Synthesis of Stealthy Deception Attacks with Limited Resources

Uday Shankar <us@cmu.edu> (Supervised by: Eunsuk Kang <eskang@cmu.edu>)

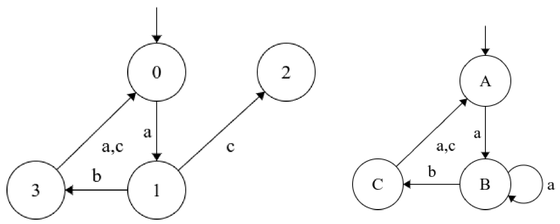
Carnegie Mellon University

Abstract

We study the security of Cyber-Physical Systems (CPS) with respect to attacks on the supervisory control layer. Specifically, we consider the sensory deception attacker model, originally proposed in [1], in which attackers aim to coerce the system into an unsafe state by altering the sensor readings received by the supervisor, while also concealing their presence from the supervisor. We modify this model by additionally placing a bound on the number of modifications made by the attacker. We present a method for the synthesis of such attackers, constructing and using an *Insertion Deletion Attack* (IDA) structure as our key technical tool.

Motivation

Cyber-Physical Systems (CPS), defined as systems where there is an interaction between computational and physical entities, are omnipresent. A typical example is the anti-lock braking system (ABS) installed in every modern car. Many CPS, such as the ABS, appear in *safety-critical* applications. In such applications, we cannot compromise on the security of our CPS designs. By studying the details of attacks on CPS and providing methods for attacker synthesis, we hope to enable CPS designers to establish provable security guarantees on their systems.



(a) System (b) Supervisor

Figure 1: A system, along with a supervisor that ensures the system never enters the unsafe state 2. Example provided by the authors of [1].

System Model

We consider CPS that can be modeled as discrete event systems. Some details of this model include:

- **Systems** are modeled as finite automata, with *states* and *transitions* between states labeled by *events*.
- **Supervisors** exist to prevent systems from reaching some unsafe states by selectively allowing or blocking events from affecting the system. They are also modeled as finite automata.

As an example, we can consider the system G and supervisor R depicted in figure 1. The system alone will be in state 2 if it executes the string of events ac, but under the control of the supervisor, state 2 is unreachable.

Attacker Model

We consider **sensory deception attackers**. Such attackers sit between the system and supervisor, as shown in figure 2. They have two main capabilities:

- They can **insert** bogus events, causing the supervisor to think that the system executed an event when it actually did not.
- They can **delete** events, causing the supervisor to think that the system did not execute an event when it actually did.

The actions of the attacker are limited to events in some subset Σ_a of the set of all events.

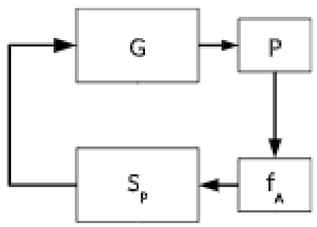


Figure 2: Interaction of the system G, the attacker f_A and the supervisor S_p. P is a projection that is relevant only for partially observed systems. Image from [1].

The Problem

The goal of the attacker is to confuse the supervisor into allowing the system to reach an unsafe state. We also impose two additional conditions on the attacker:

- **Stealth:** the attacker should not reveal their presence to the supervisor.
- **Modification Cap:** the attacker should insert at most m and delete at most n events.

The problem is to synthesize an attacker meeting the above conditions.

Building the IDA

We build a graph (called an IDA) to help the attacker make informed decisions. The nodes are classified as S**-states** (supervisor) or E**-states** (attacker). They store the following information.

- Set of possible current system states.
- Current supervisor state.
- Number of insertions left.
- Number of deletions left.

Each edge goes between two nodes of different types, and is labeled by the action of the relevant party (supervisor or attacker). The label of an edge coincides with its effect on the information stored in its endpoints (see figure 3). **Stealth** is achieved by applying the BSCP algorithm [2], which removes "bad decisions" in the graph that may cause detection.

Current Results

- A formal definition of a stealthy attacker with limited resources
- A method of construction of the IDA which accounts for partial controllability and observability.
- Proofs of correctness for the construction.
- An implementation which can build IDAs like figure 3.

Future Work

- Generalization to arbitrary costs for each attacker action.
- Define other measures of complexity for attackers.
- Synthesize minimally-complex attackers from IDAs.

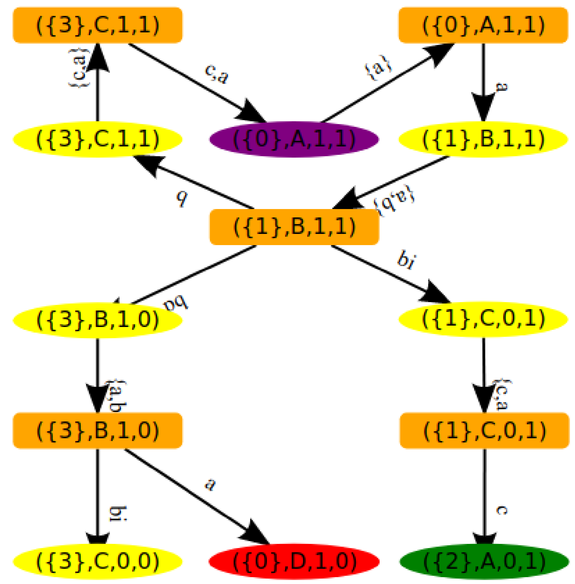


Figure 3: A slightly abbreviated IDA constructed on the example system and supervisor with $m = n = 1$, $\Sigma_a = \{b\}$. Key: S-state, $E-state$, initial state, attacker detected, and system compromise. Ex: edge from $(\{1\}, B, 1, 1)$ to $(\{1\}, C, 0, 1)$ is labeled b_i, and inserting the event b does not change the system state, changes the supervisor state from B to C, and consumes an insertion.

References

[1] R. M. Góes, E. Kang, R. Kwong, and S. Lafortune, "Stealthy deception attacks for cyber-physical systems," *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 4224–4230, 2017.

[2] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Springer Publishing Company, Incorporated, 2nd ed., 2010.