This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2015.2505066, IEEE Transactions on Automatic Control

Multi-Sensor Scheduling for State Estimation with Event-Based, Stochastic Triggers

Sean Weerakkody, *Student Member, IEEE,*
Yilin Mo, *Member, IEEE,* Bruno Sinopoli, *Member, IEEE,*
Duo Han, *Student Member, IEEE,* and Ling Shi *Member, IEEE*

***Abstract*—In networked systems, state estimation is hampered by communication limits. Past approaches, which consider scheduling sensors through deterministic event-triggers, reduce communication and maintain estimation quality. However, these approaches destroy the Gaussian property of the state, making it computationally intractable to obtain an exact minimum mean squared error estimate. We propose a stochastic event-triggered sensor schedule for state estimation which preserves the Gaussianity of the system, extending previous results from the single-sensor to the multi-sensor case.**

I. Introduction

Networked Control Systems (NCSs), spatially distributed systems where sensors, actuators, and controllers exchange information over a shared, bandlimited communication network, have become a topic of significant interest. As noted by [1], the use of NCSs in practice provides for flexible architecture and reduces costs in installation and maintenance. Thus, NCSs have been used in several applications including public transportation, health care, and mobile sensor networks. Nonetheless, remote state estimation remains a significant challenge in NCSs [2]. Traditionally, state estimates are computed at an estimation center using information from sensors which sample and send measurements periodically. While it is reasonable to assume that remote state estimation centers are well equipped, in most cases, sensors have a limited power supply and are difficult to replace. Moreover, bandwidth constraints in a communication network may restrict the number of sensors which can communicate at any given time [3], [4], [5]. One way to address these issues is to simply reduce the communication rate. This solution however degrades estimation quality. In this paper, we propose a sensor scheduling scheme which allows us to achieve a desired tradeoff between communication rate and estimation performance. Specifically, we design a stochastic multi-sensor event-based scheduler which extends the single sensor results from [6].

Before continuing, we briefly document recent attempts to address the problem of remote estimation via sensor scheduling. We first examine offline schemes where sensors are scheduled based on system parameters prior to use. Yang et. al. [7] determined that given fixed communication constraints, an optimal deterministic offline schedule should allocate sensor transmission times as uniformly as possible over a finite time horizon. Moreover, Shi et. al. [8] specifically considered the 2-sensor problem with bandwidth constraints and found that a periodic sensor schedule minimized average error covariance. In addition to offline designs, previous work has considered event-based designs, where sensor transmissions are scheduled in real time based on an occurrence related to a sensor measurement or current system parameters. Astrom and Bernhardsson [9] show that for certain systems, event based sampling offers better performance than periodic sampling. Additionally, Imer et. al. [10] consider a single sensor sequential estimation problem where communication is limited over a finite horizon and propose a stochastic solution. Furthermore, Xu et. al. [11] consider scheduling a single, smart sensor which computes and sends a local estimate of the state. The authors propose a stochastic event trigger, where the rate of transmission is a quadratic function of the state estimation error.

While not utilized in [9], [10], and [11], event-based approaches can allow the estimator to extract information about the state from the absence of a measurement, and thus improve its estimate. For instance, Ribeiro et. al. [12] require the transmission of the sign of the innovation and derive an approximate minimum mean squared error (MMSE) estimator. Also, the authors in [13] design a threshold scheme on the normalized innovation vector to trigger communication and derive an approximate MMSE estimate. Deterministic schemes as discussed by [12, 13, 14] destroy the Gaussian property of the innovation process in traditional Kalman filtering, thus rendering the closed-form derivation of the exact MMSE estimator computationally intractable. Symmetric triggers such as those proposed in [15] and [6] allow the remote estimator to compute an MMSE estimate. Here, the triggers are designed so that a priori and a posteriori estimates are identical if a measurement is dropped which implicitly requires that the sensor has access to the same information as the estimator. However, this is not feasible in the multi-sensor case without substantially increasing communication in the network.

Han et. al. in [6] incorporate a stochastic decision rule, which not only allows the remote estimator to use information contained in the absense of a measurement, but also maintains the Gaussian distribution of the current state. A key advantage of the proposed method over most deterministic triggers is that in addition to obtaining an exact MMSE estimator, by preserving Gaussianity, [6] maintains an exact distribution of the state x_k and the estimation error e_k for all time k. Thus, the proposed stochastic event-based trigger is useful in scenarios where real time error analysis is critical. In this paper, we extend the same stochastic decision rule to the multi-sensor case where there exists a unique decision variable for each of m sensors. The main contribution of this paper relative to [6], which considers a binary transmit or drop policy for a single trigger, is the derivation of a two-step estimation filter to account for multiple independent triggers, a modified optimization problem to design each trigger, and a realistic simulation example on data center energy management. For this scenario, we also obtain expressions for sensor communication rates and upper and lower bounds on the error covariance. A preliminary study for this paper was previously presented [16]. Here a three-step recursive filter is proposed which computes a state distribution conditioned on all previous information, newly received measurements, and the identity of sensors which do not transmit sequentially. In this article, we obtain an equivalent two-step recursive filter which combines the last two stages, allowing us to directly obtain an a posteriori state distribution without any intermediary steps. We also extend [16] by accounting for vector sensor measurements with correlated sensor noise as well as through our optimization problem and simulation example.

The remainder of the paper is organized as follows. Section II formulates the multi-sensor state estimation problem and proposes a stochastic event-based sensor scheduler. Section III introduces a recursive filtering algorithm to obtain the MMSE estimator of

The work by S. Weerakkody, Y. Mo, and B. Sinopoli is supported by NSF grant 0955111 CAREER: Efficient, Secure and Robust Control of Cyber Physical Systems and NSF grant 1135895 CPS: Medium: Collaborative Research: The Cyber Physical Challenges of Transient Stability and Security in Power Grids.

The work by D. Han, and L. Shi is supported by a HK RGC GRF grant 618612.

S. Weerakkody and B. Sinopoli are with the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, 15213 USA e-mail: sweerakk@andrew.cmu.edu, brunos@ece.cmu.edu

Y. Mo was with the ECE department of Carnegie Mellon University, Pittsburgh, PA, when this article was written. He is now with the School of Electrical and Electronic Engineering, Nanyang Technological University. email: ylmo@ntu.edu.sg

D. Han and L. Shi are with the ECE department of Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. e-mail: fdhanaa, eeslingg@ust.hk.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2015.2505066, IEEE Transactions on Automatic Control

the state and its error covariance. Section IV derives results about communication rate and estimation performance. Section V proposes a semi-definite program to intelligently select trigger parameters. Section VI consists of a simulation. A conclusion summarizes future work.

Notation: X' denotes the transpose of matrix X. $\mathbb{S}_+^n$ and $\mathbb{S}_{++}^n$ are the sets of $n \times n$ positive semi-definite and positive definite matrices. When $X \in \mathbb{S}_+^n$, we simply write $X \geq 0$ (or $X > 0$ if $X \in \mathbb{S}_{++}^n$). $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian distribution with mean μ and covariance matrix Σ. $\mathbb{E}[\cdot]$ denotes the expectation, $\Pr(\cdot)$ denotes the probability of a random event, $\rho(\cdot)$ denotes the spectral radius of a matrix. $\text{diag}(X_1, \cdots, X_s)$ is the block diagonal matrix with square submatrices $X_1, \cdots, X_s$. $\mathbf{1}$ and $\mathbf{0}$ denote vectors with entries 1 and 0 respectively and I_n is the identity matrix of size $n \times n$. $\mathbb{1}_{\mathcal{X}}$ is the indicator function. Finally, $\{A\}_0$ is the matrix obtained by deleting all $\mathbf{0}$ rows from A.

II. Problem Setup

We define the following linear system:

$$x_{k+1} = Ax_k + w_k, \qquad y_k^{(i)} = C^{(i)} x_k + v_k^{(i)}, \quad i = 1, \cdots, m. \quad (1)$$

Here $x_k \in \mathbb{R}^n$ is the state vector, while $y_k^{(i)} \in \mathbb{R}^{s_i}$ is the ith of m vector sensor measurements. In addition, $w_k \in \mathbb{R}^n$ and $v_k \triangleq [v_k^{(1)\prime}, \cdots, v_k^{(m)\prime}]' \in \mathbb{R}^s$ are mutually uncorrelated Gaussian noises with covariances $Q > 0$ and $R > 0$, respectively and $s = \sum_{i=1}^m s_i$. To simplify notation, we define $y_k \triangleq [y_k^{(1)\prime}, \cdots, y_k^{(m)\prime}]'$. The initial state x_0 is zero-mean Gaussian random variable with covariance matrix $\Sigma_0 > 0$, and is uncorrelated with w_k and $v_k^{(i)}$ for all $k \geq 0$. We assume that (A, C) is detectable where we define $C \triangleq [C^{(1)\prime}, \cdots, C^{(m)\prime}]'$.

To reduce the rate of sensor to estimator communication, we intelligently transmit a fraction of our sensor measurements. Note that we choose to transfer sensor measurements as opposed to local estimates. This reduces sensor computation as well as size of packets for $n > s_i$. We specify $\gamma_k^{(i)} \in \{0, 1\}$ as the binary decision variable for sensor i at time k. When $\gamma_k^{(i)} = 1$, a transmission occurs while when $\gamma_k^{(i)} = 0$, no measurement is sent. Collecting our decision variables over m sensors, we have $\gamma_k = [\gamma_k^{(1)}, \cdots, \gamma_k^{(m)}]'$. Also, suppose at each time k, l_k sensors drop their measurements and $m - l_k$ sensors transmit their measurements. The sensors which transmit have indices $p_1, \cdots, p_{m-l_k}$. Define the vector of received measurements $y_k^r \in \mathbb{R}^{m-l_k}$ at time k by $y_k^r = [y_k^{(p_1)\prime}, \cdots, y_k^{(p_{m-l_k})\prime}]'$. To obtain a MMSE estimator given all previous and current measurements, we perform a two-step process. The first step is a time update where we obtain the MMSE estimator of x_k given the information set up to time $k-1$. This is denoted by $\mathcal{I}_{k-1} \triangleq \{\gamma_0, \cdots, \gamma_{k-1}, y_0^r, \cdots, y_{k-1}^r\}$ where $\mathcal{I}_{-1} \triangleq \emptyset$. In the second step, we update our estimate of x_k, using our previous information set, the received measurements at time k, (y_k^r), and the knowledge that certain sensors did not send a measurement (γ_k). Thus, we update using $\mathcal{I}_k$.

We now define the following estimation parameters:

$$\begin{aligned} \hat{x}_k^- &\triangleq \mathbb{E}[x_k | \mathcal{I}_{k-1}], & P_k^- &\triangleq \mathbb{E}[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)' | \mathcal{I}_{k-1}], \\ \hat{x}_k &\triangleq \mathbb{E}[x_k | \mathcal{I}_k], & P_k &\triangleq \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)' | \mathcal{I}_k]. \end{aligned} \quad (2)$$

Here $\hat{x}_k^-$ is an *a priori* MMSE estimate and $\hat{x}_k$ is an *a posteriori* MMSE estimate. When all measurements are sent to the estimator, computation of $\hat{x}_k$ and P_k, the error covariance, reduces to the standard Kalman filter. As done by [6], to maintain the Gaussian distribution of x_k, we consider a stochastic trigger. A stochastic trigger takes a measurement $y_k^{(i)}$ and computes a function $\varphi^{(i)} : \mathbb{R}^{s_i} \to [0, 1]$ to determine the probability sensor i does not transmit. While deterministic triggers assign probabilities equal to 1 or 0 for each measurement, the chosen trigger assigns probabilities in $[0, 1]$. To do this, at time k, each sensor i generates an i.i.d. uniform random variable $\zeta_k^{(i)}$ over [0, 1] and computes $\gamma_k^{(i)}$.

$$\gamma_k^{(i)} = \begin{cases} 0 & \zeta_k^{(i)} \leq \varphi^{(i)}(y_k^{(i)}) \\ 1 & \zeta_k^{(i)} > \varphi^{(i)}(y_k^{(i)}) \end{cases}, \qquad \varphi^{(i)}(\alpha) \triangleq \exp\left(-\frac{1}{2}\alpha' Y^{(i)} \alpha\right). \quad (3)$$

Here $Y^{(i)} \in \mathbb{S}_{++}^{s_i}$ are trigger parameters and we define $Y \in \mathbb{S}_{++}^s$ as $Y \triangleq \text{diag}(Y^{(i)}, \cdots, Y^{(m)})$. Note that $P(\gamma_k^{(i)} = 0 | y_k^{(i)})$ has the shape of a scaled Gaussian distribution. In the next section, we will show this allows the state to remain Gaussian. For the chosen trigger we consider stable systems, i.e. $\rho(A) < 1$. If the system is unstable, any sensor i which measures an unstable state will have $y_k^{(i)}$ grow unbounded. In this case, by (3) sensor i will always transmit.

In order to maximize performance with respect to communication and mean squared error (MSE) in state estimation, the choice of the function $\varphi^{(i)}$ should be considered in conjunction with a choice of estimator. We observe that there may exist other, possibly deterministic, triggers and estimators which achieve better MSE and communication performance than the proposed trigger. Previous work [17] suggests in the scalar case, without sensor noise, a symmetric threshold based detector of the error $y_k - \hat{y}_k^-$, along with a Kalman filter is optimal among deterministic triggers.

Motivated by this result, [6] considers a closed loop trigger where $\alpha = y_k^{(i)} - \mathbb{E}[y_k^{(i)} | \mathcal{I}_{k-1}]$. This design outperforms the proposed trigger but requires estimator to sensor communication at each step, which substantially increases communication cost. As a result, we do not consider this approach. Instead, we consider $\alpha = y_k^{(i)} \approx y_k^{(i)} - \mathbb{E}[y_k^{(i)} | \mathcal{I}_{-1}]$ as $k \to \infty$. For multidimensional systems, a challenge in selecting $\varphi^{(i)}$ is the ability to obtain computationally tractable MMSE estimates. The advantage of the proposed design of $\varphi^{(i)}$ is that an exact MMSE estimator can be obtained. Moreover, from [6] as well the simulation section, the proposed trigger outperforms some known deterministic designs.

III. MMSE Estimator Design

In this section, based on the design of $\varphi^{(i)}$, we obtain a closed-form solution to the MMSE estimation problem, given recursively by the following theorem:

Theorem 1. *Consider remote state estimation with event-based scheduler* (3) *and define the matrix* $\Psi_k \in \mathbb{R}^{s \times s} \triangleq \text{diag}(\gamma_1 I_{s_1} \cdots \gamma_m I_{s_m})$ *to store the* m *decision variables. Also, for simplicity let* $Z_k \triangleq CP_k^- C' + R$. *Assume* $f(x_0 | \mathcal{I}_{-1}) \sim \mathcal{N}(0, \Sigma_0)$ *so* $\hat{x}_0^- = 0$, $P_0^- = \Sigma_0$. *Then,* $f(x_k | \mathcal{I}_k) \sim \mathcal{N}(\hat{x}_k, P_k)$ *and* $f(x_k | \mathcal{I}_{k-1}) \sim \mathcal{N}(\hat{x}_k^-, P_k^-)$ *where* $\hat{x}_k, \hat{x}_k^-$ *and* P_k, P_k^- *satisfy the following recursive equations:*
Time update:

$$\hat{x}_k^- = A\hat{x}_{k-1}, \qquad P_k^- = AP_{k-1}A' + Q, \quad (4)$$

Measurement update:

$$\hat{x}_k = \hat{x}_k^- + K_k(\Psi_k y_k - C\hat{x}_k^-), \quad (5)$$

$$P_k = P_k^- - K_k C P_k^-, \ K_k = P_k^- C'(Z_k + (I - \Psi_k)Y^{-1})^{-1}, \quad (6)$$

Proof. To simplify the proof of the theorem, we define the following notation which will allow us to distinguish among parameters associated with sent measurements versus dropped measurements. Suppose at time k, there exists l_k sensors $j_1, \cdots, j_{l_k}$ that do not trigger a transmission and $m - l_k$ sensors, $p_1, \cdots, p_{m-l_k}$ which trigger a transmission. We define the matrix $\Gamma_k \in \mathbb{R}^{\left(\sum_{1=1}^{m-l_k} s_{p_i}\right) \times s}$ and $\bar{\Gamma}_k \in$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2015.2505066, IEEE Transactions on Automatic Control

$\mathbb{R}^{m-l_k\times m}$ to select sensors which transmit and $\Lambda_k \in \mathbb{R}^{\left(\sum_{1=1}^{l_k} s_{j_i}\right)\times s}$ and $\bar{\Lambda}_k \in \mathbb{R}^{l_k\times m}$ to select sensors which do not transmit as

$$\Gamma_k = \{\Psi_k\}_0, \qquad \left(\bar{\Gamma}_k\right)_{u,v} \triangleq \begin{cases}1 & v = p_u\\ 0 & otherwise\end{cases},$$
$$\Lambda_k = \{I-\Psi_k\}_0, \qquad \left(\bar{\Lambda}_k\right)_{u,v} \triangleq \begin{cases}1 & v = j_u\\ 0 & otherwise\end{cases}. \tag{7}$$

We prove Theorem 1 using induction on the distribution $f(x_k|\mathcal{I}_{k-1}) \sim \mathcal{N}(\hat{x}_k^-, P_k^-)$.

Case $n=0$: For $n=0$, we have $\mathcal{I}_{k-1}=\emptyset$. Thus, $f(x_0|\mathcal{I}_{-1}) = f(x_0) \sim \mathcal{N}(0,\Sigma_0)$ and the initial conditions holds.

Case assume for $n=k$: We assume that $f(x_k|\mathcal{I}_{k-1}) \sim \mathcal{N}(\hat{x}_k^-, P_k^-)$.

Case prove for $n=k+1$: We first verify the measurement update step.

Measurement Update Step: Consider the joint conditional pdf of x_k and $\Lambda_k y_k$ given $\mathcal{I}_k$

$$\begin{aligned} f(x_k,\Lambda_k y_k|\mathcal{I}_k) &= f(x_k,\Lambda_k y_k|y_k^r,\bar{\Gamma}_k\gamma_k=\mathbf{1},\bar{\Lambda}_k\gamma_k=\mathbf{0},\mathcal{I}_{k-1})\\ &= f(x_k,\Lambda_k y_k|\Gamma_k y_k,\bar{\Lambda}_k\gamma_k=\mathbf{0},\mathcal{I}_{k-1}),\\ &= \frac{\Pr(\bar{\Lambda}_k\gamma_k=\mathbf{0}|x_k,y_k,\mathcal{I}_{k-1})f(x_k,\Lambda_k y_k|\Gamma_k y_k,\mathcal{I}_{k-1})}{\Pr(\bar{\Lambda}_k\gamma_k=\mathbf{0}|\Gamma_k y_k,\mathcal{I}_{k-1})}. \end{aligned} \tag{8}$$

The second equality follows since the knowledge of the values of sent measurements $\Gamma_k y_k$ implies that the decision variables $\bar{\Gamma}_k\gamma_k=\mathbf{1}$. The last equality is derived from Bayes rule.

By our induction assumption, $f(x_k,\Lambda_k y_k,\Gamma_k y_k)$ is jointly Gaussian distributed given $\mathcal{I}_{k-1}$. As a result, the distribution $f(x_k,\ \Lambda_k y_k|\Gamma_k y_k,\ \mathcal{I}_{k-1})$ is also Gaussian. We observe $f(x_k,\Lambda_k y_k,\Gamma_k y_k|\mathcal{I}_{k-1})$ has mean $[\hat{x}_k^{-\prime},(\Lambda_k C\hat{x}_k^-)',(\Gamma_k C\hat{x}_k^-)']'$ and covariance

$$\begin{bmatrix} P_k^- & P_k^- C'\Lambda_k' & P_k^- C'\Gamma_k'\\ \Lambda_k C P_k^- & \Lambda_k Z_k \Lambda_k' & \Lambda_k Z_k \Gamma_k'\\ \Gamma_k C P_k^- & \Gamma_k Z_k \Lambda_k' & \Gamma_k Z_k \Gamma_k' \end{bmatrix}. \tag{9}$$

Given a joint Gaussian distribution $f(x_k,\Lambda_k y_k,\Gamma_k y_k|\mathcal{I}_{k-1})$, it is easy to compute $f(x_k,\Lambda_k y_k|\Gamma_k y_k,\mathcal{I}_{k-1})$ which is also Gaussian [18]. The conditional means are

$$\mu_x = \hat{x}_k^- + P_k^-(\Gamma_k C)'(\Gamma_k Z_k\Gamma_k')^{-1}\Gamma_k(y_k - C\hat{x}_k^-), \tag{10}$$
$$\mu_y = \Lambda_k C\hat{x}_k^- + \Lambda_k Z_k\Gamma_k'(\Gamma_k Z_k\Gamma_k')^{-1}\Gamma_k(y_k - C\hat{x}_k^-). \tag{11}$$

Furthermore, the covariance of x_k and $\Lambda_k y_k$ given $\Gamma_k y_k$ and $\mathcal{I}_{k-1}$ is $\Phi_k = \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{xy}' & \Sigma_{yy}\end{bmatrix}$, where

$$\Sigma_{xx} = P_k^- - P_k^-(\Gamma_k C)'(\Gamma_k Z_k\Gamma_k')^{-1}(\Gamma_k C)P_k^-, \tag{12}$$
$$\Sigma_{yy} = \Lambda_k Z_k\Lambda_k' - \Lambda_k Z_k\Gamma_k'(\Gamma_k Z_k\Gamma_k')^{-1}\Gamma_k Z_k\Lambda_k', \tag{13}$$
$$\Sigma_{xy} = P_k^-(\Lambda_k C)' - P_k^-(\Gamma_k C)'(\Gamma_k Z_k\Gamma_k')^{-1}\Gamma_k Z_k\Lambda_k'. \tag{14}$$

Having obtained $f(x_k,\Lambda_k y_k|\Gamma_k y_k,\mathcal{I}_{k-1})$, we observe

$$\Pr(\bar{\Lambda}_k\gamma_k=\mathbf{0}|x_k,y_k,\mathcal{I}_{k-1}) = \exp\left(-\frac{1}{2}y_k'\Lambda_k'\Lambda_k Y\Lambda_k'\Lambda_k y_k\right).$$

Using (8), we can thus obtain the joint probability density function for the state x_k and the dropped measurements $\Lambda_k y_k$. That is we have $f(x_k,\Lambda_k y_k|\mathcal{I}_k) = \beta_k^{-1}\exp(-\frac{1}{2}\theta_k)$, where $\beta_k\in\mathbb{R}$ and $\theta_k\in\mathbb{R}$ are defined respectively as

$$\beta_k \triangleq \Pr(\bar{\Lambda}_k\gamma_k=\mathbf{0}|\Gamma_k y_k,\mathcal{I}_{k-1})\sqrt{\det(\Phi_k)(2\pi)^{n+\sum_{i=1}^{l_k}s_{j_i}}}, \tag{15}$$

$$\theta_k \triangleq \begin{bmatrix} x_k-\mu_x\\ \Lambda_k y_k-\mu_y\end{bmatrix}'\begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}^{-1}\begin{bmatrix} x_k-\mu_x\\ \Lambda_k y_k-\mu_y\end{bmatrix} + (\Lambda_k y_k)'\Lambda_k Y\Lambda_k'(\Lambda_k y_k). \tag{16}$$

We now introduce the following Lemma with proof found in [19].

Lemma 1. *The scalar $\theta_k\in\mathbb{R}$ is given by*

$$\theta_k = \begin{bmatrix} x_k-\bar{x}_k\\ \Lambda_k y_k-\bar{y}_k\end{bmatrix}'\Theta_k^{-1}\begin{bmatrix} x_k-\bar{x}_k\\ \Lambda_k y_k-\bar{y}_k\end{bmatrix} + c_k, \tag{17}$$

where $\bar{x}_k\in\mathbb{R}^n$, $\bar{y}_k\in\mathbb{R}^{\sum_{i=1}^{l_k}s_{j_i}}$, $c_k\in\mathbb{R}$ and $\Theta_k\in\mathbb{S}_{++}^{\sum_{i=1}^{l_k}s_{j_i}+n}$ are given by

$$\bar{x}_k = \hat{x}_k^- + P_k^- C'(Z_k + (I-\Psi_k)Y^{-1})^{-1}(\Psi_k y_k - C\hat{x}_k^-), \tag{18}$$
$$\bar{y}_k = \left[I + \Sigma_{yy}\Lambda_k Y\Lambda_k'\right]^{-1}\mu_y, \tag{19}$$
$$\Theta_k = \begin{bmatrix}\Theta_{xx,k} & \Theta_{xy,k}\\ \Theta_{xy,k}' & \Theta yy,k\end{bmatrix}, c_k = \mu_y'(\Sigma_{yy} + \Lambda_k Y^{-1}\Lambda_k')^{-1}\mu_y, \tag{20}$$

where

$$\begin{aligned}\Theta_{xx,k} &= P_k^- - P_k^- C'(CP_k^- C' + R + (I-\Psi_k)Y^{-1})^{-1}CP_k^-,\\ \Theta_{xy,k} &= \Sigma_{xy}(I + \Lambda_k Y\Lambda_k'\Sigma_{yy})^{-1},\\ \Theta_{yy,k} &= \left[\Sigma_{yy}^{-1} + \Lambda_k Y\Lambda_k'\right]^{-1}.\end{aligned}$$

Thus, the joint pdf of our state and unknown measurements are given as follows

$$\begin{aligned} f(x_k,\Lambda_k y_k|\mathcal{I}_k) &= \frac{1}{\beta_k}\exp\left(-\frac{c_k}{2}\right)\\ &\times\exp\left(-\frac{1}{2}\begin{bmatrix} x_k-\bar{x}_k\\ \Lambda_k y_k-\bar{y}_k\end{bmatrix}'\Theta_k^{-1}\begin{bmatrix} x_k-\bar{x}_k\\ \Lambda_k y_k-\bar{y}_k\end{bmatrix}\right). \end{aligned} \tag{21}$$

Since $f(x_k,\Lambda_k y_k|\mathcal{I}_k)$ is a pdf, its integral normalizes to one which implies that $f(x_k,\Lambda_k y_k|\mathcal{I}_k)$ are jointly Gaussian. Moreover, this implies that x_k is conditionally Gaussian given $\mathcal{I}_k$ with mean $\hat{x}_k$ and covariance P_k. Therefore, (5) and (6) hold for the measurement update step.

Time Update Step: We have proved $f(x_k|\mathcal{I}_k)\sim\mathcal{N}(\hat{x}_k,P_k)$. By the conditional independence of x_k and w_k, we see

$$f(x_{k+1}|\mathcal{I}_k) = f(Ax_k + w_k|\mathcal{I}_k) \sim \mathcal{N}(A\hat{x}_k, AP_kA' + Q). \tag{22}$$

Thus, (4) holds. By induction, $f(x_k|\mathcal{I}_{k-1})\sim\mathcal{N}(\hat{x}_k^-,P_k^-)$. Moreover, from this result, and the proof of the measurement update step, $f(x_k|\mathcal{I}_k)\sim\mathcal{N}(\hat{x}_k,P_k)$, which concludes the proof. □

Remark 1. *The estimation filter can be formulated as a Kalman filter with time-varying sensor noise $R+(I-\Psi_k)Y^{-1}$ and innovation $\Psi_k y_k - C\hat{x}_k^-$. This similarity allows for computational simplicity and easy implementation.*

Remark 2. *With an imperfect channel, the estimator will have to differentiate between intended packet drops by the sensor due to the stochastic trigger and unintended drops due to the channel. If packet drops are IID Bernoulli, the state will be distributed according to a Gaussian mixture model corresponding to each possible trajectory of γ_k. The resulting distribution however is intractable as $k\to\infty$.*

Remark 3. *In the case that we wish to model uncertainties or partial knowledge of the matrices A and C, for instance through unknown parameters ΔA and ΔC, the results of theorem 1 do not hold. As with the standard Kalman filter, uncertainty in system parameters will destroy the Gaussianity of the system state and make computation of the MMSE estimator intractable. Nonetheless, we know simply from the stability of the system that any error in the estimate due to system uncertainty will be bounded. Moreover, the state estimate will be a continuous function of parameters in A and C.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2015.2505066, IEEE Transactions on Automatic Control

IV. Performance Analysis

In proposing an event-based trigger, our goal is to address the trade-off between estimation performance and power consumed through communication by sensor nodes. The communication rate $\lambda^{(i)} \in [0,1]$ for sensor i can be defined as

$$\lambda^{(i)} \triangleq \limsup_{T\to\infty} \frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}[\gamma_k^{(i)}]. \tag{23}$$

Knowledge of the communication rate $\lambda^{(i)}$ of each sensor will allow designers to determine the required system bandwidth and to estimate the lifetime of each sensor. To obtain an expression for the communication rate $\lambda^{(i)}$ for each sensor, we first define $\Sigma \in \mathbb{S}_{++}^n, \Pi^{(i)} \in \mathbb{S}_{++}^{s_i}$ by

$$\Sigma \triangleq \lim_{k\to\infty} \mathrm{Cov}(x_k) = A\Sigma A' + Q,$$
$$\Pi^{(i)} \triangleq \lim_{k\to\infty} \mathrm{Cov}(y_k^{(i)}) = C^{(i)}\Sigma C^{(i)\prime} + R^{(i)},$$

where $R^{(i)} \triangleq \mathbb{E}[v_k^{(i)} v_k^{(i)\prime}]$. With these results, we arrive at an expression for the communicate rate of each sensor with proof in [6] .

Theorem 2. *Consider a stable linear system* (1) *with a stochastic event-based sensor schedule given by* (3). *The communication rate $\lambda^{(i)}$ for each sensor $i = 1,\cdots,m$ is given by*

$$\lambda^{(i)} = 1 - \frac{1}{\sqrt{\det\left(I + \Pi^{(i)}Y^{(i)}\right)}}. \tag{24}$$

We next verify that the properties established for the expected communication rate over several runs, apply to a single sample path, the proof of which is found in [6].

Theorem 3. *The following equality almost surely holds.*

$$\lim_{N\to\infty} \frac{1}{T+1}\sum_{k=0}^{T}\gamma_k^{(i)} \overset{a.s}{=} \lambda^{(i)}. \tag{25}$$

Furthermore, for any finite integer $l \geq 0$, define the event of l sequential packed drops over all m sensors $\overline{E}_{k,l}$ and the event of l sequential packet arrivals over all m sensors $\underline{E}_{k,l}$ as follows

$$\overline{E}_{k,l} \triangleq \{\gamma_k = \mathbf{0},\cdots,\gamma_{k+l-1} = \mathbf{0}\},$$
$$\underline{E}_{k,l} \triangleq \{\gamma_k = \mathbf{1},\cdots,\gamma_{k+l-1} = \mathbf{1}\}.$$

Then almost surely $\underline{E}_{k,l}$ and $\overline{E}_{k,l}$ happen infinitely often.

We next examine the estimation performance by analyzing the statistical properties of P_k^-.

Theorem 4. *Consider a stable system* (1) *with scheduler given by* (3). *Let*

$$g_W(X) \triangleq AXA' + Q - AXC'(CXC' + W)^{-1}CXA'.$$

1) *There exists an $M \in S_{++}^n$, such that for all k, P_k^- is uniformly bounded above by M.*
2) *For any $\epsilon > 0$, there exists an N such that for all $k \geq N$, the following inequalities hold*

$$\underline{X} - \epsilon I \leq P_k^- \leq \overline{X} + \epsilon I, \tag{26}$$

where $\underline{X}$ and $\overline{X}$ are the unique solutions $X = g_R(X)$ and $X = g_{R+Y^{-1}}(X)$ respectively.

3) *For any $\epsilon > 0$, almost surely for infinitely many $k's$, we have $P_k^- \geq \overline{X} - \epsilon I$ and almost surely for infinitely many $k's$, we have $P_k^- \leq \underline{X} + \epsilon I$.*

The first statement shows that regardless of the choice of $Y^{(i)}$ (communication rate), the error covariance is bounded. The second statement obtains upper and lower bounds while the third statement shows that during a sample path, P_k^- will approach these bounds infinitely many times, a consequence of Theorem 3, where we expect long strings of transmissions and drops.

V. Optimization of Trigger Parameters

Before we continue, we introduce the following Corollary with proof found in the appendix.

Corollary 1. *Define $\overline{P} \triangleq \overline{X} - \overline{X}C'(C\overline{X}C' + R + Y^{-1})^{-1}C\overline{X}$.*

1) *For any $\epsilon > 0$, $\exists$ an N such that for all $k \geq N$, $P_k \leq \overline{P} + \epsilon I$.*
2) *For any $\epsilon > 0$, almost surely for infinitely many $k's$, we have $P_k \geq \overline{P} - \epsilon I$*

Thus, it is a worthy goal to design $Y^{(i)}$ to limit $\overline{P}$. We address the estimation and communication tradeoff by minimizing the system communication rate subject to this bound.

Problem 1: $\displaystyle Y_*^{(i)} = \underset{Y^{(i)}\geq 0,\ i=1,\cdots,m}{\arg\min} \sum_{i=1}^{m}\lambda^{(i)}, \quad \text{s.t} \quad \overline{P} \leq \Delta.$

[1] Here, the matrix Δ serves as an upper bound on our worse case error covariance, thus providing a robust bound on our estimation quality. Unfortunately, Problem 1 is a nonconvex minimization problem which cannot easily be solved. However, we observe the following.

Lemma 2. *Define $f(x) \triangleq 1-(1+x)^{-\frac{1}{2}}$ and $g(x) = 1-\exp(x)^{-\frac{1}{2}}$. Given $\lambda^{(i)}$ from* (23), *$\Pi^{(i)} > 0$ and $Y^{(i)} > 0$, the following inequality holds*

$$f\left(\sum_{i=1}^{m}\mathrm{tr}\left(\Pi^{(i)}Y_*^{(i)}\right)\right) \leq \lambda^{opt} \leq mg\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{tr}\left(\Pi^{(i)}Y_*^{(i)}\right)\right) \tag{27}$$

where λ^{opt} is the global minimum of Problem 1.

Proof. Let $u = \sum_{i=1}^{m} u_i$ and $u_i = \mathrm{tr}\left(\Pi^{(i)}Y_*^{(i)}\right)$. We observe that

$$f(u) \leq \sum_{i=1}^{m} f(u_i) \leq \lambda^{opt} \leq \sum_{i=1}^{m} g(u_i) \leq mg\left(\frac{u}{m}\right) \tag{28}$$

The first equality holds for $u = 0$. The inequality holds since partial derivatives of $\sum_{i=1}^{m} f(u_i)$ with respect to u_i are greater than or equal to those of $f(u)$. The second and third inequalities are proved in [6]. Applying Jensen's inequality to g which is concave, we get the last inequality. □

Since the optimum value of our objective function can be bounded by two increasing functions of $\sum_{i=1}^{m}\mathrm{tr}\left(\Pi^{(i)}Y^{(i)}\right)$, we propose the following convex relaxation to Problem 1.

Problem 2: $\displaystyle Y_*^{(i)} = \underset{Y^{(i)}\geq 0,\ i=1,\cdots,m}{\arg\min} \sum_{i=1}^{m}\mathrm{tr}\left(\Pi^{(i)}Y^{(i)}\right),$
subject to $\overline{P} \leq \Delta$. (29)

There exist challenges with the constraint since $\overline{P}$ is only defined through $\overline{X}$ which itself is defined through an implicit function g_w . The following theorem allows us to obtain an equivalent set of constraints and thus formulate the problem as a semi-definite program. The proof is found in [19].

Theorem 5. *The optimal $Y^{(i)}$ satisfying Problem 2 can be found by solving the following problem.*

Solve: $\displaystyle Y_*^{(i)} = \underset{Y^{(i)}\geq 0,\ i=1,\cdots,m}{\arg\min} \sum_{i=1}^{m}\mathrm{tr}\left(\Pi^{(i)}Y^{(i)}\right),$

[1] $Y^{(i)} \geq 0$ is chosen to ensure the problem is feasible for solvers. To ensure $Y \in \mathbb{S}_{++}^m$, consider $Y^{(i)} \geq \epsilon I$ where $\epsilon > 0$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2015.2505066, IEEE Transactions on Automatic Control

$$\begin{bmatrix} Q^{-1}-S+C'R^{-1}C & Q^{-1}A & C'R^{-1} \\ A'Q^{-1} & A'Q^{-1}A+S & 0 \\ R^{-1}C & 0 & Y+R^{-1} \end{bmatrix} \geq 0,$$
$$Y^{(i)} \geq 0, \quad S \geq \Delta^{-1}.$$

VI. Numerical Analysis

To assess performance, we consider a thermal model for data centers, introduced in [20]. The size of data centers has been growing both in number and capacity, resulting in rising energy costs. To conserve energy, [20] considers the following thermal model for energy control.

$$\begin{bmatrix} \dot{T}_s^{out} \\ \dot{T}_c^{out} \\ \dot{T}_o^{out} \end{bmatrix} = \begin{bmatrix} k_s(\Psi_{ss}-1) & k_s\Psi_{sc} & k_s\Psi_{so} \\ k_c\Psi_{cs} & k_c(\Psi_{cc}-1) & k_c\Psi_{co} \\ k_o\Psi_{os} & k_o\Psi_{oc} & k_o(\Psi_{oo}-1) \end{bmatrix} \begin{bmatrix} T_s^{out} \\ T_c^{out} \\ T_o^{out} \end{bmatrix} + Bu, \tag{30}$$

$$\begin{bmatrix} T_s^{in} \\ T_c^{in} \\ T_o^{in} \end{bmatrix} = \begin{bmatrix} \Psi_{ss} & \Psi_{sc} & \Psi_{so} \\ \Psi_{cs} & \Psi_{cc} & \Psi_{co} \\ \Psi_{os} & \Psi_{oc} & \Psi_{oo} \end{bmatrix} \begin{bmatrix} T_s^{out} \\ T_c^{out} \\ T_o^{out} \end{bmatrix} + Du. \tag{31}$$

Here the state x is a collection of output temperatures of devices while the measured values y are the input temperatures of devices which require multiple sensors. The subscripts represent different nodes under consideration, where 's' corresponds to servers, 'c' corresponds to air conditioners, and 'o' corresponds to other devices. The inputs include a reference temperature for the air conditioners, power consumed, and temperature of heat sources. Ψ gives weight to how the temperature output of each node affects the temperature into each node and k is a set of thermal constants. Addressing the trade-off between estimation and communication in this example will reduce energy expenditures and data storage necessary for thermal control.

To obtain a model consistent with (1), we linearize the system around its stable equilibrium, and assume the inputs remain at or near their equilibrium values for all time, a valid assumption during the night or backup periods. Furthermore, we sample the system at a rate of $\frac{1}{150}Hz$. We consider a system with 16 servers, 3 air conditioners, and 1 other device. The matrices Q and R are generated as a product of a random matrix with entries uniform from 0 to 1 multiplied by its transpose. The matrices are scaled so that the average magnitude of error in w_k is 0.1 Kelvin and in v_k is 0.5 Kelvin.

In Fig 1, we plot the apriori mean squared error in the state estimate as a function of the average communication rate, where each data point is obtained over a run of 10,000 trials. We consider 4 main designs. We first consider a random design where for each sensor at each time step, the probability of transmission is λ_{avg}. We also consider a stochastic design where each sensor communicates at the same rate, and an optimized design from Problem 2.

Finally, for comparison we include a deterministic trigger defined according to the rule $\gamma_k^{(i)} = \mathbb{1}_{\|y_k^{(i)}\|>\delta^{(i)}}$. $\delta^{(i)}$ are chosen so sensor i communicates at the same rate as sensor i in the optimized stochastic trigger. A sub-optimal estimator is incorporated here where a posteriori estimates $\hat{x}_k$ as well as P_k are obtained using a Kalman filter for just the received measurements, (10) and (12). We note that even approximate MMSE or maximum likelihood estimators as proposed by [13] and [14] are computationally inefficient for the given trigger centered around $y_k^{(i)} = 0$ and thus cannot be used to improve the estimate. [13] requires multidimensional numerical integration across the entire state. Meanwhile, sensors which do not transmit measurements in [14] are formulated as constraints in an optimization problem. To solve, we must evaluate all 3^{l_k} possible combinations of active constraints, which becomes a computational burden for a large number of sensors. Also shown are upper and lower bounds for the un-optimized approach.

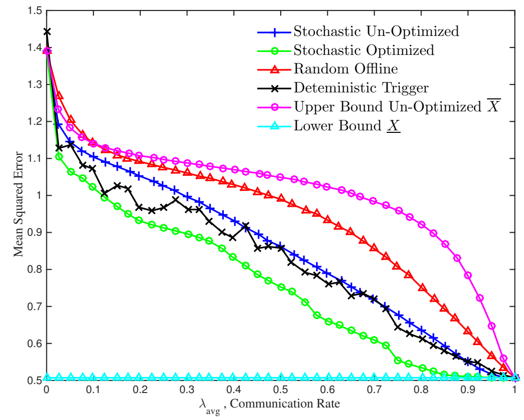

Fig. 1. Mean square error (Kelvin^2) for random, deterministic, stochastic, and stochastic optimized strategies vs λ_{avg}, the communication rate

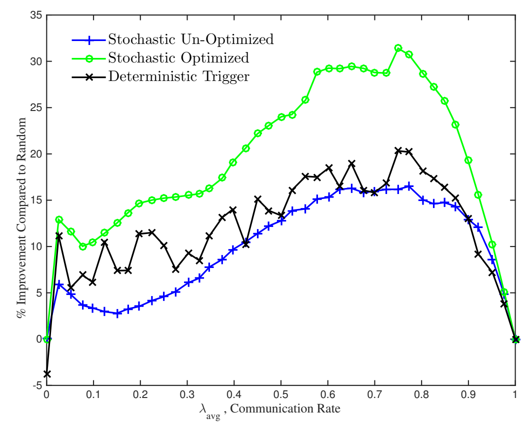

Fig. 2. Percent improvement of designed stochastic triggers and deterministic trigger compared to random triggers

In Fig 2, we plot the percent improvement of the stochastic and deterministic designs relative to the random design in terms of the mean squared error plotted in Fig 1. An un-optimized stochastic design provides as much as 15% improvement, a deterministic design offers as much as 20% improvement, and the optimized stochastic design offers as much as 30% improvement.

VII. Conclusion

In this paper we considered a stochastic event trigger for the sensor scheduling problem in multi-sensor networked systems. The stochastic trigger has inherent advantages over offline triggers which can not improve estimates using information contained by the absence of a measurement. Moreover, it maintains the Gaussian properties of the state, an advantage over previous event triggered approaches. We thus could derive a recursive filter to obtain the MMSE estimator and error covariance. Additionally, we obtained an expression for sensor communication rate as well as asymptotic bounds for our error covariance. Finally, we introduced an optimization problem that will allow designers to reduce the overall communication rate in the system subject to some upper bound on the worst case error covariance. Future work consists of considering the stochastic trigger in a system with control inputs and incorporating inter-sensor cooperation.

References

[1] J. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[2] N. Mahalik, "Sensor networks and configuration: fundamentals, standards, platforms, and applications," *Recherche*, vol. 67, p. 02, 2007.

[3] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, 2006.

[4] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 735–744, 2005.

[5] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011.

[6] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Transactions on Automatic Control, Accepted*, 2014. [Online]. Available: http://arxiv.org/pdf/1402.0599.pdf

[7] C. Yang and L. Shi, "Deterministic sensor data scheduling under limited communication resource," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 5050–5056, 2011.

[8] L. Shi and H. Zhang, "Scheduling two gauss–markov systems: An optimal solution for remote state estimation under bandwidth constraint," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2038–2042, 2012.

[9] K. J. Astrom and B. M. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *Proceedings of IEEE Conference on Decision and Conference*, no. 2, 2002, pp. 2011–2016.

[10] O. C. Imer and T. Basar, "Optimal estimation with limited measurements," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, 2005, pp. 1029–1034.

[11] Y. Xu and J. Hespanha, "Estimation under uncontrolled and controlled communication in networked control systems," in *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference*, 2005, pp. 842–847.

[12] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "Soi-kf: Distributed kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782 – 4795, 2006.

[13] J. Wu, Q. Jia, K. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 1041–1046, 2013.

[14] D. W. Shi, T. Chen, and L. Shi, "Event-triggered maximum likelihood state estimation," *Automatica*, vol. 50, no. 1, pp. 247–254, 2014.

[15] C. Ramesh, H. Sandberg, and K. H. Johansson, "Design of state-based schedulers for a network of control loops," *IEEE Transactions on Automatic Control*, vol. 58, no. 8, pp. 1962–1975, 2013.

[16] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han, and L. Shi, "Multi-sensor scheduling for state estimation with event-based stochastic triggers," in *4th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, 2013, pp. 15–22.

[17] G. Lipsa and N. Martins, "Remote state estimation with communication costs for first-order lti systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[18] L. Scharf, *Statistical Signal Processing*. Addison Wesley, 1991.

[19] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han, and L. Shi, "Multi-sensor scheduling for event-based, stochastic triggers," 2015. [Online]. Available: http://arxiv.org/pdf/1502.03068v1.pdf

[20] L. Parolini, "Models and control strategies for data center energy efficiency," Ph.D. dissertation, Carnegie Mellon University, 2012.

VIII. Appendix

Proof. (Corollary 1) To begin we define function $h(X,\Psi) : S^n_{++} \times \{0,1\}^s \to S^n_{++}$ as

$$h(X,\Psi) \triangleq X - XC'\left(CXC' + R + (I - \mathrm{diag}(\Psi))Y^{-1}\right)^{-1} CX. \tag{32}$$

Using the matrix inversion lemma

$$h(X,\Psi) = \left(X^{-1} + C'\left(R + (I - \mathrm{diag}(\Psi))Y^{-1}\right)^{-1} C\right)^{-1}.$$

This implies h is monotonically increasing in X, and maximized for $\Psi = \mathbf{0}_s$. From Theorem 1, we observe that

$$P_k = h(P_k^-, [\gamma_k^{(1)}\mathbf{1}'_{s_1} \cdots \gamma_k^{(m)}\mathbf{1}'_{s_m}]'). \tag{33}$$

By Theorem 4.2, we have that $P_k^- \leq \overline{X} + \tilde{\epsilon} I$ for $k \geq \bar{N}(\tilde{\epsilon})$. By the monotonicity of h, we obtain

$$\begin{aligned} P_k &\leq h(\overline{X} + \tilde{\epsilon} I, \mathbf{0}) \\ &= (\overline{X} + \tilde{\epsilon} I) - \\ &(\overline{X} + \tilde{\epsilon} I)C'(C(\overline{X} + \tilde{\epsilon} I)C' + R + Y^{-1})^{-1}C(\overline{X} + \tilde{\epsilon} I). \end{aligned}$$

Moreover, by the continuity of h in X, for any $\epsilon > 0$ there exists $\tilde{\epsilon} > 0$ such that

$$\begin{aligned} P_k &\leq h(\overline{X} + \tilde{\epsilon} I, \mathbf{0}) \\ &\leq \overline{X} - \overline{X}C'(C\overline{X}C' + R + Y^{-1})^{-1}C\overline{X} + \epsilon I \\ &= \overline{P} + \epsilon I, \end{aligned} \tag{34}$$

for $k \geq \bar{N}(\tilde{\epsilon}) = N(\epsilon)$. We must now show that P_k approaches this upper bound infinitely many times. To do this, define function $\bar{h} : S^n_{++} \to S^n_{++}$ as

$$\begin{aligned} &\bar{h}(X) \triangleq (AXA' + Q) - \\ &(AXA' + Q)C'(C(AXA' + Q)C' + R + Y^{-1})^{-1}C(AXA' + Q) \\ &= h(AXA' + Q, \mathbf{0}). \end{aligned} \tag{35}$$

Note that $\bar{h}$ is monotonically increasing in X since h is monotonically increasing in its first argument and $AXA' + Q$ is monotonically increasing in X. Utilizing Proposition 1 of [6], we know there exists an $l > 0$ such that

$$\bar{h}^l(0) \geq \overline{X} - \overline{X}C'(C\overline{X}C' + R + Y^{-1})^{-1}C\overline{X} - \epsilon I = \bar{P} - \epsilon I.$$

If event $\bar{E}_{k,l}$ occurs, then we know that

$$P_{k+l} = \bar{h}^l(P_k) \geq \bar{h}^l(0) \geq \bar{P} - \epsilon I. \tag{36}$$

By Theorem 3, the event $\bar{E}_{k,l}$ almost surely occurs infinitely often and thus the result holds. □