Secure Detection Using Binary Sensors

Rohan Chabukswar Yilin Mo Bruno Sinopoli

Carnegie Mellon University, Pittsburgh, PA 15213 USA
(e-mail: rchabuks@andrew.cmu.edu, ymo@andrew.cmu.edu, brunos@ece.cmu.edu)

Abstract: Cyber-physical systems employing remote sensors and actuators and sparse communication networks are pervading the infrastructure. In this paper we consider a prototypical problem of estimating a binary state using measurements provided by binary sensors. We propose a new approach to estimate the states based on sensor measurements that may have been corrupted by an attacker. The problem is formulated as a minimax problem in which a detector attempts to maximize the probability of detection in case of the worst case attempt by the attacker to minimize this probability. A fixed form of the detector is proposed in the case where the sensors are of equivalent specifications, along with a method to find the optimal detector parameters.

Keywords: Cyber-Physical Systems, SCADA, Secure, Control

1. INTRODUCTION

Cyber-Physical systems (CPS) often employ distributed networks of embedded sensors and actuators (Lee (2008)) that interact with the physical environment, and are monitored and controlled by a Supervisory Control and Data Acquisition (SCADA) system. Distributed sensors and actuator networks are often seen in varied applications, such as critical infrastructure monitoring, autonomous vehicle control, healthcare, etc.

Given the ubiquity of cyber-physical systems, and the reliance on their performance, incentives are abundant for miscreants to attack such systems, from simple economic reasons (reducing gas bills), and advantages over industrial competitors (manipulating differential electricity pricing), to political espionage and sabotage (derail national scientific and military programs) and full-fledged terrorism (cause communications breakdown, traffic disruptions). Isolation of CPS networks and controllers from the Internet can only offer a limited amount of protection, not only because of the advent of increasingly "smart" cyber-physical systems like Smart Grids, which require Internet access, but also because of the increasing deployment of sensors to remote locations where the sensors themselves, and the communications to and from them, cannot be adequately monitored for security.

Additionally, organized criminals, industrial spies, and global terrorists have proved themselves adept at introducing malware into heavily secured and isolated networks by relying on human errors. The Stuxnet worm is an example of digital warfare that was waged against Iran's nuclear program (Sanger (2012)). Stuxnet, which was chiefly used in coordination with espionage malware was introduced by infected USB flash drives, and further used peer-to-peer calls to infect other computers inside private networks (Matrosov et al. (2010)). It is evident that relying on isolation of networks and components, and in general, security with obscurity, is at best only a short-term solution.

The Stuxnet worm also brought to light serious security susceptibilities in industrial control systems. The worm was specially designed to reprogram industrial centrifuges and sabotage their outputs (Markoff (2010)). This attack resonated with a recent concern in distributed control system security, whereby an attacker could modify the software or environment of some of the networked sensors and/or actuators, to launch a coordinated attack against the system infrastructure.

A conventional method of security, is using symmetric and asymmetric encryption and decryption to secure the communications. Cryptographic keys are broken and stolen daily, but even if they were secure, an attacker could directly attack the physical environment of the components, without even touching the communication network. There are other methods of approaching CPS security, most of which rely either on the information content of the system (confidentiality, integrity, availability), or on the robustness of controllers and estimation, detection and identification algorithms. The problem with concentrating on the information content is the lack of a system model, which can blind the detector to a wide variety of attacks (for example, lowering electricity bills by bypassing the meter). On the other hand, robust controllers and algorithms tend to assume random, uncoordinated failures, which is hardly the case during an attack.

Considerable research has been devoted to constructing estimators that are not unduly affected by outliers or other small departures from model assumptions (Maronna et al. (2006), Huber and Ronchetti (2011)), which can be

⋆ This research was supported in part by CyLab at Carnegie Mellon by National Science Foundation Grants #0955111 CAREER: Efficient, Secure and Robust Control of Cyber Physical Systems, and #1135895 CPS: Medium: Collaborative Research: The Cyber-Physcial Challenges of Transient Stability and Security in Power Grids. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of CMU, CyLab, NSF, or the U.S. Government or any of its agencies.

used to nullify the effect of outliers. However, the case of an attack is quite different from randomly occurring outliers, and such methods need to be reformulated for CPS. Bad data detection has been used in power grids for a long time (Abur and Expósito (2004)). Liu et al. (2011) and Sandberg et al. (2010) consider how an attacker can design and inject inputs into measurements to change state estimation results.

In this paper, we look at the problem of secure detection for a system with a binary state and binary sensors. Although a sensor giving out just one bit of information seems too weak at the first glance, it is more than just an interesting case to look at. For systems using a multitude of distributed sensors for detecting a binary state, it is often superfluous to consider continuous readings from all sensors, and in fact, might prove to be infeasible for both the sparse and low-powered communication network, as well as the small embedded processors. It is usual on such a platform for the sensors to be programmed to make a decision based on the information they have, and only communicate this decision over the network, reducing the communication overhead. The controller then makes a decision based on these preliminary decisions.

A similar system has been previously studied by Agah et al. (2004), Alpcan and Başar (2003), Fuchs and Khargonekar (2011) and later by Vamvoudakis et al. (2012), by formulating the problem as a zero-sum partial information game in which a detector attempts to minimize the probability of error and an attacker attempts to maximize this probability. The optimal policy recommended by the authors in the latter work is a mixed strategy, where the detector chooses between two rules, based on the perceived probability of attack. This policy is dependent on the estimation of this probability of attack, which, for a lot of systems, is not only extremely difficult to analyze and estimate, but might also change widely based on several external factors.

Kodialam and Lakshman (2003) also modeled intrusion detection as a zero-sum game, albeit between the service provider and the intruder. Other game-theoretical approach to solving the problem have been proposed by Bier et al. (2007), who used the method increasing the attractiveness of some ones to the attacker, while designating others as unimportant. The chief drawback of game-theoretical approaches is that the final detection output is possibly a mixed strategy, and not a function of the just the inputs. That is, for the same inputs, the detector output can change randomly based on which policy is chosen, a behavior that may be undesirable in many systems.

Seeking a deterministic solution, we consider the behavior of such a system in the presence of a powerful attacker, without looking to estimate a probability that the adversary will attack. We consider an attack model where the adversary can attack up to a certain number of sensors, while remaining undetected. We provide an insight about what it means for an estimator to be robust in such a scenario, using sensors of different specifications. We analyze the robustness of such a detector for various capabilities of the attacker. We then focus on the case where all the sensors are equivalent, or at least, of similar specifications, and provide a procedure for choosing the detector specifications. We also explore the case where the sensors fall into 2 distinct classes, of different specifications — a case that is of special interest for infrastructures which are under modernization, replacing a few sensors at a time with better versions.

Robust detection with minimax have been previously studied by Huber (1965), along with Strassen (Huber and Strassen (1973)) and Kassam and Poor (1985), using uncertainty classes and the detector being designed as a naive-Bayes or Neymann-Pearson detector. The challenge in such and approach is constructing the least favorable distributions in the uncertainty classes, which are the classes that are supposed to be the hardest for the detector to distinguish.

This paper extends the results of Mo et al. (2012) in the case of binary sensors and binary cases. The problem of finding the sets defined in the paper has been handled, and a procedure has been proposed to construct these sets in specific cases.

The rest of this paper is organized as follows. In section 2, we define the system and the attacker strategy, and formulate the problem as a minimax problem to minimize the "worst-case" probability of detection. In section 3, we develop definitions for robustness and detection probabilities in such a scenario, and comment on the existence of robust detectors. We also look at a detection scheme in the case of a very restricted specification about the number of sensors liable to be attacked. In section 7, we concentrate on the case where all sensors are of equal specifications, and look at another restricted specification about the number of sensors. We propose a method of choosing the optimal specifications for the detector. In section 9 we explore the scenario of two sets of sensor specifications. Section 10 concludes the paper, with discussions on future work.

2. PROBLEM FORMULATION

Consider a binary random variable X, with distribution

$$X = \begin{cases} 0 & \text{with probability } P_0 \\ 1 & \text{with probability } P_1 \end{cases}, \tag{1}$$

where $P_0, P_1 \geq 0$, and $P_0 + P_1 = 1$. Without loss of generality, let $P_1 \geq P_0$.

To detect X, we have available a vector

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \{0,1\}^m \tag{2}$$

of m binary sensor measurement, each of which is conditionally independent from the others given X. Let each sensor have a probability of false alarm (α)

$$P\left(y_i = 1 \middle| X = 0\right) = \alpha_i, \tag{3}$$

$$P\left(y_i = 0 \middle| X = 0\right) = 1 - \alpha_i, \tag{4}$$

$$i = 1, 2, \ldots, m,$$

and probability of detection (β)

$$P\left(y_i = 1 \middle| X = 1\right) = \beta_i, \quad (5)$$
$$P\left(y_i = 0 \middle| X = 1\right) = 1 - \beta_i, \quad (6)$$
$$i = 1, 2, \ldots, m.$$

If any of the sensors are actually such that $\alpha_i \geq \beta_i$ for some values of i, the measurements provided by those sensors can be inverted before being used, making $\alpha_i \leq \beta_i$. Thus, without a loss of generality, we can consider $\alpha_i \leq \beta_i \ \forall i$.

In the case where there is no attack, a Bayes detection algorithm suffices.

$$P_0 \prod_{i=1}^{m} \alpha_i^{y_i} \left(1 - \alpha_i\right)^{(1-y_i)} \underset{H_0}{\overset{H_1}{\lessgtr}} P_1 \prod_{i=1}^{m} \beta_i^{y_i} \left(1 - \beta_i\right)^{(1-y_i)} \quad (7)$$

where $H_0 \equiv \hat{X} = 0$ and $H_1 \equiv \hat{X} = 1$.

2.1 Attack Strategy

It is assumed that an attacker wants to increase the probability that the detector makes an error in detecting X. The attacker has the ability to flip up to l of the m sensor measurements, but the detector does not know which of the m measurements have been manipulated. While the detector knows that at most l measurements have been manipulated, the exact number is also unknown to the detector. This means that any detection scheme $\hat{X} = f\left(y\right)$ has to rely on the original measurement vector (y) manipulated by the attack vector (y^a)

$$y^c = y \oplus y^a, \quad (8)$$

where $y^a \in \{0,1\}^m$, and $\|y^a\| \leq l$. [1] Here $\oplus$ denotes the element-wise exclusive-or operation. By selecting which bits of y^a are 1, the attacker chooses which sensors to attack.

2.2 Problem

The detection problem is formalized as a minimax problem where one wants to select an optimal detector

$$\hat{X} = f\left(y^c\right) = f\left(y \oplus y^a\right), \quad (9)$$

to minimize the probability of error (or maximize the worst-case probability of detection as derived in section 3).

2.3 Attacker Knowledge

To have the detector follow the Kerckhoffs' Principle which states that, a cryptosystem should be secure even if everything about the system, (except, of course, the key), is public knowledge, we assume that the attacker has full knowledge about f, the state of the system X, and all measurements $y_1, y_2, \ldots, y_m$.

3. ROBUSTNESS AND IMPERTURBABLE SETS

The question arises about defining robustness of a detector under such an attack. Since we are looking to maximize the probability of detection in the worst possible case, we need to look for all such sensor measurements, such that if those are the measurements provided by the sensors, the adversary can *never* affect enough of them to change the detector output.

Given a detection scheme $f(y)$, let Y_0 be defined as the set of true measurements y, for which any attack vector, which follows the above attack strategy, cannot force the estimate of X to be changed from 0 to 1. Similarly, let Y_1 be defined as the set of true measurements y, for which any attack vector, which follows the above attack strategy, cannot force the estimate of X to be changed from 1 to 0. Formally,

$$Y_0 = \left\{y \middle| f(y \oplus y^a) = 0, \forall y^a \in \{0,1\}^m, \|y^a\| \leq l\right\}, \quad (10)$$
$$Y_1 = \left\{y \middle| f(y \oplus y^a) = 1, \forall y^a \in \{0,1\}^m, \|y^a\| \leq l\right\}. \quad (11)$$

Thus, an attacker cannot affect the detection from any measurement which falls in the set $Y_0 \cup Y_1$, which is, in a sense, the "imperturbable set" for the detector.

The number of sensor measurements that fall in $Y_0 \cup Y_1$ is a measure of the robustness of the detector.

Example Consider f to be a simple voting scheme, where the detection output depends simply on the majority of the sensor values (m can be considered to be odd to break ties). Let $m = 9$, and $l = 2$. Thus,

$$f\left(y\right) = \begin{cases} 0 & \text{if } \|y\| \leq 4 \\ 1 & \text{if } \|y\| > 4. \end{cases} \quad (12)$$

It is easy to see that $Y_0 = \left\{y \middle| \|y\| \leq 2\right\}$. If $\|y\| \leq 2$, and $\|y^a\| \leq 2$, then $\|y \oplus y^a\| \leq 4$, which will force $f\left(y\right) = 0$. Similarly, it is easy to see that $Y_1 = \left\{y \middle| \|y\| \geq 7\right\}$. If $\|y\| \geq 7$, and $\|y^a\| \leq 2$, then $\|y \oplus y^a\| \geq 5$, which will force $f\left(y\right) = 1$. Thus Y_0 and Y_1, are "good sets" for the detector.

Remark 1. It is important to note that, $Y_0 \cup Y_1 \neq \{0,1\}^m$, except in the case when $l = 0$ (there is no attacker). That is, there will be measurements possible, which are neither in Y_0 nor in Y_1. For these measurements, the attacker can indeed change the output of the detector. In the above example, if the measurement y is such that $3 \leq \|y\| \leq 6$, the attacker can change the detector output to be what he chooses.

In the presence of an attacker, there will measurement values for which the attacker is able to cause an error. In a worst-case scenario, a malicious attacker will always cause errors. Thus, only the points in Y_0 and Y_1 contribute to the worst-case probability of detection. Consider $X = 0$. The probability of getting measurement $y \in Y_0$ given $X = 0$ (which will assure $f\left(y \oplus y^a\right) = 0, \forall y^a \in \{0,1\}^m, \|y^a\| \leq l$) is

$$\sum_{y \in Y_0} \left(\prod_{i=1}^{m} \alpha_i^{y_i} \cdot \prod_{i=1}^{m} \left(1 - \alpha_i\right)^{(1-y_i)} \right). \quad (13)$$

Similarly, the probability of getting measurement $y \in Y_1$ given $X = 1$ (which will assure $f\left(y \oplus y^a\right) = 1$, $\forall y^a \in \{0,1\}^m, \|y^a\| \leq l$) is

$$\sum_{y \in Y_0} \left(\prod_{i=1}^{m} \beta_i^{y_i} \cdot \prod_{i=1}^{m} \left(1 - \beta_i\right)^{(1-y_i)} \right). \quad (14)$$

Thus the total worst-case probability of detection (P) is given by

[1] In this paper, we are only dealing with binary states and sensor measurements, where both the 0-norm and the 1-norm are equivalent. Hence, for legibility we choose to drop the subscript, with the understanding that it can be either the 0-norm or the 1-norm. Indeed, the norm $\|\cdot\|$ can very well be replaced by $\|\cdot\|_p^p, 0 \leq p < \infty$ *mutatis mutandis*, without affecting any of the results.

$$P = P_0 \sum_{y\in Y_0}\left(\prod_{i=1}^{m}\alpha_i^{y_i}\cdot\prod_{i=1}^{m}(1-\alpha_i)^{(1-y_i)}\right) + P_1\sum_{y\in Y_1}\left(\prod_{i=1}^{m}\beta_i^{y_i}\cdot\prod_{i=1}^{m}(1-\beta_i)^{(1-y_i)}\right). \quad (15)$$

Thus the problem of finding the optimal detector can be formally stated as

$$\underset{Y_0,Y_1}{\text{maximize}} \quad P_0 \sum_{y\in Y_0}\left(\prod_{i=1}^{m}\alpha_i^{y_i}\cdot\prod_{i=1}^{m}(1-\alpha_i)^{(1-y_i)}\right) + P_1\sum_{y\in Y_1}\left(\prod_{i=1}^{m}\beta_i^{y_i}\cdot\prod_{i=1}^{m}(1-\beta_i)^{(1-y_i)}\right), \quad (16)$$

subject to constraints of the problem, which will be formalized in further sections.

4. NO FEWER THAN HALF THE SENSORS ATTACKED ($L \geq \lceil \frac{M}{2} \rceil$)

Theorem 2. If $l \geq \lceil \frac{m}{2} \rceil$, at least one of Y_0 and Y_1 is empty.

Proof. $l \geq \lceil \frac{m}{2} \rceil \Rightarrow m - l \leq l$. Suppose both sets are non-empty. Let

$$y^0 = \begin{pmatrix} y_1^0 & y_2^0 & \cdots & y_m^0 \end{pmatrix}^T \in Y_0, \quad (17)$$

$$y^1 = \begin{pmatrix} y_1^1 & y_2^1 & \cdots & y_m^1 \end{pmatrix}^T \in Y_1. \quad (18)$$

Consider a measurement y,

$$y = \begin{pmatrix} y_1^0 & y_2^0 & \cdots & y_l^0, y_{l+1}^1 & y_{l+2}^1 & \cdots & y_m^1 \end{pmatrix}. \quad (19)$$

Now, $y = y^0 \oplus y^a$, i.e., $y^a = y \oplus y^0$. Since the first l values in y^a are definitely zero, $\|y^a\| \leq m - l \leq l$. By the definition of Y_0 (Eq. 10), and the fact that $\|y^a\| \leq l$, it can be concluded that $f(y) = 0$. Let $y = y^1 \oplus y'^a$, i.e., $y'^a = y \oplus y^1$. Since the last $m-l$ values in y'^a are definitely zero, $\|y'^a\| \leq l$. Again by the definition of Y_1 (Eq. 11), and the fact that $\|y'^a\| \leq l$, it can be concluded that $f(y) = 1$, which contradicts the previous conclusion. Hence, one of the two sets must be empty.

Remark 3. If one of the two sets must empty, the other set can, and in general, should, contain all the possible measurements. Essentially, this scheme is equivalent to the detector disregarding the measurements and making a decision based on the prior probabilities P_0 and P_1. Thus, if $l \geq \lceil \frac{m}{2} \rceil$ and $P_0 > P_1$, the detector should always detect $\hat{X} = 0$, i.e, the set Y_1 is empty and Y_0 contains all possible measurements. Similarly, if $l \geq \lceil \frac{m}{2} \rceil$ and $P_1 > P_0$, the detector should always detect $\hat{X} = 1$, i.e, the set Y_0 is empty and Y_1 contains all possible measurements.

The conclusion of Theorem 2 is that if more than half the number of sensors are attacked, the detector should throw away all measurements and always give an output based on the *a priori* probabilities, P_0 and P_1.

Thus from this point onwards, we can consider $l \leq \lfloor \frac{m}{2} \rfloor$.

5. FEWER THAN HALF THE SENSORS ATTACKED

Define a distance metric d as follows. Given $a \in A$ and $b \in B$,

$$d(a,b) = \|a - b\|, \quad (20)$$

$$d(a,B) = \min_{b\in B}\|a - b\|, \quad (21)$$

$$d(A,B) = \min_{a\in A}\|a - B\| = \min_{a\in A, b\in B}\|a - b\|. \quad (22)$$

Lemma 4. For any Y_0, Y_1 such that $d(Y_0, Y_1) \geq 2l+1$ the detector f, $d(y, Y_0) \underset{f(y)=0}{\overset{f(y)=1}{\lessgtr}} d(y, Y_1)$, has Y_0 and Y_1 as the imperturbable sets.

Proof. We only need to prove that $f(y) = 0\ \forall y \in Y_0$ and $f(y) = 1\ \forall y \in Y_1$.

Consider $y \in Y_0$. Let $y^c = y \oplus y^a$. Since the attacker can attack at most l measurements, $\|y^a\| \leq l$. Thus, $\|y^c - y\| \leq l$. Since $y \in Y_0$, the distance metric to Y_0 can only be equal to or smaller than the distance to y, i.e., $d(y^c, Y_0) \leq l$. Since $y \in Y_0$, $d(y, Y_1) \geq 2l + 1$. Since $\|y^c - y\| \leq l$, by the triangle inequality, $d(y^c, Y_1) \geq l + 1$. Since, $d(y^c, Y_0) \leq l < 2l + 1 \leq d(y^c, Y_1)$, $f(y) = 0$ for all $y \in Y_0$.

Similarly, consider $y \in Y_1$. Let $y^c = y \oplus y^a$. Since the attacker can attack at most l measurements, $\|y^a\| \leq l$. Thus, $\|y^c - y\| \leq l$. Since $y \in Y_1$, the distance metric to Y_1 can only be equal to or smaller than the distance to y, i.e., $d(y^c, Y_1) \leq l$. Since $y \in Y_1$, $d(y, Y_0) \geq 2l + 1$. Since $\|y^c - y\| \leq l$, by the triangle inequality, $d(y^c, Y_0) \geq l + 1$. Since, $d(y^c, Y_1) \leq l < 2l + 1 \leq d(y^c, Y_0)$, $f(y) = 1$ for all $y \in Y_1$.

Remark 5. An intuitive way to see this result is that since each attacked sensors counteracts the measurement provided by an unattacked sensor, an attack on l out of m sensors essentially means that the detection is carried out using the measurements provided by $m - 2l$ sensors. Thus, $\forall y^0 \in Y_0, y^1 \in Y_1$, $\left\|y^0 - y^1\right\| \geq 2l + 1$. For example, if $m = 9$ and $l = 2$, 2 unattacked sensors will counteract the effect of 2 attacked sensors, leaving the detector to estimate $\hat{X}$ from 5 sensors. Thus $\left\|y^0 - y^1\right\| \geq 5$.

Thus the problem of finding the optimal detector can be formally stated as

$$\underset{Y_0,Y_1}{\text{maximize}} \quad P_0 \sum_{y\in Y_0}\left(\prod_{i=1}^{m}\alpha_i^{y_i}\cdot\prod_{i=1}^{m}(1-\alpha_i)^{(1-y_i)}\right) + P_1\sum_{y\in Y_1}\left(\prod_{i=1}^{m}\beta_i^{y_i}\cdot\prod_{i=1}^{m}(1-\beta_i)^{(1-y_i)}\right) \quad (23)$$

$$\text{subject to} \quad d(Y_0, Y_1) \geq 2l + 1. \quad (24)$$

6. SPECIAL CASE: $L = \frac{M-1}{2}$

The result of Lemma 4 is reduces to a simple form, for the particular case where m is odd, and $l = \frac{m-1}{2}$.

Corollary 6. If $l = \frac{m-1}{2}$, $|Y_0| = |Y_1| = 1$. Further, if $Y_0 = \{y^0\}$ and $Y_1 = \{y^1\}$, $y^0 = \bar{y^1}$.

Proof. In this case, $d(Y_0, Y_1) \geq 2l+1$. But $2l+1 = m$ and the distance between two m-dimensional binary vectors can be at most m. Thus, $d(Y_0, Y_1) = m$. Thus, for any

$y^0 \in Y_0$ and $y^1 \in Y_1$, $y^0 = \bar{y^1}$. Suppose that there is another $y'^0 \in Y_0$ such that $d\left(y'^0, y^1\right) = m$. By the triangle inequality, $d\left(y'^0, y^0\right) \leq 0$, i.e., $y'^0 = y^0$. Thus, Y_0 is a singleton set. Similarly it can be proved that Y_1 is also a singleton set.

Remark 7. If none of the sensors are "inverted", then the measurement that will form Y_0 is $y_i = 0\ \forall i$ (thus making $Y_1 = \left\{y \middle| y_i = 1\ \forall i\right\}$). To put it formally, if $\alpha_i \leq \beta_i\ \forall i$, then $Y_0 = \left\{(0\ 0\ \cdots\ 0)^T\right\}$ and $Y_1 = \left\{(1\ 1\ \cdots\ 1)^T\right\}$.

6.1 Complexity Of The Search-Space

The space of all possible measurements is $\{0,1\}^m$, i.e., there are 2^m possible values of y. Each value can be in Y_0, Y_1, or neither, thus giving rise to 3^{2^m} possible ways of designing Y_0 and Y_1, and hence, the detector.

Having said that, once one of the sets, say Y_0, is fixed, it is possible to expand Y_1 for all measurements such that $d(Y_0, Y_1) \geq 2l+1$ is not violated, by finding all all points at a distance $2l+1$ or more from each point in Y_0, and then taking the intersection of these. Even considering this reduction, there are 2^{2^m} possible ways of fixing Y_0 and Y_1.

This double-exponential behavior of the enumerations makes a brute-force search impractical beyond a very small value of m — computers will run out of memory by $m = 5$. $m = 6$ is intractable.

In the further sections, we will concentrate on reducing the search-space for some oft-encountered cases.

7. ALL SENSORS ARE EQUIVALENT

It is unlikely to ever be the case, that each sensor is unlike every other sensors — in a practical application, most, if not all, sensors would have their false alarm and detection rate equal. Even if the performance parameters are not exactly equal, they would be close enough to each other, that the sensors can be assumed to be equivalent:

$$\alpha_i = \alpha, \tag{25}$$
$$\beta_i = \beta, \tag{26}$$
$$i = 1, 2, \ldots, m.$$

Thus,

$$P = P_0 \sum_{y \in Y_0} \alpha^{\|y\|} (1-\alpha)^{(m-\|y\|)} + P_1 \sum_{y \in Y_1} \beta^{\|y\|} (1-\beta)^{(m-\|y\|)}. \tag{27}$$

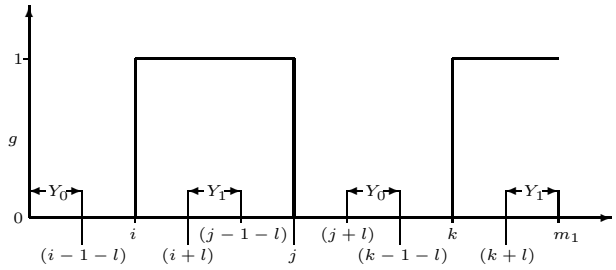
The advantage of this assumptions lies in the fact that the search for the optimal detector can be confined to only those detector functions that are symmetric in sensor values. Further, for any detector that assumes all sensors are equivalent, the detector function is a symmetric boolean function, and the output of the detector is a function of only the number of ones or zeros in the measurement y (Wegener (1987)). Thus, the detector function $f(y)$, where $y = (y_1\ y_2\ \cdots\ y_m)^T$ can be one of several types of counting functions:

$$T_k^n(y) = 1 \iff \|y\| \geq k \quad \text{(threshold functions)}$$
$$E_k^n(y) = 1 \iff \|y\| = k \quad \text{(exactly-}k\text{-functions)}$$
$$C_{k,p}^n(y) = 1 \iff \|y\| = k \mod p. \quad \text{(counting functions)}$$

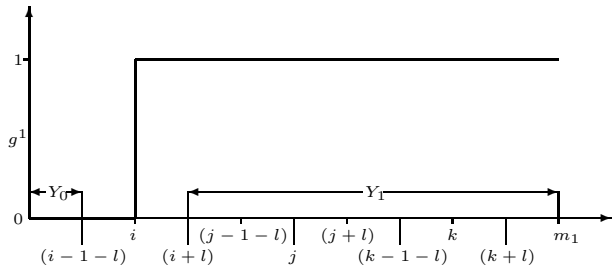
In this case, however, the optimal detector function, i.e., the function with the maximum worst-case probability of detection (among symmetric boolean functions) can be proved to be a threshold function, i.e., it is monotonically increasing.

Theorem 8. The optimal function $g(\|y\|)$, defined to be a symmetric boolean function with the maximum worst-case probability of detection, is monotonically increasing.

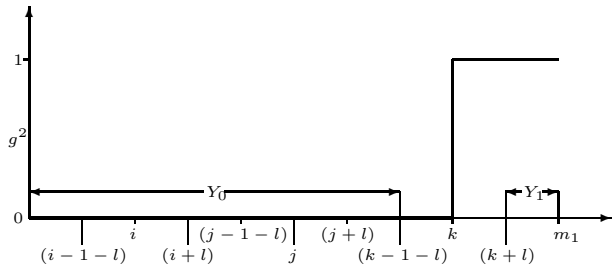
Proof.



(a) Non-Monotonic Function



(b) Monotonic Function g^1



(c) Monotonic Function g^2

Fig. 1. Detector Functions

By the assumption that none of the sensors are inverted, $g(0) = 0$ and $g(m) = 1$. Suppose that the function g is not monotonic, and has a "kink". Thus, $\exists i, j, k$, such that $0 \leq i < j < k \leq m_1 \leq m$ and

$$g(n) = \begin{cases} 0 & \text{if } 0 \leq n \leq i-1 \\ 1 & \text{if } i \leq n \leq j-1 \\ 0 & \text{if } j \leq n \leq k-1 \\ 1 & \text{if } k \leq n \leq m_1 \end{cases} \tag{28}$$

An example function g with such a "kink" is shown in Fig. 1a. Each kink in the function can be denoted by

unique values of (i, j, k, m_1). In the following argument, we consider only the kink closest to 0.

Since the detector function is given by

$$g\left(\|y\|\right) = \begin{cases} 0 & \text{if } d\left(y, Y_0\right) > d\left(y, Y_1\right) \\ 1 & \text{if } d\left(y, Y_0\right) \leq d\left(y, Y_1\right), \end{cases} \tag{29}$$

where,

$$d\left(Y_0, Y_1\right) \geq 2l + 1, \tag{30}$$

the subsets of Y_0 and Y_1 that lie in the range $[0, m_1]$ can be computed to be

$$Y_0 = \left\{y \middle| 0 \leq \|y\| \leq (i-1-l)\right\} \cup \left\{y \middle| (j+l) \leq \|y\| \leq (k-1-l)\right\} \tag{31}$$

$$Y_1 = \left\{y \middle| (k+l) \leq \|y\| \leq m_1\right\} \cup \left\{y \middle| (i+l) \leq \|y\| \leq (j-1-l)\right\} \tag{32}$$

Depending upon the value of m_1 as compared to m, there can be other subsets of Y_0 and/or Y_1 beyond the range that we consider. However, the presence of such subsets will not affect the argument.

These sets are also shown in Fig. 1a. Now consider two other functions, $g^1, g^2 \not\equiv g$ as follows:

$$g^1\left(n\right) = \begin{cases} 0 & \text{if } 0 \leq n \leq i-1 \\ 1 & \text{if } i \leq n \leq m_1 \\ g\left(n\right) & \text{if } m_1 \leq n \leq m \end{cases} \tag{33}$$

$$g^2\left(n\right) = \begin{cases} 0 & \text{if } 0 \leq n \leq k-1 \\ 1 & \text{if } k \leq n \leq m_1 \\ g\left(n\right) & \text{if } m_1 \leq n \leq m \end{cases} \tag{34}$$

The corresponding subsets of Y_0^1, Y_1^1, Y_0^2, and Y_1^2 within the range $[0, m_1]$ are given by

$$Y_0^1 = \left\{y \middle| 0 \leq \|y\| \leq (i-1-l)\right\} \tag{35}$$

$$Y_1^1 = \left\{y \middle| (i+l) \leq \|y\| \leq m_1\right\} \tag{36}$$

$$Y_0^2 = \left\{y \middle| 0 \leq \|y\| \leq (k-1-l)\right\} \tag{37}$$

$$Y_1^2 = \left\{y \middle| (k+l) \leq \|y\| \leq m_1\right\} \tag{38}$$

These two functions, along with the sets are shown in Figs. 1b and 1c. It can be seen that g^1 and g^2 are defined in a way to have only one of the two $0 \to 1$ transitions of the first kink in g. Now, using the definition of the worst-case probability of detection, the probability P_d for the detector function g can be given by

$$P_d = P_0 \sum_{n=0}^{i-1-l} \alpha^n \left(1-\alpha\right)^{m-n} + P_0 \sum_{n=j+l}^{k-1-l} \alpha^n \left(1-\alpha\right)^{m-n} + P_1 \sum_{n=i+l}^{j-1-l} \beta^n \left(1-\beta\right)^{m-n} + P_1 \sum_{n=k+l}^{m_1} \beta^n \left(1-\beta\right)^{m-n} + P_{(m_1,m)},$$

where $P_{(m_1,m)}$ denotes the contribution to the worst-case probability of detection, of the part of the function that lies beyond the range $[0, m_1]$ that we consider. Comparatively, the worst-case detection probabilities P_d^1 and P_d^2 for the constructed functions g^1 and g^2 respectively, can be calculated to be

$$P_d^1 = P_d - \underbrace{\left(P_0 \sum_{n=j+l}^{k-1-l} \alpha^n \left(1-\alpha\right)^{m-n} - P_1 \sum_{n=j+l}^{k-1-l} \beta^n \left(1-\beta\right)^{m-n}\right)}_{P_{\text{diff}}} + \underbrace{P_1 \sum_{n=j-1-l}^{j+l} \beta^n \left(1-\beta\right)^{m-n} + P_1 \sum_{n=k-1-l}^{k+l} \beta^n \left(1-\beta\right)^{m-n}}_{P_\beta},$$

and

$$P_d^2 = P_d + \underbrace{\left(P_0 \sum_{n=j+l}^{k-1-l} \alpha^n \left(1-\alpha\right)^{m-n} - P_1 \sum_{n=j+l}^{k-1-l} \beta^n \left(1-\beta\right)^{m-n}\right)}_{P_{\text{diff}}} + \underbrace{P_0 \sum_{n=i-1-l}^{i+l} \alpha^n \left(1-\alpha\right)^{m-n} + P_0 \sum_{n=j-1-l}^{j+l} \alpha^n \left(1-\alpha\right)^{m-n}}_{P_\alpha}.$$

That is,

$$P_d^1 = P_d - P_{\text{diff}} + P_\beta$$
$$P_d^2 = P_d + P_{\text{diff}} + P_\alpha.$$

We know that $P_\alpha, P_\beta \geq 0$. Now, for g to be optimal, $P_d \geq P_d^1$ and $P_d \geq P_d^2$. But,

$$\begin{aligned} & P_d \geq P_d^1 \\ \iff & P_d \geq P_d - P_{\text{diff}} + P_\beta \\ \iff & P_{\text{diff}} \geq P_\beta \\ \Rightarrow & P_{\text{diff}} \geq 0, \end{aligned} \tag{39}$$

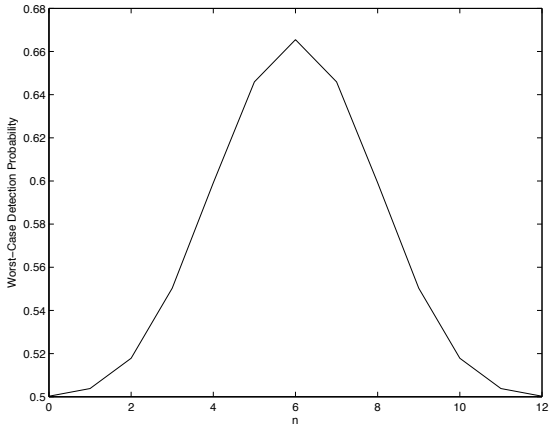
and

$$\begin{aligned} & P_d \geq P_d^2 \\ \iff & P_d \geq P_d + P_{\text{diff}} + P_\alpha \\ \iff & -P_{\text{diff}} \geq P_\alpha \\ \Rightarrow & P_{\text{diff}} \leq 0. \end{aligned} \tag{40}$$

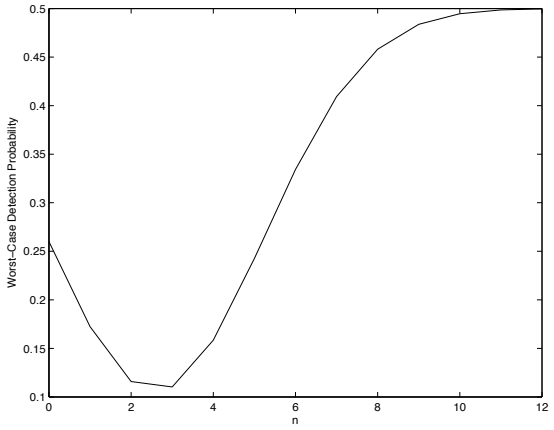
The only way these inequalities are satisfied, is if $P_{\text{diff}} = P_\alpha = P_\beta = 0$. This will be the case if $\alpha = \beta$ (in which case, all three probabilities are equal), or $i = j = k$ (there is no kink). The first case is discounted by the assumption that $\alpha < \beta$, and in the second case, all three functions g, g^1, and g^2 are equivalent, which is discounted by the assumption $g^1, g^2 \not\equiv g$. This is a contradiction.

Thus, the worst-case probability of detection of any function g can only be increased by removing the first such kink in g. If the function g has more than one kink, upon removal of the first kink in g, there will be a new "first kink" in the new function. However, the above result can be applied successively to each such kink, leading to the conclusion that the optimal g, the one that has the maximum worst-case probability of detection, has no such kinks, i.e., the optimal g has to be monotonically increasing.

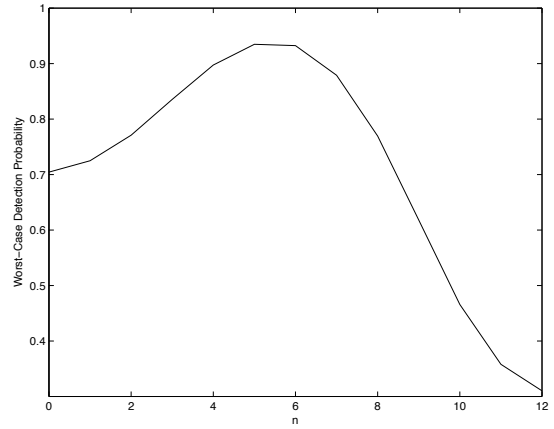
Since the the optimal detector function has only one $0 \to 1$ transition, it can be defined only by one parameter, the threshold. The results of Lemma 4 can be combined with Theorem 8, to obtain the conditions for the threshold:



(a) $P_0 = P_1 = 0.5, \alpha = 0.3, \beta = 0.7$



(b) $P_0 = P_1 = 0.5, \alpha = 0.3, \beta = 0.35$



(c) $P_0 = 0.3, P_1 = 0.7, \alpha = 0.2, \beta = 0.8$

Fig. 2. Worst-Case probability of detection P as a function of n, for $m = 9$ and $l = 3$.

Corollary 9. In a system where all m sensors have equivalent specifications, and the attacker can attack up to l sensors, the the sets Y_0 and Y_1 which maximize the worst-case probability of detection such that $d(Y_0, Y_1) \geq k$, are given by

$$Y_0 = \left\{ y \middle| \, \|y\| \leq n \right\}, \tag{41}$$

$$Y_1 = \left\{ y \middle| \, \|y\| \geq n + 2l + 1 \right\}, \tag{42}$$

for some integer n such that $0 \leq n \leq \frac{m-1}{2}$. The detector function is therefore given by

$$f(\|y\|) = \begin{cases} 0 & \text{if } \|y\| \leq n + l \\ 1 & \text{if } \|y\| \geq n + l + 1. \end{cases} \tag{43}$$

8. GENERAL VALUES OF L

We now consider other values of $l < \left\lfloor \frac{m-1}{2} \right\rfloor$. For given m and l, the worst-case probability of detection P is a function of n parametrically dependent on P_0, P_1, α and β. The shape of the function varies widely with a small change in these values, and cannot be said to be either convex or concave. For example, for $m = 9$ and $l = 3$ we get the plots of worst-case probability of detection P vs n for different values of α and β, shown in Fig. 2.

As a result, it is impossible to predict a closed form expression for n. The only solution is to do on exhaustive search for $n = 0$ through $n = m - 2l - 1$. This is a linear search and thus tractable even for large values of m and l.

9. TWO CLASSES OF SENSORS

There is an often-encountered case in practical applications, where the sensors can be grouped into two classes — "good" sensors, and "better" sensors. This is usually the case when the sensors of a legacy network are being upgraded in steps, or when the better sensors are much more expensive than the good ones to be considered worth it. In such a case, a compromise can be reached by only installing a few better sensors, while most of the network is composed of the cheaper sensors. For example, Phasor Measurement Units (PMUs) are so expensive compared to power meters, that only a few substations have them installed. Although the power grid can be considered to be in the process of being upgraded, even the best case distribution of the PMUs is expected to be around 30% of the total sensors.

$$\alpha_i = \alpha_a, \tag{44}$$

$$\beta_i = \beta_a, \tag{45}$$

$$i = 1, 2, \ldots, m_a.$$

$$\alpha_i = \alpha_b, \tag{46}$$

$$\beta_i = \beta_b, \tag{47}$$

$$i = m_a + 1, m_a + 2, \ldots, m_a + m_b = m.$$

Let

$$y = \left(\underbrace{(y_1 \;\; y_2 \;\; \cdots \;\; y_{m_a})}_{y_a} \; \underbrace{(y_{m_a+1} \;\; y_{m_a+2} \;\; \cdots \;\; y_{m=m_a+m_b})}_{y_b} \right) \tag{48}$$

The search for the optimal detector can be confined to only those detector functions that are symmetric in y_a and y_b, making $f(y_1, y_2, \ldots, y_m) = g(\|y_a\|, \|y_b\|)$.

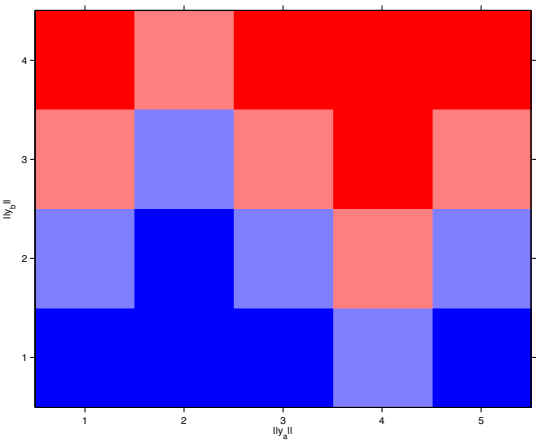


Fig. 3. Optimal Y_0 (blue) and Y_1 (red) for $m_a = 4$, $m_b = 3$, with $P_0 = P_1 = 0.5$ and $\alpha_a = 0.1$, $\beta_a = 0.9$, $\alpha_b = 0.2$, $\beta_b = 0.8$. The paler colors denote the corresponding decision when the point is neither in Y_0 nor Y_1.

$$P = P_0 \sum_{(y_a\ y_b)^T \in Y_0} \left(\alpha_a^{\|y_a\|}(1-\alpha_a)^{(m_a-\|y_a\|)} \cdot \alpha_b^{\|y_b\|}(1-\alpha_b)^{(m_b-\|y_b\|)}\right) + P_1 \sum_{(y_a\ y_b)^T \in Y_1} \left(\beta_a^{\|y_a\|}(1-\beta_a)^{(m_a-\|y_a\|)} \cdot \beta_b^{\|y_b\|}(1-\beta_b)^{(m_b-\|y_b\|)}\right). \quad (49)$$

This case reduces to a search over a 2-D space. However, equivalent conditions of monotonicity do not hold. As a counterexample, consider $m_a = 4$, $m_b = 3$, with $P_0 = P_1 = 0.5$ and $\alpha_a = 0.1$, $\beta_a = 0.9$, $\alpha_b = 0.2$, $\beta_b = 0.8$. The optimal Y_0 and Y_1 are given in Fig. 3.

Thus, the search needs to be carried over a space of $2^{(m_a+1)(m_b+1)}$ possible combinations of Y_0 and Y_1. This is a significant reduction in complexity over the double-exponential nature of the original problem, and tractable for values of $m \leq 12$.

10. CONCLUSION

We proposed a new approach to estimate a binary random variable based on a vector of sensor measurements that may have been corrupted by an attacker. The problem was formulated as a minimax problem with detector attempting to maximize the worst-case probability of detection, and the attacker attempting to minimize this probability. A tractable form of the detector was derived in the case where the sensors are either all of equivalent specifications, or belong to one of two classes of specifications.

Future work will involve reducing the search space for two classes of detectors to make higher number of sensors tractable, and extending the results to sensors with integer outputs instead of binary outputs.

REFERENCES

Abur, A. and Expósito, A.G. (2004). *Power System State Estimation*. Theory and Implementation. CRC Press.

Agah, A., Das, S.K., Basu, K., and Asadi, M. (2004). Intrusion detection in sensor networks: a non-cooperative game approach. In *Network Computing and Applications, 2004. (NCA 2004). Proceedings. Third IEEE International Symposium on*, 343–346. IEEE.

Alpcan, T. and Başar, T. (2003). A game theoretic approach to decision and analysis in network intrusion detection. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, 2595–2600 Vol.3. IEEE.

Bier, V., Oliveros, S., and Samuelson, L. (2007). Choosing What to Protect: Strategic Defensive Allocation against an Unknown Attacker. *Journal of Public Economic Theory*, 9(4), 563–587.

Fuchs, Z.E. and Khargonekar, P.P. (2011). Games, deception, and Jones' Lemma. In *American Control Conference (ACC), 2011 DO -*, 4532–4537. IEEE.

Huber, P.J. (1965). A Robust Version of the Probability Ratio Test. *The Annals of Mathematical Statistics*, 36(6), 1753–1758.

Huber, P.J. and Ronchetti, E.M. (2011). *Robust Statistics*. Wiley.

Huber, P.J. and Strassen, V. (1973). Minimax Tests and the Neyman-Pearson Lemma for Capacities. *The Annals of Statistics*, 1(2), 251–263.

Kassam, S.A. and Poor, H.V. (1985). Robust techniques for signal processing: A survey. *Proc. IEEE*, 73(3), 433–481.

Kodialam, M. and Lakshman, T.V. (2003). Detecting network intrusions via sampling: a game theoretic approach. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, 1880–1889.

Lee, E.A. (2008). Cyber Physical Systems: Design Challenges. In *Object Oriented Real-Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on DOI - 10.1109/ISORC.2008.25*, 363–369. IEEE.

Liu, Y., Ning, P., and Reiter, M.K. (2011). False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.*, 14(1), 1–33.

Markoff, J. (2010). A Silent Attack, But Not A Subtle One. *New York Times*, 160(55176), 6.

Maronna, R.A., Martin, D.R., and Yohai, V.J. (2006). *Robust Statistics*. Theory and Methods. Wiley.

Matrosov, A., Rodionov, E., Harley, D., and Malcho, J. (2010). Stuxnet under the microscope. *ESET*.

Mo, Y., Hespanha, J., and Sinopoli, B. (2012). Robust detection in the presence of integrity attacks. In *American Control Conference (ACC), 2012 DO -*, 3541–3546.

Sandberg, H., Teixeira, A., and Johansson, K.H. (2010). On Security Indices for State Estimators in Power Networks. In *First Workshop on Secure Control Systems, Cyber Physical Systems Week 2010*.

Sanger, D.E. (2012). Obama Order Sped Up Wave Of Cyberattacks Against Iran. *New York Times*, 161(55789).

Vamvoudakis, K.G., Hespanha, J.P., Sinopoli, B., and Mo, Y. (2012). Adversarial detection as a zero-sum game. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 7133–7138.

Wegener, I. (1987). The complexity of symmetric boolean functions. In E. Börger (ed.), *Lecture Notes in Computer Science*, 433–442–442. Springer, Berlin.