A Robust Reachability Review for Control System Security

Adrian N. Bishop

Abstract— **Control systems underpin the core technology in numerous critical infrastructure systems; e.g. the electricity, transportation and many defence systems. Increasingly, such systems are becoming the target of a novel type of deliberate cyber-attack. The notion of control system security is a modern idea concerned with the analysis, design and application of tools that ensure the operational goals of a particular control systems are protected from deliberate, malicious, electronic attack. The aim of this field is to reduce the likelihood of success, and the severity of impact, of a cyber-attack against control systems operating within critical infrastructure. This contribution re-examines a classical notion of control reachability through set-theoretic arguments with an additional, modern, emphasis on control system security. In particular, the reachability idea is extended to a compromised control system architecture. Classical and novel results on reachability are defined within this setting from the point-of-view of an attacker and, conversely, a system designer wanting to secure the system. The idea of reachability studied in this setting for secure control is important in both the design of robust control systems with security features and in assessing the vulnerability of particular control systems.**

I. Introduction

The problem of networked estimation and control has been considered extensively in recent years; see the expositions in [1]–[3]. This work considers *security for networked estimation and control systems*; or so-called networked cyber-physical systems. There are many networked control systems that operate critical infrastructure within strategic resource sectors[1]. The reliability and security of most systems of this nature is critical. However, the distributed and networked nature of the system often makes it vulnerable to a deliberate attack; e.g. many distributed cyber-physical systems are controlled via unsecured communication networks that may also be accessible through the world-wide-web [4]. Moreover, there are many sensitive applications involving networked control systems, e.g. power and transportation networks, and such applications are appealing targets for many malicious entities. A disruption or disconnection of these services may have significant social, environmental and/or economic consequences. Secure networked estimation and control is thus an important emerging problem[2].

A.N. Bishop is with NICTA and the Australian National University.

[1]Examples of so-called networked cyber-physical systems include critical infrastructure such as transportation networks, power and electricity grids, water distribution networks, etc. In these cases, the networked components may be spatially separated.

[2]An example of an attack on a real-world networked control system is the highly sophisticated *Stuxnet* attack on the Iranian nuclear energy program [5]. Stuxnet is a worm that spreads indiscriminately across computer networks running Microsoft Windows causing no disruption but which also includes a payload designed to target Siemens supervisory control and data acquisition (SCADA) systems configured to control and monitor very specific industrial processes (most likely certain centrifuges). In particular, Stuxnet infects the programmable logic controllers (PLC) in a SCADA system by subverting the application used to reprogram these devices. It then disrupts the operation of certain, very specific, industrial processes being driven by the controller. Stuxnet is so sophisticated that the general consensus is that it was state-sponsored and the first such act of industrial cyber-warfare [5]. Other publicised, but less dramatic, examples of intrusions and attacks on networked control systems exist; e.g. [6], [7].

The idea of secure control and estimation systems is becoming increasingly important, e.g. as a result of significant publicised attacks and the realization that a compromised system operating in a key sector may have significant social, environmental and/or economic consequences. Early studies in [4], [8] outline a framework for discussing and investigating the security of networked control and estimation systems. In [4] two classes of attack are characterized: namely *denial-of-service (DOS) attacks* where the attacker prevents the transmission of, e.g., measurement or control signals, and; *deceptive attacks* where a false (or distorted) signal is inserted into the system to disrupt the behaviour of the estimator or controller.

The problem of robust control in the presence of DOS attacks has been investigated in [8]. This work has similarities with those networked control problems that are designed to cope with limited-data rate communications and packet-dropping. However, the latter work often assumes some random model of packet failure whereas any DOS attack may be far more systematic. The problem of secure control and estimation under false-data attacks has been investigated. In [9], [10] the idea of reachability is used to analyze the security of control systems and their resilience to attack. In [11] a necessary and sufficient condition is provided under which an attacker can destabilize a system nominally controlled by an optimal control law while remaining undetected by a particular detector.

Alternatively, the problem of secure estimation and control in networked systems under false-data attacks has attracted significant attention recently; e.g. see [11]–[27]. At a high-level, the majority of this work addresses the problem of state estimation with inherent detection of false-data attacks or, conversely, a characterization of those attacks which are likely (under some modelling and algorithmic assumptions) to remain undetected in certain state estimation schemes. The classical field of fault-detection and fault-tolerant control; see [28], is related to this work also.

This contribution examines a classical notion of control reachability through set-theoretic arguments with a modern emphasis on control system security. In particular, the reachability idea is extended to a compromised control system architecture. Classical and novel results on reachability are defined within this setting from the point-of-view of an attacker and, conversely, a system designer wanting to secure the system. The idea of reachability studied in this setting for secure control is important in both the design of robust control systems with security features and in assessing the vulnerability of particular control systems.

II. A General Introduction on Robust Reachability

The introduction in this section follows the classical results in [29]–[31]. However, the notation and set-theoretic reachability results are stated, and derived, in a different fashion that suits the later framework of control security.

Consider the time-varying discrete-time system

$$\mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k) + g_k(\mathbf{w}_k) \tag{1}$$

defined for $k \in \mathbb{N} \cup \{0\} = \mathbb{N}_0$ and where $\mathbf{x}_k \in \mathcal{X}_k \subseteq \mathcal{X} \subseteq \mathbb{R}^n$ is the *state*, $\mathbf{u}_k \in \mathcal{U}_k \subseteq \mathcal{U} \subseteq \mathbb{R}^m$ is the *control vector* and $\mathbf{w}_k \in \mathcal{W}_k \subseteq \mathcal{W} \subseteq \mathbb{R}^l$ is the *system uncertainty input* or error. The functions $f_k : \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^n$ and $g_k : \mathcal{W}_k \to \mathbb{R}^n$ are known and

ISBN 978-0-85825-987-4

$f_k + g_k : \mathcal{W}_k \times \{\mathbb{R}^n \times \mathcal{U}\} \to \mathcal{X}_{k+1}$. The system (1) describes the physical system under consideration.

The following definition makes matters precise.

Definition 1. *Suppose $\mathcal{X}_0$ is given and fixed. Suppose a sequence of control functions are specified of the form $\mathbf{u}_k = u_k(\mathbf{x}_k) : \mathcal{X}_k \to \mathcal{U}_k$ where*

$$\mathcal{X}_{k+1} = \{\mathbf{x}_{k+1} \in \mathbb{R}^n : \mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k) + g_k(\mathbf{w}_k), \\ \forall \mathbf{x}_k \in \mathcal{X}_k,\ \forall \mathbf{u}_t \in \mathcal{U}_t,\ \forall \mathbf{w}_t \in \mathcal{W}_t,\ \forall t \in \{0, \ldots, k\}\}$$

*is then called the **local set of admissible state values under the sequence** $\{\mathcal{U}_t\}_{t=0}^k$ at time k for all $k > 0$.*

If we suppose that $\mathbf{u}_k$ take any value in $\mathcal{U}$ at time k and suppose $\mathcal{X}_0$ is given and fixed then

$$\mathfrak{X}_{k+1} = \{\mathbf{x}_{k+1} \in \mathbb{R}^n : \mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k) + g_k(\mathbf{w}_k), \\ \forall \mathbf{x}_k \in \mathfrak{X}_k,\ \forall \mathbf{u}_t \in \mathcal{U},\ \forall \mathbf{w}_t \in \mathcal{W}_t,\ \forall t \in \{0, \ldots, k\}\}$$

*is called the **global set of admissible state values** at time k for all $k > 0$.*

It is clear that $\mathcal{X}_k \subseteq \mathfrak{X}_k$. Note $\mathcal{X} = \cup_k \mathcal{X}_k \subseteq \mathbb{R}^n$ is defined by a particular feedback controller whereas $\mathcal{U}$ and $\mathcal{W}$ are defined by the physical properties of the system as expected. Obviously $\mathcal{X} \subseteq \cup_k \mathfrak{X}_k$.

The general control problem under consideration is stated as follows.

Problem 1. *Consider a time $0 < \tau \in \mathbb{N}$ and suppose $\mathcal{X}_0$ is given and fixed. Design a control function $\mathbf{u}_k = u_k(\mathbf{x}_k) : \mathcal{X}_k \to \mathcal{U}_k$ mapping $\mathbf{x}_k$ into $\mathcal{U}_k$, for all $k \in \{0, \ldots, \tau - 1\}$ such that the following holds*

$$f_{\tau-1} + g_{\tau-1} : \mathcal{W}_{\tau-1} \times \{\mathcal{X}_{\tau-1} \times \mathcal{U}_{\tau-1}\} \to \mathcal{X}_\tau^* \tag{2}$$

for all $\mathbf{w}_k \in \mathcal{W}_k$, $\forall k \in \{0, \ldots, \tau-1\}$. The set $\mathcal{X}_\tau^* \in \mathbb{R}^n$, $0 < \tau \in \mathbb{N}$ constitutes the desired, or target, set at time τ for the controlled system (1) and should be known to the control designer a priori.

A question in robust control then concerns the solvability of Problem 1. We introduce the following definition.

Definition 2. *A set $\mathcal{X}_\tau^* \in \mathbb{R}^n$, $0 < \tau \in \mathbb{N}$ is said to be **reachable given** $\mathcal{X}_0$ if there exists, at least, a single sequence of control functions $u_k : \mathcal{X}_k \to \mathcal{U}_k$, $\forall k \in \{0, \ldots, \tau - 1\}$ that solves Problem 1.*

Obviously $\mathcal{X}_\tau^* \subseteq \mathfrak{X}_\tau$ is a necessary condition for reachability, however, it is not necessarily true that every subset of $\mathfrak{X}_\tau$ is a reachable subset given some $\mathcal{X}_0$. Admissible is not equivalent to reachable.

Consider a set

$$\mathcal{T}_k = \{\mathbf{x} \in \mathcal{Q}_k : f_k(\mathbf{x}, \mathbf{u}_k) \in \mathcal{E}_{k+1}, \\ \text{for some } \mathbf{u}_k \in \mathcal{U}_k\} \tag{3}$$

where

$$\mathcal{E}_{k+1} = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} + g_k(\mathbf{w}_k)) \in \mathcal{Q}_{k+1}, \\ \forall \mathbf{w}_k \in \mathcal{W}_k\} \tag{4}$$

with boundary conditions $\mathcal{Q}_\tau = \mathcal{X}_\tau^*$ and $\mathcal{Q}_0 = \mathcal{X}_0$. Suppose $\mathcal{T}_0 \neq \emptyset$. Then $\mathbf{x}_0 \in \mathcal{T}_0 \subseteq \mathcal{X}_0$ implies $\mathbf{x}_1 \in \mathcal{Q}_1$. Similarly, if $\mathcal{T}_1 \neq \emptyset$ then $\mathbf{x}_1 \in \mathcal{T}_1 \subseteq \mathcal{Q}_1$ implies $\mathbf{x}_2 \in \mathcal{Q}_2$ etc.

Proposition 1. *The target set $\mathcal{X}_\tau^*$, $0 < \tau \in \mathbb{N}$ is reachable from all $\mathbf{x}_0 \in \mathcal{X}_0$ if and only if $\mathcal{Q}_k \subseteq \mathcal{T}_k$ for all $k \in \{0, \ldots, \tau - 1\}$.*

The sets $\mathcal{T}_k$ and $\mathcal{E}_k$ are, in principal, computable and can, in special cases, lead to the closed-form design of control functions $\mathbf{u}_k$ that ensure the target set is indeed reached.

Problem 2. *Consider a time $0 < \tau \in \mathbb{N}$ and suppose $\mathcal{X}_0$ is given and fixed. Design a control function $\mathbf{u}_k = u_k(\mathbf{x}_k) : \mathcal{X}_k \to \mathcal{U}_k$ mapping $\mathbf{x}_k$ into $\mathcal{U}_k$, for all $k \in \{0, \ldots, \tau - 1\}$ such that the following holds*

$$f_{k-1} + g_{k-1} : \mathcal{W}_{k-1} \times \{\mathcal{X}_{k-1} \times \mathcal{U}_{k-1}\} \to \mathcal{X}_k^* \tag{5}$$

for all $k \in \{1, \ldots, \tau\}$ and $\forall \mathbf{w}_k \in \mathcal{W}_k$. The sequence of sets $\{(\mathcal{X}_k^*, k)\}_{k=1}^\tau$, $0 < \tau \in \mathbb{N}$ constitutes a desired, or target, tube in $\mathbb{R}^n \times \mathbb{N}_1^\tau$ for the controlled system (1) and should be known to the control designer a priori.

We introduce the following definition.

Definition 3. *A sequence of sets $\{\mathcal{X}_k^*\}_{k=1}^\tau$, $0 < \tau \in \mathbb{N}$ is said to be **reachable given** $\mathcal{X}_0$ if there exists, at least, a single sequence of control functions $u_k : \mathcal{X}_k \to \mathcal{U}_k$, $\forall k \in \{0, \ldots, \tau - 1\}$ that solves Problem 2.*

In a different, but equivalent, notation we have it that $u_k : \mathcal{X}_k^* \to \mathcal{U}_k$ when $k \in \{1, \ldots, \tau - 1\}$.

Consider a set

$$\mathcal{T}_k^* = \{\mathbf{x} \in \mathcal{X}_k^* : f_k(\mathbf{x}, \mathbf{u}_k) \in \mathcal{E}_{k+1}^*, \\ \text{for some } \mathbf{u}_k \in \mathcal{U}_k\} \tag{6}$$

where

$$\mathcal{E}_{k+1}^* = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} + g_k(\mathbf{w}_k)) \in \mathcal{X}_{k+1}^*, \\ \forall \mathbf{w}_k \in \mathcal{W}_k\} \tag{7}$$

where we abuse notation and let $\mathcal{X}_0^* = \mathcal{X}_0$. Suppose $\mathcal{T}_0^* \neq \emptyset$. Then $\mathbf{x}_0 \in \mathcal{T}_0^* \subseteq \mathcal{X}_0$ implies $\mathbf{x}_1 \in \mathcal{X}_1^*$. Similarly, if $\mathcal{T}_1^* \neq \emptyset$ then $\mathbf{x}_1 \in \mathcal{T}_1^* \subseteq \mathcal{X}_1^*$ implies $\mathbf{x}_2 \in \mathcal{X}_2^*$ etc.

Proposition 2. *The sequence of target sets $\{\mathcal{X}_k^*\}_{k=1}^\tau$, $0 < \tau \in \mathbb{N}$ is reachable from all $\mathbf{x}_0 \in \mathcal{X}_0$ if and only if $\mathcal{T}_k^* \neq \emptyset$ for all $k \in \{0, \ldots, \tau - 1\}$ and $\mathcal{T}_0^* = \mathcal{X}_0$.*

III. A Compromised System Architecture

The system introduced in the previous section is depicted in Figure 1.

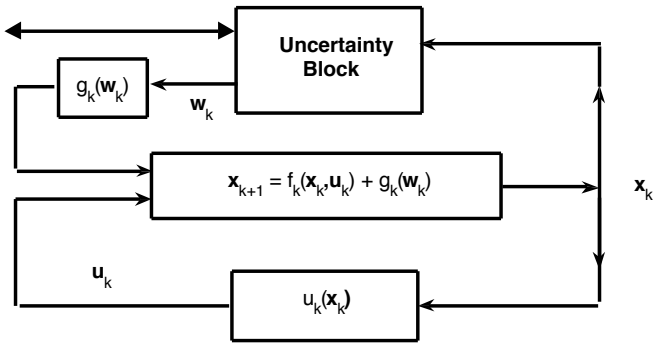

Fig. 1. A nominal robust control system.

However, suppose now an attacking, or malicious, agent places itself within the system architecture as shown in Figure 2.

The corresponding compromised time-varying discrete-time system is then

$$\widehat{\mathbf{x}}_{k+1} = f_k(\widehat{\mathbf{x}}_k, \mathbf{z}_k) + g_k(\mathbf{w}_k) + \mathbf{v}_k \tag{8}$$

defined for $k \in \mathbb{N} \cup \{0\} = \mathbb{N}_0$ and where $\widehat{\mathbf{x}}_k \in \widehat{\mathcal{X}}_k \subseteq \widehat{\mathcal{X}} \subseteq \mathbb{R}^n$ and $\mathbf{w}_k \in \mathcal{W}_k \subseteq \mathcal{W} \subseteq \mathbb{R}^l$ are the state and disturbance vectors. Now $\mathbf{z}_k \in \mathcal{Z}_k \subseteq \mathcal{Z} \subseteq \mathcal{U} \subseteq \mathbb{R}^m$ is the *compromised control*

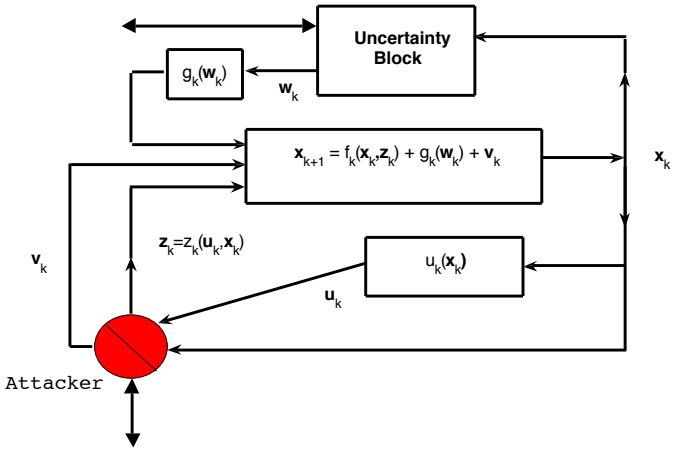

Fig. 2. A compromised robust control system. Note we will substitute $\widehat{\mathbf{x}}_k$ for $\mathbf{x}_k$ in the text of this paper in order to distinguish the trajectories of a nominal system and the compromised system for some fixed sequence of $\mathbf{u}_k$ and $\mathbf{w}_k$.

vector and $\mathbf{v}_k \in \mathcal{V}_k \subseteq \mathcal{V} \subseteq \mathbb{R}^n$ is an *explicit attack vector*[3]. Note that we suppose $\mathcal{U}$ is the entire set of admissible control inputs to the system and thus $\mathcal{Z} \subseteq \mathcal{U}$ makes sense. Thus, the functions $f_k : \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^n$ and $g_k : \mathcal{W}_k \to \mathbb{R}^n$ are known as before and

$$f_k + g_k + \mathbf{v}_k : \mathcal{W}_k \times \{\mathbb{R}^n \times \mathcal{U}\} \times \mathcal{V}_k \to \widehat{\mathcal{X}}_{k+1} \tag{9}$$

as expected. The following definition is given for completeness.

Definition 4. *Suppose $\widehat{\mathcal{X}}_0$ is given and fixed. Suppose a sequence of compromised controller functions are specified of the form $\mathbf{z}_k = z_k(\widehat{\mathbf{x}}, \mathbf{u}_k) : \widehat{\mathcal{X}}_k \times \mathcal{U}_k \to \mathcal{Z}_k$ where $\mathbf{u}_k = u_k(\widehat{\mathbf{x}}_k) : \widehat{\mathcal{X}}_k \to \mathcal{U}_k$ is a sequence of fixed nominal control vectors and where*

$$\begin{aligned}\widehat{\mathcal{X}}_{k+1} = \{\widehat{\mathbf{x}}_{k+1} \in \mathbb{R}^n : \widehat{\mathbf{x}}_{k+1} = f_k(\widehat{\mathbf{x}}_k, \mathbf{z}_k) + \\ \forall \widehat{\mathbf{x}}_k \in \widehat{\mathcal{X}}_k,\ g_k(\mathbf{w}_k) + \mathbf{v}_k,\ \ \forall \mathbf{z}_t \in \mathcal{Z}_t,\ \forall \mathbf{u}_t \in \mathcal{U}_t, \\ \forall \mathbf{w}_t \in \mathcal{W}_t,\ \forall \mathbf{v}_t \in \mathcal{V}_t,\ \forall t \in \{0, \dots, k\}\}\end{aligned}$$

*is then called the **local set of admissible compromised state values under the sequence** $\{\mathcal{Z}_t\}_{t=0}^k$ **and** $\{\mathcal{U}_t\}_{t=0}^k$ at time k for all $k > 0$.*

If we suppose that z_k take any value in $\mathcal{U}$ at time k and suppose $\widehat{\mathcal{X}}_0$ is given and fixed then

$$\begin{aligned}\widehat{\mathfrak{X}}_{k+1} = \{\widehat{\mathbf{x}}_{k+1} \in \mathbb{R}^n : \widehat{\mathbf{x}}_{k+1} = f_k(\widehat{\mathbf{x}}_k, \mathbf{z}_k) + \\ \forall \widehat{\mathbf{x}}_k \in \widehat{\mathfrak{X}}_k,\ g_k(\mathbf{w}_k) + \mathbf{v}_k,\ \forall \mathbf{z}_t \in \mathcal{U},\ \forall \mathbf{w}_t \in \mathcal{W}_t, \\ \forall \mathbf{v}_t \in \mathcal{V}_t,\ \forall t \in \{0, \dots, k\}\}\end{aligned}$$

*is called the **global set of admissible compromised state values** at time k for all $k > 0$.*

The following result is given for completeness.

Proposition 3. *Suppose $\mathcal{X}_0 = \widehat{\mathcal{X}}_0$. Then $\mathfrak{X}_k \subseteq \widehat{\mathfrak{X}}_k$ and $\mathfrak{X}_k = \widehat{\mathfrak{X}}_k$ if and only if $\mathbf{v}_t = \mathbf{0}$ for all $0 \leq t \leq k-1$.*

In general, $\mathfrak{X}_k \subset \widehat{\mathfrak{X}}_k$ implies the malicious agent may be able to distort the trajectory of the system into alternative regions of the state space $\mathbb{R}^n$ than are otherwise admissible.

[3]The main analysis will neglect such an input. This is not unreasonable since in practice an attacker is likely to systematically gain access to and distort only the nominal control input.

IV. When Can an Attack Compromise the Safety of the System?

In this section we will consider the notion of a safe operating region for a particular control system.

Definition 5. *Consider a sequence of regions $\mathcal{S}_k \subset \mathfrak{X}_k = \widehat{\mathfrak{X}}_k \subseteq \mathbb{R}^n$ for all k. We call $\mathcal{S}_k$ the **safe operating region at** k for the nominal system (1) and if $\mathbf{x}_k \notin \mathcal{S}_k$ we say the **safety of the system is compromised**.*

Note that $\cup_k \mathcal{S}_k \subset \cup_k \mathfrak{X}_k = \cup_k \widehat{\mathfrak{X}}_k \subseteq \mathbb{R}^n$. Before we look at the damage an attacker can cause under some modelling assumptions, we first look at the nominal control problem again, i.e. Problem 2.

Define a set

$$\mathcal{U}_k^+ = \{\mathbf{u} \in \mathcal{U}_k : f_k(\mathbf{x}_k, \mathbf{u}) \in \mathcal{E}_{k+1}^*, \forall \mathbf{x}_k \in \mathcal{X}_k^*\} \tag{10}$$

where we abuse notation and let $\mathcal{X}_0^* = \mathcal{X}_0$. Recall

$$\begin{aligned}\mathcal{E}_{k+1}^* = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} + g_k(\mathbf{w}_k)) \in \mathcal{X}_{k+1}^*, \\ \forall \mathbf{w}_k \in \mathcal{W}_k\}\end{aligned} \tag{11}$$

We then have the following result.

Proposition 4. *The sequence of target sets $\{\mathcal{X}_k^*\}_{k=1}^\tau$, $0 < \tau \in \mathbb{N}$ is reachable for all $\mathbf{x}_0 \in \mathcal{X}_0$ if $\mathcal{U}_k^+ \neq \emptyset$, $\forall k$.*

The proof is omitted but note the condition $\mathcal{U}_k^+ \neq 0$ is only sufficient and not necessary to solve Problem 2. In some sense, the set $\mathcal{U}_k^+$ defined above neglects the control history $\mathbf{u}_t$, for $k \in \{0, \dots, k-1\}$. We can state a necessary and sufficient condition on a set of control inputs that ensures Problem 2 is solved.

Define the set

$$\begin{aligned}\mathcal{U}_k^* = \{\mathbf{u} \in \mathcal{U}_k : f_k(\mathbf{x}_k, \mathbf{u}) \in \mathcal{E}_{k+1}^*, \\ \forall \mathbf{x}_k \in \mathcal{F}_{k-1}(\mathbf{u}_{k-1})\}\end{aligned} \tag{12}$$

where

$$\mathcal{F}_k(\mathbf{u}_k) = f_k(\mathcal{X}_k, \mathbf{u}_k) \bigoplus g_k(\mathcal{W}_k) \tag{13}$$

where $\bigoplus$ is the set-theoretic Minkowski sum. We also let $\mathcal{F}_{-1} = \mathcal{X}_0$. Using $\mathcal{U}_k^*$ we can account for a particular history of control laws.

Proposition 5. *The sequence of target sets $\{\mathcal{X}_k^*\}_{k=1}^\tau$, $0 < \tau \in \mathbb{N}$ is reachable for all $\mathbf{x}_0 \in \mathcal{X}_0$ if and only if $\mathcal{U}_k^* \neq \emptyset$, $\forall k$.*

Obviously the following result follows from the preceding two propositions.

Corollary 1. $\mathcal{U}_k^+ \subseteq \mathcal{U}_k^*$

The set $\mathcal{U}_k^*$ is, in principal, computable and can, in special cases, lead to the closed-form design of control functions $\mathbf{u}_k$. Indeed, once $\mathcal{U}_k^*$ is computed, any $\mathbf{u}_k \in \mathcal{U}_k^*$ is sufficient. (As is typical, computation of these control input sets involves a backward recursion).

Now we examine when an attack can compromise the safety of a particular system. We make the following two assumptions.

Assumption 1. *Consider the two sets $\mathcal{N}_0^\theta$ and $\mathcal{N}_{\theta+1}^\tau$ and suppose $\mathcal{X}_0$ is given. There exists a **nominal control** $\mathbf{u}_k = u_k(\mathbf{x}_k) : \mathcal{X}_k \to \mathcal{U}_k$, $\forall k \in \{0, \dots, \tau-1\}$ such that $\mathcal{X}_k = \mathcal{X}_k^* \subseteq \mathcal{S}_k$ for all $k \in \{0, \dots, \tau\}$.*

The previous assumption states that under a nominal, i.e. non-compromised control, the system will operate within its desired target tube and within its safety region.

Assumption 2. *Consider the two sets $\mathcal{N}_0^\theta$ and $\mathcal{N}_\theta^\tau$ and suppose $\mathcal{X}_0$ is given. The attacker can hijack the nominal control inputs at times $k \in \{\theta, \dots, \tau-1\}$. The **compromised control** is $\mathbf{z}_k = z_k(\widehat{\mathbf{x}}, \mathbf{u}_k)$:*

$\widehat{\mathcal{X}}_k \times \mathcal{U}_k \to \mathcal{Z}_k$. *Also,* $\mathbf{v}_k \in \mathcal{V}_k = \{\mathbf{0}\}$ *implying* $\mathfrak{X}_k = \widehat{\mathfrak{X}}_k$ *for all* k *and* $\mathcal{X}_k = \widehat{\mathcal{X}}_k$.

This assumption implies that an attacker can hijack the control input at some time $k = \theta$. Before this time the system is nominally controlled.

Definition 6. *Consider a sequence of regions* $\mathcal{S}_k \subset \mathfrak{X}_k = \widehat{\mathfrak{X}}_k \subseteq \mathbb{R}^n$ *for all* k. *We say the* ***safety of the system is compromised by an attack*** *if* $\mathbf{x}_t \notin \mathcal{S}_t$ *at some* t *and* $\exists k \in \{0, \ldots, t-1\}$ *such that* $\mathbf{z}_k \neq \mathbf{u}_k$.

Note if the safety of the system is compromised by an attack then necessarily the state of the system is outside the desired target tube since $\mathcal{X}_k^* \subseteq \mathcal{S}_k$ for all k. However, the system may be steered outside the target tube by an attacker but remain within a safe operating region for the system.

Define a set $\mathcal{A}_k = \mathfrak{X}_k \setminus \mathcal{S}_k \subset \mathfrak{X}_k = \widehat{\mathfrak{X}}_k$. This set $\mathcal{A}_k$ is a pseudo-target set for an attacker wishing to compromise the safety of the system. For simplicity, we assume it is sufficient to consider a target set $\mathcal{A}_\tau = \mathfrak{X}_\tau \setminus \mathcal{S}_\tau \subset \mathfrak{X}_\tau = \widehat{\mathfrak{X}}_\tau$.

For $k \in \{\theta, \ldots, \tau-1\}$, consider a set

$$\mathcal{Z}_k^* = \{\mathbf{z} \in \mathcal{Z}_k : f_k(\mathbf{x}_k, \mathbf{z}) \in \mathcal{H}_{k+1}, \quad \forall \mathbf{x}_k \in \mathcal{F}_{k-1}(\mathbf{z}_{k-1})\} \tag{14}$$

where

$$\mathcal{H}_{k+1} = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} + g_k(\mathbf{w}_k)) \in \mathcal{Q}_{k+1}, \quad \forall \mathbf{w}_k \in \mathcal{W}_k\} \tag{15}$$

with the constraint $\mathcal{Q}_\tau = \mathcal{A}_\tau$ and

$$\mathcal{F}_k(\mathbf{z}_k) = f_k(\mathcal{X}_k, \mathbf{z}_k) \bigoplus g_k(\mathcal{W}_k) \tag{16}$$

as before. We also let $\mathcal{F}_{\theta-1}(\mathbf{z}_{\theta-1}) = \mathcal{F}_{\theta-1}(\mathbf{u}_{\theta-1})$ and note that one could think of the relationship $\mathcal{Z}_k^* = \mathcal{U}_k^*$ for all $k \in \{0, \ldots, \theta-1\}$.

Theorem 1. *The safety of the system at time* $k = \tau$ *can be compromised by an attacker initiating an attack at some time* $k = \theta < \tau$ *if and only if* $\mathcal{Z}_k^* \neq \emptyset$ *for all* $k \in \{\theta, \ldots, \tau-1\}$.

The set $\mathcal{Z}_k^*$ is, in principal, computable and can, in special cases, lead to the closed-form design of attack functions $\mathbf{z}_k$. Indeed, once $\mathcal{Z}_k^*$ is computed, any $\mathbf{z}_k \in \mathcal{Z}_k^*$ is sufficient to drive the system outside its safety region so long as the necessary and sufficient condition of the preceding theorem holds. (As is typical, computation of these attack input sets involves a backward recursion).

The preceding theorem defines a necessary and sufficient set-theoretic condition on the attacker's control law given an initial and terminal attack time θ and τ respectively. Note that the time length $\tau - \theta$ obviously plays a role in whether or not the condition of the preceding theorem will hold. This preceding result leads to a straightforward definition concerning robust, secure control.

Definition 7. *If* $\nexists \mathbf{x}_\theta \in \mathcal{F}_{\theta-1}(\mathbf{u}_{\theta-1})$ *such that* $\exists k \in \{\theta, \ldots, \tau-1\}$ *with* $\mathcal{Z}_k^* = \emptyset$ *then the system is said to be* ***securely controllable***.

Consider a set function

$$\mathcal{P}_k(\mathcal{N}, \mathcal{L}, \mathcal{M}) = \{\mathbf{x} \in \mathcal{N} : f_k(\mathbf{x}, \mathcal{L}) \bigoplus g_k(\mathcal{W}_k) \subseteq \mathcal{M}\} \tag{17}$$

of those $\mathbf{x} \in \mathcal{N}$ such that $(f_k(\mathbf{x}, \mathbf{u}) + g_k(\mathbf{w}_k)) \in \mathcal{M}$ for every input $\mathbf{u} \in \mathcal{L}$ and $\forall \mathbf{w}_k \in \mathcal{W}_k$.

Theorem 2. *Suppose at time* θ *that* $\mathbf{x}_\theta \in \mathcal{X}_\theta^* \subseteq \mathcal{S}_\theta$. *Let* $\mathcal{F}_{\theta-1}(\mathcal{Z}_{\theta-1}) = \mathcal{X}_\theta^*$ *for notational simplicity. If*

$$\mathcal{P}_t(\mathcal{F}_{t-1}(\mathcal{Z}_{t-1}) \cap \mathcal{S}_t, \mathcal{Z}_t, \mathcal{S}_{t+1}) \neq \mathcal{F}_{t-1}(\mathcal{Z}_{t-1}) \cap \mathcal{S}_t \tag{18}$$

for some $t \geq \theta$ *then the system is securely controllable so long as* $\tau \leq t$.

The preceding theorem is the main result and provides a condition for a system designer which results in a necessary and sufficient lower bound on the number of consecutive attack signals required to compromise the safety of the system.

Corollary 2. *Suppose at time* θ *that* $\mathbf{x}_\theta \in \mathcal{X}_\theta^* \subseteq \mathcal{S}_\theta$. *Let* $\mathcal{F}_{\theta-1}(\mathcal{Z}_{\theta-1}) = \mathcal{X}_\theta^*$ *for notational simplicity. Consider the optimization problem*

$$\tau^* = \underset{\tau}{\operatorname{argmax}} \ \tau \tag{19}$$
$$\text{such that } \mu\left(\bigcap_{t=\theta}^{\tau} (\mathcal{F}_{t-1}(\mathcal{Z}_{t-1}) \cap \mathcal{S}_t)\right) - \mu\left(\bigcap_{t=\theta}^{\tau} \mathcal{P}_t(\mathcal{F}_{t-1}(\mathcal{Z}_{t-1}) \cap \mathcal{S}_t, \mathcal{Z}_t, \mathcal{S}_{t+1}\right) > 0$$

where the operation $\mu(\cdot)$ *gives the Lebesque measure of its argument. Then there exists a suitable attack sequence of length* $\tau^* - \theta$ *that can compromise the safety of the system at time* τ^*. *Conversely, there is no attack sequence of a shorter duration that can compromise the safety of the system.*

The preceding two results, for example, provide computable functions on the known system parameters, e.g. the model and the input constraints etc, which can aid the system designer in securing the control system. In practice, if an attacker wanted to drive the state outside some safety region and the attacker had an arbitrarily long time period within which to distort the system controller then (apart from some special cases) it is unreasonable to expect this to be impossible.

The preceding two results lead to a necessary and sufficient lower bound on the time period in which an attacker can take the state from a safety region to a unsafe region. The system designer can then design an attack detection algorithm or a monitoring process etc that attempts to verify the integrity of the controller within this time period. It has also been discussed in, e.g., [8] that an attacker may only intermittently be able to distort a control signal. Thus, a designer may be able to configure the system such that the lower bound introduced in this paper is increased.

V. Concluding Remarks

This paper briefly explored the classical notion of control system reachability as applied within a secure control theoretic framework.

Note that all the set-theoretical results introduced in this paper are, in principal, computable; e.g. it is possible to design controller and attack input sets that drive the state of the system to the respective target sets (so long as those sets are reachable). For linear time-varying systems, where the control, attack and disturbance sets are ellipsoidal etc, the set-theoretical results introduced in this work can be well-approximated using ellipsoidal calculus.

Acknowledgement

This work is supported by NICTA which is funded by the Australian Government via the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. This work is also supported by the US Air Force via the Asian Office of Aerospace Research and Development (AOARD) and contract USAF-AOARD-10-4102.

References

[1] G.N. Nair, F. Fagnani, S. Zampieri, and R.J. Evans. Feedback control under data rate constraints: an overview. *Proceedings of the IEEE*, 95(1):108–137, March 2007.

[2] J.P. Hespanha, P. Naghshtabrizi, and Y. Xu. A survey of recent results in networked control systems. *Proceedings of the IEEE*, 95(1):138–162, March 2007.

[3] A.S. Matveev and A.V. Savkin. *Estimation and Control over Communication Networks*. Birkhauser, Boston, 2008.

[4] A.A. Cárdenas, S. Amin, and S.S. Sastry. Secure control: Towards survivable cyber-physical systems. In *Proceedings of the 1st International Workshop on Cyber-Physical Systems*, pages 495–500, Beijing, China, June 2008.

[5] K. O'Brien. A silent attack, but not a subtle one. *The New York Times*, September 26 2010.

[6] J. Meserve. Staged cyber attack reveals vulnerability in power grid. *CNN*, September 26 2007.

[7] S. Gorman. Electricity grid in U.S. penetrated by spies. *The Wall Street Journal*, April 8 2009.

[8] S. Amin, A.A. Cárdenas, and S.S. Sastry. Safe and secure networked control systems under denial-of-service attack. In *Hybrid Systems: Computation and Control*, pages 31–45. Lecture Notes in Computer Science, Springer, April 2009.

[9] P. Mohajerin, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. Cyber attack in a two-area power system: Impact identification using reachability. In *Proc. of the 2010 American Control Conference*, Baltimore, MD, June 2010.

[10] P. Mohajerin, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. A robust policy for automatic generation control cyber attack in two area power network. In *Proc. of the 49th IEEE Conf. on Decision and Control*, pages 5973–5978, Atlanta, GA, December 2010.

[11] Y. Mo and B. Sinopoli. False data injection attacks in cyber physical systems. In *Proceedings of the 1st Workshop on Secure Control Systems*, Stockholm, Sweden, April 2010.

[12] S. Sundaram and C.N. Hadjicostis. Distributed function calculation via linear iterations in the presence of malicious agents - part I: Attacking the network. In *Proceedings of the 2008 American Control Conference*, pages 1350–1355, Seatle, WA, USA, June 2008.

[13] S. Sundaram and C.N. Hadjicostis. Distributed function calculation via linear iterations in the presence of malicious agents - part II: Overcoming malicious behavior. In *Proceedings of the 2008 American Control Conference*, pages 1356–1361, Seatle, WA, USA, June 2008.

[14] Y. Liu, M.K. Reiter, and P. Ning. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, pages 21–32, November.

[15] R.B. Bobba, K.M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T.J. Overbye. Detecting false data injection attacks on DC state estimation. In *Proceedings of the 1st Workshop on Secure Control Systems*, Stockholm, Sweden, April 2010.

[16] H. Sandberg, A. Teixeira, and K.H. Johansson. On security indices for state estimators in power networks. In *Proceedings of the 1st Workshop on Secure Control Systems*, Stockholm, Sweden, April 2010.

[17] O. Kosut, L. Jia, R.J. Thomas, and L. Tong. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In *Proc. of the 1st IEEE Int. Conference on Smart Grid Communications*, pages 220–225, Gaithersburg, MD, USA, October 2010.

[18] G. Dán and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *Proceedings of the 1st IEEE International Conference on Smart Grid Communications*, pages 214–219, Gaithersburg, MD, USA, October 2010.

[19] B. Zhu and S. Sastry. SCADA-specific intrusion detection/prevention systems: A survey and taxonomy. In *Proceedings of the 1st Workshop on Secure Control Systems*, Stockholm, Sweden, April 2010.

[20] Le Xie, Yilin Mo, and B. Sinopoli. False data injection attacks in electricity markets. In *Proceedings of the 1st IEEE International Conference on Smart Grid Communications*, pages 226–231, Gaithersburg, MD, USA, October 2010.

[21] A. Teixeira, H. Sandberg, and K.H. Johansson. Networked control systems under cyber attacks with applications to power networks. In *Proc. of the 2010 American Control Conference*, pages 3690–3696, Baltimore, MD, June 2010.

[22] F. Pasqualetti, R. Carli, A. Bicchi, and F. Bullo. Identifying cyber attacks via local model information. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 5961–5966, Atlanta, GA, USA, December 2010.

[23] S. Sundaram, M. Pajic, C.N. Hadjicostis, R. Mangharam, and G.J. Pappas. The wireless control network: Monitoring for malicious behavior. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 5979–5984, Atlanta, GA, USA, December 2010.

[24] A. Teixeira, S. Amin, H. Sandberg, K.H. Johansson, and S.S. Sastry. Cyber security analysis of state estimators in electric power systems. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 5991–5998, Atlanta, GA, USA, December 2010.

[25] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 5967–5972, Atlanta, GA, USA, December 2010.

[26] S. Zheng, T. Jiang, and J.S. Baras. Robust state estimation under false data injection in distributed sensor networks. In *Proceedings of the 2010 IEEE Global Telecommunications Conference*, Miami, FL, USA, December 2010.

[27] A.N. Bishop and A.V. Savkin. On false-data attacks in robust multi-sensor-based estimation. In *Proceedings of the 9th IEEE International Conference on Control and Automation*, Santiago, Chile, December 2011.

[28] R. Isermann. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Verlag, 2006.

[29] H. Witsenhausen. A minimax control problem for sampled linear systems. *IEEE Transactions on Automatic Control*, 13(1):5–21, February 1968.

[30] D.P. Bertsekas and I.B. Rhodes. On the minimax reachability of target sets and target tubes. *Automatica*, 7(2):233–247, March 1971.

[31] J. Glover and F. Schweppe. Control of linear dynamic systems with set constrained disturbances. *IEEE Transactions on Automatic Control*, 16(5):411–423, October 1971.