Distributed Resilient Estimation over Directed Graphs

Shamik Bhattacharyya, Kiran Rokade, Rachel Kalpana Kalaimani

Department of Electrical Engineering, IIT Madras, Chennai - 36, India

Abstract

This paper addresses the problem of estimating an unknown static parameter by a network of sensor nodes in a distributed manner over a directed communication network in the presence of adversaries. We introduce an algorithm, Resilient Estimation through Weight Balancing (REWB), which ensures that all the nodes, both normal and malicious, asymptotically converge to the value to be estimated provided less than half of them are affected by adversaries. We discuss how our REWB algorithm is developed using the concepts of weight balancing of directed graphs, and the consensus+innovations approach for linear estimation. Numerical simulations are presented to show the performance of our algorithm over directed graphs and its resilience to sensor attacks.

Key words: Resilient Consensus; Distributed Estimation; Directed Graphs; Weight-balancing.

1 Introduction

An advancement in wireless sensor networks (WSNs) has diversified their areas of application to agriculture [10], healthcare [9], renewable energy [8], astronomy [14] to name a few. As a result the scale and complexity of the networks is also on the rise [1], [2]. This necessitates the use of more distributed approaches to signal processing. Distributed estimation is a key aspect of it, which deals with determining a parameter of interest locally at each sensor node with cooperation between neighboring nodes [12]. With an increase in areas of application, WSNs have in turn become more vulnerable to adversarial attacks [11]. A major mode of such attacks are aimed to manipulate the normal functioning of the sensor nodes and thus disrupt the overall signal processing capability of the WSNs. Some common threat models used to represent such adversarial attacks are Byzantine [19], malicious [5], Denial of Service (DoS) [3], etc. So the distributed estimation algorithms need to be resilient to adversarial attacks in order to be more effective.

Here we are concerned with distributed estimation based on consensus algorithms of MAS. Different resilient consensus algorithms have been designed in recent years which ensure consensus even in the presence of adversaries. One such family of algorithms are based on the Mean-Subsequence-Reduced (MSR) algorithm [6], [17]. In these algorithms, each sensor node updates its own value based on the average of a sorted list of values received from its neighbours after discarding certain unwanted values. On the other hand, the Median Consensus Algorithm (MCA) introduced in [20], updates the state of the sensor node based on the median of the values in the sorted list without discarding any of them. In these resilient consensus algorithms, only the unaffected nodes achieve consensus while the sensor nodes under attack may have arbitrary values.

The consensus+innovations approach introduced in [15] uses the consensus framework to design algorithms for linear estimation. Based on this approach the Constant weight Saturated Innovation Update (CSIU) algorithm, introduced in [4], is a distributed estimation algorithm resilient to sensor attacks. In [4] a new term resilience index ι was used Along with that a to model the fraction of sensor nodes under attack. The CSIU algorithm ensures that all the agents are able to estimate the parameter of interest provided less than three-tenth of the total agents are under attack. This was further improved in [5], where the Saturated Innovation Update (SIU) algorithm ensures all the agents’ estimate converge to the the desired parameter value provided the adversaries attack less than half of the total agents. Both the above algorithms are designed on undirected graphs representing bidirectional communication links between the agents.

In this paper we look into resilient distributed estimation over directed graphs. In many practical scenar-
ios, the power levels at which sensor nodes broadcast information or their interference and noise patterns differ from node to node [7], [18]. The communication between nodes in such cases is unidirectional which is aptly represented by a directed graph. To the best of our knowledge this is the first work that deals with a directed network topology for resilient distributed estimation where all the nodes, both the normal ones and those under attack, asymptotically estimate the desired parameter value. A time-invariant network topology is used and the fault-model is considered to be malicious. We present an algorithm, Resilient Estimation through Weight Balancing (REWB), which ensures that all agents asymptotically converge to the value to be estimated provided less than half of the total number of agents are affected by adversaries. The REWB algorithm ensures only the local information available to them. The agents operate in a distributed manner using only the local information available to them. The agents under attack by the adversaries are not known a-priori by the normal agents. Moreover we allow for a more general scenario where the adversaries may attack different agents over time. The REWB algorithm ensures that not only the normal agents, but also the malicious ones asymptotically reach consensus over the required estimation parameter.

**Notations.** \( \mathbb{R} \) denotes the set of real numbers, and \( \mathbb{R}^N \) represents the \( N \)-dimensional Euclidean space. For any set \( S \), the cardinality of the set is denoted by \( |S| \). \( \mathbb{I} \) denotes a vector of all 1s, \( \mathbb{I} = (1, 1, \ldots, 1) \), and \( \mathbb{O} \) denotes a vector of all 0s, \( \mathbb{O} = [0, 0, \ldots, 0] \), of appropriate dimensions. For a real-valued vector \( v \), \( v^T \) denotes the transpose of the vector, \( ||v|| \) denotes its \( l_2 \)-norm and \( ||v||_\infty \) denotes its \( \infty \)-norm. Similarly for a real-valued matrix \( M \), \( M^T \) denotes the transpose of the matrix, and \( ||M|| \) denotes its spectral norm. Among the eigenvalues of \( M \), \( \lambda_2(M) \) denotes the second lowest eigenvalue of \( M \), while \( \lambda_{\text{max}}(M) \) denotes its largest eigenvalue. For a real-valued vector \( v \), \( \text{diag}(v) \) represents a diagonal matrix with \( v \) as the main diagonal.

The rest of the paper is organised as follows. Section-2 discusses the details of the problem like the inter-agent communication network, the threat model of the adversaries and the concept of resilience index. Section-3 starts with the development of the REWB algorithm using the weight-balancing approach, followed by the details of the algorithm, finally leading to our main result. Some numerical simulations are presented in Section-4 to validate the performance of the REWB algorithm. Finally the conclusions are presented in Section-5.

2 Problem Formulation

2.1 System Model

Consider a system of \( N \) agents where each agent is equipped with sensing, computing and communication capabilities - it can record measurements using its sensor, can perform computations using its own data and the information received from its neighbouring agents, and can also share its data with the neighbours. The aim of each agent is to estimate an unknown parameter \( \theta^* \in \mathbb{R} \) in a distributed manner even while some agents’ sensor measurements are corrupted by adversaries.

The communication among the agents is modelled as a directed graph \( \Gamma = (V, E) \), where the vertex set \( V = \{1, 2, \ldots, N\} \) represents the set of \( N \) agents. The set of directed edges \( E \subset V \times V \) represents the information exchange links between the agents, where \( (i, j) \in E \) if agent-\( j \) can send information to agent-\( i \). A directed path between vertices \( i \) and \( j \) is the sequence of directed edges \((i, i_1), (i_1, i_2), \ldots, (i_p, j)\). The set of in-neighbours of agent-\( j \) is defined as \( N_j = \{i \in V : (i, j) \in E\} \), and the corresponding in-degree is denoted as \( d_j^{\text{in}} = |N_j| \). The set of out-neighbours of agent-\( j \) is defined as \( O_j = \{i \in V : (j, i) \in E\} \), and the corresponding out-degree is denoted as \( d_j^{\text{out}} = |O_j| \). A corresponding diagonal matrix is defined as \( D^{\text{out}} = \text{diag}(d_1^{\text{out}}, \ldots, d_N^{\text{out}}) \). The adjacency matrix, \( A \) is a square matrix of size \( N \times N \) defined as \( A = [a_{ij}] \) where \( a_{ij} = 1 \) if \( (j, i) \in E \), and \( a_{ij} = 0 \) otherwise. The Laplacian, \( L \) is defined as \( L := (D^{\text{out}} - A) \).

**Definition 1.** (Strongly Connected) A directed graph is said to be strongly connected if there exists a directed path between every pair of vertices in the graph.

The flow of information is such that each agent \( i \) is able to receive information from its in-neighbours \( (N_i) \), and send its own data to its out-neighbours \( (O_i) \). In order to ensure that the information about every agent \( i \) is received by every other agent \( j \) (\( i \neq j \)), either directly through a directed communication link between them, or indirectly via intermediate agent(s), we introduce the following assumption.

**Assumption 1.** The directed graph \( \Gamma \) is Strongly Connected.

Now we proceed to model the effect of the adversaries, which attack the agents with a motive to disrupt the estimation process thus trying to prevent them from correctly estimating the value \( \theta^* \). At every time-step \( t \geq 0 \), the agents which are under attack by the adversaries are termed as the the set of Bad (or malicious) agents, denoted as \( B_t \). The remaining agents form the set of Good (or normal) agents, denoted as \( G_t \). The set of bad agents can vary with time, and are also not known a-priori to the set of good agents. So for each \( t \geq 0 \), the set \( V \) is partitioned into \( G_t \) and \( B_t \). Thus \( G_t \cup B_t = V, \forall \ t \geq 0 \). Here we consider the mode of attack where the adversaries corrupt the sensor measurements of the agents either by introducing noise into the measurement or by completely replacing the actual measurement by an arbitrary value. Accordingly, we model the sensor mea-
The resilience index (\( s \)) is defined as an upper bound on the fraction of agents which are under attack by the adversaries at any time-step \( t \). So, \( s \geq \frac{|B_t|}{|V|} \) for all \( t \geq 0 \), \( s \in \mathbb{R} \). Thus \( s = 0 \) would indicate the trivial case where bad agents are totally absent. Having \( s = 1 \) allows for the possibility of all the agents being under attack at any time-step \( t \).

In the sequel, we initially proceed to design an algorithm which provides us with a suitable value of \( u_i(t) \) \( \forall t \geq 0 \), for all \( i \in V \). Then we present our main result on how the newly designed algorithm, under certain assumptions, helps us achieve (4).

### 3 Results

The aim of each agent in the multi-agent system under consideration, is to estimate an unknown static parameter in a distributed manner, as given in (4). The technique used for the distributed estimation of \( \theta^* \) is based on the consensus+innovations approach [15]. Based on this approach we proceed to design an algorithm such that the desired objective is achieved through fulfilling the following two smaller goals simultaneously as \( t \rightarrow \infty \):

- \( G1: \) the state of each agent, \( x_i(t) \) approaches the combined average of the states of all the agents, \( \bar{x}(t) \)
- \( G2: \) \( \bar{x}(t) \) approaches the unknown value to be estimated, \( \theta^* \).

#### 3.1 Weight Balancing

In a MAS, the communication network is usually modelled as a graph, with the nodes of the graph representing the agents and the edges between the nodes representing the corresponding communication links between the agents. When the flow of information between agents is bi-directional, the model used is an undirected graph and its edges represent the corresponding bidirectional communication links. On the other hand, when the flow of information between agents is unidirectional, a directed graph (or digraph) is required to model it. The directed edges of the digraph represent the unidirectional communication links while preserving their direction of information flow.

In case of an undirected graph, the amount of information flowing into a node is equal to that flowing out of it. But this balance in information flow does not necessarily hold true in the case of a digraph. To overcome this imbalance, we scale the information sent by agent \( i \) by weight \( w_i \in \mathbb{R} \) for all \( i \in V \). These weights are said to balance the graph if \( \sum_{j \in N_i} w_j = \sum_{j \in N_i} w_i \). The notion of a balanced graph is formally defined below -

**Definition 2.** (Balancing Weights) The node weights \( w \in \mathbb{R}^{N} \), balance the directed graph of \( N \) nodes, \( \Gamma \) if the
The weights \( \{w_1, w_2, \ldots, w_N\} \) which balance a given digraph are called the balancing weights of the corresponding digraph [18]. Note that an undirected graph is inherently balanced with \( w = I \) as the vector of balancing weights. On the other hand, for a strongly connected digraph the vector of balancing weights is also unique to the given digraph [18]. For example, consider the strongly connected digraph shown in Fig.1. For this graph the vector of balancing weights is \( w = [0.5, 1.5, 1]^T \), which is non-trivial and unique.

The SIU algorithm proposed in [5] for resilient estimation does not work in general for directed graphs. This is illustrated through a numerical example in Fig.4 in Section 4. We propose to use the idea of balancing weights described above to achieve resilient estimation over digraphs. For the directed graph \( \Gamma \), we use the following update rule, proposed in [18], to iteratively compute a set of balancing weights. Let \( w_i(t) \in \mathbb{R} \) denote the weight at node \( i \) at time-step \( t \). The initial set of weights assigned to the agents satisfy : \( w_i(0) \leq (1/d^*)^{2B+1} \), where \( d^* \) represents the maximum out-degree and \( D \) represents the diameter of the concerned digraph [18]. Then for all \( t \geq 0 \),

\[
\begin{align*}
    w_i(t+1) &= \frac{1}{2} w_i(t) + \frac{1}{d^*_{out}} \left( \sum_{j \in N_i} \frac{1}{2} w_j(t) \right) \quad (6)
\end{align*}
\]

Let \( w(t) = (w_1(t), w_2(t), \ldots, w_N(t)) \) represent the vector of node-weights at time-step \( t \). Then the corresponding vector representation of (6) is given by :

\[
    w(t+1) = P w(t) \quad (7)
\]

where \( P := 0.5 \left( I + (D^\text{out})^{-1} A \right) \). So for the limiting case, \( \lim_{t \to \infty} w(t) = \lim_{t \to \infty} P^t w(0) \). Now from Lemma 1 of [18] we know that \( \lim_{t \to \infty} P^t \) exists, and that the sequence \( \{w(t)\}_{t \geq 0} \) converges to the vector of balancing weights. Here we define the vector of weights which balances the digraph \( \Gamma \) as

\[
    w^\infty := \lim_{t \to \infty} w(t) = \lim_{t \to \infty} P^t w(0) \quad (8)
\]

The time-varying weighted Laplacian matrix is represented as

\[
    L(t) = (D^\text{out} - A) W(t), \quad \text{where} \quad W(t) = \text{diag}(w(t)) \quad (9)
\]

Then the Laplacian matrix for the limiting case can be defined using the result from (8) in (9) as

\[
    L^\infty := (D^\text{out} - A) W^\infty, \quad \text{where} \quad W^\infty = \text{diag}(w^\infty) \quad (10)
\]

Now as \( w^\infty \) balances the digraph, \( L^\infty \) satisfies the desired balancing condition expressed in (5). By definition of \( L(t) \) we have \( 1^T L(t) = 0 \) for all \( t \geq 0 \). So to arrive at the desired balanced graph condition, we need \( L(t) \parallel = 0 \) which is eventually achieved with \( L(t) \) converging to \( L^\infty \) as \( t \to \infty \). We state a lemma which provides an explicit rate for this convergence and additionally provides the rate of decay of \( L(t) \parallel \) to \( \parallel \).

**Lemma 1.** Given \( L(t) = (D^\text{out} - A) W(t) \) and \( L^\infty = (D^\text{out} - A) W^\infty \), there exists constants \( C > 0 \) and \( \eta \in (0, 1) \), such that \( \|L(t) - L^\infty\| \leq C \lambda^t \), \( \|L(t) \parallel \| \leq C \lambda^t \) for all \( t \geq 0 \).

### 3.2 Algorithm

Now we introduce our algorithm - Resilient Estimation through Weight Balancing (REWB). It consists of two main update steps - one for the state of the agents, and the other for the node weights.

The updates performed by agent \( i \) at time-step \( t \) are :

i) Updating the estimate

\[
    x_i(t+1) = \left(1 - \beta(t) w_i(t) d^\text{out}_{i} \right) x_i(t) \\
    + \beta(t) \left( \sum_{j \in N_i} w_j(t) x_j(t) \right) + \alpha(t) k_i(t) (y_i(t) - x_i(t)) \quad (11)
\]

ii) Updating the weight

\[
    w_i(t+1) = \frac{1}{2} w_i(t) + \frac{1}{d^*_{out}} \left( \sum_{j \in N_i} \frac{1}{2} w_j(t) \right) \quad (12)
\]

The update law (11) used by agents to update their estimate of \( \theta^* \) is based upon the consensus+innovation approach. The first two terms dealing with the agent’s own and neighbours’ estimates and the corresponding node-weights constitute the consensus part of the update law. The third term involving the measurements, \( y_i(t) \) and a scaling factor, \( k_i(t) \) constitute the innovation part. These two parts working simultaneously through the same update law help in achieving the smaller goals \( G1 \) and \( G2 \) mentioned before. The above update law uses step-size parameters \( \beta(t) \) and \( \alpha(t) \) to assign proper
weightage to its consensus and innovation parts respectively. The parameters are defined as:

\[ \alpha(t) = \frac{\alpha_0}{(1 + t)^{\alpha_1}}, \beta(t) = \frac{\beta_0}{(1 + t)^{\beta_1}} \]  

where \( 0 < \alpha_0 \leq \frac{1}{2 \alpha_1}, \ 0 < \beta_0 < \psi, \ 0 < \beta_1 < \alpha_1 < 1. \)

The constant \( \psi \) is defined as \( \psi := \frac{\lambda_2((L^\infty)^2 + L^\infty)}{\lambda_{\text{max}}((L^\infty)^2 L^\infty)} \).

The scaling factor, \( k_i(t) \), is used in the innovation part in order to ensure that the effect of the adversaries on the state of an agent always remains bounded.

\[ k_i(t) := \begin{cases} 1 & \text{if } |y_i(t) - x_i(t)| \leq \gamma(t) \\ \frac{\mu_0 - x_i(t)}{\gamma(t)} & \text{otherwise} \end{cases} \]  

where \( \gamma(t) \) is the output of a dynamical system defined as

\[ \gamma(t) := \gamma_1(t) + \gamma_2(t) \]  

The dynamics of \( \gamma_1(t) \) and \( \gamma_2(t) \) are defined as

\[ \gamma_1(t + 1) := (1 - c_1 \mu(t) + \sqrt{N} \alpha(t)) \gamma_1(t) \]  

\[ + (1 + \sqrt{N} \alpha(t)) \gamma_2(t) + c_2 \eta(t)^2 \]  

\[ \gamma_2(t + 1) := \alpha(t) \gamma_1(t) + (1 - \alpha(t)(1 - 2s)) \gamma_2(t) \]  

where, \( \mu(t) = \frac{\mu_0}{(t + 1)^{\psi}}, \mu_0 > 0, \beta_1 < \mu_1 < \alpha_1, c_1 > 0, c_2 > 0. \) The above time-varying system in two variables plays a crucial role in proving our main result. From the definition of \( k_i(t) \) in (14), a corresponding diagonal matrix is defined as

\[ K(t) := \text{diag}(k_1(t), k_2(t), \ldots, k_N(t)) \]  

Let \( x(t) = (x_1(t), x_2(t), \ldots, x_N(t)) \) represent the vector of the states of the agents at time-step \( t \). Now we summarise our REWB algorithm as follows:

**Algorithm 1 : REWB**

**Given** : Graph \( \Gamma, \Theta \geq \theta^* \), Resilience index \( s \)

**Initialize** : \( 0 < \alpha_0 \leq 1/(1 - 2s), \ 0 < \beta_0 < \psi, \mu_0 > 0, \gamma_1(0) = 0, \gamma_2(0) = \Theta, \ x(0) = 0, \ w_i(0) \leq \left( \frac{1}{\mu_0} \right)^{2(\beta_2 + 1)} \)

**Choose** : \( 0 < \beta_1 < \mu_1 < \alpha_1 < 1 \)

for \( t = 0, 1, \ldots \) do

- record \( y(t) \)
- exchange \( x(t) \) among neighbouring agents
- update \( x(t) : \)
  \[ x(t + 1) = (I - \beta(t)L(t))x(t) + \alpha(t)K(t)(y(t) - x(t)) \]
- update \( w(t) : \)
  \[ w(t + 1) = P(w(t) \]
- update \( \gamma(t) : \) using equations (15), (16) & (17)

end for

### 3.3 Main Result

The following theorem states our main result on resilient distributed estimation using the REWB algorithm.

**Theorem 1.** Suppose Assumptions 1 and 2 hold, and the effect of the adversaries is modelled as in (1). Then the REWB algorithm ensures that the state of every agent, \( x_i(t) \) converges to \( \theta^* \), provided \( s \in [0, \frac{1}{2}) \). In particular,

\[ \lim_{t \to \infty} (t + 1)^{\delta_1} (x_i(t) - \theta^*) = 0, \text{ for all } i \in \mathcal{V} \]  

where \( 0 \leq \delta_1 \leq \alpha_1 - \beta_1 \)

The proof of the Theorem-1 is given in Appendix-C. Here we provide some remarks on the above theorem.

**Remark 1.** From Theorem-1 it can be inferred that as long as the number of bad agents is less than half the total number of agents, the REWB algorithm ensures that each agent, regardless of it being affected by adversaries, is able to correctly estimate \( \theta^* \).

### 4 Simulation Results

We evaluate the performance of our proposed REWB algorithm through numerical simulations. We generate a random network consisting of 100 agents with directed edges which models the communication network among the agents. Each agent estimates a parameter \( \theta^* \in \mathbb{R} \), the value of which is upper bounded as \( |\theta^*| \leq \Theta = 50 \). The required parameters are chosen as : \( \alpha_0 = 0.01, \alpha_1 = 0.25, \beta_0 = 0.02, \beta_1 = 0.025, \mu_0 = 0.025, \mu_1 = 0.05, c_1 = 10, c_2 = 10 \), and \( \eta = 0.01 \).

The noise term \( \zeta_i(t) \) models the effect of the adversaries on the sensor measurements of agent-\( i \). For each bad agent \( i \in \mathcal{B}_t \), at every time step, \( \zeta_i(t) \) takes on a random value uniformly distributed between 0 and \( -\Theta \). Note that the REWB algorithm works for any other range also. We select the base resilience index to be \( s = 0.405 \), and correspondingly choose \( |\mathcal{B}_t| = 40 \). At first we consider two cases with respect to the set of agents under attack and observe the performance of the REWB algorithm. In Fig.2, the plot on the left is the case where \( \mathcal{B}_t \) has a fixed set of agents, while that on the right is where \( \mathcal{B}_t \) is allowed to vary with time. Both the plots show the error in estimation of \( \theta^* \) by the agents, given by \( ||x(t) - \theta^*|| \). From the proof of Theorem-1 we have \( |x_i(t) - \theta^*| \leq \gamma(t), \forall i \in \mathcal{V}, \forall t \geq 0 \). Then for a set of \( N \) agents, we have \( |x_i(t) - \theta^*|| \leq \sqrt{N}\gamma(t) \). Fig.2 shows that, regardless of adversaries attacking a fixed or varying set of agents, the REWB algorithm ensures that the local estimation error always remains bounded by \( \sqrt{N}\gamma(t) \), and consequently dies down asymptotically.
Next we use two different variations in operating conditions compared to the one used in Fig. 2 and observe their effect in the performance of the REWB algorithm in Fig. 3. For the plot on the left in Fig. 3, the resilience index is decreased to $s = 0.255$ and correspondingly we choose $|\mathcal{B}_t| = 25$. As can be observed the estimation error dies down much faster with a decrease in $s$. Next for the plot on the right in Fig. 3, we simulate an increase in the degree of manipulation done by the adversaries on the sensor measurements by increasing the noise level. We assign $\zeta_i(t) = 5\Theta \forall i \in \mathcal{B}_t, \forall t \geq 0$. As is evident from the plot on the right, a high value of $\zeta_i(t)$ is also quite efficiently handled by the REWB algorithm with the estimation error remaining bounded by $\sqrt{N} \gamma(t)$ at all time and eventually converging to 0. From Fig. 2 and Fig. 3 it is evident that the REWB algorithm ensures that even the bad agents are able to correctly estimate the the true value of $\theta^*$, along with the good agents. This is in accordance with Remark-1 stated in Section-3.3.

In Section-3.1, we mentioned that the SIU algorithm in [5] does not give convergence in general when applied over a directed network of agents. In Fig. 4, we see how the SIU algorithm performs over a directed graph of 100 agents with $s = 0.405$ compared to the proposed REWB algorithm. The two plots on the top show how the states of the agents behave with time, while the two plots at the bottom show the net estimation error. Fig. 4 shows how on applying the SIU algorithm, the states of the agents diverge away from each other and never achieve consensus, leading to a constant estimation error. On the other hand, Fig. 4 also shows how our REWB algorithm not only ensures the agents reach consensus but they also correctly estimate the value of $\theta^*$. This is made possible by the introduction of the weight balancing idea while designing the REWB algorithm. The dynamics of the time-varying weights ensure that the weighted graph eventually approaches a balanced condition, and thus consensus is achieved.

5 Conclusion

In this paper we propose the Resilient Estimation through Weight Balancing (REWB) algorithm which is a distributed estimation algorithm designed to work for a network of sensor nodes with directed communication links. The REWB algorithm is designed to be resilient to adversarial attacks on sensor nodes while estimating an unknown parameter of interest. It is developed based on the consensus+innovation approach and uses the weight-balancing idea to ensure consensus over directed graph. Through numerical simulations it is shown that the proposed algorithm accurately estimates an unknown parameter provided less than half of the agents are under adversarial attack. Future direction of work is to implement REWB algorithm over an event-based communication strategy.

References

[1] "A survey on sensor networks. IEEE Communications Magazine, 40(8):102–114, 2002.
[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. Computer Networks, 38(4):393–422, 2002.
[3] Bo Chen, Li Yu, Daniel W.C. Ho, and Wen-An Zhang. Networked fusion estimation under denial-of-service attacks. IFAC-PapersOnLine, 50(1):3835–3840, 2017. 20th IFAC World Congress.
[4] Yuan Chen, Soummya Kar, and Jose M. F. Moura. Resilient distributed estimation : Exponential convergence under sensor attacks. Proceedings for the IEEE Conference on Decision and Control 2018, pages 7275–7282, 2018.
3.1. A Proof of Lemma-1

Following is the proof of Lemma-1 stated in section-3.1.

Proof. By definition, $P$ is a primitive matrix with spectral radius 1. Then by properties of primitive matrices: $\lim_{t \to \infty} P^t$ exists, and $\lim_{t \to \infty} P^t = P^\infty = uu^T$, where $u, v$ are the right and left eigen-vectors of $P$ corresponding to eigen-value 1, and $v^Tu = 1$. Then from (6) and (8) we have:

$$w(t) - w^\infty = (P^t - P^\infty)w(0) \quad (A.1)$$

$$\quad (P - P^\infty)^2 = P^2 - PP^\infty - P^\infty P + P^\infty P^\infty$$

$$\quad = P^2 - Puu^T - uu^TP + uu^Tuu^T$$

$$\quad = P^2 - uu^T - uu^T + uu^T$$

$$\quad : (P - P^\infty)^2 = P^2 - P^\infty$$

Proceeding similarly we have $(P - P^\infty)^t = P^t - P^\infty$, for all $t \geq 1$. Then from (A.1) we have:

$$w(t) - w^\infty = (P^t - P^\infty)w(0) = (P - P^\infty)^t w(0)$$

So by Theorem 8.3 in [13], there exists $c > 0, \eta < 1$ such that for all $t \geq 0$

$$\|w(t) - w^\infty\| \leq c\eta^t \|w(0)\| \quad (A.2)$$

Using (9) and (10), and applying the properties of sub-multiplicativity of spectral norm we have

$$\|L(t) - L^\infty\| \leq \|D^\text{out} - A\|\|W(t) - W^\infty\| \quad (A.3)$$

Now $W(t), W^\infty$ are diagonal matrices and for any diagonal matrix $M = \text{diag}(m_1, \ldots, m_N)$ we have $\|M\| \leq \|m\|_\infty \leq \|m\|$. Applying this to (A.3) and using the result from (A.2) we get

$$\|L(t) - L^\infty\| \leq C_L\eta^t \quad (A.4)$$

where $C_L = c\|D^\text{out} - A\|\|w(0)\| > 0$. From (9) using the sub-multiplicativity property of the norm and the result from (A.4) we get

$$\|L(t)\| = \|(L(t) - L^\infty) + L^\infty\| \leq \sqrt{N}C_L\eta^t \quad (A.5)$$

By choosing $C \geq \sqrt{N}C_L$ and applying to (A.4) and (A.5) we get:

$$\|L(t) - L^\infty\| \leq C\eta^t, \|L(t)\| \leq C\eta^t$$
B Intermediate Lemmas

Here we introduce some intermediate lemmas which will be useful in the proof of Theorem-1. Before proceeding to a time-varying system in two variables, we first analyse the dynamics of a scalar time-varying system. Consider a linear scalar time-varying system -

\[ v_{t+1} = (1 - r_2(t))v_t + r_1(t) \]  

(B.1)

where

\[ r_1(t) = \frac{c_1}{(1 + t)^{\delta_1}} \]
\[ r_2(t) = \frac{c_2}{(1 + t)^{\delta_2}} \]  

(B.2)

where \( c_1, c_2, \delta_2 \) are positive constants, and \( 0 \leq \delta_1 \leq 1 \).

The following result is based upon the results introduced in Lemma 25 in [16] and Lemma 3 in [5]. It provides a relation between \( \delta_1 \) and \( \delta_2 \) under which the dynamics of the scalar time-varying system in (B.1) is bounded. It also gives the condition under which the system dynamics converges to zero, and the corresponding rate of convergence.

**Proposition 1.** Consider the system given in (B.1) where \( r_1(t), r_2(t) \) is given by (B.2). Then if \( \delta_1 = \delta_2 \), there exists \( B > 0 \), such that for sufficiently large non-negative integers \( j < t \),

\[
0 \leq \sum_{k=j}^{t-1} \left( \prod_{l=k+1}^{t-1} (1 - r_1(l)) \right) r_2(k) \leq B
\]

Moreover the constant \( B \) can be chosen independently of \( t, j \). Also, if \( \delta_1 > \delta_2 \), then for arbitrary fixed \( j \),

\[
\lim_{t \to \infty} \sum_{k=j}^{t-1} \left( \prod_{l=k+1}^{t-1} (1 - r_1(l)) \right) r_2(k) = 0
\]

and correspondingly

\[
\lim_{t \to \infty} (t + 1)^{\delta_0} v_t = 0
\]

for all \( 0 \leq \delta_0 < \delta_1 - \delta_2 \), and for all initial conditions \( v_0 \).

The following result provides the rate of convergence of a scalar system modified from (B.1).

**Proposition 2 (Lemma 4 in [5]).** Consider the scalar time-varying linear system :

\[ v_{t+1} = (1 - c_3 r_2(t) + c_4 r_1(t))v_t + c_5 r_1(t) \]  

(B.3)

where \( r_1(t), r_2(t) \) are given by :

\[ r_1(t) = \frac{c_1}{(1 + t)^{\delta_1}}, r_2(t) = \frac{c_2}{(1 + t)^{\delta_2}} \]

where \( c_1, c_2, \ldots, c_5 > 0 \), and \( 0 < \delta_2 < \delta_1 < 1 \). The system in (B.3) satisfies

\[
\lim_{t \to \infty} (t + 1)^{\delta_0} v_t = 0
\]

for all \( 0 \leq \delta_0 < \delta_1 - \delta_2 \), and for all initial conditions \( v_0 \).

Now by using the above results we introduce the following two lemmas. The proof of the main theorem depends on the convergence of \( \gamma_1(t) \) and \( \gamma_2(t) \) introduced in (16), (17). So we first look into the dynamics of a linear time-varying system with two variables similar to \( \gamma_1(t), \gamma_2(t) \). Lemma-2, adapted from Lemma 5 in [5], provides a bound on the system dynamics. Lemma-3, adapted from Lemma 1 in [5], describes the convergence of the system.

Consider a linear time-varying system :

\[
\nu(t + 1) = (1 - \sigma_1 a(t))\nu(t) + \sigma_2 a(t)\omega(t)
\]
\[
\omega(t + 1) = (1 - \sigma_3 b(t) + \sigma_4 a(t))\omega(t) + \sigma_5 a(t)\nu(t) + \sigma_6 b(t)
\]

(B.4)

with non-zero initial conditions \( \nu(0), \omega(0) \neq 0 \), and \( \sigma_1, \ldots, \sigma_6 > 0 \), \( 0 < \rho < 1 \), and \( a(t), b(t) \) follow

\[
a(t) = \frac{a_0}{(1 + t)^{\delta_a}}, b(t) = \frac{b_0}{(1 + t)^{\delta_b}}, \text{ where } \delta_a > \delta_b
\]

**Lemma 2.** The system in (B.4) satisfies

\[
\sup_{t \geq 0} |\nu(t)| < \infty \hspace{1cm} (B.5)
\]
\[
\sup_{t \geq 0} |\omega(t)| < \infty \hspace{1cm} (B.6)
\]

**Proof.** Step 1 : As \( a(t), b(t) \) are decreasing in \( t \), and \( \delta_a > \delta_b \), there exists a finite \( T > 0 \) such that for all \( t > T \)

\[
0 \leq 1 - \sigma_1 a(t) \leq 1
\]
\[
0 \leq 1 - \sigma_3 b(t) + \sigma_4 a(t) \leq 1
\]

(B.7)

From (B.4) we can express \( \nu(t) \) as -

\[
\nu(t) = \prod_{\tau = T}^{t-1} (1 - \sigma_1 a(\tau))\nu(T) + \sum_{\tau = T}^{t-1} \left( \prod_{j = \tau + 1}^{t-1} (1 - \sigma_1 a(j)) \right) \sigma_2 a(\tau)\omega(\tau)
\]

(B.8)

So using Proposition-1 and (B.7) in (B.8)

\[
|\nu(t)| \leq |\nu(T)| + \sigma_7 \sup_{t \in [T, t]} |\omega(t)|
\]

(B.9)
for some constant $\sigma > 0$

**Step 2 :** From (B.4) we can say

$$|\omega(t+1)| \leq |1 - \sigma_3 b(t) + \sigma_4 a(t)||\omega(t)| + \sigma_5|\nu(t)| + \sigma_6 \rho^t$$

Now by using (B.7) we have

$$|\omega(t+1)| \leq (1 - \sigma_3 b(t) + \sigma_4 a(t) \sup_{l\in[T,t]} |\omega(l)| + \sigma_5|\nu(t)| + \sigma_6 \rho^t$$

Finally applying (B.9) we get

$$|\omega(t+1)| \leq (1 - \sigma_3 b(t) + \sigma_8 a(t)) \sup_{l\in[T,t]} |\omega(l)| + \sigma_9 a(t) + \sigma_6 \rho^t$$

where $\sigma_8 = \sigma_4 + \sigma_3 \sigma_7$ and $\sigma_9 = \sigma_5 |\nu(T)|$

Now as $0 < \rho < 1$, there exists $\sigma_{10} > 0$ such that for all $t > 0$

$$\sigma_9 a(t) + \sigma_6 \rho^t < \sigma_{10} a(t) \quad \text{(B.10)}$$

$$\therefore |\omega(t+1)| \leq (1 - \sigma_3 b(t) + \sigma_8 a(t)) \sup_{l\in[T,t]} |\omega(l)| + \sigma_{10} a(t)$$

We define a new system -

$$\eta(t+1) = \max(\eta(t), (1 - \sigma_3 b(t) + \sigma_8 a(t))\eta(t) + \sigma_{10} a(t)) \quad \text{(B.11)}$$

for all $t > T$ and initial condition $\eta(T) = \omega(T)$.

So by definition of $\eta(t)$ we have :

$$\eta(t) \geq \sup_{l\in[T,t]} |\omega(l)| \quad \text{(B.12)}$$

We define another new system :

$$\tilde{\eta}(t+1) = (1 - \sigma_3 b(t) + \sigma_8 a(t))\tilde{\eta}(t) + \sigma_{10} a(t) \quad \text{(B.13)}$$

for all $t > T$ and initial condition $\tilde{\eta}(T) = \eta(T) = \omega(T)$. By definition $\tilde{\eta}(T) \geq 0$. Also for $t > T$, from (B.7) we have $1 - \sigma_3 b(t) + \sigma_8 a(t) \geq 0$. Then $\tilde{\eta}(t) \geq 0$ for all $t \geq T$.

Now using Proposition-1 and (B.13) we have

$$\lim_{t \to \infty} \tilde{\eta}(t) = 0$$

**Step 3 :** By virtue of $\tilde{\eta}(t)$ being a non-negative sequence which converges to 0, there exists a time $T_1 \geq T$ such that $\tilde{\eta}(T_1) \leq \tilde{\eta}(T_1')$. We choose the smallest value among all such possible $T_1 \geq T$.

Then from the definition of $T_1$ we have $\tilde{\eta}(T) \leq \tilde{\eta}(T+1) < \ldots < \tilde{\eta}(T_1)$. So from (B.11), $\eta(t) = \tilde{\eta}(t)$ for all $t \in [T,T_1]$.

$$\therefore \eta(t) \leq \eta(T_1) \quad \text{for all } t \in [T,T_1] \quad \text{(B.14)}$$

Also by definition of $T_1, \eta(t)$ we have $\eta(T_1 + 1) = \eta(T_1)$.

Let for all $t \geq T_1$

$$\pi(t) := \eta(T_1) - (1 - \sigma_3 b(t) + \sigma_8 a(t))\eta(T_1) - \sigma_{10} a(t)$$

By algebraic manipulation

$$\pi(t) = \left(\frac{\sigma_{11}}{(t+1)^{\delta_b}} - \frac{\sigma_{12}}{(t+1)^{\delta_b}}\right) \eta(T_1)$$

where $\sigma_{11} = \sigma_3 b_0 > 0, \sigma_{12} = \left(\frac{\sigma_5 |\nu(T)|}{\delta_b}\right) a_0 > 0$.

Now $\eta(T_1) = 0$, and since $\eta(T_1 + 1) = \eta(T_1)$ we have

$$\pi(T_1) \geq 0 \iff T_1 \geq \left(\frac{\sigma_{12}}{\sigma_{11}}\right)^{1/(\delta_a - \delta_b)} - 1$$

So we have $\pi(t) \geq 0$ for all $t \geq T_1$. Then using (B.11) we have

$$\eta(t) = \eta(T_1), \text{ for all } t \geq T_1 \quad \text{(B.15)}$$

So combining the results from (B.12), (B.14) and (B.15) we get

$$\sup_{t \geq T} |\omega(t)| \leq \eta(T_1) < \infty \quad \text{(B.16)}$$

As $T < \infty$, we have -

$$\sup_{t \in [0,T]} |\omega(t)| < \infty \quad \text{(B.17)}$$

So combining (B.16) and (B.17) we can infer (B.6).

**Step 4 :** Let $\sup_{t \in [0,T]} |\omega(t)| = B_\omega < \infty$.

Then by (B.9) we have

$$\sup_{t \geq T} |\nu(t)| \leq |\nu(T)| + \sigma_T B_\omega < \infty \quad \text{(B.18)}$$

As $T < \infty$, we have

$$\sup_{t \in [0,T]} |\nu(t)| < \infty \quad \text{(B.19)}$$

So combining (B.18) and (B.19) we can infer (B.5).

**Lemma 3.** The system in (B.4) satisfies

$$\lim_{t \to \infty} (t+1)^{\delta_0} \nu(t) = 0 \quad \text{(B.20)}$$

$$\lim_{t \to \infty} (t+1)^{\delta_0} \omega(t) = 0 \quad \text{(B.21)}$$

where $0 \leq \delta_0 < \delta_a - \delta_b$

**Proof.** Step 1 : From Lemma-2 we have $|\nu(t)| \leq B_\nu < \infty$. Then for sufficiently large $t$, from (B.4) we have

$$|\omega(t+1)| \leq |1 - \sigma_3 b(t) + \sigma_4 a(t)||\omega(t)| + \sigma_5 a(t)B_\nu + \sigma_6 \rho^t$$

9
Now as $0 < \rho < 1$, there exists $C_\rho > 0$ and $T_\rho > 0$ such that for all $t > T_\rho$
\[
\sigma_3 B_\rho a(t) + \sigma_6 \beta^t < C_\rho a(t) \tag{B.22}
\]
For a suitable choice of $C_\rho = \sigma_{13}$, (B.22) holds for all $t > 0$.
\[\therefore \omega(t+1)| \leq |1 - \sigma_3 b(t) + \sigma_4 a(t)| \omega(t)| + \sigma_{13} a(t) \tag{B.23}\]
As (B.23) falls under the purview of Proposition-2, we can infer (B.21).
\[\text{Step 2 : As a consequence of Lemma-2, there exists } R_\omega > 0 \text{ such that } |\omega(t)| < R_\omega (t+1)^{b_0} \text{ for all } 0 \leq \delta_0 < \delta_a - \delta_b. \text{ Thus for sufficiently large } t \text{ we have -}
\]
\[|\nu(t+1)| \leq (1 - \sigma_1 a(t))|\nu(t)| + \frac{\sigma_2 a_0 R_\omega}{(t+1)^{b_0 + \delta_0}} \tag{B.24}\]
As (B.24) falls under the purview of Proposition-1, we have
\[
\lim_{t \to \infty} (t+1)^{\delta_0} \nu(t) = 0
\]
for all $0 \leq \delta_0 < \delta_a$. By making $\delta_0$ arbitrarily close to $\delta_a - \delta_b$ we get (B.20).

Let $J := I - \frac{1}{N} \mathbb{1}^T$. A bound for $\|J - \beta(t)L^\infty\|$ is obtained from the following lemma.

Lemma 4. Given $c_1 > 0$, $L^\infty = (D^{out} - A)W^\infty$ where $W^\infty = \text{diag}(w^\infty)$, $\beta(t) = \frac{\beta_0}{(1+t)^{b_1}}$, $\mu(t) = \frac{\mu_0}{(1+t)^{b_2}}$ where $0 < \beta_0 < \psi$, $\mu_0 > 0$ and $0 < \beta_1 < \mu_1 < 1$, there exists $T > 0$ such that $\|J - \beta(t)L^\infty\| \leq 1 - c_1 \mu(t) < 1$ for all $t \geq T$.

Proof. Using the property $\mathbb{1}^T L^\infty = 0$ we can write
\[
\|J - \beta(t)\| = \sup_{x \in \mathbb{R}^N} x^T (J - \beta(t)M_2 + \beta^2(t)M_3) x
\]
\[
= \lambda_{\text{max}} (J - \beta(t)M_2 + \beta^2(t)M_3) \leq \lambda_{\text{max}} (J - \beta(t)M_2 + \beta^2(t)M_3)
\]
where $M_2 = (L^\infty)^T L^\infty$ and $M_3 = (L^\infty)^T L^\infty$. Now by definition $L^\infty \mathbb{1} = 0$. So we have
\[
M_2 \mathbb{1} = 0 ; \ M_3 \mathbb{1} = 0 \tag{B.26}
\]
Also, $M_2$ and $M_3$ are the Laplacians of the corresponding undirected graph. As the graphs are connected, so we can infer :
\[
\lambda_2 (M_2) > 0 ; \ \lambda_2 (M_3) > 0
\]
i.e the 2nd lowest eigen-value of each of the Laplacians is strictly positive.

Let $x \in \text{span}\{\mathbb{1}\} \equiv x = \alpha \mathbb{1}, \alpha \in \mathbb{R}$. Then using (B.26) we have -
\[
x^T \left( I - \frac{1}{N} \mathbb{1}^T - \beta(t)M_2 + \beta^2(t)M_3 \right) x = 0 \tag{B.27}
\]
Now suppose $x \in \text{span}\{\mathbb{1}^\perp\} \equiv x^T \mathbb{1} = 0$. Then
\[
x^T \left( I - \frac{1}{N} \mathbb{1}^T - \beta(t)M_2 + \beta^2(t)M_3 \right) x
\]
\[
= x^T x - \beta(t)x^T M_2 x + \beta^2(t)x^T M_3 x
\]
\[
\leq x^T x - \beta(t) \lambda_2 (M_2) + \beta^2(t) \lambda_{\text{max}} (M_3)
\]
\[
\leq (1 - \beta(t) \lambda_m - \beta^2(t) \lambda_M) x^T x
\]
Having $\beta_0 < \psi$ and $\beta_1 < 1$ ensures that $\beta(t) < \frac{\beta_0}{\lambda_M}$, and in turn $\|I - \frac{1}{N} \mathbb{1}^T - \beta(t)L^\infty\|^2 < 1 \forall t \geq 0$. We choose an $\epsilon$ such that $0 < \epsilon < \lambda_m - \beta(t) \lambda_M$. Then for all $t \geq 0$
\[
\beta(t) \lambda_m - \beta^2(t) \lambda_M \geq \beta(t) \epsilon > 0 \tag{B.29}
\]
Then from (B.25), (B.27), (B.28) and (B.29) we have :
\[
\|J - \beta(t)L^\infty\| \leq \sqrt{1 - \beta(t) \epsilon} < 1 \tag{B.30}
\]
Now, as $\mu_1 > \beta_1$, there exists time $T > 0$ such that for all $t > T$
\[
\frac{1}{(1+t)^{\mu_1}} \leq \frac{\epsilon \beta_0}{2c_1 \mu_0} \implies 2c_1 \mu(t) \leq \epsilon \beta(t)
\]
\[
\implies 1 - \epsilon \beta(t) \leq 1 - 2c_1 \mu(t) + c_1 \mu(t)^2
\]
\[
\implies \sqrt{1 - \epsilon \beta(t)} \leq 1 - c_1 \mu(t) \tag{B.31}
\]
Also as $c_1 > 0$, we have $1 - c_1 \mu(t) < 1$. Then from (B.30) and (B.31) we have
\[
\|J - \beta(t)L^\infty\| \leq 1 - c_1 \mu(t) < 1 , t \geq T
\]

C Proof of Theorem-1

Now we present the proof of our main result, Theorem-1.

Proof. Let $p(t)$ denote the difference between the state of the agents and their average, and $q(t)$ denote the difference between the average value and $\theta^*$.
\[
p(t) := x(t) - \bar{x}(t) \mathbb{1}, p(t) \in \mathbb{R}^N, q(t) := \bar{x}(t) - \theta^*, q(t) \in \mathbb{R} \tag{C.1}
\]
\[\|x_i(t) - \theta^*\| \leq \|x_i(t) - \bar{x}(t)\| + \|\bar{x}(t) - \theta^*\| \leq \|x(t) - \bar{x}(t)\| + \|\bar{x}(t) - \theta^*\| \]  
\[\therefore \|x_i(t) - \theta^*\| \leq \|p(t)\| + \|q(t)\| \text{ for all } i \in \mathcal{V}\]

We wish to show that the states asymptotically converge to \(\theta^*\). This can be shown by expressing \(\|x_i(t) - \theta^*\| \leq \gamma(t)\) and showing that \(\gamma(t)\) has a dynamics that asymptotically converges to 0. In what follows, we use an approach where firstly we use the method of induction to show that

\[\|p(t)\| \leq \gamma_1(t) \text{ and } \|q(t)\| \leq \gamma_2(t) \text{ for all } t \geq 0 \]  

and in the process also define the dynamics of \(\gamma_1(t)\) and \(\gamma_2(t)\). After that we express these dynamics as a linear time-varying system which is asymptotically stable. From there, using (C.2), (15) and (C.3) we arrive at our desired result.

**Dynamics of** \(x(t) - \bar{x}(t)\) :

\[p(t + 1) = x(t + 1) - \bar{x}(t+1) = Jx(t+1)\]

\[\therefore p(t + 1) = M_1 + \alpha(t)JK(t)(y(t) - x(t))\]  
\[(C.4)\]

where \(J := I - \frac{1}{\|T\|} T^T\), and \(M_1 := J(I - \beta(t)L(t))x(t)\). Expanding \(M_1\) and using \(T^T L(t) = 0\) followed by \(J\|1 = 0\) we get

\[M_1 = (J - \beta(t)L(t))(x(t) - \bar{x}(t)\| - \beta(t)L(t)\|\bar{x}(t)\| = (J - \beta(t)L_\infty)\|p(t) - \beta(t)L(t) - L_\infty\| - \beta(t)L(t)\|\bar{x}(t)\|\]

\[\therefore M_1 = (J - \beta(t)L_\infty)\|p(t) - \beta(t)L(t)q(t)\| - \beta(t)L(t) - L_\infty\) - \beta(t)L(t)\theta^*\| \]  
\[(C.5)\]

Applying \(l_2\)-norm to (C.5) and using the properties of triangle-inequality and sub-multiplicativity

\[\|M_1\| \leq \|(J - \beta(t)L_\infty)\|p(t)\| + \beta(t)\|L(t)\|q(t)\| + \|(L(t) - L_\infty)\| + \|L(t)\|\theta^*\| \]

\[(C.6)\]

Now by applying Lemmas 1 and 4, and Assumption-2 in (C.6) we get

\[\|M_1\| \leq (1 - c_1\mu(t))\|p(t)\| + \beta(t)C\theta^*\|q(t)\| + \beta(t)C\theta^*\| \]

\[\therefore \|M_1\| \leq (1 - c_1\mu(t))\|p(t)\| + C\beta(t)\theta^*\|q(t)\| + C(1 + \theta)\beta(t)\theta^*\]

\[(C.7)\]

Now applying \(l_2\)-norm to (C.4) and using the properties of triangle-inequality followed by sub-multiplicativity and \(\|J\| = 1\) we get

\[\|p(t + 1)\| \leq \|M_1\| + \alpha(t)\|K(t)(y(t) - x(t))\| \]

Applying the inequality from (C.7) and the definition of \(K(t)\) from (18) to the above equation we get

\[\|p(t + 1)\| \leq (1 - c_1\mu(t))\|p(t)\| + C\beta(t)\theta^*\|q(t)\| + C\beta(t)\theta^*\|1 + \Theta\|v\alpha(t)\gamma(t)\]

\[(C.8)\]

**Dynamics of** \(x(t) - \theta^*\) :

\[q(t + 1) = x(t + 1) - \theta^*\]

\[= \bar{x}(t) - \theta^* + \frac{\alpha(t)}{N}t^T K(t)(y(t) - x(t))\]  
\[(C.9)\]

We define two diagonal matrices \(K_G(t), K_B(t)\) as

\[\|K_G(t)\|_{ij} := \begin{cases} k_i(t), & \text{if } j = i \& i \in \mathcal{B}_t \\ 0, & \text{otherwise} \end{cases} \]

\[\|K_B(t)\|_{ij} := \begin{cases} k_i(t), & \text{if } j = i \& i \in \mathcal{G}_t \\ 0, & \text{otherwise} \end{cases} \]

\[\therefore q(t + 1) = (1 - \frac{\alpha(t)}{N})t^T K_G(t)\|p(t)\| + \frac{\alpha(t)}{N}t^T K_B(t)(y(t) - x(t))\]

\[(C.10)\]

Using (C.10), (1) and (C.1) in (C.9) we get

\[q(t + 1) = q(t) + \frac{\alpha(t)}{N}t^T (K_G(t)(\theta^*\| - x(t))) + \frac{\alpha(t)}{N}(K_B(t)(y(t) - x(t)))\]

\[\therefore q(t + 1) = (1 - \frac{\alpha(t)}{N})t^T K_G(t)p(t) + \frac{\alpha(t)}{N}t^T K_B(t)(y(t) - x(t))\]

\[(C.11)\]

Applying \(l_2\)-norm to (C.11) and using the properties of triangle-inequality and sub-multiplicativity we get

\[\|q(t + 1)\| \leq \|1 - \frac{\alpha(t)}{N}\| \sum_{i \in \mathcal{V}} k_i(t)\|\theta^*\| \]

\[+ \frac{\alpha(t)}{N}\| \sum_{i \in \mathcal{G}_t} k_i(t)(x_i(t) - \bar{x}(t))\| \]

\[+ \frac{\alpha(t)}{N}\| \sum_{i \in \mathcal{B}_t} k_i(t)(y(t) - x(t))\| \]

\[(C.12)\]

**Dynamics of** \(\gamma_1(t)\) and \(\gamma_2(t)\) via method of Induction : By the method of induction we wish to show (C.3), and in the process arrive at the dynamics of \(\gamma_1(t)\) and \(\gamma_2(t)\).

**Step 1 : at** \(t = 0\),

\[\|p(0)\| = \|x(0) - \bar{x}(0)\| = 0 \quad \therefore x_i(0) = 0 \forall i \in \nu\]

\[\|q(0)\| = \|\bar{x}(0) - \theta^*\| \leq \Theta \quad \therefore \|\theta^*\| < \Theta\]
We choose $\gamma_1(0) = 0$ and $\gamma_2(0) = \Theta$, and thus we have
\[ \|p(0)\| \leq \gamma_1(0), \|q(0)\| \leq \gamma_2(0) \quad (C.13) \]

**Step 2**: for some $t > 0$ we assume that
\[ \|p(t)\| \leq \gamma_1(t), \|q(t)\| \leq \gamma_2(t) \quad (C.14) \]

**Step 3**: based on the assumption (C.14) from Step-2, we need to show that
\[ \|p(t+1)\| \leq \gamma_1(t+1), \|q(t+1)\| \leq \gamma_2(t+1) \quad (C.15) \]

Applying (C.14) to (C.8) and using (15) we have
\[
\begin{align*}
\|p(t+1)\| &\leq (1 - c_1 \mu(t)^2) \gamma_1(t) + C \beta(t) \eta^2 \gamma_2(t) \\
&\quad + C(1 + \Theta) \beta(t) \eta^2 + \sqrt{N} \alpha(t)(\gamma_1(t) + \gamma_2(t)) \\
&\quad + (C \beta(t) \eta^2 + \sqrt{N} \alpha(t)) \gamma_2(t) + C(1 + \Theta) \beta(t) \eta^2
\end{align*}
\]
\[
\therefore \|p(t+1)\| \leq (1 - c_1 \mu(t) + \sqrt{N} \alpha(t)) \gamma_1(t) \\
&\quad + (1 + \sqrt{N}) \alpha(t) \gamma_2(t) + C \beta(t) \eta^2
\]
\[ (C.16) \]

Now as $\eta < 1$, there exists $c_2 > 0$ and $T > 0$ such that for all $t > T$
\[ C(1 + \Theta) \beta(t) \eta^2 \leq c_2 \eta^2, \text{ and } C \beta(t) \eta^2 \leq \alpha(t) \quad (C.16) \]

By appropriate choice of $\beta_0 < \frac{\Omega}{2}$, $c_2 > C(1 + \Theta) \beta_0$, (C.16) holds for all $t > 0$
\[ \therefore \|p(t+1)\| \leq (1 - c_1 \mu(t) + \sqrt{N} \alpha(t)) \gamma_1(t) \\
&\quad + (1 + \sqrt{N}) \alpha(t) \gamma_2(t) + c_2 \eta^2 \quad (C.17) \]

We define the dynamics of $\gamma_1(t)$ as -
\[ \gamma_1(t+1) := (1 - c_1 \mu(t) + \sqrt{N} \alpha(t)) \gamma_1(t) \\
&\quad + (1 + \sqrt{N}) \alpha(t) \gamma_2(t) + c_2 \eta^2 \quad (C.18) \]

Using the assumption (C.14) and equations (15), (2) we have
\[ k_i(t) = 1 \text{ for all } i \in G_i \quad [\because y_i(t) = \theta^* \forall i \in G_i] \quad (C.19) \]

So from (C.12) and (C.19), and further using (15) we get
\[
\begin{align*}
\|q(t+1)\| &\leq \|1 - \frac{\alpha(t)}{N} |G_i| \|q(t)\| \\
&\quad + \frac{\alpha(t)}{N} (|G_i| \|p(t)\| + |B_i| \gamma(t)) \\
&\quad \leq \|1 - \alpha(t)(1 - s)\| \gamma_2(t) + \alpha(t)(1 - s) \gamma_1(t) \\
&\quad + \alpha(t)s(\gamma_1(t) + \gamma_2(t))
\end{align*}
\]

Now as $\alpha_0 < \frac{1}{2s}$ and $s < \frac{1}{2}$ we have
\[ \|q(t+1)\| \leq (1 - \alpha(t)(1 - 2s)) \gamma_2(t) + \alpha(t) \gamma_1(t) \quad (C.20) \]

We define the dynamics of $\gamma_2(t)$ as -
\[ \gamma_2(t+1) := \alpha(t) \gamma_1(t) + (1 - \alpha(t)(1 - 2s)) \gamma_2(t) \quad (C.21) \]

Then from (C.17), (C.18) and (C.20), (C.21) we can infer
(C.15).
Thus combining steps 1, 2 and 3 we have shown (C.3).

**Asymptotic stability of $\gamma_1(t)$ and $\gamma_2(t)$**:

Consider a linear time-varying system with state-variables $\gamma_1(t)$ and $\gamma_2(t)$, and the state dynamics given by (C.18) and (C.21) respectively.

Now using Lemma-3 we can say that the system is asymptotically stable, i.e. $\lim_{t \to \infty} (t + 1)^{\beta_0} \gamma_1(t) = 0$ and $\lim_{t \to \infty} (t + 1)^{\beta_0} \gamma_2(t) = 0$. 

\[ \square \]