Dissertation for the Faculty of Computer Science (FIN),
Otto von Guericke University (OVGU), Magdeburg

Data Confidentiality for Distributed Sensor Fusion

Marko Ristic

August 18, 2022

Reviewers:

Prof. Dr.-Ing. Benjamin Noack (OVGU, Magdeburg, Germany)

. . .

Contents

Acknowledgements

Acks go here.

Abstract

Distributed sensing and fusion algorithms are increasingly present in public computing networks and have led to a natural concern for data security in these environments. This thesis aims to present generalisable data fusion algorithms that simultaneously provide strict cryptographic guarantees on user data confidentiality. While fusion algorithms providing some degrees of security guarantees exist, these are typically either provided at the cost of solution generality or lack formal security proofs. Here, novel cryptographic constructs and state-of-the-art encryption schemes are used to develop formal security guarantees for new and generalised data fusion algorithms. Industry-standard Kalman filter derivates are modified and existing schemes abstracted such that novel cryptographic notions capturing the required communications can be formalised, while simulations provide an analysis of practicality. Due to the generality of the presented solutions, broad applications are supported, including autonomous vehicle communications, smart sensor networks and distributed localisation.

Kurzfassung

German abs go here.

Notation

Complete the notation here.

1. Introduction

Sensor data processing, state estimation and data fusion have long been active areas of research and continue to find applications in modern systems [1, 2]. As distributed networks have become more prevalent over the years, greater stress has been put on the need for broadly applicable algorithms that support varying types of measurements, estimate accuracies and communication availabilities [3, 4], supporting use cases including localisation, weather forecasting, mapping, cooperative computing and cloud computing []. The use of Bayesian estimation methods such as the popular Kalman filter and its non-linear derivatives have become especially prevalent in application due to their recursive, often optimal, estimation properties and their suitability for modelling cross-correlations between local estimates [5, 6]. The handling of these cross-correlations, especially when they are not known in advance, is a common difficulty in state estimation and is tied to the challenges within the field [7]. Much of the work in this field, including some of the methods presented in this thesis, handle these challenges.

While the challenges in data fusion and state estimation that relate to correlation errors are a well-established field of research, widespread advancements in distributed computing and uses of public networks for sensor communication have led the additional requirements of data privacy and state secrecy to become particularly relevant in recent years [8, 9].

Problems that require careful consideration of correlations errors while simultaneously guaranteeing a level of security for the participants involved are therefore

Typical cryptographic secrecy involves hiding all transferred data such that unauthorised parties gain no information from acquired encryptions. This can often be achieved irrespective of the estimation algorithms by using common symmetric and public-key encryption schemes such as the Advanced Encryption Standard (AES) [] and the Rivest-Shamir0Adleman cryptosystem (RSA) [], respectively.

—

more complicated examples have different requirements where these are not suitable

more complicated schemes

differential privacy as a statistical security

Due to the nature of cryptographic goals and proofs in distributed environments, security goals in estimation and fusion are often very context-specific and have led to numerous solutions for various scenarios and security desires

leads us into the state-of-the-art

1.1. State-of-the-Art and Research Questions

As mentioned in the previous section, the nature of cryptographic notions in distributed environments typically requires communications between participating parties to be known exactly. This, in turn, has led to many, otherwise, general estimation algorithms, being restricted in some way to make communication and security easier to discuss.

note privacy-preserving expression in terms of data confidentiality

For example, [aristov] presents a distributed Kalman filter, namely an Information filter, where sensor measurements and measurement errors are kept private to the measuring sensors only, while the final estimation update (sum of this information) is leaked to an estimator. For this to be achieved, however, sensors must form a hierarchical communication structure and measurement models must be linear, restricting the otherwise more broadly applicable information filter and its non-linear variants.

[proloc]

Pisac and pwsah papers

aggregation papers

differentially private Kalman filtering

privacy-preserving optimisation with security based on statistical estimation

added noise estimation

–

privacy-preserving image-based localisation

eavesdropper paper with a secure return channel and a lossier channel for eavesdroppers

GPS

chaotic system paper

physical layer noise paper (similar to chaotic noise paper)

–

The two different approaches, restricting existing broad estimation methods in some ways to make cryptographic analysis plausible and ignoring formal security when assumptions and conclusions are intuitive demonstrate a gap in the existing literature and bring us to the target research topics this thesis aims to explore.

- dot point topics

These broad topics aim to fulfil the goal of generalisable but cryptographically provable estimation and fusion methods in distributed environments and lead to the concrete problems tackled in this work

1.2. Contributions

The contributions tackle the research topics in section .. by considering three concrete problems that coincide with the broader problems in the field

1.3. Thesis Structure

Each chapter includes relevant related literature to the specific problem and a formal problem formalisation before presenting the novel solutions.

2. Preliminaries

When introducing the novel estimation and cryptographic methods in this thesis, we make use of several existing algorithms. In this section, an overview of these methods is given

2.1. Estimation Preliminaries

As introduced in section .., the estimation and fusion algorithms considered are Bayesian and based on the Kalman filter and derivatives. The linear KF and linearising EKF are used explicitly and are the focus of the presented preliminaries.

2.1.1. Kalman Filter

2.1.2. Kalman Filter Optimality

2.1.3. Extended Kalman Filter

2.1.4. Information Filter

2.1.5. Extended Information Filter

2.1.6. Fast Covariance Intersection

Covariance Intersection (CI), introduced in [julierNondivergentEstimationAlgorithm1997], provides a consistent state estimate fusion algorithm when cross-correlations are not known. The resulting fused estimate $\underline{\hat{x}}$ and covariance $\mathbf{P}$ can be easily derived from its equations

$$\mathbf{P}^{-1} = \sum_{i=1}^{n} \omega_i \mathbf{P}_i^{-1},\ \mathbf{P}^{-1}\underline{\hat{x}} = \sum_{i=1}^{n} \omega_i \mathbf{P}_i^{-1}\underline{\hat{x}}_i \ . \tag{2.1}$$

Note that (2.1) computes the fusion of the information vectors and information matrices defined in [niehsenInformationFusionBased2002] and reduces the fusion to a weighted sum. Values for weights ω_i must satisfy

$$\omega_1 + \omega_2 + \cdots + \omega_n = 1,\ 0 \leq \omega_i \leq 1 \ , \tag{2.2}$$

which guarantees consistency of the fused estimates. They are chosen in a way to speed up convergence and minimize the error by minimizing a certain specified property of the resulting fused estimate covariance. One such property, the covariance trace, requires the solution to

$$\underset{\omega_1,\ldots,\omega_n}{\arg\min}\{\mathrm{tr}(\mathbf{P})\} = \underset{\omega_1,\ldots,\omega_n}{\arg\min}\left\{\mathrm{tr}\left(\left(\sum_{i=1}^{n}\omega_i\mathbf{P}_i^{-1}\right)^{-1}\right)\right\} \tag{2.3}$$

for computing weights ω_i. However, minimizing this non-linear cost function can be very computationally costly and has led to the development of faster approximation techniques.

The Fast Covariance Intersection (FCI) algorithm from [niehsenInformationFusionBased2002] is a non-iterative method for approximating the solution to (2.3) without the loss of guaranteed consistency. It is computed by defining a new constraint

$$\omega_i\,\mathrm{tr}(\mathbf{P}_i) - \omega_j\,\mathrm{tr}(\mathbf{P}_j) = 0,\ \ i,j = 1,2,\ldots,n \tag{2.4}$$

on ω_i and solving the resulting equations instead. In the two sensor case, this results in the solving of

$$\omega_1\,\mathrm{tr}(\mathbf{P}_1) - \omega_2\,\mathrm{tr}(\mathbf{P}_2) = 0,\ \ \omega_1 + \omega_2 = 1\ \ . \tag{2.5}$$

When computed for n sensors, the highly redundant (2.4) can have its largest linearly independent subset represented by

$$\omega_i\,\mathrm{tr}(\mathbf{P}_i) - \omega_{i+1}\,\mathrm{tr}(\mathbf{P}_{i+1}) = 0,\ \ i = 1,2,\ldots,n-1\ \ , \tag{2.6}$$

and requires the solution to the linear problem

$$\begin{bmatrix} \mathcal{P}_1 & -\mathcal{P}_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \mathcal{P}_{n-1} & -\mathcal{P}_n \\ 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_{n-1} \\ \omega_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}\ \ , \tag{2.7}$$

where we let $\mathcal{P}_i = \mathrm{tr}(\mathbf{P}_i)$.

Our proposed filter aims to solve FCI fusion, namely (2.1) and (2.7), using only encrypted values from each sensor i, and leaking only the weight values $\omega_1,\ldots,\omega_n$.

2.2. Encryption Preliminaries

Used cryptographic notions and schemes are summarised here. In addition, the encoding of floating point numbers to integers suitable for encryption by the presented schemes is introduced as well.

2.2.1. Meeting Cryptographic Notions

2.2.2. Paillier Homomorphic Encryption Scheme

The Paillier encryption scheme [paillierPublicKeyCryptosystemsBased1999] is an additively homomorphic encryption scheme that bases its security on the decisional composite residuosity assumption (DCRA) and meets the security notion of IND-CPA. Key generation of the Paillier scheme is performed by choosing two sufficiently large primes p and q, and computing $N = pq$. A generator g is also required for encryption, which is often set to $g = N + 1$ when p and q are of equal bit length [katzIntroductionModernCryptography2008]. The public key is defined by (N, g) and the secret key by (p, q).

Encryption of a plaintext message $a \in \mathbb{Z}_N$, producing ciphertext $c \in \mathbb{Z}^*_{N^2}$, is computed by

$$c = g^a \rho^N \pmod{N^2} \tag{2.8}$$

for a randomly chosen $\rho \in \mathbb{Z}_N$. Here, ρ^N can be considered the noise term which hides the value $g^a \pmod{N^2}$, which due to the scheme construction, is an easily computable discrete logarithm. The decryption of a ciphertext is computed by

$$a = \frac{L(c^\lambda \pmod{N^2})}{L(g^\lambda \pmod{N^2})} \pmod{N} \tag{2.9}$$

where $\lambda = \mathsf{lcm}(p-1, q-1)$ and $L(\psi) = \frac{\psi - 1}{N}$.

In addition to encryption and decryption, the following homomorphic functions are provided by the Paillier scheme. $\forall a_1, a_2 \in \mathbb{Z}_N$,

$$\mathcal{D}(\mathcal{E}(a_1)\mathcal{E}(a_2) \pmod{N^2}) = a_1 + a_2 \pmod{N}\,, \tag{2.10}$$

$$\mathcal{D}(\mathcal{E}(a_1)g^{a_2} \pmod{N^2}) = a_1 + a_2 \pmod{N}\,, \tag{2.11}$$

$$\mathcal{D}(\mathcal{E}(a_1)^{a_2} \pmod{N^2}) = a_1 a_2 \pmod{N}\,. \tag{2.12}$$

2.2.3. Joye-Libert Aggregation Scheme

The Joye-Libert privacy-preserving aggregation scheme [joyeScalableSchemePrivacyPreserving2013] is a scheme defined on time-series data and meets the security notion of Aggregator Obliviousness (AO) [shiPrivacyPreservingAggregationTimeSeries2011]. Similarly to the Paillier scheme, it bases its security on the DCRA. A notable difference to a public-key encryption scheme is its need for a trusted party to perform the initial key generation and distribution.

Key generation is computed by first choosing two equal-length and sufficiently large primes p and q, and computing $N = pq$. A hash function $H : \mathbb{Z} \to \mathbb{Z}^*_{N^2}$ is defined and the public key is set to (N, H). n private keys are generated by choosing sk_i, $i \in \{1, \ldots, n\}$, uniformly from $\mathbb{Z}_{N^2}$ and distributing them to n participants (whose values are to be aggregated), while the last key is set as

$$sk_0 = -\sum_{i=1}^{n} sk_i\,, \tag{2.13}$$

and sent to the aggregator.

Encryption of plaintext $a_i^{(t)} \in \mathbb{Z}_N$ to ciphertext $c_i^{(t)} \in \mathbb{Z}_{N^2}$ at instance t is computed by user i as

$$c_i^{(t)} = (N+1)^{a_i^{(t)}} H(t)^{sk_i} \pmod{N^2}\,. \tag{2.14}$$

Here, we can consider $H(t)^{sk_i}$ the noise term which hides the easily computable discrete logarithm $g^{a_i^{(t)}} \pmod{N^2}$, where $g = N + 1$ (as with the Paillier scheme above).

When all encryptions $c_i^{(t)}$, $i \in \{1, \ldots, n\}$ are sent to the aggregator, summation and decryption of the aggregated sum are computed by the functions

$$c^{(t)} = H(t)^{sk_0} \prod_{i=1}^{n} c_i^{(t)} \pmod{N^2} \tag{2.15}$$

and

$$\sum_{i=1}^{n} a_i^{(t)} = \frac{c^{(t)} - 1}{N} \pmod{N}\,. \tag{2.16}$$

Correctness follows from $\sum_{i=0}^{n} sk_i = 0$, and thus

$$\begin{aligned}
&H(t)^{sk_0} \prod_{i=1}^{n} c_i^{(t)} \pmod{N^2} \\
\equiv &H(t)^{sk_0} \prod_{i=1}^{n} (N+1)^{a_i^{(t)}} H(t)^{sk_i} \pmod{N^2} \\
\equiv &H(t)^{\sum_{j=0}^{n} sk_j} \prod_{i=1}^{n} (N+1)^{a_i^{(t)}} \pmod{N^2} \\
\equiv &(N+1)^{\sum_{i=1}^{n} a_i^{(t)}} \pmod{N^2}\,,
\end{aligned}$$

removing all noise terms.

2.2.4. Lewi Order-Revealing Encryption Scheme

For ORE, we use the Lewi symmetric-key Left-Right ORE scheme as it has the added property of only allowing certain comparisons between cyphertexts. This property can be used to decide which values may not be compared, which will be shown in the section ... It is described as follows: two encryption functions allow integers to be encrypted as either a "Left" (L) or "Right" (R) encryption by

$$\begin{aligned}
\mathrm{enc}^{\mathrm{L}}_{\mathrm{ORE}}(\mathrm{k}, \mathrm{x}) = \mathcal{E}^{\mathrm{L}}_{\mathrm{ORE,k}}(\mathrm{x}) \ , \\
\mathrm{enc}^{\mathrm{R}}_{\mathrm{ORE}}(\mathrm{k}, \mathrm{y}) = \mathcal{E}^{\mathrm{R}}_{\mathrm{ORE,k}}(\mathrm{y}) \ ,
\end{aligned} \tag{2.17}$$

and only comparisons between an L and an R encryption are possible, by

$$\mathrm{cmp}_{\mathrm{ORE}}(\mathcal{E}^{\mathrm{L}}_{\mathrm{ORE}}(\mathrm{x}),\ \mathcal{E}^{\mathrm{R}}_{\mathrm{ORE}}(\mathrm{y})) = \mathrm{cmp}(\mathrm{x}, \mathrm{y}) \ . \tag{2.18}$$

Note that no decryption function is provided as only encryptions are required to provide a secure comparison. The Lewi ORE encryption scheme provides security against the simulation-based security model [chenettePracticalOrderRevealingEncryption2016] but is not secure against the IND-OCPA model.

2.2.5. Encoding Numbers for Encryption

In both the Paillier and Joye-Libert schemes, as well as the one we introduce, meaningful inputs a are bounded to $a \in \mathbb{Z}_N$. For this reason, real-valued estimation variables require quantisation and integer mapping for encryption and aggregation. We will rely on a generalised Q number encoding [oberstarFixedPointRepresentationFractional2007] due to implementation simplicity and applicability.

We will consider a subset of rational numbers in terms of a range $M \in \mathbb{N}$ and fractional precision $\phi \in \mathbb{N}$. This contrasts with the common definition in terms of total and fractional bits [oberstarFixedPointRepresentationFractional2007, schulzedarupEncryptedCooperativeControl2019, farokhiSecurePrivateControl2017], but allows for a direct mapping to integer ranges which are not a power of two. A rational subset $\mathbb{Q}_{M,\phi}$ is then given by

$$\mathbb{Q}_{M,\phi} = \left\{ o \;\middle|\; \phi o \in \mathbb{N} \wedge -\left\lfloor \frac{M}{2} \right\rfloor \leq \phi o < \left\lfloor \frac{M}{2} \right\rfloor \right\}, \tag{2.19}$$

and we can quantize any real number a by taking the nearest rational $o \in \mathbb{Q}_{M,\phi}$, that is, $\arg\min_{o \in \mathbb{Q}_{M,\phi}} |a - o|$. In this form, mapping rationals $\mathbb{Q}_{M,\phi}$ to an encryption range $\mathbb{Z}_N$ is achieved by choosing $M = N$ and handling negatives by modulo arithmetic. Additionally, we note that the Q number format requires a precision factor ϕ to be removed after each encoded multiplication. This is captured by a third parameter d; the number of additional precision factors present in encodings.

The function for *combined* quantisation and encoding, $\mathsf{E}_{M,\phi,d}(a)$, of a given number $a \in \mathbb{R}$ and with an integer range $\mathbb{Z}_M$, precision ϕ and scaling for d prior encoded multiplications is given by

$$\mathsf{E}_{M,\phi,d}(a) = \left\lfloor \phi^{d+1} a \right\rceil \pmod{M}. \tag{2.20}$$

Decoding of an integer $u \in \mathbb{Z}_M$, is given by

$$\mathsf{E}^{-1}_{M,\phi,d}(u) = \begin{cases} \dfrac{u \pmod{M}}{\phi^{d+1}}, & u \pmod{M} \leq \left\lfloor \dfrac{M}{2} \right\rfloor \\ -\dfrac{M - u \pmod{M}}{\phi^{d+1}}, & \text{otherwise} \end{cases}. \tag{2.21}$$

This encoding scheme provides the following homomorphic operations,

$$\begin{aligned} \mathsf{E}_{M,\phi,d}(a_1) + \mathsf{E}_{M,\phi,d}(a_2) \pmod{M} = \\ \mathsf{E}_{M,\phi,d}(a_1 + a_2) \end{aligned} \tag{2.22}$$

and

$$\begin{aligned} \mathsf{E}_{M,\phi,d}(a_1)\mathsf{E}_{M,\phi,d}(a_2) \pmod{M} = \\ \mathsf{E}_{M,\phi,d+1}(a_1 a_2), \end{aligned} \tag{2.23}$$

noting that when $M = N$, the operations and modulus correspond with those in the Paillier homomorphic operations (2.10), (2.11) and (2.12), and the Joye-Libert sum (2.16).

In general, the choice of a large precision parameter ϕ may reduce quantisation errors introduced in (2.20), but risks overflow after too many multiplications. Given the largest number of encoded multiplications, d_{max}, and the largest value to be encoded a_{max}, the parameter should be chosen such that

$$\left|\phi^{d_{max}+1} a_{max}\right| < \left\lfloor \frac{M}{2} \right\rfloor . \tag{2.24}$$

In practice, N is typically very large ($N > 2^{1024}$) and this condition can be ignored when $M = N$, as ϕ can be made sufficiently large to make quantisation errors negligible.

3. Estimate Fusion on an Untrusted Cloud

3.1. Problem Formulation

3.2. Related Literature

3.3. Confidential Cloud Fusion Leaking Fusion Weights

3.4. Confidential Cloud Fusion Without Leakage

3.5. Conclusions

4. Distributed Non-Linear Measurement Fusion with Untrusted Participants

4.1. Problem Formulation

4.2. Related Literature

4.3. Confidential Range-Only Localisation

4.3.1. Unidirectional Alternative

4.3.2. Solvable Sub-Class of Non-Linear Measurement Models

4.4. Conclusions

5. Provable Estimation Performances

5.1. Problem Formulation

5.2. Related Literature

5.3. Covariance Privilege

5.4. Privileged Estimation for Linear Systems

5.4.1. Extension to Non-Linear Systems

5.5. Fusion in Privileged Estimation Environments

5.6. Conclusions

6. Conclusion

A. Linear-Combination Aggregator Obliviousness

B. Cryptographic Proof of LCAO Scheme Security

List of Figures

List of Tables

Bibliography

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Dover Publications, 1979.

[2] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* John Wiley & Sons, 2006.

[3] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems.* CRC press, 1998.

[4] M. Liggins, C. Y. Chong, D. Hall, and J. Llinas, *Distributed Data Fusion for Network-Centric Operations.* CRC Press, 2012.

[5] C. Y. Chong, "Forty Years of Distributed Estimation: A Review of Noteworthy Developments," in *IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2017)*, 2017, pp. 1–10.

[6] A. J. Haug, *Bayesian Estimation and Tracking: A Practical Guide.* John Wiley & Sons, 2012.

[7] B. Noack, J. Sijs, M. Reinhardt, and U. D. Hanebeck, *Treatment of Dependent Information in Multisensor Kalman Filtering and Data Fusion.* CRC Press, 2017, pp. 169–192.

[8] 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)*, 2011, pp. 114–119.

[9] K. Ren, C. Wang, and Q. Wang, "Security Challenges for the Public Cloud," *IEEE Internet Computing*, vol. 16, no. 1, pp. 69–73, 2012.