Delay Performance of Distributed Physical Layer Authentication Under Sybil Attacks

Henrik Forssell, Ragnar Thobaben, James Gross
Div. of Information Science and Engineering, KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden
E-mail: {hefo, ragnart, jamesgr}@kth.se

***Abstract*—Physical layer authentication (PLA) has recently been discussed in the context of URLLC due to its low complexity and low overhead. Nevertheless, these schemes also introduce additional sources of error through missed detections and false alarms. The trade-offs of these characteristics are strongly dependent on the deployment scenario as well as the processing architecture. Thus, considering a feature-based PLA scheme utilizing channel-state information at multiple distributed radio-heads, we study these trade-offs analytically. We model and analyze different scenarios of centralized and decentralized decision-making and decoding, as well as the impacts of a single-antenna attacker launching a Sybil attack. Based on stochastic network calculus, we provide worst-case performance bounds on the system-level delay for the considered distributed scenarios under a Sybil attack. Results show that the arrival-rate capacity for a given latency deadline is increased for the distributed scenarios. For a clustered sensor deployment, we find that the distributed approach provides 23% higher capacity when compared to the centralized scenario.**

I. Introduction

Security in emerging ultra-reliable low-latency communications (URLLC), as envisioned for application in industrial automation or vehicle-to-vehicle communications, is very challenging since the room for security overhead (e.g., additional signaling for key agreement, encryption, or authentication) often is limited. A promising solution for combating impersonation-based attacks in such systems is to use feature-based physical layer authentication (PLA). These schemes, as opposed to crypto- and tag-based authentication protocols, are based on transmitter- or location-specific features at the PHY layer (e.g., frequency or impulse responses [1, 2], or multiple-antenna channels [3]), and do not require the extensive overhead that can be detrimental in low-latency applications.

Multiple-antenna transceivers along with distributed antenna architectures, as a way of exploiting the spatial diversity of the wireless channel, are considered potential enablers for the strict reliability requirements of URLLC [4, 5]. Interestingly, the diversity leveraged by distributed antenna architectures can also be exploited for enhanced feature-based PLA; with feature observations from multiple reception points the probability of an attacker successfully impersonating the legitimate feature is significantly reduced. However, for PLA in a distributed antenna setting, the question remains open whether PLA decision-making should rather be centralized (i.e., by combining the raw features from each reception point), or if each distributed receiver should perform independent decisions that are fused centrally. Moreover, while the reduced overhead from PLA schemes is a major benefit, in general PLA introduces the possibility of erroneous authentication decisions (i.e., false-alarms and missed detections), which potentially degrades performance. Altogether, these system-level costs of using PLA in a distributed antenna system need to be properly quantified and weighed against the detection performance benefits.

PLA has previously been proposed for security in URLLC [6, 7], vehicle-to-roadside communications [8], and for industrial internet-of-things [9]. Many of these previous works acknowledge that latency is critical in the considered systems; however, none of them particularly quantify the PLA-induced delay impacts. To address this open issue, in our previous work [10, 11] we have derived delay performance bounds for feature-based PLA for single- and multiple-antenna receivers. Moreover, for a distributed antenna architecture, we have provided security bounds (i.e., worst-case bounds on missed detection probability) in [12] and preliminary work on delay analysis in [13]. However, these works are lacking the delay performance analysis under different distributed decision-making scenarios (i.e., centralized vs. decentralized decision-making). Moreover, our previous works do not consider the delay impacts of active attack strategies against the distributed PLA schemes.

With these open issues in mind, this paper compares the delay-impacts of distributed PLA with varying degrees of distributed processing, ranging from completely centralized to distributed processing with hard decision fusing, so as to answer the question under which system configurations distributed PLA is a preferable option for a time-sensitive application. We compare the performance impacts under normal system operation as well as under a Sybil attack, i.e., when an attacker claims multiple forged identities to deplete network resources. The main contributions of this paper can be summarized as follows: (i) We provide a queueing model of distributed PLA that includes both centralized and decentralized PLA and decoding decisions. The modeled PLA scheme is based on the single-input multiple-output (SIMO) channel-states observed at multiple distributed radio-heads, but it could easily encompass other PLA schemes with closed-form receiver operating characteristics. (ii) We derive performance guarantees for the distributed PLA system in terms of bounds on the probability that the stochastic queueing delay violates

This work is supported in part by the Swedish Civil Contingencies Agency, MSB, through the CERCES project.

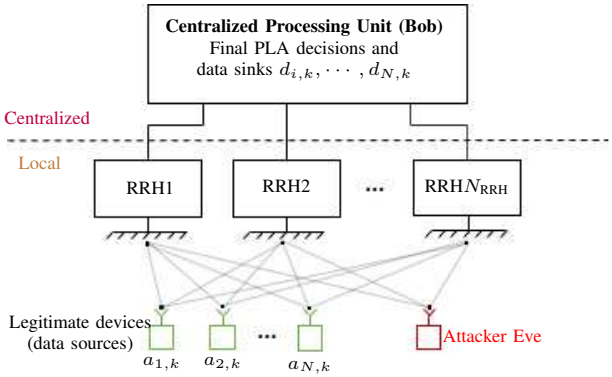

Fig. 1. Considered physical layer authentication system: Remote radio-heads connected to a centralized processing unit, multiple single-antenna transmit devices, and the attacker Eve.

a pre-defined deadline. The performance bounds are derived through stochastic network calculus [14]. (iii) We provide numerical evaluations of the derived bounds in an example factory automation scenario where we compare different deployment scenarios and choices of distributed processing. The results show benefits of the distributed PLA scheme under both centralized and decentralized decision-making when compared to the single-RRH case with the same number of antennas.

The rest of this paper is organized as follows: Section II presents the system model. Section III introduces the queueing model and the PLA scheme. In Section IV, we provide the main theoretical result which is the delay performance bound based on stochastic network calculus. Section V provides numerical evaluations and Section VI concludes the paper.

II. System Model

We consider a wireless system consisting of N single-antenna transmitters and N_{RRH} remote radio-heads (RRHs) with N_{Rx} antennas each. Each transmitting device has a local buffer of data that needs to be communicated to Bob reliably and with integrity. Moreover, we assume that there is a single-antenna attacker Eve that is trying to impersonate legitimate devices to harm the system operation. An illustration of the considered system setup can be seen in Fig. 1.

Medium Access Control: We assume a frame-based transmission protocol based on time-division multiple-access with frames consisting of a periodically transmitted broadcast beacon, followed by a management period (MGMT) and a data transmission period (DTP). In the MGMT period, devices can request resources for transmission of data payload. Moreover, we assume that devices can be inactive while being connected (i.e., connected to the network but idle waiting for data to transmit). We denote by $\mathcal{I}_{\text{DTP}}(k)$ the set of devices that are granted data resources in frame k and we assume that the network expects at most one request from each device. We assume the DTP has a fixed length of N_{Frame} complex symbols that are time-shared between the devices in the set $\mathcal{I}_{\text{DTP}}(k)$[1].

[1]We note that other resource allocation schemes like FDMA can also work in our model as long as channel realizations can be assumed independent.

An equal division of device resources is assumed, where the number of channel-uses each device gets is denoted by[2]

$$N_k = \left\lfloor \frac{N_{\text{Frame}}}{|\mathcal{I}_{\text{DTP}}(k)|} \right\rfloor, \tag{1}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than x.

Physical Layer: We let $\mathbf{Y}_k^{(j)} = [\mathbf{Y}_{i_1,k}^{(j)} \cdots \mathbf{Y}_{i_{[\mathcal{I}_{\text{DTP}}(k)]},k}^{(j)}]$ denote the $(N_{\text{Rx}} \times N_{\text{Frame}})$ received complex symbols at radio-head j in frame k. Furthermore, we let $\mathbf{y}_{i,k}^{(j)}(n)$ denote the nth column of $\mathbf{Y}_{i,k}^{(j)}$ (i.e., the observation of the nth symbol received from device i). We adopt a narrowband single-input multiple-output (SIMO) channel model according to

$$\mathbf{y}_{i,k}^{(j)}(n) = \mathbf{h}_{i,k}^{(j)} x_{i,k}(n) + \mathbf{w}_{i,k}^{(j)}(n), \tag{2}$$

for $n \in \{1, \cdots, N_k\}$, where $\mathbf{h}_{i,k}^{(j)}$ represent the channel-state vector between device i and the N_{Rx} antennas at radio-head j, $x_{i,k}(n)$ are the transmitted data symbols scaled to unit power, and $\mathbf{w}_{i,k}^{(j)}(n) \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{N_{\text{Rx}}})$ is the additive noise represented by a circular-symmetric complex Gaussian (CSCG) random vector. We will assume that data symbols $x_{i,k}(n)$ are encoded with a fixed rate R.

Channel Assumptions: We assume time-variant Rice fading channels that follow a complex Gaussian distribution defined by $\mathbf{h}_{i,k}^{(j)} \sim \mathcal{CN}(\boldsymbol{\mu}_i^{(j)}, \boldsymbol{\Sigma}_i^{(j)})$. This model is relevant in situations where there is a dominating line-of-sight or reflective path from the transmitter to the receiver. The channel-specific distribution parameters $\boldsymbol{\mu}_i^{(j)}, \boldsymbol{\Sigma}_i^{(j)}$ constitute features that will be used for verifying the authenticity of a given transmission based on observed channel states in the PLA scheme (to be defined in Section III-B). The fading realizations $\mathbf{h}_{i,k}^{(j)}$ are assumed independent from frame to frame and between radio-heads. Moreover, we also note that the path-loss and distance dependent received power is included in the magnitude of $\mathbf{h}_{i,k}^{(j)}$. Finally, throughout this paper we will assume perfect knowledge of the channel states $\mathbf{h}_{i,k}^{(j)}$ at each radio-head.

Sybil Attacker Model: The attacker Eve is assumed to be transmitting from an unknown location with channels distributed according to $\mathcal{CN}(\boldsymbol{\mu}_{\text{E}}^{(j)}, \boldsymbol{\Sigma}_{\text{E}}^{(j)})$. We presuppose the worst-case assumption that Eve knows everything about the system implementation, i.e., medium access and PHY-layer protocols. Eve's MAC-layer strategy is to impact the system performance by launching a Sybil attack, conducted by transmitting multiple DTA requests with fraudulent IDs referred to as Sybil IDs. The result is that the per-frame available resources N_k (given by (1)) are reduced, which depletes resources available to the other legitimate devices. Assuming there is no cryptographic-based authentication of the requests and data payload transmissions, or secret key information has leaked to Eve, Bob has to rely on the PLA scheme to distinguish the legitimate requests from the ones originating from Eve.

[2]Note that N_k in general is a random variable depending on the number of users allocated in the frame, and that N_k can get very small if many devices request resources at the same time.

Decoding \ Authentication	Centralized	Local
Centralized	**Scenario A** Centralized / soft decision decoding and soft PLA	**Scenario B** Local RRH decoding + soft PLA decision
Local	**N/A**	**Scenario C** Local RRH decoding + independent hard PLA

Fig. 2. The modeled decoding and authentication scenarios.

III. PLA Models and Problem Formulation

Based on received samples $\mathbf{y}_{i,k}^{(j)}(n)$, Bob needs to decode the transmission so that reliable reception of the device's data is achieved. However, due to the uncertainty to whether a given transmission originated from the legitimate device or from Eve, Bob additionally needs to perform PLA based on the channel-state information. We consider three scenarios for how decision making (i.e., PLA and message decoding) can be distributed in the considered multiple-RRH system. The scenarios are illustrated in Fig. 2 and defined as follows:

Scenario A: We assume completely centralized processing in the sense that each RRH uses maximum-ratio combining (MRC) and forwards soft information[3] that is centrally combined at Bob, yielding an effective signal-to-noise ratio (SNR) $\gamma(\tilde{\mathbf{h}}_k) = \frac{\|\tilde{\mathbf{h}}_k\|^2}{N_{\text{RRH}}N_0}$ with $\tilde{\mathbf{h}}_k = [\tilde{\mathbf{h}}_{1,k}^T, \cdots, \tilde{\mathbf{h}}_{N_\text{a},k}^T]^T$. We assume Bob can successfully decode the transmission given that $C_k > R$, where in this paper, we adopt the Shannon capacity $C_k = \log(1+\gamma_k)$. The centralized PLA scheme utilizes forwarded soft information from the RRHs and is defined in Section III-B.

Scenario B: We assume that decoding is performed locally at the RRHs, but PLA decisions are made centrally based on forwarded soft information. For the local decoding, we assume that RRH j independently uses $\mathbf{y}_k^{(j)}(n)$ for MRC and achieves the local SNR $\gamma_k^{(j)} = \frac{\|\mathbf{h}_k^{(j)}\|^2}{N_0}$. Given that $R < C_k^{(j)}$, with $C_k^{(j)} = \log(1+\gamma_k^{(j)})$, the RRH can successfully decode the transmission and forward it to Bob, who successfully decodes the transmission if *any* RRH is able to decode it.

Scenario C: We again assume that decoding is performed locally and modeled equivalently to Scenario B. However, here the PLA scheme instead performs local hard decisions (i.e., binary decision to accept or reject the message) which need to be fused into a final decision at Bob. The local PLA procedure will also be defined in Section III-B.

Discussion on RRH Data Links: Note that the three considered scenarios essentially correspond to varying degrees of bandwidth on the links from the RRHs to Bob. Scenario A demands $2N_k+1$ real-valued samples per transmission and RRH link (i.e., with MRC of the received samples at the RRHs and a soft value for PLA). Scenario B demands RN_k bits and one real-valued sample (i.e., the data bits and the PLA soft decision) while Scenario C demands only RN_k+1 bits per transmission and link. In practice, the signaling from the RRHs to Bob would have additional impacts on the system-level delay performance; however, in this work we mainly focus on the direct impacts of the distributed PLA decision making. Therefore, we make the ideal assumption of error- and latency-free links from the RRHs to Bob.

[3]Note that due to the assumption of independent RRH channels, the soft-combining is in our case equivalent to Bob having $\mathbf{y}_{1,k}(n), \cdots, \mathbf{y}_{N_\text{RRH},k}(n)$ and $\tilde{\mathbf{h}}_{1,k}, \cdots, \tilde{\mathbf{h}}_{N_\text{RRH},k}$ centrally available.

A. Queueing Model

With the aim of quantifying the delay performance of the different distributed decision-making scenarios, we let infinite-buffer queues model the flow of data from each device to Bob. The data-link from device i (the data source) to Bob (the data sink) are modeled by the bivariate stochastic processes $A_i(\tau,t) = \sum_{k=\tau}^{t} a_{k,i}$ and $D_i(\tau,t) = \sum_{k=\tau}^{t} d_{k,i}$, that denote the cumulative arrivals to and departures from link i, respectively, for all $0 \le \tau \le t$. Here $a_{k,i}$ represents the instantaneous arrivals to the device's buffer in frame k measured in bits (see Fig 1). In this work, we will assume a constant arrival-rate of α bits per device and frame. Similarly, $d_{k,i}$ represents the instantaneous departures from the queue (i.e., information successfully received, decoded, and authenticated at Bob). The link's ability to transfer data from the local buffer to Bob is characterized by the cumulative service process $S_i(\tau,t) = \sum_{k=\tau}^{t} s_{k,i}$, where $s_{k,i}$ represents the instantaneous service in frame k; that is, $s_{k,i}$ is the amount of information bits that can be reliably communicated over the wireless channels to the RRHs and successfully received, decoded, and authenticated at Bob.

The queueing delay at time point t is defined as

$$W_i(t) \triangleq \inf\{u > 0; A_i(0,t) \le D_i(0,t+u)\}, \tag{3}$$

representing the frames required to serve the bits in the queue at time t. This delay is randomly varying due to the stochastic service process. A widely used measure on the queueing system's ability to meet delay requirements is the *delay violation probability* [15], defined as $p_i(w) = \mathbb{P}(W_i(t) > w)$, i.e., the probability that information on the link from device i is not received within a defined deadline w.

For a general model that can be used for Scenario A-C, we employ the following definition of the instantaneous service

$$s_k = \begin{cases} RN_k & \text{if} \quad X_{k,i} = 1 \\ 0 & \text{if} \quad X_{k,i} = 0 \end{cases}, \tag{4}$$

where $X_{k,i}$ is a stochastic indicator of successful reception at Bob. This model (4) gives us two simple interfaces, N_k and $X_{k,i}$, whose distributions will determine the performance of the link. In the rest of this paper we will denote the distribution of $X_{k,i}$ using $p_X = \mathbb{P}(X_{k,i}=0) = 1 - \mathbb{P}(X_{k,i}=1)$.

B. Physical Layer Authentication Scheme

For ease of notation, we drop the dependence of the frame index k in this section. PLA in this work is performed on a message m received in a frame that potentially is originating from Eve. We denote by $\tilde{\mathbf{h}}_m^{(j)}$ the SIMO channel-states associated with the message, which are assumed to be observed without estimation error at the RRHs. The authentication procedure extracts the message identifier $\text{ID}(m)$ and defines the binary hypotheses $\mathcal{H}_0$, representing that the message is

indeed from device ID(m), and $\mathcal{H}_1$, representing that the message is from the attacker Eve.

Now let us introduce the centralized PLA scheme:

Definition 1 (Centralized PLA)*: Bob makes a decision according to the following binary hypothesis test,*

$$d_i(\tilde{\mathbf{h}}_m) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T, \quad \textit{with} \quad d_i(\tilde{\mathbf{h}}_m) = \sum_{j=1}^{N_{RRH}} d_i^{(j)}(\tilde{\mathbf{h}}_m^{(j)}), \tag{5}$$

where T is a design threshold and $d_i^{(j)}(\tilde{\mathbf{h}}_m) = 2\|\tilde{\mathbf{h}}_m - \boldsymbol{\mu}_i^{(j)}\|^2_{\{\boldsymbol{\Sigma}_i^{(j)}\}^{-1}}$ are soft decisions received from the RRHs.

In contrast to [11] and [12], in this paper we also consider local PLA with decision fusing, defined by

Definition 2 (Local PLA with hard decision fusing)*: RRH j makes a decision $b^{(j)} \in \{0,1\}$ according to*

$$d_i^{(j)}(\tilde{\mathbf{h}}_m^{(j)}) \underset{b^{(j)}=0}{\overset{b^{(j)}=1}{\gtrless}} T. \tag{6}$$

Upon receiving the hard decisions $b^{(j)}$, Bob makes a final decision $b = \mathcal{D}(b^{(1)}, \cdots, b^{(N_{RRH})})$, where $\mathcal{D} : \{0,1\}^{N_{RRH}} \to \{0,1\}$ is a binary decision rule.

For an optimal decision rule design, Bob would need the receiver operating characteristics of each hypothesis test, and such information is generally not available at Bob due to the unknown distributions of the channels from Eve. In this work, we restrict ourselves to the three straightforward decision rules: (i) an *all-accept* rule, where a message is accepted only if $\mathbf{b}_i = 1$ for all RRHs, (ii) an *all-reject* rule, where a message is rejected only if $\mathbf{b}_i = 0$ for all RRHs, and (iii) a *majority-vote* rule, where a message is accepted if the majority of $\mathbf{b}_i = 1$.

Error Probabilities: Two types of errors can occur in the presented binary authentication tests: a *false alarm* when a legitimate message is rejected, and a *missed detection* when an adversary message is accepted. The probability of false alarm and missed detection are defined as $p_{\text{FA}}(T) = \mathbb{P}(d(\tilde{\mathbf{h}}_m) > T|\mathcal{H}_0)$ and $p_{\text{MD}}(T) = \mathbb{P}(d(\tilde{\mathbf{h}}_m) < T|\mathcal{H}_1)$, respectively. It is easy to show that $d(\tilde{\mathbf{h}}_m)|\mathcal{H}_0 \sim \chi^2_{2N_{\text{RRH}}N_{\text{Rx}}}$, i.e., the discriminant function is following a central χ^2 distribution. Hence, the false alarm probability is given by

$$p_{\text{FA}}(T) = 1 - F_{\chi^2_{2N_{\text{RRH}}N_{\text{Rx}}}}(T). \tag{7}$$

Based on (7), the threshold T can be computed for a corresponding false alarm probability. Similarly, the individual thresholds of local PLA decisions can be computed from (7) except with $2N_{\text{Rx}}$ degrees of freedom.

Power Manipulation Attack: Given that Eve knows that Bob is employing the PLA scheme, she can launch a counter-strategy at the PHY-layer. In this strategy, called a power manipulation attack, we assume that Eve manipulates the transmit power and phase at her single-antenna transmitter by employing a complex scaling factor $\rho_E e^{j\varphi_E}$ such that the channel state observed at Bob becomes $\eta_E e^{j\psi_E}\mathbf{h}_{\text{E}}$.

C. Problem Formulation

The system-level delay performance impacts of PLA will be influenced by: (i) the choice of authentication threshold T and the resulting tradeoff between false alarms and missed detections; and (ii) the configuration of decoding and authentication decisions modeled by Scenario A-C. The delay impacts are quantified by the delay violation probability $p_i(w)$, which unfortunately is very complicated to derive in closed form. Deriving an upper bound on $p_i(w)$ is however possible and such bounds prove useful as performance guarantees in a mission-critical communications, i.e., a real system operating under the assumed conditions will with high probability achieve a better delay performance. In the following Section IV, we provide such bounds for the considered distributed PLA system model that hold for Scenarios A-C and under the Sybil attack.

IV. Delay Bound for Distributed PLA

In this section, we derive an upper bound on the delay violation probability $p_i(w)$. First, we introduce the relevant tools from stochastic network calculus.

A. Stochastic Network Calculus

Stochastic network calculus (SNC) is a mathematical framework that allows us to analyze input-output relationships of stochastic queueing systems through, for example, performance bounds on delay or backlog given arrival and service distributions. For a complete overview of stochastic network calculus, we refer to [14]. In [15], the stochastic network calculus framework was developed to study wireless fading links by introducing the exponential transformation $\mathcal{A}(\tau,t) \triangleq e^{A(\tau,t)}$, $\mathcal{S}(\tau,t) \triangleq e^{S(\tau,t)}$ and $\mathcal{D}(\tau,t) \triangleq e^{D(\tau,t)}$. This transformation allows the characterization of the random service process in terms of the varying instantaneous SNR due to fading. This is referred to as transforming the bit-domain processes into the SNR-domain since the processes become linear in the instantaneous SNR γ_k instead of logarithmic. Arrival processes in the SNR-domain can then be seen as instantaneous SNR demands. The performance bounds, which can be seen as variations of moment bounds, are derived in terms of Mellin transforms of the involved queueing processes. The Mellin transform of a random variable X, closely related to the moment-generating function (MGF), is defined as $\mathcal{M}_X(s) = \mathbb{E}[X^{s-1}]$.

The upper bound on the delay violation probability we utilize in this paper is given by the following lemma:

Lemma 1*: For $s > 0$ and under the stability condition $\mathcal{M}_{\mathcal{A}}(1+s)\mathcal{M}_{\mathcal{S}}(1-s) < 1$,*

$$p(w) \leq \min_s \frac{\mathcal{M}_{\mathcal{S}}(1-s)^w}{1 - \mathcal{M}_{\mathcal{A}}(1+s)\mathcal{M}_{\mathcal{S}}(1-s)} \tag{8}$$

where $\mathcal{M}_{\mathcal{S}}(s) \triangleq \mathbb{E}[e^{s_k(s-1)}]$ and $\mathcal{M}_{\mathcal{A}}(s) \triangleq \mathbb{E}[e^{a_k(s-1)}]$ are the Mellin transforms of the independent SNR-domain service and arrival processes, respectively.

Proof. See [15, Theorem 1]. □

The upper bound (8) can be shown to be a convex function for every s in the stability interval $\mathcal{M}_{\mathcal{A}}(1+s)\mathcal{M}_{\mathcal{S}}(1-s) <$

1 [16, Theorem 1]. However, no analytical tools from convex optimization can be applied, and therefore, one typically resorts to a numerical grid search.

In the rest of this section, we will characterize the Mellin-transforms $\mathcal{M}_{\mathcal{S}}(s)$ and $\mathcal{M}_{\mathcal{A}}(s)$.

B. Service Process Mellin Transform

First, let us provide a general expression for the Mellin transform of the service process in the following lemma:

Lemma 2 (Service Process Mellin Transform)*:*

$$\mathcal{M}_{\mathcal{S}}(s) = (1 - p_X) \sum_n e^{Rn(s-1)} p_N(n) + p_X \tag{9}$$

Proof. Clearly, $\mathbb{E}[e^{s_k(s-1)} | X_k = 1, N_k] = e^{RN_k(s-1)}$. Taking expectation over N_k and X_k yields (9). □

Obtaining an exact expression for p_X is difficult. We will instead provide an upper bound in the following theorem:

Theorem 1 (Service outage probability bound)*: The probability of service outage for Scenario $\mathcal{S} \in \{A, B, C\}$ is upper bounded by*

$$p_X \leq p_{FA}^{(\mathcal{S})} + p_{out}^{(\mathcal{S})}, \tag{10}$$

where $p_{FA}^{(\mathcal{S})}$ and $p_{out}^{(\mathcal{S})}$ are the corresponding false-alarm and SNR-outage probabilities, respectively, given by

$$p_{FA}^{(A)} = p_{FA}^{(B)} = p_{FA}(T), \tag{11}$$

$$p_{FA}^{(C)} = \sum_{n=1}^{K} \binom{N_{RRH}}{n} (1 - p_{FA}(T))^n p_{FA}(T)^{N_{RRH}-n}, \tag{12}$$

and

$$p_{out}^{(A)} = \mathbb{P}(\gamma_k < 2^R - 1), \quad p_{out}^{(B)} = p_{out}^{(C)} = \prod_{j=1}^{N_{RRH}} p_{out}^{(j)}(R), \tag{13}$$

for $p_{out}^{(j)}(R) = \mathbb{P}(\gamma_k^{(j)} < 2^R - 1)$. The limit K in (12) *is given by $K = \lfloor N_{RRH}/2 \rfloor$ (majority-vote), $K = 1$ (all-accept), or $K = N_{RRH}$ (all-reject) depending on the decision rule.*

Proof. We start by considering Scenario A. Note that we have

$$p_X = \mathbb{P}\left(\{d(\mathbf{h}) > T\} \cup \{\gamma_k < 2^R - 1\}\right) \tag{14}$$

$$\leq \mathbb{P}(d(\mathbf{h}) > T) + \mathbb{P}(\gamma_k < 2^R - 1) \tag{15}$$

from the union bound. Clearly, $\mathbb{P}(d(\mathbf{h}) > T) = p_{\text{FA}}(T)$, so (10) and (11) follow. In Scenario B, the union bound can be similarly applied as in (15), however, in this scenario an SNR outage happens if every RRH channel is in outage, and hence, the SNR-outage probability is $p_{\text{out}}^{(\text{B})} = \mathbb{P}\left(\cap_{j=1}^{N_{\text{RRH}}} \{\gamma_k^{(j)} < 2^R - 1\}\right)$ from which (13) follows. Again, the union bound can be applied in Scenario C; however, here a false-alarm occurs if a majority of RRHs suffer from false alarms. Hence the fusing false alarm probability follows a binomial distribution according to (12). □

The SNR-outage probabilities in (13) are generally not tractable for the considered distributed-RRH Rice-fading channel model and, hence, we leave them without closed-form expressions in Theorem 1. For the numerical results in Section V, we will use a moment-matching central χ^2 approximation to evaluate these probabilities.

C. Analysis for Sybil Attack

In a Sybil attack, we assume that $|\mathcal{I}_{\text{DTP}}| = K_{\text{Active}} + K_{\text{Sybil}}$ where K_{Active} and K_{Sybil} represent the number of active and Sybil devices that are granted DTP resources, respectively. Based on these assumptions, we recapitulate the distribution of N_k from [11] in the following lemma:

Lemma 3 (Resource Distribution Under Sybil Attack)*:*

$$p_{N_k}(n) = \sum_{l=0}^{\lfloor \frac{N_{Frame}}{n} \rfloor} p_{K_{Active}}(l) p_{K_{Sybil}}\left(l - \left\lfloor \frac{N_{Frame}}{n} \right\rfloor\right), \tag{16}$$

where $p_{K_{Active}}(\cdot)$ is the probability mass function (PMF) of scheduled active devices and $p_{K_{Sybil}}(\cdot)$ is the PMF of successful Sybil IDs. The latter can be approximated as

$$p_{K_{Sybil}}(k) \approx \sum_{B \in A_k} \prod_{i \in B} p_{MD}^{(\mathcal{S})}(i) \times \prod_{j \in B^c} (1 - p_{MD}^{(\mathcal{S})}(j)), \tag{17}$$

where A_k denotes the set of all size k subsets of D_{Sybil} and $p_{MD}^{(\mathcal{S})}(j, T)$ denotes the missed detection probability given that Eve impersonates device i in Scenario $\mathcal{S} \in \{A, B, C\}$.

Proof. This approximation stems from an assumption that the events $\{d_i(\mathbf{h}_E) < T\}_{i \in D_{\text{Sybil}}}$ can be approximated as independent, in which case K_{Sybil} is Poisson-binomial distributed. See [11] for further details. □

Next, we provide the missed detection probabilities:

Theorem 2*: The missed detection probabilities are given by $p_{MD}^{(A)}(i) = p_{MD}^{(B)}(i) = \mathbb{P}\left(d_i(\mathbf{h}_E) < T\right)$ and*

$$p_{MD}^{(C)}(i) = \sum_{n=1}^{L} \binom{N_{RRH}}{n} \left(1 - p_{MD}^{(loc)}(j)\right)^n p_{MD}^{(loc)}(j)^{N_{RRH}-n}, \tag{18}$$

for $p_{MD}^{(loc)} = F_{\chi^2_{2N_{Rx}}(\nu_i)}(T/\lambda_i)$, where $F_{\chi^2_{2N_{Rx}}(\nu_i)}(\cdot)$ is the CDF of a non-central χ^2 distribution with $2N_{Rx}$ degrees of freedom and non-centrality parameter $\nu_i = 2(\boldsymbol{\mu}_E - \boldsymbol{\mu}_i)^\dagger \boldsymbol{\Sigma}_E^{-1} (\boldsymbol{\mu}_E - \boldsymbol{\mu}_i)$. Again, the limit L in (18) *is given by $L = \lfloor N_{RRH}/2 \rfloor$ (majority-vote), $L = N_{RRH}$ (all-accept), or $L = 1$ (all-reject) depending on the decision rule.*

Proof. The first expression follows by definition. The second result (18) follows by observing that the local missed detection events $\{d_i^{(j)}(\mathbf{h}_{\text{E}}) < T\}$ are independent, and thus, the number of missed detections will follow a binomial distribution which cumulative distribution function is given by (18). Finally, the χ^2 distribution of $p_{\text{MD}}^{(\text{loc})}$ is a standard result. □

Finally, we close this section by providing the saddle-point approximation technique from [12, Theorem 2] which we will use to compute $p_{\text{MD}}^{(\text{A,B})}(i)$ under the power manipulation strategy:

Lemma 4 (Missed Detection Probability for $N_{\text{RRH}} \geq 2$)*:*

$$p_{MD}^{(A,B)}(i) \approx -\frac{1}{2\pi} e^{s(z_0)} e^{-j\angle s''(z_0)} \sqrt{\frac{2\pi}{|s''(z_0)|}}, \tag{19}$$

where

$$s(z) = -\mathbf{b}^\dagger \left(\mathbf{I} + \frac{1}{z}\mathbf{D}^{-1}\right)^{-1} \mathbf{b} - \ln(z) - \ln(|\mathbf{I} + z\mathbf{D}|), \tag{20}$$

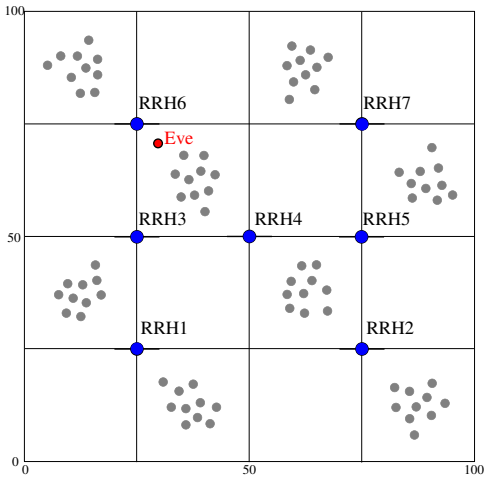

Fig. 3. Considered wireless industrial automation scenario.

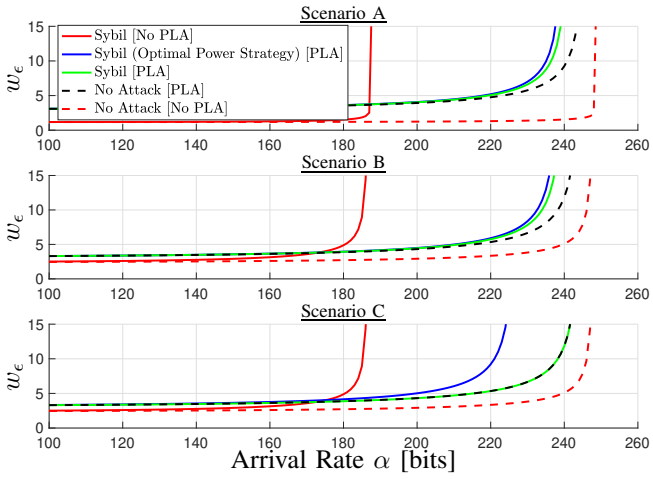

Fig. 4. Comparison of the system-level delay performance under distributed decision-making Scenarios A-C.

$\mathbf{b} = \mathbf{Q}_E^{\dagger}\boldsymbol{\mu}_E$, $\mathbf{Q}_E$ *and* $\mathbf{D}$ *are the Cholesky factorization and eigenvalues of* $\boldsymbol{\Sigma}_E$*, respectively, and* z_0 *is a stationary point such that* $s'(z_0) = 0$.

V. Numerical Results

In this section, we study the impacts of the PLA scheme in Scenario A-C in a wireless industrial automation deployment. Recall that Scenario A represents the completely centralized system, Scenario B represents centralized PLA and local decoding, and Scenario C represents local PLA and local decoding. As depicted in Fig. 3, we consider a 100 m $\times$ 100 m factory floor with 7 radio-head deployment locations. Sensor devices are deployed either uniformly across the factory floor or in the 8 sensor clusters illustrated in Fig. 3.

Channel Model: The channel distributions $\mathcal{CN}(\boldsymbol{\mu}_i^{(j)}, \boldsymbol{\Sigma}_i^{(j)})$ are based on the relative positions in the two-dimensional deployment area. We assume a uniform linear receive array for each radio-head location and compute $\boldsymbol{\mu}_i^{(j)}$ based on the distance and angle-of-arrival (AoA) with respect to each radio-head. For complete details regarding this model, we refer to [12]. In this section, we have assumed uncorrelated fading (i.e., $\boldsymbol{\Sigma}_i^{(j)} = \sigma^2\mathbf{I}$), Rice factor $K_{\text{Rice}} = 6$ dB and path-loss exponent $\beta = 2$. The RRHs have normalized antenna separation $\Delta_r = 0.5$ and the system is assumed to operate at $f_c = 2.4$ GHz carrier frequency with transmission rate $R = 1.5$ bits/channel-use. The average receive SNR from the devices was observed to vary between 10-12 dB.

a) Sybil Attack Impact: We first fix a prescribed delay violation probability $p_i(w_{i,\epsilon}) = 10^{-6}$ and solve for the corresponding delay guarantee $w_{i,\epsilon}$ under different scenarios. In Fig. 4, the worst-case delay is shown for varying device-buffer arrival rates α measured in bits. The results in Fig. 4 are based on a uniformly random deployment of 80 devices out of which $K_{\text{Active}} = 60$ are active and $|D_{\text{Sybil}}| = 20$ are subject to Sybil impersonation by Eve, $N_{\text{RRH}} = 2$ with arrays positioned at RRH2 and RRH4), and $N_{\text{Rx}} = 4$. Eve is positioned at (30 m, 70 m). Firstly, for low arrival rates we observe a fixed delay cost due to the false alarm probability, which was fixed to $p_{\text{FA}} = 10^{-2}$ via a constant threshold T computed from (7). We observe that the system-level delay under Sybil attack and without PLA grows infinite around $\alpha = 190$ bits, meaning that for arrival rates $\alpha > 190$ Eve will be able to use the considered Sybil attack to cause denial-of-service for at least one device. With the PLA scheme activated, we can see that the system has a higher arrival-rate capacity, closer to the baseline capacity of the system without the attack. In all scenarios, we see that Eve can enhance the Sybil attack impact by choosing the optimal power manipulation strategy (i.e., this shows the worst-case attack impact), where this effect is most pronounced for Scenario C. However, the system capacity is still larger than without the PLA scheme. For Scenario C, the all-accept decision rule (see Section III-B) was used since it achieved the best performance.

b) RRH Deployments: In Fig. 5(a), we plot the worst-case delay, but instead considering different deployment configurations of a total of 8 antennas. Eve is positioned at (30 m, 70 m) and is using the optimal power manipulation strategy. The curve colors correspond to: (i) a completely centralized deployment with $N_{\text{RRH}} = 1$ (RRH4), (ii) a $N_{\text{RRH}} = 2$ deployment (RRH3 and RRH5), and (iii) a $N_{\text{RRH}} = 4$ deployment (RRH1, RRH2, RRH6 and RRH7). We observe that all the distributed deployments outperform the centralized single-RRH system. There is a degradation in performance for Scenario C compared to Scenario A/B which is expected since detection performance is lost due to hard-decision fusing, which again is based on the all-accept rule. In Fig. 5(b), we show the results for the clustered device deployment. Eve is launching the Sybil attack, again for $|D_{\text{Sybil}}| = 20$ devices although inside the two clusters in the top-left corner of Fig. 3. Eve is again positioned at (30 m, 70 m) close to the attacked clusters and the all-accept rule was used for Scenario C. Here, we observe a large performance benefit from the distributed scenarios. For instance, considering a latency deadline of 10 frames in Fig. 5(b), the arrival-rate capacity for the distributed scenarios is 23% higher than for the centralized

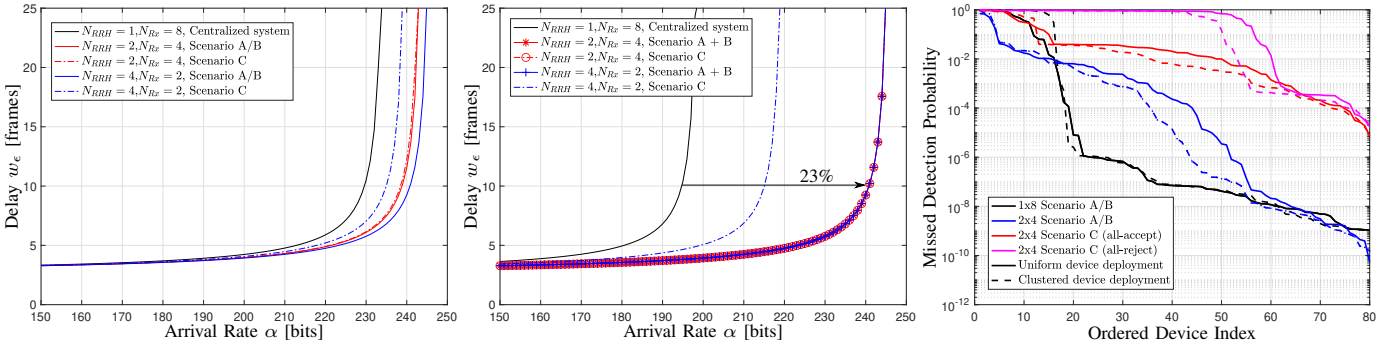

Fig. 5. Delay performance impacts of PLA under different deployment strategies and (a) uniform device deployment, and (b) clustered device deployment. (c) Detection performance of distributed PLA scheme.

system. The reason is that the single RRH cannot distinguish Eve from the nearby devices, resulting in many missed Sybil impersonations.

c) Discussion: In Scenario A and B, our results indicate a benefit in Sybil-attack resilience from distributing the antennas (i.e., increasing N_{RRH}). The performance difference between Scenario A and B is very small. This can be explained by the observation that the local decoding outage probabilities were significantly lower than the false-alarm probability. Hence, for the studied scenario, the false alarms dominates the delay impacts, and more generally it is $\max(p_{\text{FA}}, p_{\text{out}})$ that limits the performance. This indicates that Scenario B is preferred for the given system due to the lower bandwidth requirement on the RRH-to-Bob links. A comparison of Fig. 5(a) and 5(b) indicates that the Sybil-attack impact is larger when Eve is targeting the clustered device deployment. This is expected since Eve more easily can impersonate a group of devices in a close neighbourhood. This is furthermore explained by Fig. 5(c) where we show the missed detection probability when Eve is impersonating different devices in the network (ordered by descending missed detection probability). We see that the detection performance for the worst-case devices improves from distributing the radio-heads which explains the smaller Sybil attack impact. However, in Scenario C (i.e., with hard-decision fusing) we can observe a substantial performance degradation. Fig. 5(c) also compares the *all-accept* and *all-reject* decision rules and we observe that the all-accept rule provides better performance. The performance under Scenario C can potentially be further improved by investigating other decision fusing rules.

VI. Conclusion

In this paper, we have studied the delay performance impacts of PLA in a distributed antenna system. We have characterized the queueing service process under three scenarios of distributed PLA decision making. Moreover, we have analyzed the impact of a Sybil attack launched from a single-antenna transmitter with optimal power and phase rotation. Results have shown benefits of the distributed PLA schemes in terms of delay-performance impacts, most significant for centralized authentication. Authentication based on fusing local hard decisions was found to result in performance degradations, although still superior to the single-RRH system.

References

[1] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Fingerprints in the ether: Using the physical layer for wireless authentication," in *IEEE Int. Conference on Communications*, June 2007, pp. 4646–4651.

[2] F. J. Liu, X. Wang, and H. Tang, "Robust physical layer authentication using inherent properties of channel impulse response," in *Proc. of Military Communications Conference*, Nov 2011, pp. 538–542.

[3] P. Baracca, N. Laurenti, and S. Tomasin, "Physical layer authentication over MIMO fading wiretap channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2564–2573, July 2012.

[4] S. R. Panigrahi, N. Bjorsell, and M. Bengtsson, "Feasibility of large antenna arrays towards low latency ultra reliable communication," in *Proc. of IEEE International Conference on Industrial Technology*, 2017, pp. 1289–1294.

[5] M. Alonzo, P. Baracca, S. R. Khosravirad, and S. Buzzi, "URLLC for factory automation: an extensive throughput-reliability analysis of D-MIMO," in *Proc. of ITG Workshop on Smart Antennas*, 2020, pp. 1–6.

[6] A. Weinand, M. Karrenbauer, R. Sattiraju, and H. Schotten, "Application of machine learning for channel based message authentication in mission critical machine type communication," in *Proc. of European Wireless Conference*, May 2017, pp. 1–5.

[7] A. Weinand, R. Sattiraju, M. Karrenbauer, and H. D. Schotten, "Supervised learning for physical layer based message authentication in URLLC scenarios," in *Proc. of IEEE Vehicular Technology Conference*, Sep. 2019, pp. 1–7.

[8] A. Abdelaziz, C. E. Koksal, F. Barickman, R. Burton, J. Martin, and J. Weston, "Enhanced authentication based on angle of signal arrivals," *IEEE Trans. on Vehicular Technology*, vol. 68, no. 5, pp. 1–1, 2019.

[9] Z. Gu, H. Chen, P. Xu, Y. Li, and B. Vucetic, "Physical layer authentication for non-coherent massive SIMO-based industrial IoT communications," in *Proc. of IEEE Wireless Communications and Networking Conference*, 2020, pp. 1–6.

[10] H. Forssell, R. Thobaben, H. Al-Zubaidy, and J. Gross, "On the impact of feature-based physical layer authentication on network delay performance," in *IEEE Global Communications Conference*, Dec 2017.

[11] H. Forssell, R. Thobaben, H. Al-Zubaidy, and J. Gross, "Physical layer authentication in mission-critical MTC networks: A security and delay performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 795–808, April 2019.

[12] H. Forssell and R. Thobaben, "Worst-case detection performance for distributed SIMO physical layer authentication," *CoRR*, vol. 1609.03109, 2020.

[13] H. Forssell, R. Thobaben, and J. Gross, "Performance analysis of distributed SIMO physical layer authentication," in *Proc. of IEEE International Conference on Communications*, May 2019, pp. 1–6.

[14] M. Fidler and A. Rizk, "A guide to the stochastic network calculus," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 92–105, Firstquarter 2015.

[15] H. Al-Zubaidy, J. Liebeherr, and A. Burchard, "Network-layer performance analysis of multihop fading channels," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 204–217, Feb 2016.

[16] N. Petreska, H. Al-Zubaidy, R. Knorr, and J. Gross, "Bound-based power optimization for multi-hop heterogeneous wireless industrial networks under statistical delay constraints," *Elsevier Computer Networks*, vol. 148, pp. 262–279, 2019.