Secure Task Offloading and Resource Allocation Design for Multi-Layer Non-Terrestrial Networks

Alejandro Flores*, Isabella W. G. da Silva+, Vu Nguyen Ha*, Konstantinos Ntontin*,
Hien Quoc Ngo+, Michail Matthaiou+ and Symeon Chatzinotas*
*Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg
+Centre for Wireless Innovation (CWI), Queen's University Belfast, U.K.
e-mails:*{alejandro.flores, vu-nguyen.ha, kostantinos.ntontin, symeon.chatzinotas}@uni.lu;
+{iwgdasilva01, hien.ngo, m.matthaiou}@qub.ac.uk

***Abstract*—Remote and resource-constrained Internet-of-Things (IoT) deployments often lack terrestrial connectivity for task offloading, motivating non-terrestrial networks (NTNs) with onboard multiaccess edge computing (MEC) capabilities. Nevertheless, in the presence of malicious actors, authentication needs to be performed to avoid non-authorized nodes from draining the computing resources of the NTN nodes. As a solution, we propose a four-layer MEC-enabled NTN with unmanned aerial vehicles (UAVs) acting as access nodes, a high altitude platform station (HAPS) acting as coordinator and authenticator, and a constellation of low-Earth orbit satellites (LEOSats) acting as remote MEC servers. We consider a tag-based physical-layer authentication (PLA) scheme to authenticate legitimate users, and formulate a joint task offloading decision and resource allocation for the admitted tasks, which is solved via block coordinate descent. Numerical results show that the PLA scheme is efficient and performs better than the benchmark schemes. We also demonstrate that the proposed scheme is robust against malicious attacks even under relaxed false-alarm constraints.**

***Index Terms*—Multiaccess edge computing, non-terrestrial networks, physical layer authentication, task offloading.**

I. Introduction

Internet-of-Things (IoT) devices enable continuous data collection and processing across diverse domains, e.g., Industry 4.0 and smart agriculture. Nonetheless, these devices are typically constrained by limited computing capabilities, limiting their ability to process data locally in a timely manner. To address this limitation, task offloading to remote servers (using multiaccess edge computing (MEC) platforms) may be an alternative to achieve low-latency and secure processing. However, even though there are several benefits in employing MEC platforms, the current coverage provided by terrestrial networks (TNs) is limited to approximately 15% of the Earth's surface. This leaves arge areas, such as remote regions, disaster zones, or conflict areas, without reliable connectivity [1].

To overcome this limitation, non-terrestrial networks (NTNs) have emerged as a promising solution for achieving global coverage by utilizing aerial nodes (e.g., unmanned aerial vehicles (UAVs), high altitude platform station (HAPSs)) and spaceborne assets (e.g., low-Earth orbit satellites (LEOSats)). The dynamic deployment and flexible positioning of these NTN nodes enable seamless service provisioning to IoT devices operating beyond the reach of conventional TN infrastructure [2]. For instance, in [3], the authors explored partial task offloading strategies involving local devices and remote ground servers interconnected via ultra-dense LEOSats backhauls. In the recent work of [4], the authors considered a HAPS as a central entity, and multiple UAVs for coordination between the IoT devices and LEOSats in a multi-layer MEC-enabled NTN, for efficient service provisioning to IoT deployments over large areas. It was shown that the use of multiple layers can underpin a simplified coordination across wide-spanning IoT deployments. Also, the use of intermediate nodes, such as UAVs and HAPS, for task offloading to MEC-enabled LEOSats can provide higher communication rates to the LEOSats than the direct access from the IoT nodes, resulting in reduced transmission delays.

Nonetheless, the same openness and wide-area accessibility that make MEC-enabled NTNs attractive for remote IoT applications also expose the offloading interface to malicious nodes, which could attempt to inject task requests and drain the limited onboard computing and backhaul resources, effectively creating a denial-of-service (DoS) condition for legitimate devices. Thus, it is of utmost importance to authenticate the devices for a proper use of the resources of the NTN. However, conventional cryptographic schemes can be computationally infeasible for constrained IoT devices [5]. To address this challenge, physical-layer authentication (PLA) offers a lightweight alternative by leveraging features of the received signal for rapid verification, making it well-suited for massive IoT access [6]. In the context of NTNs, PLA has been investigated [7], such as in [8], where the authors exploited the Doppler-shift and the received power as features for a PLA scheme to authenticate the LEOSat constellations. Also, in [9], the optimal detection thresholds to differentiate between legitimate and spoofing satellites were evaluated. To the best of our knowledge, existing NTN-PLA studies do not quantify how authentication decisions impact task admission, offloading decisions, and MEC resource allocation in multi-layer NTN computing architectures.

Motivated by the above and acknowledging the benefits of multiple NTN layers for IoT-based task offloading, we study secure task offloading in a four-layer MEC-enabled NTN where ground IoT devices intend to offload computation tasks in the presence of malicious nodes. The NTN comprises

multiple UAVs that forward the received signals in an amplify-and-forward (AF) manner to a HAPS, which serves as a central coordinator and authenticator and subsequently offloads admitted tasks to MEC-enabled LEOSats. Our key contributions are threefold: (i) we integrate a tag-based PLA mechanism into the uplink offloading pipeline and derive expressions for the test statistic distribution, the probability of false alarm (PFA), the probability of detection (PD), and the optimized threshold under a fixed false-alarm constraint; (ii) we couple the PLA-driven admission outcome with a joint task offloading and LEO computing-resource allocation formulation that prevents unauthenticated tasks from consuming MEC resources; and (iii) we propose an iterative block-coordinate descent solution and demonstrate via simulations that the proposed PLA-enabled framework improves feasible task admission and robustness against malicious offloading compared with benchmarks schemes. To the best of our knowledge, this is the first work that co-designs PLA-based authentication and MEC offloading/resource allocation in multi-layer NTNs.

Notation: Bold uppercase and lowercase letters denote matrices and vectors, respectively; $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and Hermitian transpose; $\|\cdot\|$ and $|\cdot|$ denote the Euclidean norm and absolute value; $\mathbb{E}\{\cdot\}$ is the expectation operator, and $\mathbf{a}(\cdot,\cdot)$ denotes the array response of a half-wavelength spaced uniform planar array (UPA); $\mathbb{R}\cdot$ extracts the real part, and $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ indicates a circularly symmetric complex Gaussian vector with covariance $\mathbf{C}$.

II. System Model - Joint Authentication, Offloading, and Resource Allocation Framework

As illustrated in Fig. 1, we consider a heterogeneous network where I remotely-located legitimate resource-constrained IoT devices offload computing tasks to an NTN centrally coordinated by a HAPS, denoted by a, with the assistance of U UAVs and in the presence of M malicious nodes. The HAPS acts as a CPU that offloads tasks to a set of MEC-enabled LEOSats to provide computing capacity and it manages the synchronization of the UAVs' transmissions for coherent reception. The malicious nodes attempt to exploit network resources by generating computation-intensive tasks and offloading them to the system, thereby depleting resources available to legitimate devices. To counteract this, the HAPS first authenticates devices and then allocates computing resources. By using a secure channel, the IoT devices and the HAPS agree on a time-varying key-based PLA scheme, which allows the HAPS to verify that the received message signals are indeed from the intended devices in advance [10]. Specifically, the ith device uses its unique secret key $\mathbf{l}_i$, unknown to the malicious nodes, to generate an authentication tag that is superimposed onto the message signal. Although the malicious nodes are assumed to be aware of the authentication mechanism, their lack of knowledge about the secret keys prevents successful impersonation attacks. Moreover, since the keys are assumed to be time-varying, if Eve attempts to replay a previous message from one of the IoT devices, it will not be accepted by the HAPS.

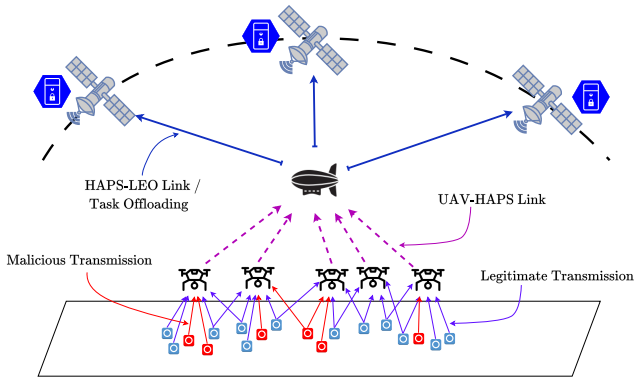

Fig. 1. Illustration of the considered system model.

Let $\mathcal{I} = \{1, \ldots, I\}$, $\mathcal{U} = \{1, \ldots, U\}$, $\mathcal{M} = \{1, \ldots, M\}$, $\mathcal{U} = \{1, \ldots, U\}$, and $\mathcal{K} = \{1, \ldots, K\}$ denote the set of IoT devices, malicious nodes, UAVs and LEOSats, respectively. In this system, the HAPS is assumed to be equipped with N_A^R receiving antennas to communicate with the UAVs and N_A^T transmit antennas to communicate with the LEOSats. Likewise, the UAVs are equipped with N_U^R receiving antennas to communicate with the IoT devices and N_U^T transmitting antennas to communicate with the HAPS. In particular, we assume $N_U^T = N_U^R = N_U$. Moreover, both the legitimate and malicious nodes are assumed to have a single antenna.

A. Authentication Framework

1) Transmission Phase: Let $\mathbf{s}_i = [s_{i,1}, \ldots, s_{i,L}]^T \in \mathbb{C}^{L\times 1}$ denote the message signal from the ith IoT device in a particular transmission block with length L. An authentication tag $\mathbf{t}_i = [t_{i,1}, \ldots, t_{i,L}] \in \mathbb{C}^{L\times 1}$, which depends on $\mathbf{s}_i$ and $\mathbf{l}_i$, is created as

$$\mathbf{t}_i = g(\mathbf{s}_i, \mathbf{l}_i), \tag{1}$$

where $g(\cdot)$ is a secure hash function [11]. Here, one assumes that $\mathbb{E}\{\|\mathbf{s}_i\|^2\} = \mathbb{E}\{\|\mathbf{t}_i\|^2\} = L$ and $\mathbb{E}\{s_{i,l}\} = \mathbb{E}\{t_{i,l}\} = 0, \forall i \in \mathcal{I}, \forall l \in \{1, \ldots, L\}$. In addition, the message signals are not correlated to the tags [12], i.e., $\mathbb{E}\{\mathbf{t}_i^H \mathbf{s}_i\} = 0$, and since the secret keys are unique to each device, the authentication tags of different devices are independent, i.e., $\mathbb{E}\{\mathbf{t}_i^H \mathbf{t}_{i'}\} = 0, \forall i' \neq i, i' \in \mathcal{I}$. Then, the tagged signal transmitted by the ith IoT device is written as

$$\mathbf{x}_i = \rho_{\mathrm{s},i}\mathbf{s}_i + \rho_{\mathrm{t},i}\mathbf{t}_i, \tag{2}$$

where $\rho_{\mathrm{s},i}$ and $\rho_{\mathrm{t},i}$ are the power allocation coefficients of the message and of the tag signal of device i, respectively, which should satisfy $\rho_{\mathrm{t},i} \ll \rho_{\mathrm{s},i}$ and $\mathbb{E}\{\|\mathbf{x}_i\|^2\} \leq L$, and hence, $\rho_{\mathrm{s},i}^2 + \rho_{\mathrm{t},i}^2 \leq 1$ [6]. Consequently, if $\rho_{\mathrm{s},i} = 1$ (then $\rho_{\mathrm{t},i} = 0$), it indicates that device i is not transmitting the authentication tag during the transmission block.

Each IoT device transmits its data to the UAVs over orthogonal subchannels with bandwidth B_i. Thus, the $N_U \times L$ received signal at UAV u from device i is written as

$$\mathbf{Y}_u = \sqrt{p_i}\mathbf{g}_{i,u}\mathbf{x}_i^H + \mathbf{\Omega}_u, \tag{3}$$

where p_i is the maximum normalized transmit power of device i, $\mathbf{\Omega}_u \in \mathcal{C}^{N_u \times L}$ is the noise matrix at UAV u, which is assumed to have independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ entries. Also, $\mathbf{g}_{i,u}=\sqrt{\beta_{i,u}}\mathbf{h}_{i,u}$ is the channel vector from device i to UAV u, where $\beta_{i,u}$ is the air-to-ground large-scale fading, while $\mathbf{h}_{i,u}$ is the small-scale Ricean fading component [13].

Then, the UAVs relay their received signal to the HAPS in a half-duplex AF manner to reduce onboard signal processing and energy consumption at the UAVs. Thus, the aggregated received signal at the HAPS is given by

$$\mathbf{Y}_a = \sum_{u\in\mathcal{U}} \sqrt{q_{u,i}}\mathbf{G}_{u,a}^H \mathbf{Y}_u + \mathbf{\Omega}_a, \tag{4}$$

where $\mathbf{\Omega}_a \in \mathcal{C}^{N_A^R \times L}$ is the noise matrix at the HAPS, modeled as i.i.d. with $\mathcal{CN}(0,1)$ entries, $\mathbf{G}_{u,a}$ is the $\mathbb{C}^{N_u \times N_A^R}$ channel matrix between UAV u and the HAPS, modeled as

$$\mathbf{G}_{u,a}=\sqrt{\beta_{u,a}}e^{-j\frac{2\pi}{\lambda}d_{u,a}}\mathbf{a}_a(\vartheta_{u,a}^{\mathrm{AoA}},\varphi_{u,a}^{\mathrm{AoA}})\mathbf{a}_u^H(\vartheta_{u,a}^{\mathrm{AoD}},\varphi_{u,a}^{\mathrm{AoD}}), \tag{5}$$

where $\beta_{u,a}$, λ and $d_{u,a}$ are the large-scale fading, the wavelength, and the distance between UAV u and the HAPS, respectively. In addition, $\vartheta_{u,a}^{\mathrm{AoA}}$ and $\varphi_{u,a}^{\mathrm{AoA}}$ denotes the azimuth and elevation angles of arrival at the HAPS from UAV u, and $\vartheta_{u,a}^{\mathrm{AoD}}$ and $\varphi_{u,a}^{\mathrm{AoD}}$ are the azimuth and elevation angles of departure from UAV u to the HAPS. Finally, $q_{u,i}$ is the amplification factor at UAV u for device i, which can be obtained from the following transmit power constraint

$$\mathbb{E}\{\|\sqrt{q_{u,i}}\mathbf{Y}_u\|_F^2\} \le Lp_u,$$
$$q_{u,i} = \frac{p_u}{N_U(p_i\beta_{i,u}+1)}, \forall i \in \mathcal{I}, \forall u \in \mathcal{U}, \tag{6}$$

where p_u is UAV u's maximum normalized transmit power.

2) Authentication Phase: Let us assume that the HAPS has perfect knowledge of the weighted effective channel sum between the UAVs and device i, $\mathbf{g}_{i,a} \triangleq \sum_{u\in\mathcal{U}} \sqrt{q_{u,i}p_i}\mathbf{G}_{u,a}^H\mathbf{g}_{i,u}$, and employs this knowledge to equalize the received signal and detect $\mathbf{x}_i$. Specifically, the estimated $\hat{\mathbf{x}}_i$ at the HAPS is given by

$$\hat{\mathbf{x}}_i = \frac{\mathbf{g}_{i,a}^H}{\|\mathbf{g}_{i,a}\|^2}\left(\sum_{u\in\mathcal{U}}\sqrt{q_{u,i}p_i}\mathbf{G}_{u,a}^H\mathbf{g}_{i,u}\mathbf{x}_i^H + \bar{\mathbf{\Omega}}_a\right), \tag{7}$$

where $\bar{\mathbf{\Omega}}_a \triangleq \sum_{u\in\mathcal{U}}\sqrt{q_{u,i}}\mathbf{G}_{u,a}^H\mathbf{\Omega}_u + \mathbf{\Omega}_a$. Given $\rho_{\mathrm{t},i} \ll \rho_{\mathrm{s},i}$, the presence of the authentication tag in $\hat{\mathbf{x}}_i$ can be ignored, and the estimation of the message signal from device i, $\hat{\mathbf{s}}_i$, can be obtained with a tolerable bit error rate (BER) [14].

Then, based on $\hat{\mathbf{s}}_i$ and $\mathbf{l}_i$, the HAPS generates an expected authentication tag with the same secure hash function as the one employed by device i as $\tilde{\mathbf{t}}_i = g(\hat{\mathbf{s}}_i, \mathbf{l}_i)$. The HAPS also recovers the received tag from the residual signal as

$$\mathbf{r}_i = (\hat{\mathbf{x}}_i - \rho_{\mathrm{s},i}\hat{\mathbf{s}}_i)/\rho_{\mathrm{t},i}. \tag{8}$$

To verify whether the estimated tag of device i is present in the received signal, a hypothesis test is conducted as

$$\begin{cases} \mathcal{H}_0 : \tilde{\mathbf{t}}_i \text{ is not present in } \mathbf{r}_i, \\ \mathcal{H}_1 : \tilde{\mathbf{t}}_i \text{ is present in } \mathbf{r}_i, \end{cases} \tag{9}$$

where $\mathcal{H}_1$ indicates the presence of the correct tag in the received signal, whereas $\mathcal{H}_0$ shows that $\mathbf{r}_i$ does not contain the correct tag. Hence, the threshold test for each legitimate device is written as

$$\lambda_i = \mathbb{R}\{\tilde{\mathbf{t}}_i^H\mathbf{r}_i\} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \theta_i, \forall i \in \mathcal{I}, \tag{10}$$

where θ_i is the decision threshold of device i and λ_i is the test statistic, which is constructed by match-filtering $\mathbf{r}_i$ with $\tilde{\mathbf{t}}_i$, that is, $\lambda_i = \mathbb{R}\{\tilde{\mathbf{t}}_i^H\mathbf{r}_i\}$. Given the nature of the hash function, if $\mathbf{s}_i$ is recovered with errors, then $\tilde{\mathbf{t}}_i \neq \mathbf{t}_i$ with high probability. Thus, we consider that $\hat{\mathbf{s}}_i = \mathbf{s}_i$ and $\tilde{\mathbf{t}}_i = \mathbf{t}_i$, and derive the expressions for the test statistics when device i transmits only the message signal ($\lambda_i|\mathcal{H}_0$), and when a tagged signal is received from device i ($\lambda_i|\mathcal{H}_1$) as follows

$$\lambda_i|\mathcal{H}_0 = \mathbf{t}_i^H\mathbf{r}_i = \mathbf{t}_k^H\mathbf{n}_i/\rho_{\mathrm{t},i}, \tag{11}$$
$$\lambda_i|\mathcal{H}_1 = \|\mathbf{t}_i\|^2 + \mathbf{t}_k^H\mathbf{n}_i/\rho_{\mathrm{t},i}, \tag{12}$$

where $\mathbf{n}_i \triangleq (\mathbf{g}_{i,a}^H\bar{\mathbf{\Omega}}_a)/\|\mathbf{g}_{i,a}\|^2$.

According to the hypotheses, if the HAPS accepts $\mathcal{H}_1$ when $\mathcal{H}_0$ is true, a false alarm will be raised. On the other hand, if $\mathcal{H}_1$ is accepted when $\mathcal{H}_1$ is true, a correct detection has occurred. In the following proposition, we derive the expressions for the PFA and the PD as well as the optimal decision threshold for each device i, θ_i^*, to maximize the corresponding PD for a given PFA rate, $\mathsf{P}_{\mathrm{FA},i}$.

Proposition 1. *The closed-form expressions for the PFA, the PD, and the optimal decision threshold of the ith IoT device are given, respectively, by*

$$\mathrm{P}_{\mathrm{FA},i} = Q\left(\frac{\theta_i}{\sqrt{L\sigma_{n_i}^2/(2\rho_{t,i}^2)}}\right), \tag{13}$$

$$\mathrm{P}_{\mathrm{D},i} = Q\left(\frac{\theta_i - L}{\sqrt{L\sigma_{n_i}^2/(2\rho_{t,i}^2)}}\right), \tag{14}$$

$$\theta_i^* = Q^{-1}(\mathsf{P}_{\mathrm{FA},i})\sqrt{L\sigma_{n_i}^2/(2\rho_{t,i}^2)}, \tag{15}$$

where $Q(\cdot)$ represents the Q-function, and $\sigma_{n_i}^2$ denotes the per-symbol variance of $\mathbf{n}_i$.

Proof: The proof follows the methodologies in [15, Proposition 1], and the details are omitted due to page constraints. ∎

B. Joint Offloading and Resource Allocation

The requested computing tasks from device i is embedded onto its message signal $\mathbf{s}_i$, and is denoted by

$$\boldsymbol{\psi}_i \triangleq [\delta_i, c_i, \tau_i^{\max}]^T, \forall i \in \mathcal{I}, \tag{16}$$

where δ_i is the bit number, c_i is the computing density, and $\tau_i^{\max}$ is the maximum tolerable delay. Specifically, we consider that the total delay experienced by task $\boldsymbol{\psi}_i$ is given by

$$\tau_i = \tau_{i,a} + \sum_{k\in\mathcal{K}} b_i^k\left(\tau_{a,k} + \tau_k^{\mathrm{prop}} + \tau_i^k\right), \tag{17}$$

where b_i^k is a binary variable that indicates if task i is offloaded to LEOSat k ($b_i^k = 1$) or not ($b_i^k = 0$). Moreover, $\tau_{i,a} \triangleq \delta_i / R_{i,a}$ and $\tau_{a,k} \triangleq \sum_{i \in \mathcal{I}} b_i^k \delta_i / R_{a,k}$ are, respectively, the transmission delay between device i and the HAPS and between the HAPS and LEOSat k, in which $R_{i,a}$ and $R_{a,k}$ are the corresponding communication rates, respectively, i.e.,

$$R_{i,a} = B_i \log_2 \left(1 + \frac{\rho_{\mathrm{s},i}^2 p_i}{\rho_{\mathrm{t},i}^2 p_i + \sigma_{n_i}^2}\right), \tag{18}$$

$$R_{a,k} = B_a \log_2 \left(1 + \rho_{a,k} \frac{\omega_{a,k}^2}{\sigma_k^2}\right), \tag{19}$$

where B_a is the bandwidth of the HAPS-LEO link, $\rho_{a,k}$ is the transmit power of the HAPS for the corresponding link, σ_k^2 is the noise power at the LEO and $\omega_{a,k}$ is the non-zero singular value of the rank-deficient line-of-sight (LoS) MIMO channel matrix between a and k that reflects the array gain, the path-loss and the atmospheric losses [16]. Moreover, $\tau_k^{\mathrm{prop}} \triangleq \frac{1}{c} d_{a,k}$ is the propagation delay between the HAPS and LEOSat k, where c is the speed of light, and $d_{a,k}$ is the Euclidean distance between the HAPS and the LEOSat k. Finally, $\tau_i^k \triangleq \delta_i c_i / f_i^k$ is the computing delay of the task from device i at LEOSat k, where f_i^k are the resources allocated by LEOSat k to compute the task offloaded by device i. We assume a regular offloading pattern in time, for which the optimization is performed once, thus, we do not consider the algorithm runtime delay as an additional source of delay.

C. Problem Formulation

This work aims to minimize the maximum weighted delay across tasks by selecting the optimized MEC server and computing the optimized share of computing resources for the tasks. Reasonably, the resource allocation should be intended only for legitimate device tasks. Thus, let us define an admission binary variable α_i, that indicates whether the task of device i is admitted as legitimate ($\alpha_i = 1$) or rejected ($\alpha_i = 0$) based on the PLA-based scheme presented in Sec. II-A. Specifically, $\alpha_i = 1$ if the test statistic λ_i of device i surpasses the optimized detection threshold, θ_i^*. After computing the respective α_i for all $i \in \mathcal{I} \cup \mathcal{M}$, the joint offloading and resource allocation optimization problem can be formulated as

$$\begin{aligned}
(\mathcal{P}): \ \min_{\mathbf{F},\mathbf{B}} \quad & \max_{i \in \mathcal{I} \cup \mathcal{M}} \alpha_i \frac{\tau_i}{\tau_i^{\max}} && \text{(20a)}\\
\text{s.t.} \quad & \alpha_i \tau_i \le \tau_i^{\max}, \ \ \forall i \in \mathcal{I} \cup \mathcal{M} && \text{(20b)}\\
& \sum_{i \in \mathcal{I} \cup \mathcal{M}} f_i^k \le F_k^{\max}, \ \ \forall k \in \mathcal{K}, && \text{(20c)}\\
& \sum_{k \in \mathcal{K}} b_i^k = \alpha_i, \ \ \forall i \in \mathcal{I} \cup \mathcal{M}, && \text{(20d)}\\
& b_i^k \in \{0,1\}, \ \ \forall i \in \mathcal{I} \cup \mathcal{M}, \ \forall k \in \mathcal{K}, && \text{(20e)}\\
& f_i^k \ge 0, \ \ \forall i \in \mathcal{I} \cup \mathcal{M}, k \in \mathcal{K}, && \text{(20f)}
\end{aligned}$$

where $\mathbf{F} \in \mathbb{R}_+^{I \times K}$ and $\mathbf{B} \in \{0,1\}^{I \times K}$ gather the computing resources and offloading decision variables, respectively. Also, $F_k^{\max}$ is LEOSat k's maximum computing resource.

III. Proposed Solution

To solve the joint authentication, offloading, and resource allocation problem $(\mathcal{P})$ for the admitted users, we start by rewriting it in epigraph form as

$$\begin{aligned}
(\mathcal{P}'): \ \min_{\mathbf{F},\mathbf{B},\mu} \quad & \mu && \text{(21a)}\\
\text{s.t.} \quad & \alpha_i \tau_i \le \mu \tau_i^{\max}, \ \ \forall i \in \mathcal{I} \cup \mathcal{M}, && \text{(21b)}\\
& 0 \le \mu \le 1, && \text{(21c)}\\
& \text{(20c), (20d), (20e), (20f).}
\end{aligned}$$

Given the non-convex nature of $(\mathcal{P}')$, we propose a multi-layer multi-stage optimization framework. An iterative block-coordinate descent algorithm is performed over the offloading decision variables and the computing resources to find the solution to $(\mathcal{P}')$, as described in the following.

A. Offloading Decision

For the admitted tasks $i \in \mathcal{I}$, and given the set of computing resources $\mathbf{F}$, the following optimization problem is solved to obtain the optimized offloading decisions

$$\begin{aligned}
(\mathcal{P}^o): \ \min_{\mathbf{B},\mu} \quad & \mu && \text{(22a)}\\
\text{s.t.} \quad & \sum_{k \in \mathcal{K}} b_i^k = 1, \quad \forall i \in \mathcal{I}, && \text{(22b)}\\
& b_i^k \in \{0,1\}, \ \ \forall i \in \mathcal{I}, \ \forall k \in \mathcal{K}, && \text{(22c)}\\
& \text{(21b), (21c).}
\end{aligned}$$

Problem $(\mathcal{P}^o)$ is not convex due to constraints (21b) and (22c). To deal with these, we first relax constraint (22c) to allow the decision variables b_i^k to be continuous as $0 \le b_i^k \le 1$, and solve the problem iteratively with proximity constraints, such as $b_i^{k,(n)} - \varrho \le b_i^k \le b_i^{k,(n)} + \varrho$, where ϱ is chosen appropriately for convergence and the $^{(n)}$ suffix indicates that the corresponding term is the value from the previous iteration.

Moreover, a quadratic penalty function $\phi(b_i^k) = (1 - b_i^k) b_i^k$ is included in the objective function to drive the solution to binary values. However, given that this function is not convex, we linearize it via a first-order Taylor approximation as $\hat{\phi}(b_i^k) = b_i^k (1 - 2 b_i^{k,(n)})$ and define the binary penalty function as $\Phi(\mathbf{B}) = \chi \sum_{k \in \mathcal{K}, i \in \mathcal{I}} \hat{\phi}(b_i^k)$, where χ is increasingly set at each iteration. Then, to handle constraint (21b), we start by rewriting the non-convex quadratic term in (17) as

$$\frac{b_i^k \sum_{j \ne i} \delta_j b_j^k}{R_{a,k}} = \frac{\left(\sum_{j \ne i} \delta_j b_j^k + b_i^k\right)^2 - \left(\sum_{j \ne i} \delta_j b_j^k - b_i^k\right)^2}{4 R_{a,k}}. \tag{23}$$

Then, we apply the quadratic transform (QT) [17] to each quadratic term in (23), such that

$$\frac{\left(\sum_{j \ne i} \delta_j b_j^k \pm b_i^k\right)^2}{R_{a,k}} \Rightarrow 2 z_{\pm,i} \left|\sum_{j \ne i} \delta_j b_j^k \pm b_i^k\right| - z_{\pm,i}^2 R_{a,k}, \tag{24}$$

$$z_{\pm,i} = \left|\sum_{j \ne i} \delta_j b_j^{k,(n)} \pm b_i^{k,(n)}\right| / R_{a,k}^{(n)}. \tag{25}$$

By enforcing $\sum_{j\neq i}\delta_j b_j^k \geq \delta_i$ for all $i \in \mathcal{I}$, we can remove the absolute value from (24) and (25), and rewrite (17) as

$$\tau_i = \zeta_i + \sum_{k\in\mathcal{K}}\sum_{j\neq i} b_j^k \zeta_{i,j}^k + \sum_{k\in\mathcal{K}} b_i^k \zeta_i^k, \tag{26}$$

where

$$\zeta_i \triangleq \frac{\delta_i}{R_{i,a}} - \sum_{k\in\mathcal{K}}\sum_{j\neq i}\delta_j b_i^{k,(n)} b_j^{k,(n)}/R_{a,k}, \tag{27}$$

$$\zeta_{i,j}^k \triangleq \delta_j b_i^{k,(n)}/R_{a,k}, \tag{28}$$

$$\zeta_i^k \triangleq \delta_i c_i/f_i^k + \delta_i/R_{a,k} + \sum_{j\neq i}\delta_j b_j^{k,(n)}/R_{a,k} + 2\tau_k^{\text{prop}}. \tag{29}$$

Then, let $\boldsymbol{\tau}^{\max} = [\tau_1^{\max},\ldots,\tau_I^{\max}]^T$, $\mathbf{b}^k = [b_1^k,\ldots,b_I^k]^T$, $\mathbf{b} = [\left(\mathbf{b}^1\right)^T,\ldots,\left(\mathbf{b}^K\right)^T,\mu]^T$, $\mathbf{d} = [\delta_1,\ldots,\delta_I]^T$, $\mathbf{b}_L = \mathbf{b}^{(n)} + \varrho$, $\mathbf{b}_H = -\mathbf{b}^{(n)} + \varrho$, $\mathbf{B}_I = \text{diag}\,(\mathbf{d}) - \mathbf{d}\otimes\mathbf{1}_I^T + \mathbf{I}_I$, $\mathbf{B} = [\text{blkdiag}(\mathbf{B}_I)_K, \mathbf{0}_{IK}]_{IK\times IK+1}$, $\mathbf{C}_L = [\text{blkdiag}(\mathbf{I}_I)_K, \mathbf{0}_{IK}]_{IK\times IK+1}$, $\mathbf{C}_U = -\mathbf{C}_L$, $\boldsymbol{\zeta} = [\zeta_1,\ldots,\zeta_I]^T$ and $\mathbf{A} = [\mathbf{A}_1,\ldots,\mathbf{A}_I,\boldsymbol{\tau}^{\max}]_{I\times IK+1}$, with

$$\mathbf{A}_k = \begin{bmatrix}\zeta_{1,1}^k & \cdots & \zeta_{1,I}^k \\ \cdots & \cdots & \cdots \\ \zeta_{I,1}^k & \cdots & \zeta_{I,I}^k\end{bmatrix}, \forall k \in \mathcal{K}. \tag{30}$$

Finally, by defining $\boldsymbol{\Upsilon} = [\mathbf{A}^T,\mathbf{B}^T,\mathbf{C}_L^T,\mathbf{C}_U^T,\mathbf{C}_L^T,\mathbf{C}_U^T]_{5IK+I\times IK+1}^T$, $\mathbf{c} = [\boldsymbol{\zeta}^T,\mathbf{0}_{IK}^T,\bar{\mathbf{b}}_L^T,\bar{\mathbf{b}}_U^T,\mathbf{1}_{IK}^T,\mathbf{0}_{IK}^T]_{5IK+I\times 1}^T$ and $\boldsymbol{\upsilon} = [\chi\left(\mathbf{1}_{IK} - 2\bar{\mathbf{b}}^{(n)}\right)^T, 1]_{IK+1\times 1}^T$, where $\bar{\mathbf{b}}^{(n)}$, $\bar{\mathbf{b}}_U$ and $\bar{\mathbf{b}}_L$ correspond to $\mathbf{b}^{(n)}$, $\mathbf{b}_U$ and $\mathbf{b}_L$ without their last element, the problem $(\mathcal{P}^o)$ can be relaxed to a linear program (LP) written as

$$(\mathcal{P}_{\text{LP}}^o): \quad \min_{\mathbf{b}} \boldsymbol{\upsilon}^T\mathbf{b} \text{ s.t. } \boldsymbol{\Upsilon}\mathbf{b} \leq \mathbf{c}, \tag{31}$$

which can be readily solved with the convex programming toolbox CVX. If $(\mathcal{P}_L^o)$ is not feasible due to maximum delay constraints of the tasks, we reject the task with the largest pull for resources from the system ($i^* = \text{argmax}_{i\in\mathcal{I}}\{\frac{\delta_i c_i}{\tau_i^{\max}}\}$) until the problem becomes feasible for the remaining tasks.

B. Remote Computing Resource Allocation

The optimization problem to obtain the optimized computing resources at every LEOSat can be written as

$$(\mathcal{P}^f): \quad \min_{\mathbf{F},\mu} \quad \mu \tag{32a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} b_i^k \delta_i c_i/f_i^k \leq \mu\tau_i^{\max} - \varepsilon_i, \quad \forall i \in \mathcal{I} \tag{32b}$$

$$(21\text{c}), (20\text{c}), (20\text{f}),$$

where $\varepsilon_i = \delta_i/R_{i,a} + \sum_{k\in\mathcal{K}} b_i^k\left(\sum_{j\in\mathcal{I}} b_j^k\delta_j/R_{a,k} + 2\tau_k^{\text{prop}}\right)$.

This problem is convex and can be solved with dedicated software. Accordingly, problems $(\mathcal{P}^o)$ and $(\mathcal{P}^f)$ are solved iteratively in a block coordinate descent framework, with increasing binary penalty χ until convergence. A description of the joint authentication, offloading and resource allocation framework can be seen in **Algorithm 1**.

Algorithm 1: JOINT AUTHENTICATION, OFFLOADING, AND RESOURCE ALLOCATION OPTIMIZATION

```
foreach j ∈ I ∪ M do
    Compute x̂_j as in (7) and r_j as in (8);
    Compute optimal decision threshold θ_j^* as in (15);
    Obtain test statistics λ_j as in (10);
    Admit or reject j as per (10)
Set n_it := 1 and initialize B^(1) and F^(1) for admitted tasks;
repeat
    Compute Υ, υ and c;
    Solve P_LP^o and obtain B;
    if P_LP^o is unfeasible then
        Reject task i* = argmax_i{δ_i c_i / τ_i^max} ;
        Go back to 8;
    Solve P^f and obtain F;
    If ||B − B^(n_it)||_F ≤ ε/(IK), return;
    Adjust χ and ϱ;
    Set B^(n_it+1) = B and F^(n_it+1) = F;
    Set n_it := n_it + 1;
until n_it ≥ N_it;
```

IV. PERFORMANCE EVALUATION

To evaluate the performance of our PLA-based scheme we consider a system with $I = 50$ and $M = 25$, spread over a 10×10 km^2 area following a binomial point process. Both legitimate and malicious nodes generate tasks with a bit number of 10 Kb and computing density of 200 CPU cycles per bit. The computing capacity at the LEOSats is set as $F_k^{\max} = 10$ Gcycles/s. As per the 3GPP recommendations [18], the carrier frequency for the HAPS-LEO links is chosen in the Ka band, as 28 GHz, with a bandwidth $B_a = 100$ MHz. For the IoT-UAV-HAPS access, we consider NB-IoT-compliant subchannels of bandwidth $B_i = 200$ KHz over the carrier at 2.1 GHz. The altitude of the UAVs and the HAPS is set as 120 m and 20 Km, respectively. Moreover, all UAVs are considered to be uniformly spread in a grid over the area. A set of $K = 3$ LEOSats is considered within a constellation above the HAPS, with the satellites distributed at an Earth-central angular separation of 5 degrees. The constellation is confined within a maximum Earth-central angle of 20 degrees, guaranteeing visibility within the HAPS horizon. Furthermore, unless stated otherwise, $U = 25$, $\rho_{\text{t},i}^2 = 0.01$, $\mathsf{P}_{\text{FA}} = 10^{-6}$, $N_u = 64$, and $N_A^R = N_A^T = 256$, $p_i = 20$ dBm $\forall i \in \mathcal{I}$, $\rho_u = 30$ dBm $\forall u \in \mathcal{U}$, and $\rho_{a,k} = 33$ dBm $\forall k \in \mathcal{K}$. Moreover, ϱ is chosen such that $\varrho = 0.01$ for the first iteration, and is uniformly increased to $\varrho = 0.5$ for the last iteration.

In Fig. 2 we show the proportion of feasible tasks achieved under different authentication schemes as a function of the maximum delay constraint $\tau_i^{\max}$. Specially, we consider the following benchmarks: the bilinear pairing-based public-key (BP-PK) scheme from [19], the ID-based blockchain (ID-BC) scheme from [20] and the lightweight MSR-based blockchain

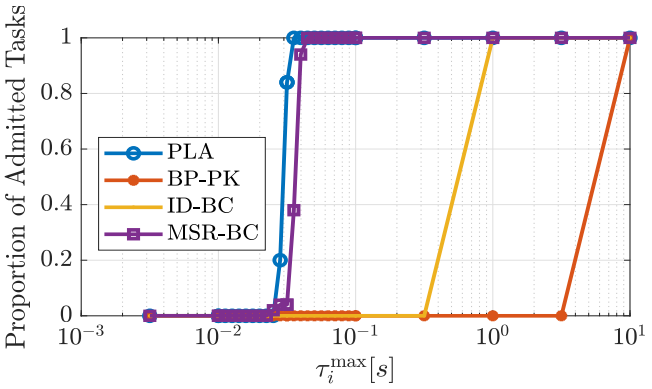

Fig. 2. Proportion of feasible tasks for varying $\tau_i^{\max}$.

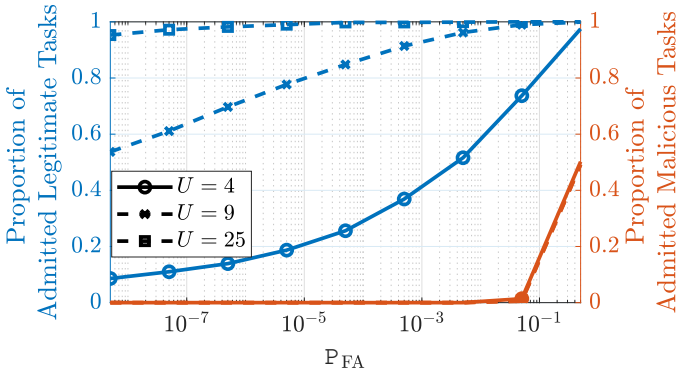

Fig. 3. Proportion of legitimate and malicious tasks admitted due to authentication versus the fixed PFA rate, $\mathrm{P_{FA}}$, for different number of UAVs.

(MSR-BC) scheme from [21]. The computation times from these schemes are taken from [21]. Given that the operations for the proposed PLA scheme are performed in the physical layer with minimal signaling, it can be seen that the PLA-based results achieve task feasibility much earlier than all the other considered authentication schemes. This occurs because the time incurred by the authentication schemes entails an extra source of delay that further constraints the tasks to be computed within the maximum time $\tau_i^{\max}$. It is worthwhile to mention that although the performance of the MSR-BC scheme approximates to the one observed for the PLA scheme, it requires an additional blockchain infrastructure, which adds significant complexity to the system.

Figure 3 illustrates the proportion of legitimate and malicious tasks admitted during the authentication process as a function of the allowable PFA, $\mathrm{P_{FA}}$, under varying numbers of UAVs. As expected, increasing $\mathrm{P_{FA}}$ leads to a higher admission rate for legitimate tasks, due to the corresponding decrease in the optimal detection threshold θ_i^* for all $i \in \mathcal{I}$. However, this relaxation also slightly raises the acceptance rate of malicious tasks. Notably, even at $\mathrm{P_{FA}} = 0.05$, the proportion of admitted malicious tasks remains as low as 1%. Furthermore, increasing the number of UAVs significantly enhances the detection of legitimate tasks: with $U = 25$, the admission rate for legitimate tasks exceeds 95% even under a

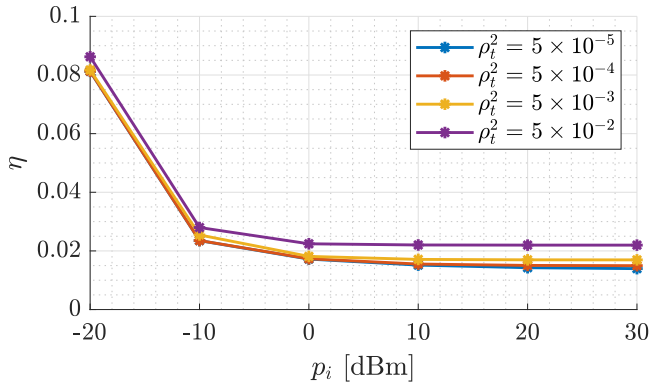

Fig. 4. Maximum relative delay, η, with varying IoT transmit power, p_i. for different tag allocation power ,$\rho_{\mathrm{t},i}^2$.

stringent false alarm constraint.

In Fig. 4, we illustrate the maximum relative delay η versus the transmit power of devices, p_i, for different values of the tag allocation power, $\rho_{\mathrm{t},i}^2$. Note that by increasing p_i, the IoT device-HAPS communication rates also increase, thereby resulting in a lower offloading delay and consequently lower values of η. However, for $p_i \geq 0$ dBm, there is no significant impact on η. This occurs because as p_i increases above this value, the contribution of $\tau_{i,a}$ to the total task delay is dominated by the HAPS–LEO transmission latency and computing delays. Furthermore, decreasing $\rho_{\mathrm{t},i}^2$ has minimal effect on η, thereby enabling a greater portion of power to be allocated to message signal detection.

V. Conclusions

In this work, we developed a tag-based PLA framework designed to authenticate IoT devices offloading tasks to a multi-layer NTN for remote processing. We derived the expressions for the PFA and PD for each legitimate user and defined the optimal decision threshold, which was used to admit the tasks for offloading. A joint task offloading and resource allocation optimization problem was formulated for the admitted tasks and solved via block coordinate descent. Our results showed that the PLA scheme is efficient and allows for better task feasibility under stringent task requirements than the compared benchmark schemes. They also showed that increasing the number of UAVs leads to the reliable admission of legitimate tasks even under strict false-alarm requirements.

A. Future Works

Potential extensions could include adapting the power allocation to device-specific requirements. Furthermore, it is worth looking into more practical scenarios with imperfect knowledge of the channel state information and imperfect message detection.

Acknowledgement

This work is funded by the Horizon Europe under the Marie Sklodowska Curie actions: ELIXIRION (GA 101120135) and by the U.K. Engineering and Physical Sciences Research

Council (EPSRC) grant (EP/X04047X/2) for TITAN Telecoms Hub. The work of H. Q. Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/X010635/1. The work of M. Matthaiou was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101001331). The work of I. W. G. da Silva, H. Q. Ngo and M. Matthaiou was also supported by a research grant from the Department for the Economy Northern Ireland under the US-Ireland R&D Partnership Programme.

REFERENCES

[1] K. Ntontin *et al.*, "A vision, survey, and roadmap toward space communications in the 6G and beyond era," *Proc. IEEE*, vol. 113, no. 9, pp. 987–1023, Sep. 2025.

[2] M. M. Saad, M. A. Tariq, M. T. R. Khan, and D. Kim, "Non-terrestrial networks: An overview of 3GPP release 17 & 18," *IEEE Internet Things Mag.*, vol. 7, no. 1, pp. 20–26, Jan. 2024.

[3] Z. Tang, K. Yu, G. Yang, L. X. Cai, and H. Zhou, "New bridge to cloud: An ultra-dense LEO assisted green computation offloading approach," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 552–564, Sept. 2023.

[4] A. Flores, K. Ntontin, A. Bandi, V. N. Ha, and S. Chatzinotas, "Low-complexity resource allocation for task offloading in hierarchical non-terrestrial networks," *IEEE Internet of Things Journal*, pp. 1–1, 2026.

[5] Y. Lee, J. Yoon, J. Choi, and E. Hwang, "A novel cross-layer authentication protocol for the Internet of things," *IEEE Access*, vol. 8, pp. 196 135–196 150, Oct. 2020.

[6] N. Xie, C. Chen, and Z. Ming, "Security model of authentication at the physical layer and performance analysis over fading channels," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 1, pp. 253–268, Jan. 2021.

[7] N. H. S. Suhaimi, N. H. Kamarudin, M. N. A. Khalid, I. Tahir, and M. A. A. Mohamed, "State-of-the-art authentication measures in satellite communication networks: A comprehensive analysis," *IEEE Access*, vol. 12, pp. 142 241–142 264, 2024.

[8] M. Abdrabou and T. A. Gulliver, "Physical layer authentication for satellite communication systems using machine learning," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 2380–2389, Dec. 2022.

[9] ——, "Game theoretic spoofing detection for space information networks using physical attributes," *IEEE Trans. Commun.*, vol. 72, no. 7, pp. 3947–3956, Feb. 2024.

[10] P. Zhang, Y. Teng, Y. Shen, X. Jiang, and F. Xiao, "Tag-based PHY-layer authentication for RIS-assisted communication systems," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 6, pp. 4778–4792, Nov. 2023.

[11] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. CRC press, 2018.

[12] P. L. Yu, J. S. Baras, and B. M. Sadler, "Physical-layer authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 1, pp. 38–51, Mar. 2008.

[13] M. M. Azari, F. Rosas, K.-C. Chen, and S. Pollin, "Ultra reliable uav communication using altitude and cooperation diversity," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 330–344, Aug. 2018.

[14] P. L. Yu, G. Verma, and B. M. Sadler, "Wireless physical layer authentication via fingerprint embedding," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 48–53, Jun. 2015.

[15] I. W. da Silva, Z. Mobini, H. Q. Ngo, and M. Matthaiou, "Cell-free massive MIMO-based physical-layer authentication," *arXiv preprint arXiv:2508.19931*, 2025.

[16] ITU-R, "Propagation data and prediction methods required for the design of earth-space telecommunication systems," ITU-R, Tech. Rep. ITU-R P.618-14 (08/2023), 2023.

[17] K. Shen and W. Yu, "Fractional programming for communication systems—Part i: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, Mar. 2018.

[18] 3GPP, "Satellite access radio frequency (rf) and performance requirements (release 18)," 3GPP, Tech. Rep. TS 38.101-5 V18.7.0 (2024-09), 2024.

[19] J. Ni, X. Lin, and X. S. Shen, "Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 644–657, Mar. 2018.

[20] M. Shen *et al.*, "Blockchain-assisted secure device authentication for cross-domain industrial IoT," *IEEE J. Selected Areas Commun.*, vol. 38, no. 5, pp. 942–954, Mar. 2020.

[21] X. Yang *et al.*, "Blockchain-based secure and lightweight authentication for Internet of things," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3321–3332, Jul. 2022.