<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in . This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Li, Z., Oechtering, T. (2017)

Privacy-Constrained Parallel Distributed Neyman-Pearson Test

IEEE Transactions on Signal and Information Processing over Networks, 3(1): 77-90

<https://doi.org/10.1109/TSIPN.2016.2623092>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-192963>

Privacy-Constrained Parallel Distributed Neyman-Pearson Test

Zuxing Li, *Student Member, IEEE*, and Tobias J. Oechtering, *Senior Member, IEEE*

Abstract—In this paper, the privacy leakage problem in an eavesdropped parallel distributed binary hypothesis test network is considered. A novel Neyman-Pearson test-operational privacy leakage measure is proposed and a privacy-constrained distributed Neyman-Pearson test problem is formulated. Such privacy-constrained distributed Neyman-Pearson test network is designed to optimize the Neyman-Pearson test performance and meanwhile to satisfy a desired suppression constraint on the privacy leakage. This study characterizes the privacy-constrained distributed Neyman-Pearson test network design and particularly identifies the sufficiency of deterministic likelihood-ratio test for optimality. These results help to simplify the optimal design problem of a privacy-constrained distributed Neyman-Pearson test network. Numerical results are presented to show the trade-off between the test performance and privacy leakage in privacy-constrained distributed Neyman-Pearson test networks.

Index Terms—Eavesdropper, cyber-physical system, likelihood-ratio test, person-by-person optimality, physical-layer secrecy.

I. INTRODUCTION

The future cyber-physical system (CPS), such as smart house, eHealth, assisted living, and etc., will surely bring significant benefits on our environment and our lives. However, these systems will inevitably collect sensitive personal data, for which privacy has to be ensured. Therefore, the privacy problems of CPS attract much attention nowadays. Local distributed data processing is the core functionality of many proposed CPS applications, e.g., eHealth where sensors collect, process raw data, and then transmit local results to a terminal; the terminal makes the final conclusion for the user to monitor the health condition. Such local distributed data processing can be seen as an inference on a variable over a network.

An inference network can be modeled as a distributed detection/hypothesis test. In [1], [2], classic distributed detection problems were summarized in different formulations: Bayesian, Neyman-Pearson, or information-theoretic criterion, parallel, serial, or tree topology, one-shot or sequential processing, hard or soft decision, and so on. The basic distributed detection concepts were extended and reviewed in [3].

These established theories and results have to be modified when the privacy issue is taken into account in the distributed

detection network design. There are two pioneering works [4], [5] making efforts in this direction. In [4], it assumed that the eavesdropper is interested in the state of data transmission between remote decision makers and the fusion node. With an asymptotic assumption of an infinite number of remote decision makers and using the Kullback-Leibler divergence as an information-theoretic non-operational privacy leakage measure, it has been proved in [5] that likelihood-ratio tests (LRTs) are asymptotically optimal decision strategies for remote decision makers under a constraint on the eavesdropper's performance. A recent work [6] characterized an achievable rate-error-equivocation region of a distributed Neyman-Pearson test network with communication and privacy constraints. Unfortunately, the converse proof was only given for a special case. Some researches devised privacy-preserving schemes based on the assumption that the eavesdropper is uninformed about certain parameters. In [7], [8], they studied stochastic encryption methods where sensors intentionally generate errors to confuse the eavesdropper who is uninformed about the statistic of the induced errors. Similarly, works [9], [10] proposed channel aware encryption methods to design the transmission scheme between the sensors and fusion node based on the channel states which are not known by the eavesdropper. Besides the passive eavesdropping problems, active adversaries are also studied. Among them, the problem of Byzantines in a distributed detection network gains more attention recently [11], [12]. In [13], we studied two privacy-aware parallel distributed detection problems where the detection performance and eavesdropping privacy risk are evaluated by Bayesian detection-operational measures. For a privacy-constrained distributed Bayesian detection design, the decision strategy of an eavesdropped remote decision maker can be a deterministic LRT or a randomized strategy of two LRTs which achieves an operation point determined by the privacy constraints only.

Besides privacy problems in distributed detections/hypothesis tests, the statistical inference risks have been discussed in several other contexts. In particular, research on the differential privacy has gained much interest recently. The differential privacy problem [14] was studied to guarantee the statistical privacy by adding Laplace noise when an adversary observes two neighbor databases. In [15], the degradation of differential privacy level under adaptive interactions was characterized. In [16], the methods were developed to approximate a filter by its differentially private version. A survey of the differential privacy in machine learning was given in [17]. Relations between different formalisms for statistical inference privacy were discussed in

The work was supported in part by the Swedish Research Council (VR) project CLONE under Grant E0628201 and in part by the Strategic Research Agenda Program, Information and Communication Technology - The Next Generation (SRA ICT - TNG), through the Swedish Government.

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm 100 44, Sweden (e-mail: zuxing@kth.se; oech@kth.se).

Fig. 1. The studied model of an eavesdropped parallel distributed hypothesis test network. For the purpose of illustration, it shows that the first $n - 1$ remote decisions are intercepted. In the following discussion, we assume that the first s , $1 \leq s \leq n$, remote decisions are intercepted.

[18].

In this work, we study the privacy leakage problem of an eavesdropped distributed hypothesis test network. Different from the asymptotically optimal design in [5], we are interested in the design of the optimal privacy-constrained hypothesis test network from the Neyman-Pearson test-operational perspective. The eavesdropper in this study is assumed to be informed, i.e., the statistics of the distributed hypothesis test network are assumed to be known. Therefore, the eavesdropper considered here can raise severer privacy challenges than the eavesdropping problems studied in [7]–[10]. We characterize the privacy-constrained distributed Neyman-Pearson test network design and identify the sufficiency of deterministic LRTs for an optimal design. In our initial work [19], we studied a privacy-constrained distributed Neyman-Pearson detection problem in a simple network and under a priori restricting to consider deterministic fusion strategies only. In this work, we drop this restriction.

The remaining of this paper is organized as follows: The studied general model of an eavesdropped parallel distributed binary hypothesis test network is introduced in Section II; the distributed Neyman-Pearson test problem is briefly recapitulated in Section III; in Section IV, a Neyman-Pearson test-operational privacy leakage measure is proposed and a privacy-constrained distributed Neyman-Pearson test problem is formulated and studied; numerical studies on the trade-off between the test performance and privacy leakage in privacy-constrained distributed Neyman-Pearson test networks are illustrated in Section V; and the conclusion is given in Section VI.

In this paper, we will denote a random variable by a capital letter, e.g., X , its realization by the lower-case letter, e.g., x , and its definition set by the corresponding calligraphic letter, e.g., $\mathcal{X}$. Let X^i , x^i , and $\mathcal{X}^i$ denote a random sequence $(X_1, \dots, X_i)$, its realization sequence $(x_1, \dots, x_i)$, and the definition set $\mathcal{X}_1 \times \dots \times \mathcal{X}_i$ respectively. Let X_j^i , x_j^i , and $\mathcal{X}_j^i$ denote a random sequence $(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_i)$, its realization sequence $(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_i)$, and the definition set $\mathcal{X}_1 \times \dots \times \mathcal{X}_{j-1} \times \mathcal{X}_{j+1} \times \dots \times \mathcal{X}_i$.

II. STUDIED SYSTEM MODEL

The studied model of an eavesdropped parallel distributed hypothesis test network is shown in Fig. 1. Let H denote a binary hypothesis. W.l.o.g., the hypothesis is defined on the set $\mathcal{H} = \{0, 1\}$. There are n , $n \geq 2$, remote decision makers. A remote decision maker DM_i , $1 \leq i \leq n$, receives a continuous random observation Y_i which is defined on the set $\mathcal{Y}_i$ and is conditionally independently generated following a conditional p.d.f.¹ $f_{Y_i|H}$ such that

$$f_{Y^n|H}(y^n|h) = \prod_{i=1}^n f_{Y_i|H}(y_i|h).$$

Then, DM_i independently makes a binary remote decision M_i which is defined on the set $\mathcal{M}_i = \{0, 1\}$ following a remote decision strategy

$$\gamma_i : \mathcal{Y}_i \rightarrow \mathcal{M}_i.$$

Let Γ_i denote the set of all possible remote decision strategies of DM_i . A remote decision strategy γ_i can be represented by a conditional p.m.f. $p_{M_i|Y_i}$ and can further be mapped to an operation point of (p_{Fi}, p_{Di}) , where the detection probability p_{Di} and false-alarm probability p_{Fi} of DM_i are defined in terms of $p_{M_i|Y_i}$ as follows:

$$p_{Di} = p_{M_i|H}(1|1) = \int_{\mathcal{Y}_i} p_{M_i|Y_i}(1|y_i) f_{Y_i|H}(y_i|1) dy_i,$$

$$p_{Fi} = p_{M_i|H}(1|0) = \int_{\mathcal{Y}_i} p_{M_i|Y_i}(1|y_i) f_{Y_i|H}(y_i|0) dy_i.$$

In the following, let γ_i , $1 \leq i \leq n$, denote a set of remote decision strategies $\gamma_1, \dots, \gamma_i$, and γ_{ij} , $1 \leq j \leq i$, denote a set of remote decision strategies $\gamma_1, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_i$. Correspondingly, define $\Gamma_i = \Gamma_1 \times \dots \times \Gamma_i$ and $\Gamma_{ij} = \Gamma_1 \times \dots \times \Gamma_{j-1} \times \Gamma_{j+1} \times \dots \times \Gamma_i$ such that $\gamma_i \in \Gamma_i$ and $\gamma_{ij} \in \Gamma_{ij}$.

All remote decisions M^n are transmitted to the fusion center (FC) through error-free, rate-limited links. FC makes a binary fusion decision H_F defined on the set $\mathcal{H}$ to infer on the realization of H following a fusion strategy

$$\gamma_F : \mathcal{M}^n \rightarrow \mathcal{H}.$$

Let Γ_F denote the set of all possible fusion strategies. A fusion strategy can be represented by a conditional p.m.f. $p_{H_F|M^n}$.

W.l.o.g., the first s , $1 \leq s \leq n$, remote decisions are intercepted by an eavesdropper (EVE), i.e., EVE knows M^s . Based on the intercepted remote decisions, EVE is assumed to make a binary decision H_E defined on the set $\mathcal{H}$ to infer on the realization of the hypothesis H following an eavesdropping strategy

$$\gamma_E : \mathcal{M}^s \rightarrow \mathcal{H}.$$

Let Γ_E denote the set of all possible eavesdropping strategies. Similarly, an eavesdropping strategy can be equivalently represented by a conditional p.m.f. $p_{H_E|M^s}$.

Note that the (randomized) decision strategies of all remote decision makers, FC, and EVE are assumed to be implemented independently in this paper.

¹The conditional p.d.f. $f_{Y_i|H}$ contains no point masses of probability and the likelihood ratio $\frac{f_{Y_i|H}(\cdot|0)}{f_{Y_i|H}(\cdot|1)}$ contains no point masses.

In this work, we are mostly interested in theoretic analysis on the privacy leakage and developing the basic principles and understandings of the privacy-preserving design based on the studied eavesdropped parallel distributed hypothesis test network. An application driver for such research on the distributed hypothesis test network is emerging sensor network technology [20]. The sensor network technology can be widely used for CPS applications for monitoring or sensing. For instance, the parallel distributed binary hypothesis test network in Fig. 1 can be used to model an energy-limited sensor network for a personal eHealth application which measures the physiological indexes by embedded or wearable sensors and feeds back the user with the health condition to be “good” or “bad”. Due to the energy-saving concern, sensors can simply make binary messages and forward them to the fusion node, e.g., a simple “on”/“off” transmission scheme was pointed out in [21] to be energy-efficient since the “off” state is communicated at zero energy cost. The privacy-sensitive personal health condition and the vulnerable nature of links between sensors and fusion node to eavesdropping motivate this study to suppress privacy leakage through the method of privacy-by-design of distributed hypothesis test network. In a sensor network for a personal eHealth application, the links between sensors and fusion node can be open-access or protected. The messages transmitted over the open-access links are modeled as the s intercepted remote decisions in the studied eavesdropped distributed hypothesis test network.

In the next section, we briefly recapitulate the distributed Neyman-Pearson test problem which is used to model the hypothesis test and privacy leakage in the studied eavesdropped distributed hypothesis test network.

III. DISTRIBUTED NEYMAN-PEARSON TEST

In the context of the eavesdropped distributed hypothesis test network introduced in Section II, important concepts and results of the distributed Neyman-Pearson test problem are reviewed in this section as fundamentals of the extended discussion of the distributed Neyman-Pearson test problem with a privacy leakage suppression constraint.

Define the false-alarm and miss probabilities of the fusion decision as

$$p_{\text{FF}} = p_{H_F|H}(1|0) = \sum_{m^n \in \mathcal{M}^n} p_{H_F|M^n}(1|m^n) \prod_{i=1}^n p_{M_i|H}(m_i|0),$$

$$p_{\text{MF}} = p_{H_F|H}(0|1) = \sum_{m^n \in \mathcal{M}^n} p_{H_F|M^n}(0|m^n) \prod_{i=1}^n p_{M_i|H}(m_i|1),$$

where $p_{H_F|M^n}$ depends on γ_F and $p_{M_i|H}$ depends on γ_i . It is obvious that decreasing p_{FF} and decreasing p_{MF} are conflicting design objectives of a distributed hypothesis test network. The distributed Neyman-Pearson test problem is formulated to take into account the two conflicting objectives as

$$\min_{\gamma_n \in \Gamma_n, \gamma_F \in \Gamma_F} p_{\text{MF}}, \text{ s.t. } p_{\text{FF}} \leq \lambda, \quad (1)$$

i.e., a distributed Neyman-Pearson test network is designed to minimize the miss probability of the fusion decision under an upper bound constraint on the false-alarm probability.

Given a set of remote decision strategies $\gamma_n \in \Gamma_n$, the fusion observation likelihood $p_{M^n|H} = p_{M_1|H} \cdots p_{M_n|H}$ is determined and the problem (1) reduces to a centralized Neyman-Pearson test problem

$$\min_{\gamma_F \in \Gamma_F} p_{\text{MF}}, \text{ s.t. } p_{\text{FF}} \leq \lambda. \quad (2)$$

In this case, a fusion strategy γ_F can be represented by the corresponding probabilities p_{MF} and p_{FF} . Given $\gamma_n \in \Gamma_n$, denote the operation region of FC by $\mathcal{R}_F(\gamma_n)$ which consists of operation points of $(p_{\text{FF}}, p_{\text{MF}})$ achieved by all fusion strategies. In the following, we always use $\hat{\cdot}$ and $\check{\cdot}$ to denote the upper boundary and lower boundary of an operation region. In $\mathcal{R}_F(\gamma_n)$, the lower boundary $\check{\mathcal{R}}_F(\gamma_n)$ can be seen as a parameterized function of

$$\check{\mathcal{R}}_F : \Gamma_n \times [0, 1] \rightarrow [0, 1]$$

and maps p_{FF} to p_{MF} on the lower boundary $\check{\mathcal{R}}_F(\gamma_n)$. In [1], it has been shown that, out of all deterministic fusion strategies, only deterministic LRTs of the form of

$$\frac{p_{M^n|H}(m^n|0)}{p_{M^n|H}(m^n|1)} \underset{h_F=1}{\overset{h_F=0}{\geq}} \varepsilon, \varepsilon \in \mathbb{R} \quad (3)$$

achieve points on $\check{\mathcal{R}}_F(\gamma_n)$. Based on the result, the properties of $\check{\mathcal{R}}_F(\gamma_n)$ summarized in Propositions 1-3 can be proved by constructing these deterministic LRTs through change of the test threshold ε from $-\infty$ to ∞ . In Fig. 2, a lower boundary $\check{\mathcal{R}}_F(\gamma_2)$ is illustrated.

Proposition 1. *Given $\gamma_n \in \Gamma_n$, $\check{\mathcal{R}}_F(\gamma_n, p_{\text{FF}})$ is a convex, decreasing function of p_{FF} , $0 \leq p_{\text{FF}} \leq 1$. $\check{\mathcal{R}}_F(\gamma_n, 0) = 1$ and $\check{\mathcal{R}}_F(\gamma_n, 1) = 0$.*

Given $1 \leq i \leq n$, let

$$r^i : \{1, \dots, |\mathcal{M}^i|\} \rightarrow \mathcal{M}^i$$

denote a likelihood ratio sort function which maps an index k , $1 \leq k \leq |\mathcal{M}^i|$, on the realization m^i with the k -th smallest likelihood ratio. Therewith, we have the following sorted likelihood ratio chain

$$\frac{p_{M^i|H}(r^i(1)|0)}{p_{M^i|H}(r^i(1)|1)} \leq \dots \leq \frac{p_{M^i|H}(r^i(|\mathcal{M}^i|)|0)}{p_{M^i|H}(r^i(|\mathcal{M}^i|)|1)}.$$

Given $1 \leq j \leq i \leq n$, let

$$r_j^i : \{1, \dots, |\mathcal{M}_j^i|\} \rightarrow \mathcal{M}_j^i$$

denote a likelihood ratio sort function which maps an index k , $1 \leq k \leq |\mathcal{M}_j^i|$, on the realization m_j^i with the k -th smallest likelihood ratio. Therewith, we have the following sorted likelihood ratio chain

$$\frac{p_{M_j^i|H}(r_j^i(1)|0)}{p_{M_j^i|H}(r_j^i(1)|1)} \leq \dots \leq \frac{p_{M_j^i|H}(r_j^i(|\mathcal{M}_j^i|)|0)}{p_{M_j^i|H}(r_j^i(|\mathcal{M}_j^i|)|1)}.$$

Proposition 2. *The lower boundary $\check{\mathcal{R}}_F(\gamma_n)$ consists of $|\mathcal{M}^n| = 2^n$ connected line segments. From $(0, 1)$ to $(1, 0)$, denote the k -th line segment by $\mathcal{L}_F(\gamma_n, k)$, $1 \leq k \leq 2^n$, such that p_{FF} increases by $p_{M^n|H}(r^n(k)|0)$ along the line segment while p_{MF} decreases by $p_{M^n|H}(r^n(k)|1)$.*

Fig. 2. Illustration of a lower boundary $\check{\mathcal{R}}_F(\gamma_2)$ for a distributed hypothesis test network with $n = 2$ remote decision makers. All circles are achieved by deterministic LRTs. Any other point, e.g., the cross point, can be achieved by a randomized fusion strategy of two deterministic LRTs.

For $1 \leq i \leq n$ and $h \in \{0, 1\}$, define a cumulative sorted likelihood function as

$$c(i, k, h) = \begin{cases} \sum_{j=1}^k p_{M^i|H}(r^i(j)|h), & \text{if } k \geq 1 \\ 0, & \text{if } k = 0 \end{cases}.$$

Based on Proposition 2, the end points of $\mathcal{L}_F(\gamma_n, k)$ can be identified as: $(c(n, k-1, 0), 1 - c(n, k-1, 1))$ and $(c(n, k, 0), 1 - c(n, k, 1))$.

Proposition 3. On $\check{\mathcal{R}}_F(\gamma_n)$, an intersection point between two neighbor line segments of different slopes is achieved by a deterministic LRT of the form of (3). The end points $(0, 1)$ and $(1, 0)$ can be achieved by deterministic LRTs with thresholds $-\infty$ and ∞ . Any other point on $\check{\mathcal{R}}_F(\gamma_n)$ can be achieved by a randomized fusion strategy of two deterministic LRTs.

Given $\gamma_n \in \Gamma_n$, the properties of $\check{\mathcal{R}}_F(\gamma_n)$ imply that the fusion strategy of the Neyman-Pearson test problem (2) achieves an operation point $(\lambda, \check{\mathcal{R}}_F(\gamma_n, \lambda))$ and

$$\check{\mathcal{R}}_F(\gamma_n, \lambda) = \min_{\gamma_F \in \Gamma_F} p_{MF}, \text{ s.t. } p_{FF} \leq \lambda. \quad (4)$$

Therefore, the distributed Neyman-Pearson test problem (1) can be equivalently reformulated as

$$\min_{\gamma_n \in \Gamma_n} \check{\mathcal{R}}_F(\gamma_n, \lambda). \quad (5)$$

In the following, we always use a superscript $*$ to denote an optimal decision strategy. Given the remote decision strategies γ_n^* in the distributed Neyman-Pearson test network, the fusion strategy γ_F^* is a deterministic LRT or a randomized test of two LRTs which achieves an operation point $(\lambda, \check{\mathcal{R}}_F(\gamma_n^*, \lambda))$.

Denote the operation region of DM_i by $\mathcal{R}_i$ which consists of operation points of (p_{Fi}, p_{Di}) achieved by all remote decision strategies. In Fig. 3, an operation region $\mathcal{R}_i$ is illustrated. The properties of $\mathcal{R}_i$ shown in [22] are summarized in the following.

Proposition 4. $\mathcal{R}_i$ is a convex set. The upper boundary $\hat{\mathcal{R}}_i$, which is known as the receiver operating characteristic (ROC) curve, is concave, increasing, above the line $p_{Fi} = p_{Di}$, with

Fig. 3. Illustration of an operation region $\mathcal{R}_i$. On the boundaries $\hat{\mathcal{R}}_i$ and $\check{\mathcal{R}}_i$, all points are achieved by deterministic LRTs.

end points $(0, 0)$, $(1, 1)$, and consists of operation points achieved by deterministic LRTs of the form of

$$\frac{f_{Y_i|H}(y_i|0)}{f_{Y_i|H}(y_i|1)} \underset{m_i=1}{\overset{m_i=0}{\geq}} \epsilon, \epsilon \in \mathbb{R}.$$

The lower boundary $\check{\mathcal{R}}_i$, which is known as the inverse ROC curve, consists of operation points achieved by deterministic LRTs of the form of

$$\frac{f_{Y_i|H}(y_i|0)}{f_{Y_i|H}(y_i|1)} \underset{m_i=0}{\overset{m_i=1}{\geq}} \zeta, \zeta \in \mathbb{R}.$$

The boundaries are point symmetric with respect to $(0.5, 0.5)$.

In the following discussion, we will reuse $\hat{\mathcal{R}}_i$ and $\check{\mathcal{R}}_i$ to denote functions

$$\hat{\mathcal{R}}_i : [0, 1] \rightarrow [0, 1]$$

$$\check{\mathcal{R}}_i : [0, 1] \rightarrow [0, 1]$$

which map p_{Fi} to p_{Di} on the upper boundary $\hat{\mathcal{R}}_i$ and the lower boundary $\check{\mathcal{R}}_i$ respectively.

IV. PRIVACY-CONSTRAINED DISTRIBUTED NEYMAN-PEARSON TEST

In this section, we first model the hypothesis test and privacy leakage as distributed Neyman-Pearson tests. A novel Neyman-Pearson test-operational privacy leakage measure is then proposed. Taking into account both the design objectives to improve the hypothesis test performance and to suppress the privacy leakage, a privacy-constrained distributed Neyman-Pearson test problem is formulated and studied.

A. Models of Hypothesis Test and Privacy Leakage

The studied system introduced in Section II consists of two distributed hypothesis test networks: the fusion hypothesis test network and the eavesdropping hypothesis test network.

To assess the hypothesis test performance of a fusion hypothesis test network design $(\gamma_n, \gamma_F) \in \Gamma_n \times \Gamma_F$, the used hypothesis test performance measure is the miss probability of the fusion decision p_{MF} under an upper bound constraint on the false-alarm probability $p_{FF} \leq \lambda$. A smaller value of the

measure p_{MF} means a better hypothesis test performance, and vice versa.

In the eavesdropping hypothesis test network, we make the following assumptions:

- 1) EVE is informed about the observation likelihoods $f_{Y_1|H}, \dots, f_{Y_s|H}$ and the used decision strategies γ_s of the remote decision makers in the eavesdropping hypothesis test network.
- 2) The eavesdropping strategy design is modeled as a centralized Neyman-Pearson test, i.e., the informed EVE is greedy to use the Neyman-Pearson test to minimize the miss probability of the eavesdropping decision p_{ME} under an upper bound constraint on the false-alarm probability p_{FE} .

The probabilities of the eavesdropping decision p_{ME} , p_{FE} are similarly defined as p_{MF} , p_{FF} as

$$p_{FE} = p_{H_E|H}(1|0) = \sum_{m^s \in \mathcal{M}^s} p_{H_E|M^s}(1|m^s) \prod_{i=1}^s p_{M_i|H}(m_i|0),$$

$$p_{ME} = p_{H_E|H}(0|1) = \sum_{m^s \in \mathcal{M}^s} p_{H_E|M^s}(0|m^s) \prod_{i=1}^s p_{M_i|H}(m_i|1).$$

By modeling the privacy leakage as a Neyman-Pearson test made by the informed and greedy EVE, the worst privacy leakage scenario is considered. The Neyman-Pearson test-operational model also leads to a concrete assessment of the privacy leakage and a hypothesis test network design with a concrete privacy leakage suppression guarantee.

B. Neyman-Pearson Test-Operational Privacy Leakage Measure

To assess the privacy leakage suppression performance of a design of remote decision makers in the eavesdropping hypothesis test network $\gamma_s \in \Gamma_s$, we use a Neyman-Pearson test-operational privacy leakage measure:

$$p_{ME}^{\min}(\sigma) = \min_{\gamma_E \in \Gamma_E} p_{ME}, \text{ s.t. } p_{FE} \leq \sigma, \quad (6)$$

which is the minimal miss probability of the eavesdropping decision under an upper bound constraint on the false-alarm probability. A larger value of the measure $p_{ME}^{\min}(\sigma)$ means a better privacy leakage suppression performance. Note that the value of the upper bound σ is set following the privacy leakage suppression interest of the network designer.

Given a set of remote decision strategies $\gamma_s \in \Gamma_s$, denote the operation region of EVE by $\mathcal{R}_E(\gamma_s)$. All the properties of $\check{\mathcal{R}}_F(\gamma_n)$ summarized in Propositions 1-3 can also be applied to $\check{\mathcal{R}}_E(\gamma_s)$. Then, Proposition 1 implies that

$$\check{\mathcal{R}}_E(\gamma_s, \sigma) \leq 1 - \sigma, \quad \forall \gamma_s \in \Gamma_s,$$

where the equality can be achieved if the given set of remote decision strategies γ_s lead to an equal eavesdropping observation likelihood ratio chain as

$$\frac{p_{M^s|H}(r^s(1)|0)}{p_{M^s|H}(r^s(1)|1)} = \dots = \frac{p_{M^s|H}(r^s(2^s)|0)}{p_{M^s|H}(r^s(2^s)|1)}.$$

Similar as (4), given a set of remote decision strategies $\gamma_s \in \Gamma_s$, we have

$$p_{ME}^{\min}(\sigma) = \check{\mathcal{R}}_E(\gamma_s, \sigma). \quad (7)$$

Therefore, the value of the privacy leakage measure $p_{ME}^{\min}(\sigma)$ is bounded as

$$p_{ME}^{\min}(\sigma) \leq 1 - \sigma.$$

C. Privacy-Constrained Distributed Neyman-Pearson Test Problem

In Section III, the reviewed distributed Neyman-Pearson test problem leads to an optimal distributed hypothesis test network design without privacy leakage concerns. In this section, we extend the classic problem formulation by taking into account the privacy leakage suppression in the distributed hypothesis test network design optimization. The privacy-constrained distributed Neyman-Pearson test problem is formulated as

$$\min_{\gamma_n \in \Gamma_n, \gamma_F \in \Gamma_F} p_{MF}, \text{ s.t. } p_{FF} \leq \lambda, \quad p_{ME}^{\min}(\sigma) \geq \delta. \quad (8)$$

In the privacy-constrained distributed Neyman-Pearson test problem, the hypothesis test performance is optimized by minimizing the measure p_{MF} under an upper bound constraint on the false-alarm probability $p_{FF} \leq \lambda$ and a privacy leakage suppression constraint $p_{ME}^{\min}(\sigma) \geq \delta$. Since the value of the privacy leakage measure is upper bounded as $p_{ME}^{\min}(\sigma) \leq 1 - \sigma$, the privacy-constrained distributed Neyman-Pearson test problem (8) is feasible if $\delta \leq 1 - \sigma$.

Given a set of remote decision strategies $\gamma_n \in \Gamma_n$, we have the equations (4) and (7). Therefore, the privacy-constrained distributed Neyman-Pearson test problem (8) can be equivalently reformulated as

$$\min_{\gamma_n \in \Gamma_n} \check{\mathcal{R}}_F(\gamma_n, \lambda), \text{ s.t. } \check{\mathcal{R}}_E(\gamma_s, \sigma) \geq \delta. \quad (9)$$

The optimization problem (8) leads to a privacy-constrained distributed Neyman-Pearson test network design. In the next part, properties of a privacy-constrained distributed Neyman-Pearson test network design will be studied.

D. Privacy-Constrained Distributed Neyman-Pearson Test Network Design

Here, we will characterize the privacy-constrained distributed Neyman-Pearson test network design. First, a general property of remote decision strategies is given. Then, privacy-constrained distributed Neyman-Pearson test design of a simple eavesdropped hypothesis test network is studied. Finally, the work is extended to study the privacy-constrained distributed Neyman-Pearson test design of the eavesdropped hypothesis test network introduced in Section II.

Now, let γ_n^* denote remote decision strategies in a privacy-constrained distributed Neyman-Pearson test network design. An important tool to be used in the following discussion of γ_n^* is the person-by-person optimality argument.

Remark 1. Given remote decision strategies γ_i^* , γ_{ni}^* in a privacy-constrained distributed Neyman-Pearson test network,

γ_i^* is a person-by-person optimal remote decision strategy:
if $1 \leq i \leq s$,

$$\gamma_i^* = \arg \min_{\gamma_i \in \Gamma_i} \check{\mathcal{R}}_F(\gamma_i, \gamma_{ni}^*, \lambda), \text{ s.t. } \check{\mathcal{R}}_E(\gamma_i, \gamma_{si}^*, \sigma) \geq \delta; \quad (10)$$

if $s < i \leq n$,

$$\gamma_i^* = \arg \min_{\gamma_i \in \Gamma_i} \check{\mathcal{R}}_F(\gamma_i, \gamma_{ni}^*, \lambda), \text{ s.t. } \check{\mathcal{R}}_E(\gamma_s^*, \sigma) \geq \delta. \quad (11)$$

Proof: If γ_i^* is not person-by-person optimal, there exists a remote decision strategy $\gamma_i \in \Gamma_i$ which, jointly with remote decision strategies γ_{ni}^* , achieves a smaller value of the hypothesis test performance measure p_{MF} under the constraint $p_{FF} \leq \lambda$ and the privacy leakage suppression constraint. Then, the assumption that $\gamma_i^*, \gamma_{ni}^*$ are the remote decision strategies of a privacy-constrained distributed Neyman-Pearson test network is violated. Therefore, γ_i^* is person-by-person optimal. ■

Note that a remote decision strategy $\gamma_i \in \Gamma_i$ can be represented by an operation point $(p_{Fi}, p_{Di}) \in \mathcal{R}_i$. In (10) and (11), the functions of $\gamma_i, \check{\mathcal{R}}_F(\gamma_i, \gamma_{ni}^*, \lambda)$ and $\check{\mathcal{R}}_E(\gamma_i, \gamma_{si}^*, \sigma)$, depend on the probabilities p_{Fi} and p_{Di} corresponding to γ_i . Therefore, we reuse the notations and study the following person-by-person optimization problem

$$(p_{Fi}^*, p_{Di}^*) = \arg \min_{(p_{Fi}, p_{Di}) \in \mathcal{R}_i(\gamma_{ni}^*)} \check{\mathcal{R}}_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, \lambda), \quad (12)$$

where $\mathcal{R}_i(\gamma_{ni}^*)$ denotes the privacy-constrained operation region of DM_i and is defined as:

if $1 \leq i \leq s$,

$$\mathcal{R}_i(\gamma_{ni}^*) = \mathcal{R}_i \cap \{(p_{Fi}, p_{Di}) : \check{\mathcal{R}}_E(p_{Fi}, p_{Di}, \gamma_{si}^*, \sigma) \geq \delta\};$$

if $s < i \leq n$,

$$\mathcal{R}_i(\gamma_{ni}^*) = \mathcal{R}_i \cap \{(p_{Fi}, p_{Di}) : \check{\mathcal{R}}_E(\gamma_s^*, \sigma) \geq \delta\} = \mathcal{R}_i.$$

The remote decision strategy γ_i^* achieves the operation point (p_{Fi}^*, p_{Di}^*) .

In a privacy-constrained distributed Neyman-Pearson test network design, the following general property of a remote decision strategy γ_i^* is satisfied.

Theorem 1. *It is sufficient to consider a remote decision strategy which achieves an operation point on the boundary of $\mathcal{R}_i(\gamma_{ni}^*)$ for γ_i^* in a privacy-constrained distributed Neyman-Pearson test network.*

The proof of Theorem 1 is given in Appendix A. Based on Theorem 1 and Proposition 4, we can further specify the decision strategy γ_i^* of DM_i whose decision M_i is not intercepted.

Corollary 1. *If M_i is not intercepted, it is sufficient to consider a deterministic LRT which achieves an operation point on the boundary of $\mathcal{R}_i$ for γ_i^* in a privacy-constrained distributed Neyman-Pearson test network.*

Proof: Since M_i of DM_i is not intercepted, we have $\mathcal{R}_i(\gamma_{ni}^*) = \mathcal{R}_i$. Theorem 1 shows that it is sufficient to consider a decision strategy which achieves an operation point

Fig. 4. Illustration of a privacy-constrained operation region $\mathcal{R}_1(\gamma_2^*)$ when $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$. The privacy leakage suppression constraint $\check{\mathcal{R}}_E(p_{F1}, p_{D1}, \sigma) \geq \delta$ leads to linear constraint boundaries: A – B – C on the upper boundary, A' – B' – C' on the lower boundary.

on the boundary of $\mathcal{R}_i$ for γ_i^* and Proposition 4 shows that all operation points on the boundary of $\mathcal{R}_i$ are achieved by deterministic LRTs. Then, we can conclude the property of γ_i^* with $s < i \leq n$ in Corollary 1. ■

1) *Design of a Simple Eavesdropped Hypothesis Test Network:* The studied simple eavesdropped hypothesis test network consists of 2 remote decision makers DM_1, DM_2 and only the remote decision M_1 of DM_1 is intercepted by EVE, i.e., $n = 2$ and $s = 1$.

The privacy-constrained distributed Neyman-Pearson test design of the simple eavesdropped hypothesis test network is

$$\gamma_2^*, \gamma_F^* = \arg \min_{\gamma_2 \in \Gamma_2, \gamma_F \in \Gamma_F} p_{MF}, \text{ s.t. } p_{FF} \leq \lambda, p_{ME}^{\min}(\sigma) \geq \delta. \quad (13)$$

Based on Corollary 1, it is sufficient to consider a deterministic LRT for γ_2^* . Then, we focus on characterizing γ_1^* by using the person-by-person optimality argument.

The remote decision strategy γ_1^* is also person-by-person optimal and achieves an operation point (p_{F1}^*, p_{D1}^*) as

$$(p_{F1}^*, p_{D1}^*) = \arg \min_{(p_{F1}, p_{D1}) \in \mathcal{R}_1(\gamma_2^*)} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \quad (14)$$

where the privacy-constrained operation region $\mathcal{R}_1(\gamma_2^*)$ can be specified as

$$\mathcal{R}_1(\gamma_2^*) = \mathcal{R}_1 \cap \{(p_{F1}, p_{D1}) : \check{\mathcal{R}}_E(p_{F1}, p_{D1}, \sigma) \geq \delta\}.$$

In our initial work [19], the privacy-constrained operation region $\mathcal{R}_1(\gamma_2^*)$ has been studied. It is shown that the privacy leakage suppression constraint $\check{\mathcal{R}}_E(p_{F1}, p_{D1}, \sigma) \geq \delta$ can lead to linear constraint boundaries of $\mathcal{R}_1(\gamma_2^*)$. The properties of $\mathcal{R}_1(\gamma_2^*)$ are summarized as follows.

Proposition 5. *The upper and lower boundaries of $\mathcal{R}_1(\gamma_2^*)$ are point symmetric with respect to (0.5, 0.5). If $\delta \leq 1 - \hat{\partial}\mathcal{R}_1(\sigma)$, $\mathcal{R}_1(\gamma_2^*) = \mathcal{R}_1$. If $\delta = 1 - \sigma$, $\mathcal{R}_1(\gamma_2^*)$ reduces to the line segment between (0, 0) and (1, 1). If $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$, $\mathcal{R}_1(\gamma_2^*)$ is a non-convex set.*

When $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$, a privacy-constrained operation region $\mathcal{R}_1(\gamma_2^*)$ is illustrated in Fig. 4. The linear constraint boundaries A – B – C on $\hat{\partial}\mathcal{R}_1(\gamma_2^*)$ and A' – B' – C'

on $\partial\mathcal{R}_1(\gamma_2^*)$ are determined by the parameters σ and δ as: $B = (\sigma, 1-\delta)$; $B' = (1-\sigma, \delta)$; A is an intersection point of $\partial\mathcal{R}_1$ and the line through points $B, (1, 1)$; C is an intersection point of $\partial\mathcal{R}_1$ and the line through points $B, (0, 0)$; A' is an intersection point of $\partial\mathcal{R}_1$ and the line through points $B', (0, 0)$; and C' is an intersection point of $\partial\mathcal{R}_1$ and the line through points $B', (1, 1)$. Note that an operation point on the linear constraint boundaries can be achieved by a randomized decision strategy of two LRTs, e.g., any operation point on $A - B$ can be achieved by a randomized decision strategy of the LRTs which achieve the operation points A and $(1, 1)$.

For the problem (13), we have the following theorems about γ_1^* in the privacy-constrained distributed Neyman-Pearson test network design.

Theorem 2. *If $\lambda = \sigma$, it is sufficient to consider a deterministic LRT for γ_1^* in the privacy-constrained distributed Neyman-Pearson test problem (13).*

Theorem 3. *If $\lambda \neq \sigma$, it is sufficient to consider a remote decision strategy, which achieves an operation point on $\partial\mathcal{R}_1 \cap \partial\mathcal{R}_1(\gamma_2^*)$, $A - B - C$ with p_{F1} ranging between λ and σ , $\partial\mathcal{R}_1 \cap \partial\mathcal{R}_1(\gamma_2^*)$, or $A' - B' - C'$ with p_{F1} ranging between $1 - \lambda$ and $1 - \sigma$, for γ_1^* in the privacy-constrained distributed Neyman-Pearson test problem (13).*

The proof of Theorems 2-3 is given in Appendix B. Theorem 3 implies that γ_1^* in the privacy-constrained distributed Neyman-Pearson test problem (13) with $\lambda \neq \sigma$ can be a deterministic LRT or a randomized decision strategy of two LRTs. By characterizing the sufficiency of deterministic LRT for optimality in the privacy-constrained distributed Neyman-Pearson test design of the simple eavesdropped hypothesis test network, the optimization problem (13) can be significantly simplified.

2) *Design of the General Model of Eavesdropped Hypothesis Test Network:* Here, we study the privacy-constrained distributed Neyman-Pearson test design of the eavesdropped hypothesis test network introduced in Section II:

$$\gamma_n^*, \gamma_F^* = \arg \min_{\gamma_n \in \Gamma_n, \gamma_F \in \Gamma_F} p_{MF}, \text{ s.t. } p_{FF} \leq \lambda, p_{ME}^{\min}(\sigma) \geq \delta. \quad (15)$$

Corollary 1 has shown that it is sufficient to consider a deterministic LRT for γ_j^* with $s < j \leq n$. Then, we focus on characterizing a remote decision strategy γ_i^* with $1 \leq i \leq s$.

The person-by-person optimality argument implies $\gamma_i^*, 1 \leq i \leq s$, achieves an operation point (p_{Fi}^*, p_{Di}^*) as

$$(p_{Fi}^*, p_{Di}^*) = \arg \min_{(p_{Fi}, p_{Di}) \in \mathcal{R}_i(\gamma_{ni}^*)} \check{\partial}\mathcal{R}_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, \lambda), \quad (16)$$

where

$$\mathcal{R}_i(\gamma_{ni}^*) = \mathcal{R}_i \cap \{(p_{Fi}, p_{Di}) : \check{\partial}\mathcal{R}_E(p_{Fi}, p_{Di}, \gamma_{si}^*, \sigma) \geq \delta\}.$$

Lemma 1 in Appendix A shows that there are $z(n)$ valid sorted fusion observation likelihood ratio chains given γ_{ni}^* and $z(s)$ valid sorted eavesdropping observation likelihood ratio chains given γ_{si}^* . Given a valid sorted fusion observation likelihood ratio chain with a fusion observation likelihood ratio sort function r^n and supposing that the k -th line segment consists

of an operation point with $p_{FF} = \lambda$, the objective function $\check{\partial}\mathcal{R}_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, \lambda)$ can be specified and denoted as

$$m_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, r^n, k) = 1 - c(n, k - 1, 1) - \frac{\lambda - c(n, k - 1, 0)}{p_{M^n|H}(r^n(k)|0)} p_{M^n|H}(r^n(k)|1) \quad (17)$$

with a feasible set

$$\mathcal{R}_i(\gamma_{ni}^*, r^n, k) = \left\{ (p_{Fi}, p_{Di}) : \frac{p_{M^n|H}(r^n(1)|0)}{p_{M^n|H}(r^n(1)|1)} \leq \dots \leq \frac{p_{M^n|H}(r^n(2^n)|0)}{p_{M^n|H}(r^n(2^n)|1)} \right. \\ \left. c(n, k - 1, 0) \leq \lambda, c(n, k, 0) \geq \lambda \right\}.$$

Given a valid sorted eavesdropping observation likelihood ratio chain with an eavesdropping observation likelihood ratio sort function r^s and supposing that the v -th line segment consists of an operation point with $p_{FE} = \sigma$, the function $\check{\partial}\mathcal{R}_E(p_{Fi}, p_{Di}, \gamma_{si}^*, \sigma)$ can be specified and denoted as

$$m_E(p_{Fi}, p_{Di}, \gamma_{si}^*, r^s, v) = 1 - c(s, v - 1, 1) - \frac{\sigma - c(s, v - 1, 0)}{p_{M^s|H}(r^s(v)|0)} p_{M^s|H}(r^s(v)|1) \quad (18)$$

with a feasible set

$$\mathcal{R}_i(\gamma_{si}^*, r^s, v) = \left\{ (p_{Fi}, p_{Di}) : \frac{p_{M^s|H}(r^s(1)|0)}{p_{M^s|H}(r^s(1)|1)} \leq \dots \leq \frac{p_{M^s|H}(r^s(2^s)|0)}{p_{M^s|H}(r^s(2^s)|1)} \right. \\ \left. c(s, v - 1, 0) \leq \sigma, c(s, v, 0) \geq \sigma \right\}.$$

Then, we can study the person-by-person optimization problem (16) with specified objective function and feasible set as

$$(p_{Fi}^\#, p_{Di}^\#) = \arg \min_{(p_{Fi}, p_{Di}) \in \mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v)} m_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, r^n, k), \quad (19)$$

where

$$\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v) = \mathcal{R}_i \cap \mathcal{R}_i(\gamma_{ni}^*, r^n, k) \cap \mathcal{R}_i(\gamma_{si}^*, r^s, v) \\ \cap \{(p_{Fi}, p_{Di}) : m_E(p_{Fi}, p_{Di}, \gamma_{si}^*, r^s, v) \geq \delta\}.$$

If $\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v) \neq \emptyset$, the problem (19) is feasible and the operation point $(p_{Fi}^\#, p_{Di}^\#)$ can be seen as a candidate of (p_{Fi}^*, p_{Di}^*) . Given all combinations of $(r^n, r^s, k, v) \in \mathcal{C} = \{1, \dots, z(n)\} \times \{1, \dots, z(s)\} \times \{1, \dots, 2^n\} \times \{1, \dots, 2^s\}$, the operation point (p_{Fi}^*, p_{Di}^*) in (16) can be obtained as

$$(p_{Fi}^*, p_{Di}^*) = \min_{(r^n, r^s, k, v) \in \mathcal{C}} \arg \min_{(p_{Fi}, p_{Di}) \in \mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v)} m_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, r^n, k). \quad (20)$$

There are a huge number of inner optimizations in (20). The following conclusion can be used to skip inner optimizations and to reduce the complexity of (20).

Remark 2. *If $\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v) \cap (\partial\mathcal{R}_i \cup \check{\partial}\mathcal{R}_i) = \emptyset$ and $\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v) \cap \{(p_{Fi}, p_{Di}) : m_E(p_{Fi}, p_{Di}, \gamma_{si}^*, r^s, v) = \delta\} = \emptyset$, the corresponding inner optimization can be skipped.*

Proof: Theorem 1 shows that it is sufficient to consider a remote decision strategy which achieves an operation point on the boundary of $\mathcal{R}_i(\gamma_{ni}^*)$ for γ_i^* . If the

conditions $\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v) \cap (\hat{\partial}\mathcal{R}_i \cup \check{\partial}\mathcal{R}_i) = \emptyset$ and $\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v) \cap \{(p_{Fi}, p_{Di}) : m_E(p_{Fi}, p_{Di}, \gamma_{si}^*, r^s, v) = \delta\} = \emptyset$ are satisfied, it means that no operation point in $\mathcal{R}_i(\gamma_{ni}^*, r^n, r^s, k, v)$ is on the boundary of $\mathcal{R}_i(\gamma_{ni}^*)$ which is determined by the boundary of $\mathcal{R}_i$ and the privacy leakage suppression constraint boundary $\{(p_{Fi}, p_{Di}) : \check{\partial}\mathcal{R}_E(p_{Fi}, p_{Di}, \gamma_{si}^*, \sigma) = \delta\}$. Therefore, the corresponding inner optimization can be skipped. ■

We have proposed a quantitative measure of privacy leakage and studied the privacy-constrained distributed Neyman-Pearson test problem. Based on them, a distributed hypothesis test modeled privacy-sensitive CPS application, e.g., a personal eHealth system, can be optimally designed and satisfies a concrete Neyman-Pearson test-operational privacy leakage suppression constraint. The theoretic analysis in this work characterizes the optimal remote decision strategies and shows the sufficiency of deterministic LRTs for optimality in certain cases. These results can simplify the privacy-constrained distributed Neyman-Pearson test network design of privacy-sensitive CPS applications since only the optimal test threshold needs to be found.

V. NUMERICAL STUDY

In this section, we illustrate the trade-off between hypothesis test performance and privacy leakage in privacy-constrained distributed distributed Neyman-Pearson test designs of two eavesdropped distributed hypothesis test networks which are denoted as Network 1 and Network 2. We assume the same settings for the parallel distributed binary hypothesis test network: two remote decision makers DM_1 , DM_2 and FC. The remote observation Y_1 is assumed to be distorted through a Rayleigh fading channel and the remote observation Y_2 is assumed to be distorted by an additive Gaussian noise. They are conditionally independently generated following:

$$\begin{aligned} Y_1|h=0 &\sim \text{Rayleigh}(1), \quad Y_2|h=0 \sim \mathcal{N}(1,1), \\ Y_1|h=1 &\sim \text{Rayleigh}(2), \quad Y_2|h=1 \sim \mathcal{N}(2,1). \end{aligned}$$

The two eavesdropped distributed hypothesis test networks are different in the eavesdropping settings: Network 1 with only the remote decision Y_1 intercepted and Network 2 with both of the remote decisions Y_1 , Y_2 intercepted.

In Fig. 5, it shows the trade-off between hypothesis test performance and privacy leakage of privacy-constrained distributed Neyman-Pearson test designs of Network 1 and Network 2 when the false-alarm probability upper bounds of fusion decision and eavesdropping decision are set as $\lambda = \sigma = 0.2$. The privacy-constrained distributed Neyman-Pearson test designs of Network 1 when δ is smaller than the value of the point in the cyan ellipse and Network 2 when δ is smaller than the value of the point in the black ellipse have the same hypothesis test performance measure p_{MF} which does not change with the increase of the privacy leakage measure lower bound δ . First, the distributed Neyman-Pearson test designs of Network 1 and Network 2 without a privacy leakage suppression constraint, i.e., $\delta = 0$, are the same given the same false-alarm probability upper bound of fusion decision $\lambda = 0.2$. Second, the distributed Neyman-Pearson test designs

Fig. 5. Trade-off between hypothesis test performance and privacy leakage of privacy-constrained distributed Neyman-Pearson test designs of Network 1 and Network 2. The false-alarm probability upper bounds of fusion decision and eavesdropping decision are set as $\lambda = \sigma = 0.2$.

of Network 1 and Network 2 without a privacy leakage suppression constraint lead to different values of the privacy leakage measure $p_{ME}^{\min}(\sigma)$ under the same false-alarm probability upper bound of eavesdropping decision $\sigma = 0.2$ and different eavesdropping settings, i.e., the distributed Neyman-Pearson test design without a privacy leakage suppression constraint is also a privacy-constrained distributed Neyman-Pearson test design when the privacy leakage measure lower bound δ is smaller than a certain value under the eavesdropping setting of Network 1 or Network 2. As the privacy leakage measure lower bound δ increases to be larger than the value of the point in the black ellipse, the hypothesis test performance measure p_{MF} increases linearly for the privacy-constrained distributed distributed Neyman-Pearson test designs of Network 2. Note that the values of the hypothesis test performance measure p_{MF} and the privacy leakage measure $p_{ME}^{\min}(\sigma)$ are always the same given the same false-alarm probability upper bounds of fusion decision and eavesdropping decision $\lambda = \sigma = 0.2$ for the privacy-constrained distributed Neyman-Pearson test designs of Network 2. The observation implies that we can always find a privacy-constrained distributed Neyman-Pearson test design of Network 2 such that $p_{MF} = p_{ME}^{\min}(\sigma) = \delta$ when δ is larger than the value of the point in the black ellipse. As the privacy leakage measure lower bound δ increases to be larger than the value of the point in the cyan ellipse, the hypothesis test performance measure p_{MF} increases for the privacy-constrained distributed distributed Neyman-Pearson test designs of Network 1. On the two sides of the point in the red ellipse, we can observe that the hypothesis test performance measure p_{MF} of privacy-constrained distributed distributed Neyman-Pearson test designs of Network 1 increases in different patterns. That is because the operation point (p_{F1}^*, p_{D1}^*) of the privacy-constrained distributed distributed Neyman-Pearson test design of Network 1 jumps between the two separated $\hat{\partial}\mathcal{R}_1 \cap \hat{\partial}\mathcal{R}_1(\gamma_2^*)$ segments on the upper boundary of the privacy-constrained operation region of DM_1 .

In Fig. 6, the curves correspond to three different settings on the false-alarm probability upper bounds of fusion decision and eavesdropping decision: $(\lambda = 0.1, \sigma = 0.2)$, $(\lambda = 0.2, \sigma = 0.2)$, and $(\lambda = 0.3, \sigma = 0.2)$. Given a

Fig. 6. Comparison of hypothesis test performance of privacy-constrained distributed Neyman-Pearson test designs of Network 1 and Network 2 with different settings of false-alarm probability upper bounds of fusion decision and eavesdropping decision: $(\lambda = 0.1, \sigma = 0.2)$, $(\lambda = 0.2, \sigma = 0.2)$, and $(\lambda = 0.3, \sigma = 0.2)$.

Fig. 7. Comparison of hypothesis test performance of privacy-constrained distributed Neyman-Pearson test designs of Network 1 and Network 2 with different settings of false-alarm probability upper bounds of fusion decision and eavesdropping decision: $(\lambda = 0.2, \sigma = 0.1)$, $(\lambda = 0.2, \sigma = 0.2)$, and $(\lambda = 0.2, \sigma = 0.3)$.

setting $(\lambda = 0.1, \sigma = 0.2)$ or $(\lambda = 0.3, \sigma = 0.2)$, the analysis and comparison of the privacy-constrained distributed Neyman-Pearson test designs of Network 1 and Network 2 are similar as the discussions of the case $(\lambda = 0.2, \sigma = 0.2)$. Here, we focus on the comparison of privacy-constrained distributed Neyman-Pearson test designs of Network 1 or Network 2 given different settings of λ . Consider the curves corresponding to the privacy-constrained distributed Neyman-Pearson test designs of Network 1. By fixing a value for δ , the hypothesis test performance measure p_{MF} is larger given a smaller value of λ and the fixed value of $\sigma = 0.2$. That is because a smaller value of λ leads to a stronger upper bound constraint on the false-alarm probability of fusion decision and further leads to a privacy-constrained distributed Neyman-Pearson test design of Network 1 with a worse hypothesis test performance when the other constraints are the same. We have a similar observation of privacy-constrained distributed Neyman-Pearson test designs of Network 2 given different settings of λ .

In Fig. 7, the curves correspond to three different settings on the false-alarm probability upper bounds of fusion

decision and eavesdropping decision: $(\lambda = 0.2, \sigma = 0.1)$, $(\lambda = 0.2, \sigma = 0.2)$, and $(\lambda = 0.2, \sigma = 0.3)$. Given a setting $(\lambda = 0.2, \sigma = 0.1)$ or $(\lambda = 0.2, \sigma = 0.3)$, the analysis and comparison of the privacy-constrained distributed Neyman-Pearson test designs of Network 1 and Network 2 are similar as the discussions of the case $(\lambda = 0.2, \sigma = 0.2)$. Here, we focus on the comparison of privacy-constrained distributed Neyman-Pearson test designs of Network 1 or Network 2 given different settings of σ . Consider the curves corresponding to the privacy-constrained distributed Neyman-Pearson test designs of Network 1. By fixing a value for δ , the hypothesis test performance measure p_{MF} is larger corresponding to a larger value of σ and the fixed value of $\lambda = 0.2$. That is because a larger value of σ leads to a weaker upper bound constraint on the false-alarm probability of eavesdropping decision, or equivalently a stronger privacy leakage suppression constraint, and further leads to a privacy-constrained distributed Neyman-Pearson test design of Network 1 with a worse hypothesis test performance when the other constraints are the same. We have a similar observation of privacy-constrained distributed Neyman-Pearson test designs of Network 2 given different settings of σ .

In Fig. 5, Fig. 6, and Fig. 7, the hypothesis test performance of the privacy-constrained distributed Neyman-Pearson test design of Network 1 is always better than or equal to that of the privacy-constrained distributed Neyman-Pearson test design of Network 2 given the same settings of false-alarm probability upper bounds λ, σ and privacy leakage measure lower bound δ . This observation is reasonable since EVE has less information in Network 1.

VI. CONCLUSION

In this paper, the privacy-constrained distributed Neyman-Pearson test design of an eavesdropped binary hypothesis test network is studied. The proposed Neyman-Pearson test-operational privacy leakage measure provides a concrete assessment of the privacy leakage and leads to an operational privacy leakage suppression guarantee of the privacy-constrained distributed Neyman-Pearson test network design. The decision strategies in a privacy-constrained distributed Neyman-Pearson test network design are characterized: for a remote decision maker whose decision is not intercepted, it is sufficient to consider a deterministic LRT for its decision strategy; for an intercepted remote decision maker, it is sufficient to consider a decision strategy which achieves an operation point on the boundary of the privacy-constrained operation region. For the considered simple eavesdropped hypothesis test network, we show that it is sufficient to consider a deterministic LRT for the intercepted remote decision maker if the same false-alarm probability upper bound is applied to Neyman-Pearson tests of FC and EVE. If the false-alarm probability upper bounds are different, we also identify the non-LRT candidates for the optimal remote decision strategy. These results help to simplify the design of a privacy-constrained distributed Neyman-Pearson test network. We believe that the privacy-constrained distributed Neyman-Pearson test design is an interesting approach for trustworthy CPS applications.

APPENDIX A
PROOF OF THEOREM 1

In a privacy-constrained distributed Neyman-Pearson test network, the person-by-person optimality argument in Remark 1 shows that the remote decision strategy γ_i^* has to satisfy the properties of the person-by-person optimal remote decision strategy given other remote decision strategies γ_{ni}^* .

In the person-by-person optimization problem (12), the objective function of (p_{Fi}, p_{Di}) , $\partial \mathcal{R}_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, \lambda)$, is not specified. To specify the objective function, a sorted fusion observation likelihood ratio chain and the line segment which consists of an operation point with $p_{FF} = \lambda$ need to be fixed.

Note that not all sorted fusion observation likelihood ratio chains are feasible given remote decision strategies γ_{ni}^* . Here, we define valid sorted fusion observation likelihood ratio chains such that any non-valid sorted fusion observation likelihood ratio chain must be infeasible. Therefore, only the valid sorted fusion observation likelihood ratio chains need to be considered in the following discussion.

Given remote decision strategies γ_{ni}^* , a sorted likelihood ratio chain can be specified as

$$\frac{p_{M_i^n|H}(r_i^n(1)|0)}{p_{M_i^n|H}(r_i^n(1)|1)} \leq \dots \leq \frac{p_{M_i^n|H}(r_i^n(2^{n-1})|0)}{p_{M_i^n|H}(r_i^n(2^{n-1})|1)}.$$

Then, the 2^n fusion observation likelihood ratios satisfy the following two inequality chains:

$$\begin{aligned} \frac{p_{M_i^n|H}(r_i^n(1), 0|0)}{p_{M_i^n|H}(r_i^n(1), 0|1)} &\leq \dots \leq \frac{p_{M_i^n|H}(r_i^n(2^{n-1}), 0|0)}{p_{M_i^n|H}(r_i^n(2^{n-1}), 0|1)}, \\ \frac{p_{M_i^n|H}(r_i^n(1), 1|0)}{p_{M_i^n|H}(r_i^n(1), 1|1)} &\leq \dots \leq \frac{p_{M_i^n|H}(r_i^n(2^{n-1}), 1|0)}{p_{M_i^n|H}(r_i^n(2^{n-1}), 1|1)}. \end{aligned} \quad (21)$$

If $p_{Fi} \leq p_{Di}$, it means the inequality $\frac{p_{M_i^n|H}(1|0)}{p_{M_i^n|H}(1|1)} \leq \frac{p_{M_i^n|H}(0|0)}{p_{M_i^n|H}(0|1)}$ and the inequalities

$$\frac{p_{M_i^n|H}(r_i^n(k), 1|0)}{p_{M_i^n|H}(r_i^n(k), 1|1)} \leq \frac{p_{M_i^n|H}(r_i^n(k), 0|0)}{p_{M_i^n|H}(r_i^n(k), 0|1)}, \quad \forall 1 \leq k \leq 2^{n-1}. \quad (22)$$

If $p_{Fi} \geq p_{Di}$, it means the inequality $\frac{p_{M_i^n|H}(0|0)}{p_{M_i^n|H}(0|1)} \leq \frac{p_{M_i^n|H}(1|0)}{p_{M_i^n|H}(1|1)}$ and the inequalities

$$\frac{p_{M_i^n|H}(r_i^n(k), 0|0)}{p_{M_i^n|H}(r_i^n(k), 0|1)} \leq \frac{p_{M_i^n|H}(r_i^n(k), 1|0)}{p_{M_i^n|H}(r_i^n(k), 1|1)}, \quad \forall 1 \leq k \leq 2^{n-1}. \quad (23)$$

Given remote decision strategies γ_{ni}^* , a sorted fusion observation likelihood ratio chain is defined to be valid if it satisfies the inequalities in (21), (22) or it satisfies the inequalities in (21), (23).

Lemma 1. *Given remote decision strategies γ_{ni}^* , there are*

$$z(n) = 2 + \sum_{j=2}^{2^{n-1}} \sum_{l_1+\dots+l_j=2^{n-1}} \sum_{v_1=1}^{l_1} \dots \sum_{v_{j-1}=1}^{l_{j-1}+\dots+l_{j-1}-v_1-\dots-v_{j-2}} 2$$

valid sorted fusion observation likelihood ratio chains.

Proof: Given remote decision strategies γ_{ni}^* , a valid sorted fusion observation likelihood ratio chain, which satisfies the inequalities in (21), (22), can be constructed as follows.

Step 1: Set an integer j , $1 \leq j \leq 2^{n-1}$.

Step 2: Generate two sets of positive integers $(l_1, \dots, l_j)$, $(v_1, \dots, v_j) \in \mathbb{Z}^{+j}$ such that

$$\begin{aligned} \sum_{t=1}^k l_t &\geq \sum_{t=1}^k v_t, \quad \forall 1 \leq k \leq j-1, \\ \sum_{t=1}^j l_t &= \sum_{t=1}^j v_t = 2^{n-1}. \end{aligned} \quad (24)$$

Step 3: Divide the sorted fusion observation sequence realizations $(r_i^n(1), 1), \dots, (r_i^n(2^{n-1}), 1)$ into j segments. The length of the k -th segment is l_k , i.e., the k -th segment is $(r_i^n(1 + \sum_{t=1}^{k-1} l_t), 1), \dots, (r_i^n(\sum_{t=1}^k l_t), 1)$.

Step 4: Divide the sorted fusion observation sequence realizations $(r_i^n(1), 0), \dots, (r_i^n(2^{n-1}), 0)$ into j segments. The length of the k -th segment is v_k , i.e., the k -th segment is $(r_i^n(1 + \sum_{t=1}^{k-1} v_t), 0), \dots, (r_i^n(\sum_{t=1}^k v_t), 0)$.

Step 5: Generate a valid sorted fusion observation likelihood ratio chain

$$\frac{p_{M_i^n|H}(r_i^n(1)|0)}{p_{M_i^n|H}(r_i^n(1)|1)} \leq \dots \leq \frac{p_{M_i^n|H}(r_i^n(2^n)|0)}{p_{M_i^n|H}(r_i^n(2^n)|1)}$$

where the sorted fusion observation sequence realizations $r^n(1), \dots, r^n(2^n)$ can be divided into $2 \cdot j$ segments such that the $(2 \cdot k - 1)$ -th segment is the k -th segment of $(r_i^n(1), 1), \dots, (r_i^n(2^{n-1}), 1)$ and the $(2 \cdot k)$ -th segment is the k -th segment of $(r_i^n(1), 0), \dots, (r_i^n(2^{n-1}), 0)$.

The construction method shows that a division manner generated in Steps 1-2 corresponds to a unique valid sorted fusion observation likelihood ratio chain. Given $j = 1$, there is only one division manner $l_1 = v_1 = 2^{n-1}$. Given $2 \leq j \leq 2^{n-1}$ and a set of positive integers $(l_1, \dots, l_j)$ such that $\sum_{t=1}^j l_t = 2^{n-1}$, there are $\sum_{v_1=1}^{l_1} \dots \sum_{v_{j-1}=1}^{l_{j-1}+\dots+l_{j-1}-v_1-\dots-v_{j-2}} 1$ sets of positive integers $(v_1, \dots, v_j)$ which satisfy the constraints in (24). Therefore, there are

$$1 + \sum_{j=2}^{2^{n-1}} \sum_{l_1+\dots+l_j=2^{n-1}} \sum_{v_1=1}^{l_1} \dots \sum_{v_{j-1}=1}^{l_{j-1}+\dots+l_{j-1}-v_1-\dots-v_{j-2}} 1$$

valid sorted fusion observation likelihood ratio chains satisfying the inequalities in (21), (22).

A valid sorted fusion observation likelihood ratio chain satisfying the inequalities in (21), (23) can be constructed using a similar method as follows.

Step 1: Set an integer j , $1 \leq j \leq 2^{n-1}$.

Step 2: Generate two sets of positive integers $(l_1, \dots, l_j)$, $(v_1, \dots, v_j) \in \mathbb{Z}^{+j}$ satisfying the constraints in (24).

Step 3: Divide the sorted fusion observation sequence realizations $(r_i^n(1), 0), \dots, (r_i^n(2^{n-1}), 0)$ into j segments. The length of the k -th segment is l_k , i.e., the k -th segment is $(r_i^n(1 + \sum_{t=1}^{k-1} l_t), 0), \dots, (r_i^n(\sum_{t=1}^k l_t), 0)$.

Step 4: Divide the sorted fusion observation sequence realizations $(r_i^n(1), 1), \dots, (r_i^n(2^{n-1}), 1)$ into j segments. The length of the k -th segment is v_k , i.e., the k -th segment is $(r_i^n(1 + \sum_{t=1}^{k-1} v_t), 1), \dots, (r_i^n(\sum_{t=1}^k v_t), 1)$.

Step 5: Generate a valid sorted fusion observation likelihood ratio chain

$$\frac{p_{M_i^n|H}(r_i^n(1)|0)}{p_{M_i^n|H}(r_i^n(1)|1)} \leq \dots \leq \frac{p_{M_i^n|H}(r_i^n(2^n)|0)}{p_{M_i^n|H}(r_i^n(2^n)|1)}$$

where the sorted fusion observation sequence realizations $r^n(1), \dots, r^n(2^n)$ can be divided into $2 \cdot j$ segments such that the $(2 \cdot k - 1)$ -th segment is the k -th segment of $(r_i^n(1), 0), \dots, (r_i^n(2^{n-1}), 0)$ and the $(2 \cdot k)$ -th segment is the k -th segment of $(r_i^n(1), 1), \dots, (r_i^n(2^{n-1}), 1)$.

Based on the construction method, it can be shown that there are

$$1 + \sum_{j=2}^{2^{n-1}} \sum_{l_1+\dots+l_j=2^{n-1}} \sum_{v_1=1}^{l_1} \dots \sum_{v_{j-1}=1}^{l_1+\dots+l_{j-1}-v_1-\dots-v_{j-2}} 1$$

valid sorted fusion observation likelihood ratio chains satisfy the inequalities in (21), (23).

Therefore, it can be concluded that there are

$$2 + \sum_{j=2}^{2^{n-1}} \sum_{l_1+\dots+l_j=2^{n-1}} \sum_{v_1=1}^{l_1} \dots \sum_{v_{j-1}=1}^{l_1+\dots+l_{j-1}-v_1-\dots-v_{j-2}} 2$$

valid sorted fusion observation likelihood ratio chains given remote decision strategies γ_{ni}^* . ■

Lemma 2. *Given remote decision strategies γ_{ni}^* , a valid sorted fusion observation likelihood ratio chain, a constant coefficient $a \in [0, 1]$, and an integer $1 \leq k \leq 2^n$, the linear function of p_{Di} , $-c(n, k - 1, 1) - a \cdot p_{M^n|H}(r^n(k)|1)$, is non-increasing if the valid sorted fusion observation likelihood ratio chain satisfies the constraint $p_{Fi} \leq p_{Di}$ or is non-decreasing if the valid sorted fusion observation likelihood ratio chain satisfies the constraint $p_{Fi} \geq p_{Di}$.*

Proof: First consider the case that the given valid sorted fusion observation likelihood ratio chain satisfies the constraint $p_{Fi} \leq p_{Di}$. Note that $p_{Fi} \leq p_{Di}$ means the inequalities in (22). Therefore, the given valid sorted fusion observation likelihood ratio chain under the constraint $p_{Fi} \leq p_{Di}$ satisfies the inequalities in (21), (22). Based on the construction method of a valid sorted fusion observation likelihood ratio chain satisfying the inequalities in (21), (22) in the proof of Lemma 1, it can be shown that the coefficient of the variable p_{Di} in the linear function $-c(n, k - 1, 1) - a \cdot p_{M^n|H}(r^n(k)|1)$ is non-positive, i.e., $-c(n, k - 1, 1) - a \cdot p_{M^n|H}(r^n(k)|1)$ is a non-increasing linear function of p_{Di} if the given valid sorted fusion observation likelihood ratio chain satisfies the constraint $p_{Fi} \leq p_{Di}$. Similarly, it can be shown that $-c(n, k - 1, 1) - a \cdot p_{M^n|H}(r^n(k)|1)$ is a non-decreasing linear function of p_{Di} if the given valid sorted fusion observation likelihood ratio chain satisfies the constraint $p_{Fi} \geq p_{Di}$. ■

Based on Lemma 2, the proof of Theorem 1 is given in the following.

Proof: The optimization problem (12) can be reduced to different problems corresponding to different valid sorted fusion observation likelihood ratio chains or different line segments consisting of an operation point with $p_{FF} = \lambda$.

Given a valid sorted fusion observation likelihood ratio chain under the constraint $p_{Fi} \leq p_{Di}$, the 2^n sorted line segments can be specified in terms of p_{Fi} and p_{Di} . Suppose that the k -th line segment consists of an operation point with

$p_{FF} = \lambda$. Then, the objective function in (12) reduces to

$$\begin{aligned} \partial \mathcal{R}_F(p_{Fi}, p_{Di}, \gamma_{ni}^*, \lambda) = & 1 - c(n, k - 1, 1) - \\ & \frac{\lambda - c(n, k - 1, 0)}{p_{M^n|H}(r^n(k)|0)} p_{M^n|H}(r^n(k)|1) \end{aligned} \quad (25)$$

where $c(n, k - 1, 0)$, $p_{M^n|H}(r^n(k)|0)$ are linear functions of p_{Fi} , and $c(n, k - 1, 1)$, $p_{M^n|H}(r^n(k)|1)$ are linear functions of p_{Di} . The corresponding feasible set of the reduced optimization problem is a subset of the upper part of $\mathcal{R}_i(\gamma_{ni}^*)$ above the line $p_{Fi} = p_{Di}$ and consists of all operation points of (p_{Fi}, p_{Di}) which satisfy the constraints introduced by the given valid sorted fusion observation likelihood ratio chain and the k -th line segment having an operation point with $p_{FF} = \lambda$.

Fixing a feasible value for p_{Fi} , the term $\frac{\lambda - c(n, k - 1, 0)}{p_{M^n|H}(r^n(k)|0)}$ is a constant in the range of $[0, 1]$ because of the constraint that the k -th line segment consists of an operation point with $p_{FF} = \lambda$. Then, the objective function further reduces to a linear function of p_{Di} . Based on Lemma 2, the linear objective function of p_{Di} is non-increasing, i.e., by fixing a feasible value for p_{Fi} , it is sufficient to maximize p_{Di} in order to minimize the non-increasing linear objective. Note that this result applies to all reduced optimization problems corresponding to different valid sorted fusion observation likelihood ratio chains under the constraint $p_{Fi} \leq p_{Di}$ or different line segments consisting of an operation point with $p_{FF} = \lambda$. Therefore, it is sufficient to consider a decision strategy which achieves an upper boundary point of $\mathcal{R}_i(\gamma_{ni}^*)$ to minimize the objective of (12) if only the upper part of $\mathcal{R}_i(\gamma_{ni}^*)$ above the line $p_{Fi} = p_{Di}$ is considered.

Given a valid sorted fusion observation likelihood ratio chain under the constraint $p_{Fi} \geq p_{Di}$, supposing that the k -th line segment consists of an operation point with $p_{FF} = \lambda$, and fixing a feasible value for p_{Fi} , a similar analysis shows that the objective function of (12) reduces to a non-decreasing linear function of p_{Di} , i.e., by fixing a feasible value for p_{Fi} , it is sufficient to minimize p_{Di} in order to minimize the objective. Therefore, it is sufficient to consider a decision strategy which achieves a lower boundary point of $\mathcal{R}_i(\gamma_{ni}^*)$ to minimize the objective of (12) if only the lower part of $\mathcal{R}_i(\gamma_{ni}^*)$ below the line $p_{Fi} = p_{Di}$ is considered.

Then, we can conclude the property of γ_i^* in Theorem 1. ■

APPENDIX B PROOF OF THEOREMS 2-3

Theorem 1 shows that it is sufficient to consider a remote decision strategy which achieves an operation point on the boundary of $\mathcal{R}_1(\gamma_2^*)$ for γ_1^* . Due to the privacy leakage suppression constraint $\partial \mathcal{R}_E(p_{F1}, p_{D1}, \sigma) \geq \delta$, boundary of $\mathcal{R}_1(\gamma_2^*)$ may consist of linear constraint boundary segments other than $\partial \mathcal{R}_1 \cap \partial \mathcal{R}_1(\gamma_2^*)$ and $\partial \mathcal{R}_1 \cap \partial \mathcal{R}_1(\gamma_2^*)$ segments. Therefore, the following proof of Theorem 2 is to show that it is sufficient to consider a deterministic LRT which achieves an operation point on $\partial \mathcal{R}_1 \cap \partial \mathcal{R}_1(\gamma_2^*)$ or $\partial \mathcal{R}_1 \cap \partial \mathcal{R}_1(\gamma_2^*)$ boundary segments of $\mathcal{R}_1(\gamma_2^*)$ for γ_1^* when $\lambda = \sigma$.

Proof: If $\delta \leq 1 - \partial \mathcal{R}_1(\sigma)$, Proposition 5 shows that $\mathcal{R}_1(\gamma_2^*) = \mathcal{R}_1$. In this case, it is sufficient to consider a

deterministic LRT which achieves an operation point on the boundary of $\mathcal{R}_1$ for γ_1^* .

If $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$, there are 4 linear constraint boundary segments as shown in Fig. 4: $A - B$, $B - C$ on the upper boundary, $A' - B'$, $B' - C'$ on the lower boundary. Let us consider the optimization problems of (p_{F1}, p_{D1}) over linear constraint boundary segments. Then, we have the following results under the constraint $\lambda = \sigma$:

$$\begin{aligned} A &= \arg \min_{(p_{F1}, p_{D1}) \in A-B} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C &= \arg \min_{(p_{F1}, p_{D1}) \in B-C} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ A' &= \arg \min_{(p_{F1}, p_{D1}) \in A'-B'} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C' &= \arg \min_{(p_{F1}, p_{D1}) \in B'-C'} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda). \end{aligned} \quad (26)$$

Here, we present the analysis on the first two optimizations in (26). Since the optimization of (p_{F1}, p_{D1}) is over the linear segments $A - B$ and $B - C$ on the upper boundary of $\mathcal{R}_1(\gamma_2^*)$, there is a constraint $p_{F1} \leq p_{D1}$ in the discussion. The form of the objective $\check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda)$ depends on the sorted fusion observation likelihood ratio chain and the line segment which consists of an operation point with $p_{FF} = \lambda$. If the remote decision strategy γ_2^* achieves an operation point (p_{F2}^*, p_{D2}^*) with $p_{F2}^* \leq p_{D2}^*$, the construction method in the proof of Lemma 1 leads to $\frac{z(2)}{2} = 2$ valid sorted fusion observation likelihood ratio chains under the constraint $p_{F1} \leq p_{D1}$:

$$\frac{p_{M^2|H}(1, 1|0)}{p_{M^2|H}(1, 1|1)} \leq \frac{p_{M^2|H}(1, 0|0)}{p_{M^2|H}(1, 0|1)} \leq \frac{p_{M^2|H}(0, 1|0)}{p_{M^2|H}(0, 1|1)} \leq \frac{p_{M^2|H}(0, 0|0)}{p_{M^2|H}(0, 0|1)}$$

which can be expressed in terms of p_{F1} , p_{D1} , p_{F2}^* , and p_{D2}^* as

$$\frac{p_{F1}p_{F2}^*}{p_{D1}p_{D2}^*} \leq \frac{p_{F1}(1 - p_{F2}^*)}{p_{D1}(1 - p_{D2}^*)} \leq \frac{(1 - p_{F1})p_{F2}^*}{(1 - p_{D1})p_{D2}^*} \leq \frac{(1 - p_{F1})(1 - p_{F2}^*)}{(1 - p_{D1})(1 - p_{D2}^*)},$$

and

$$\frac{p_{M^2|H}(1, 1|0)}{p_{M^2|H}(1, 1|1)} \leq \frac{p_{M^2|H}(0, 1|0)}{p_{M^2|H}(0, 1|1)} \leq \frac{p_{M^2|H}(1, 0|0)}{p_{M^2|H}(1, 0|1)} \leq \frac{p_{M^2|H}(0, 0|0)}{p_{M^2|H}(0, 0|1)}$$

which can be expressed in terms of p_{F1} , p_{D1} , p_{F2}^* , and p_{D2}^* as

$$\frac{p_{F1}p_{F2}^*}{p_{D1}p_{D2}^*} \leq \frac{(1 - p_{F1})p_{F2}^*}{(1 - p_{D1})p_{D2}^*} \leq \frac{p_{F1}(1 - p_{F2}^*)}{p_{D1}(1 - p_{D2}^*)} \leq \frac{(1 - p_{F1})(1 - p_{F2}^*)}{(1 - p_{D1})(1 - p_{D2}^*)}.$$

For each valid sorted fusion observation likelihood ratio chain, depending on the line segment which consists of an operation point with $p_{FF} = \lambda$, the objective $\check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda)$ reduces to a function of p_{F1} and p_{D1} as specified in (25). The corresponding feasible set is a subset of the upper part of $\mathcal{R}_1(\gamma_2^*)$ above the line $p_{F1} = p_{D1}$ and consists of operation points of (p_{F1}, p_{D1}) which satisfy the constraints introduced by the given valid sorted fusion observation likelihood ratio chain and the given line segment having an operation point with $p_{FF} = \lambda$. An important observation of the reduced objective functions is:

- 1) Given a reduced objective function whose feasible set of (p_{F1}, p_{D1}) has a non-empty intersection with the set $\{(p_{F1}, p_{D1}) : 0 \leq p_{F1} \leq \lambda\}$, the reduced objective is a non-decreasing function of p_{F1} along any line through

points $(1, 1)$ and (λ, x) with $p_{F1} \leq \lambda$ and $\lambda < x < \hat{\partial}\mathcal{R}_1(\lambda)$;

- 2) Given a reduced objective function whose feasible set of (p_{F1}, p_{D1}) has a non-empty intersection with the set $\{(p_{F1}, p_{D1}) : \lambda \leq p_{F1} \leq 1\}$, the reduced objective is a non-increasing function of p_{F1} along any line through points $(0, 0)$ and (λ, y) with $p_{F1} \geq \lambda$ and $\lambda < y < \hat{\partial}\mathcal{R}_1(\lambda)$.

Note the condition $\lambda = \sigma$. Then, any reduced objective function whose feasible set of (p_{F1}, p_{D1}) has a non-empty intersection with the set $\{(p_{F1}, p_{D1}) : 0 \leq p_{F1} \leq \lambda\}$ is a non-decreasing function of p_{F1} along the linear constraint boundary segment $A - B$; and any reduced objective function whose feasible set of (p_{F1}, p_{D1}) has a non-empty intersection with the set $\{(p_{F1}, p_{D1}) : \lambda \leq p_{F1} \leq 1\}$ is a non-increasing function of p_{F1} along the linear constraint boundary segment $B - C$. Then, we can conclude the first two equations in (26). If the remote decision strategy γ_2^* achieves an operation point with $p_{F2}^* \geq p_{D2}^*$, we have the same observation of the reduced objective functions and conclusion. The above analysis on the first two optimizations in (26) verifies the first two equations.

Let us consider the last two problems in (26). Since the optimization of (p_{F1}, p_{D1}) is over the linear segments $A' - B'$ and $B' - C'$ on the lower boundary of $\mathcal{R}_1(\gamma_2^*)$, there is a constraint $p_{F1} \geq p_{D1}$ in the discussion. Given the remote decision strategy γ_2^* , a valid sorted fusion observation likelihood ratio chain satisfying the constraint $p_{F1} \geq p_{D1}$, and a line segment which consists of an operation point with $p_{FF} = \lambda$, the objective function $\check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda)$ reduces to a function of p_{F1} and p_{D1} as specified in (25) with a corresponding feasible set of (p_{F1}, p_{D1}) . Then, we have an observation of these reduced objective functions as:

- 1) Given a reduced objective function whose feasible set of (p_{F1}, p_{D1}) has a non-empty intersection with the set $\{(p_{F1}, p_{D1}) : 0 \leq p_{F1} \leq 1 - \lambda\}$, the reduced objective is a non-decreasing function of p_{F1} along any line through points $(1, 1)$ and $(1 - \lambda, x)$ with $p_{F1} \leq 1 - \lambda$ and $\hat{\partial}\mathcal{R}_1(\lambda) < x < 1 - \lambda$;
- 2) Given a reduced objective function whose feasible set of (p_{F1}, p_{D1}) has a non-empty intersection with the set $\{(p_{F1}, p_{D1}) : 1 - \lambda \leq p_{F1} \leq 1\}$, the reduced objective is a non-increasing function of p_{F1} along any line through points $(0, 0)$ and $(1 - \lambda, y)$ with $p_{F1} \geq 1 - \lambda$ and $\hat{\partial}\mathcal{R}_1(\lambda) < y < 1 - \lambda$.

Following a similar analysis, we can verify the last two equations in (26).

Note that points A , C , A' , and C' are also on $\hat{\partial}\mathcal{R}_1$ or $\check{\partial}\mathcal{R}_1$. Therefore, we can conclude that it is sufficient to consider a deterministic LRT which achieves an operation point on $\hat{\partial}\mathcal{R}_1 \cap \check{\partial}\mathcal{R}_1(\gamma_2^*)$ or $\check{\partial}\mathcal{R}_1 \cap \hat{\partial}\mathcal{R}_1(\gamma_2^*)$ boundary segments of $\mathcal{R}_1(\gamma_2^*)$ for γ_1^* under the constraint $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$.

If $\delta = 1 - \sigma$, Proposition 5 shows that $\mathcal{R}_1(\gamma_2^*)$ reduces to the line segment between $(0, 0)$ and $(1, 1)$. Optimization of remote decision strategy over the line segment between $(0, 0)$ and $(1, 1)$ can be seen as the problems in (26) where the points are specified as $B = (\sigma, \sigma)$, $A = C' = (0, 0)$, $A' = C = (1, 1)$, and $B' = (1 - \sigma, 1 - \sigma)$. Therefore, it is sufficient to consider a

Fig. 8. Illustration of a privacy-constrained operation region $\mathcal{R}_1(\gamma_2^*)$ when $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$ and $\lambda > \sigma$. The operation point D is the intersection of the linear constraint boundary segment B – C and the line $p_{F1} = \lambda$. The operation point D' is the intersection of the linear constraint boundary segment B' – C' and the line $p_{F1} = 1 - \lambda$. The red line segments show the linear constraint boundary A – B – C with p_{F1} ranging between σ and λ and the linear constraint boundary A' – B' – C' with p_{F1} ranging between $1 - \lambda$ and $1 - \sigma$.

Fig. 9. Illustration of a privacy-constrained operation region $\mathcal{R}_1(\gamma_2^*)$ when $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$ and $\lambda < \sigma$. The operation point D is the intersection of the linear constraint boundary segment A – B and the line $p_{F1} = \lambda$. The operation point D' is the intersection of the linear constraint boundary segment A' – B' and the line $p_{F1} = 1 - \lambda$. The red line segments show the linear constraint boundary A – B – C with p_{F1} ranging between λ and σ and the linear constraint boundary A' – B' – C' with p_{F1} ranging between $1 - \sigma$ and $1 - \lambda$.

deterministic LRT which achieves an end of the line segment, $(0, 0)$ or $(1, 1)$, for γ_1^* . ■

The following proof of Theorem 3 is based on the observation of reduced objective functions in the proof of Theorem 2.

Proof: If $\delta \leq 1 - \hat{\partial}\mathcal{R}_1(\sigma)$, Proposition 5 shows that $\mathcal{R}_1(\gamma_2^*) = \mathcal{R}_1$. In this case, it is sufficient to consider a deterministic LRT which achieves an operation point on $\hat{\partial}\mathcal{R}_1$ or $\check{\partial}\mathcal{R}_1$ for γ_1^* .

If $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$, there are 4 linear constraint boundary segments as shown in Fig. 4. Let us consider the optimization problems of (p_{F1}, p_{D1}) over these linear constraint boundary segments. The observation of reduced objective functions in the proof of Theorem 2 implies the following results.

In the case shown in Fig. 8 (a) with $\lambda > \sigma$,

$$\begin{aligned} A &= \arg \min_{(p_{F1}, p_{D1}) \in A-B} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C &= \arg \min_{(p_{F1}, p_{D1}) \in D-C} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ A' &= \arg \min_{(p_{F1}, p_{D1}) \in A'-B'} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C' &= \arg \min_{(p_{F1}, p_{D1}) \in D'-C'} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \end{aligned}$$

where the operation point D is the intersection of the linear constraint boundary segment B – C and the line $p_{F1} = \lambda$; and the operation point D' is the intersection of the linear constraint boundary segment B' – C' and the line $p_{F1} = 1 - \lambda$. In the case shown in Fig. 8 (b) with $\lambda < \sigma$,

$$\begin{aligned} A &= \arg \min_{(p_{F1}, p_{D1}) \in A-B} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ A' &= \arg \min_{(p_{F1}, p_{D1}) \in A'-B'} \check{\partial}\mathcal{R}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda). \end{aligned}$$

In the case shown in Fig. 9 (a) with $\lambda < \sigma$,

$$\begin{aligned} A &= \arg \min_{(p_{F1}, p_{D1}) \in A-D} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C &= \arg \min_{(p_{F1}, p_{D1}) \in B-C} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ A' &= \arg \min_{(p_{F1}, p_{D1}) \in A'-D'} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C' &= \arg \min_{(p_{F1}, p_{D1}) \in B'-C'} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \end{aligned}$$

where the operation point D is the intersection of the linear constraint boundary segment A – B and the line $p_{F1} = \lambda$; and the operation point D' is the intersection of the linear constraint boundary segment A' – B' and the line $p_{F1} = 1 - \lambda$. In the case shown in Fig. 9 (b) with $\lambda < \sigma$,

$$\begin{aligned} C &= \arg \min_{(p_{F1}, p_{D1}) \in B-C} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda), \\ C' &= \arg \min_{(p_{F1}, p_{D1}) \in B'-C'} \check{\mathcal{R}}_F(p_{F1}, p_{D1}, \gamma_2^*, \lambda). \end{aligned}$$

Then, we can conclude that it is sufficient to consider a remote decision strategy which achieves an operation point on $\hat{\partial}\mathcal{R}_1 \cap \hat{\partial}\mathcal{R}_1(\gamma_2^*)$, or linear constraint boundary A – B – C with p_{F1} ranging between λ and δ , or $\hat{\partial}\mathcal{R}_1 \cap \hat{\partial}\mathcal{R}_1(\gamma_2^*)$, or linear constraint boundary A' – B' – C' with p_{F1} ranging between $1 - \lambda$ and $1 - \delta$ for γ_1^* under the constraint $1 - \hat{\partial}\mathcal{R}_1(\sigma) < \delta < 1 - \sigma$.

If $\delta = 1 - \sigma$, Proposition 5 shows that $\mathcal{R}_1(\gamma_2^*)$ reduces to the line segment between (0, 0) and (1, 1). Based on the observation of reduced objective functions, we can conclude that it is sufficient to consider a deterministic LRT which achieves an end of the line segment, (0, 0) or (1, 1), for γ_1^* . ■

REFERENCES

- [1] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag New York, Inc., 1996.
- [2] J. N. Tsitsiklis, "Decentralized detection," in *Proceedings of Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.
- [3] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II - Advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.
- [4] S. Marano, V. Matta, and P. K. Willett, "Distributed detection with censoring sensors under physical layer secrecy," *Signal Processing, IEEE Transactions on*, vol. 57, no. 5, pp. 1976–1986, 2009.
- [5] V. S. S. Nadendla, H. Chen, and P. K. Varshney, "Secure distributed detection in the presence of eavesdroppers," in *Proceedings of ASIOMAR 2010*, 2010, pp. 1437–1441.
- [6] M. Mhanna and P. Piantanida, "On secure distributed hypothesis testing," in *Proceedings of ISIT 2015*, 2015, pp. 1605–1609.
- [7] V. Nadendla, "Secure distributed detection in wireless sensor networks via encryption of sensor decision," Master's thesis, Louisiana State University, 2009.
- [8] R. Soosahabi, M. Naraghi-Pour, D. Perkins, and M. A. Bayoumi, "Optimal probabilistic encryption for secure detection in wireless sensor networks," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 3, pp. 375–385, 2014.
- [9] H. Jeon, D. Hwang, J. Choi, H. Lee, and J. Ha, "Secure type-based multiple access," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 763–774, 2011.
- [10] H. Jeon, J. Choi, S. W. McLaughlin, and J. Ha, "Channel aware encryption and decision fusion for wireless sensor networks," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 4, pp. 619–625, 2013.
- [11] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Optimal distributed detection in the presence of Byzantines," in *Proceedings of ICASSP 2013*, 2013, pp. 2925–2929.
- [12] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attacks," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 16–29, 2009.
- [13] Z. Li and T. J. Oechtering, "Privacy-aware distributed Bayesian detection," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, no. 7, pp. 1345–1357, 2015.
- [14] C. Dwork, "Differential privacy," in *Proceedings of ICALP 2006*. Springer, 2006, pp. 1–12.
- [15] S. Oh and P. Viswanath, "The composition theorem for differential privacy," eprint arXiv:1311.0776.
- [16] J. Le Ny and G. J. Pappas, "Differentially private filtering," *Automatic Control, IEEE Transactions on*, vol. 59, no. 2, pp. 341–354, 2014.
- [17] Z. Ji, Z. C. Lipton, and C. Elkan, "Differential privacy and machine learning: A survey and review," eprint arXiv:1412.7584.
- [18] R. Foygel Barber and J. C. Duchi, "Privacy and statistical risk: Formalisms and minimax bounds," eprint arXiv:1412.4451.
- [19] Z. Li, T. J. Oechtering, and J. Jaldén, "Parallel distributed Neyman-Pearson detection with privacy constraints," in *Proceedings of ICC 2014 Workshop*, 2014, pp. 765–770.
- [20] J. F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, 2007.
- [21] S. Verdú, "On channel capacity per unit cost," *Information Theory, IEEE Transactions on*, vol. 36, no. 5, pp. 1019–1030, 1990.
- [22] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. Wiley-Interscience, 2001.

Zuxing Li (S'14) received his B.Eng. degree in Information Security in 2009 from Shanghai Jiao Tong University, P. R. China, his M.Sc. degrees in Electrical Engineering in 2013 from Technical University of Catalonia, Spain and KTH Royal Institute of Technology, Sweden. He is currently working toward the Ph.D. degree with the Communication Theory Lab. at KTH Royal Institute of Technology. His research interests include information security, distributed detection, information theory, and networked control.

Tobias J. Oechtering (S'01-M'08-SM'12) received his Dipl.-Ing degree in Electrical Engineering and Information Technology in 2002 from RWTH Aachen University, Germany, his Dr.-Ing degree in Electrical Engineering in 2007 from the Technische Universität Berlin, Germany, and his Docent degree in Communication Theory in 2012 from KTH Royal Institute of Technology, Sweden. In 2008, he joined the Communication Theory Lab. at KTH Royal Institute of Technology and has been an Associate Professor since May 2013. He is currently Associate Editor of IEEE Transactions on Information Forensic and Security since June 2016 and was editor for IEEE Communications Letters during 2012–2015. Dr. Oechtering received the "Förderpreis 2009" from the Vodafone Foundation. His research interests include network information theory, statistical signal processing, physical-layer privacy and security, as well as communication for networked control.