Perspectives on Probabilistic Graphical Models

DONG LIU

Doctoral Thesis in Electrical Engineering
Stockholm, Sweden 2020

Devision of Information Science and Engineering
TRITA-EE XXXX KTH, School of Electrical Engineering and and Computer Science
ISSN ....
ISBN ....
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i Elektroteknik onsdag den 13 september 2017 klockan 13.15 i F3, Lindstedtsvägen 26, Stockholm.

© 2020 Dong Liu, unless otherwise noted.

Tryck: Universitetsservice US AB

To my beloved

Abstract

Sammanfattning

Acknowledgements

This thesis could not be finished without the help and support from many professors, colleagues, friends, and my family. It is my pleasure to acknowledge people who give me help, guidance, and encouragement.

First and foremost, I would like to thank my main supervisor Assoc. Prof. Tobias Oechtering. You offer me the opportunity to pursue Ph.D. degree, always provide me strong support and helpful guidance in the research, and share me positive philosophy.

I am deeply grateful to my co-supervisor Assoc. Prof. Joakim Jaldén for the valuable discussions in the DeWiNe group meetings. I would like to thank Assoc. Prof. Deniz Gündüz for hosting my visit of ICL and supervising my research in the COPES project.

I would like to express my sincere gratitude to Assoc. Prof. Stefano Marano from University of Salerno for acting as the opponent, and to the grading board members: Assoc. Prof. Edith Ngai from Uppsala University, Assoc. Prof. Aikaterini Mitrokotsa from Chalmers University of Technology, Assoc. Prof. Pablo Piantanida from Supelec Télécom, and Prof. Mats Bengtsson. I would like to thank Prof. Mikael Skoglund for being the defense chair and Prof. Magnus Jansson for advance thesis review.

I must thank all my colleagues for creating the enjoyable working environment. I would like to thank Dr. Kittipong Kittichokechai for supervising my master thesis project and helping me in the beginning of my Ph.D. study. I feel grateful to work with the seniors: Prof. Peter Händel, Prof. Lars Kildehøj, Assoc. Prof. Ming Xiao, Assoc. Prof. Ragnar Thobaben, Assoc. Prof. James Gross, Assoc. Prof. Markus Flierl, Assis. Prof. Saikat Chatterjee, Dr. Satyam Dwivedi, Dr. Isaac Skog, Dr. Jinfeng Du, Dr. Ali Zaidi, Dr. Hieu Do, Dr. Ricardo Blasco Serrano, Dr. Mattias Andersson, Dr. Jalil Taghia, Dr. Amirpasha Shirazinia, Dr. Dennis Sundman, Dr. Frédéric Gabry, Dr. Hamed Farhadi, Dr. Maksym Girnyk, Dr. Iqbal Hussain, Dr. Sheng Huang, Dr. Zhao Wang, Dr. Efthymios Stathakis, Dr. Tai Do, Dr. Haopeng Li, Dr. Leefke Grosjean, Dr. Hadi Ghauch, Dr. Majid Gerami, Dr. Alla Tarighati, Dr. Nima Najari Moghadam, Dr. Hussein Mohammed Al Zubaidy, Dr. Germán Bassi, Dr. Rami Mochaourab, Dr. Antonios Pitarokoilis, Dr. Qiwen Wang, Mr. Peter Larsson, and Mr. Ahti Ainomäe. I devote special thanks to Ms. Raine Tiivel, Ms. Dora Söderberg, and Ms. Tove Schwartz for careful and efficient

administrative support.

It is a pleasure to share the office with my talented officemate Minh Thanh Vu. I really enjoy the relaxed lunch time with Guang Yang, Bing Li, Nan Qi, Le Phuong Cao, Dr. Lin Zhang, Yu Ye, Zhengquan Zhang, Dong Liu, and Shaocheng Huang. I am grateful to have Marie Maros as my teaching partner, and to have Pol del Aguila Pla, Arash Owrang, Håkan Carlsson as my candy corner partners. It is my pleasure to have the nice fellows: C V Ramana Reddy Avula, Baptiste Cavarec, Hasan Basri Celebi, Henrik Forssell, Jin Huang, Xinyue Liang, Sahar Imtiaz, Robin Larsson Nordström, Du Liu, Nan Li, Alireza Mahdavi Javid, Sina Molavipour, Boules Atef Mouris, Sebastian Schiessl, Arun Venkitaraman, Johan Wahlström, and Hanwei Wu.

I am always indebted to my master study adviser Prof. Miguel Ángel Lagunas in UPC for encouraging me to do Ph.D. research. I would like to express my sincere gratitude to all professors who taught me during my master study in UPC and KTH. I would like to thank the master program assistant Ms. Lise Vierning for her Catalan warmth.

Finally, I would like to express my gratitude to my parents, my elder brother, and my precious sister for their love and support.

Zuxing Li
Stockholm, July 2017

Contents

Acronyms and Notations

Acronyms

AD	adversary
CPS	cyber-physical system
EMU	energy management unit
EP	energy provider
ES	energy storage
EVE	eavesdropper
FC	fusion center
GDPR	General Data Protection Regulation
i.i.d.	identically independently distributed
KKT	Karush-Kuhn-Tucker
LRC	likelihood-ratio chain
LRT	likelihood-ratio test
MDP	Markov decision process
p.d.f.	probability density function
p.m.f.	probability mass function
PBPO	person-by-person optimization
RES	renewable energy source
ROC	receiver operating characteristic

Notations

X	random variable
x	realization of the random variable X
$\mathcal{X}$	alphabet of the random variable X
X_i^k	random sequence $(X_i, \ldots, X_k)$
x_i^k	realization of the random sequence X_i^k
$\mathcal{X}_i^k$	alphabet of the random sequence X_i^k
X^k	random sequence $(X_1, \ldots, X_k)$
x^k	realization of the random sequence X^k
$\mathcal{X}^k$	alphabet of the random sequence X^k
$X_i^{k\setminus n}$	random sequence $(X_i, \ldots, X_{n-1}, X_{n+1}, \ldots, X_k)$
$x_i^{k\setminus n}$	realization of the random sequence $X_i^{k\setminus n}$
$\mathcal{X}_i^{k\setminus n}$	alphabet of the random sequence $X_i^{k\setminus n}$
$X^{k\setminus n}$	random sequence $(X_1, \ldots, X_{n-1}, X_{n+1}, \ldots, X_k)$
$x^{k\setminus n}$	realization of the random sequence $X^{k\setminus n}$
$\mathcal{X}^{k\setminus n}$	alphabet of the random sequence $X^{k\setminus n}$
$\|\|\cdot\|\|$	set cardinality
f_X	p.d.f. of the continuous random variable X
p_X	p.m.f. of the discrete random variable X
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathrm{D}(\cdot\|\|\cdot)$	Kullback-Leibler divergence
$\mathrm{D}_\tau(\cdot\|\|\cdot)$	τ-th order Rényi divergence
$\mathrm{C}(\cdot,\cdot)$	Chernoff information
$\mathrm{E}[\cdot]$	expectation
$\partial\cdot$	boundary of a closed set

$\hat{\partial}\cdot$	upper boundary of a two-dimensional closed set
$\check{\partial}\cdot$	lower boundary of a two-dimensional closed set
$\log(\cdot)$	natural logarithm

Chapter 1

Introduction

1.1 Privacy Challenge in Cyber-Physical System

A cyber-physical system (CPS) consists of two major components: a physical process and a cyber system. As shown in Figure **??**, the physical process, which can be a natural phenomenon or a man-made physical system, is monitored and controlled by the cyber system, which typically is a networked system of several tiny devices with sensing, computation, and communication capabilities [37]. There have been a large number of proposed CPS applications, such as smart house, smart grid, eHealth, assisted living, and etc. They are envisioned to form a smart environment which will greatly benefit the users. A typical CPS often collects a huge amount of privacy-sensitive information for data analysis and decision making. The information enables the system to make smart decisions through sophisticated algorithms. However, a privacy leakage could potentially happen in any stage(s) of data collection, data transmission, data processing, or data storage. On the other hand, there is an increasing demand to protect privacy, e.g., the EU General Data Protection Regulation (GDPR) [1] is to replace the Data Protection Directive 95/46/EC and is designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens data privacy, and to reshape the way organizations across the region approach data privacy. Therefore, research on privacy protection for CPSs attracts much attention nowadays. In the next, researches on privacy protection are briefly reviewed.

1.2 Literature Review

In face of a variety of privacy threats, a large amount of fruitful works have been done. As a conventional privacy-preserving tool, cryptography was reported in many papers, e.g., [27], to protect the private data. Although it is an effective method, cryptography is not always applicable because of its demands of high computation capability, high power consumption, and complicated key management.

Some studies investigated the privacy-protection in the multi-hop routing, e.g., the reputation-based scheme [5] and the broadcast authentication [21].

All these aforementioned technologies use additive privacy functionality blocks and cannot protect privacy against the authorized data recipient. GDPR calls for an authorized data recipient to hold and process only the data absolutely necessary for the completion of its duties as well as limiting the access to personal data to those needing to act out the processing [1]. To this end, GDPR advocates the *p*rivacy-by-design approach which can "inherently" preserve privacy through the inclusion of data protection from the onset of the designing of systems rather than an addition afterward. Depending on the physical-layer operations, privacy-by-design approaches can further be categorized into different classes.

Until now, most privacy-by-design approaches focus on the data transmission stage, which corresponds to sensing and communication in the physical layer of a CPS. The study on the wire-tap channel [38] derives the secrecy capacity. Based on the theory of wire-tap channel, people have developed privacy schemes, such as artificial noise [8] and cooperative jamming [32]. Recently, secure data compression in source coding also attracts much attention [16].

Besides data transmission, there are other physical-layer operations for a CPS application. Consider the eHealth system in Figure **??** which consists of wearable or embedded sensors and a handheld terminal. The sensors collect different raw data, e.g., heart rate, body temperature, or blood pressure, process the raw data to make sensor decisions, and then transmit the sensor decisions to the terminal. The terminal makes the final conclusion of the user health condition based on the received sensor decisions. This eHealth CPS operation can be seen as a statistical inference on the user health condition through a sensor network. The statistical inference operation in the physical layer can be modeled as a hypothesis test[1] or an estimation.

A statistical inference can be done centrally or distributively. The corresponding hypothesis testing theorems have been well established. A brief introduction of the centralized hypothesis testing problems based on the Bayesian approach and Neyman-Pearson approach was presented in [34]. In [33, 36], the basic distributed hypothesis testing problems were summarized of different formulations, topologies, processing and communication constraints. Extended distributed hypothesis testing problems were reviewed in [4].

When the privacy-by-design approach is used, the privacy-preserving objective is taken into account in the hypothesis tests and the established hypothesis testing theorems need to be revised. There are two pioneering works [22, 26] which considered a distributed hypothesis testing network in the presence of an eavesdropper. In [22], the eavesdropper is assumed to be only interested in the data transmission state between the remote decision makers and the fusion center. However, an eavesdropper in reality can be more aggressive and tries to intercept the

[1]Note that there are other terms, e.g., detection and classification, used to refer to a hypothesis test in many literatures.

transmitted information for malicious purposes. When the eavesdropper makes a hypothesis test based on the intercepted decisions of remote decision makers, it has been shown in [26] that a likelihood-ratio test (LRT) is an asymptotically optimal decision strategy of a remote decision maker under a constraint on the eavesdropping performance measured by a Kullback-Leibler divergence. Whereas, it was not provided how to design the asymptotically optimal privacy-preserving distributed hypothesis testing network. A recent work [23] characterized an achievable rate-error-equivocation region of a distributed hypothesis testing network with communication and privacy constraints. Unfortunately, the converse proof was only given for a special case. Some works devised privacy schemes based on the assumption that the eavesdropper is uninformed of certain parameters. In [25, 29], they studied stochastic encryption methods where remote decision makers intentionally generate error to confuse the eavesdropper which is uninformed of the error statistics. Similarly, works [11,12] proposed channel aware encryption methods to design the transmission scheme between the remote decision makers and the fusion center based on the channel states which are not known by the eavesdropper.

The privacy issue is also addressed in a distributed estimation context. Some works used a similar method of stochastic encryption, e.g., a stochastic cipher was utilized in [2] to protect privacy in a network where both the fusion center and eavesdropper make maximum-likelihood estimations. In [10], the optimal power allocation scheme was studied in a decentralized minimum mean square error estimation susceptible to eavesdropping.

In the aforementioned privacy-by-designs of statistical inference operation in the physical layer, the privacy leakage can also be modeled as a statistical inference made by an eavesdropper or adversary. Besides hypothesis test and estimation, the statistical inference risks have been discussed and evaluated in computer science through the differential privacy. Differential privacy [6] was proposed to guarantee the statistical privacy by sanitizing mechanisms when an adversary has access to two neighbor databases. In [14], the degradation of the differential privacy level under adaptive interactions was characterized. In [28], for any statistical estimator and input distribution satisfying a regularity condition, it was proved that there exists a differentially private estimator with the same asymptotic output distribution. In [17], 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 [13]. Relations between different formalisms for statistical inference privacy were discussed in [3].

Smart grid is one of the most attractive CPS applications. Real-time information about energy demands and advanced control and communication technologies enable more efficient energy generation and distribution in smart grids [31]. Real-time energy demand information is provided to the energy provider by the smart meters installed at consumer premises. While high-resolution meter readings are essential for monitoring and control tasks, they also reveal sensitive private information about the consumers [24,30]. A number of privacy-preserving technologies have been developed for the smart meter privacy problem in the recent years. In [18], an

encryption method was proposed to protect the privacy of an individual consumer through data aggregation in the neighborhood. In [15], a privacy scheme was devised by scheduling delay-tolerable appliances to hide the energy demand profiles of others. Most of the literature focuses on the manipulation of meter readings to preserve privacy, e.g., adding a noise sequence on the meter readings. There is a growing interest in guaranteeing privacy by directly altering the energy demands from the energy provider. This can be achieved by an energy management of renewable energy supplies or energy storage charge/discharge flows to filter the real energy demand characteristics. Information-theoretic approaches to these problems have been studied in [7, 9, 31, 35]. In these works, the privacy leakage measure is the mutual information rate between the energy demand sequence and the energy supply sequence. The information-theoretic measure can be adopted regardless of the real adversary behavior. However, it lacks an operational meaning. An optimal privacy-preserving energy management of the renewable energy supply or the energy storage charge/discharge is designed to minimize the mutual information rate. Based on the observation that a constant meter reading sequence does not leak any privacy, another privacy-preserving idea was proposed in [39] to utilize an energy storage to minimize the variance of random energy supplies from the energy provider. Similar to the information-theoretic privacy leakage measure, a variance does not have a clear operational meaning. Recently, a hypothesis testing measure of smart meter privacy was proposed in [20] and the discussion of a privacy-preserving energy management was based on an infinite-capacity energy storage. The hypothesis testing measure has a clear operational meaning while it limits the adversary behavior to be a hypothesis test. System memory is commonly inevitable in the energy management problems, e.g., the utilization of a finite-capacity energy storage. In [19, 40, 41], a privacy-preserving energy management in the presence of an energy storage was cast to a Markov decision process framework.

1.3 Thesis Outline

The general research question of this work is how to realize privacy-preserving CPSs. In this thesis, the privacy-by-design approaches are investigated in the contexts of a distributed sensor network and a smart meter system. The sensing data or the energy data processed in the CPS is driven by a privacy-sensitive unknown physical process, e.g., health condition or life style of the user. A such unknown physical process can be seen as a hypothesis. In this work, hypothesis tests are assumed for the physical-layer operation of a CPS and the privacy leakage. The following questions are to be discussed in the remaining chapters:

Chapter 2

In this chapter, the basics of hypothesis tests are recapitulated. The Bayesian and Neyman-Pearson hypothesis testing approaches are introduced. The optimality

of deterministic likelihood-based test (or LRT) is testified in centralized and distributed hypothesis tests by using the analysis tools of hypothesis testing operation region and person-by-person optimality argument. Depending on the hypothesis testing approach, it is shown that the asymptotic hypothesis testing performance can be characterized by a Kullback-Leibler divergence or a Chernoff information.

Chapter 2

Hypothesis Testing Problems

Bibliography

[1] URL `http://www.eugdpr.org/`.

[2] T. C. Aysal and K. E. Barner. Sensor data cryptography in wireless sensor networks. *IEEE Transactions on Information Forensics and Security*, 3(2): 273–289, 2008.

[3] R. F. Barber and J. C. Duchi. Privacy and statistical risk: Formalisms and minimax bounds. eprint arXiv:1412.4451.

[4] R. S. Blum, S. A. Kassam, and H. V. Poor. Distributed detection with multiple sensors: Part II - Advanced topics. *Proceedings of the IEEE*, 85(1):64–79, 1997.

[5] Z. Cao, J. Hu, Z. Chen, M. Xu, and X. Zhou. Feedback: Towards dynamic behavior and secure routing for wireless sensor networks. In *Proceedings of AINA 2006*, pages 160–164, 2006.

[6] C. Dwork. Differential privacy. In *Proceedings of ICALP 2006*, pages 1–12, 2006.

[7] G. Giaconi, D. Gündüz, and H. V. Poor. Smart meter privacy with an energy harvesting device and instantaneous power constraints. In *Proceedings of ICC 2015*, pages 7216–7221, 2015.

[8] S. Goel and R. Negi. Guaranteeing secrecy using artificial noise. *IEEE Transactions on Wireless Communications*, 7(6):2180–2189, 2008.

[9] D. Gündüz and J. Gómez-Vilardebó. Smart meter privacy in the presence of an alternative energy source. In *Proceedings of ICC 2013*, pages 2027–2031, 2013.

[10] X. Guo, A. S. Leong, and S. Dey. Estimation in wireless sensor networks with security constraints. *IEEE Transactions on Aerospace and Electronic Systems*, PP(99):1–1, 2017.

[11] H. Jeon, J. Choi, S. W. McLaughlin, and J. Ha. Channel aware encryption and decision fusion for wireless sensor networks. *IEEE Transactions on Information Forensics and Security*, 8(4):619–625, 2013.

[12] H. Jeon, D. Hwang, J. Choi, H. Lee, and J. Ha. Secure type-based multiple access. *IEEE Transactions on Information Forensics and Security*, 6(3):763–774, 2011.

[13] Z. Ji, Z. C. Lipton, and C. Elkan. Differential privacy and machine learning: A survey and review. eprint arXiv:1412.7584.

[14] P. Kairouz, S. Oh, and P. Viswanath. The composition theorem for differential privacy. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1–10, 2015.

[15] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Proceedings of SmartGridComm 2010*, pages 232–237, 2010.

[16] K. Kittichokechai, T. J. Oechtering, and M. Skoglund. Secure source coding with action-dependent side information. In *Proceedings of ISIT 2011*, pages 1678–1682, 2011.

[17] J. Le Ny and G. J. Pappas. Differentially private filtering. *IEEE Transactions on Automatic Control*, 59(2):341–354, 2014.

[18] F. Li, B. Luo, and P. Liu. Secure information aggregation for smart grids using homomorphic encryption. In *Proceedings of SmartGridComm 2010*, pages 327–332, 2010.

[19] S. Li, A. Khisti, and A. Mahajan. Structure of optimal privacy-preserving policies in smart-metered systems with a rechargeable battery. In *Proceedings of SPAWC 2015*, pages 375–379, 2015.

[20] Z. Li and T. J. Oechtering. Privacy on hypothesis testing in smart grids. In *Proceedings of ITW 2015 Fall*, pages 337–341, 2015.

[21] D. Liu, P. Ning, S. Zhu, and S. Jajodia. Practical broadcast authentication in sensor networks. In *Proceedings of MobiQuitous 2005*, pages 118–129, 2005.

[22] S. Marano, V. Matta, and P. K. Willett. Distributed detection with censoring sensors under physical layer secrecy. *IEEE Transactions on Signal Processing*, 57(5):1976–1986, 2009.

[23] M. Mhanna and P. Piantanida. On secure distributed hypothesis testing. In *Proceedings of ISIT 2015*, pages 1605–1609, 2015.

[24] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 61–66, 2010.

[25] V. Nadendla. Secure distributed detection in wireless sensor networks via encryption of sensor decision. Master's thesis, Louisiana State University, 2009.

[26] V. S. S. Nadendla, H. Chen, and P. K. Varshney. Secure distributed detection in the presence of eavesdroppers. In *Proceedings of ASILOMAR 2010*, pages 1437–1441, 2010.

[27] S. Schmidt, H. Krahn, S. Fischer, and D. Wätjen. A security architecture for mobile wireless sensor networks. In *Proceedings of the First European Conference on Security in Ad-hoc and Sensor Networks*, pages 166–177, 2005.

[28] A. Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Fourty-Third Annual ACM Symposium on the Theory of Computing*, pages 813–822, 2011.

[29] R. Soosahabi, M. Naraghi-Pour, D. Perkins, and M. A. Bayoumi. Optimal probabilistic encryption for secure detection in wireless sensor networks. *IEEE Transactions on Information Forensics and Security*, 9(3):375–385, 2014.

[30] F. Sultanem. Using appliance signatures for monitoring residential loads at meter panel level. *IEEE Transactions on Power Delivery*, 6(4):1380–1385, 1991.

[31] O. Tan, D. Gündüz, and H. V. Poor. Increasing smart meter privacy through energy harvesting and storage devices. *IEEE Journal on Selected Areas in Communications*, 31(7):1331–1341, 2013.

[32] E. Tekin. The Gaussian multiple access wire-tap channel: Wireless secrecy and cooperative jamming. In *Proceedings of ITA 2007*, pages 404–413, 2007.

[33] J. N. Tsitsiklis. Decentralized detection. In *Proceedings of Advances in Statistical Signal Processing*, pages 297–344, 1993.

[34] H. L. van Trees. *Detection, Estimation, and Modulation Theory, Part I.* Wiley-Interscience, 2001.

[35] D. Varodayan and A. Khisti. Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage. In *Proceedings of ICASSP 2011*, pages 1932–1935, 2011.

[36] P. K. Varshney. *Distributed Detection and Data Fusion.* Springer-Verlag New York, Inc., 1996.

[37] E. K. Wang, Y. Ye, X. Xu, S. M. Yiu, L. C. K. Hui, and K. P. Chow. Security issues and challenges for cyber physical system. In *Proceedings of GreenCom-CPSCom 2010*, pages 733–738, 2010.

[38] A. D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54(8): 1355–1387, 1975.

[39] L. Yang, X. Chen, J. Zhang, and H. V. Poor. Optimal privacy-preserving energy management for smart meters. In *Proceedings of INFOCOM 2014*, pages 513–521, 2014.

[40] J. Yao and P. Venkitasubramaniam. On the privacy-cost tradeoff of an in-home power storage mechanism. In *Proceedings of Allerton 2013*, pages 115–122, 2013.

[41] J. Yao and P. Venkitasubramaniam. The privacy analysis of battery control mechanisms in demand response: Revealing state approach and rate distortion bounds. In *Proceedings of CDC 2014*, pages 1377–1382, 2014.