

As a widely used ML-security baseline for jamming detection in wireless communications, Arjoune et al. report $P_D = 97.5\%$ with $P_{FA} = 5.6\%$ for a Random-Forest detector [38]. Table 4 summarizes these reference results alongside the internal benchmark obtained under matched false-alarm constraints.

Table 4 should be interpreted under the fixed low-false-alarm operating point $P_{FA} \approx 0.05$, which is required for an onboard integrity monitor. Hence, the lower replay detection at $\beta = 0.7$ is not a flaw but an expected consequence of a physically consistent photon-counting setting: replay preserves much of the nominal control-segment structure and therefore remains statistically close to H_0 within Poisson variability and channel fluctuations, making moderate replay a limiting hard case for the given feature set and segment length. This conservative behavior indicates that the benchmark does not artificially inflate separability and helps localize the dominant residual risk. Table 4 also provides the requested performance comparison by aligning the internal benchmark with representative AI/security reference results and by contrasting learning-based detectors with conventional Poisson photon-counting baselines under the same unified model.

Table 4. Quantitative comparison (external security-detection methods vs. internal benchmark at $P_{FA} \approx 0.05$).

Method	Metric(s)	Numbers	Comment
Cooperative VLC attack detection [37]	Detection accuracy/detection rate	91% accuracy; min 84% detection rate	Closest VLC attack detection reference; scenario-dependent metrics
Random Forest jamming detector [38]	P_{FA}; P_D	P_D = 97.5% P_{FA} = 5.6%	Cross-domain ML reference (non-VLC); lacks photon-count modeling
This paper: LR (ML detector)	P_{FA}; P_D at fixed attack levels	P_{FA} = 5.08% P_D ($\Delta\lambda_{DC}$ = 6.0) = 89.7% P_D ($\Delta\lambda_{pulse}$ = 11.0) = 87.3% P_D (α = 1.15) = 93.7% P_D (β = 0.7) = 62.8%	Highest sensitivity for DC/spoof; physics-informed ML
This paper: AI-assisted fusion		P_{FA} = 4.50% P_D ($\Delta\lambda_{DC}$ = 6.0) = 83.3% P_D ($\Delta\lambda_{pulse}$ = 11.0) = 87.3% P_D (α = 1.15) = 93.3% P_D (β = 0.7) = 55.3%	Lower false alarms; robust multi-detector fusion

5. Conclusions

This study addressed the problem of protecting an artificial intelligence software subsystem that monitors a service visible light communication channel based on an LED emitter as part of a compact multichannel SORA system under limited onboard computational resources. A threat profile was formulated, including impacts of the DC-jamming, pulsed, spoofing, and replay classes, and an interpretable anomaly detection architecture was proposed. The architecture is based on a Poisson photon-counting model over the "preamble + control bit sequence" segment, a feature-mapping procedure for count-based features, and an ensemble of "lightweight" detectors executable on a central processing unit without graphical acceleration.

The simulation results support the working hypothesis that attacks realized as controlled modifications of the counting-process statistics over the service segment of the packet lead to reproducible shifts in the feature space. These shifts are sufficient for stable discrimination between the nominal operating hypothesis and the presence of an impact under fixed thresholds tuned only on a nominal-mode sample. Comparative analysis of the sensitivity curves further indicates non-uniform observability of different attack classes within the selected feature set: energy-based impacts manifest differently from protocol-level impacts, and replay scenarios require additional "contextual" feature informativeness associated with the timeliness and temporal consistency of service messages.

Under the fixed low-false-alarm operating point obtained by threshold calibration on nominal data only, the final "2-out-of-3" ensemble yields an empirical nominal-mode false-alarm probability of 0.045 on an independent nominal test sample. Under the same fixed operating point, the target detection probability of at least 0.90 is achieved for DC-jamming once the mean background count is increased by about 7.6 counts per interval, for pulsed interference at an increase of about 12.9 counts per interval, and for spoofing at a small signal-scaling mismatch of about 2%. In contrast, for replay, the target detection level is not reached within the tested replay-strength range, which confirms replay as the limiting hard case for the considered feature set at a low false-alarm constraint. The corresponding per-packet decision time scale is determined by the adopted obser-

vation window (20 intervals of 1 microsecond each), which supports real-time CPU-only deployment without graphical acceleration.

These simulation results provide actionable implications for protecting the AI decision layer in resource-constrained airborne systems. The reported sensitivity curves link the attack impact strength in the service-segment photon-count statistics to the achievable detection probability at the selected low-false-alarm operating point, which enables design-level tuning of operating margins for a target false-alarm rate. The protection stage remains lightweight because it relies on Poisson-counting features extracted from short service packets and a fixed "2-out-of-3" ensemble of three simple detector scores, so it can be deployed per service VLC link (or per channel) in a multi-channel airborne radar station without graphical acceleration. Scaling to larger configurations is therefore achieved by parallel per-link deployment with nominal-data calibration, while keeping the feature definitions and voting logic unchanged. Beyond the specific SORA service VLC setting, the same pattern can be transferred to other edge AI subsystems in which integrity can be monitored through short-segment count statistics of service messages under adversarial perturbations.

Future work should extend the model by explicitly parameterizing degradation and instability of the LED emitter and by augmenting the features and/or protocol mechanisms to improve observability of replay scenarios (e.g., via freshness and temporal-consistency verification of service messages).

Author Contributions: Conceptualization, V.P.K. and V.A.N.; methodology, V.P.K.; software, V.P.K.; validation, V.A.N., S.S.D. and O.V.V.; formal analysis, V.P.K.; investigation, V.P.K.; resources, V.A.N. and O.V.V.; data curation, V.P.K.; writing—original draft preparation, V.P.K.; writing—review and editing, V.P.K., V.A.N., S.S.D. and O.V.V.; visualization, V.P.K.; supervision, V.A.N.; project administration, V.P.K.; funding acquisition, V.A.N. All authors have read and agreed to the published version of the manuscript.

Funding: The research of Vadim A. Nenashev and Vladimir P. Kuzmenko was supported by a grant from the Russian Science Foundation (project No. 24-79-10259).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SORA	Small Onboard Radar Station
MIMO	Multiple-Input Multiple-Output
ESA	Effective Scattering Area
OWC	Optical Wireless Communication
LED	Light-Emitting Diode
VLC	Visible Light Communication
LiFi	Light Fidelity
RF	Radio Frequency
AI	Artificial Intelligence
LOS	Line-of-Sight
CRC	Control Bit Sequence (Cyclic Redundancy Check)
OC	One-Class
LR	Logistic Regression

Appendix A. Mathematical and Modeling Details

This appendix collects intermediate mathematical details that are omitted from Sections 2 and 3 to improve the flow of the main text while preserving reproducibility. The main text retains only the basic Poisson counting model, the attack-driven parameter transformations, and the final feature definitions, whereas Appendix A provides the expanded nominal probability formulation, the optional physical power-to-count relationship, and the detector scoring functions with threshold calibration.

Appendix A.1. Nominal Poisson Formulation and Background Component Details

The following expression is provided for completeness and specifies the nominal counting statistics at the level of discrete-time photon counts over the observation window.

The probability mass function of observing a specific set of count values $N = (N_1, N_2, \ldots, N_L)$ in the absence of attacks (hypothesis H_0) and given parameters $\lambda = (\lambda_1, \ldots, \lambda_L)$, assuming conditional independence of the values N_k, is given by the probability function (A1):

$$p(N|H_0, \lambda) = \prod_{k=1}^{L} \frac{\lambda_k^{N_k} e^{-\lambda k}}{N_k!}, \tag{A1}$$

where $p(N|H_0, \lambda)$ (dimensionless) is the probability mass function of observing the set of counting values N under the nominal operating hypothesis H_0;

$\lambda = (\lambda_1, \ldots, \lambda_L)$ is the vector of parameters of the discrete-time photon-counting process;

$N_k!$ (dimensionless) is the factorial of the number of registered photons in the k-th interval.

This formulation clarifies the probabilistic meaning of the nominal hypothesis H_0 used throughout the paper and serves as the reference model for defining feature statistics and false-alarm control.

The model above operates in the photon-count domain through the intensity parameters $\lambda_k^{(s)}$. If a physical interpretation in terms of received optical power is required, $\lambda_k^{(s)}$ can be related to the optical power over the same time interval as follows.

If required, the physical relationship between the expected number of photons registered in the k-th time interval of duration ΔT, due to the useful optical signal of the LED transmitter during the transmission of a service bit packet $\lambda_k^{(s)}$, and the corresponding optical power can be specified as in (A2):

$$\lambda_k^{(s)} = \eta \frac{P_k^{(opt)} \Delta T}{h\nu}, \tag{A2}$$

where $\eta \in (0, 1)$ (dimensionless) is the quantum efficiency of the photodetector;

$P_k^{(opt)}$ (W) is the average optical power of the light-emitting diode in the k-th interval;

h (J·s) is Planck's constant;

ν (Hz) is the optical frequency of the LED radiation.

In the main text, the detection algorithm and the AI subsystem operate directly with the expected photon-count parameters λ_k and the derived intra-packet features; therefore, the power-domain relationship is included here only as an optional physical link.

The optical LOS channel is characterized by an effective channel gain G_{ch} (dimensionless), defined as the ratio between the received optical power at the photodetector aperture and the transmitted optical power. Under the standard Lambertian emission model, this gain is given in (A3) and (A4):

$$G_{ch} = \frac{(m+1)A_r}{2\pi\ell^2} \cos^m(\varphi) \cos(\psi) T_s(\psi) g(\psi), \tag{A3}$$

$$0 \leq \psi \leq \Psi_c, \tag{A4}$$

where m is the Lambertian order of the transmitter (dimensionless);

A_r is the photodetector active area (m^2);

ℓ is the transmitter–receiver distance (m);

φ and ψ are the irradiance and incidence angles (rad);

$T_s(\psi)$ is the optical filter gain (dimensionless);

$g(\psi)$ is the concentrator gain (dimensionless);

Ψ_c is the receiver field-of-view (rad).

By definition, the received optical power at the photodetector aperture in the k-th interval is $P_r(k) = G_{ch}\, P_k$ is the transmitted LED optical power in that interval. Substituting $P_r(k)$ into the photon–power relation yields the useful-signal contribution to the Poisson intensity, and the total intensity λ_k is formed by adding the background component as defined in Section 3.

Appendix A.2. Detector Scoring Functions and Threshold Calibration

This section provides the explicit detector scoring functions and the threshold calibration logic used by the ensemble described in Section 3. In all cases, thresholds are selected using nominal-mode realizations to ensure that the empirical false-alarm probability does not exceed the prescribed level.

Throughout this appendix, the label is used only as a reference annotation for simulated realizations for training, threshold calibration, and evaluation. In online operation of the AI subsystem, only the feature vector is available, whereas the reference label is not observed and is used solely to quantify false-alarm and missed-detection probabilities.

First, the one-class detector (OC) is defined; it provides a lightweight deviation score under the nominal operating mode using nominal feature statistics. For this purpose, a scalar scoring function in the form of a quadratic form of the Mahalanobis-distance type is introduced, as given in (A5):

$$s_{OC}(x;\theta_{OC}) = (x - \mu_0)^T \Sigma_0^{-1} (x - \mu_0), \tag{A5}$$

where $s_{OC}(x;\theta_{OC})$ (dimensionless) is a quadratic form of the Mahalanobis-distance type;

$\theta_{OC} = (\mu_0, \Sigma_0, \gamma_{OC})$—are the parameters of the one-class detector, including the estimates μ_0, Σ_0, and the threshold γ_{OC}.

μ_0—is the sample estimate of the mean of the feature vector in the nominal operating mode, computed as the arithmetic mean of all $x^{(m)}$ with $y^{(m)} = 0$, μ_0 has the same units as the corresponding feature.

Σ_0—is the sample covariance matrix of the features in the nominal operating mode, i.e., the matrix of variances and covariances computed over the same set of $x^{(m)}$; the diagonal elements have units equal to the squared units of the corresponding features, while the off-diagonal elements have units equal to the product of the units of the corresponding feature pairs.

The local decision of the one-class detector is defined by the rule in (A6):

$$f_{OC}(x;\, \theta_{OC}) \in \{0,\, 1\}, \tag{A6}$$

where $f_{OC}(x;\, \theta_{OC}) = 1$ corresponds to a decision indicating the presence of an anomaly or attack;

γ_{OC} (dimensionless) is a threshold selected based on the sample of nominal realizations such that the empirical false-alarm probability of this detector does not exceed a prescribed level $\overline{P}_{FA}^{(OC)}$.

While the OC detector captures multivariate deviations through the nominal covariance structure, the energy detector provides a simpler aggregate deviation measure based on the joint magnitude of the feature vector. Within the implementation of the energy detector, the following scoring function is used, as given in (A7):

$$S_{EN}(x;\theta_{EN}) = \sum_{i=1}^{5} x_i^2, \tag{A7}$$

where $S_{EN}(x;\theta_{EN})$ (counts^2, if the components x_i are expressed via counting characteristics) is the sum of squares of the components of the feature vector, used as an integral measure of the joint deviation of all features from their typical level in the nominal operating mode;

$\theta_{EN} = (\gamma_{EN})$ is the threshold parameter of the energy detector, determined in an analogous manner;

γ_{EN} (counts^2)—is the threshold tuned on the nominal-mode sample according to the target level $\overline{P}_{FA}^{(EN)}$.

Finally, to incorporate labeled attack examples and exploit discriminative feature combinations, the LR detector operating on the same feature vector is introduced. The LR detector is constructed using the labeled dataset $\{(x^{(m)}, y^{(m)})\}$, which contains realizations of both the nominal operating mode ($y^{(m)} = 0$), and attacks ($y^{(m)} = 1$).

The scalar output of the detector is defined as in (A8):

$$\pi_{LR}(x;\theta_{LR}) = \sigma\left(w^T x + b\right), \tag{A8}$$

where $\pi_{LR}(x;\theta_{LR}) \in \{0, 1\}$ (dimensionless) is an estimate of the conditional attack probability $P(H_1|x)$ under the logistic regression model;

$\sigma(z) = \frac{1}{1+e^{-z}}$ (dimensionless function) is the logistic link function that monotonically maps any real-valued argument z (dimensionless) to the interval (0, 1);

$w \in \mathrm{R}^5$ is the vector of feature weights (dimensionless);

$b \in R$ is the scalar bias term of the logistic model (dimensionless) added to the linear combination $w^{\mathrm{T}}x$; increasing b makes the detector more "sensitive" to attacks (all else being equal, it more often yields larger values of $\pi_{LR}(x;\theta_{LR})$, whereas decreasing b makes it more "conservative" (more often tending to remain in the nominal-mode decision);

$\theta_{LR} = (w, b, \tau_{LR})$ is the set of parameters of the logistic detector, including the threshold τ_{LR} applied to the estimated attack probability;

$\tau_{LR} \in \{0, 1\}$ (dimensionless) is the threshold on the estimated attack probability selected such that the empirical false-alarm probability of the logistic detector does not exceed a prescribed level $\overline{P}_{FA}^{(LR)}$.

The local binary decisions produced by the OC detector, the energy detector, and the LR detector are combined in the main text using the "2-out-of-3" voting rule, yielding the final ensemble decision used for attack detection and for initiating adaptive countermeasures. Under this fusion rule, the ensemble declares an attack if at least two of the three local detector decisions indicate an anomaly/attack; otherwise, a nominal decision is retained.

References

1. An, D.; Ge, B.; Wang, W.; Chen, L.; Feng, D.; Zhou, Z. Research on the technology of airborne multi-channel wide angle staring SAR ground moving target indication. *J. Radars* **2023**, *12*, 1179–1201. [CrossRef]
2. Liu, K.; Li, Y.; Xu, Z.; Zhou, Z.; Jin, T. Airborne multi-channel forward-looking radar super-resolution imaging using improved fast iterative interpolated beamforming algorithm. *Remote Sens.* **2024**, *16*, 4121. [CrossRef]
3. Nguyen, H.; Al, I.; Jang, Y.M. Survey of next-generation optical wireless communication technologies for 6G and beyond 6G. *ICT Express* **2025**, *11*, 576–589. [CrossRef]

4. Gupta, S.; Roy, D.; Bose, S.; Dixit, V.; Kumar, A. Illuminating the future: A comprehensive review of visible light communication applications. *Opt. Laser Technol.* **2024**, *177*, 111182. [CrossRef]
5. Wang, W.; Zeng, Z.; Chen, C.; Wang, D.; Liu, M.; Haas, H. UAV Array-Aided Visible Light Communication with Enhanced Angle Diversity Transmitter. *Sensors* **2025**, *25*, 5752. [CrossRef]
6. Nenashev, V.A.; Shepeta, A.P.; Kryachko, A.F. Fusion radar and optical information in multi-position on-board location systems. In Proceedings of the 2020 Wave Electronics and Its Application in Information and Telecommunication Systems (WE-CONF), St. Petersburg, Russia, 1–5 June 2020; pp. 1–5. [CrossRef]
7. Nenashev, V.A.; Bestugin, A.R.; Rabin, A.V.; Solenyi, S.V.; Nenashev, S.A. Modified nested Barker codes for ultra-wideband signal-code constructions. *Sensors* **2023**, *23*, 9528. [CrossRef]
8. Ullmann, I.; Bonfert, C.; Grathwohl, A.; Lahmeri, M.A.; Mustieles-Pérez, V.; Kanz, J.; Sterk, E.; Bormuth, F.; Ghasemi, R.; Fenske, P.; et al. Towards detecting climate change effects with UAV-borne imaging radars. *IEEE J. Microw.* **2024**, *4*, 881–893. [CrossRef]
9. Nenashev, V.A.; Khanykov, I.G. Formation of fused images of the land surface from radar and optical images in spatially distributed on-board operational monitoring systems. *J. Imaging* **2021**, *7*, 251. [CrossRef] [PubMed]
10. Tarasenkov, M.V.; Poznakharev, E.S.; Fedosov, A.V. Non-line-of-sight atmospheric optical communication in the visible wavelength range between UAV and the ground surface. *Atmosphere* **2024**, *15*, 21. [CrossRef]
11. Gupta, A.; Dhawan, D.; Gupta, N. Review on UAV-based FSO links: Recent advances, challenges, and performance metrics. *Opt. Eng.* **2023**, *63*, 041204. [CrossRef]
12. Rajahrajasingh, H.; Jayakody, D.N.K. Unmanned aerial vehicle-assisted terahertz–visible light communication systems: An in-depth performance analysis. *Sensors* **2024**, *24*, 4080. [CrossRef]
13. Liu, L.; Wang, A.; Wu, J.; Lu, J.; Li, J.; Sun, G. Secure and energy-efficient unmanned aerial vehicle-enabled visible light communication via a multi-objective optimization approach. *arXiv* **2024**, arXiv:2403.15410. [CrossRef]
14. Liu, Y.; Lu, J.; Sun, G.; Liu, L.; Zhang, J. Optical power coverage optimization for UAV-enabled visible light communication. In Proceedings of the ICC 2022—IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; pp. 1–6.
15. Li, Y.; Wu, L.; Zhang, Z.; Dang, J.; Zhu, B.; Zhang, X.; Wu, Y. Sensing assisted optical wireless communication for UAVs. *IEEE Trans. Veh. Technol.* **2024**, *73*, 18620–18634. [CrossRef]
16. Guaña-Moya, J.; Román Cañizares, M.; Palacios Játiva, P.; Sánchez, I.; Ruminot, D.; Lobos, F.V. Comprehensive survey on VLC in e-healthcare: Channel coding schemes and modulation techniques. *Appl. Sci.* **2024**, *14*, 8912. [CrossRef]
17. Loureiro, P.A.; Guiomar, F.P.; Monteiro, P.P. Visible Light Communications: A Survey on Recent High-Capacity Demonstrations and Digital Modulation Techniques. *Photonics* **2023**, *10*, 993. [CrossRef]
18. Zhang, X.; Klevering, G.; Lei, X.; Hu, Y.; Xiao, L.; Tu, G.-H. The security in optical wireless communication: A survey. *ACM Comput. Surv.* **2023**, *55*, 329. [CrossRef]
19. Luo, B.; Cao, H.; Cui, J.; Lv, X.; He, J.; Li, H.; Peng, C. SAR-PATT: A Physical Adversarial Attack for SAR Image Automatic Target Recognition. *Remote Sens.* **2025**, *17*, 21. [CrossRef]
20. Liu, C.; Feng, L.; Wang, J.; Sun, H.; Hu, R.Q.; Lu, H. IRS-VLC physical layer security scheme: A dual-strategy against eavesdropping and attacks. *Appl. Opt.* **2025**, *64*, 703–711. [CrossRef]
21. Saxena, V.N.; Dwivedi, V.; Gupta, J. Machine learning in visible light communication system: A survey. *Wirel. Commun. Mob. Comput.* **2023**, *2023*, 3950657. [CrossRef]
22. Al, I.; Chowdhury, M.; Joha, M.D.; Rahman, M.M.; Jang, Y.M. Machine learning and deep learning in VLC systems: A comprehensive survey. *IEEE Open J. Commun. Soc.* **2025**. [CrossRef]
23. Wadud, A.; Basalamah, A. Optical wireless communications for next-generation radio access networks. *ICT Express* **2025**, *11*, 721–727. [CrossRef]
24. Khazane, H.; Ridouani, M.; Salahdine, F.; Kaabouch, N. A holistic review of machine learning adversarial attacks in IoT networks. *Future Internet* **2024**, *16*, 32. [CrossRef]
25. Vitorino, J.; Oliveira, N.; Praça, I. Adaptative perturbation patterns: Realistic adversarial learning for robust intrusion detection. *Future Internet* **2022**, *14*, 108. [CrossRef]
26. Jamiri, H.; Zyane, A. Adversarial attacks in IoT: A performance assessment of ML and DL models. *Eng. Proc.* **2025**, *112*, 15. [CrossRef]
27. Alenezi, M.N. Significance of machine learning-driven algorithms for effective discrimination of DDoS traffic within IoT systems. *Future Internet* **2025**, *17*, 266. [CrossRef]
28. Mir, M.S.; Guzman, B.G.; Varshney, A.; Giustiniano, D. LiFi for Low-Power and Long-Range RF Backscatter. *IEEE/ACM Trans. Netw.* **2024**, *32*, 2237–2252. [CrossRef]
29. Wei, Z.; Wang, Z.; Zhang, J.; Li, Q.; Zhang, J.; Fu, H.Y. Evolution of optical wireless communication for B5G/6G. *Prog. Quantum Electron.* **2022**, *83*, 100398. [CrossRef]

30. Dhasmana, G.; Sharma, V.; Boob, N.S.; Aravind, K.; Reddy, R.A.; Yadav, K. Securing Wireless and Optical Networks: Advanced Strategies for Network and Information Security in Modern Communication Systems. In Proceedings of the 2025 International Conference on Computational, Communication and Information Technology (ICCCIT), Indore, India, 7–8 February 2025; pp. 956–961. [CrossRef]
31. Marin-Garcia, I.; Guerra, V.; Rabadan, J.; Pérez-Jiménez, R. Visible Light Communication vs. Optical Camera Communication: A Security Comparison Using the Risk Matrix Methodology. *Photonics* **2025**, *12*, 1201. [CrossRef]
32. Al Hasnawi, R.; Marghescu, I. A Survey of Vehicular VLC Methodologies. *Sensors* **2024**, *24*, 598. [CrossRef] [PubMed] [PubMed Central]
33. Sikder, P.; Rahman, M.T.; Bakibillah, A.S.M. Advancements and Challenges of Visible Light Communication in Intelligent Transportation Systems: A Comprehensive Review. *Photonics* **2025**, *12*, 225. [CrossRef]
34. Fang, J.; Pan, J.; Huang, X.; Lin, J.; Jiang, C. Integrated physical-layer secure visible light communication and positioning system based on polar codes. *Opt. Express* **2023**, *31*, 41756–41772. [CrossRef] [PubMed]
35. Huang, S.; Chitnis, D.; Chen, C.; Haas, H.; Khalighi, M.-A.; Henderson, R.K.; Safari, M. Single-photon counting receivers for optical wireless communications in future 6G networks. *IEEE Commun. Mag.* **2024**, *62*, 54–60. [CrossRef]
36. Solyonyj, S.V.; Rabin, A.V.; Solenaya, O.Y.; Kuzmenko, V.P.; Rysin, A.V. Definition and approximation of the light flux degradation of an LED lamp. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *734*, 012197. [CrossRef]
37. Park, S.-H.; Joo, S.; Lee, I.-G. Secure Visible Light Communication System via Cooperative Attack Detecting Techniques. *IEEE Access* **2022**, *10*, 20473–20485. [CrossRef]
38. Arjoune, Y.; Salahdine, F.; Islam, M.S.; Ghribi, E.; Kaabouch, N. A Novel Jamming Attacks Detection Approach Based on Machine Learning for Wireless Communication. In Proceedings of the 2020 International Conference on Information Networking (ICOIN), Barcelona, Spain, 7–10 January 2020; pp. 459–464. [CrossRef]