CEA Comité Español de Automática
XLVI JORNADAS AUTOMÁTICA CARTAGENA 2025
UNIVERSIDADE DA CORUÑA

Jornadas de Automática

Mecanismo de enmascaramiento seguro para algoritmos de consenso en sistemas multiagente no lineales

Cecilia, Andreu[a,*], Baldomà-Mitjans, Pol[a], Puig, Vicenç[a], Costa-Castelló, Ramon[a]

[a] *Universitat Politècnica de Catalunya, Avinguda Diagonal, 647, 08028 Barcelona, España.*

To cite this article: Cecilia, Andreu, Baldomà-Mitjans, Pol, Puig, Vicenç, Costa-Castelló, Ramon 2025. Secure masking mechanism for consensus algorithms in nonlinear multiagent systems
Jornadas de Automática, 46. https://doi.org/10.17979/ja-cea.2025.46.12167

Resumen

Este trabajo presenta un protocolo de enmascaramiento diseñado para mejorar la seguridad de los protocolos de consenso en sistemas multiagente no lineales. El enfoque consiste en agregar una señal de enmascaramiento a la salida de cada agente, que luego se elimina mediante un filtro de desenmascarado en el agente receptor. En este trabajo, establecemos condiciones suficientes para garantizar que este protocolo de seguridad preserva el consenso de salida. Además, las simulaciones numéricas validan la capacidad del protocolo para prevenir ataques de escucha y de inyección de datos falsos

Palabras clave: Sistemas multi-agente, Control y estimación distribuidos, Detección de fallos e implementación de tolerancia a fallos (FDI y FTC) en sistemas en red, Detección de fallos en sistemas no lineales, Estabilidad de sistemas no lineales

Secure masking mechanism for consensus algorithms in nonlinear multiagent systems

Abstract

This paper presents a masking protocol designed to improve the security of consensus protocols for nonlinear multi-agent systems. The approach involves introducing a masking signal to each agent's output, which is then removed using a de-masking filter at the receiving agent. We provide sufficient conditions to guarantee that this security protocol maintains output consensus. Additionally, numerical simulations validate the protocol's ability to prevent eavesdropping and false data injection attacks.

Keywords: Multi-agent systems, Distributed control and estimation, FDI and FTC for networked systems, FDI for nonlinear Systems, Stability of nonlinear systems

1. Introducción

Las infraestructuras críticas, como los sistemas de energía eléctrica, las redes de distribución de agua, las redes de telecomunicaciones, los sistemas de transporte y los procesos industriales, son hoy en día sistemas a gran escala que están interconectados no sólo a nivel físico, sino también a través de toda una red cibernética asociada a la infraestructura de comunicación entre las distintas unidades de control.

Esta última red implica la presencia de vulnerabilidades debido a posibles ciberataques externos. Dado que el uso de dispositivos de baja potencia computacional en entornos interconectados dificulta la aplicación de métodos criptográficos clásicos, el control/estimación/optimización segura se ha convertido en un tema popular para dar respuesta a la necesidad del nivel físico de asegurar y preservar la información mientras se cumplen los estándares de rendimiento de la automatización industrial.

En este artículo nos centramos en el campo de los algoritmos distribuidos, específicamente en los protocolos de control de consenso, que son omnipresentes en los sistemas multiagente cuando se debe lograr un acuerdo determinado. Dado que este tipo de sistemas intercambian información a través

*Autor para correspondencia: andreu.cecilia@upc.edu
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

de redes de comunicación compartidas, son susceptibles a ciberataques. Dos amenazas muy comunes en el contexto del control seguro son los ataques de escucha y los ataques de engaño (Teixeira et al., 2015). En los últimos años han surgido múltiples mecanismos de seguridad orientados al control para superar tales eventos, entre los que se incluyen el diseño de esquemas de preservación de privacidad (Mo and Murray, 2017), mecanismos de programación de transmisiones (Leong et al., 2019; Le Wang et al., 2022) contra los espías, o métodos resilientes contra ataques de inyección de datos falsos (FDIA) (Lu and Jia, 2023). Actualmente, la mayoría de los métodos están restringidos a sistemas multiagente lineales en entornos estocásticos, con resultados limitados para sistemas no lineales. Además, en el contexto de los sistemas multiagente, la mayoría de los algoritmos tienen dificultades para garantizar el consenso, presentando un compromiso entre seguridad y rendimiento del sistema (Kawano and Cao, 2020).

Para abordar estos desafíos, este trabajo propone un algoritmo inspirado en el mecanismo de privacidad para observadores no lineales introducido en (Cecilia et al., 2023; Mitjans et al., 2025). El enfoque consiste en agregar una señal de enmascaramiento a la señal transmitida, asegurando que un espía con acceso a los datos del sensor no pueda inferir el estado interno de la planta. Además, el algoritmo tiene propiedades de convergencia demostrables que pueden aprovecharse para detectar ataques de inyección de datos falsos. Considerando esto, las principales contribuciones de este trabajo son:

- Adaptamos el mecanismo diseñado en (Cecilia et al., 2023) al caso de algoritmos de consenso distribuidos, es decir, proponemos un protocolo distribuido de enmascaramiento/desenmascaramiento;
- Mostramos que dicho mecanismo detecta con éxito los ataques de inyección de datos falsos (FDI) y previene los ataques de escucha en redes, mientras preserva el consenso. Además, demostramos que nuestros protocolos son capaces de detectar de manera separada qué comunicación ha sido atacada.

Notación

$\mathbb{R}$ es el conjunto de números reales y $\mathbb{R}_{\geq 0} = [0, \infty)$. Considerando $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. I_n es la identidad de dimensión n y $\mathbf{1}_n$ es un vector columna de dimensión n donde cada elemento es 1. $A \otimes B$ es el producto de Kronecker de las matrices A y B.

Teoría de grafos

Las conexiones entre agentes en un sistema multiagente se representan mediante un grafo no dirigido $\mathcal{G} = \mathcal{V}, \mathcal{E}, \mathcal{A}$, donde $\mathcal{V} = v_1, ..., v_N$ es el conjunto de vértices, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ es el conjunto de aristas y $\mathcal{A} \in \mathbb{R}^{N \times N}$ es la matriz de adyacencia. Cada vértice $v_k \in \mathcal{V}$ representa un agente, y existe un flujo de información del agente i al agente j si $(v_i, v_j) \in \mathcal{E}$. Si v_j envía información a v_i, entonces v_j es un vecino entrante de v_i. El conjunto de índices de los vecinos del agente i se denota como $\mathcal{N}i$. La matriz de adyacencia $\mathcal{A} = \{a_{ij}\}$ es simétrica y se define de tal forma que $a_{ij} = a_{ji} = 1$ cuando hay flujo de información de v_j a v_i, y $a_{ij} = a_{ji} = 0$ en caso contrario. El grado del nodo deg(vi) representa la suma del número de nodos que envían información al nodo i. La matriz de grados $\mathcal{D}$ del grafo $\mathcal{G}$ se define como diag(deg(v_k)). La matriz Laplaciana de $\mathcal{G}$ se define como $\mathcal{L} = \mathcal{D} - \mathcal{A}$. La matriz de incidencia firmada $E = \{e_{ij}\}$ del grafo $\mathcal{G}$ se define como $e_{ij} = 1$ cuando hay flujo de información de v_j a v_i, $e_{ij} = -1$ cuando hay flujo de información de v_i a v_j, y $e_{ij} = 0$ cuando no se intercambia información. Un camino de v_i a v_j es un conjunto de aristas que conecta v_i con v_j.

2. Formulación del Problema

Considera un sistema multi-agente homogéneo de N agentes y asociado a un grafo invariante en el tiempo $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$. Las dinámicas de cada agente se describen de la siguiente forma:

$$\dot{x}_i = f(x_i) + Bu_i, \quad y_i = Cx_i, \tag{1}$$

donde $i \in \{1, \ldots, N\}$. El factor $x_i \in \mathbb{R}^{n_x}$ es el vector de estados del agente i, $u_i \in \mathbb{R}^{n_u}$ son las entradas de control, $y_i \in \mathbb{R}^{n_y}$ son las salidas del sistema que se transmite entre los agentes. Finalmente, se asume $n_u = n_y$. El campo vectorial $f : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ es suficientemente suave y localmente Lipschitz. Cada agente cumple la siguiente propiedad de pasividad:

Suposición 1. *Existe una matriz simétrica y positiva $P \in \mathbb{R}^{n_x \times n_x}$ que cumple, por todo $x \in \mathbb{R}^{n_x}$,*

$$P\frac{\partial f}{\partial x}(x) + \frac{\partial f}{\partial x}(x)^T P \preceq 0, \qquad PB = C^T. \tag{2}$$

Como se explica en (Pavlov and Marconi, 2008; Pavlov et al., 2022), la condición (2) establece que cada agente (1) es incrementalmente pasivo, con una función de almacenamiento $V_i = (x_i - x_i')^\top P(x_i - x_i')$ para cada $i \in 1, \ldots, N$, es decir, $\dot{V}_i \leq (y_i - y_i')^\top (u_i - u_i')$, donde $x_i, x_i' \in \mathbb{R}^{n_x}$ son cualquier par de soluciones de (1) correspondientes a las entradas u_i, u_i', y y_i, y_i' son las salidas correspondientes.

En este marco, al elegir cada entrada local u_i en la forma de un acoplamiento difusivo, es decir,

$$u_i = -\kappa \sum_{j=1}^{N} a_{ij}(y_i - y_j), \tag{3}$$

donde a_{ij} son los elementos de la matriz de adyacencia $\mathcal{A}$ del grafo conexo $\mathcal{G}$, se garantiza el consenso de salida de la red para cualquier $\kappa > 0$, es decir,

$$\lim_{t \to \infty} |y_i(t) - y_j(t)| = 0, \quad \forall i, j = 1, ..., N \tag{4}$$

ver, entre otros, (Stan and Sepulchre, 2007; Zhang et al., 2014; Pavlov et al., 2022). En este artículo, para simplificar el análisis, restringimos nuestra atención a grafos no dirigidos, como se indica a continuación.

Suposición 2. *El grafo $\mathcal{G}$ es no dirigido y conectado, es decir, existe un camino que conecta cualquier pareja de vértices v_i y v_j.*

2.1. Política de ataques y objetivos

La transmisión de la señal de medición de salida entre agentes en (3) se realiza a través de una red de comunicación abierta, la cual es susceptible a ciberataques. En este trabajo, consideramos dos tipos de ataques.

Definición 1 (Ataque de escucha)**.** *Ciberataque donde el intruso accede a los datos y_{ij} que se transmite entre los nodos v_i y v_j.*

Definición 2 (Ataque de inyección de datos falsos)**.** *Un ciberataque en el cual el atacante es capaz de modificar los datos transmitidos en una arista entre dos agentes v_i y v_j, es decir, $\hat{y}_{ij} := y_{ij} + a$, donde a es la señal de ataque. En este trabajo, también consideramos un ataque FDI más fuerte, en el que el atacante además bloquea la comunicación entre los agentes, es decir, $a = -y_{ij} + \bar{a}$, donde $\bar{a}$ es la señal de ataque.*

El objetivo de este trabajo es diseñar un mecanismo de seguridad que pueda utilizarse para prevenir ataques de escucha y detectar ataques FDI. Para el diseño de este mecanismo, se han impuesto las siguientes restricciones:

1. **Consenso**: el mecanismo debe preservar el consenso de salida, es decir, debe cumplirse la condición (4);
2. **Modularidad**: el mecanismo debe poder añadirse o eliminarse sin modificar el esquema subyacente del sistema multiagente. Más precisamente, se asumirá que ni el protocolo de control (3), ni la función f, ni las matrices B y C pueden ser modificadas;
3. **Restricción de recursos**: no se pueden añadir canales adicionales al esquema de comunicación. Esto impide el uso de técnicas de codificación/decodificación similares a las propuestas, por ejemplo, en (Joo et al., 2021);
4. **Asincronía**: los puntos extremos de comunicación no pueden contener elementos que se asuman sincronizados. Esto impide, por ejemplo, el uso de algoritmos de generación de seudoaleatoriedad sincronizados entre nodos de comunicación, como en (Li et al., 2023).

3. Propuesta

La principal propuesta de este trabajo consiste en añadir un generador de enmascaramiento en la salida de cada agente, transmitir una señal enmascarada (y segura) a través de la red, y utilizar un filtro en el agente receptor para eliminar dicho enmascaramiento. Este mecanismo se ilustra en la Figura 1.

3.1. Protocolo de enmascaramiento

En particular, la salida de cada agente, es decir, la señal transmitida entre agentes, se modifica de la siguiente manera:

$$\hat{y}i = y_i + \gamma(\Gamma_i \omega_i), \tag{5}$$

donde $\Gamma_i \in \mathbb{R}^{n_y \times 2n\omega i}$ es la matriz de salida de la señal de enmascaramiento del agente i, con $i \in 1, \ldots, N$, $\gamma : \mathbb{R}^{n_y} \to \mathbb{R}^{n_y}$ es una función no lineal continua y monótona que satisface

$$(a-b)^T(a-b) \leq (a-b)^T(\gamma(a) - \gamma(b)), \quad \forall, a, b \in \mathbb{R}^{n_y}, \tag{6}$$

y $\omega_i \in \mathbb{R}^{2n_{\omega i}}$ es el estado interno del generador de la señal de enmascaramiento del agente i, que evoluciona según un sistema exógeno de la forma

$$\dot{\omega}_i = \Phi_i \omega_i, \qquad i \in 1, \ldots, N. \tag{7}$$

Las matrices Φ_i y Γ_i se definen como sigue:

$$\begin{aligned} \Phi_i &= I_{n_y} \otimes S_i, & \Gamma_i &= I_{n_y} \otimes G_i, \\ S_i &= \text{blckdiag}(S_{i,1}, \ldots, S_{i,n_{\omega i}}), & G_i &= \text{col}(\underbrace{g, \ldots, g}_{n_{wi} \text{ times}}), \\ S_{i,j} &= \begin{pmatrix} 0 & \omega_{i,j} \\ -\omega_{i,j} & 0 \end{pmatrix}, & g &= \begin{pmatrix} 1 & 0 \end{pmatrix}. \end{aligned} \tag{8}$$

Por construcción, las matrices Φ_i son matrices antisimétricas, es decir,

$$\Phi_i + \Phi_i^T = 0, \qquad \forall i \in 1, \ldots, N. \tag{9}$$

Los estados internos de enmascaramiento de cada agente consisten en un conjunto desacoplado de $n_{\omega i}$ osciladores con frecuencias $\omega_{i,j}$. Las $n_{\omega i}$ señales oscilatorias resultantes se combinan mediante la matriz Γ_i, generando una señal escalar para cada una de las n_y salidas de (1).

3.2. Protocolo de des-enmascaramiento

Con el fin de desenmascarar la señal para el controlador (3), se incorpora un filtro en el extremo de la comunicación, el cual elimina todas las señales de enmascaramiento recibidas de los vecinos del agente. Este filtro se retroalimenta con la salida del propio agente. Esta acción de retroalimentación garantiza la sincronización asintótica entre el nodo enmascarador y el nodo desenmascarador, sin necesidad de canales adicionales.

Más precisamente, recordando que $\mathcal{N}_i$ es el conjunto de vecinos del agente i, el filtro requerido para el agente i se define como

$$\begin{aligned} \dot{\xi}_i &= \tilde{\Phi}_i \xi_i - \tilde{\Gamma}_i^T z_i, \\ z_i &= -\hat{d}_i + \nu_i, \\ \hat{d}i &= \gamma(\tilde{\Gamma} i^o \xi_i). \end{aligned} \tag{10}$$

donde el estado interno del filtro $\xi_i = \text{col}(\{\xi_{ki}\}_{k \in \mathcal{N}_i})$ representa todas las estimaciones que el agente i realiza sobre los estados internos de enmascaramiento de sus vecinos, y las matrices del filtro se definen como

$$\begin{aligned} \tilde{\Phi}_i &= \text{blckdiag}(\{\Phi_k\}_{k \in \mathcal{N}_i}), \quad \tilde{\Gamma}_i = \text{col}(\{\Gamma_k\}_{k \in \mathcal{N}_i}), \\ \tilde{\Gamma}_i^o &= \text{blckdiag}(\{\Gamma_k\}_{k \in \mathcal{N}_i}). \end{aligned} \tag{11}$$

La señal z_i es la señal de entrada filtrada que se utiliza en el control de sincronización (3). Finalmente, las entradas ν_i de los filtros son

$$\nu_i = -\sum_{j=1}^{N} a_{ij}(y_i - \hat{y}_j) = d_i - \sum_{j=1}^{N} a_{ij} C(x_i - x_j) \tag{12}$$

donde d_i se define como $d_i = \sum_{j=1}^{N} a_{ij} \gamma(\Gamma_j \omega_j)$. Las señales de salida del filtro se utilizan en la ley de control del agente i, que queda entonces definida como

$$u_i = \kappa z_i = \kappa\Big[(d_i - \hat{d}_i) - \sum_{j=1}^{N} a_{ij} C(x_i - x_j)\Big] \tag{13}$$

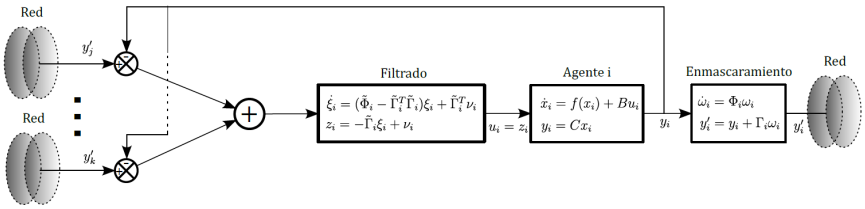

Figura 1: Esquema del protocolo de enmascaramiento.

Dadas estas descripciones, en las siguientes secciones se presentarán las condiciones necesarias para que los agentes del sistema multiagente alcancen consenso (4).

Nota 1. *Cabe destacar que la elección de utilizar distintos protocolos de enmascaramiento entre los agentes, es decir, diferentes frecuencias $\omega_{i,j}$ y matrices de salida Γ_i, está motivada por los puntos 3) y 4) de nuestros objetivos. Además, esta elección facilita la identificación del origen de un posible ataque FDI. Es decir, al contar con filtros distintos, cada agente es capaz de identificar qué canal ha sido afectado.*

3.3. Protocolo de consenso modificado

En esta sección presentamos el resultado principal de este artículo, demostrando que el controlador modificado de cada agente puede eliminar la señal d_i generada por los enmascaramientos de sus vecinos entrantes y garantizar la condición (4). Este resultado requiere una condición de detectabilidad sobre un sistema extendido que considera tanto la dinámica del agente (1) como la dinámica del filtro (10).

Suposición 3. *Considera que f, C están definidos como en* (1) *y $(\tilde{\Phi}_i, \tilde{\Gamma}_i^o)$ como en* (11)*. El sistema extendido*

$$\begin{pmatrix}\dot{x}\\ \dot{\xi}\end{pmatrix} = \begin{pmatrix}f(x)\\ \tilde{\Phi}_i\xi\end{pmatrix}, \qquad y = \begin{pmatrix}C & \tilde{\Gamma}_i^o\end{pmatrix}\begin{pmatrix}x\\ \xi\end{pmatrix}, \tag{14}$$

es asintóticamente detectable para cada $i \in 1, \dots, N$. Es decir, para todo par de soluciones $(x(t), \xi(t))$, $(x'(t), \xi'(t))$ del sistema (14) *y sus correspondientes trayectorias de salida $y(t), y'(t)$, se cumple que*

$$y(t) = y'(t) \;\; \forall t \geq 0 \implies \lim_{t\to\infty} \left\| \begin{bmatrix} x(t) \\ \xi(t) \end{bmatrix} - \begin{bmatrix} x'(t) \\ \xi'(t) \end{bmatrix} \right\| = 0.$$

La detectabilidad en el sistema extendido garantiza que el protocolo de enmascaramiento/desenmascaramiento propuesto no oculta información necesaria para lograr el consenso. La combinación de la condición de pasividad con la detectabilidad asintótica es suficiente para probar que dos trayectorias de estado —una formada por un agente y su señal de enmascaramiento, y la otra por otro agente vecino y su filtro— convergerán eventualmente. Bajo la Suposición 2, esta propiedad se cumple para todos los agentes del grafo, logrando así consenso en la salida. El siguiente teorema establece el resultado principal del artículo.

Teorema 1. *Considere la red $\mathcal{G}$ de un sistema multi-agente homogéneo formada por N agentes con dinámica* (1)*, y asume que se cumplen las Suposiciones 1, 2 y 3. Entonces, el sistema en lazo cerrado con la ley de retroalimentación* (7)*,* (10) *y* (13) *garantiza sincronización de estados, es decir, para todo i, j,*

$$\lim_{t\to\infty} |x_i(t) - x_j(t)| = 0. \tag{15}$$

Demostración. Las leyes de control del agente i, con la inclusión del enmascaramiento y los filtros, se convierten en (13). La ley de control global del sistema multi-agente se expresa como

$$u = -\kappa(\mathcal{L} \otimes C)x + \kappa\tilde{d}, \tag{16}$$

donde $\mathcal{L}$ es la matriz Laplaciana del grafo $\mathcal{G}$, y

$$u := (u_1, \dots, u_N), \quad x := (x_1, \dots, x_N),$$
$$\tilde{d} := (d_1 - \hat{d}_1, \dots, d_N - \hat{d}_N).$$

Considerando la dinámica del sistema multi-agente completo con la ley de control modificada (16), se obtiene

$$\dot{x} = F(x) - \kappa(\mathcal{L} \otimes BC)x + \kappa(I_N \otimes B)\tilde{d}, \tag{17}$$

con $F(x) = (f(x_1), \dots, f(x_N))$. Ahora, definimos la dinámica del error en los bordes de x como

$$\tilde{x} := (E^T \otimes I_{n_x})x, \tag{18}$$

donde $E \in \mathbb{R}^{N\times M}$ es la matriz de incidencia con signo del grafo $\mathcal{G}$, y M es el número de aristas del grafo $\mathcal{G}$, o cardinalidad de $\mathcal{E}$.

Recordando que $\mathcal{L} = EE^T$, la dinámica del error en los bordes se define como

$$\begin{aligned}\dot{\tilde{x}} &= (E^T \otimes I_{n_x})F(x) - \kappa(E^T\mathcal{L} \otimes BC)x + \kappa(E^T \otimes B)\tilde{d}\\ &= \tilde{F}(x) - \kappa(E^TE \otimes BC)\tilde{x} + \kappa(E^T \otimes B)\tilde{d},\end{aligned} \tag{19}$$

donde se usó $(E^T\mathcal{L} \otimes BC) = (E^TE \otimes BC)(E^T \otimes I_n)$ y se definió $\tilde{F}(x) = (E^T \otimes I_{n_x})F(x)$.

A continuación, definimos el error de estimación del estado de la señal de enmascaramiento como $\tilde{\xi}_i := \Omega_i - \xi_i$, con dinámica

$$\dot{\tilde{\xi}}_i = \tilde{\Phi}_i\tilde{\xi}_i - \tilde{\Gamma}_i^T\tilde{d}_i - \tilde{\Gamma}_i^T(\mathbf{1}_{M_i}^T \otimes C)\tilde{x}_i \tag{20}$$

donde M_i es la cardinalidad de $\mathcal{N}_i$ y $\Omega_i := \text{col}(\{\xi_k\}_{k\in\mathcal{N}_i})$, donde $\tilde{x}_i$ es un vector fila que contiene el error de estado entre el agente i y sus vecinos entrantes.

Ahora construimos la función de Lyapunov:

$$V = V_{\tilde{x}} + \sum_{i=1}^{N} V_{\tilde{\xi}_i} = \frac{1}{2}\tilde{x}^T(I_M \otimes P)\tilde{x} + \sum_{i=1}^{N}\frac{1}{2}\kappa\tilde{\xi}_i^T\tilde{\xi}_i, \tag{21}$$

donde P cumple (2). La derivada de $V_{\tilde{x}}$ es:

$$\begin{aligned}\dot{V}_{\tilde{x}} = {}& \tilde{x}^T(I_M \otimes P)\tilde{F}(x) - \kappa\tilde{x}^T(E^TE \otimes PBC)\tilde{x}\\ &+ \kappa\tilde{x}^T(E^T \otimes PB)\tilde{d}.\end{aligned} \tag{22}$$

Usando el teorema fundamental del cálculo, para cualquier par $x_i, x_j \in \mathbb{R}^{n_x}$:

$$f(x_j) - f(x_i) = \left(\int_0^1 \frac{\partial f}{\partial x} \Big(s x_j + (1-s) x_i \Big) ds \right) (x_j - x_i).$$

En consecuencia, se tiene que para todos $x_i, x_j \in \mathbb{R}^{n_x}$ y $\tilde{x}_{ji} = x_j - x_i$, donde $(i, j) \in \mathcal{E}$:

$$\begin{aligned} &\tilde{x}_{ji}^T P(f(x_j) - f(x_i)) = \\ &\tilde{x}_{ji}^T \left(\int_0^1 \frac{1}{2} \left(P \frac{\partial f}{\partial x}(*) + \frac{\partial f}{\partial x}^T (*) P \right) ds \right) \tilde{x}_{ji} \leq 0 \end{aligned}$$

donde la última desigualdad se deriva de (2). Aplicando esta desigualdad recursivamente a cada componente del primer término de la derivada de la función de Lyapunov (22):

$$\tilde{x}^T (I_M \otimes P) \tilde{F}(x) \leq 0.$$

Además, recordando de (2) que $PB = C^T$, obtenemos:

$$\begin{aligned} \dot{V}_{\tilde{x}} &\leq -\kappa \tilde{x}^T (E^T E \otimes C^T C) \tilde{x} + \kappa \tilde{x}^T (E^T \otimes C^T) \tilde{d} \\ &\leq -\kappa \tilde{x}^T (E^T \otimes C^T)(E \otimes C) \tilde{x} + \kappa \tilde{x}^T (E^T \otimes C^T) \tilde{d}. \end{aligned}$$

La derivada de cada elemento ξ_i del segundo término en el lado derecho de (21), siguiendo la dinámica (20) y combinando con (9), da:

$$\begin{aligned} \dot{V}_{\tilde{\xi}_i} &= \frac{\kappa}{2} \tilde{\xi}_i^T (\tilde{\Phi}_i + \tilde{\Phi}_i^T) \tilde{\xi}_i - \kappa \tilde{\xi}_i^T \tilde{\Gamma}_i^T \tilde{d}_i - \kappa \tilde{\xi}_i^T \tilde{\Gamma}_i^T (\mathbf{1}_{M_i}^T \otimes C) \tilde{x}_i \\ &= -\kappa (\tilde{\Gamma}_i \tilde{\xi}_i)^T \tilde{d}_i - \kappa (\tilde{\Gamma}_i \tilde{\xi}_i)^T (\mathbf{1}_{M_i}^T \otimes C) \tilde{x}_i. \end{aligned}$$

Nota que el término $(\mathbf{1}_{M_i}^T \otimes C)\tilde{x}_i$ es igual a:

$$(\mathbf{1}_{M_i}^T \otimes C) \tilde{x}_i = \sum_j^{\mathcal{N}_i} C(x_j - x_i) = (l_i \otimes C) x$$

donde l_i es la i-ésima fila de la Laplaciana $\mathcal{L}$ del grafo. Entonces, la suma completa de las derivadas de V_{ξ_i} es:

$$\begin{aligned} \sum_{i=1}^N \dot{V}_{\tilde{\xi}_i} &= -\kappa \sum_{i=1}^N (\tilde{\Gamma}_i \tilde{\xi}_i)^T \tilde{d}_i - \kappa \sum_{i=1}^N (\tilde{\Gamma}_i \tilde{\xi}_i)^T (\mathbf{1}_{M_i}^T \otimes C) \tilde{x}_i \\ &= -\kappa \Delta^T \tilde{d} - \kappa \Delta^T (E \otimes C) \tilde{x} \end{aligned}$$

con $\Delta = (\tilde{\Gamma}_1 \tilde{\xi}_1, \ldots, \tilde{\Gamma}_N \tilde{\xi}_N)$. Entonces:

$$\begin{aligned} \dot{V} = &-\kappa \tilde{x}^T (E^T \otimes C^T)(E \otimes C) \tilde{x} + \kappa \tilde{x}^T (E^T \otimes C^T) \tilde{d} \\ &- \kappa \Delta^T \tilde{d} - \kappa \Delta^T (E \otimes C) \tilde{x} \end{aligned}$$

La propiedad de monotonía de la función γ en (6) implica que:

$$-\Delta^T \tilde{d} \leq -\Delta^T \Delta,$$

lo que lleva a:

$$\begin{aligned} \dot{V} \leq &-\kappa \tilde{x}^T (E^T \otimes C^T)(E \otimes C) \tilde{x} + \kappa \tilde{x}^T (E^T \otimes C^T) \tilde{d} \\ &- \kappa \Delta^T \Delta - \kappa \Delta^T (E \otimes C) \tilde{x} \\ = &-\kappa \left((E \otimes C) \tilde{x} - \Delta \right)^T \left((E \otimes C) \tilde{x} - \Delta \right) \\ = &-\kappa \Big((\mathcal{L} \otimes I_{n_y}) y - \Delta \Big)^T \Big((\mathcal{L} \otimes I_{n_y}) y - \Delta \Big) \end{aligned}$$

Este resultado implica que el sistema converge al conjunto:

$$\mathcal{M} := \{ y, \Delta \in \mathbb{R}^{N n_y} : (\mathcal{L} \otimes I_{n_y}) y - \Delta = 0 \}.$$

Dentro de este conjunto y considerando (16), se tiene que $u = 0$, por lo tanto, la dinámica resultante del sistema multiagente es:

$$\dot{x} = F(x), \quad \dot{\tilde{\xi}}_i = \tilde{\Phi}_i \tilde{\xi}_i,$$

para todo $i = 1, ..., N$. Por lo tanto, bajo la Suposición 3,

$$\lim_{t \to \infty} \left| (\mathcal{L} \otimes I_{n_y}) y - \Delta \right| = 0 \implies \lim_{t \to \infty} \eta(t) = 0,$$

siendo $\eta = \begin{pmatrix} \tilde{x}^T & \tilde{\xi}^T \end{pmatrix}^T$, lo que significa que el estado de error entre cualquier par de trayectorias de estado de vecinos en $\mathcal{G}$ converge a 0. Esto implica que las trayectorias de estado de dos agentes vecinos convergen asintóticamente entre sí, lo que implica (15) mediante la Suposición 2. □

3.4. *Seguridad del protocolo*

El protocolo de seguridad opera añadiendo una señal de enmascaramiento a las señales de comunicación originales entre los agentes. Cada agente cuenta con un generador de enmascaramiento único, conocido únicamente por él mismo y por sus vecinos salientes (aquellos que reciben su información). Similar al caso del observador (Cecilia et al., 2023), si un agente malicioso intenta realizar un ataque de espionaje sin conocer el generador de enmascaramiento, no podrá recuperar la información verdadera y_i a partir de la señal enmascarada. Además, si se intenta un ataque por inyección de datos falsos y la señal modificada no coincide con los estados del filtro, los filtros de desenmascaramiento detectarán la perturbación.

4. Caso Práctico

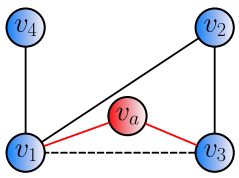

Figura 2: Topología de la red en el caso práctico.

Consideramos un sistema de $N = 4$ agentes, que están conectados como se muestra en la Figura 2. La dinámica de cada agente ubicado en el vértice v_i es

$$\begin{aligned} \dot{x}_i &= \begin{pmatrix} x_{i2} + x_{i3} \\ -x_{i1} \\ -x_{i3} + \arctan(x_{i3}) + x_{i1} \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} u_i \\ y_i &= \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} x_i + r_i \end{aligned} \tag{23}$$

siendo $x_i = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} \end{pmatrix}^T$, $u_i \in \mathbb{R}$ y $y_i \in \mathbb{R}$ el estado, la entrada de control y la salida medida del agente v_i, respectivamente, y r_i es ruido gaussiano con covarianza 0,0001. Las dinámicas (23) son pasivas según la Suposición 1, con $P = I_3$, y asintóticamente detectables según la Suposición 3. Las leyes de control se diseñan como en (3), con $\kappa = 1$.

El agente rojo v_a mostrado en el esquema de la red de la Figura 2 representa un agente adversario, que tiene conocimiento de la planta. El objetivo del atacante es infiltrarse en la dinámica de la red realizando un ataque de tipo *Man in the Middle*, que es una combinación de ataque de escucha y ataque de inyección de datos falsos. El procedimiento para llevar a cabo dicho ataque es el siguiente:

- Interceptar los datos enviados de v_3 a v_1 a partir de $t = 40$ s, de modo que el agente v_a pueda sincronizarse con el consenso promedio de toda la red (Espionaje).
- Denegar el enlace de comunicación de v_3 a v_1 y sustituirlo por la ruta $((v_3, v_a), (v_a, v_1))$ en $t = 100$ s (Inyección de datos falsos).

El sistema de enmascaramiento ha sido diseñado para cada agente con la matriz Φ_i elegida según (8) con 3 frecuencias para cada agente. Específicamente, para el agente i, $\omega_i = k_i, k_i\sqrt{2}, k_i\sqrt{3}$, con $k_1 = 5$, $k_2 = 6$, $k_3 = 7$ y $k_4 = 8$. La función monótona utilizada para la generación final de la señal es $\gamma(s) = \arctan\left(\frac{1}{3}s^3 + \frac{1}{2}s^2 + s\right)$. En la Figura 3, presentamos las salidas (antes del enmascaramiento) de cada agente del sistema con enmascaramiento. Podemos observar que el protocolo de enmascaramiento no impide el consenso entre los agentes.

Luego, en la Figura 4 mostramos el comportamiento de un atacante cuando está realizando un ataque de espionaje. Cuando se implementa el enmascaramiento, el atacante no puede recuperar la información original y no logra la sincronización.

En la última prueba, el atacante, después de intentar sincronizarse, intenta inyectar sus propios datos (sin enmascaramiento) en la red. La Figura 5 muestra las señales internas del filtro de desenmascaramiento sin y con el ataque FDI. En la Figura 5, podemos ver que el filtro es capaz de detectar qué señal está enviando datos falsos en el tiempo 100.

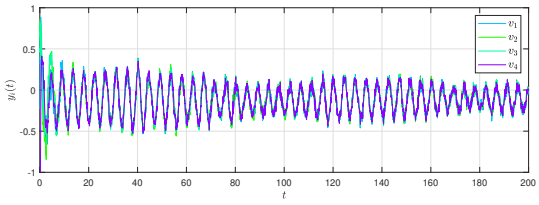

Figura 3: Salidas y_i del sistema con enmascaramiento.

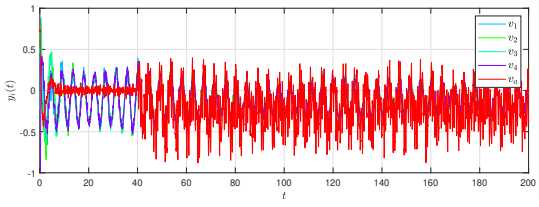

Figura 4: Salidas y_i dels sistema en un ataque de escucha.

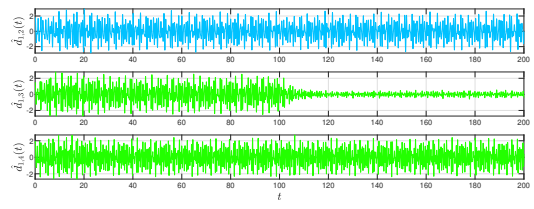

Figura 5: Señales de error en un ataque de inyección de datos.

5. Conclusiones

Este trabajo propone un mecanismo de enmascaramiento para algoritmos de consenso en sistemas multi-agente no lineales. La convergencia de la arquitectura ha sido formalmente demostrada y validada mediante simulaciones numéricas. Sin embargo, observamos que la incorporación de este mecanismo afecta el desempeño del sistema para lograr el consenso de salida, modificando la superficie de sincronización. Una dirección prometedora para investigaciones futuras es analizar la relación entre la privacidad de los datos y la variedad de sincronización.

Agradecimientos

This publication is part of the project ACROBA, financed by European Union NextGeneration-EU, the Recovery Plan, Transformation and Resilience, through INCIBE. This work is part of the Project MAFALDA (PID2021-126001OB-C31) funded by MCIN/ AEI /10.13039/501100011033 and by "ERDF A way of making Europe".

Referencias

Cecilia, A., Astolfi, D., Casadei, G., Costa-Castelló, R., Nešić, D., 2023. A masking protocol for private communication and attack detection in nonlinear observers. In: 62nd IEEE Conference on Decision and Control. pp. 7495–7500.

Joo, Y., Qu, Z., Namerikawa, T., 2021. Resilient control of cyber-physical system using nonlinear encoding signal against system integrity attacks. IEEE Transactions on Automatic Control 66 (9), 4334–4341.

Kawano, Y., Cao, M., 2020. Design of privacy-preserving dynamic controllers. IEEE Transactions on Automatic Control 65 (9), 3863–3878.

Le Wang, Cao, X., Zhang, H., Sun, C., Zheng, W. X., 2022. Transmission scheduling for privacy-optimal encryption against eavesdropping attacks on remote state estimation. Automatica 137, 110145.

Leong, A. S., Quevedo, D. E., Dolz, D., Dey, S., 2019. Transmission scheduling for remote state estimation over packet dropping links in the presence of an eavesdropper. IEEE Transactions on Automatic Control 64 (9), 3732–3739.

Li, T., Wang, Z., Zou, L., Chen, B., Yu, L., 2023. A dynamic encryption–decryption scheme for replay attack detection in cyber–physical systems. Automatica 151, 110926.

Lu, X., Jia, Y., 2023. Bipartite byzantine–resilient event–triggered consensus control of heterogeneous multi–agent systems. International Journal of Robust and Nonlinear Control 33 (1), 282–310.

Mitjans, P. B., Cecilia, A., Castelló, R. C., 2025. Protocolo de enmascaramiento para observador en pila de combustible. Revista Iberoamericana de Automática e Informática industrial 22 (2), 104–111.

Mo, Y., Murray, R. M., 2017. Privacy preserving average consensus. IEEE Transactions on Automatic Control 62 (2), 753–765.

Pavlov, A., Marconi, L., 2008. Incremental passivity and output regulation. Systems & Control Letters 57 (5), 400–409.

Pavlov, A., Steur, E., van de Wouw, N., 2022. Nonlinear integral coupling for synchronization in networks of nonlinear systems. Automatica 140, 110202.

Stan, G.-B., Sepulchre, R., 2007. Analysis of interconnected oscillators by dissipativity theory. IEEE Transactions on Automatic Control 52 (2), 256–270.

Teixeira, A., Shames, I., Sandberg, H., Johansson, K. H., 2015. A secure control framework for resource-limited adversaries. Automatica 51, 135–148.

Zhang, F., Trentelman, H. L., Scherpen, J. M., 2014. Fully distributed robust synchronization of networked lur'e systems with incremental nonlinearities. Automatica 50 (10), 2515–2526.