SECURE ANALYTICS AND RESILIENT INFERENCE FOR THE INTERNET OF THINGS

Yuan Chen, Soummya Kar, and José M. F. Moura

Carnegie Mellon University
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213 USA

ABSTRACT

Internet of Things (IoT) applications for Smart Cities, such as systems for traffic control and pollution monitoring, increasingly rely on trustworthy and secure data analytics. Proper countermeasures are needed to ensure that IoT applications function reliably under security threats. This paper studies secure analytics and resilient inference for IoT in the context of recursive parameter estimation. A team of devices makes noisy measurements of an unknown parameter, and an attacker manipulates the measurement data of a subset of the devices. We present a resilient recursive estimation algorithm that processes the measurement streams to recover the value of the parameter, even when a subset of the devices fall under attack. The estimator is guaranteed to be strongly consistent – that is, the estimate converges almost surely to the value of the parameter – as long as less than half of the devices fall under attack. We illustrate the performance of the estimator through numerical examples.

Index Terms— Estimation, Security, Internet of Things

1. INTRODUCTION

The growing pervasivness of the Internet of Things (IoT) has introduced numerous applications to improve the functionality and well-being of smart cities, including traffic control [1], health analytics [2] and pollution monitoring [3]. For example, the Array of Things project aims to deploy a network of sensors throughout the city of Chicago to observe weather conditions, pedestrian and vehicular traffic, and air quality [4]; the Clean Air Nairobi initiative examines the efficacy of using low cost sensors to measure air pollution in Nairobi [5]; and reference [3] develops a mobile sensor network using UAVs for monitoring air quality in Beijing. IoT systems also incorporate personal devices to complement static sensing infrastructure (such as the sensing nodes in the Array of Things). A crowdsensing air quality monitoring system, for example, uses data from individuals' smart phones and wearable gadgets to improve the spatial and temporal resolution of static pollutant monitoring sensor networks [6].

These applications require the processing of data streams from a collection of sensors or smart devices to recover unknown information: in pollution monitoring, for example, we need to process each sensor's data stream to construct a pollutant concentration heatmap over a city. Due to their limited computing capacity, energy-consumption constraints, and spatial distribution (particularly in Smart City applications), IoT devices are vulnerable to adversarial attack [7, 8]. This paper focuses on *data integrity* attacks, where an adversary pathologically manipulates the measurement streams of a subset of the devices. For example, in crowdsensing, an adversarial user may attack the system by intentionally transmitting falsified data from his or her personal device. To ensure the reliability of IoT applications, which depend on trustworthy data and analytics, it is necessary to develop countermeasures against these intrusions. In this paper, we study secure and resilient analytics for IoT in the context of security countermeasures for parameter estimation.

We consider a team of sensors or devices measuring an unknown parameter. The devices relay their measurement streams to a fusion center (FC), which processes the data to estimate the unknown parameter. In context of [3], for example, we may have a team of drones monitoring air pollutant concentrations in a neighborhood. The drones transmit their sensor data to the cloud, which then estimates an air quality heatmap. All of the sensors are affected by noise, and a fraction of the sensors are under adversarial attack. When a sensor is under attack, it may produce any arbitrary measurement value (as determined by the adversary). This paper presents a resilient recursive estimation algorithm: the estimator processes the measurement streams on-line and is guaranteed to be strongly consistent (i.e., as more measurement samples are processed, the estimate converges almost surely to the true value of the parameter) as long as less than half of the sensors are under attack.

Existing reactive countermeasures for parameter estimation focus on explicity detecting security intrusions and identifying which sensors are under attack. Once an attack has been identified, the FC may take corrective actions in computing its estimate. Reference [9] proposes a metric to characterize the quality of measurements collected from IoT devices. Instead of just characterizing the quality of data, this paper, unlike [9], presents an algorithm to resilently *process* at that may be compromised by an adversary. Our previous work [10] presents a fully distributed algorithm to explicitly detect adversarial devices. The algorithm in [10] raises alarms when it detects adversarial devices whose behavior would prevent the remaining nonadversarial devices from correctly estimating the unknown parameter. In contrast, this paper presents a decentralized estimation algorithm that correctly estimates the unknown parameter without explicitly detecting adversarial behavior. References [11] and [12] propose algorithms to identify compromised sensors (when the adversary is restricted to use certain probabilistic attack strategies) and use data from the remaining uncompromised sensors for inference. For general attack strategies, i.e., attackers who are not restricted to use probabilistic strategies, attack identification becomes a combinatorial problem [13, 14].

This material is based upon work supported by the Department of Energy under award number DE-OE0000779, by DARPA under agreement numbers DARPA FA8750-12-2-0291 and DARPA HR00111320007, and by the National Science Foundation under award number CCF 1513936. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

Unlike [11] and [12], which restrict attackers to follow probabilistic strategies, in this paper, we consider attackers who behave arbitrarily. To avoid the computational burden of attack identification, the algorithm in this paper does not need to identify the compromised sensors. Instead, the estimator processes the measurements in a way so as to *implicitly* mitigate the damage from the attack, without explicitly identifying the attacker's behavior. Regardless of how an attack is carried out, the estimator in this paper is guaranteed to be strongly consistent as long as less than half of the sensors are compromised. Our previous work [15, 16] developed an attack-resilient algorithm for distributed parameter recovery that did not need to explicitly identify the attack, but these algorithms required noiseless measurements. In contrast, the estimator presented in this paper handles measurement noise and attacks simultaneously, which is challenging because a smart adversary may hide the attack within the measurement noise.

The rest of this paper is organized as follows. In Section 2, we provide the sensor measurement and attack models, and we define the resilient recursive estimation problem. Section 3 presents a resilient recursive parameter estimation algorithm that iteratively updates a running estimate based on new measurements available at each time step. The estimator is strongly consistent, i.e., the estimate converges almost surely to the true value of the parameter, as long as less than half of the sensors are compromised. In Section 4, we analyze the estimator's performance and provide a proof sketch of its strong consistency. We illustrate the performance of the estimator through numerical examples (based on air quality monitoring) in Section 5, and we conclude in Section 6.

2. BACKGROUND

Consider a set of N sensors, $\{1, 2, \ldots, N\}$, where each sensor n makes a stream of measurements

$$y_n(t) = \theta^* + w_n(t) \tag{1}$$

of an unknown static parameter $\theta^* \in \mathbb{R}^M$, where $w_n(t)$ is measurement noise. We assume that the parameter θ^* is static. In practice, we are often interested estimating parameters that change over time, for example, the air pollutant concentration in an environment or traffic conditions at a road intersection [7], but these parameters may be approximated to be static over short time windows. We assume that the measurement noise $w_n(t)$ is independently and identically distributed (i.i.d.) over time with mean $\mathbb{E}[w_n(t)] = 0$ and finite covariance $\mathbb{E}[w_n(t)w_n(t)^\intercal] = \Sigma_n$ and independent across sensors.

A fusion center (FC) collects the raw measurement streams of each of the sensors. The goal of the FC is to recursively process the measurement streams and construct a sequence of estimates x_t of the parameter θ^*. A malicious adversary attempts to prevent the FC from estimating θ^* by falsifying a subset of the measurement streams. That is, the adversary attacks a subset of the sensors and replaces their measurements with arbitrary values $y_n^a(t)$. We model the effect of the sensor attack as follows

$$y_n^a(t) = \theta^* + w_n(t) + a_n(t). \tag{2}$$

Note that, through the appropriate choice of $a_n(t)$, (2) models *all* possible values of $y_n^a(t)$. The adversary may choose $y_n^a(t)$ (or $a_n(t)$) arbitrarily; we do not restrict the adversary to follow specific strategies (unlike, e.g., [11, 12], where the attacker must follow certain probabilitistic strategies).

We may partition the set of all sensors into a set of sensors that fall under attack, $\mathcal{A}$, and a set that does not fall under attack, $\mathcal{N}$. We define $\mathcal{A}$ as $\mathcal{A} = \left\{ n \in \{1, \ldots, N\} \,\middle|\, \exists t, a_n(t) \neq 0 \right\}$. That is, $\mathcal{A}$ is the set of sensors n that *may* fall under attack at any time t, but, for some t, the attacker may choose not to attack ($a_n(t) = 0$). The FC does not know which sensors belong to $\mathcal{A}$ and which sensors belong to $\mathcal{N}$. We assume that the adversary may only attack a strict subset of the sensors, i.e., there exists $0 \leq S < N$ such that $|\mathcal{A}| < S$. For the measurement model (1), Theorem 3.2 from [14] states that, in the absence of measurement noise (i.e., $w_n(t) = 0$), it is impossible to reliably estimate the value of θ^* if at least half of the sensors are under attack. In the sequel, we will compare the performance of our algorithm, which accounts for measurement noise, against this theoretical bound. Intuitively, it is difficult to simultaneously deal with measurement noise and attacks, since a smart adversary may be able to hide the attack in the noise.

3. RESILENT RECURSIVE ESTIMATION

In this section, we describe an iterative algorithm for the FC to achieve its goal of estimating θ^* when some of the measurement streams may be under attack.

3.1. Estimation Algorithm

The FC maintains a running estimate x_t and follows a three step procedure for updating the estimate.

1. Measurement Averaging: For each sensor n, the FC computes a time-averaged measurement

$$\overline{y}_n(t) = \frac{1}{t+1} y_n(t) + \frac{t}{t+1} \overline{y}_n(t). \tag{3}$$

   Note that, if n is not under attack (i.e., $n \in \mathcal{N}$), then $\overline{y}_n(t) = \theta^* + \overline{w}_n(t)$, where $\overline{w}_n(t) = \frac{1}{t+1} \sum_{j=0}^{t} w_n(j)$.

2. Gain Calculation: For each sensor n, the FC computes a scalar gain

$$K_n(t) = \min\left(1, \frac{\gamma_t}{\|\overline{y}_n(t) - x_t\|_2}\right), \tag{4}$$

   where γ_t is a threshold sequence to be defined shortly.

3. Estimate Update: The FC updates the estimate x_t as

$$x_{t+1} = x_t + \alpha_t \sum_{n=1}^{N} K_n(t) \left(\overline{y}_n(t) - x_t\right), \tag{5}$$

   where α_t is a weight sequence to be defined shortly.

Following (5), the FC updates its estimate as a weighted sum of its current estimate and the innovations component of each sensor, i.e., the difference between the (time-averaged) measurement of each sensor and the current estimate. The estimate and measurement dependent gain $K_n(t)$ limits the impact of measurements that deviate too much from the current estimate. It ensures that the ℓ_2-norm of the weighted innovations term $K_n(t)\left(\overline{y}_n(t) - x_t\right)$ does not exceed the value of the threshold γ_t.

The gain $K_n(t)$ provides resilience to the estimation algorithm by limiting the contribution of measurements that may have been compromised. Depending on the value of the theshold γ_t, this procedure may, however, also limit the contributition of sensors that are *not* compromised. For example, if we have $\gamma_t = 0$, then $K_n(t) = 0$ for all sensors n, and the FC will be unable to estimate θ^*, since it ignores all measurements. The challenge in designing this resilient

recursive estimator is to choose $K_n(t)$ and γ_t to limit the impact of compromised sensors without overly affecting the contribution of uncompromised sensors.

We adopt the following procedure for selecting the weight sequence α_t and threshold sequence γ_t:

1. Select the sequence α_t as $\alpha_t = \frac{a}{(t+1)^{\tau_1}}$, where $0 < a \leq \frac{1}{N}$ and $0 < \tau_1 \leq 1$.
2. Select the sequence γ_t as $\gamma_t = \frac{\Gamma}{(t+1)^{\tau_\gamma}}$ where $\Gamma > 0$ and $0 < \tau_\gamma < \frac{1}{2}$.

That is, we choose the sequences α_t and γ_t to decay over time. Intuitively, the decaying γ_t means that we allow less deviation in the measurements (from the current estimate) over time. We illustrate the effect of choosing different values of Γ (in γ_t) in Section 5.

3.2. Estimation Performance

The main result of the paper characterizes the performance of the recursive estimator under measurement attacks.

Theorem 1. *If $\frac{|\mathcal{A}|}{N} < \frac{1}{2}$, then, under update rule* (5)*, the estimate x_t satisfies*

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\tau_0}\left\|x_t-\theta^*\right\|_2=0\right)=1, \tag{6}$$

for every $0 \leq \tau_0 < \min\left(\tau_\gamma, \frac{1}{2}-\tau_\gamma\right)$.

Theorem 1 states that the algorithm can tolerate attacks on up to half of the sensors and still ensure that x_t converges to θ^* almost surely, regardless of *how* they are attacked (i.e., regardless of the specific values of $a_n(t)$). The estimator achieves the theoretical bound on resilience (in terms of number of tolerable attacked sensors) without identifying which sensors are compromised and avoids the associated combinatorial computational expense [13, 14].

4. PERFORMANCE ANALYSIS

We provide a proof sketch of Theorem 1 and omit details due to space constraints. the proof sketch depends on several intermediate results, the proofs of which are omitted due to space constraints.

The following lemma from [10] characterizes the behavior of time-averaged measurement noise.

Lemma 1. *Let $v_1, v_2, v_3, \ldots$ be i.i.d.random variables with mean $\mathbb{E}[v_t] = 0$ and covariance $\mathbb{E}[v_t v_t^{\mathrm{T}}] = \Sigma$. The time averaged mean $\overline{v}_t = \frac{1}{t+1}\sum_{j=0}^{t} v_j$ satisifies*

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)_0^{\delta}\left\|\overline{v}_t\right\|_2=0\right)=1, \tag{7}$$

for every $0 \leq \delta_0 < \frac{1}{2}$.

To analyze the estimator's performance, we need to characterize the behavior of the following, scalar, time-varying dynamical systems:

$$w_{t+1} = (1 - r_1(t))\, w_t + r_2(t), \tag{8}$$

$$\begin{aligned} \widehat{m}_{t+1} &= \left(1 - \frac{r_1(t)}{m_t + c_3}\right) m_t + r_2(t), \\ m_{t+1} &= \max\left(\left|\widehat{m}_{t+1}\right|, |m_t|\right), \end{aligned} \tag{9}$$

with initial conditions $w_0, m_0 \geq 0$, where $r_1(t) = \frac{c_1}{(t+1)^{\delta_1}}$, $r_2(t) = \frac{c_2}{(t+1)^{\delta_2}}$, $c_1, c_2, c_3 > 0$, $0 < \delta_1 \leq 1$, and $\delta_1 < \delta_2$.

Lemma 2. *The system in* (8) *satisfies $\lim_{t\to\infty}(t+1)^{\delta_0} w_t = 0$, for every $0 \leq \delta_0 < \delta_2 - \delta_1$.*

Lemma 3. *The system in* (9) *satisfies $\sup_{t\geq 0} m_t < \infty$.*

The performance of the estimator depends on the gains $K_n(t)$ for sensors n that are uncompromised. For $n \in \mathcal{N}$, $K_n(t)$ depends on the relationship between the estimation error $e_t = x_t - \theta^*$ and the threshold γ_t.

Lemma 4. *If $\frac{|\mathcal{A}|}{N} < \frac{1}{2}$, then, for any $0 < \epsilon_W < \frac{1}{2}$, almost surely, there exists $T_0 \geq 0$ and $0 < W < \infty$, such that:*

1. *for all $n \in \{1, 2, \ldots, N\}$, $\left\|\overline{w}_n(t)\right\|_2 \leq \frac{W}{(t+1)^{\frac{1}{2}-\epsilon_W}}$, and*
2. *if for some $T_1 \geq T_0$, we have $\left\|e_{T_1}\right\|_2 \leq \overline{\gamma}_{T_1}$, then, $\left\|e_t\right\|_2 \leq \overline{\gamma}_t$ for all $t \geq T_1$, where*

$$\overline{\gamma}_t = \frac{\Gamma - W(t+1)^{\tau_\gamma - \frac{1}{2} + \epsilon_W}}{(t+1)^{\tau_\gamma}}. \tag{10}$$

Lemma 4 states that, as long as less than half of the sensors are under attack, for large enough T_1, if the ℓ_2 norm of the estimation error e_{T_1} is upper bounded by $\overline{\gamma}_{T_1} < \gamma_{T_1}$, then, almost surely, the upper bound will hold for *all* times $t \geq T_1$.

(Proof Sketch of Theorem 1). We need to show that x_t converges to θ^* almost surely. By Lemma 4, almost surely, there exists finite T_0 such that, if at any time $T_1 \geq T_0$, $\left\|x_{T_1} - \theta^*\right\|_2 \leq \overline{\gamma}_{T_1}$, then, for all $t \geq T_1$, $\|x_t - \theta^*\| \leq \gamma_t$. We examine the evolution of $e_t = x_t - \theta^*$ along sample paths $\omega \in \Omega$ for which Lemma 4 holds (such a set of sample paths has measure 1). For a sample path ω, if there exists $T_1 \geq T_0$ such that $\left\|e_{T_1,\omega}\right\|_2 \leq \overline{\gamma}_{T_1,\omega} \leq \gamma_{T_1}$, then, we have

$$\lim_{t\to\infty}(t+1)^{\tau_0}\left\|e_{t,\omega}\right\|_2 \leq \lim_{t\to\infty}(t+1)^{\tau_0}\gamma_t = 0, \tag{11}$$

for every $0 \leq \tau_0 < \tau_\gamma$.

If no such T_1 exists, it means that, for all $t \geq T_0$, $\left\|e_{t,\omega}\right\|_2 > \overline{\gamma}_{t,\omega}$. Then, defining

$$K_{t,\omega} = \frac{\overline{\gamma}_t + W_\omega (t+1)^{-\frac{1}{2}+\epsilon_W}}{\left\|e_t\right\|_2 + W_\omega (t+1)^{-\frac{1}{2}+\epsilon_W}}, \tag{12}$$

where $(t+1)^{\frac{1}{2}-\epsilon_W}\left\|\overline{w}_n(t,\omega)\right\| \leq W_\omega$ for all $n \in \mathcal{N}$, we have $K_n(t,\omega) > K_{t,\omega}$ for all $n \in \mathcal{N}$. It can be shown that

$$\left\|e_{t+1,\omega}\right\|_2 \leq (1 - \alpha_t \kappa K_{t,\omega})\left\|e_{t,\omega}\right\|_2 + \frac{\alpha_t W_\omega}{(t+1)^{\frac{1}{2}-\epsilon_W}}, \tag{13}$$

where $\kappa = 1 - \frac{2|\mathcal{A}|}{N}$. Note that $K_{t,\omega} \geq \gamma_t \left(\sup_{j\in[0,t]}\left\|e_j\right\|_2 + W_\omega\right)^{-1}$. Then, from (13), we can show that $\sup_{j\in[0,t]}\left\|e_j\right\|_2 \leq m_t$, where m_t follows

$$\begin{aligned} \overline{m}_{t+1} &= \left(1 - \frac{\alpha_t \kappa \gamma_t}{m_t + W_\omega}\right) m_t + \frac{\alpha_t W_\omega}{(t+1)^{\frac{1}{2}-\epsilon_W}}, \\ m_{t+1} &= \max\left(\overline{m}_{t+1}, m_t\right). \end{aligned} \tag{14}$$

The system in (14) falls under the purview of Lemma 3, which means that $\sup_{t\geq 0}\left\|e_t\right\|_2$ is bounded above and there exists $K_\omega > 0$ such that $K_{t,\omega} \geq K_\omega (t+1)^{-\tau_\gamma}$. Replacing $K_{t,\omega}$ with $K_\omega (t+1)^{-\tau_\gamma}$, (13) then falls under the purview of Lemma 2, which means that

$$\lim_{t\to\infty}(t+1)^{\tau_0}\left\|e_{t,\omega}\right\|_2 = 0, \tag{15}$$

for every $0 \leq \tau_0 < \tau_\gamma - \frac{1}{2}$. The set of sample paths $\omega \in \Omega$ for which either (11) or (15) holds has measure 1, which yields (6). □

5. NUMERICAL EXAMPLES

In our numerical examples, we consider a team of $N = 50$ devices measuring ozone concentration in the same physical neighborhood. These devices may be static sensor nodes (e.g., [4, 5]), mobile sensors such as drones [3], or smart phones and wearable gadgets in crowdsensing applications [6]. These devices have different accuracy and measurement noise: typical sensors for ozone measurement range in accuracy ± 5 parts per billion (ppb) to ± 0.25 ppb, depending on cost [6]. To accomodate varying levels of accuracy, we present simulations for a range of measurement signal-to-noise ratios (SNR), specified below. In air quality monitoring, we are interested in determining if pollutant concentration levels adhere to health standards. For example, the United States Environmental Protection Agency (EPA) establishes that a safe concentration of ozone is 70 ppb [17]. In our numerical examples, we let the true value of the parameter $\theta^* = 37$ ppb.[1] In the first example, we let 15 sensors be compromised. The compromised sensors ($n \in \mathcal{A}$) report measurement streams with mean 100 ppb. The measurements of all devices are corrupted by additive white Gaussian noise (AWGN) with local SNR values of 17 dB, 11 dB, 5 dB, and 2 dB.

We run the resilient recursive estimator for 250 iterations with weights $a = 0.02, \tau_1 = 0.005, \Gamma = 10, \tau_\gamma = 0.40$. We compare the performance of the resilient estimator against an unsecure standard recursive estimator, which does not use the adaptive gain $K_n(t)$ when updating its estimate. We repeat the simulation 200 times and compute the average root mean square error (RMSE) in ppb at each iteration.

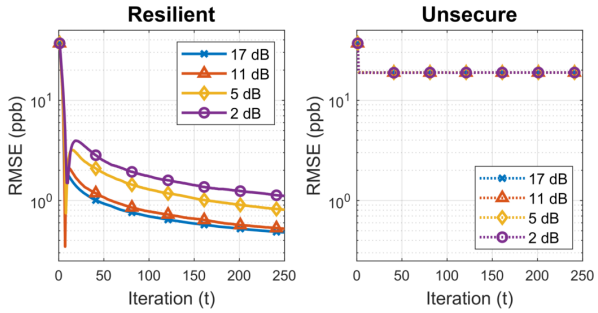

Fig. 1: RMSE (in ppb) versus iteration averaged over 200 trials of the resilient estimator (left) and unsecure estimator (right) when $|\mathcal{A}| = 15$ of the $N = 50$ sensors are compromised.

Figure 1 shows that the resilient estimator successfully copes with the compromised devices. Even when 15 out of the 50 devices report falsified measurement data, the RMSE of the resilient estimator decreases and the estimate moves closer to the true value of the parameter θ^* as the number of iterations increases. On the other hand, the unsecure estimator is unable to cope with the attack. A non zero RMSE persists in the unsecure estimate, and it does not decrease with increasing number of iterations.

Next, we study the performance of the resilient estimator as we vary the number of compromised devices (with fixed threshold weight $\Gamma = 10$) and as we vary the the threshold weight Γ (with a fixed number of compromised devices $|\mathcal{A}| = 15$). Figure 2 shows that, as more sensors become compromised, the RMSE of the resilient estimator after 250 iterations increases. This demonstrates the

[1] Data from the EPA shows that, in Pittsburgh, Pennsylvania, from January 2018 to March 2018, the daily maximum ozone concentration ranged from 1 ppb to 87 ppb, with a mean of 37 ppb [18].

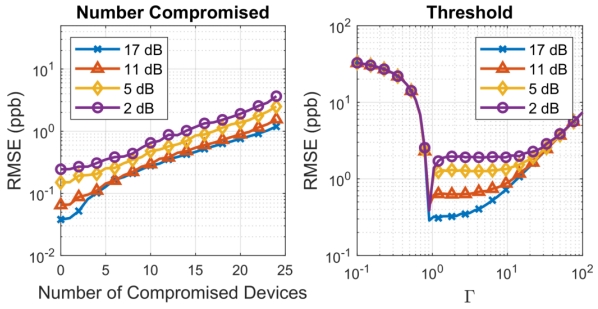

Fig. 2: RMSE (in ppb) after 250 iterations of the resilient estimator versus number of compromised sensors (left) and the threshold weight Γ (right), averaged over 200 trials.

tradeoff that exists between resilience and estimation performance. Keeping the number of iterations fixed, the estimator's performance becomes worse (in terms of RMSE) as the number of compromised sensors increases. As a consequence, to achieve a given level of performance (in terms of RMSE), the resilient estimator needs more iterations with an increasing number of compromised sensors.

Figure 2 also shows the effect of choosing different values of Γ (in the threshold γ_t) on the estimation performance after a fixed number of iterations. If Γ is too small (e.g., in Figure 2, $\Gamma < 1$), then the thresholding limits the impact of noncompromised measurements, resulting in slower convergence. For large enough Γ (e.g., in Figure 2, $1 \leq \Gamma \leq 10$), the resilient estimator limits the contribution of compromised measurements without overly affecting the noncompromised measurements, resulting in faster convergence and better performance in terms of RMSE. As Γ becomes too large (e.g., in Figure 2, $\Gamma > 10$), the compromised measurements have greater impact on the estimator, resulting in slower convergence. For *all* (positive) choices of the threshold weight Γ, we may always improve the estimator performance (decrease RMSE) by increasin the number of iterations.

6. CONCLUSION

In this paper, we have studied secure analytics for IoT type setups in the context of resilient recursive estimation. We considered a team of sensors or devices that each measure an unknown static parameter over time. The parameter may represent, for example, traffic conditions at a certain intersection or pollutant concentrations in a particular neighborhood. The sensors transmit their noisy measurement streams to a fusion center in the cloud, which processes the data online to estimate the parameter. Due to security vulnerabilities of IoT devices, a subset of the measurement streams fall under attack and take arbitrary value. We presented a recursive estimation algorithm that is resilient to such attacks: as long as less than half of the sensors are compromised, our algorithm constructs a sequence of estimates that converges *almost surely* to the value of the parameter. Finally, we demonstrated the performance of our resilient estimator through numerical examples based on ozone pollution monitoring. In future work, we plan to address resilient recursive estimation in fully distributed setups, where individual devices must collaborate over a peer-to-peer communication network and estimate the unknown parameter without a central coordinator.

7. REFERENCES

[1] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark., "Connected vehicles: Solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[2] J. Venkatesh, B. Aksanli, C. S. Chan, A. S Akyurek, and T. S. Rosing, "Modular and personalized smart health application design in a smart city environment," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1–10, June 2017.

[3] Y. Yang, Z. Zheng., K. Bian, L. Song, and Z. Han, "Real-time profiling of fine-grained air quality index distribution using UAV sensing," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 186–198, Feb. 2018.

[4] "What is the Array of Things?," URL: http://www.arrayofthings.github.io/faq.html, Accessed Oct. 19, 2018.

[5] P. deSouza, V. Nthusi, J. M. Klopp, B. E. Shaw, W. O Ho, J. Safell, R. Jones, and C. Ratti, "A Nairobi experiment in using low cost air quality monitors," *Clean Air Journal*, vol. 27, no. 2, pp. 1–31, 2017.

[6] W. Y. Yi, K. M. Lo, T. Mak, K. S. Leung, Y. Leung, and M. L. Meng, "A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems," *Sensors*, vol. 15, no. 12, pp. 31392–31427, 2015.

[7] Y. Chen, S. Kar, and J. M. F. Moura, "The Internet of Things: Secure Distributed Inference," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 64–75, Sept. 2018.

[8] J. Zhang, R. S. Blum, and H. V. Poor, "Approaches to Secure Inference in the Internet of Things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 50–63, Sept. 2018.

[9] S. B. Azmy, N. Zorba, and H. S. Hassanein, "Robust quality metric for scarce mobile crowd-sensing scenarios," in *Proc. of the 2018 IEEE International Conf. on Communications Workshops*, Kansas City, MO, May 2018, pp. 1–5.

[10] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation through adversary detection," *IEEE Trans. Signal Process.*, vol. PP, no. 99, pp. 1–15, Mar. 2018.

[11] J. Zhang, R. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1086–1101, Mar. 2015.

[12] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "Distributed Bayesian detection in the presence of Byzantine data," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5250–5263, Oct. 2015.

[13] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[14] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, June 2014.

[15] Y. Chen, S. Kar, and J. M. F. Moura, "Attack Resilient Distributed Estimation: A Consensus+Innovations Approach," in *Proc. 2018 American Control Conf. (ACC)*, Milwaukee, WI, June 2018, pp. 1015–1020.

[16] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation: Sensor attacks," *ArXiv e-prints*, pp. 1–8, Mar. 2018.

[17] United States Environmental Protection Agency, "NAAQS Table," URL: https://www.epa.gov/criteria-air-pollutants/naaqs-table, Accessed Oct. 24, 2018.

[18] United States Environmental Protection Agency, "Outdoor Air Quality Data," URL: https://www.epa.gov/outdoor-air-quality-data/download-daily-data, Accessed Oct. 24, 2018.