同源传感器攻击下的多智能体系统分布式安全状态估计

工业设备中的网络攻击

- 2021年，美国石油管道公司受勒索软件的攻击
- 2011年，伊朗攻击GPS以劫持美国RQ-170无人机
- 2015年，乌克兰电网遭受攻击

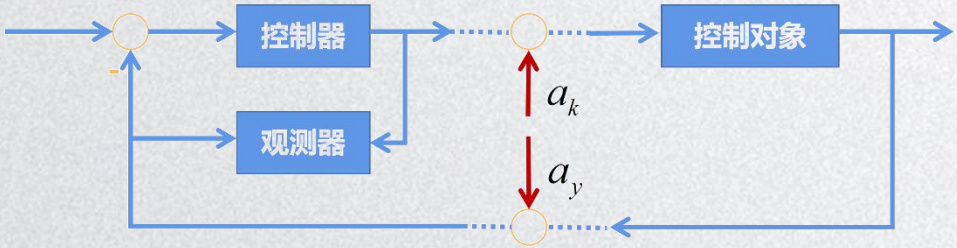
重大危害和频发事故，凸显控制系统抗攻击的必要性！



工业设备中的网络攻击攻击

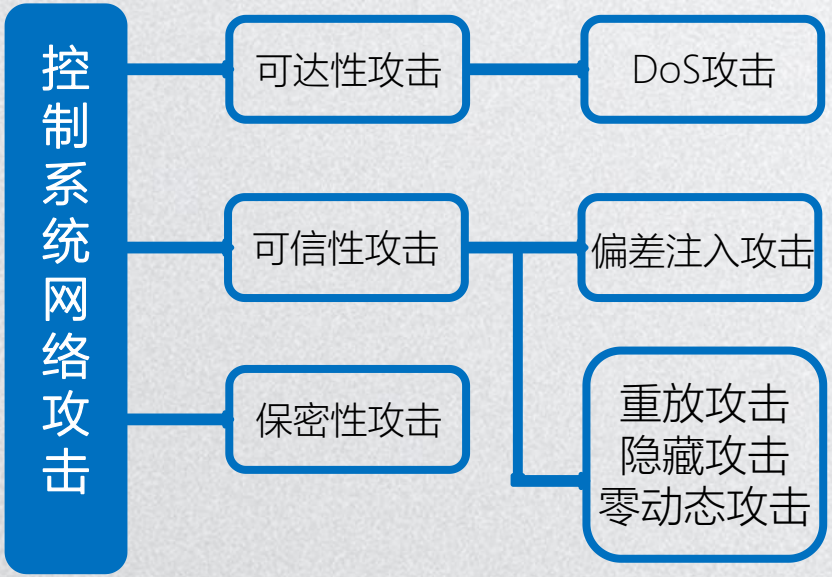
控制系统网络攻击

- 网络攻击会干扰工业控制系统，进而影响社会稳定。
- 数字化控制系统依赖计算机技术，因而无法绝对安全。
- 需要从控制科学的角度提供相应的解决方案。



受攻击的工业控制系统

网络攻击的分类



- 可达性攻击使得系统无法进行正常的网络通信
- 可信性攻击着力于修改通信数据
- 攻击数据也可以避免观测器或检测器辨别

安全状态估计

- 从被错误数据攻击的输出中重构系统状态

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + a(k)$$

- 对传统单一系统来说，攻击信号的自由度过大，会导致系统完全无法观测
- 因此攻击必须是稀疏的，即部分攻击维度为零
- 找出未受攻击的维度可以利用暴力搜索或优化算法

[1] Y. Shoukry and P. Tabuada, “Event-triggered state observers for sparse sensor noise/attacks," IEEE Transactions on Automatic Control, vol. 61, no. 8, pp. 2079-2091, Aug. 2016.

多智能体系统与分布式一致性

- 多智能体一致性：不同智能体趋于相同的状态
- 在多智能体协同中，一致性占据着重要地位
- 一致性可以实现传感器融合，从而得到协同状态估计

$$\hat{x}_i(k+1) = A\hat{x}_i(k) + M_i(C_i\hat{x}_i(k) - y_i(k)) + \sum_{j\in\mathcal{N}_i} N_{ij}(C_{ij}\hat{x}_j(k-\tau_{ij}(k)) - C_{ij}\hat{x}_i(k-\tau_{ij}(k)))$$

传感器网络中常见的分布式观测器算法

[2] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," Proceedings of the IEEE, vol. 95, no. 1, pp. 215-233, Jan. 2007.

多智能体系统中的同源传感器攻击

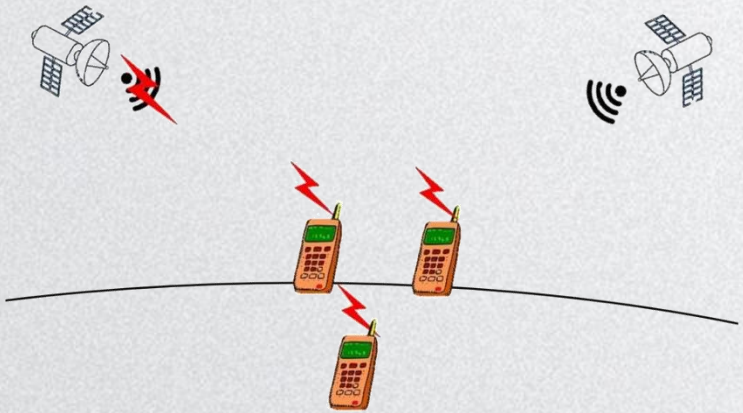
同源攻击：来源于相同攻击源，对系统中不同智能体产生相关影响的攻击信号。

$$x_i(k+1) = A_i x_i(k) + B_i u_i(k)$$

$$y_i(k) = C_i x_i(k) + a(k)$$

对GPS系统的攻击是一种典型的同源攻击。

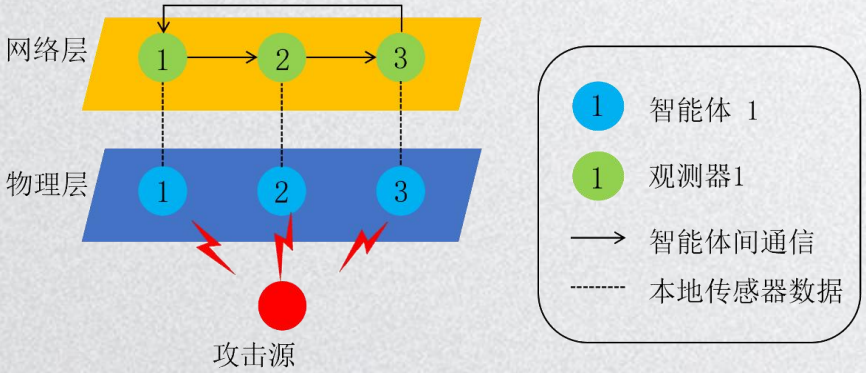
GPS系统中的同源攻击



GPS攻击示意图

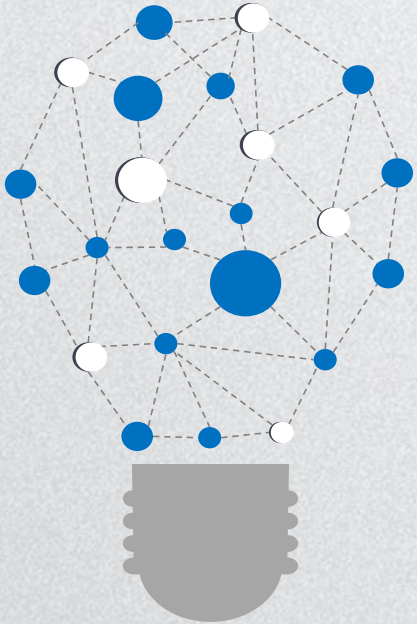
当GPS卫星信号受到攻击时，相邻区域内的接收器会产生相同的定位偏移。

分布式安全状态估计的系统结构



观测器系统图

- 系统中存在一个攻击源对智能体产生影响
- 每个智能体对应一个观测器来估计自身状态和攻击信号
- 观测器之间可以相互通信以形成协同估计



$$x_i(k+1) = A_i x_i(k) + B_i u_i(k)$$

$$y_i(k) = C_i x_i(k) + a(k)$$

系统同源攻击下完全能观

已知一段有限时间窗内各智能体的输入和输出，能唯一的确定各智能体的初始状态和攻击信号

[3] Y. Shi, C. Liu, and Y. Wang, “Secure state estimation of multiagent systems with homologous attacks using average consensus,” IEEE Transactions on Control of Network Systems, vol. 8, no. 3, pp. 1293–1303, 2021.

双环离线观测器设计

Algorithm 1: The Design of Two-loop Observer.

Initialize $m = 0, n = 0, \hat{z}^{(0,n^{(0)})} = 0$
While $m < M$ **Do**
  Do equation (15)
  While $n < n^{(m+1)}$ **Do**
    Do equation (16)
    Do $\hat{x}^{i(m+1,n+1)} = \hat{x}^{i(m+1,n)}$ for all i
    $n = n + 1$
  Endwhile
  $m = m + 1, n = 0$
Endwhile

- 利用离线时间窗内的输出数据重构状态
- 将一致性应用为内环，残差更新为外环
- 估计误差随观测器的运行次数逐渐收敛至零

[3] Y. Shi, C. Liu, and Y. Wang, “Secure state estimation of multiagent systems with homologous attacks using average consensus,” IEEE Transactions on Control of Network Systems, vol. 8, no. 3, pp. 1293–1303, 2021.

双环离线通用框架设计

Algorithm 2: The Design of Framework.

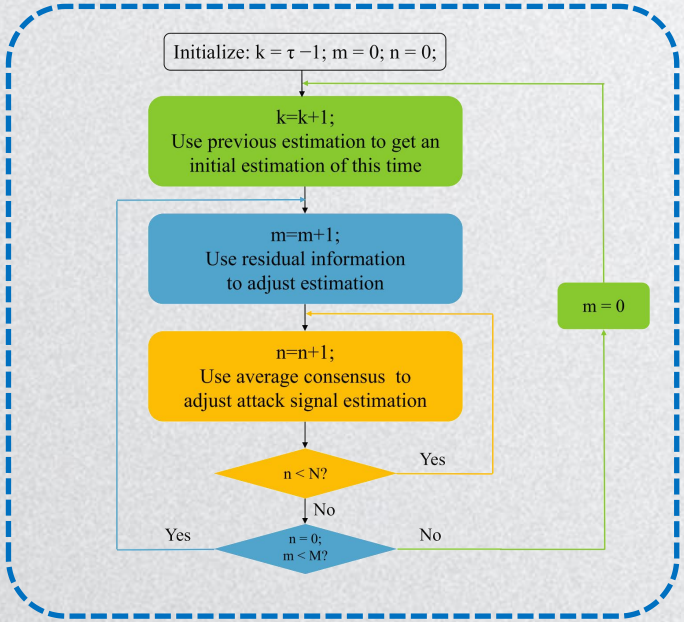
Initialize $m = 0, \hat{z}^{(0)} = 0$
While $m < M$
  Do equation (47)
  Do equation (48)
  $m = m + 1$
Endwhile

- 可以兼容多数一致性算法
- 估计误差随观测器的运行次数逐渐收敛至有界值

然而该研究仍停留在离线应用场景。

[3] Y. Shi, C. Liu, and Y. Wang, “Secure state estimation of multiagent systems with homologous attacks using average consensus,” IEEE Transactions on Control of Network Systems, vol. 8, no. 3, pp. 1293–1303, 2021.

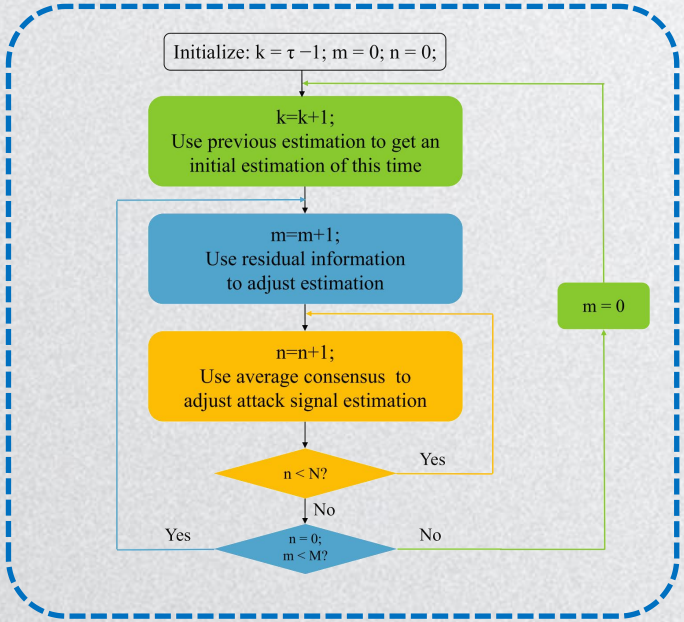
三环在线观测器设计



- 在静态观测器的基础上叠加时间更新步骤作为外环
- 每个循环的循环次数是固定的，相关证明可计算得出保证观测器收敛的循环次数。
- 估计误差随时间逐渐收敛至零。

[4] Y. Shi and Y. Wang, “Online secure state estimation of multiagent systems using average consensus," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 5, pp. 3174-3186, May 2022.

三环在线观测器设计



然而三观观测器具备如下不足：

- 循环层数过多，计算量消耗大，设计复杂。
- 同时上述的具有多环结构的观测器着重于分析循环次数而没有分析观测器增益的设计，因此观测器效率低下。

[4] Y. Shi and Y. Wang, “Online secure state estimation of multiagent systems using average consensus," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 5, pp. 3174-3186, May 2022.

最新研究

- 设计计算量小，通信量小的无循环的观测器
- 能够在复杂网络环境中保持较高的性能

解决方法

- 通过调节观测器增益以提高效率
- 针对通信网络中存在和不存在通信延迟两种情况分别设计观测器。（设计独立的无延迟观测器能够在合适的情况下提供更好的适配）

通信网络不存在延迟的观测器

$$x_i(k+1) = A_i x_i(k) + B_i u_i(k)$$

$$y_i(k) = C_i x_i(k) + a(k)$$

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\mathcal{N}_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

通信网络不存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\aleph_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

- 步骤一、二、三分别得到对攻击信号的初步估计、对攻击信号的最终估计和对状态的估计 。

通信网络不存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\aleph_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

- 由于时间窗的作用，利用上个时刻攻击信号估计可以得到部分本时刻的中间估计。
- 利用自身的状态和输出来对最新时刻攻击信号的中间估计。

通信网络不存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\aleph_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

- 由于最新时刻攻击信号的中间估计是计算出来的，因此在计算残差时，这一部分的残差为零。

通信网络不存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\aleph_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

- 重新生成残差的关键：对同源攻击信号估计应相同。
- 未得到真实值时，攻击初步估计不同。可以利用这一差异重新生成残差。
- 生成的残差保留在时间窗内在下时刻发挥作用

通信网络不存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\aleph_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

- 步骤二利用一致性 $\sum_{j\in\aleph_i'} p_{ij}(\bullet)$ 使得系统对攻击信号的估计值趋于一致，并得到攻击信号估计。
- 利用邻居残差信息 $Y_j(k) - \tilde{Y}_j(k)$ 调整攻击估计值，其中 $\tilde{Y}_j(k)$ 由 $\tilde{E}_i(k)$ 计算。

通信网络不存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\mathcal{N}_i} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

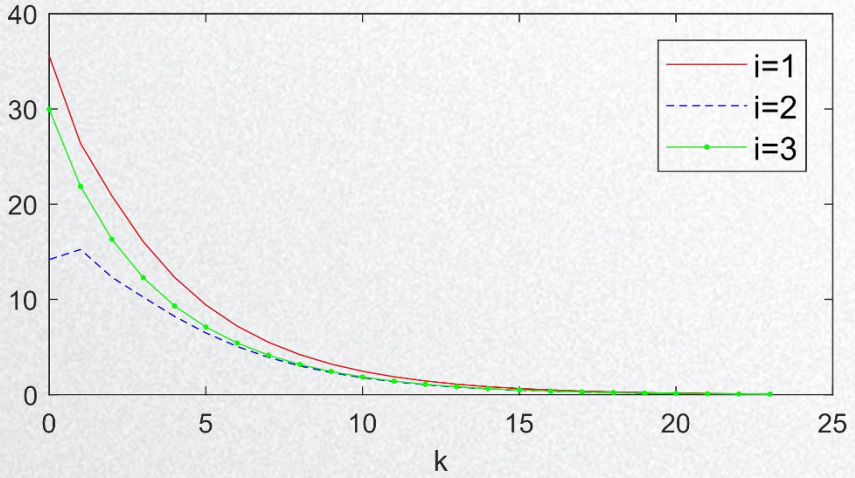
- 步骤三利用一致性生成的残差信息，调整状态估计值。该步骤是一个经典的Luenberger观测器。
- 观测器整体为分布式观测器

定理1：无延迟观测器的收敛性

考虑利用无延迟观测器的受攻击系统。若如下矩阵不等式对所有的智能体都成立，则无延迟观测器的估计误差渐近收敛到零。

$$\begin{bmatrix} Q_i^{-1} & * & * & * \\ \sqrt{p_{1i}}\xi_{1i}\left(H_i-\bar{L}_iN_i\right) & X_1 & * & * \\ \vdots & 0 & \ddots & * \\ \sqrt{p_{ji}}\xi_{ji}\left(H_i-\bar{L}_iN_i\right) & 0 & 0 & Q_j \end{bmatrix} \succ 0 (j\in\bar{\aleph}_i)$$

*矩阵不等式可以对非线性项增加限制进而转化为LMI问题求解。



无延迟观测器中各智能体估计误差的二范数

通信网络存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \tilde{E}_i(k) + L_E^i\left(Y_i(k) - \tilde{Y}_i(k)\right) + \sum_{j\in\aleph_i} W_{ij}\left(\hat{E}_i\left(k - t_{ij}(k)\right) - \hat{E}_j\left(k - t_{ij}(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

- 有时延迟观测器和无延迟观测器的步骤一和步骤三的设计思路相同。
- 其不同之处在于步骤二应用的一致性算法不同。

通信网络存在延迟的观测器

$$\tilde{E}_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \tilde{E}_i(k) + L_E^i\left(Y_i(k) - \tilde{Y}_i(k)\right) + \sum_{j\in\aleph_i} W_{ij}\left(\hat{E}_i\left(k - t_{ij}(k)\right) - \hat{E}_j\left(k - t_{ij}(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

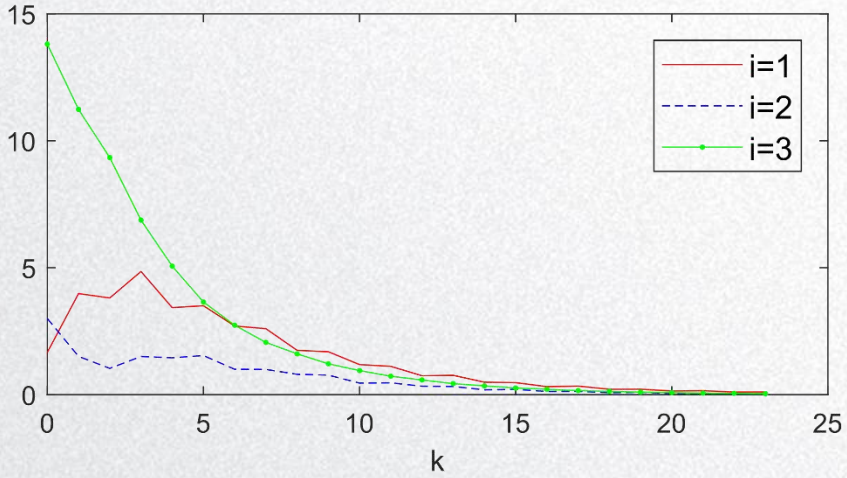
- 无延迟观测器取加权平均方法，其结构简单，易分析。
- 有延迟观测器将一致性作为额外项，其形式便于处理时间延迟。

定理2：有延迟观测器的收敛性

考虑利用有延迟观测器的受攻击系统。若如下矩阵不等式对所有的智能体都成立，则无延迟观测器的估计误差渐近收敛到零。

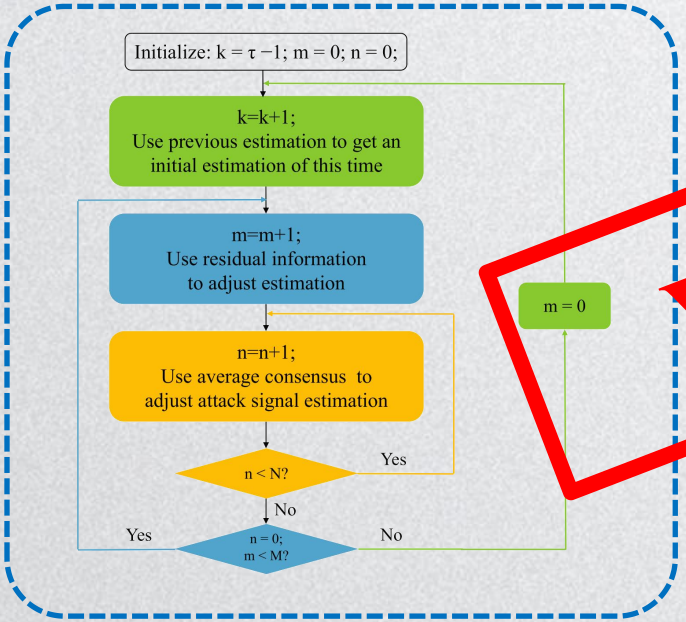
$$\begin{bmatrix} \Gamma & \Phi & 0 & \Omega^T & t_{\max}\left(\Omega^T - I\right) \\ * & \Xi & \Phi^T & \Pi^T & t_{\max}\Pi^T \\ * & * & \Psi & 0 & 0 \\ * & * & * & -Q^{-1} & 0 \\ * & * & * & * & -\frac{1}{l}P_2^{-1} \end{bmatrix} \prec 0$$

*矩阵不等式可以对非线性项增加限制进而转化为LMI问题求解。



有延迟观测器中各智能体估计误差的二范数

对比试验



三环在线观测器

VS

$$L_i(k) = S\hat{E}_i(k-1) + G_i\hat{x}_i(k) + T_i y_i(k)$$

$$\hat{E}_i(k) = \sum_{j\in\aleph_i'} p_{ij}\left(\tilde{E}_j(k) + L_E^j\left(Y_j(k) - \tilde{Y}_j(k)\right)\right)$$

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i u_i(k) + L_x^i\left(Y_i(k) - \tilde{Y}_i(k)\right)$$

有延迟观测器

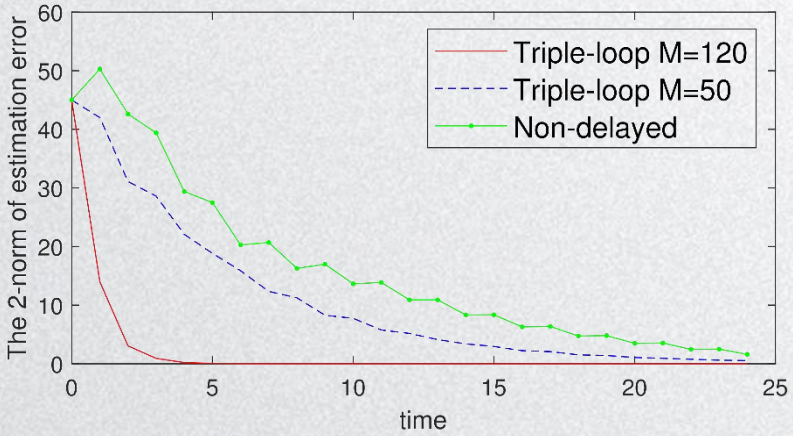
计算量对比试验

- 三环观测器每一时刻的循环次数M分别为120和50
- 无延迟观测器的计算量相当于M=1时三环观测器的计算量。

从计算量的角度出发可以看出无延迟观测器具有极低的计算量。

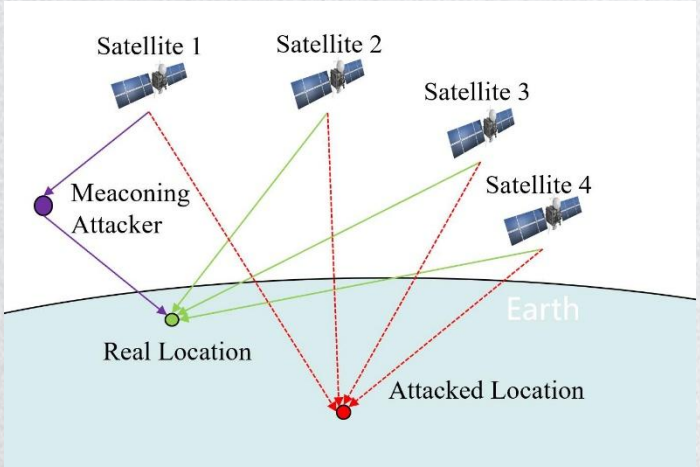
收敛速度对比试验

- M=120的三环观测器收敛速度最快但是计算量消耗极高。无延迟观测器的收敛速度接近M=50的三环观测器，但计算量却更小。



不同观测器的系统整体估计误差的收敛速度

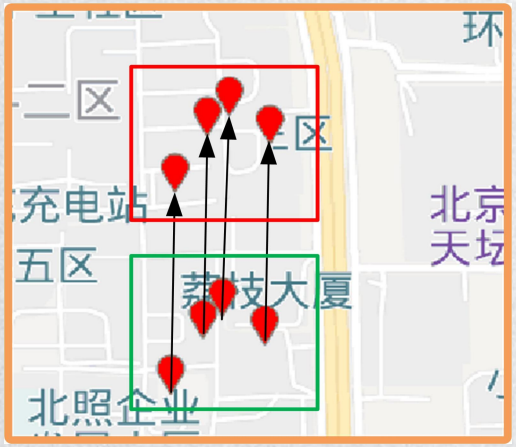
受攻击的GPS系统



GPS攻击原理

- 存在恶意攻击者拦截并记录了某颗卫星的信号。
- 在一段时间后将记录到的信号再次发送。
- 由于信号接受时间改变，使得伪距改变，进而产生定位误差

受攻击的GPS系统校准



GPS攻击导致的定位偏移

- 实验说明了在上述攻击下，相邻区域的接收器产生相同的定位偏差。
- 将真实定位视作状态、定位偏差视作攻击信号，本文的观测器可以实现GPS定位校准

受攻击的GPS校准实验

GPS攻击实验设备

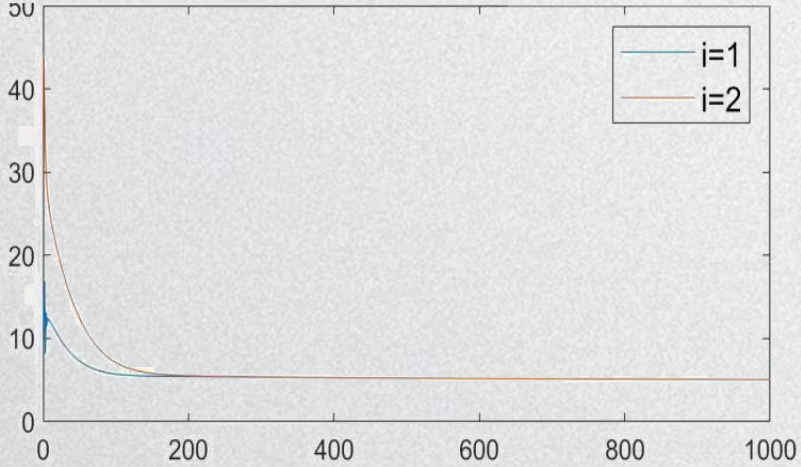
1、建立无人车动态模型
2、采集实际GPS信号
3、施加转发攻击
4、利用观测器重构定位

受攻击的GPS校准实验

	实际经度	实际纬度	攻击后经度	攻击后纬度	经度偏移量	纬度偏移量
智能体1	19	16	18.617	1.571	0.383	14.429
	24.043	23.752	23.71	9.29	0.333	14.462
	31.498	30.451	30.94	16.131	0.558	14.32
	43.359	40.792	43.081	26.295	0.278	14.497
智能体2	37.929	34.988	37.546	20.558	0.383	14.43
	19.535	23.9	19.202	9.437	0.333	14.463
	12.079	14.896	11.52	0.576	0.559	14.32
	7.525	9.249	7.247	-5.249	0.278	14.498

GPS攻击实验数据

受攻击的GPS校准实验



利用双环离线观测器实现的GPS校准实验

- 利用双环离线观测器对定位进行校准。横轴为观测器运行次数
- 在校准前定位误差分别为40米和30米。校准后定位误差小于10米。
- 最终10米定位偏差主要由于模型准确度、噪声和攻击细微的不同源导致。

总结与展望

- 在GPS攻击中实现全部观测器的测试，分析其在实际环境下的表现。
- 在石油化工过程中寻找可能存在的同源攻击或同源未知信号，利用分布式协同方法提高工业设备的安全性。

- Y. Shi, C. Liu and Y. Wang, "Secure State Estimation of Multiagent Systems With Homologous Attacks Using Average Consensus," IEEE Transactions on Control of Network Systems, vol. 8, no. 3, pp. 1293-1303, Sept. 2021.
- Y. Shi and Y. Wang, "Online Secure State Estimation of Multiagent Systems Using Average Consensus," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 5, pp. 3174-3186, May 2022.
- Y. Shi, C. Liu and Y. Wang, "Asymptotically Stable Filter for MVU Estimation of States and Homologous Unknown Inputs in Heterogeneous Multiagent Systems," IEEE Transactions on Automation Science and Engineering, vol. 19, no. 2, pp. 884-894, April 2022.
- Y. Shi, Y. Wang and J. Tuo, "Distributed secure state estimation of multi-agent systems under homologous sensor attacks," IEEE/CAA Journal of Automatica Sinica. doi: 10.1109/JAS.2022.105920

谢谢