Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-24 (2018) 963–968

Model-based Detection of Cyber-Attacks in Networked MPC-based Control Systems ⋆

Angelo Barboni *, **Francesca Boem** ****, **Thomas Parisini** *,**,***

* *Dept. of Electrical and Electronic Engineering, Imperial College London, UK*
** *KIOS Research and Innovation Centre of Excellence, University of Cyprus*
*** *Dept. of Engineering and Architecture, University of Trieste, Italy*
**** *Dept. of Electronic and Electrical Engineering, University College London, UK*

Abstract: In this preliminary work, we consider the problem of detecting cyber-attacks in a linear system equipped with a Model Predictive Controller, where the feedback loop is closed over a non-ideal network, and the process is subject to a random Gaussian disturbance. We adopt a model-based approach in order to detect anomalies, formalizing the problem as a binary hypothesis test. The proposed approach exploits the analytical redundancy obtained by computing partially overlapping nominal system trajectories over a temporal sliding window, and propagating the disturbance distributions along them. The recorded data over such window is then used to define a probabilistic consistency index at each time step in order to make a decision about the presence of possible attacks. Preliminary simulation results show the effectiveness of the proposed attack-detection method.

© 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Networked control, Cyber-attacks, Model-based detection, probabilistic, prediction methods, uncertain linear systems

1. INTRODUCTION

With the extensive diffusion of embedded computing, networked control systems have become more and more ubiquitous (see the surveys Baillieul and Antsaklis (2007), Hespanha et al. (2007)), playing a key role in areas like autonomous vehicles, UAVs, process industry, transportation systems and other critical infrastructure such as, for example, power networks. In the years to come, this tendency is set to increase even more along with the progress in robotics and autonomous systems and with the increasing safety requirements in industrial automation systems and critical infrastructures. While very useful to enable large-scale control, the presence of links over communication networks can potentially create entry points for malicious activities – to which control systems are vulnerable – as shown by the famous Stuxnet attack (see Langner (2011)), or by the more recent incidents targeting the Ukrainian power grid (Lee et al. (2016)).

In this connection, the problem of secure control is becoming of crucial importance and it is receiving a significant attention in the very recent literature (see, for instance, the special issue editorials Sandberg et al. (2015); Cheng et al. (2017) and the related papers).

Therefore, of utmost importance are (i) the problem of timely detection of cyber-attacks on communication links of networked control system and (ii) the problem of ensuring robustness to these attacks of the control system. Fault diagnosis (FD) methodologies in a networked scenario (see, for instance the seminal paper Patton et al. (2007), the recent work Boem et al. (2017) and the references cited therein) are potentially very relevant. Unfortunately, in a typical cyber-attack scenario characterized by *intelligent adversaries* the standard FD methodologies may not be well suited, at least in a first instance (Cárdenas et al. (2008)). A comprehensive survey of the literature is out of the scope of the present paper and hence in the following only a glimpse of some noticeable works is given.

There is a whole branch of research that deals with the design of "optimal" attacks able to maximize potential damage to system dynamics, while remaining undetected: Mo and Sinopoli (2010) determine conditions under which stealthy attacks against linear state estimators are possible, and Guo et al. (2016) investigate their optimality from the attacker's perspective. An approach including modeling of the attacker is presented in Teixeira (2014), where several deceptive strategies are shown. In Dutta and Langbort (2016), an output injection attack is presented, where an attacker with limited yet realistic information maximizes the estimation error in a Model Predictive Control (MPC) architecture. De Persis and Tesi (2015) do not consider signal attacks, but provide stability conditions for different strategies of Denial of Service attacks.

⋆ This work has been partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE). The first author is supported by the EPSRC Centre for Doctoral Training in High Performance Embedded and Distributed Systems (HiPEDS, Grant Reference EP/L016796/1).

2405-8963 © 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.09.691

Another batch of literature, that partially overlaps with the attack design case, tackles the problem of detecting adversarial activity. The focus has often been on identifying system vulnerabilities and monitoring limitations (see for example Pasqualetti et al. (2013)). In terms of notable strategies, Mo et al. (2015) exploit the idea of watermarking signals as proof of integrity, by leveraging their non-repeatability, and Forti et al. (2016) set the problem in a Bayesian framework.

The methodology illustrated in this paper falls into this latter category, since it is aimed at detecting malicious activities by a *model-based approach* in which – as in standard FD approaches – the knowledge of the nominal dynamic model of the monitored controlled system entails an *analytical redundancy* that can be exploited for on-line detection of cyber-attacks while the nominal controller is acting on the plant.

In this preliminary paper, we propose the integration of a novel attack-detection scheme with a separately designed networked Model Predictive Control (MPC) scheme. More specifically, we consider a linear stochastic plant controlled over a non-ideal link, where an attacker can delay network packets through rerouting or buffering, thus possibly affecting system performance quite severely. The availability of a MPC controller and the knowledge of the nominal model of the system allows the collection of past state and control trajectories that can be used by a custom maximum likelihood scheme to detect the possible presence of a malicious attack on the communication link. To the best of the Author‘s knowledge, it is the first time that an attack detection method is proposed for a networked MPC-based control system. In Dutta and Langbort (2016), a similar MPC framework is considered, but the aim of that paper is the attack design.

The paper is organized as follows. In the next section, the problem formulation is provided while in Section 3 the proposed approach is illustrated. In Section 4, extensive simulation results are given showing the effectiveness of the proposed technique.

2. PROBLEM FORMULATION

Consider a discrete-time Linear Time Invariant (LTI) system, whose state dynamics is described by [1]

$$x_{k+1} = Ax_k + B\tilde{u}_k + v_k\,, \tag{1}$$

where $x_k \in \mathbb{R}^n$ denotes the state vector, $\tilde{u}_k \in \mathbb{R}^m$ the applied control input, and $v_k \in \mathbb{R}^n$ denotes the realization of a white noise process disturbance $v \sim N(0, V)$ with zero mean and covariance matrix V.

System (1) is controlled by a Networked Control System (NCS) based on a MPC controller described in the following. The basic layout of the NCS with multiple loops sharing a packet-based communication network is depicted in Fig. 1, where, in order to distinguish the time delays in the sensor-to-controller and controller-to-actuator links, the network has been partitioned in two segments affected by delays $\tau_{sc}(k)$ and $\tau_{ca}(k)$, respectively. In this paper, the following assumption is in place.

[1] In this preliminary work, we only consider the case of perfectly measurable state variables.

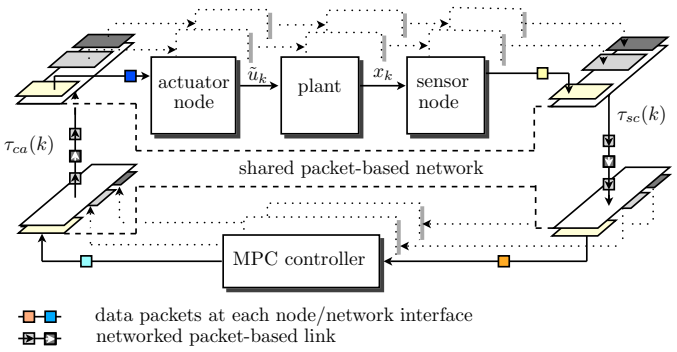


Fig. 1. Scheme of a NCS with multiple loops closed through a shared packet-based network with delayed data transmission.

Assumption 1. We assume that the sensor-to-controller link is perfect with no delays, that is $\tau_{sc}(k) = 0$, $\forall k \geq 0$. Moreover, we suppose that the random delay $\tau_{ca}(k)$ introduced by the controller-to-actuator link can be modeled as an exponentially distributed delay with mean rate $\bar{\delta}$, which is the typical choice for modeling network traffic (Bertsekas and Gallager (1987)). ◁

Assumption 2. There is no packet loss in the communication. ◁

The MPC controller implements a receding-horizon strategy, that is, at every time step k, it solves the problem of finding a control sequence $\mathbf{u}$, over a finite horizon N, that minimizes a cost function J defined as

$$J(x, k, \mathbf{u}) = \sum_{t=k}^{k+N-1} l(x_t, u_t) + F(x_{k+N})\,, \tag{2}$$

where l is a quadratic function in both its arguments and F is a terminal cost, chosen to guarantee closed loop stability. This minimization is subject to the state dynamics (1), and in this work we consider an unconstrained scenario.

Once this problem is solved, only the first element u_k of $\mathbf{u}$ is used as input to the plant via a networked connection. Due to the presence of delays, the input actually used by the plant is

$$\tilde{u}_k = u_{k-\tau_{ca}(k)}. \tag{3}$$

We assume the controller is designed offline, possibly incorporating the knowledge of the average delay, so to guarantee stability properties thanks to its intrinsic margins. However, in this paper, the stochastic nature of the delay is not taken into account by the controller, which relies on the nominal model of the system and the network. Moreover, the actuator follows a zero-order hold strategy thanks to the use of a buffer, i.e. it uses the latest received sample whenever a delay event happens. This makes the choice of a continuous random distribution for the delay compatible with our sampled time setting.

We consider an attacker that can alter the nominal parameters of the network by rerouting or buffering the original signal, therefore imposing an arbitrary delay rate. This strategy is similar to the wormhole attack described in Hu et al. (2006), and it is particularly appealing for an attacker as it does not require any particular information

about the plant or the controller model, and because it can be performed even on confidential and authenticated communications. The attacker's aim is that of degrading the dynamic performance of the controlled system, for example by inducing oscillation within a certain tolerance, possibly while remaining undetected.A complete analysis of the attack detectability and design optimization is out of the scope of this early and preliminary paper and will be subject of future research efforts.

While approaching the problem from a time-delay perspective is difficult, this attack can be considered from a signal point of view, which also allows for a more general framework should we consider other cases of data injections. To realize the effect of a d samples-long delay, with $d > \bar{\delta}$, the attacker may utilize an internal buffer of appropriate size, shifting at each time step the collected samples, storing the current one, and giving as output the following signal

$$a_k = u_{k-d} - u_{k-\tau_{ca}(k)}\,, \tag{4}$$

for k within the attack interval, if u_{k-d} is available in the buffer. According to this design, the attack can start after d time steps of input recordings. In view of these considerations, when an attack is underway, (3) can be rewritten as

$$\tilde{u}_k = u_{k-\tau_{ca}(k)} + a_k\,, \tag{5}$$

which is the form typically considered for attack models (Teixeira (2014)).

Given this setting, we want to design a detection strategy able to discriminate between the nominal and attacked behaviors. Ideally, we want to exploit the knowledge of the model and of the controller local data streams for enforcing the integrity of the closed loop, thus limiting the degrees of freedom of a possible attacker.

To address the aforementioned problem, we define the following hypotheses:

$\mathcal{H}_0$: "The system is attack-free"
$\mathcal{H}_1$: "The system is under attack".

The aim of the present preliminary paper is to design an on-line attack-detection model-based algorithm able to build a statistics allowing the rejection of the null hypothesis $\mathcal{H}_0$ when an attacker is tampering with the communication network between controller and actuators.

3. A MOVING-WINDOW ATTACK-DETECTION METHODOLOGY

We develop a novel strategy for detection that exploits a moving window, called *observation window* (see Figure 2), whose length L is fixed and represents a design parameter. The idea of using a sliding window based on past samples is not new, as it has been used for estimation (Alessandri et al. (2003)), fault detection (see for instance (Gatzke and Doyle, 2002) and Lauricella et al. (2017)), and even for detecting and correcting attacks on sensors by Shoukry et al. (2017) within the context of formal methods. Our approach is novel as we simulate a bundle of trajectories within the sliding window at every time step. These virtual trajectories are based on the nominal model and on nominal computed input values, and therefore they embody the expected behavior that the actual trajectory should have, within a certain error due to the disturbance.

To better understand the concept, let us consider the controller $\mathcal{C}$: at each time k it produces a forward sequence of length N of predicted *open loop* states

$$\mathbf{x}_N^f(k) = \left\{x^f_{k+1|k}, \dots, x^f_{k+N|k}\right\} = \{x^f_{n|k}\}_{n=k+1}^{k+N} \tag{6}$$

and inputs

$$\mathbf{u}_N(k) = \{u_{n|k}\}_{n=k}^{k+N-1} \tag{7}$$

over the control horizon. We use the notation $u_{n|k}$ to explicitly remark the dependence of the entire sequence on the information available at time k. At the next time step, the sequences $\mathbf{x}_N^f(k+1)$ and $\mathbf{u}_N(k+1)$ will be available, and so on. The elements of (6) and (7) are related by the controller's disturbance-free model

$$x^f_{t+1|k} = Ax^f_{t|k} + Bu_{t|k}\,, \tag{8}$$

for a given k and for $t \in [k, k+N-1]$. $x_{k|k}$ is the initial state from which the finite horizon optimal control problem is solved. As shown in Figure 2, these sequences shift at each iteration, each time with an updated initial state.

For monitoring reasons, we adjust the MPC predictions in the past observation window of length L, by using the control inputs $u_{t|t}$ actually communicated by the MPC controller to the actuators at each time step, that we recorded between $k-L$ and $k-1$, that is

$$\hat{x}_{t+1|j} = A\hat{x}_{t|j} + Bu_{t|j} + B(u_{t|t} - u_{t|j}), \tag{9}$$

for $j \in [k-L, k-1]$ and $t \in [j, k-1]$, and with $\hat{x}_{j|j} = x_j$ being the measured state. This sequences are highlighted in green in Figure 2. In this way, in the observation window, the monitoring unit has available a set of sequences based on the nominal model, the actual initial states, and it is driven by the control inputs actually computed at each time step. Under the assumption that there are no attacks acting on the network, these adjusted sequences differ from the real system trajectories only due to process disturbances. By exploiting the linearity of the control law and the Gaussian distribution of the disturbances, we can propagate the disturbance distributions ahead in time, in order to characterize those of prediction errors.

Let $\mu_{n|k} \doteq E[\hat{x}_{n|k}]$ and $Q_{n|k} \doteq \mathrm{Var}[\hat{x}_{n|k}]$ denote the mean and variance of the prediction values $\hat{x}_{n|k}$ seen as stochastic variables with respect to time n. Then, using the same notation as in (9), the mean and variance can be characterized by

$$\mu_{t+1|j} = A\mu_{t|j} + Bu_{t|t} + E[v_t] \tag{10}$$

$$Q_{t+1|j} = AQ_{t|j}A^T + \mathrm{Var}[v_t]. \tag{11}$$

Remark 1. The adjusted state sequences $\hat{x}$ only depend on the actual control actions fed to the system and the measured initial conditions at every time step. Therefore, they could be obtained even independently of the specific controller implementation, as we only rely on the superposition principle granted by the system linearity.

Even though the previous equations have been written in feed-forward form for the sake of clarity and notational convenience, we stress that (9) is computed on past predictions by means of the recorded inputs. That said, (10) and (11) allow to parametrize the distribution of each data point given the initial condition and time step of the prediction, thus we can associate with a measured sample

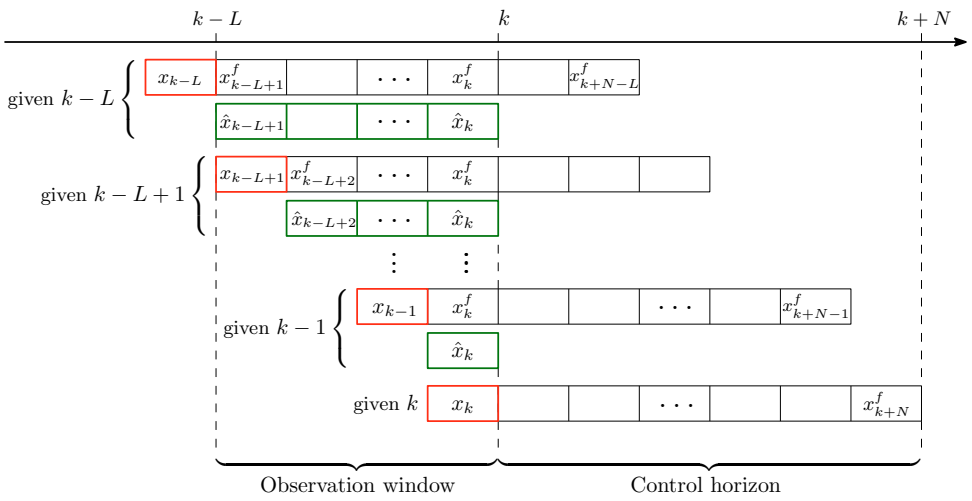


Fig. 2. Intuitive diagram illustrating the MPC open-loop prediction sequences x^f (boxed in black) and the adjusted sequences $\hat{x}$ (in green) available at time k, considering an MPC horizon of length N and an observation window of length L. In the red boxes the actual samples recorded in the considered time span are shown.

$x_t, t \in [k-L+1, k]$ an index $\mathcal{I}(x_t; \mathcal{H}_0, j), j \in [k-L, t-1]$ related to the probability of observing the data assuming the null hypothesis is true. For example, such index could be the likelihood of an observation for a time t in the window, given a past initial state within the same window.

$$\mathcal{L}(x_t; \mathcal{H}_0, j) = p(x_t; \mu_{t|j}, Q_{t|j}) \tag{12}$$

where $p(x)$ is the probability density function. We remark that since each expected distribution is the result of successive iterations, then its validity depends on the null hypothesis being true during the entire interval in which the uncertainty has been propagated using the nominal model.

However, to be representative the likelihoods in (12) need to be weighed into a normalized sum, so that their combination preserves the properties of a probability density. We first notice that in the observation window previously defined, at time $t = k - i, i \in [0, L-1]$, there exist $L - i$ distributions available for comparison. This can be visualized again in Figure 2, by looking at the adjusted sequences "column-wise". We introduced the forward index i relative to the window, as opposed to the absolute index j in order to simplify the following sums. We define weight vectors $\mathbf{d}_t$ such that the sum of their elements satisfies

$$\sum_{s=1}^{L-i} d_t^{[s]} = 1, \tag{13}$$

where the superscript $[s]$ denotes the s-th component. Once all the weight vectors $\mathbf{d}_t$ are found, one can obtain the weighted indices for all time steps in the window

$$\mathcal{I}_t(x_t; \mathcal{H}_0) = \sum_{s=1}^{L-i} d_t^{[s]} \mathcal{I}(x_t; \mathcal{H}_0, t-s) \tag{14}$$

While (13) gives some degrees of freedom as to the optimal choice of $\mathbf{d}_t$, in this work we consider decaying weights, based on the rationale that older predictions become less and less accurate as the time difference increases. One possible candidate that agrees with this principle and that does not excessively penalize the tail terms is the geometric progression with common ratio $d_t < 1$ and initial index $m = 1$, so we choose $d_t^{[s]} = d_t^s$, subject to

$$\frac{d_t - d_t^{L-i+1}}{1 - d_t} = 1. \tag{15}$$

The equation above has a unique solution in d_t in the open interval $(0, 1)$ for each $t \in [k-L+2, k]$. For $t = k-L+1$, (15) is the partial sum of the geometric progression, and the trivial solution is $d = 1$. The decay terms then need to be correctly ordered so that the vectors $\mathbf{d_t}$ are consistent with the previous considerations. Furthermore, these solutions can be found offline, because they depend on the remaining length of the observation window, which in turn depends only on its size.

The last step consists in aggregating the individual indices in (14) in order to obtain a scalar index for the entire observation window:

$$\mathcal{I}_L(k) = \frac{1}{L} \sum_{t=k-L+1}^{k} \mathcal{I}_t(x_t; \mathcal{H}_0). \tag{16}$$

Let X_N^f and U_N^f be the collections of $\{\mathbf{x}_N^f(k-L), \ldots, \mathbf{x}_N^f(k-1)\}$ and $\{\mathbf{u}_N(k-L), \ldots, \mathbf{u}_N(k-1)\}$ respectively, for example arranged in matrix form. We summarize in Algorithm 1 the steps described in this section in a procedural fashion. The two outer loops mirror the two phases of the method, namely the propagation of the uncertainty, and the evaluation of the index $\mathcal{I}$.

The index produced is finally compared to a threshold in order to decide between the two hypotheses.

4. SIMULATION RESULTS

To evaluate the effectiveness of the proposed method, we implemented the architecture illustrated in Section 2. We consider the discrete-time linear system

$$A = \begin{bmatrix} 1.2214 & 0.1681 \\ 0 & 0.5488 \end{bmatrix}, B = \begin{bmatrix} 0.0178 \\ 0.1504 \end{bmatrix}. \tag{17}$$

The choice of an unstable system is motivated by the fact that we adopt a ZOH strategy for the system inputs,

Algorithm 1 Computation of $\mathcal{I}_L(k)$

Require: $k > L$

procedure ComputeIndex(X_N^f, U_N^f)
  for $j \leftarrow k-L, k-1$ **do**
    for $t \leftarrow j+1, k$ **do**
      compute $\hat{x}_{t|j}$ as in (9)
      compute $\mu_{t|j}, Q_{t|j}$ as in (10), (11)
    end for
  end for
  for $t \leftarrow k-L+1, k$ **do**
    compute $\mathcal{I}_t$ as in (14) ▷ use offline $\mathbf{d}_t$
  end for
  return $\mathcal{I}_L(k) \leftarrow$ Mean($[\mathcal{I}_{k-L+1}, \dots, \mathcal{I}_k]$)
end procedure

that is the latest control input is applied when a sample is not received, which can potentially cause the state to diverge. As a result, the system is particularly sensitive to communication effects in a networked setting. The control objective is to steer the state to the origin from an initial condition $x_0 = [1, 1]^T$. In this case we choose $\bar{\delta} = 0.25T_s$, with T_s the sampling interval, in order to limit the delay variance and obtain reasonable nominal conditions. The system is controlled via unconstrained MPC with $N = 15$ samples, designed on the system with no delay, given the choice of $\bar{\delta}$. The simulation time is 80 s, and the observation window has size $L = N$.

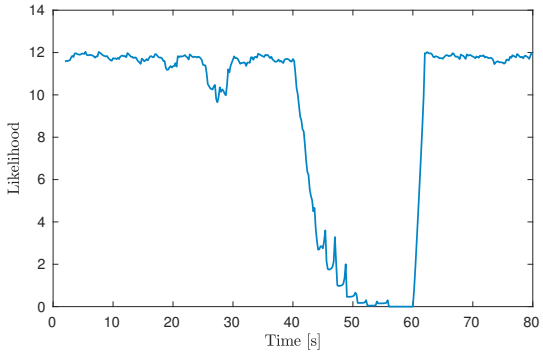
In the case of unconstrained MPC it is possible to determine the equivalent linear feedback gain K by means of the Fake Algebraic Riccati Equation (see Bitmead et al. (1990) for a detailed discussion on the topic). This allows to assess stability margins, for example via state augmentation techniques, under the assumption that the dwell time between model switches is sufficiently long. In this scenario, the closed loop is stable for delay values up to $\delta_{max} = 2$ samples. In the time interval $[40, 60]$ s, an attacker applies strategy (4) imposing a fixed delay rate of $d = \delta_{max} + 1$ samples.

In Figure 3, we show how the index computed via (12)–(16) is sensitive to the attack action, as a steep drop with respect to nominal conditions can be noted. The graph also shows smaller drops that correspond to random non-malicious delays. Future research efforts will be devoted at defining – in a model-based way – suitable thresholds for detecting the change and distinguish normal random deviations from attacks. In this paper, we consider the cumulative density function parametrized by (10) and (11) and evaluate the double-tailed probability (Casella and Berger (2002)) of an event; namely, for a sample x_t in the observation window, we evaluate

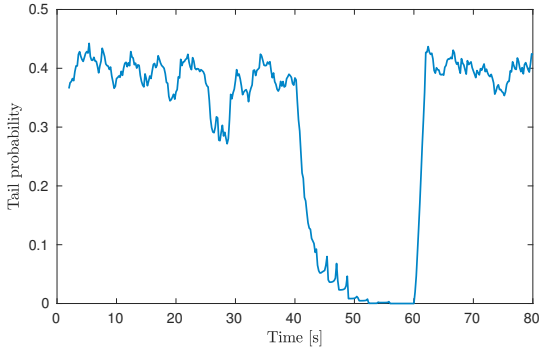
$$\mathcal{I}_t(x_t; \mathcal{H}_0, j) = 2P(X \leq -|x_t|), \tag{18}$$

where $X \sim N(\mu_{t|j}, Q_{t|j})$, for $j \in [k-L, t-1]$, and process them as described in (14) and (16). We then compute $\mathcal{I}_L(k)$ as described in the previous section, and the trend of this index is shown for one simulation in Figure 3(b), where a behavior similar to that of the likelihood can be noticed. We set a detection threshold α for $\mathcal{I}_L(k)$, and reject the null hypothesis when $\mathcal{I}_L(k) < \alpha$.

Given the stochastic nature of the scenario, we ran a batch of 100 simulations for a given variance matrix. In



(a) Likelihood index



(b) Tail probability

Fig. 3. Trends for a randomly chosen simulation in the batch for $V = 0.001I$. 3(a) shows the behavior of the index $\mathcal{I}$ when chosen to be the likelihood, 3(b) when using tail probabilities instead.

Table 1. Performance of the proposed method evaluated on the simulation batch for variance $V = 0.001I$

α	FPR	FNR	Precision	Recall	DD [s]
.10	.0125	.2008	.9587	.7992	4.2
.15	.0178	.1556	.9448	.8444	3.2
.20	.0229	.1238	.9318	.8762	2.6
.25	.0307	.0946	.9127	.9054	2.2

Table 2. Performance of the proposed method evaluated on the simulation batch for variance $V = 0.01I$

α	FPR	FNR	Precision	Recall	DD [s]
.10	.0173	.2081	.9428	.7919	4.4
.15	.0216	.1625	.9318	.8375	3.4
.20	.0263	.1281	.9210	.8719	2.8
.25	.0358	.0977	.8982	.9023	2.2

Table 1, some rates for the simulated batch are reported; specifically, for different choices of α we compute mean values of False Positive Rate (FPR), False Negative Rate (FNR), precision, recall, and Detection Delay (DD) over the batch. The detection delay is rounded up to the next discrete time step. Notice that increasing the threshold causes the system to be more reactive and increases the performance with respect to false negatives and recall, at the expense of an higher rate of false positives. In this case, recall is of particular importance, since it represents the ability of detecting an attack when it is taking place.

In Table 2, we repeat the experiments for a batch of the same size, but using a noise signal with larger variance $V = 0.01I$. Whereas the performance relative to the parameter α shows a consistent behavior as commented for the previous case, we notice that the larger uncertainty results in worsened performance, especially in case of lower thresholds.

5. CONCLUDING REMARKS

In this work, some preliminary results for a novel strategy to detect cyber-attacks in networked MPC-based control systems have been presented. A model-based approach has been illustrated in which the inherent analytical redundancy provided by the MPC controller is exploited. Extensive simulation results show the effectiveness of the proposed technique.

Future research efforts will be devoted to the important issue of attack detectability, more realistic control architectures will be considered removing, for example, the assumption on perfect state accessibility thus allowing for considering possible attacks also in the sensor-to-controller communication links. The issue of deriving analytically probabilistic detection thresholds will also be considered as well as the extension to the more difficult case of distributed large-scale systems controlled via decentralized schemes. Finally, the proposed approach will also be compared with other techniques of the state of the art, and will be tested in those cases where traditional approaches have been proven to fail.

REFERENCES

Alessandri, A., Baglietto, M., and Battistelli, G. (2003). Receding-horizon estimation for discrete-time linear systems. *IEEE Transactions on Automatic Control*, 48(3), 473–478.

Baillieul, J. and Antsaklis, P. (2007). Control and communication challenges in networked real-time systems. *Proc. of the IEEE*, 95, 9–28.

Bertsekas, D.P. and Gallager, R.G. (1987). *Data networks*, volume 2.

Bitmead, R.R., Gevers, M., and Wertz, V. (1990). *Adaptive optimal control: The thinking man's GPC*. Prentice Hall New York.

Boem, F., Ferrari, R.M.G., Keliris, C., Parisini, T., and Polycarpou, M.M. (2017). A distributed networked approach for fault detection of large-scale systems. *IEEE Transactions on Automatic Control*, 62, 18–33.

Cárdenas, A.A., Amin, S., and Sastry, S. (2008). Secure control: Towards survivable cyber-physical systems. In *Proceedings - International Conference on Distributed Computing Systems*, 495–500.

Casella, G. and Berger, R.L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Cheng, P., Shi, L., and Sinopoli, B. (2017). Control and communication challenges in networked real-time systems. *IEEE Trans. on Control of Network Systems*, 4, 1–3.

De Persis, C. and Tesi, P. (2015). Input-to-state stabilizing control under denial-of-service. *IEEE Transactions on Automatic Control*, 60(11), 2930–2944.

Dutta, A. and Langbort, C. (2016). Stealthy output injection attacks on control systems with bounded variables. *International Journal of Control*, 7179(October), 1–14.

Forti, N., Battistelli, G., Chisci, L., and Sinopoli, B. (2016). A Bayesian approach to joint attack detection and resilient state estimation. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 1192–1198.

Gatzke, E.P. and Doyle, F.J. (2002). Use of multiple models and qualitative knowledge for on-line moving horizon disturbance estimation and fault diagnosis. *Journal of Process Control*, 12(2), 339–352.

Guo, Z., Shi, D., Johansson, K.H., and Shi, L. (2016). Optimal Linear Cyber-Attack on Remote State Estimation. *IEEE Transactions on Control of Network Systems*, (99), 1–10.

Hespanha, J.P., Naghshtabrizi, P., Xu, Y., and Naghshtabrizi, Y.X. (2007). A survey of recent results in networked control systems. *Proceedings of the IEEE*, 95(1), 138–172.

Hu, Y.C., Perrig, A., and Johnson, D.B. (2006). Wormhole attacks in wireless networks. *IEEE journal on selected areas in communications*, 24(2), 370–380.

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3), 49–51.

Lauricella, M., Farina, M., Schneider, R., and Scattolini, R. (2017). A distributed fault detection and isolation algorithm based on moving horizon estimation. In *20th IFAC World Congress, Toulouse, France*.

Lee, R.M., Assante, M.J., and Conway, T. (2016). Analysis of the cyber attack on the Ukrainian power grid. *SANS Industrial Control Systems*.

Mo, Y. and Sinopoli, B. (2010). False Data Injection Attacks in Cyber Physical Systems. In *First Workshop on Secure Control Systems*.

Mo, Y., Weerakkody, S., and Sinopoli, B. (2015). Physical Authentication of Control Systems. *IEEE Control Systems Magazine*, 35(1), 93–109.

Pasqualetti, F., Dörfler, F., and Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11), 2715–2729.

Patton, R., Kambhampati, C., Casavola, A., Zhang, P., Ding, S., and Sauter, D. (2007). Control and communication challenges in networked real-time systems. *European Journal of Control*, 13, 280–296.

Sandberg, H., Amin, S., and Johansson, K.H. (2015). Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems*, 35(1), 20–23.

Shoukry, Y., Nuzzo, P., Puggelli, A., Sangiovanni-Vincentelli, A.L., Seshia, S.A., and Tabuada, P. (2017). Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*.

Teixeira, A. (2014). *Toward Cyber-Secure and Resilient Networked Control Systems*. Ph.D. thesis, KTH Royal Institute of Technology.