23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Security Analysis of Quantized Bayesian Estimators

Farhad Farokhi, Daniel D. Selvaratnam, and Iman Shames

***Abstract*— Security of networked Bayesian source localization algorithms is analysed in this paper. The Bayesian estimators construct the probability density function of the location of the source from quantised measurements. Understanding fundamental limits in the performance of an adversary for manipulating the posterior of the Bayesian estimator is the main focus of the paper. The analysis is performed for cases where the estimator is not aware of the presence of the adversary. The results are then generalized to the case where the estimator is aware of the attacker.**

I. Introduction

Finding the source of leakages and contaminants within an environment is an application of source localization. Source localization can be done using multi-agent robotic systems in a distributed manner [1]–[3]. In these distributed approaches, each agent relies on the measurements of other agents to estimate the location of the source collaboratively. While more efficient in the use of resources and faster in convergence, the distributed nature of these methods leaves them vulnerable to cyber-security attacks. For instance, if one of the agents is high-jacked, an adversary can feed wrong data to the other agents in an strategic manner so that they cannot localize (or mistakenly localize) the source. This paper studies such attacks and provides fundamental bounds on their impacts.

An important family of source localization algorithms rely on the Bayesian estimators to construct the probability density function of obtaining a positive signal detection for binary measurements or the distribution of chemical pollutants for more detailed quantised measurements. Security analysis of such localization methods is of interest in this paper. Specifically, understanding the fundamental limits in the performance of an adversary for manipulating the posterior of the Bayesian estimator is the main topic of this paper. First, the analysis is performed for the cases where the estimator is not aware of the presence of the adversary. This is important for analysis of covert attacks or quantifying effectiveness of security attacks prior to the detection of an adversary. Subsequently, the case where the estimator is aware of the attacker is studied. In this case, the estimator, unsurprisingly, acts more conservatively while processing the received information.

The topic of this paper has close connections to the Bayesian persuasion [4], [5] and cheap talk [6], [7] literature in economics. In these studies, a well-informed sender is passing a message to an uninformed receiver with the intention of manipulating its decision. Contrary to this paper, in those studies, the receiver is always aware of the presence of the adversary. This knowledge turns out to play a significant role on the ability of the adversary.

This paper also has close connections to the topic of security of cyber-physical systems within the control and signal processing community [8]–[11]. This is an ever expanding field of research motivated by the impact of security flaws in networked devices on the underlying dynamical systems. The analysis of the security of the Bayesian estimators for source localization has been missing from this field of research.

The rest of the extended abstract is organized as follows. The problem formulation is presented in Section II. Section III contains fundamental bounds on the performance of the adversary for fooling the Bayesian estimator. Finally, the paper is concluded in Section IV.

II. Problem Formulation

Consider the problem in which we are interested in estimating the distribution of a random variable X that takes values in a finite set $\mathbb{X}$, e.g., containing all the possible locations for a contaminant. The probability function of X is given by $p_X : \mathbb{X} \to [0,1]$, i.e., $\mathbb{P}\{X = x\} = p_X(x)$. This is the prior for X. The estimator is interested in finding the posterior $p_{X|M}(x|m) := \mathbb{P}\{X = x \,|\, M = m\}$ based on the received (possibly stochastic) message M, which takes values in a finite set $\mathbb{M}$ e.g., a binary value determining if traces of the contaminant is sensed (or not). The conditional probability of the message is given by $p_{M|X}(m|x) := \mathbb{P}\{M = m \,|\, X = x\}$. The receiver follows the Bayes' rule to compute the posterior:

$$p_{X|M}(x|m) = \frac{p_{M|X}(m|x)p_X(x)}{\sum_{x' \in \mathbb{X}} p_{M|X}(m|x')p_X(x')}. \tag{1}$$

Note that, since the message M is a random variable, the value of the conditional density function $p_{X|M}(x|m)$ at a given $x \in \mathbb{X}$ is itself a random variable.

A. Covert adversary

Consider the case where the adversary injects a message $M = m$, but the estimator is unaware of this attack so it mistakenly follows (1). It is clearly of interest to see whether

The authors are with the Department of Electrical and Electronic Engineering at the University of Melbourne, Parkville, Australia. e-mails: dselvaratnam@student.unimelb.edu.au and {ffarokhi, ishames}@unimelb.edu.au

The work of F. Farokhi was supported by the McKenzie Fellowship from the University of Melbourne, the VESKI Fellowship from the Victorian State Government, and a grant (MyIP: ID6874) from Defence Science and Technology Group (DSTG). The work of D. Selvaratnam and I. Shames was supported by a grant (MyIP: ID6874) from Defence Science and Technology Group (DSTG). D. Selvaratnam is further supported by a PhD scholarship from the University of Melbourne.

it is possible to ensure that the posterior of (1) matches a desired[1] posterior $g(x)$. This implies m must satisfy

$$\forall x \in \mathbb{X},\ p_{X|M}(x|m) = g(x).$$

Establishing the existence of solutions to this set of $|\mathbb{X}|$ equations, and finding them, is a difficult problem. We therefore relax the problem by only requiring $p_{X|M}(x|m)$ to take the desired values in expectation. Suppose the adversary knows the parameter value for a given realization is $X = x$, and transmits stochastic messages M according to $\bar{p}_{M|X}(m|x)$.

Problem 1: For a given $g : \mathbb{X} \to [0, 1],\ \sum_{x\in\mathbb{X}} g(x) = 1$, find a probability function $\bar{p}_{M|X} : \mathbb{M} \times \mathbb{X} \to [0, 1]$ such that

$$\forall x, x' \in \mathbb{X},\ \mathbb{E}_M\{p_{X|M}(x'|M) \mid X = x\} = g(x'). \quad (2)$$

Note that, in the above equation, x' corresponds to a dummy variable for evaluating $g(x')$, and x corresponds to the realization of the random variable X. Problem 1 can be relaxed further: instead of requiring (2) to hold for every possible realization $x \in \mathbb{X}$, we only require that it holds after taking a further expectation over X. This yields

$$\forall x' \in \mathbb{X},\ \mathbb{E}_{M,X}\{p_{X|M}(x'|M)\} = g(x'). \quad (3)$$

B. Covert adversary with side information

It is often the case that the estimator has access to some side information, e.g., its own measurements or measurements from honest agents. A random variable $N \in \mathbb{M}$ with conditional probability function $p_{N|X}(n|x) := \mathbb{P}\{N = n \,|\, X = x\}$ is used to show this side-channel information. It is assumed that, conditioned on the realization of X, messages M and N are statistically independent, i.e., $\mathbb{P}\{M = m, N = n|X = x\} = \mathbb{P}\{M = m|X = x\}\mathbb{P}\{N = n|X = x\}$. In this case, the update rule in (1) must be adapted to

$$p_{X|M,N}(x|m,n) = \frac{p_{N|X}(n|x)p_{M|X}(m|x)p_X(x)}{\sum_{x'\in\mathbb{X}} p_{N|X}(n|x')p_{M|X}(m|x')p_X(x')}. \quad (4)$$

As before, the adversary may follow the conditional probability function $\bar{p}_{M|X}(m|x)$ to generate its messages (for deceiving the receiver) while the receiver unknowingly follows (4). This leaves us with the following problem.

Problem 2: For a given $g : \mathbb{X} \to [0, 1],\ \sum_{x\in\mathbb{X}} g(x) = 1$, find a probability distribution $\bar{p}_{M|X} : \mathbb{M} \times \mathbb{X} \to [0, 1]$ such that

$$\forall x, x' \in \mathbb{X},\ \mathbb{E}_{M,N}\{p_{X|M,N}(x'|M,N) \mid X = x\} = g(x').$$

Similarly, this can be further relaxed to

$$\forall x' \in \mathbb{X},\ \mathbb{E}_{M,N,X}\{p_{X|M,N}(x'|m,n)\} = g(x').$$

C. Detected adversary

For the case where the receiver is aware of the adversary and its malicious intentions, the problem formulation needs to be adapted as the receiver no longer update its belief based on (1). The receiver however follows

$$p_{X|M}(x|m) = \frac{\bar{p}_{M|X}(m|x)p_X(x)}{\sum_{x'\in\mathbb{X}} \bar{p}_{M|X}(m|x')p_X(x')}. \quad (5)$$

This is similar to the Bayesian persuasion approach studied in [4]. We consider a similar question as in Problem 1, however, (3) is calculated noting the receiver's ability to follow (5).

Definition 1: The pair $(\bar{p}_{M|X}, p_{X|M})$ constitutes an equilibrium if it simultaneously satisfies $\forall x' \in \mathbb{X},\ \forall m' \in \mathbb{M}$,

$$\mathbb{E}_M\{p_{X|M}(x'|M)\} = g(x'), \quad (6a)$$

$$p_{X|M}(x|m) = \frac{\bar{p}_{M|X}(m|x)p_X(x)}{\sum_{x'\in\mathbb{X}} \bar{p}_{M|X}(m|x')p_X(x')}. \quad (6b)$$

An interesting problem in this case is to establish the existence of the equilibrium, i.e., capture the family of functions $g(x)$ for which the equilibrium exists.

With these problem formulations at hand, we are now ready to present the results of the paper.

III. Main Results

The first result of the paper discusses the case where the receiver is not aware of the presence of the adversary.

Proposition 1: In Problem 1 if the receiver is not aware of the presence of the adversary, $\mathbb{E}_M\{p_{X|M}(x'|M)|X = x\} = g(x'), \forall x, x' \in \mathbb{X}$, can be ensured *if and only if* the following set of linear equations in $(\bar{p}_{M|X}(m|x))_{m\in\mathbb{M},x\in\mathbb{X}}$ admits a non-negative solution:

$$\sum_{m\in\mathbb{M}} \alpha(m, x')\bar{p}_{M|X}(m|x) = g(x'),\ \forall x, x' \in \mathbb{X}, \quad (7a)$$

$$\sum_{m\in\mathbb{M}} \bar{p}_{M|X}(m|x) = 1, \forall x \in \mathbb{X}, \quad (7b)$$

where

$$\alpha(m, x') := \frac{p_X(x')p_{M|X}(m|x')}{\sum_{x''\in\mathbb{X}} p_{M|X}(m|x'')p_X(x'')}. \quad (8)$$

Proof: Note that

$$\begin{aligned}
&\mathbb{E}_M\{p_{X|M}(x'|M)|X = x\} \\
&\quad= \sum_{m\in\mathbb{M}} p_{X|M}(x'|m)\mathbb{P}\{M = m|X = x\} \\
&\quad= p_X(x') \sum_{m\in\mathbb{M}} \frac{p_{M|X}(m|x')\mathbb{P}\{M = m|X = x\}}{\sum_{x''\in\mathbb{X}} p_{M|X}(m|x'')p_X(x'')} \quad (9) \\
&\quad= p_X(x') \sum_{m\in\mathbb{M}} \frac{p_{M|X}(m|x')\bar{p}_{M|X}(m|x)}{\sum_{x''\in\mathbb{X}} p_{M|X}(m|x'')p_X(x'')} \\
&= \sum_{m\in\mathbb{M}} \alpha(m, x')\bar{p}_{M|X}(m|x).
\end{aligned}$$

Therefore, if we want $\mathbb{E}_M\{p_{X|M}(x'|M)|X = x\}$ to be equal to $g(x')$, the set of linear equations in (7a) must admit a solution. The rest follows from that $\bar{p}_{M|X}$ is a conditional density function so it must sum to one. ∎

Proposition 1 provides a constructive approach for checking the existence of the messaging policy $\bar{p}_{M|X}$ by the adversary to make sure that in average the posterior matches

[1] from the perspective of the adversary.

the intended function $g(x)$. In the next result, $\mathrm{supp}(\cdot)$ is used to denote the support set of the probability function of a random variable.

Corollary 1: The set of linear equations in Proposition 1 admits a solution only if $\mathrm{supp}(g) \subseteq \mathrm{supp}(p_X)$.

Proof: From (9), it is evident that $\mathbb{E}_M\{p_{X|M}(x'|M)\} = 0$ if $p_X(x') = 0$. Thus, a necessary condition for realizability of $g(x)$ is $\mathrm{supp}(g) \subseteq \mathrm{supp}(p_X)$. ∎

Corollary 1 states the-rather-obvious result that it is not possible to fool the Bayesian estimator to believe that an event with zero probability has happened.

Conjecture 1: The set of linear equations in Proposition 1 admits a solution only if $|\mathbb{M}| \geq |\mathbb{X}| + 1$.

Conjecture 1 follows from that the number of the equations in (7a) is equal to $|\mathbb{X}|(\mathbb{X}+1)$ while the number of the unknowns is $|\mathbb{X}||\mathbb{M}|$.

Proposition 2: In Problem 1 if the receiver is not aware of the presence of the adversary, $\mathbb{E}_M\{p_{X|M}(x|M)\} = g(x), \forall x \in \mathbb{X}$ can be ensured *if and only if* the following set of linear equations in $(\bar{p}_{m|x}(m'|x''))_{m'\in\mathbb{M},x''\in\mathbb{X}}$ admits a non-negative solution:

$$\sum_{m\in\mathbb{M}}\sum_{x'\in\mathbb{X}} \alpha(m,x,x')\bar{p}_{M|X}(m'|x') = g(x),\ \forall x \in \mathbb{X}, \quad (10a)$$

$$\sum_{m\in\mathbb{M}} \bar{p}_{M|X}(m|x) = 1, \forall x \in \mathbb{X}, \quad (10b)$$

where

$$\alpha(m,x,x') := \frac{p_{M|X}(m|x)p_X(x')p_X(x)}{\sum_{x''\in\mathbb{X}} p_{M|X}(m|x'')p_X(x'')}. \quad (11)$$

Proof: The proof is similar to that of Proposition 1. ∎

Now, we are ready to consider the case wide side-channel information.

Proposition 3: In Problem 2 if the receiver is not aware of the presence of the adversary, $\mathbb{E}_{M,N}\{p_{X|M,N}(x'|M,N)|X = x\} = g(x'), \forall x, x' \in \mathbb{X}$ can be ensured *if and only if* the following set of linear equations in $(\bar{p}_{M|X}(m|x))_{m'\in\mathbb{M},x\in\mathbb{X}}$ admits a non-negative solution:

$$\sum_{m\in\mathbb{M}} \beta(m,x,x')\bar{p}_{M|X}(m|x) = g(x'),\ \forall x, x' \in \mathbb{X}, \quad (12a)$$

$$\sum_{m\in\mathbb{M}} \bar{p}_{M|X}(m|x) = 1,\ \forall x \in \mathbb{X}, \quad (12b)$$

where

$$\beta(m,x,x') = \sum_{n\in\mathbb{M}} \frac{p_{N|X}(n|x')p_{M|X}(m|x')p_X(x')p_{N|X}(n|x)}{\sum_{x''\in\mathbb{X}} p_{N|X}(n|x'')p_{M|X}(m|x'')p_X(x'')}. \quad (13)$$

Proof: Note that

$$\begin{aligned}
&\mathbb{E}_{M,N}\{p_{X|M,N}(x'|M,N)\} \\
&= \sum_{m,n} p_{X|M,N}(x|m,n)\mathbb{P}\{M=m, N=n|X=x\} \\
&= \sum_{m,n} \frac{p_{N|X}(n|x')p_{M|X}(m|x')p_X(x')p_{N|X}(n|x)\bar{p}_{M|X}(m|x)}{\sum_{x''\in\mathbb{X}} p_{N|X}(n|x'')p_{M|X}(m|x'')p_X(x'')}.
\end{aligned}$$

The rest of the proof follows from the same linear reasoning as in the proof of Proposition 2. ∎

Proposition 3 shows that the addition of the side-channel information does not add to the complexity of crafting the attack. This is intuitively because the receiver can first update its prior based on the side channel information and then processes the adversary's measurement in which case attack is simply done on the posterior of the estimator based on the side-channel information.

Proposition 4: An equilibrium in the sense of Definition 1 *if and only if* $g(x) = p_X(x), \forall x \in \mathbb{X}$.

Proof: It can be shown that

$$\begin{aligned}
\mathbb{E}_M\{p_{X|M}(x|M)\} &= \sum_{m\in\mathbb{M}} \Bigg[\frac{\bar{p}_{M|X}(m|x)p_X(x)}{\sum_{x'\in\mathbb{X}} \bar{p}_{M|X}(m|x')p_X(x')} \\
&\qquad \times \sum_{x''\in\mathbb{X}} \bar{p}_{M|X}(m|x'')p(x'') \Bigg] \\
&= \sum_{m\in\mathbb{M}} p_X(x)\bar{p}_{M|X}(m|x) \\
&= p_X(x)\Bigg[\sum_{m\in\mathbb{M}} \bar{p}_{M|X}(m|x)\Bigg] \\
&= p_X(x).
\end{aligned}$$

This concludes the proof. ∎

Proposition 4 is a negative result (from the perspective of the adversary) illustrating that fooling a cautious Bayesian estimator is a difficult task *in average*.

IV. Conclusions and Future Work

In this paper, we analysed the security of the networked Bayesian source localization algorithms. In fact, fundamental limits in the performance of an adversary for manipulating the posterior of the Bayesian estimator were provided. It was shown that if the receiver is aware of the adversary (i.e., if it knows that the system is under attack) changing the posterior of the Bayesian estimator in average. This observation points to the fact that the adversary cannot fake the location of the source to a known place. However, it might be able to obfuscate the location of the source at random by maximizing the variance of the posterior. This is an important problem that should be studied in the future.

References

[1] J. Hu, L. Xie, K.-Y. Lum, and J. Xu, "Multiagent information fusion and cooperative control in target search," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1223–1235, 2013.

[2] L. Brinón-Arranz and L. Schenato, "Consensus-based source-seeking with a circular formation of agents," *European Control Conference*, pp. 2831–2836, 2013.

[3] R. Fabbiano, F. Garin, and C. Canudas de Wit, "Distributed source seeking without global position information," *IEEE Transactions on Control of Network Systems*. In Press.

[4] M. Gentzkow and E. Kamenica, "Bayesian persuasion," *American Economic Review*, vol. 101, no. 6, pp. 2590–2615, 2011.

[5] R. Alonso and O. Camara, "Bayesian persuasion with heterogeneous priors," *Journal of Economic Theory*, vol. 165, pp. 672–706, 2016.

[6] V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica: Journal of the Econometric Society*, pp. 1431–1451, 1982.

[7] M. Battaglini, "Multiple referrals and multidimensional cheap talk," *Econometrica*, vol. 70, no. 4, pp. 1379–1401, 2002.
[8] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 911–918, 2009.
[9] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.
[10] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Proceedings of the 28th International Conference on Distributed Computing Systems Workshops*, pp. 495–500, 2008.
[11] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Systems*, vol. 35, no. 1, pp. 20–23, 2015.