Privileged Estimate Fusion With Correlated Gaussian Keystreams

Marko Ristic
Autonomous Multisensor Systems Group (AMS),
Institute for Intelligent Cooperating Systems (ICS),
Otto von Guericke University (OVGU),
Magdeburg, Germany
Email: marko.ristic@ovgu.de

Benjamin Noack
Autonomous Multisensor Systems Group (AMS),
Institute for Intelligent Cooperating Systems (ICS),
Otto von Guericke University (OVGU),
Magdeburg, Germany
Email: benjamin.noack@ovgu.de

***Abstract*—Providing cryptographic privacy guarantees in a distributed state estimation problem has been a growing topic of research since the ubiquity of modern public networks. One such guarantee is having different levels of estimation performance achievable by trusted and untrusted users within a sensor network. In the presence of multiple sensor measurements, guaranteeing better estimation performance by the usual means of adding removable noise to measurements is complicated by an alternative for untrusted users to improve their performance: fusing more measurements. Our novel method adds correlated noise at different sensors, restricting the performance gained from fusing additional measurements while guaranteeing better performance to those that can remove it. We extend a cryptographic framework for defining estimation privilege and use this to prove the scheme's security goals, while simulations demonstrate the effects of parameters in a concrete estimation problem. A scheme that can ensure such differences in estimation performance between types of estimators can find applications in priority-based or subscription-based performances in environments where more than one sensor is present.**

I. Introduction

Sensor data processing and state estimation have long been active areas of research and continue to find applications in modern systems [1], [2]. In the context of distributed sensing environments such as decentralised autonomous vehicles or distributed weather stations, estimation methods relying on Kalman filters and derivatives [3] are particular prevalent due to their recursive, often optimal, estimation properties and their suitability to modelling measurement cross-correlations typically required for data fusion [4], [5]. In recent years, the ubiquity of distributed public networks has seen the additional requirements of preserving algorithm participants' privacies, such as individual contributions or identifying information, become increasingly relevant and has led to an active field of research [6], [7].

While hiding transmitted information to eavesdroppers can be achieved using common encryption schemes that make ciphertexts indistinguishable [8], distributed estimation tasks that preserve participants' privacies require information to remain hidden during partial or complete processing of the task and often justify some leakage [9], [10]. The security goals of these problems are context-specific and have produced a variety of solutions. In [11], non-Bayesian localisation is performed using homomorphic encryption such that individual sensor information and measurements remain private, while in [12], similar goals are achieved in a Bayesian setting by fusing linear measurements when sensors form a hierarchical network. A combination of homomorphic and order-revealing encryption schemes are used in [9] to solve conservative Gaussian estimate fusion while leaking only the ratios of estimate covariances, and in [10], [13], cryptographic aggregation schemes are introduced and used to leak only total consumptions in an energy grid while hiding individual participant power-usages. In addition to these examples of privacy, achieved through hiding and leaking quantifiable information, optimal estimation performance itself can be considered leakage, which introduces a privilege in estimation where leakage defines the difference in performance between trusted and untrusted estimators. In the original Global Positioning System (GPS) [14], this was achieved with a secondary encrypted channel that allowed better estimation performance to parties that held an encryption key. Similarly, in [15], a synchronised chaotic system is used to add noise to measurements which can only be removed by estimators knowing its properties. This idea of estimation privilege is further explored in [16], where a formal cryptographic definition is given and a scheme for a single sensor presented. Here, a synchronised pseudorandom Gaussian keystream adds measurement noise only removable by estimators holding the stream key.

In this work, we consider this definition of estimation privilege, presented in [16], and introduce a modified scheme suitable for an environment of multiple sensors, where both holding the secret key and fusing additional measurements can lead to better estimation performance. Our contribution consists of a generalised notion of estimation privilege, the introduction of security requirements for privileged estimation in a well defined multisensor environment and a scheme that satisfies them. Along with a cryptographic proof sketch, simulation results are provided to demonstrate the effects of scheme parameters and how they can be chosen to provide varying amounts of privilege. Use-cases for varying performance in this way include subscription or priority models where some users are provided better results than others. For example, subscription-based weather forecasts using measurements from

spatially distributed stations or modular mass-produced sensors that differ in accuracy dependent on cost.

In section II, the multisensor estimation privilege problem is presented. Relevant preliminaries are introduced in section III and the estimation privilege fusion scheme itself in section IV. A cryptographic analysis and the simulation results are then given in sections V and VI, respectively, before the concluding remarks in section VII.

A. Notation

Lowercase underlined characters $\underline{v}$ are vectors and uppercase bold characters $\mathbf{M}$ are matrices, while $\underline{0}$, $\mathbf{0}$ and $\mathbf{I}$ are the zero vector, zero matrix and identity matrix, respectively, with sizes inferrable from context. $\mathbf{M} \succ 0$ and $\mathbf{M} \succeq 0$ denote positive definitness and semi-definitness, respectively, and $\mathbf{M} \succ \mathbf{N}$ is short for $\mathbf{M} - \mathbf{N} \succ 0$. Given an indexed vector $\underline{v}_i$, notation $\underline{v}^{(a:b)}$, $a < b$, will denote the stacked vector $[\underline{v}_a \ \underline{v}_{a+1} \ \cdots \ \underline{v}_b]^\top$, while for indexed matrices $\mathbf{M}_i$, $\mathbf{M}^{(a:b)}$ will denote the block-diagonal matrix with the indexed matrices along the diagonal. Function $\mathsf{Cov}[\cdot]$ computes the covariance of a random vector, $\sim$ denotes distribution, $\dot{\sim}$ denotes pseudorandom distribution and, in a cryptographic context, $\mathcal{A}(\underline{v})$ denotes the output of an arbitrary algorithm $\mathcal{A}$ given inputs $\underline{v}$.

II. Problem Formulation

We are interested in environments where multiple sensors are present and the fusion of their measurements can lead to better state estimation accuracy of the system they are measuring. In these environments, we want to provide levels of privilege to state estimators such those with a higher privilege can perform better than those with a lower one, while taking into consideration the estimation benefits from fusing additional measurements. We will consider linear and Gaussian models, where a state $\underline{x}_k \in \mathbb{R}^n$, at an integer timestep k, follows a system model given by

$$\underline{x}_k = \mathbf{F}_k \underline{x}_{k-1} + \underline{w}_k \,, \tag{1}$$

with white noise term $\underline{w}_k \sim \mathcal{N}(\underline{0}, \mathbf{Q}_k)$ and a known covariance $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$. Similarly, measurements $\underline{y}_{k,i} \in \mathbb{R}^m$ from each sensor i, $1 \leq i \leq N$, follow the measurement models

$$\underline{y}_{k,i} = \mathbf{H}_{k,i} \underline{x}_k + \underline{v}_{k,i} \,, \tag{2}$$

with white noise term $\underline{v}_{k,i} \sim \mathcal{N}(\underline{0}, \mathbf{R}_{k,i})$ and a known covariance $\mathbf{R}_{k,i} \in \mathbb{R}^{m \times m}$. In addition to these models, we assume that all sensors i are synchronised in timesteps k to simplify later cryptographic evaluation. In practice, this would restrict the presented schemes to scenarios where synchronisation is easier to guarantee, such as ones where measurements are taken infrequently.

Each of the sensors also holds a secret key sk_i, which can be made available to estimators of appropriate privilege. The privileges that we consider, in terms of access to keys and measurements, will be defined by sequential sensors. That is, in the presence of N sensors, we consider exactly N possible privilege levels, where each privilege $p > 0$ corresponds to holding the sequential secret keys sk_j, $1 \leq j \leq p$, while being unprivileged, $p = 0$, corresponds to holding none. We assume that estimators have access to all "privileged measurements", those from sensors whose keys they hold, and can additionally fuse "unprivileged measurements" from those whose keys they do not hold. To simplify notation, we will consider access to unprivileged measurements to be sequential as well, and can therefore capture their capabilities by letting $\mathsf{e}^{[p,q]}$ denote an estimator with privilege p and access to measurements from $q \geq p$ sensors, $1 \leq i \leq q$.

To cryptographically guarantee the difference between estimation performances, we will use the notion of covariance privilege [16]. As we want better performance for higher privilege estimators as well as to limit the gained performance from fusing unprivileged measurements, two differences will be considered.

Different Keys Lower Bound We want to guarantee a lower bound on the difference between the best unprivileged estimator $\mathsf{e}^{[0,N]}$ and the best privilege-p estimator $\mathsf{e}^{[p,p]}$. This will, therefore, remain a lower-bound when unprivileged estimators have access to fewer unprivileged measurements or privileged estimators have access to more.

Same Keys Upper Bound We want to guarantee an upper bound on the difference between the best privilege-p estimators $\mathsf{e}^{[p,p]}$ and $\mathsf{e}^{[p,N]}$. Here, the bound remains an upper bound on the gained estimation performance of any privilege-p estimator fusing additional unprivileged measurements.

The resulting scheme should be such that two parameters are responsible for choosing the values of these two bounds.

Remark. We stress that the two bounds that will be guaranteed only bound the performances of honest-but-curious estimators. That is, nothing is said about estimators which may corrupt sensors to obtain keys beyond their privilege. Bounds on leakage caused by learning additional keys are beyond the scope of this work.

III. Preliminaries

When defining our scheme and discussing its cryptographic guarantees, we will make use of Gaussian keystreams and the notion of cryptographic estimation covariance privilege. These have been summarised below.

A. Gaussian keystreams

A Gaussian keystream is a sequence of pseudorandom multivariate Gaussian samples $\underline{g}_k \in \mathbb{R}^m \dot{\sim} \mathcal{N}(\underline{0}, \mathbf{S})$, $k > 0$, for some covariance matrix $\mathbf{S}$, generated using a secret key. The sequence is indistinguishable from a truly random multivariate Gaussian sequence to any observer who does not hold the key while being reproducible exactly by those that do. The keystream can be constructed from a typical cryptographic stream cipher [8, Ch. 3.6], by using any common method for random floating-point generation [17] to create a stream of pseudorandom uniform samples $u_x \dot{\sim} \mathcal{U}(0,1)$, $x > 0$. Here, we make the assumption that floating-point representations of

real numbers are sufficiently similar to actual real numbers, made reasonable in practice due to the insignificance of their difference in many applications and their prevalence in estimation theory.

To construct the Gaussian keystream $\underline{g}_k$, $k > 0$, from the uniform one u_x, $x > 0$, a vector of uncorrelated standard Gaussian samples $\underline{z}_k$ is first generated using the Box-Muller transform,

$$\underline{z}_k = \begin{bmatrix} z_{(k-1)m+1} & \cdots & z_{km} \end{bmatrix}^\top \tag{3}$$

where

$$z_x = \begin{cases} \sqrt{-2\ln(u_{2x})}\cos(2\pi u_{2x+1}), & x \text{ is odd} \\ \sqrt{-2\ln(u_{2x})}\sin(2\pi u_{2x+1}), & x \text{ is even} \end{cases},$$

and then correlated by

$$\underline{g}_k = \mathbf{S}^{\frac{1}{2}}\underline{z}_k\,, \tag{4}$$

where $\mathbf{S}^{\frac{1}{2}}$ is any matrix such that $\mathbf{S}^{\frac{1}{2}}\mathbf{S}^{\frac{1}{2}\top} = \mathbf{S}$.

B. Cryptographic Estimation Privilege

The formal definition of a privileged estimation scheme and accompanying notion of covariance privilege are introduced in [16] and capture a reduction in estimation performance that can always be achieved from a probabilistic polynomial-time (PPT) estimator knowing only unprivileged measurements, and holding no scheme key, to one knowing both the key and privileged measurements. We will use a slight generalisation of this definition to capture an arbitrary difference that can always be achieved between the estimators (rather than only a reduction). A privileged estimation scheme is defined by a pair of probabilistic algorithms $(\mathsf{Setup}, \mathsf{Noise})$:

$\mathsf{Setup}(\mathcal{M}_S, \mathcal{M}_M, \kappa)$ Given system and measurement models $\mathcal{M}_S$ and $\mathcal{M}_M$, and the security parameter κ, public parameters pub and a secrect key sk are returned.

$\mathsf{Noise}(\mathsf{pub}, \mathsf{sk}, k, \mathcal{M}_S, \mathcal{M}_M, \underline{y}_1, \ldots, \underline{y}_k)$ Given public parameters pub, secret key sk, timestep k, models $\mathcal{M}_S$ and $\mathcal{M}_M$, and measurements $\underline{y}_1, \ldots, \underline{y}_k$, modified privileged and unprivileged measurements, $\underline{y}_k^{[p]}$ and $\underline{y}_k^{[up]}$, respectively, are returned.

Using the defintions from [16] for an *estimator* and *negligible covariance* $\mathsf{negl}\mathsf{Cov}_m(\kappa)$, the notion of covariance privilege is defined as follows.

Definition III.1. A privileged estimation scheme meets notion $\{\mathbf{D}_1, \mathbf{D}_2, \ldots\}$-*Covariance Privilege for Models* $\mathcal{M}_S$ *and* $\mathcal{M}_M$ if for any PPT estimator $\mathcal{A}$, there exists a PPT estimator $\mathcal{A}'$, such that

$$\begin{aligned} &\mathsf{Cov}\left[\mathcal{A}\left(k, \kappa, \mathsf{pub}, \mathcal{M}_S, \mathcal{M}_M, \underline{y}_1^{[up]}, \ldots, \underline{y}_k^{[up]}\right) - \underline{x}_k\right] \\ &- \mathsf{Cov}\left[\mathcal{A}'\left(k, \kappa, \mathsf{pub}, \mathcal{M}_S, \mathcal{M}_M, \underline{y}_1^{[p]}, \ldots, \underline{y}_k^{[p]}\right) - \underline{x}_k\right] \\ &\succeq \mathbf{D}_k - \mathsf{negl}\mathsf{Cov}_m(\kappa) \end{aligned} \tag{5}$$

for all $k > 0$, some negligible covariance and where matrices $\mathbf{D}_k$ are semi-definite, *i.e.*, $\mathbf{D}_k \preceq \mathbf{0}$ or $\mathbf{D}_k \succeq \mathbf{0}$. Here, $\mathcal{A}$ and $\mathcal{A}'$ are running in polynomial-time with respect to the parameter κ and all probabilities are taken over randomness introduced in models $\mathcal{M}_S$ and $\mathcal{M}_M$, estimators $\mathcal{A}$ and $\mathcal{A}'$, and algorithms Setup and Noise.

We note that in the generalised notion, definition III.1, $\mathbf{D}_k \preceq \mathbf{0}$ is allowed, and unprivileged estimators may now perform better than privileged ones, albeit by a bounded amount. This feature will be useful when discussing additional estimation performance gainable from fusing unprivileged measurements. Lastly, a sign is also corrected, such that negligible covariances are subtracted (rather than added), correctly capturing the additional negligible performance gainable by an unprivileged estimator.

IV. Privileged Fusion

The idea behind our privileged estimation fusion scheme is to add *correlated* Gaussian keystreams to the measurements from each sensor. These noises can be computed and subtracted by estimators holding respective sensor keys, while their correlation limits the additional information gained from fusing unprivileged measurements.

A. Noise Generation

Similarly to the Gaussian keystream generation in (4), pseudorandom samples can be correlated in this way even when generated using different stream cipher keys. To parameterise the correlation between noises at each sensor, we introduce a fully correlated component $\mathbf{Z} \in \mathbb{R}^{m\times m}$ and an uncorrelated component $\mathbf{Y} \in \mathbb{R}^{m\times m}$ and define a noise cross-correlation matrix for x noises as $\mathbf{S}^{(x)} \in \mathbb{R}^{xm\times xm}$,

$$\mathbf{S}^{(x)} = \begin{bmatrix} \mathbf{Z} & \cdots & \mathbf{Z} \\ \vdots & \ddots & \vdots \\ \mathbf{Z} & \cdots & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{Y} & \mathbf{0} & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{Y} \end{bmatrix}, \tag{6}$$

and $\mathbf{S}^{(1)} = \mathbf{Z} + \mathbf{Y}$. The generation of all N multivariate Gaussian noises at timestep k, $\underline{g}_k^{(1:N)}$, can now be written as

$$\underline{g}_k^{(1:N)} = \begin{bmatrix} \underline{g}_{k,1} \\ \vdots \\ \underline{g}_{k,N} \end{bmatrix} = \mathbf{S}^{(N)\frac{1}{2}} \cdot \begin{bmatrix} \underline{z}_{k,1} \\ \vdots \\ \underline{z}_{k,N} \end{bmatrix}, \tag{7}$$

where each $\underline{z}_{k,i}$ is computed as $\underline{z}_k$ in (3) using uniform samples generated with key sk_i, and $\mathbf{S}^{(N)\frac{1}{2}}$ is a matrix such that $\mathbf{S}^{(N)\frac{1}{2}}\mathbf{S}^{(N)\frac{1}{2}\top} = \mathbf{S}^{(N)}$. Notably, it is important that the first p noises $\underline{g}_{k,i}$, $1 \le i \le p$, in (7), denoted $\underline{g}_k^{(1:p)}$, can be reproduced by an estimator of privilege p, holding only the keys sk_i, $1 \le i \le p$. One case where this is possible is when a lower-triangular decomposition, such as the the Cholesky decomposition, is used to compute $\mathbf{S}^{(x)\frac{1}{2}}$ from $\mathbf{S}^{(x)}$. Here, each correlated Gaussian sample $\underline{g}_{k,i}$ is computable from preceeding uniform samples $\underline{z}_{k,j}$, $j \le i$ only and the generalised noise generation equation

$$\underline{g}_k^{(1:p)} = \mathbf{S}^{(p)\frac{1}{2}} \cdot \begin{bmatrix} \underline{z}_{k,1} \\ \vdots \\ \underline{z}_{k,p} \end{bmatrix}, \tag{8}$$

generates the same first p noises $\underline{g}_k^{(1:p)}$ as would be obtained from (7) since $\mathbf{S}^{(p)\frac{1}{2}} \in \mathbb{R}^{pm\times pm}$ is equal to the top left block of matrix $\mathbf{S}^{(N)\frac{1}{2}}$ when using the Cholesky decomposition.

With (8), at every timestep k, $\underline{g}_k^{(1:N)}$ can be generated using all N keys and used to modify sensor measurements, while the subset $\underline{g}_k^{(1:p)}$ can be generated by estimators of privilege p using the only keys they hold.

B. Measurement Modification

With a way to generate noises for sensors and estimators, we can introduce the means of measurement modification and the observable measurement models for different estimators. Measurement modification is peformed by adding noises $\underline{g}_k^{(1:N)}$ to measurements from each sensor i before making them public, resulting in modified measurement equations for each sensor,

$$\begin{aligned}\underline{y}'_{k,i} &= \underline{y}_{k,i} + \underline{g}_{k,i} \\ &= \mathbf{H}_{k,i}\underline{x}_k + \underline{v}_{k,i} + \underline{g}_{k,i}\,,\end{aligned} \tag{9}$$

with real measurement noise $\underline{v}_{k,i} \sim \mathcal{N}(\underline{0}, \mathbf{R}_{k,i})$ and the vector of all added noises $\underline{g}_k^{(1:N)} \dot{\sim} \mathcal{N}(\underline{0}, \mathbf{S}^{(N)})$. As we assume that sensors are synchronised, we can capture the correlation between these modified measurements exactly by considering the stacked measurement model for any estimator with access to q measurements, $\mathsf{e}^{[p,q]}$, at time k, given by

$$\underline{y}_k^{(1:q)} = \mathbf{H}_k^{(1:q)}\underline{x}_k + \underline{v}_k^{(1:q)} + \underline{g}_k^{(1:q)} \tag{10}$$

where $\underline{v}_k^{(1:q)} \sim \mathcal{N}(\underline{0}, \mathbf{R}_k^{(1:q)})$ and $\underline{g}_k^{(1:q)} \dot{\sim} \mathcal{N}(\underline{0}, \mathbf{S}^{(q)})$, with

$$\underline{y}_k^{(1:q)} = \begin{bmatrix}\underline{y}'_{k,1}\\ \vdots \\ \underline{y}'_{k,q}\end{bmatrix},\ \mathbf{H}_k^{(1:q)} = \begin{bmatrix}\mathbf{H}_{k,1}\\ \vdots \\ \mathbf{H}_{k,q}\end{bmatrix},\ \underline{v}_k^{(1:q)} = \begin{bmatrix}\underline{v}_{k,1}\\ \vdots \\ \underline{v}_{k,q}\end{bmatrix},$$

$$\mathbf{R}_k^{(1:q)} = \begin{bmatrix}\mathbf{R}_{k,1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{k,q}\end{bmatrix}$$

and $\mathbf{S}^{(q)} \in \mathbb{R}^{qm\times qm}$ defined by (6).

Since we are using a cryptographically sound stream cipher to generate the added Gaussian keystream, the pseudorandom samples are indistinguishable from truly random ones to estimators without appropriate keys, which leads us to three observable measurement models, *i.e.*, the models that capture all the information available to an estimator exactly, for three types of mutually exclusive estimators.

Estimators of the form $\mathsf{e}^{[0,q]}$ Here, no keys are held by the unprivileged estimators and the generated noise $\underline{g}_k^{(1:q)}$ is indistinguishable from noise from the truly random distribution $\mathcal{N}(\underline{0}, \mathbf{S}^{(q)})$. For these estimators, we can rewrite the measurement equation (10) as the observed measurement model

$$\underline{y}_k^{[0,q]} = \mathbf{H}_k^{(1:q)}\underline{x}_k + \underline{v}'_k\,, \tag{11}$$

with truly Gaussian term $\underline{v}'_k \sim \mathcal{N}(\underline{0}, \mathbf{R}_k^{(1:q)} + \mathbf{S}^{(q)})$.

Estimators of the form $\mathsf{e}^{[p,p]}$ Estimators with keys for all the sensors to which they have access can generate all added noises and subtract them from the recieved measurements. That is, $\underline{g}_k^{(1:p)}$ can be generated and $\underline{y}_k^{[p,p]} = \underline{y}_k^{(1:p)} - \underline{g}_k^{(1:p)}$ computed to give the observed measurement model equal to recieving unmodified measurements only,

$$\underline{y}_k^{[p,p]} = \mathbf{H}_k^{(1:p)}\underline{x}_k + \underline{v}_k^{(1:p)}\,, \tag{12}$$

where $\underline{v}_k^{(1:p)} \sim \mathcal{N}(\underline{0}, \mathbf{R}_k^{(1:p)})$.

Estimators of the form $\mathsf{e}^{[p,q]}$, $p < q$ Lastly, we want the observed measurement model when only some accessible measurements can have have their noises removed. Here, care must be taken when giving the observed model, as the noises from sensors $i > p$, which cannot be removed, are conditionally dependant on the known noises $\underline{g}_k^{(1:p)}$. Since we can generate the noises $\underline{g}_k^{(1:p)}$ and know that $\underline{g}_k^{(1:q)} \dot{\sim} \mathcal{N}(\underline{0}, \mathbf{S}^{(q)})$, we can write

$$\begin{aligned}\underline{g}_k^{(1:q)} &= \begin{bmatrix}\underline{g}_k^{(1:p)}\\ \underline{g}_k^{(p+1:q)}\end{bmatrix} \\ &\dot{\sim} \mathcal{N}\left(\begin{bmatrix}\underline{0}\\ \underline{0}\end{bmatrix}, \mathbf{S}^{(q)} = \begin{bmatrix}\mathbf{S}^{(p)} & \mathbf{Z}'\\ \mathbf{Z}'^{\top} & \mathbf{S}^{(q-p)}\end{bmatrix}\right),\end{aligned} \tag{13}$$

where $\mathbf{Z}' \in \mathbb{R}^{pm\times(q-p)m}$ is a block matrix with every element equal to $\mathbf{Z}$, and compute the conditional pseudorandom Gaussian distribution

$$\begin{aligned}&\underline{g}_k^{(p+1:q)} \mid \underline{g}_k^{(1:p)} \\ &\quad\dot{\sim} \mathcal{N}\Big(\mathbf{Z}'^{\top}\mathbf{S}^{(p)-1}\underline{g}_k^{(1:p)}, \\ &\qquad \mathbf{S}^{(q-p)} - \mathbf{Z}'^{\top}\mathbf{S}^{(p)-1}\mathbf{Z}'\Big)\,.\end{aligned} \tag{14}$$

Now, subtracting the known noises $\underline{g}_k^{(1:p)}$ and the means of unknown noises (14) from recieved measurements,

$$\underline{y}_k^{[p,q]} = \underline{y}_k^{(1:q)} - \begin{bmatrix}\underline{g}_k^{(1:p)}\\ \mathbf{Z}'^{\top}\mathbf{S}^{(p)-1}\underline{g}_k^{(1:p)}\end{bmatrix}, \tag{15}$$

and accounting for unknown pseudorandom noises being indistinguishable from random, a zero-mean observed measurement model can be written as

$$\underline{y}_k^{[p,q]} = \mathbf{H}_k^{(1:q)}\underline{x}_k + \underline{v}'_k \tag{16}$$

where

$$\underline{v}'_k \sim \mathcal{N}\left(\underline{0}, \begin{bmatrix}\mathbf{R}_k^{(1:p)} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{(q-p)} - \mathbf{Z}'^{\top}\mathbf{S}^{(p)-1}\mathbf{Z}' + \mathbf{R}_k^{(p+1:q)}\end{bmatrix}\right).$$

Remark. Recalling that access to unprivileged measurements are sequential to simplify notation, (13), (15) and (16) can be generalised when having access to arbitrart $q-p$ non-sequential unprivileged measurements $\underline{y}_{k,i}$, $p < i \leq q$, by appropriately replacing elements $\mathbf{Z}'^{\top}$ and $\mathbf{S}^{(q-p)}$ in (13).

From the observed measurement models (11), (12) and (16) we can tell that the parameters $\mathbf{Z}$ and $\mathbf{Y}$ control the difference in estimation performance between the three types of estimators. More specifically, while both components affect estimation, the fully correlated component $\mathbf{Z}$ predominantly affects the difference in estimation between estimators of the form $\mathsf{e}^{[0,q]}$ and $\mathsf{e}^{[p,p]}$, while the uncorrelated component $\mathbf{Y}$ affects the difference between estimators $\mathsf{e}^{[p,p]}$ and $\mathsf{e}^{[p,q]}$. These two differences exactly correspond to the difference that we want to guarantee cryptographically from section II. The effects of choosing $\mathbf{Z}$ and $\mathbf{Y}$ will be more formally explored in the cryptography and simulation sections, V and VI, respectively.

C. Noise Distribution

While we have described a method for generating noises that modify N measurements resulting in different observed measurement models depending on estimator privilege, we have not discussed where the noise is generated and how it is distributed to the sensors. To handle the inherent correlation of the noises $\underline{g}_{k,i}$, $1 \leq i \leq N$, used to modify measurements, they can be either generated centrally before distribution to sensors or sequentially at the sensors themselves, given previously generated values.

Central noise generation To compute noises centrally, (8) can be computed at a single central processor and each noise $\underline{g}_{k,i}$, $1 \leq i \leq N$ sent to respective sensors i before modifying their local measurement by (9).

Sequential noise generation To compute the same noises sequentially, at each timestep k, sensor 1 can generate its noise independently using its current standard Gaussian sample $\underline{z}_{k,1}$, by $\underline{g}_{k,1} = \mathbf{S}^{(1)\frac{1}{2}}\underline{z}_{k,1}$. Each following sensor $i > 1$ can generate its noise $\underline{g}_{k,i}$ given the preceeding noises $\underline{g}_k^{(1:i-1)}$ using the same conditional pseudorandom Gaussian breakdown as in (14) allowing for the generation of $\underline{g}_{k,i}$ with the transformation

$$\begin{aligned}\underline{g}_{k,i} =& \mathbf{Z}'^{\top}\mathbf{S}^{(i-1)-1}\underline{g}_k^{(1:i-1)} + \\ & (\mathbf{S}^{(1)} - \mathbf{Z}'^{\top}\mathbf{S}^{(i-1)-1}\mathbf{Z}')^{\frac{1}{2}}\underline{z}_{k,i}\,.\end{aligned} \tag{17}$$

After local noise generation, each sensor i sends its result $\underline{g}_{k,i}$ as well as preceding results $\underline{g}_k^{(1:i-1)}$ to the next sensor. This method has the clear downside of increasing communication costs with each successive generation but does not require a single central communicator.

In both cases above, the computation of all noises $\underline{g}_k^{(1:N)}$ can be performed offline such that sensors have sufficient additive noises already computed prior to commencement of measurement.

V. Cryptographic Privilege

To give sketch proofs of the cryptographic guarantees provided by the presented scheme, we first recall some made assumptions. We consider floating-point numbers to be sufficiently close to real random numbers that real-number proofs still hold, and that all sensors are synchronised in k such that observed measurement models (11), (12) and (16) are exactly correct. Using these assumptions, the proofs rely on the optimality of the linear Kalman filter (KF) [3] to produce a series of covariances for optimal estimators and take their difference to obtain the *Different Keys Lower Bound* and *Same Keys Upper Bound* from section II. Similarly to [16], these series give $\mathbf{D}_1, \mathbf{D}_2, \ldots$ in definition III.1 for the two bounds, and demonstrate that the existence of an estimator violating the notion implies the existence of a linear estimator with lower mean squared error (MSE) than the KF. This guarantees the bound by contrapositive.

A. Different Keys Lower Bound

First we consider the lower bound to the difference in estimation between estimators of the form $\mathsf{e}^{[0,N]}$ and $\mathsf{e}^{[p,p]}$. The observed measurement models for the two estimators are given by (11) and (12), respectively, when $q = N$.

With these measurement models and the system model (1), the KF can be used to recursively compute the optimal estimate covariances $\mathbf{P}_k$ at each timestep k for an estimator given an initial estimate covairance. Since the KF also preserves initial covariance order,

$$\mathbf{P}_k \succeq \mathbf{P}'_k \implies \mathbf{P}_{k+1} \succeq \mathbf{P}'_{k+1}\,, \tag{18}$$

we can compute a lower-bound for all possible covariances when letting $\mathbf{P}_0 = \mathbf{0}$.

- We can then write the combined Kalman predict and update equation as the equation with params from the previous one.
- Due to the KF preserving error covariance order, by setting $P_0 = 0$ we get a series of covariance such that no unprivileged estimator can estimate with error covariances less than or equal to the series (lower-bound).
- Similarly we can do the same for an estimator to which we want to lower-bound the estimation difference with the unprivileged estimator.
- A privileged estimator of privilege j (access to the first j keys) and that can only access the first j sensors, has the appropriate measurement equation. Similarly, setting the initial covariance to zero gives the lower bound series on the best possible estimation error achievable by the privileged estimator using only the measurements from the sensors to which it holds keys.
- Taking the difference of the two estimator bounding series' produces the difference series.
- In the context of cryptographic privileged estimation scheme, the Setup and Noise algorithms are given accordingly. The optimality of the KF can then be used to achieve the security notion.
- In the above, we assume the unprivileged estimator has access to all n sensors, while the j privilege estimator has access to only the first j sensors. In the case when the unprivileged estimator has access to fewer sensors or the privileged one to more, their difference in estimation can only increase, thus keeping the computed lower-bound

a lower-bound and the cryptographic guarantee does not change (albeit the definitions of Setup and Noise will, to capture the now available sets of measurements).

B. Same Keys Upper Bound

- We can use a similar approach to separately guarantee the largest possible benefit in estimation available to an estimator of privilege j (access to first j keys) when fusing measurements from sensors to which they do not hold keys to get a better state estimate.
- We can again write stacked estimation models for the two estimators. The estimator of privilege j with access to the first j measurements still follows the model (given in the previous subsection).
- The estimator with access to the additional l measurements, however, does not have zero-mean measurement noise as seen in (equation for the additional measurements estimator given before) required for the Kalman filter. As the mean is known at each k, from the known noises $p_{i,k}$, $1 \leq i \leq j$, remaining measurements can be offset to produce the equivalent measurement model with zero-mean noise (given here).
- Again, using the combined predict and update equations of the KF and setting $P_0 = 0$, we can use the optimality of the KF to give the best possible performances of the estimators as a series of covariances.
- Taking the difference of the bounds now gives a bound on how much better an estimate can become when unprivileged measurements are fused with privileged ones.
- In the context of a cryptographic privileged estimation scheme, the Setup and Noise algorithms can be given as follows. KF optimality can then be used to achieve the security notion.
- We note that unlike in the unprivileged adversary case, or the previous paper, here we do not use unmodified measurements for the privileged estimates and the resulting algorithms provide better estimation for the adversary than the privileged estimator. In this form, we bound how much better an adversary can estimate when using unprivileged estimates resulting in series D_k consisting of negative-definite rather than positive definite matrices. This makes use of the generalised cryptographic definitions earlier.

VI. Simulation

- In addition to showing how to derive the bounds to the benefits of using unprivileged measurements, we have simulated concrete scenarios to demonstrate the methods.
- We consider the following linear system, and linear measurement models for 4 sensors. Correlated and uncorrelated noise covariances are of the form $Y = \sigma_y I$ and $Z = \sigma_z I$.
- Implementation details.
- First figure shows 4 plots. Each plot shows three traces. The traces are the trace of the unprivileged estimator (no keys, all measurements), the trace of a privilege j estimator (j measurements) and the trace of a fusing privilege j estimators (all measurements). j is equal to $1, 2, 3$ and 4 for each plot, respectively. Here, the difference between the first two traces in figure 4 and the second two traces in figure 1 is equal to the traces of the two D_k series from the cryptography section, respectively. Values for σ_y, σ_z are fixed.
- The second figure focuses on a single plot from the first figure but varies σ_y and σ_z. Again, 4 plots are shown, with σ_y increased in two and σ_z increased in two, such that they are both increased in only one plot. This will show that while both parameters affect the estimation of unprivileged and additional fusion estimators, σ_y and therefore Y predominantly affects the unprivileged estimator and σ_z, Z, affects the additional fusion estimator.

VII. Conclusion

- Concluding remarks.
- Future work includes exploring key subsets that do not need to be sequential, decentralised methods for multi-key correlated noise generation, and the effects and cryptographic guarantees of sensors falling out of synchronisation.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Dover Publications.

[2] D. Simon, *Optimal State Estimation: Kalman, H Infinity and Nonlinear Approaches*. Wiley-Interscience.

[3] A. J. Haug, *Bayesian Estimation and Tracking: A Practical Guide*. John Wiley & Sons.

[4] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. CRC press.

[5] M. Liggins, C. Y. Chong, D. Hall, and J. Llinas, *Distributed Data Fusion for Network-Centric Operations*. CRC Press.

[6] K. Ren, C. Wang, and Q. Wang, "Security Challenges for the Public Cloud," vol. 16, no. 1, pp. 69–73.

[7] M. Brenner, J. Wiebelitz, G. von Voigt, and M. Smith, "Secret Program Execution in the Cloud Applying Homomorphic Encryption," in *5th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, pp. 114–119.

[8] J. Katz and Y. Lindell, *Introduction to Modern Cryptography: Principles and Protocols*. Chapman & Hall.

[9] M. Ristic, B. Noack, and U. D. Hanebeck, "Secure Fast Covariance Intersection Using Partially Homomorphic and Order Revealing Encryption Schemes," vol. 5, no. 1, pp. 217–222.

[10] E. Shi, T.-H. H. Chan, and E. Rieffel, "Privacy-Preserving Aggregation of Time-Series Data," p. 17.

[11] A. Alanwar, Y. Shoukry, S. Chakraborty, P. Martin, P. Tabuada, and M. Srivastava, "PrOLoc: Resilient Localization with Private Observers Using Partial Homomorphic Encryption," in *16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 41–52.

[12] M. Aristov, B. Noack, U. D. Hanebeck, and J. Müller-Quade, "Encrypted Multisensor Information Filtering," in *21st International Conference on Information Fusion (Fusion 2018)*, pp. 1631–1637.

[13] M. Joye and B. Libert, "A Scalable Scheme for Privacy-Preserving Aggregation of Time-Series Data," in *International Conference on Financial Cryptography and Data Security*, ser. Lecture Notes in Computer Science. Springer, pp. 111–125.

[14] P. D. Groves, "Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems," vol. 30, no. 2, pp. 26–27.

[15] C. Murguia, I. Shames, F. Farokhi, and D. Nešić, "Information-Theoretic Privacy Through Chaos Synchronization and Optimal Additive Noise," in *Privacy in Dynamical Systems*. Springer, pp. 103–129.

[16] M. Ristic, B. Noack, and U. D. Hanebeck, “Cryptographically Privileged State Estimation With Gaussian Keystreams,” vol. 6, pp. 602–607.
[17] F. Goualard, “Generating Random Floating-Point Numbers by Dividing Integers: A Case Study,” vol. 12138, pp. 15–28.