Conflict-driven Hybrid Observer-based Anomaly Detection

Zheng Wang[1], Farshad Harirchi[2], Dhananjay Anand[3], CheeYee Tang[3], James Moyne[1], Dawn Tilbury[1]

***Abstract*— This paper presents an anomaly detection method using a hybrid observer – which consists of a discrete state observer and a continuous state observer. We focus our attention on anomalies caused by intelligent attacks, which may bypass existing anomaly detection methods because neither the event sequence nor the observed residuals appear to be anomalous. Based on the relation between the continuous and discrete variables, we define three conflict types and give the conditions under which the detection of the anomalies is guaranteed. We call this method conflict-driven anomaly detection. The effectiveness of this method is demonstrated mathematically and illustrated on a Train-Gate (TG) system.**

I. Introduction

Cyber-Physical Systems (CPS) are systems that are shaped by a combination of computing devices, communication networks, and physical processes [1]. The integration of these systems into our every-day life is inevitable. The performance and functionality of many critical infrastructures such as power, traffic and health-care networks and smart cities rely on the advances on CPS. A fault or an attack on one of these critical systems, may affect a large portion of society with serious and lethal consequences. As such, the safety and reliability of CPS becomes more and more crucial every day. Fault, attack and anomaly detection mechanisms play a vital role in providing such reliability and safety to CPS. In this paper, we propose an anomaly detection approach that provides formal detection guarantees for an extended class of anomalies in CPS. Similar to [2], we refer to any occurrence that is different from what is standard, normal, or expected as *anomaly*. In this paper, we utilize the rich dynamical behavior of mixed continuous and discrete (i.e., hybrid) systems [3] as our modeling framework to describe CPS. Even though the design and implementation of anomaly detection methods is significantly more challenging on hybrid models, we leverage these models, because of their advantage in better representing the real-world CPS.

Our motivational example is a Train-Gate (TG) system, consisting of a train and a gate with a road crossing the track, as shown in Fig.1. It is an abstracted example that captures one of the important characteristics of a railway system which is railway level crossing control system. The TG system is a hybrid system. The train with an internal controller for the train speed is the continuous system. An external controller changes the reference train speed based on the measured train position such that the train passes the gate at a lower speed. The gate is a discrete system, which is raised or lowered by a controller using two presence sensors located on both sides of the road. If sensor 1 detects the train, the gate must be lowered down to stop traffic on the road. If sensor 2 detects the train, the gate must be raised up to allow traffic on the road. Two monitors are used to detect anomalies. One monitor detects anomalies in the continuous train system, which uses the continuous system model and compares the measured variables with the estimated ones. The other monitor detects anomalies in the discrete gate system, which uses the discrete system model and compares the expected discrete event sequence with observed one. If an anomaly is detected from either of these monitors, some actions should be taken to mitigate its impact.

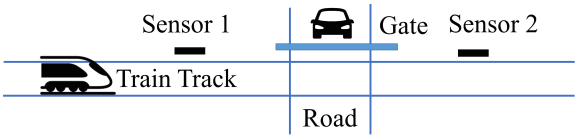

Fig. 1: TG Schematic

However, an attacker can launch an attack to cause an anomaly bypassing both monitors. For example, an anomalous ramp signal could be added to the measured train position without increasing the difference between the measured and the estimated variables. The drifted measured position can make the train pass the gate with a high speed, causing insufficient time to lower the gate. A driver may pass the gate, causing an accident.

In order to detect this type of anomaly, we propose a higher level monitor to augment the previous two monitors. This new monitor uses a hybrid model of the system, and estimates both the continuous and the discrete variables. For the anomaly of a ramp signal injection on the train position, although the continuous system is anomalous, the discrete system is normal. If sensor 1 detects the train but the estimated train position indicates that the train is far away from sensor 1, a "conflict" between the continuous and the discrete variables occurs. This new monitor expands the types of anomalies that can be detected by checking the occurrence of conflicts, called conflict-driven method. Both mathematical demonstrations and simulation results illustrate the effectiveness of the conflict-driven method.

II. Background and Contributions

Various model-based anomaly detection methods have been developed for both continuous systems and discrete systems [4], [5]. Even though discrete model-based anomaly detection methods are computationally efficient [6], they cannot provide sufficient resolution of continuous degradations

[1] Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA *zhengwa, moyne, tilbury@umich.edu*

[2] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA *harirchi@umich.edu*

[3] Software and Systems Divison, National Institute of Standards and Technology, Gaithersburg, MD, USA *dhananjay.anand, cheeyee.tang@nist.gov*

for hybrid systems [7]. Continuous model-based methods are impractical for the diagnosis of hybrid systems with a large number of discrete states because multiple continuous models need to run in parallel, each model corresponding to one discrete state [8].

As most CPS are hybrid, consisting of both continuous dynamics and discrete behavior, hybrid model-based approaches are promising in anomaly detection. Hybrid model-based anomaly detection includes set membership-based methods [9] and observer-based methods [10]. Given a data trajectory, set membership-based methods check whether the trajectory can be generated by the model. Even though these methods provide necessary and sufficient conditions in some cases for anomaly detection, they are computationally demanding, as they require costly set calculations or mixed integer programming. The set membership-based methods are also utilized in active fault diagnosis, where the goal is to design a minimal excitation that guarantees the detection of anomalous behavior [11], [12], [13].

Observer-based methods assume the continuous component of the hybrid model is observable under both normal and anomalous operations. For most observer-based methods, a residual, which is the difference between the estimated output and the actual output, is analyzed to determine the occurrence of an anomaly. State estimation problem is directly related to observer-based methods. Among various hybrid state estimation methods, a hybrid observer is better for real-time computation since it can reduce the computational complexity [10]. A hybrid observer consists of two components: a discrete state observer identifying what is the current discrete state, and a continuous state observer estimating the continuous state [10], [14]. With the hybrid observer framework, various traditional residual-based anomaly detection methods can be applied for hybrid systems, including different residual generation methods, such as the dedicated and generalized observer scheme [15], [16], and some residual evaluation methods, such as adaptive threshold [17].

A. Contributions

Even though the residual-based methods are efficient, intuitive and easy to implement, they can easily be tricked by a smart attacker or by sensor faults that make the continuous system unobservable, causing anomalies. An example of such class of anomalies is described in Section I. In this paper, we propose a conflict-driven anomaly detection approach with three conflict types defined based on the relation between the discrete and the continuous variables of the hybrid systems and in addition to faults that can be detected by traditional observer-based and residual-based methods, it is capable of providing guarantees on the detection of attacks and faults that are undetectable using these methods.

III. Problem Formulation & Solution

In this section, we describe the modeling framework that we consider and the anomaly types that are of interest. Also, a review of utilized hybrid observer is given.

A. Notation

Let $\|\cdot\|$ denote ∞-norm, $\tilde{\cdot}$ denote estimated variables, $\cup$ denote disjoint union, and $\square\sigma$ denote the ball of center 0 and of radius σ. In addition, $\mathbf{x} \in \mathbb{R}^n$ represents a vector, where its i^{th} element is indicated by $\mathbf{x}^{(i)}$. $\mathbf{A} \in \mathbb{R}^{n\times m}$ represents a matrix. The linear span of a set of vectors is denoted by $span(\cdot)$. For a set $X \subset \mathbb{R}^n$, we denote its closure, interior, and boundary by $\overline{X}$, X^o and ∂X respectively. Clearly, $\partial X = \overline{X} \backslash X^o$. The volume of the closed set $\overline{X}$ is denoted by $Vol(\overline{X})$.

B. Modeling Framework

1) Hybrid Model: A hybrid system can be modeled as a hybrid automaton $\mathcal{H} = (\mathcal{X}, \mathcal{U}, \mathcal{Y}, Init, field, E, \phi, \eta)$, where each element is defined as

- $\mathcal{X} = Q \times X$: a set of discrete and continuous states
- $\mathcal{U} = \Psi \times U$: a set of discrete and continuous inputs
- $\mathcal{Y} = \Omega \times Y$: a set of discrete and continuous outputs
- $Init = (q(t_0), \mathbf{x}(t_0)) \in \mathcal{X}$: an initial state
- $field : \mathcal{X} \times \mathcal{U} \to X$: a time invariant vector field
- $E = \Psi \cup \Omega$: a set of discrete events
- $\phi : Q \times \Psi \to Q$: a set of discrete transitions
- $\eta : \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$: an output map consisting of a discrete output map ζ and a continuous output equation h
  $\zeta : Q \times \Psi \to \Omega$: a discrete output map
  $h : \mathbf{y}(t) = \mathbf{C}_q \mathbf{x}(t) + \mathbf{v}(t)$: a continuous output equation

The hybrid models considered in this paper capture both nominal system model with a set of nominal discrete states Q_n and anomaly models with a set of anomalous discrete states Q_f. The set of all discrete states is defined as $Q = Q_n \cup Q_f$. The nominal hybrid system $\mathcal{H}_n$ can be derived by removing Q_f and the events and transitions connecting Q_f. The initial state $Init$, which is a combination of initial discrete state $q(t_0) \in Q_n$ and initial continuous state $\mathbf{x}(t_0)$, is not required to be known.

For each discrete state $q \in Q$, we consider continuous dynamics that can be represented by a Linear Time Invariant (LTI) model, subject to process and measurement noise.

$$\begin{aligned} field : \mathbf{x}(t+1) &= \mathbf{A}_q \mathbf{x}(t) + \mathbf{B}_q \mathbf{u}(t) + \mathbf{w}(t), \\ h : \mathbf{y}(t) &= \mathbf{C}_q \mathbf{x}(t) + \mathbf{v}(t), \end{aligned} \tag{1}$$

where $\mathbf{A}_q \in \mathbb{R}^{n\times n}, \mathbf{B}_q \in \mathbb{R}^{n\times n_u}, \mathbf{C}_q \in \mathbb{R}^{n_y \times n}$ are system matrices, $\mathbf{x} \in X \subset \mathbb{R}^n$, $\mathbf{u} \in U \subset \mathbb{R}^{n_u}$ and $\mathbf{y} \in Y \subseteq \mathbb{R}^{n_y}$ are continuous states, inputs and outputs, respectively. The process and measurement noise are represented by $\mathbf{w} \sim \mathcal{N}(0, \mathbf{W})$ and $\mathbf{v} \sim \mathcal{N}(0, \mathbf{V})$, respectively, where $\|\mathbf{w}\| \le w$ and $\|\mathbf{v}\| \le v$. Each entry of the process and measurement noise has its bound, i.e., $|\mathbf{w}^{(i)}| \le w_i$ and $|\mathbf{v}^{(i)}| \le v_i$. The continuous dynamical models of the system in anomalous discrete states are not required to be known. To simplify the notation, we assume:

Assumption 1: The output matrix $\mathbf{C}_q$ is an identity matrix in all discrete states, i.e., $\forall q \in Q, \mathbf{C}_q = \mathbf{I}$.

We can easily extend our work to general $\mathbf{C}$ matrix assuming the continuous system is observable.

Discrete events E can be partitioned into observable events E_o and unobservable events E_{uo}, i.e., $E = E_o \cup E_{uo}$. Only observable events can be detected by an observer. We denote the set of observable input events as Ψ_o and a set of

unobservable input events as Ψ_{uo}. Obviously, all of the output events are observable.

The i^{th} discrete event occurs at time t_i. The continuous evolutions occur in time $t \in [t_{i-1}+1, t_i], \forall i = 1,2,....$ In reality, discrete events may occur between two adjacent sample times. We assume

Assumption 2: The occurrence of the discrete events can be captured at sample times. At most one input event occurs within one sampling period. An output event occurs simultaneously with an input event.

Note that the discrete state is changed one time step after a discrete input event occurs, that is $\phi(q(t_i), \psi) = q'(t_i + 1)$, where $q(t_i), q'(t_i+1) \in Q$. To each discrete transition $\phi(q, \psi) = q'$, we associate a guard:

$$G(q, q', \psi) = \{\mathbf{x} : s_G \mathbf{x}^{(i_G)} \geq c_G\}, \tag{2}$$

where c_G is a constant value and s_G is either -1 or 1. A guard is a closed half-space divided by the hyperplane

$$\mathcal{P}(q, q', \psi) = \{\mathbf{x} : \mathbf{x}^{(i_G)} = s_G c_G\}. \tag{3}$$

A guard $G(q, q', \psi)$ indicates that the transition ψ will take place if and only if the i_G^{th} state variable of $s_G\mathbf{x}$ is no smaller than c_G in discrete state q.

To each discrete state $q \in Q$, we associate an invariant:

$$Inv_q = \{\mathbf{x} : \forall i = 1, ..., n, \underline{\beta}_i \leq \mathbf{x}^{(i)} \leq \overline{\beta}_i, \} \subseteq X, \tag{4}$$

where $\underline{\beta}_i$ and $\overline{\beta}_i$ are constant values. An invariant is a hyperrectangle with bounded intervals on each continuous state variable. An invariant Inv_q indicates that the system can remain in the discrete state q if and only if the continuous state $\mathbf{x} \in Inv_q \backslash \bigcup_j G(q, q_j, \psi_j)$.

Our definitions of guard $G(q, q', \psi)$ and invariant Inv_q indicate that c_G is between the lower and upper bounds of the state variable $\mathbf{x}^{(i_G)}$ of the invariant Inv_q, i.e., $\underline{\beta}_{i_G} \leq c_G \leq \overline{\beta}_{i_G}$. We define a neighbor hyperplane of guard $G(q, q', \psi)$ as

Definition 1: (Neighbor hyperplane of guard $G(q, q', \psi)$) is one of the hyperplanes forming the boundary of the invariant ∂Inv_q, which is defined as follows:

$$\begin{aligned}\mathcal{L}(q, q', \psi) = \{\mathbf{x} \in X : &|\mathbf{x}^{(i_G)} - \mathbf{x}'^{(i_G)}| = \min(c_G - \underline{\beta}_{i_G}, \overline{\beta}_{i_G} - c_G) \\ &\wedge \mathbf{x}^{(i_G)} \in \{\underline{\beta}_{i_G}, \overline{\beta}_{i_G}\}, \mathbf{x}' \in \mathcal{P}(q, q', \psi)\}.\end{aligned} \tag{5}$$

An example of neighbor hyperplane of $G(q, q', \psi)$ is shown in Fig.2. To simplify notation, we denote $c_{\mathcal{L}}$ as the value of $\mathbf{x}^{(i_G)}$, where $\mathbf{x} \in \mathcal{L}(q, q', \psi)$. If $\mathcal{P}(q, q', \psi)$ forms one of the hyperplanes of ∂Inv_q, then $\mathcal{L}(q, q', \psi) = \mathcal{P}(q, q', \psi)$. Otherwise, $\mathcal{L}(q, q', \psi) \cap \mathcal{P}(q, q', \psi) = \varnothing$. Discontinuities may exist in continuous variables due to discrete transitions in general hybrid systems. However, in our hybrid system formalism, no discontinuities exist in the continuous variables. This is imposed without any reset maps.

The hybrid observer used in this paper is proposed in [14], which is designed based on the Finite State Machine (FSM) associated with the nominal hybrid model. The FSM $\mathcal{M}_n$ is derived by removing all of the continuous dynamics in $\mathcal{H}_n$, and is represented by tuple $(Q, \Psi, \Omega, q(t_0), E, \phi, \zeta)$. In order to get a unique estimate of the discrete state with the hybrid observer after finite observable events, we assume

Assumption 3: The FSM $\mathcal{M}_n$ is current-state observable. Current-state observable is defined in [14].

Definition 2: (Current-State Observable) A FSM is current-state observable if there exists an integer k such that for any unknown initial discrete state, the discrete state at i can be determined from the observed input/output event pairs sequence up to i, i.e., $i \geq k$.

Note that one input/output event pair is considered as one input event to the hybrid observer. Thus, after the k^{th} input/output event pair occurs, the hybrid observer can give a unique estimated discrete state. The necessary and sufficient condition of current state observability is given in [14].

2) Nominal Discrete States: We partition the invariants of the nominal discrete states into an intermediate region and several normal operating regions. The intermediate region $\mathcal{R}_{in}$ is the union of all the intersections between the invariants of any two nominal discrete states

$$\mathcal{R}_{in} = \{\mathbf{x} \in X : \forall q, q' \in Q_n, q \neq q' \wedge \mathbf{x} \in Inv_q \cap Inv_{q'}\}. \tag{6}$$

For discrete state $q \in Q_n$, we define a normal operating region as the set of continuous states that are in the invariant but not the intermediate region,

$$\mathcal{R}_{no,q} = \overline{Inv_q} \backslash \overline{\mathcal{R}_{in}}. \tag{7}$$

We pose an assumption on the hybrid model to restrict state-space abstraction method. This assumption helps select the appropriate hybrid model of the system with which the conflict-driven method can provide detection guarantees.

Assumption 4: The intermediate region is bounded by the hyperplane $\mathcal{P}(q, q', \psi)$ corresponding to the guard $G(q, q', \psi)$ and the neighbor hyperplane of $G(q, q', \psi)$ and ∂Inv_q in each discrete state.

$$\begin{aligned}\mathcal{R}_{in} \subset \bigcup_{q \in Q_n} \{\mathbf{x} \in Inv_q : \forall q_j \in Q_n : \exists \mathcal{P}(q, q_j, \psi_j), \\ \min(c_G, c_{\mathcal{L}}) \leq \mathbf{x}^{(i_G)} \leq \max(c_G, c_{\mathcal{L}})\}.\end{aligned} \tag{8}$$

The visualization of this assumption on a 2-dimensional system is shown in Fig.2. Assumption 4 indicates $\mathcal{P}(q, q', \psi)$ is one of the hyperplanes forming $\partial Inv_{q'}$.

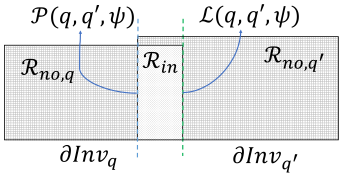

Fig. 2: Normal operating and intermediate regions

The basic principle of the conflict-driven method is to check at each time step, whether or not the sets of continuous states, which are calculated based on the estimated continuous state, intersect with the invariant of the estimated discrete state. The sets of continuous states include an initial set with the estimated continuous state as the center, and a forward reachable set which is the set of all continuous states that can be reached along trajectories starting in the initial set. Reachable set calculation requires following assumptions.

Assumption 5: The continuous system in each nominal discrete state is open-loop stable or marginally stable, i.e., $|\lambda(\mathbf{A}_q)| \leq 1$, where $\lambda(\mathbf{A}_q)$ are the eigenvalues of $\mathbf{A}_q$.

Assumption 6: The continuous input signal is bounded, and the bound is known, i.e., $\|\mathbf{u}\| \leq \mu$.

A great deal of attention has been given to algorithms and software developed for analysis of hybrid systems. To date, the most efficient way to compute the reachable set is to use zonotopes [18]. A zonotope is a Minkowski sum of a finite set of line segments, defined as

Definition 3: (Zonotope Z) is a set such that:

$$\begin{aligned} Z &= (\mathbf{x_c}, <\mathbf{g_1}, ..., \mathbf{g_p}>) \\ &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{x_c} + \Sigma_{i=1}^{i=p} b_i \mathbf{g_i}, -1 \leq b_i \leq 1\}, p \geq n, \end{aligned} \tag{9}$$

where $\mathbf{x_c}, \mathbf{g_i} \in \mathbb{R}^n$ are the center and generators, respectively. Both p and n determine the maximum number of vertices and facets.

C. Hybrid Observer

Given the nominal hybrid model $\mathcal{H}_n$, we can design a hybrid observer to estimate both the discrete state and the continuous state of the system using the method in [14]. The hybrid observer $\mathcal{O}$ consists of a discrete state observer $\mathcal{D}$ and a continuous state observer $\mathcal{C}$, as shown in Fig.3. The discrete state observer receives discrete input/output event pair (ψ, ω) and gives $\tilde{q}$. The estimated discrete state $\tilde{q}$ contains a set of estimated discrete states before the occurrence of the k^{th} observable input/output event pair. After the occurrence of the k^{th} observable input/output event pair, $\tilde{q}$, which contains a unique estimate, is passed to the corresponding continuous state observer. Then the continuous state observer gives an estimated continuous state $\tilde{\mathbf{x}}$ using the continuous input $\mathbf{u}$ and output $\mathbf{y}$.

The discrete state observer is represented by a FSM which is a tuple $\mathcal{D} = (\tilde{Q}, E_{\mathcal{D}}, -, Q_n, E_{\mathcal{D}}, \tilde{\phi}, -)$, where $E_{\mathcal{D}} = (\Psi, \Omega)$ is the set of discrete input/output event pairs of $\mathcal{M}_n$. The discrete state observer is tracking the set of possible discrete states that the system can be in. Therefore, no discrete output events or discrete map are defined for discrete state observer.

The construction of $\mathcal{D}$ starts from $\tilde{q}(t_0)$: with unknown initial discrete state of $\mathcal{M}_n$, $\tilde{q}(t_0) = Q_n$. Then for each discrete state $\tilde{q} \in \tilde{Q}$, we identify the input/output event pairs (ψ, ω), that label all the transitions out of any state q' in $\tilde{q}$. These events are called active event set of $\tilde{q}$. For each pair (ψ, ω) in the active event set, we identify $q \in Q_n$ that can be reached from $q' \in \tilde{q}$, and these states return as a new $\tilde{q}$ in $\tilde{Q}$. This transition is added to $\tilde{\phi}$ satisfying

$$\tilde{\phi} := \{q \in Q_n : \exists q' \in \tilde{q}, q \in \phi(q', \psi) \wedge \omega = \zeta(q', \psi)\}. \tag{10}$$

Repeat this step until no new $\tilde{q}$ and $\tilde{\phi}$ can be added to $\mathcal{D}$.

To reduce the effect of system noise on state estimation, we use a Kalman filter as the continuous state observer, with Kalman gain $\mathbf{K}_{q(t)}$. It is well known that the Kalman gain will converge in a few steps in practice if the system is observable [19]. We can use the steady state Kalman gain given in [19], with which the eigenvalues of $(\mathbf{A}_q - \mathbf{K}_q \mathbf{A}_q)$ are stable. Note that we have different Kalman gains for different discrete states. Let us define

Definition 4: (Dwell time Δt) is the minimum time to guarantee the convergence of the estimation error.

Dwell time Δt should satisfy the condition in section 3.2 in [14], Then we assume:

Assumption 7: The time gap between any two consecutive transitions is greater than dwell time, i.e., $t_i - (t_{i-1} + 1) > \Delta t$. With bounded noise, we design Kalman filter such that the estimation error $\mathbf{x_e}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}(t)$ is bounded when the Kalman filter reaches its steady state, i.e., $\exists t_{ss}, \|\mathbf{x_e}(t)\| \leq \theta, t > t_{ss}$. The residual $\mathbf{r}$ of the system is defined as the difference between the measure output and the estimated output,

$$\mathbf{r}(t) = \mathbf{y}(t) - \tilde{\mathbf{x}}(t). \tag{11}$$

In the nominal discrete states, the residual $\mathbf{r}(t), t > t_{ss}$ is bounded by $\theta + v$ because of bounded estimation error and noise. If $\|\mathbf{r}(t)\| > \theta + v, t > t_{ss}$, then the system is in an anomalous discrete state.

D. Anomalous Discrete States

An anomaly $f \in \Psi_{uo}$ is defined as an unobservable input event that transits the system from a nominal discrete state $q_n \in Q_n$ to an anomalous discrete state $q_f \in Q_f$. Arguably, the multiplicative anomalies can be represented by additive anomaly models (e.g., Section 3.5 in [4]). Thus, we restrict our attention to additive anomaly models as follows.

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t) + \mathbf{\Gamma}\boldsymbol{\gamma}(t), \tag{12}$$

where $\mathbf{\Gamma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with binary variables. The i^{th} diagonal variable is 1 if and only if the i^{th} output is added with an anomalous signal $\boldsymbol{\gamma}(t) \in \mathbb{R}^n$. Then the residual in anomalous discrete states is changed to

$$\mathbf{r}(t) = \mathbf{x_e}(t) + \mathbf{v}(t) + \mathbf{\Gamma}\boldsymbol{\gamma}(t). \tag{13}$$

The conflict-driven method is guaranteed to detect the anomalies that are not consistent with the continuous dynamics of the system, i.e, the anomalies that make the residual greater than threshold $\theta + v$. This is because of leveraging continuous state observer that is described in the previous subsection. Additionally, the proposed method extends the types of anomalies that can be detected compared to the methods mentioned in Section II.

Perfectly attackable systems are defined by Mo, et al. in [19] as continuous systems for which anomalies caused by certain attacks can remain undetected, i.e., the residual will not increase. One of the conditions for a continuous system to be perfectly attackable is that the state matrix $\mathbf{A}_{q_f}$ has at least one unstable or marginally stable eigenvalue. If the continuous system only has stable eigenvalues, anomalies on the system will increase the residual. The smart attacks that cannot be detected in perfectly attackable systems are called False Data Injection Attack (FDIA) as defined and demonstrated in [19]. One of the conditions of FDIA is that the eigenvector $\boldsymbol{\xi}$ corresponding to an unstable or marginally stable eigenvalue of $\mathbf{A}_{q_f}$ is in the span of $\mathbf{\Gamma}$, i.e., $\boldsymbol{\xi} \in span(\mathbf{\Gamma})$. If $\boldsymbol{\xi} \notin span(\mathbf{\Gamma})$, the anomaly will increase the residual and will be detected by the Kalman filter implemented as the continuous state observer in conflict-driven method.

As mentioned before, detecting FDIA type anomalies is challenging, as their effect cannot be observed in the value of residual. In addition to anomalies that can be detected

by checking the residual, our main contribution is to also guarantee the detection of this particular type of anomalies, if they satisfy certain conditions (explained in Section IV). Let us define Type-C_u anomalies for the hybrid systems as:

Definition 5: (Type-C_u anomaly) is an anomaly that is caused by False Data Injection Attack. If an anomaly occurs at time t_f, it satisfies the following two conditions.

1) The input-output sequence generated from the anomalous discrete state satisfies the continuous dynamics of the nominal discrete states for $t \geq t_f$, that is, the residual does not grow larger than the threshold $\theta + v$.

2) The occurrence of the anomaly results in:

$$\text{for } t \geq t_f, \text{ if } q \in Q_f \implies \|\mathbf{x_e}(t)\| > \theta.$$

Our objective is to extend the detection guarantees to the class of Type-C_u anomalies. In order to establish the goal, we also assume that:

Assumption 8: An anomaly occurs after the continuous state observer enters its steady state, i.e., $t_f \geq t_{ss}$.

IV. Conflict-driven Anomaly Detection Method

In the conflict-driven method, we define three conflict types. This method checks the occurrence of the conflicts to detect anomalies. The work flow diagram is shown in Fig.3. Note that this method is used after the hybrid observer is in the steady state, i.e., $t \geq t_{ss}$.

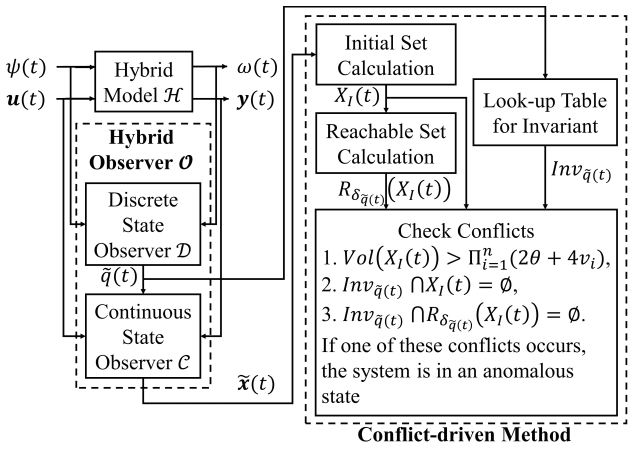

Fig. 3: Conflict-driven anomaly detection method

The conflict-driven method has three steps:

1) Calculate an initial set $X_I(t)$: An initial set $X_I(t)$ is constructed as a zonotope based on $\tilde{\mathbf{x}}(t)$ and $\mathbf{r}(t)$, as $X_I(t) = (\tilde{\mathbf{x}}(t), < \mathbf{g_1}, ..., \mathbf{g_n} >)$. The i^{th} generator $\mathbf{g_i}^{(i)} = |\mathbf{r}^{(i)}(t)| + v_i$. Other entries of vector $\mathbf{g_i}$ are zero. Based on Equation (11), we have $|\mathbf{x_e}^{(i)}(t)| \leq |\mathbf{r}^{(i)}(t)| + v_i$ in nominal discrete states. Thus, we can ensure $\mathbf{x}(t) \in X_I(t)$ when the system is in nominal discrete states. The initial set is changing at each time step because of the changes in the estimated continuous state and the residual.

2) Calculate the reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$: The $\delta_{\tilde{q}(t)}$ time-step forward reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$ is calculated starting from $X_I(t)$ constructed in Step 1. It satisfies

$$R_{\delta_{\tilde{q}(t)}}(X_I(t)) \subseteq \mathbf{A}_{\tilde{q}(t)}^{\delta_{\tilde{q}(t)}} X_I(t) + \square\sigma_{\tilde{q}(t)} \tag{14}$$

where $\sigma_{\tilde{q}(t)} = \frac{1-\|\mathbf{A}_{\tilde{q}(t)}\|^{\delta_{\tilde{q}(t)}}}{1-\|\mathbf{A}_{\tilde{q}(t)}\|}(\|\mathbf{B}_{\tilde{q}(t)}\|\mu + w)$. For more details about reachable set calculation using zonotopes, refer to [18].

3) Check conflicts: We define three conflict types in this paper, as shown in Fig.4:

Conflict A. The volume of the initial set is larger than the bound, i.e., $Vol(X_I(t)) > \Pi_{i=1}^{n}(2\theta + 4v_i)$

Conflict B. The initial set has no intersection with the invariant of the estimated discrete state $(X_I(t) \cap Inv_{\tilde{q}(t)} = \emptyset)$

Conflict C. The $\delta_{\tilde{q}(t)}$ time steps forward reachable set has no intersection with the invariant of the estimated discrete state, i.e., $R_{\delta_{\tilde{q}(t)}}(X_I(t)) \cap Inv_{\tilde{q}(t)} = \emptyset$.

If one of these conflicts occurs, the system is in an anomalous state.

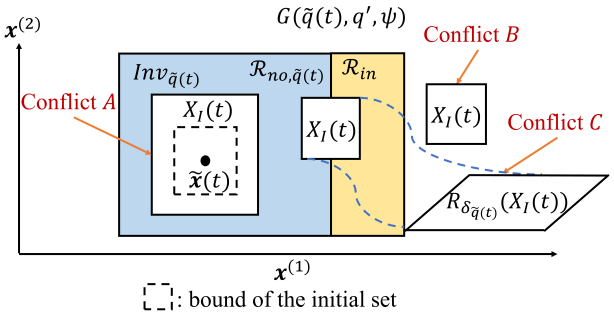

Fig. 4: Three conflict types

Note that for Step 2, we do not consider the discrete behavior in reachability analysis. The reachable set could be completely outside the invariant if $\delta_{\tilde{q}(t)}$ is too large, causing false alarms. To avoid false alarms and provide detection guarantees, we determine the time steps δ_q for each discrete state with given nominal hybrid model $\mathcal{H}_n$ of the system according to the following steps:

1) In Inv_q, starting from the intersection of the hyperplane corresponding to the i^{th} guard $G(q, q_i, \psi_i)$ as defined by (3) and Inv_q, we find the minimum time steps $\delta_{q,i}$ which satisfies

$$R_{\delta_{q,i}+1}(\mathcal{P}(q, q_i, \psi_i) \cap Inv_q) \cap \mathcal{L}(q, q_i, \psi_i) \neq \emptyset \tag{15}$$

Note that $\delta_{q,i}$ may be different for different guards in the same discrete state. The reason we use $\delta_{q,i} + 1$ is that the continuous system is a discrete-time model and we want to ensure the $\delta_{q,i}$ time-step forward reachable set, starting from any possible real continuous state when a transition occurs, has intersection with Inv_q in nominal discrete states.

2) Let $\delta_q = \min_i(\delta_{q,i})$. If the distance between $\mathcal{P}(q, q_i, \psi_i)$ and $\mathcal{L}(q, q_i, \psi_i)$ is small, δ_q may be 0. Then we only need to check Conflicts A and B in discrete state q.

Following proposition and theorem demonstrate the effectiveness of the conflict-driven method. We give some intuitions first. Proposition 1 gives the upper bound for the volume of the initial set. Based on Assumption 1, in a nominal discrete state, the estimation error of the continuous state, as well as the residual, should converge. Therefore, an upper bound exists for the volume of the initial set $Vol(X_I(t)), t > t_{ss}$, as demonstrated in Proposition 1. The increase of $Vol(X_I(t))$ indicates the increase of the residual. Conflict A can detect anomalies that increase the residual. Since the main contribution of this paper is focusing on the

detection of Type-C_u anomaly which does not increase the residual, finding the lower bound of the anomalous signal which causes conflict A and the conditions under which a residual-based method is equivalent to checking Conflict A are part of our future work.

Proposition 1: Given a nominal hybrid automaton $\mathcal{H}_n$ and a hybrid observer $\mathcal{O}$ with bounded estimation error in steady state, i.e., $\forall t > t_{ss}, \|\mathbf{x_e}(t)\| \leq \theta$, the volume of the initial set is also bounded, i.e., $Vol(X_I(t)) \leq \Pi_{i=1}^{n}(2\theta + 4v_i)$.

Proof: In steady state, $\forall t > t_{ss}$,

$$\begin{aligned} Vol(X_I(t)) =& \Pi_{i=1}^{n}(2(|\mathbf{r}^{(i)}(t)| + v_i)) \leq \Pi_{i=1}^{n}(2(\|\mathbf{x_e}(t)\| + 2v_i)) \\ \leq& \Pi_{i=1}^{n}(2\|\mathbf{x_e}(t)\| + 4v_i) \leq \Pi_{i=1}^{n}(2\theta + 4v_i) \end{aligned}$$

■

As discussed before, Type-C_u anomaly affects the continuous outputs of the system, but can remain undetectable by residual-based methods and unobserved by discrete state observer. In order to detect this type of anomaly, we leverage the estimated states from both continuous and discrete observers, and take advantage of observation of a discrete event. This enables us to employ the contradictions among estimated continuous and discrete states and the model parameters such as guards and invariants to detect these challenging anomalies. These contradictions are formalized in Conflicts B and C. In what follows, we set the stage to present the main contribution of this paper, which is Theorem 1. This theorem provides sufficient conditions on the lower bound of the anomalous signal, under which the conflict-driven method is guaranteed to detect Type-C_u anomalies. Towards this goal, we first find the lower bound of the estimation error that creates one of Conflicts B or C, and then relate this bound to the lower bound on the anomalous signal according to (13).

Let us assume that a Type-C_u anomaly occurs at time t_f which causes a large estimation error on the i_G^{th} state variable, i.e., $|\mathbf{x}_e^{(i_G)}| > \theta$, and a discrete event ψ occurs at time t_e which associates a guard with condition on the i_G^{th} state variable, i.e., $\{\mathbf{x} \in Inv_q : s_G \mathbf{x}^{(i_G)} \geq c_G\}$. Without loss of generality, we assume that the projection of $\mathcal{R}^o_{no,q}$ onto $\mathbf{x}_e^{(i_G)}$ is bounded above by c_G, i.e., $\mathbf{H}_{i_G}\mathcal{R}^o_{no,q} \leq c_G$ (because $s_G = 1$), where $\mathbf{H}_{i_G} \in \mathbb{R}^n$ is the projection row vector with the i_G^{th} entry "1" and "0" elsewhere. The procedure for the case where $\mathbf{H}_{i_G}\mathcal{R}^o_{no,q} \geq -c_G$ (because $s_G = -1$) is identical. When this event occurs, we can only have two possibilities for the estimated state at time t_e, either $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}^o_{no,q}$, or $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}^o_{in} \cap Inv_q$. Based on our definitions of guard, invariant, neighbor hyperplane of the guard, and Assumption 4, along the i_G^{th} state variable, the upper bound of Inv_q is $c_{\mathcal{L}}$ and the lower bound of $Inv_{q'}$ is c_G. For brevity in notation and as in this section we mainly consider $G(q, q', \psi)$, we refer to it as G.

First, consider the case where $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}^o_{no,q}$, that is, when the real continuous state satisfies the guard, the estimated state is in the normal operating region of discrete state q. The goal is to find the lower bound of the estimation error along the i_G^{th} state variable, such that:

- The initial set $X_I(t_e + 1)$ has no intersection with $Inv_{q'}$.

We denote such minimum estimation error corresponding to G by z_G^*. To find z_G^*, it suffices to find the minimum z such that for all $\tilde{\mathbf{x}}(t_e + 1)$ the upper bound of $X_I(t_e + 1)$ is smaller than the lower bound of $Inv_{q'}$ along the i_G^{th} state variable,

$$\mathbf{H}_{i_G}\tilde{\mathbf{x}}(t_e + 1) + \theta + 2v < c_G. \tag{16}$$

Note that at time t_e, the continuous state of the system along the i_G^{th} state variable is greater than or equal to c_G, and smaller than the maximum value of the one time step forward reachable set from $\mathcal{P}(q, q', \psi) \cap Inv_q$ along the i_G^{th} state variable, i.e., $c_G \leq \mathbf{H}_{i_G}\mathbf{x}(t_e) < \varepsilon$, where $\varepsilon = \max(\mathbf{H}_{i_G}R_1(\mathcal{P}(q, q', \psi) \cap Inv_q))$. After the occurrence of event ψ, the state equation of the anomalous discrete state is changed to $(\mathbf{A}_{q'}, \mathbf{B}_{q'})$ and the estimated discrete state is changed to q' at time $t_e + 1$. Then the set of all possible continuous states at time $t_e + 1$ can be represented by:

$$\begin{aligned} &\forall \mathbf{x}(t_e) \in Inv_q, c_G \leq \mathbf{H}_{i_G}\mathbf{x}(t_e) < \varepsilon, \\ &\qquad \mathbf{x}(t_e + 1) \in R_1(\mathbf{x}(t_e)) \subseteq \mathbf{A}_{q'}\mathbf{x}(t_e) + \square\sigma_{q'}, \end{aligned} \tag{17}$$

where $\sigma_{q'} = \|\mathbf{B}_{q'}\|\mu + w$.

Now, we can pose the problem of finding z_G^* as a robust optimization problem.

$$\begin{aligned} z_G^* = \min_{z} \quad & z \\ \text{s. t.} \quad & z \geq 0,\ z \geq \mathbf{H}_{i_G}\mathbf{A}_{q'}\mathbf{x} + \sigma_{q'} + \theta + 2v - c_G \\ & \forall \mathbf{x} \in Inv_{q'}, c_G \leq \mathbf{H}_{i_G}\mathbf{x} \leq \varepsilon, \end{aligned} \tag{18}$$

By utilizing methods from robust optimization literature, e.g., [20], we can convert (18) to a linear programming problem as follows:

$$\begin{aligned} z_G^* = \min_{\mathbf{J}, z} \quad & z \\ \text{s. t.} \quad & \begin{bmatrix} 1 \\ 1 \end{bmatrix} z - \begin{bmatrix} \mathbf{J}^\top \boldsymbol{\rho}_\mathbf{1} \\ 0 \end{bmatrix} \geq \begin{bmatrix} \sigma_{q'} + \theta + 2v - c_G \\ 0 \end{bmatrix} \\ & \mathbf{\Lambda}^\top \mathbf{J} \geq (\mathbf{H}_{i_G}\mathbf{A}_{q'})^\top, \mathbf{J} \geq \mathbf{0} \end{aligned} \tag{19}$$

where $\mathbf{0} \in \mathbb{R}^{2n \times 1}$ is a zero vector. $\mathbf{x}$ is in a polytopic uncertain set, i.e., $\mathbf{\Lambda x} \leq \boldsymbol{\rho}_\mathbf{1}$ for problem (18), where $\mathbf{\Lambda} \in \mathbb{R}^{2n \times n}$, $\boldsymbol{\rho}_\mathbf{1} \in \mathbb{R}^{2n \times 1}$ and $\mathbf{J} \in \mathbb{R}^{2n \times 1}$ is a variable of the optimization problem.

For the second possibility, i.e., $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}^o_{in} \cap Inv_q$, we are seeking the lower bound of the estimation error along the i_G^{th} state variable such that it satisfies the following:

- The reachable set for δ_q time steps from any point within the initial set $X_I(t_e)$ of the estimated continuous state has no intersection with Inv_q.

Considering the worst case that the continuous state is the furthest to the upper bound of ∂Inv_q along the i_G^{th} state variable, i.e., $\mathbf{H}_{i_G}\mathbf{x}(t_e) = c_G$, our objective can be equivalently changed to find the minimum distance between c_G and $\mathbf{H}_{i_G}\tilde{\mathbf{x}}(t_e)$. We denote this minimum distance by d_G^*. Define $d = |\mathbf{H}_{i_G}\tilde{\mathbf{x}}(t_e) - c_G|$ as the distance between $\mathcal{P}(q, q', \psi)$ and the estimated state along the i_G^{th} state variable. With this definition, the initial set at time t_e can be represented as $X_I(t_e) = \{\mathbf{x} : \mathbf{H}_{i_G}\mathbf{x} \in [c_G + d - \theta - 2v, c_G + d + \theta + 2v]\}$. Starting from this initial set $X_I(t_e)$, the projection of the reachable set for δ_q time steps forward onto the i_G^{th} state variable becomes $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} \pm \sigma_q,\ \forall \mathbf{x} \in X_I(t_e)$, where $\sigma_q = \frac{1 - \|\mathbf{A}_q\|^{\delta_q}}{1 - \|\mathbf{A}_q\|}(\|\mathbf{B}_q\|\mu + w)$. If $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} - \sigma_q > c_{\mathcal{L}},\ \forall \mathbf{x} \in X_I(t_e)$, then it is guaranteed that the δ_q time-step forward reachable set starting from this

initial set $X_I(t_e)$ has no intersection with the invariant Inv_q. We can pose the problem of finding d_G^* as the following robust optimization problem.

$$\begin{aligned} d_G^* = \min_{d} \quad & d \\ \text{s. t.} \quad & d \geq 0,\ \mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} - \boldsymbol{\sigma}_q \geq c_{\mathcal{L}} \\ & \forall \mathbf{x} \in Inv_q, \mathbf{x} \in X_I(t_e). \end{aligned} \tag{20}$$

With a change of variables and by employing the robust optimization techniques [20], we can write an equivalent problem to (20) as a linear program.

$$\begin{aligned} d_G^* = \min_{\mathbf{D},\mathbf{J}} \quad & \mathbf{H}_{i_G}\mathbf{D} \\ \text{s. t.} \quad & \begin{bmatrix} \mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q} \\ \mathbf{H}_{i_G} \end{bmatrix}\mathbf{D} - \begin{bmatrix} \mathbf{J}^\top \boldsymbol{\rho}_2 \\ 0 \end{bmatrix} \geq \begin{bmatrix} \boldsymbol{\sigma}_q + c_{\mathcal{L}} \\ 0 \end{bmatrix} \\ & \boldsymbol{\Lambda}^\top \mathbf{J} \geq -(\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q})^\top, \mathbf{J} \geq \mathbf{0}, \mathbf{D} \geq \mathbf{0} \end{aligned} \tag{21}$$

where $\mathbf{0}$ is a zero vector with proper dimension, and $\mathbf{D} \in \mathbb{R}^n$ is a vector with the i_G^{th} entry d and other entries "0". $\mathbf{x}$ is in a polytopic uncertain set, i.e., $\boldsymbol{\Lambda}\mathbf{x} \leq \boldsymbol{\rho}_2$, where $\boldsymbol{\rho}_2 \in \mathbb{R}^{2n\times 1}$ and $\mathbf{J} \in \mathbb{R}^{2n\times 1}$ is the dual variable.

Now that we have introduced z_G^* and d_G^*, we can present the main result of the paper.

Theorem 1: Given a nominal hybrid automaton $\mathcal{H}_n$. Assume a Type-C_u anomaly f occurs at time t_f. If an event $\psi \in \Psi_o$ occurs at $t_e > t_f$, which is supposed to transit the system from discrete state q to q', and the guard $G(q,q',\psi)$ is a condition on the real continuous state which is affected by the anomaly f, i.e., $G(q,q',\psi): s_G\mathbf{x}^{(i_G)} \geq c_G$ and $|\mathbf{x}_\mathbf{e}^{(i_G)}| \geq \theta$, then the conflict-driven method is guaranteed to detect the anomaly, if the anomaly satisfies:

$$\|\boldsymbol{\Gamma}\boldsymbol{\gamma}(t)\| > \max(z_q^* + \theta + 2v, d_q^* + \theta + 2v), \tag{22}$$

where $z_q^* = \max_{q'} z_G^*$ and $d_q^* = \max_{q'} d_G^*$ can be derived by solving the robust optimization problems (18) and (20), respectively for all possible q'.

Proof: The solution z_G^* is the lower bound of the estimation error which ensures $X_I(t_e+1) \cap Inv_{q'} = \varnothing$, i.e. Conflict *B*. The values of z_G^* varies from one guard to another. Therefore, by considering z_q^*, we guarantee that at the discrete state q, regardless of guard, Conflict *B* occurs, if $\|\boldsymbol{\Gamma}\boldsymbol{\gamma}(t)\| > z_q^* + \theta + 2v$. On the other hand, the solution d_G^* is the lower bound of the estimation error, which ensures $R_{\delta_q}(X_I(t_e)) \cap Inv_q = \varnothing$, i.e., Conflict *C*. The values of d_G^* varies for different guards, hence, we similarly take the maximum of these values for all possible q', which is d_q^*. Based on the relation between the estimation error and anomalous signal in (13), it is guaranteed that if $\|\boldsymbol{\Gamma}\boldsymbol{\gamma}(t)\| > d_q^* + \theta + 2v$, regardless of guard, Conflict *C* occurs. By combining the two conditions obtained on the magnitude of anomalous signal for the two possibilities, we can conclude that the proposed conflict-driven method provides detection guarantees on the detection of anomalous signals that satisfy condition (22), regardless of where the estimated state is located in the Inv_q at the time of event. This concludes the proof. ■

V. Simulation Result

In this section, we revisit the TG system. We present the nominal hybrid model of the TG system and compare the conflict-driven method with a residual-based method under a Type-C_u anomaly.

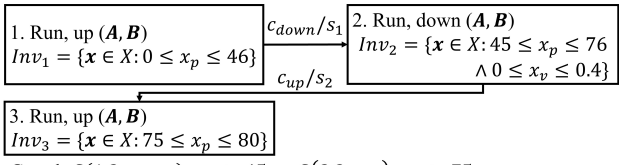

Fig. 5: Hybrid automaton $\mathcal{H}_n$ of the TG system.

The graphic representation of the nominal hybrid model $\mathcal{H}_n$ of the TG system is shown in Fig.5. The train automaton has one discrete state "run". The gate automaton has two discrete states: "up" and "down" (The time of raising and lowering the gate is ignored for simplicity). Although the automata product results in two discrete states, we additionally partition discrete state "run, up" to two discrete states to ensure hyperrectangle invariants as defined in (4). The discrete transitions between discrete states are determined by discrete input events c_{up} and c_{down}, where c_{up} means "raise the gate" and c_{down} means "lower the gate". When sensor 1 detects the train and emits discrete output event s_1, the gate controller sends out c_{down}. When sensor 2 detects the train and emits discrete output event s_2, the gate controller sends out c_{up}. For each transition, we associate a guard. The invariants of the discrete states and the guards are given in Fig.5. The continuous state of the TG system is $\mathbf{x} = [x_p \quad x_v]^\top$, where x_p, x_v are the train position and the train speed, respectively. The continuous output of the TG system is $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$. If the train is within $16m$ of the gate, the reference speed is $0.2m/s$. Otherwise, it is $1m/s$. The desired operation is that the train speed is no faster than $0.4m/s$ when the train is within $12m$ of the gate. The TG system is current state observable. Based on Assumption 8, we will only focus on the observer's steady state [1,2].

The intersections of the invariants give the intermediate region $\mathcal{R}_{in}$ as $\mathcal{R}_{in} = \{\forall \mathbf{x} \in X : (45 \leq x_p \leq 46 \vee 75 \leq x_p \leq 76) \wedge 0 \leq x_v \leq 0.4\}$. Then we can determine the normal operating regions of the three discrete states, as shown in Fig.6,

$$\begin{aligned} \mathcal{R}_{no,1} &= \{\forall \mathbf{x} \in X : 0 \leq x_p < 45 \vee (45 \leq x_p \leq 46 \wedge x_v > 0.4)\}, \\ \mathcal{R}_{no,2} &= \{\forall \mathbf{x} \in X : 46 < x_p < 75 \wedge 0 \leq x_v \leq 0.4\}, \\ \mathcal{R}_{no,3} &= \{\forall \mathbf{x} \in X : 76 < x_p \leq 80 \vee (75 \leq x_p \leq 76 \wedge x_v > 0.4)\}. \end{aligned} \tag{23}$$

The neighbor hyperplane of each guard is then:

$$\mathcal{L}(1,2,c_{down}): x_p = 46, \quad \mathcal{L}(2,1,c_{up}): x_p = 76, \tag{24}$$

With the invariants, guards and neighbor hyperplanes, we can determine the time step for reachability analysis of each discrete state, which is $\delta_1 = \delta_2 = 9, \delta_3 = 0$.

[1]More parameters: track length: $80m$; gate, sensor 1, sensor 2 locations: $60m$, $45m$, $75m$; sampling period: $0.1s$; Upper bounds of noise: $v = 0.1$, $w = 0.01$ (Units depend on the state variable with larger noise); estimation error upper bound in the observer's steady state in nominal discrete states: $\theta = 0.05$ (The unit depends on the state variable with larger estimation error at sample times).

[2]In reality, the train track intersects with multiple roads at different locations. The discrete state observer gives a unique estimated discrete state after passing the first road. We only focus on the track segment when the observer is in steady state.

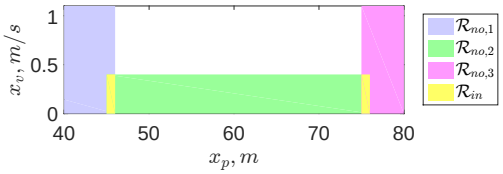

Fig. 6: The normal operating regions and the intermediate region of the TG system.

In these three discrete states, state matrices $(\mathbf{A}, \mathbf{B})$ are the same. The eigenvalues of $\mathbf{A}$ are 1 and 0.95. The eigenvector $\boldsymbol{\xi}$ corresponding to the marginally stable eigenvalue is $[1 \quad 0]^\top$. The non-zero element of $\boldsymbol{\xi}$ corresponds to the measured train position. A Type-C_u anomaly scenario is a ramp anomalous signal with slope $0.02m/s$ added to the measured train position. The anomaly starts at $0s$ and runs until the end of the simulation, which makes the system violate its desired operation at $180.8s$ with position $71.55m$ and speed $0.41m/s$.

The comparison of the detection performance of the conflict-driven method and a residual-based method under the anomaly mentioned above is shown in Fig.7. The threshold of the residual-based method is $\theta + v = 0.15$ (The unit depends on θ). The residual-based method fails to detect the anomaly because the residual does not increase. The conflict-driven method detects this anomaly at time $48.2s$ when Conflict C occurs. The estimated discrete state is 1, but the reachable set $R_{\delta_1}(X_I(482)) \cap Inv_1 = \varnothing$. At $48.2s$, the norm of the anomalous signal is $0.96m$, which is lower than the lower bound $0.98m$ calculated by solving robust optimization problems. That means the conflict-driven method may detect the anomalies with norm lower than the lower bound which we can provide detection guarantees.

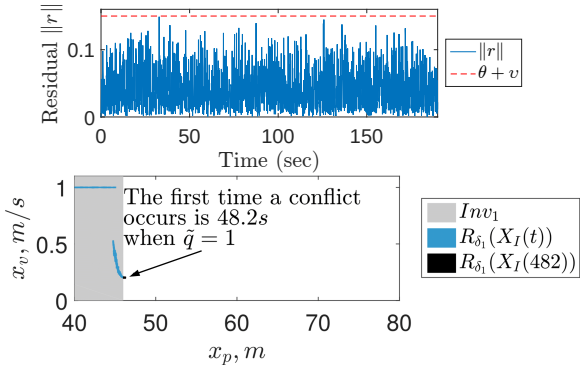

Fig. 7: Simulation result under the Type-C_u anomaly: (Top) Residual; (Bottom) The occurrence of Conflict C.

VI. Conclusion and Future Work

In this paper, we propose a conflict-driven method, which uses the discrete and continuous variables and the hybrid model of the system, to provide detection guarantees for anomalies that are undetectable with traditional residual-based methods in addition to anomalies that can be detected with these methods. We define three different conflict types. If any one of the conflicts occurs, the anomaly is detected. Both mathematical demonstration and simulation result illustrate the effectiveness of the conflict-driven method.

More work needs to be done about the conflict-driven method. One future work is to improve the hybrid observer design such that we can apply the conflict driven method to more general hybrid systems with reset maps. One potential solution is to use the Convergence Ratio method in [16], which calculates the estimation error of the continuous state with two continuous state observers. Other future work includes the analysis of the conflict-driven method in detecting anomalies that affect both the continuous and discrete variables of the system.

Acknowledgment

This work was supported by the National Institute of Standards and Technology under Award No.70NANB16H205. We thank Isaac Spiegel, a Ph.D. student at University of Michigan for his discussion on hybrid system definition.

References

[1] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *CPSSW*, 2009, p. 5.

[2] F. Lopez, M. Saez, Y. Shao, E. Balta, J. Moyne, M. Mao, K. Barton, and D. Tilbury, "Categorization of anomalies in smart manufacturing systems to support the selection of detection mechanisms," in *IEEE CASE*, 2017.

[3] K. Wan, D. Hughes, K. L. Man, and T. Krilavičius, "Composition challenges and approaches for cyber physical systems," in *IEEE NESEA*, 2010, pp. 1–7.

[4] S. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[5] M. Sayed-Mouchaweh, *Discrete Event Systems: Diagnosis and Diagnosability*. Springer Science & Business Media, 2014.

[6] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. C. Teneketzis, "Failure diagnosis using discrete-event models," *IEEE Trans. on control systems technology*, vol. 4, no. 2, pp. 105–124, 1996.

[7] C. H. Goodrich and J. Kurien, "Continuous measurements and quantitative constraints: Challenge problems for discrete modeling techniques," 2001.

[8] F. Zhao, X. Koutsoukos, H. Haussecker, J. Reich, and P. Cheung, "Monitoring and fault diagnosis of hybrid systems," *IEEE Trans. on Systems, Man, and Cybernetics-B*, vol. 35, no. 6, pp. 1225–1240, 2005.

[9] F. Harirchi, S. Z. Yong, E. Jacobsen, and N. Ozay, "Active model discrimination with applications to fraud detection in smart buildings," in *IFAC WC, Toulouse, France*, 2017.

[10] M. W. Hofbaur and B. C. Williams, "Hybrid estimation of complex systems," *IEEE Trans. on Systems, Man, and Cybernetics-B*, vol. 34, no. 5, pp. 2178–2191, 2004.

[11] S. L. Campbell and R. Nikoukhah, *Auxiliary signal design for failure detection*. Princeton University Press, 2015.

[12] R. Nikoukhah and S. L. Campbell, "Auxiliary signal design for active failure detection in uncertain linear systems with a priori information," *Automatica*, vol. 42, no. 2, pp. 219–228, 2006.

[13] F. Harirchi, Z. Luo, and N. Ozay, "Model (in) validation and fault detection for systems with polynomial state-space models," in *ACC*, 2016, pp. 1017–1023.

[14] A. Balluchi, L. Benvenuti, M. D. Di Benedetto, and A. L. Sangiovanni-Vincentelli, "Design of observers for hybrid systems," in *HSCC*. Springer, 2002, pp. 76–89.

[15] R. N. Clark, "Instrument fault detection," *IEEE Trans. on Aerospace and Electronic Systems*, no. 3, pp. 456–465, 1978.

[16] Z. Wang, D. Anand, J. Moyne, and D. Tilbury, "Improved sensor fault detection, isolation, and mitigation using multiple observers approach," *Systems Science & Control Engineering*, vol. 5, no. 1, pp. 70–96, 2017.

[17] A. Emami-Naeini, M. M. Akhter, and S. M. Rock, "Effect of model uncertainty on failure detection: the threshold selector," *IEEE Trans. on Automatic Control*, vol. 33, no. 12, pp. 1106–1115, 1988.

[18] A. Girard, "Reachability of uncertain linear systems using zonotopes," in *HSCC*, vol. 5. Springer, 2005, pp. 291–305.

[19] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *1st Workshop on SCS*, 2010.

[20] D. Bertsimas and F. J. de Ruiter, "Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds," *INFORMS Journal on Computing*, vol. 28, no. 3, pp. 500–511, 2016.