Vulnerability Analysis and Protection of Cyber-Physical Systems from a Control Perspective

By

HEMANGI LAXMAN GAWAND
ENGG01201004029

BHABHA ATOMIC RESEARCH CENTER

A thesis submitted to the

Board of Studies in Engineering Sciences

In partial fulfillment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

August 2018

Homi Bhabha National Institute[1]

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Ms. Hemangi Laxman Gawand** entitled **"Vulnerability Analysis and Protection of Cyber-Physical Systems from a Control Perspective"** and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman –Dr. A.P. Tiwari		**Date:** 10/08/2018
Guide - Dr. A. K. Bhattacharjee	AKBhattacharjee	**Date:** 10/08/2018
Guide- Dr. Kallol Roy	Kallol Roy	**Date:** 10/08/18
Examiner - Dr. Amitava Gupta	Amitava Gupta	**Date:** 10-08-2018
Member 1- Dr. V. H. Patankar	V. H. Patankar	**Date:** 10.08.2018
Member 2- Dr. S.Kar	SKar	**Date:** 10.08.2018

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 28/08/2018

Place:

AKBhattacharjee
<Signature>
Guide Dr. A. K. Bhattacharjee

Kallol Roy
<Signature>
Guide Dr.Kallol Roy

[1] This page is to be included only for final submission after successful completion of viva voce.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Hemangi Gawand

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Hemangi Gawand

List of Publications arising from the thesis

Accepted paper

Journal

a. **"Securing Cyber Physical System Using LSA and Computational Geometric Approach"** Nuclear Engineering and Technology, Vol 49, Issue 3, Elsevier 2017, DOI:http://dx.*doi*.org/10.1016/j.net.2016.10.009

b. "**Investigation of control theoretic cyber attacks on controllers**", International. Journal of Systems, Control and Communications, Inderscience, 2016, DOI: http://dx.doi.org/10.1504/IJSCC.2016.077410.

Conference

a. "**Confirmation of Theoretical Results Regarding Control Theoretic Cyber Attacks on Controllers** " Accepted paper in 10th IFAC International Symposium on Dynamics and Control of Process Systems (DYCOPS 2013),The International Federation of Automatic Control, December 18-20, 2013. Mumbai, India, DOI 10.3182/20131218-3-IN-2045.00081

b. **"Real Time Jitters And Cyber Physical System"**- 24-27 Sept. 2014, International Conference on Advances in Computing, Communications and Informatics (ICACCI-2014) - Page(s): 2004 - 2008,ISBN:978-1-4799-3078-4,AccessionNumber: 14779107, DOI:10.1109/ ICACCI.2014, .6968505.IEEE – **Best Paper Award**

c. **"Control Aware Techniques for Protection of Industrial Control System"** , Date of Conference:11-13 Dec. 2014,Pages:1 - 6,Print ISBN:978-1-4799-5362-2,INSPEC Accession Number:14904139, INDICON 2014 DOI - 10.1109/INDICON.2014.7030660

d. "**Online Monitoring of a Cyber Physical System against Control Aware Cyber Attacks**", 4th International Conference on Eco-friendly Computing and Communication Systems (ICECCS), Elsevier Procedia, November 2015, page - 238 – 244, DOI :10.1016/j.procs.2015.10.079

Hemangi Gawand

Dedicated to my mother.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my guides, Dr. A.K. Bhattacharjee and Dr. Kallol Roy for their excellent guidance, encouragement and support in every stage of my research. I would also like to thank my doctoral committee head Dr. A. P. Tiwari for his valuable suggestions and guidance.

I feel a deep sense of gratitude towards my family and friends for their tremendous support and motivation.

“Trust, but verify.”

-Ronald Reagan

CONTENTS

Synopsis

Cyber Physical System (CPS) includes network of devices that receives and perform physical actions while simultaneously being controlled and monitored by computational and communication software. These systems maintain an ongoing relationship with a physical system governed by laws of natural sciences. Such systems are often vulnerable to cyber-attacks due to the weakness in the design of the system for which security was never considered as a requirement in the same level as functional requirements. Several instances of cyber-attacks have been reported targeting specifically the critical infrastructures like National Power Grids. It is very challenging to secure cyber physical systems from such attacks and protect the system under control. Protecting against such attacks is challenging as the attack payloads are configured with deep knowledge of the controller and system under control. Established diagnostic and monitoring algorithms based on sequential analysis techniques like Sequential probability ratio test (SPRT), Cumulative Sum (CUSUM), Maximum likelihood (MLE), etc., are promising to be useful for detection of anomalies in controller characteristics. Data mining techniques used for streaming data can be used for protecting and securing control systems.

In our research work, we have explored control aware techniques for protecting the cyber physical systems. Our work focuses on theoretical analysis of such cyber-attacks with a postulate that such attacks could be detected by statistical techniques like SPRT, CUSUM, and GLR etc. These are studied closely on lab-scale experimental setups. Design of security monitors was studied to detect anomalous changes in behavior of controller output.

List of figures

List of tables

List of Abbreviations

1. **CPS –** Cyber Physical System
2. **SPRT –** Sequential Probability Ratio Test
3. **ICS**- Industrial Control Systems.
4. **CERT** - Cyber Emergency Response Team
5. **COTS** - Commercial off the shelf
6. **CIA**- confidentiality, integrity and availability
7. **SCADA**- Supervisory Control And Data Acquisition
8. **LTI** - linear and time invariant.
9. **DCS** - Distributed control system
10. **DOS** - Denial of Service
11. **CUSUM –** Cumulative sum
12. **GLR –** Generalized likelihood ratio
13. **MLE** - Maximum likelihood Estimation
14. **LSA** - Least Square approximation
15. **SISO** - single input single output
16. **MIMO** - Multiple input multiple output
17. **ICS** - Industrial Control Systems
18. **IT –** Information Technology
19. **MITM –** Man in the middle

List of Definitions

Data Integrity: It ensures that all the information generated and exchanged during the system's operation is accurate and complete without any alterations.

Data Confidentiality: It ensures that all sensitive information generated within the system is disclosed only to those who are supposed to. Confidentiality requires the ability to hide data.

Authentication: It ensures that the system knows the identities of all the entities interacting with it.

Authorization: It ensures that any entity trying to access particular information from the system is able to access only that information which they are entitled to.

Availability: It ensures that any entity that uses the data and services and resources of the system are able to do when required.

Asset: An item of economic value owned by an organization is termed as asset.

Threat: A threat can be defined as any agent, circumstance, or situation that could cause harm or loss to an asset.

Attack: The execution of the threats is called an attack.

Attackers: The entities, which execute the threats, are called as attackers.

Insider Attack: An insider attack is a malicious attack perpetrated on a network or computer system by a person with authorized system access.

Eavesdropping: The attacker can intercept any information communicated by the system. It is a passive attack.

Passive Attack: The attacker does not interfere with the working of the system and simply observes its operation.

Hacker: - A "**hacker**" is a person who compromises one or all the security goal (confidentiality, integrity and availability) so as to achieve the desire outcome.

False Positive: A false positive is where you receive a positive result for a test, when you should have received a negative result.

False Negative: A false negative is where a negative test result is wrong. Expected test result was positive.

Engineered System: an engineered system is a combination of components that work in synergy to collectively perform a useful function.

Non-repudiation: is the means by which a recipient can ensure the identity of the sender and that neither party can deny having sent or received the message.

Chapter 1: Introduction

1.1 Cyber-Physical System (CPS)

Cyber-Physical Systems are characterized by the tight interaction between a digital computing component (the Cyber part) and a continuous-time dynamic system (the Physical part). Cyber Physical System (CPS) includes network of devices that receives and perform physical actions while simultaneously being controlled and monitored by computational and communication software. CPSs are core for critical infrastructures like industrial plants and are significantly important to public and nation infra structure.

Cyber system and physical system are intertwined in CPS including components such as

- Computing elements with software.
- Communication networks for data transmission.
- Electrical systems for power.
- Considerable amount of electrical wiring between sensors to computing systems.

Few examples of CPS system are SCADA (Supervisory Control and Data Acquisition), HVAC (Heat Ventilation and Air Conditioning), Health monitoring system, Air craft etc.

One of the challenging threats to CPS is from "**targeted attacks**" where the attacker has deep knowledge of the targeted system and can aim for maximum damage. There is no rollback in CPS because of its integration in the physical world. Cyber-attacks can result into equipment and production damage as well as compliance violation. Few well known attacks on industrial control systems are –Ukrainian Power Grid attack in 2016, Stuxnet in 2010, Pennsylvania water-filtering plant in 2006, Davis-Besse power plant in Oak Harbor, Ohio in 2003 and Maroochy Shire Sewage attack in 2000.

Securing CPS is a challenge due to its inherent complexity. Vulnerabilities in the CPS systems may be due to defects emanating from imprecise understanding of software elements, architecture exposing hacking points including network and sensor elements. CPS vulnerabilities assessment cannot be done by information technology tools alone. Information security focuses on confidentiality of the data and strength of the cryptographic algorithms. It is not the knowledge of computing that forms the basis of a targeted CPS attack, but rather the comprehensive knowledge about the sensors, wiring, control algorithms, software components and the network technology used for the design.

Figure 1 shows a simplified CPS network. An actuator is an energy device that moves or controls another mechanism. Data received by the actuator causes necessary actions on the physical system. The sensor provides measurement of physical parameters, which are in turn processed by the controller to provide the output. Sensors measure physical system states and transmit them to the distributed controllers. A control action is a reactive process and the failure of any non- redundant sensor, algorithm or actuator breaks the reactive action that causes irreparable damage to the system under control.

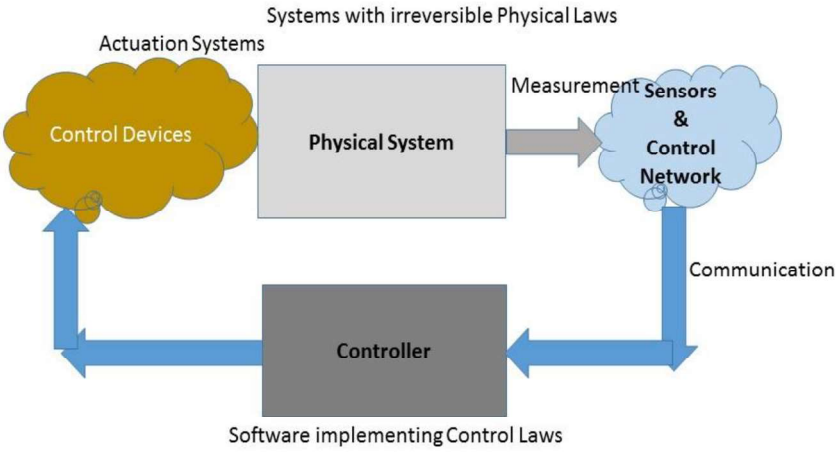


Figure 1. The general architecture of cyber physical systems

1.2 CPS Security

Data security has three primary goals - Confidentiality, Integrity and Availability. **Confidentiality** is a measure taken to prevent disclosure of information or data to unauthorized individuals or systems.

Integrity refers to methods and actions taken so as to protect the information from unauthorized alteration when it is in transmit or rest phase.

Availability refers that the data is available to the legitimate user when required. Attack on the availability is referred as '**Denial of Service**' (DoS). CPSs used in mission-critical system have availability as its prime criteria for design.

CPS is vulnerable to cyber security due to:-

1. Its rapid adoption of commercial off the shelf technology (COTS).

2. Remote access for support and operations.
3. Availability of security information for Industrial Control Systems through internet and blogs.

1.2.1 Differences between Control and Information Technology Security

Software update methods are not suitable for the control system. In CPS patching or upgrading a system can take extended periods of time, as the system must be shut down prior to upgrade. Following key points need to be considered while developing security measures for the CPS: -

1. **Risk assessment**: - Estimating the amount of damage an attack can cause to a system under control.
2. **Detection algorithms**: - compromised state of the CPS can be identified.
3. **Attack-response algorithms**: - required so that the system components can survive in the attack period as well as after the attack without operational loss.

1.3 Motivation and Assumptions

A CPS is composed of a set of networked programmable digital systems with interfaces to sensors, actuators, and communication units. Challenging threats to CPS is from "**targeted**" attacks where the attackers have deep knowledge of the targeted controlled process and hence can tune their attacks with the aim of maximum damage from safety and economic perspectives. Traditional information security focuses more on confidentiality of the data and strength of its cryptographic algorithms, while cyber-attack on CPS target on the underlying physical/chemical/biochemical processes by tampering

the control algorithms and its associated system configuration. Hence it is not enough to ascertain the security of the individual CPS components in isolation.

Recently there has been lot of research activities in assessing CPS security from a combined view of control theory and computing. The focus of this research is to understand effect on CPS from a cyber attack perspective with knowledge of the control algorithms. In CPS, the changes resulted due to controller actions are irreversible and hence detection algorithm needs to quickly detect the changes and take corrective action instantaneously.

Thus, monitoring control system is as important as securing the embedded element from possible targeted attacks on its computational elements. Monitoring requires task decomposition and constraints validation. It can be integrated with the control system to provide real time as well as historic information. The rapid process of output generation from multiple sensors and controllers create a large data log files in a short time span. Log files are useful in the analysis of the control plant behavior in safe as well as in the attack period. However, these files are bulky and difficult for manual inspection. Various data mining techniques such as Least Square Approximation (LSA) and Computational methods can be used in the data log analysis and take preventive actions when required. Research work in this thesis highlights few statistical methodologies that are used in algorithm design for effective monitoring that can be used for security and diagnostic purpose.

Assumptions:

Following assumptions are made:

1. The plant under control is amenable to a LTI model.

2. The attacker has complete knowledge of the control plant model. All targeted attacks are assumed to be internal attacks.
3. There is no distinction between failure and attack for a controller and hence all failures are assumed as an attack.

1.4 Problem Formulation and Contribution of the Thesis

The main problems studied in this thesis are

1. The various possible impacts that an attack can create through a detailed study of the system architecture and control algorithms, and
2. Design of an effective monitoring scheme to detect an anomaly in the behavior of the control system with appropriate measurements.

The focus in this thesis is in the analysis of various postulated CPS attacks using computational techniques for fault and change detection. Our attack analysis method relies on control theoretic notion of a LTI controller. Continuous monitoring of the system under control is an essential task to be performed for generating alarm on detection of anomalous controller behavior and study its impact on the physical process. Hence designing algorithms for effective monitoring and studying their characteristic behavior is another research focus for us.

Assuming CPS as a LTI system and controller acting as its core, our research work aims at analyzing various CPS attack models with respect to state space model and designing effective monitoring system for its corrective action. Statistical analysis tools like Sequential Probability Ratio Test (SPRT), Cumulative Sum (CUSUM), Generalized Likelihood ratio Test (GLR) are used in fault diagnostics in signal system and reused in our CPS vulnerability analysis.

Detection and correction go hand in hand. Hence a corrective system is designed using an effective monitor for diagnosing abrupt changes in the characteristic properties of the object under observation. Available research focuses on observing properties like threshold etc. that are varying over time. Log files can be analyzed for failure detection and taking corrective action. Data mining techniques like Least Square Approximation (LSA), computational geometric approach is helpful in the analysis of huge log files. Log files can be analyzed for failure and system behavior change detection and taking corrective action. In real world there are many parameters that a control system handle. Critical parameters are identified prior to monitoring any control system, so as to generate alerts when the observed parameter exceed the threshold limit.

The research work presented in this thesis, identifies various targeted control aware cyber-attacks and techniques for securing control systems. It has been shown that by incorporating changes as stated in attack model, one is able to simulate various attacks (based on four tank model [14]) that were detected by innovation value change (Innovation values indicate difference in actual and expected value), using SPRT, CUSUM and other techniques like co-relation and covariance.

SPRT technique has been extended using GLR and CUSUM method for comparative analysis. State space method is used in continuous and discrete time domain for detecting safety properties.

Our thesis examines and proposes how an intelligent monitor can be designed for attack detection and takes corrective action using knowledge of general control system framework and computational geometric methods. Computational geometric approach

like LSA and convex hull method has an advantage of analysis of complex data obtained from LTI plant. This approach helps in securing CPS from cyber attacks.

The main contributions of our thesis are:

1. Various targeted control aware cyber-attacks have been identified. Assuming CPS as a LTI system and controller acting as its core, attack model has been developed for various attacks like stealth attack, replay attack, and covert attack.
2. The theoretical aspects of these attacks have been analyzed and it has been postulated that such attacks can be detected by statistical techniques like SPRT, CUSUM, Kalman filter etc.
3. Techniques of designing security monitors extending the diagnostic features in signal analysis for synchronous detection of fault or attack in the control system under observations have been explored. The research work presented in the thesis highlights methodologies and algorithms so as to develop an effective monitor. This thesis work provides explanation for various features an ideal monitor needs to support. Traditional approach of selecting single filter or monitor parameter is replaced with complex parameters selection for analysis using computational methods based on GLR, CUSUM, LSA and convex hull approach.

1.5 Outline of the dissertation

- Chapter 1 provides introduction to CPS and formulate problem for research work.
- Chapter 2 gives literature survey of the existing research and related work with respect to CPS.
- Chapter 3 is a foundation material and the reader must know this.

- Chapter 4 analyses various CPS vulnerabilities and model them.
- Chapter 5 provides statistical analysis for attack analysis.
- Chapter 6 designs online monitor for taking corrective action on the CPS attacks using geometric approach.

Chapter 2

Literature Survey

CPS is a complex field that requires knowledge of multiple disciplines to solve the arising challenges. Fei Hu [21] explains about the principle for design of CPSs' and its challenges. He describes about various design theories and modeling methods for a practical CPS. Byres et. al. [4] explains about Stuxnet attack along with its means and methods to spread to the desired control devices. Alvaro et. al in [5,6,7] demonstrates the threat to a control system that reprograms the controller to behave out of specified boundaries. They have analyzed various control system attacks by using example of Tennessee Eastman plant [22] as a system under attack. Eric Byres et. al. [4] has highlighted various incidents of PLC attacks like waste management system attack in Australia, East Coast paper mill attack and 'Data Storm' [8].There are other incidents like Taum Sauk project- Missouri by David W. Lord and Davis – Besse nuclear power plant – Oak harbor Ohio.

Peter Maybeck [29] has proposed stochastic methods for understanding control system behavior. Yu-Lun Huang et. al. [22] has shown that by incorporating knowledge of the physical systems under control, it is possible to detect the change in behavior of the targeted control system. He describes threat models for false data injection attack. Yilin Mo et.al [30] has analyzed the effect of the replay attacks on control system in a steady state LTI system. Marco Caselli et.al [9] has demonstrated the basis for semantic attacks.

Generalized likelihood ratio (GLR) test to detect changes in system dynamics and sensor jumps, has been proposed by Willksy [42]. Michelle Basseville et.al. [3] has analyzed various detection methods like SPRT, CUSUM methods to detect abrupt changes in the continuous time domain signal. Basseville in [1, 2] have explained various algorithms like Hinkley CUSUM to detect changes in signal by observing its mean value. Zhang et. al. [44] has analyzed techniques for early warning for systems behavior changes. Julian Rrushi [35] has classified attacks on the distributed system based on the data transferred on the network. He further analyzed various intrusion detection techniques for it.

Alwyn Goodloe et.al [19] has discussed about monitor designing in real time, its properties and architectures in a distributed environment. There are various other methods for data analysis like clustering method [25, 26], Expectation maximization [27], sampling method, statistical method and continuous data stream querying. Yunyue Zhu et.al. [45] has analyzed large data streams generated by using statistical parameters like correlation co-efficient, auto correlation as well as beta factor and to observe the change in data pattern behavior. Shapiro [36] explains about Real time functions that can be used to represent solid figure (three dimensional figures). Izchak Sharfman [37] has described about various search engines having different mirrors for data monitoring. He has listed few of the monitor design approaches like frequency, feature and computational method. Fabio Pasqualetti.et al. [32] has used computational geometric approach for large data stream analysis. Toussaint, Godfried [40] has analyzed convex hulls intersection algorithm using rotating caliper method. Four tank model has been considered as an example of decentralized control system by Manuel Mazo Espinosa [14] for analyzing CPS. Daniel Perez Huertas [23] has analyzed control system security against malicious

attack in water tank system. R.K.Shyamasundar in [38] has described about big data approach for the protection of control systems by analyzing its large log files.

2.1 CPS Attack Histories

Few of the well-known CPS attacks are -

1. In 1998, East Coast paper mill faced a major problem when the machine operator lost its control on the motor, controlled by the PLC. After troubleshooting it was noticed that, the DCS to PLC communications gateways was overloaded every 5 minutes. One of the ex-employee (officer) had loaded a small program onto one of the DCS graphics stations that requested all DCS devices to dump their data every five minutes.

2. One of the target attacks is of "**Maroochy Shire**" in 2000, when a sacked employee took control via his notebook PC and radio transmitter of the waste water facility and successfully spilled 264,000 gallons of sewage into nearby streams and rivers by re-programming the control plant behavior.

3. On 14thAugust 2003, a large portion of the Midwest and Northeast United States, Ontario, Canada, experienced an electric power blackout. It was caused due to modification in the power grid meters that introduce bad measurements and affected state estimation.

4. On 19th August 2006, at Browns Ferry nuclear power plant, Unit 3 went at high-risk state when both 3A and 3B reactor re-circulating pumps failed placing the plant in high power and low flow condition. This unfavorable condition was controlled only by manual shutdown of the reactor. After investigation it was noticed that variable frequency drive (VFD) controllers controlling re-circulating pump and Unit 3

demineralizer controller has simultaneously failed. It was caused due to extensive increase in the data traffic that resulted in re-circulating pump failure. This attack is referred as '**Data Storm**' attacks.

5. A Stuxnet attack [43] made on Iran nuclear power plant was zero-day vulnerability. It was detected in 2010 and was specific for Siemens SCADA system only. The virus had the ability to modify the data, send to and received from PLC's without being noticed by the PLC operator.
6. In 2014, attacker through its phishing emails took control of the production systems in a German steel mill. It disabled various alarms and safety mechanisms, and triggered an emergency shutdown of a blast furnace, causing a massive damage.
7. In December 2015, Thousands of homes in Western Ukraine were black out due to "Black Energy" malware that attacked energy sub-station. January 2016, same "Black energy" virus attacked Kiev airport.

In November 2016, A Distributed Denial of Service (DDoS) attack resulted in the loss of heating to two buildings in the city of Lappeenranta in eastern Finland. The building management system was flooded with bogus internet traffic causing system to restart every few minutes.

2.2 Research Gaps Addressed

Security of the CPS is a challenge due to its inherent complexity due to reactive interaction with physical systems, large software implementing the control laws. The gap lies between the general software designs for IT systems and software intended to do control communication, coordination and intelligence. It must be borne in mind that the cyber attack is to cause an intended failure in the system as opposed to random failure in

traditional electronic systems. The resiliency to intended failure needs to be addressed borrowing techniques from fault detection, isolation, signal processing and statistical techniques used for anomaly detection.

2.3 Conclusion

The chapter has provided the foundation for research in this thesis by discussing related works by the researchers. The focus is in understanding the present research scenarios in using statistical techniques to detect such cyber-attacks on the software implemented controllers.

Chapter 3

Control System and Statistical Analysis Theory

In this chapter the basic background of Linear Time Invariant (LTI) systems and state space techniques relevant to security are discussed. A discussion on Bayesian estimation with respect to cyber security analysis of controllers has been made.

3.1 Linear State space Equation

Consider a linear time invariant system (LTI) with A, B, C and D being state space matrices. A state-space model of a system with input vector 'u' and output vector 'y' takes the following form in continuous time.

$$x(k+1) = Ax(k) + Bu(k) \quad (1)$$

$$y(k) = Cx(k) + Du(k) \quad (2)$$

where $x(k) \in R^n$, $y(k) \in R^p$, $A \in R^{nxn}$, $B \in R^{nxm}$, $C \in R^{pxn}$ and $D \in R^{pxm}$

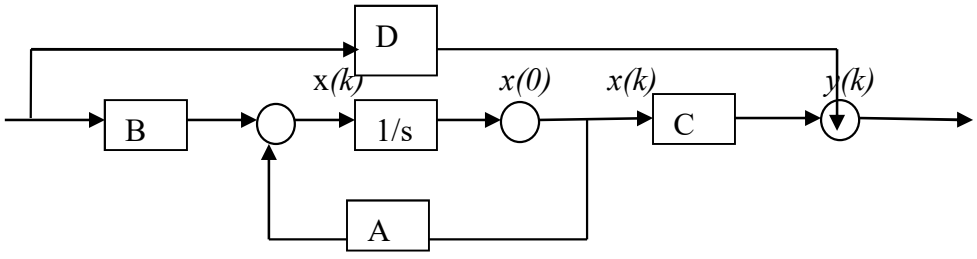


Figure 2: - State Change diagram

Figure 2 gives pictorial representation of output signal '*y(k)'* due to state matrix and input state '*x(k)'* signal interaction. State space equations (1) and (2) are further used for various control aware attacks analysis in chapter 4.

3.2 Control System

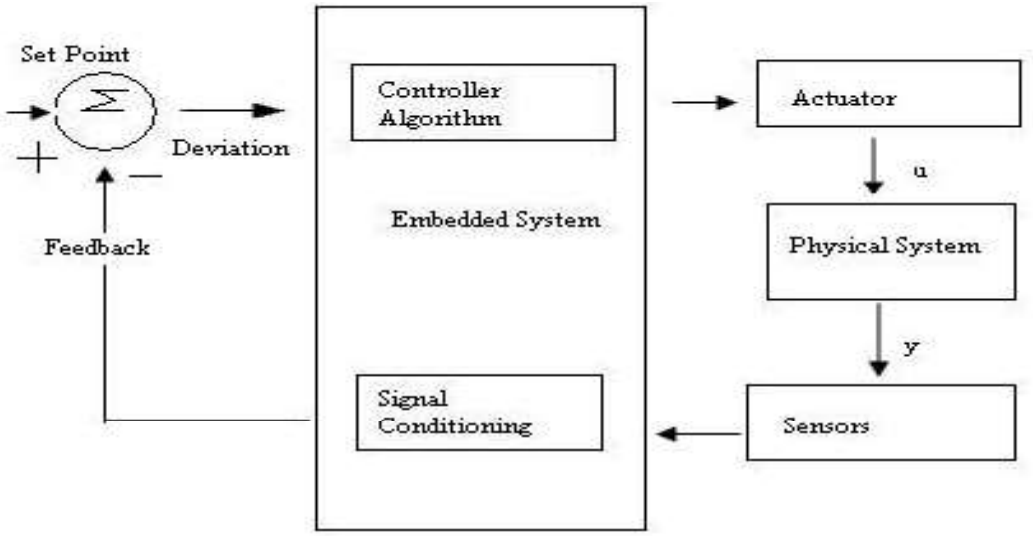


Figure 3. Abstraction of CPS

Basic elements of the control system are sensors, computation and actuators. A basic control system is represented using Figure 3. The difference in the set point and feedback value acts as an input to the controller algorithm that resides in embedded electronic system. Controller algorithm controls the actuator output that acts as an input (u) to the physical system. Physical system output (y) acts as an input to the sensor that in turn feed to signal conditioning to generate feedback signal so as to recalculate the deviation signal [31].

3.3 KALMAN Filter

Kalman Filter is a predictive /corrective filter that uses state model equations for estimating the next state and minimizing square mean error. The predictive corrective

model helps in indicating any change in the state of the control system (as explained in Chapter 4) and identifies any control aware attack.

A research paper was published in 1960 by R.E. Kalman titled "*A New Approach to Linear Filtering and Prediction Problems*" describing a recursive solution to the discrete-data linear filtering problem. Kalman filter is a set of mathematical equations that provides an efficient computational recursive means to estimate the state of a process, in a way that can minimizes the square mean error. It is a powerful estimator that can estimate the past, the present, and the future states of a control system. Filter address as the problem of estimating the states x $\in R^n$ *of* discrete time-controlled process governed by linear stochastic equations as stated by equation (3).

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1} + Bu_{k-1} + \mathrm{w}_{k-1} \qquad (3)$$

In the above Eq. (3) 'A' is *n x n* matrix, and represents state matrix that relates the previous states $\hat{x}_{k-1}$ at step 'k-1' to the current state $\hat{x}_k$ at the step 'k'. 'A' matrix is assumed to be constant. 'B' is *n x l input* matrix that controls control input $u_k \in R^l$ to state $\hat{x}_k$. $\hat{x}_k^-$ represent a priori state estimate at step 'k'. $\hat{x}_k$ is the posterior state estimate at step 'k' for a given measurement y_k' . $\hat{x}_{k|k-1}$ is system state at 'k' step when moved from 'k-1' step.

Measurement y $\in R^n$ is given by

$$\mathrm{y}_k' = \mathrm{C}\,\mathrm{x}_k' + \mathrm{v}_k \qquad (4)$$

'C' is *m x n* matrix that relates measurement, y_k' to state x_k' and is assumed to be constant. Random variable 'w_k'and 'v_k' represents the process and measurement noise respectively and assumed to be white Gaussian. Process noise covariance 'Q' and

measurement noise covariance 'R' are assumed to be constant and their normal distribution is given by:

p(w) ~N(0,Q),

p(v) ~N(0,R).

Covariance matrix is defined in Eq. (5).

$$P_k = AP_{k-1}A^T + Q \quad (5)$$

Kalman gain is defined as Eq. (6).

$$K = P_k C^T (CP_k C^T + R)^{-1} \quad (6)$$

Error in prior and posterior estimate is given by Eq.(7) and (8) respectively:

$$e_k^- = x_k - \hat{x}_{k+}^- \quad (7)$$

$$e_k^+ = x_k - \hat{x}_{k+} \quad (8)$$

Covariance matrix and Kalman gain value are used to estimate the updated state and covariance value as given by -

$$\hat{x}_k = A\hat{x}_{k|k-1} + K_k e_k \quad (9)$$

$$P_{k+1|k} = P_k - KCP_k \quad (10)$$

Equation (6), (7) and (8) are called as estimator and (9) and (10) are called as corrector for state variables [20].

In Chapter 4, we extend Kalman Filter knowledge for further analysis, where we have subjected output from the controller (connected to Kalman Filter) to various control aware attacks.

Following section detail about generalized likelihood ratio (GLR) [42], SPRT [3], CUSUM [3] etc.

3.4 Generalized likelihood Ratio test

Likelihood ratio test is a statistical test used to compare the best fit of two models, the null model and the alternative model.

Considering mean value (μ) for test, Null and alternative models can be defined as below:

Null model (H_0): $\mu \leq \mu_0$

Where as in

Alternative model (H_A): $\mu > \mu_0$

μ_0 represent first assumed reference mean value. The symbols μ and σ^2 stands for mean and variance respectively.

We are using likelihood function (Λ) as given by Eq. (11) to obtain the peak (mode) for the conditional density function $f(x_1 \ldots . x_n|\mu\sigma)$.

Likelihood ratio is as given by equation (11): -

$$\Lambda = \frac{\max f(x_1 \ldots . x_n|\mu\sigma \ \ for \ \mathrm{H_0}}{\max f(x_1 \ldots . x_n|\mu\sigma \ \ for \ \mathrm{H_0} \cup \mathrm{H_A}} \qquad (11)$$

$$\Lambda = \frac{\max \prod_{i=1}^{n} f(x_i|\mu\sigma \ \ for \ \mathrm{H_0}}{\max \prod_{i=1}^{n} f(x_i|\mu\sigma \ \ for \ \mathrm{H_0} \cup \mathrm{H_A}} \qquad (12)$$

f(x) represent probability density function (p.d.f). Since the x_i are independent their joint p.d.f is the product of the individual p.d.f's

Replacing f(x) = $(1/\sigma\sqrt{2\Pi})e^{(-(\frac{((x_i-\mu^2)}{2\sigma^2}))}$ for each value of x_i in Equation (12).

$$\Lambda = \frac{\max \prod_{i=1}^{n} (1/\sigma\sqrt{2\Pi})e^{(-(\frac{((x_i-\mu^2)}{2\sigma^2}))} \ for \ \mathrm{H_0}}{\max \prod_{i=1}^{n} (1/\sigma\sqrt{2\Pi})e^{(-(\frac{((x_i-\mu^2)}{2\sigma^2}))} \ for \ \mathrm{H_0} \cup \mathrm{H_A}} \qquad (13)$$

$$\Lambda = \frac{\max \prod_{i=1}^{n} (1/\sigma\sqrt{2\Pi})e^{(-(\frac{((x_i-\mu)^2)}{2\sigma^2}))} \ for \ \mathrm{H_0}}{\max \prod_{i=1}^{n} (1/\sigma\sqrt{2\Pi})e^{(-(\frac{((x_i-\mu)^2)}{2\sigma^2}))} \ for \ \mathrm{H_0} \cup \mathrm{H_A}} \qquad (14)$$

$$lik = \prod_{i=1}^{n}(1/\sigma\sqrt{2\Pi})e^{(-(\frac{((x_i-\mu^2)}{2\sigma^2}))} \quad (15)$$

Hence the solution obtained is as given by Eq.(16) and (17):-

$$\text{If } \hat{x} \leq \mu o, then\ \mu\ \mathrm{H}_{0mle} = \hat{x}\ and\ \sigma\ \mathrm{H}_{0mle} = \frac{\sum_{i=1}^{n}(x_i-\hat{x})^2}{n} \quad (16)$$

$$\text{If } \hat{x} \geq \mu o, then\ \mu\ \mathrm{H}_{0mle} = \hat{\mu}\ and\ \mathrm{H}_{0mle} = \frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{n} \quad (17)$$

GLR has been further extended for SPRT analysis in the following section.

3.5 Sequential Probability Ratio Test (SPRT)

SPRT is a hypothesis testing developed by Abraham Wald like GLR between two statistical hypotheses termed as null and alternative hypothesis. SPRT considers the likelihood ratio as a function of the number of observations. Maximum likelihood Estimation (MLE) can be used to estimate the change in mean and variance by threshold parameter '$\mathcal{I}$' as shown in equation (18)

$$\mathcal{I} = \frac{\max_{Ho} f\ (x_1 \ldots\ldots x_n|\mu,\rho)}{\max_{HoUHa} f\ (x_1 \ldots\ldots x_n|\mu,\rho)} \quad (18)$$

H_o and H_a are decision criteria as per value of '$\mathcal{I}$'. H_o (null hypothesis) is true if $\mathcal{I} < \lambda$. H_o symbolizes no Attack. H_1 (alternate hypothesis) is true if $\mathcal{I} > \lambda$. H_1 symbolizes change in the behavior due to possible attack. Assuming probability density function as Gaussian, '$\mathcal{I}$' can be expressed, with respect to change in mean and variance.

To find maximum value for numerator, we use equations (19) and (20):-

$$If\ \bar{x} \leq \mu_0\ then\ \widehat{\mu_{HoMLE}} = \bar{x}\ ,\ \widehat{\rho_{HoMLE}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} \quad (19)$$

$$If\ \bar{x} > \mu_0\ then\ \widehat{\mu_{HoMLE}} = \mu_0\ ,\ \widehat{\rho_{HoMLE}} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \quad (20)$$

For analysis, consider a time series sequence *{z(1); z(2); . . . ; z(N)}* for '*N*' samples and a decision 'd_N' between two hypotheses: H_0 and H_1. Assuming that the observations *'$z_j(k)$'*

under *'H_j'* are generated with a probability distribution *'p_j'* , the SPRT algorithm can be described for the sequence by the equation (21):

$$S(k+1) = log\frac{p_1(z(k))}{p_0(z(k))} + S(k) \quad (21)$$

d_N = H_1 if S(k+1)≥ upper limit

d_N = H_0 if S(k+1) ≤ upper limit

For understanding let's define

$$z_i(k) = \left|\left|\widehat{y_i(k)} - y_i(k)\right|\right| - b_i \quad (22)$$

where $\widehat{y_i(\mathrm{k})}$ is observed measurement value while '$y_i(\mathrm{k})$' represent actual value and '$\mathrm{b_i}$' is a small positive value added to make the correction in the difference between actual and observed value. The expected value can be expressed as equation (23): -

$$\mathrm{E}\left(\left|\left|\widehat{y_i(\mathrm{k})} - y_i(\mathrm{k})\right|\right| - \mathrm{b_i}\right) < 0 \quad (23)$$

The nonparametric CUSUM statistic for sensor 'i' is given by equation.

$$\mathrm{S(k)} = \mathrm{S(k-1)} + \ \mathrm{z_i}\ (\mathrm{k}) \text{ where } \mathrm{S(0)} = 0 \quad (24)$$

The above equation acts as a foundation for various future controls aware attacks analysis like Surge attacks, geometric attacks described in chapter 4.

3.6 Hinkley's 'CUSUM' tests

Variation in the mean value of the controller output signal can be used for analysis of controller behavior. 'CUSUM' test checks for variations in the mean value of the controller output signal by observing its deviations (signal drift), with respect to the maximum and minimum of the cumulative sum values. Signal drift can be due to noise, other parameters or a result of the targeted attack. The mean value 'μ' before and after attack is unknown and so the first mean 'μ_0' and magnitude of change is assumed as 'γ'. An increase or decrease in the mean value, outside the acceptable limit generates an alarm.

Consider decreasing mean value (indicating change in output signal) due to changes in the controller characteristics. Let 'S_n' be the cumulative sum and 'M_n' be maximum cumulative sum for the controller signal output. 'S_n' is define by equation (25) -

$$S_n = \sum_{i=1}^{n}(y_i - \mu_0 - {}^{\gamma_m}/_2) \quad \text{where } S_0 = 0 \tag{25}$$

$$M_n = \max_{0 \le k \le n} S_k \tag{26}$$

Alarm (for exceeding the threshold) is raised if the above equation satisfies:

$$M_n - S_n \geq threshold \tag{27}$$

For increasing mean value for increasing controller output, let 'U_n' be the cumulative sum and 'N_n' be minimum cumulative sum for the controller signal output. ' U_n 'can be defined by equation (28)-

$$U_n = \sum_{i=1}^{n}(y_i - \mu_0 - {}^{\gamma_m}/_2) \text{ where } U_0 = 0 \tag{28}$$

$$N_n = \min_{0 \le k \le n} U_k \tag{29}$$

Alarm is raised if the above equation satisfies: -

$$U_k - N_n \geq threshold \quad (30)$$

CUSUM method is effective for continuous comparing live data with the maximum and minimum CUSUM values.

3.7 Cyber Security Aspects in Bayesian Estimation

Estimation of unmeasured states and monitoring of changes in the statistical parameters of the residues/innovations, form an important approach towards both model-based fault detection & diagnosis (FDD) and for *deliberate introduction* of faults (considered as attacks). This requires the formulation of system dynamics in the state-space framework

$$x_k = A_{k|k-1}x_{k-1} + B_{k-1}u_{k-1} + w_{k-1}$$

$$z_k = H_k x_k + D_k u_k + v_k$$

wherein the conditional probability density function (*pdf*) of the state-vector (***X***), conditioned on the measurement, z$^{p(x_k|z_k)}$,

is propagated through a predictor-corrector process to obtain the optimum estimate of the state while minimizing its error covariance

$$E[(\hat{x}_k - x_k)^T(\hat{x}_k - x_k)] = E[\tilde{x_k}^T\tilde{x_k}]$$

The Bayesian formulation yields the conditional *pdf* of the k_{th} state, which is equated to the likelihood function & the prior

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)\cdot p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}$$

$$\text{posterior} = \frac{\text{likelihood}\cdot\text{prior}}{\text{evidence}}$$

and it is this formulation which governs the Bayesian estimation methodology. The cyber security aspects, envisaged in such a situation is the deliberate intrusion in the computer system, for altering the *prior* or *likelihood* functions, which would necessarily result in

wrong prediction/ computation of the prior. It may also be noted that the fundamental assumptions, based on which the entire Bayesian framework is built, are as follows:

$$E[\mathbf{x}_0] = \boldsymbol{\mu}_0^{\mathbf{x}}$$

$$E[\mathbf{w}_k] = 0 \forall k$$

$$E[\mathbf{v}_k] = 0 \forall k$$

$$\mathrm{cov}\{\mathbf{w}_k, \mathbf{w}_j\} = \mathbf{Q}_k \delta_{kj}$$

$$\mathrm{cov}\{\mathbf{v}_k, \mathbf{v}_j\} = \mathbf{R}_k \delta_{kj}$$

$$\mathrm{cov}\{\mathbf{x}_0, \mathbf{x}_0\} = \mathbf{P}_0$$

$$\mathrm{cov}\{\mathbf{w}_k, \mathbf{v}_j\} = 0, \forall k,$$

$$\mathrm{cov}\{\mathbf{x}_0, \mathbf{w}_k\} = 0, \forall k,$$

$$\mathrm{cov}\{\mathbf{x}_0, \mathbf{v}_j\} = 0, \forall j$$

Any deviations in the above assumptions, either deliberately or by engineered methods, would result in non-optimal solution of the Bayesian prediction structure or finally contribute towards improper results.

3.8 Cyber Security Aspects in the Computations Involving a Kalman Filter in a Non-Bayesian Framework

The entire concept is based on the primary foundations of a recursive least square approach, wherein the errors in the measurement vector $\varepsilon_y = y - H\hat{x}$

from the basic measurement equation

$$y = Hx + v$$

can be minimized by assumption of a cost function ***J***,

$$J = \varepsilon_{y_1}^2 + \varepsilon_{y_2}^2 + \ldots + \varepsilon_{y_k}^2$$

$$J = \varepsilon_y^T \varepsilon_y$$

Wherein

$$J = (y - H\hat{x})^T (y - H\hat{x})$$

$$\frac{dJ}{d\hat{x}} = -H^T y - y^T H + 2\hat{x}^T H^T H = 0$$

$$H^T y = H^T H \hat{x}$$

$$\hat{x} = (H^T H)^{-1} H^T y$$

Resulting in a recursive formulation, for estimation of $\boldsymbol{x}$, given in the equation below :

$$\hat{x}_k = \hat{x}_{k-1} + K_k(y_k - H_k \hat{x}_{k-1})$$

Now, by effective minimization of the covariance, $\boldsymbol{P_k}$, explained below, we obtain :

$$E[x_k] = \hat{x}_k$$
$$E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] = P_k$$

And after subsequent minimization of variance and computation of mean, a Kalman filter can be effectively cast into the given generalized framework of a Bayesian estimation problem, resulting in a closed form solution, provided the dynamics are linear and the uncertainties are Gaussian.

The *prediction* equations, wherein the mean and the variance components are separated

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$$

$$P_k^- = AP_{k-1}A^T + Q$$

and the *update* equations, where the Kalman gain is computed and the error used for final computation of the mean and covariance, are as given below :

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

$$P_k = (I - K_k H)P_k^-$$

Such applications of Kalman filter are common in any optimization problem, where the estimated value of $\mathbf{X_k}$ is utilized in the Riccatti equation, for the final solution of the control & trajectory problem.

As can be observed from the above equations, there are a number of steps, involving matrix inversions, which implies that within the micro-computer registers and accumulators, there are a number of to-&-fro operations, involving basic math operation utilizing numeric-co-processors. Any knowledgeable intruder can always intercept the interim results and make minor alterations in the same, causing the final estimated value to be different from the expected value and thereby resulting in the optimal trajectory to be different.

Typically, the conventional statistical paradigms, used for FDD, can also be used in Cyber Security Scenarios, wherein by study of the behavior of the estimated states, $\mathbf{X_k}$ & the residues

$$(z_k - H\hat{x}_k^-)$$

along with the convergence of the error covariance matrix

$$P_k = (I - K_k H)P_k^-$$

can be studied by use of multiple-model filtering, GLR (generalized likelihood ratio) methods, sequential probability ratio tests (SPRT) on the residues, etc. and any deviations in the same can be suitably interpreted.

3.9 Conclusion

Statistical techniques like SPRT, CUSUM and GLR etc. are useful in the analysis of the controller output and assess any small unintended changes. In the analysis of controller behavior with a focus on cyber security, change detection is of importance. The focus of

the research here lies in development of algorithms for change detection and their applications.

Chapter 4

Modeling Attacks on a CPS and their Simulation

4.1 Introduction

In this chapter, we investigate few targeted attack models using control theory and statistical techniques as explained in chapter 3. We use a case study on a standard four tank system. We develop a mathematical model and simulate in simulation software. The objective is in analyzing various CPS attack models with respect to state space equations and taking corrective action using statistical techniques as discussed in Chapter 3.

The main contributions of this chapter are as follows:

1. Classification of various control aware attacks (targeted attacks) using state space equation.
2. Simulation of control aware attacks on four tank model.
3. Detection of attack using statistical techniques like SPRT, CUSUM, correlation, co-variance etc.

4.2 Attack Classification

An attack on a CPS can cause damage to the system under control by manipulating the controller characteristic parameters. Attacks on CPS are classified as: -

1. **Non-Targeted Attacks**: - Attacker is unaware of the damage that is going to be result of his act. E.g. Attack on water filtering plant in Pennsylvania in 2006, Davis-Besse power plant (2003) in Oak Harbor- Ohio infected with slammer worm.
2. **Targeted Attacks**: - Attacker is aware of the targeted control system and the attack strategy is well planned. These attacks can also be termed as **'insider attack'** as attack is by authorized person. E.g. Stuxnet, Maroochy Shire incident (Slay, J. and Miller, M. (2007) [39].

A targeted attack in CPS can be classified based on the strategy of capturing and altering input, output or state of the control system.

1. **Input Data Attack**: - Control signal is targeted in this attack that result in the output or state of the system under control to change.
2. **Output Data Attack**: - Measurement signal of the target system E.g Power plant output is varied.
3. **State Attack**: - Attacker tries to manipulate the controller states that result in the change of the controller as well as target system output [32].

Targeted control aware cyber-attacks as listed below are explained using state space equations and control theoretic models [5,6,7].

4.3 GENERALIZED CONTROL AWARE ATTACK MODEL

4.3.1 Input Data Attack

Control-data attacks

It is a type of control signal attack where *'u'* **(**input for controlling physical system) is manipulated. Consider normal plant behavior where next state *x(k+1)* and output. *y(k)* is given by linear time invariant (LTI) equation (31) and (32) [31] for time interval 'k' such that $\forall k \notin K_a$ and $K_a = \{ k_s, \ldots k_e\}$ represent attack duration. k_s, k_e are start and end time of attack respectively. Equation (31) and (32) are identical to Equation (1) and (2).

$$x_1(k+1) = Ax(k) + Bu(k) \qquad \forall k \notin K_a \text{ and } k_s \leq k \leq k_e \qquad (31)$$

$$y_1(k) = Cx(k) + Du(k) \qquad \forall k \notin K_a \text{ and } k_s \leq k \leq k_e \qquad (32)$$

For simplicity of understanding noise term and *Du(k)* are ignored in future discussion. Let's assume that the control signal is modified to 'u_1' in attack period. Change in control signal can be represented by state space equations (33) and (34).

$$x_1(k+1) = Ax(k) + Bu_1(k) \qquad \forall k \in K_a \text{ and } k_s \leq k \leq k_e \qquad (33)$$

$$y_1(k) = Cx(k) \qquad \forall k \in K_a \text{ and } k_s \leq k \leq k_e \qquad (34)$$

Where $x_1(k+1), x(k), u(k) \in R^n$ *and* $u_1(k) \in R^n$.

$x_1(k+1)$ and $x(k)$ represents next and current state of the system at k^{th} time interval, while *y(k)* represents measurement output and *u(k)* as control input [5].

4.3.2 OUTPUT DATA ATTACK

In this attack, measurement signal 'y' (input to controller) is targetted. For analysis let's consider that the modified data signal to be *'*$y_1(k)$*'* at k^{th} time interval where $y_1(k) \in R$ *and* $k \in K_a$. This attack is achieved by either change in 'C' matrix or by adding noise 'v_k' or both as shown in equations (35), (36) and (37).

$$y_1(k) = C_1 x(k) \quad , \forall k \in K_a \quad \text{and} \quad k_s \leq k \leq k_e \qquad (35)$$

$$y_1(k) = Cx(k) + v_k \quad , \forall k \in K_a \quad \text{and} \quad k_s \leq k \leq k_e \qquad (36)$$

$$y_1(k) = C_1 x(k) + v_k \quad , \forall k \in K_a \quad \text{and} \quad k_s \leq k \leq k_e \qquad (37)$$

The attacker can change the probability distribution function (PDF) of the output by modulating the noise term (v_k as shown in Equation (36) and (37)) so that the output changes from Gaussian as in Figure 4 to non-Gaussian in Figure 5 respectively. Figure 4 represents probability density function of the normal control plant and Figure 5 depicts plant behavior when noise is mixed with it.

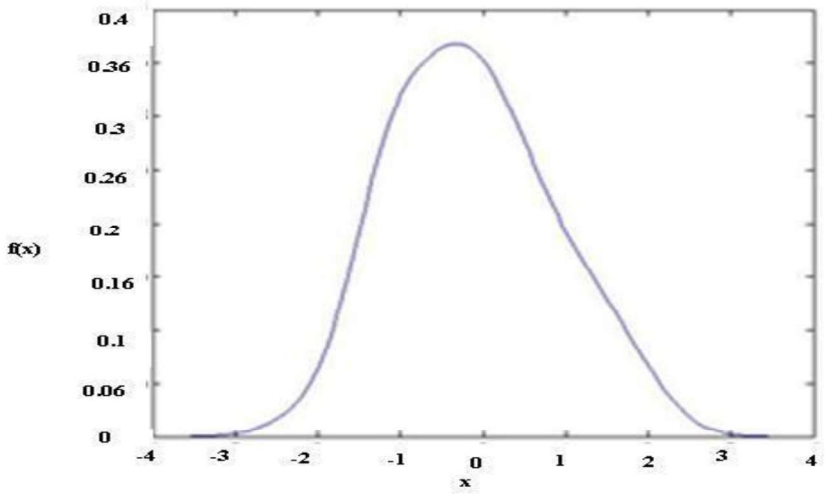


Figure 4. Probability distribution function of normal plant output without noise.

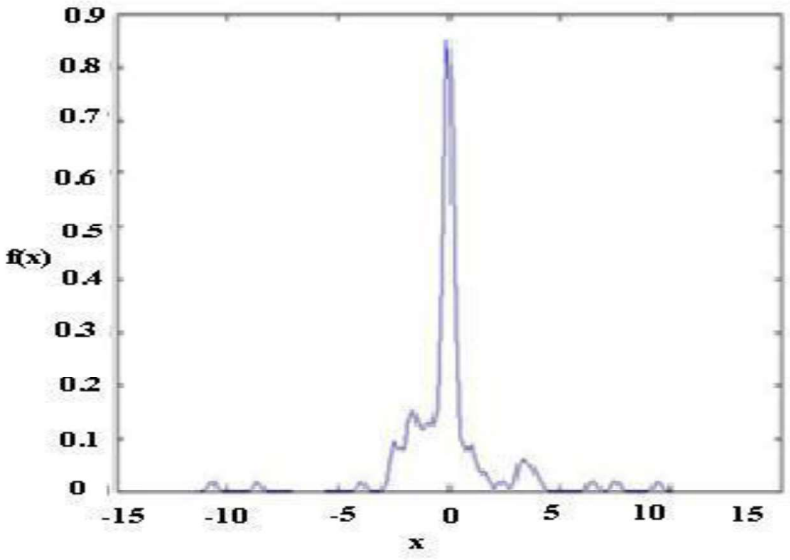


Figure 5. Probability Distribution function of plant output with output attack.

Data attacks can take place in the network as shown in Figure 6, where intruder corrupts the channel data. Data set entering the channel is replaced with a new identical data set such that the changes remain undetectable. Such an attack is also known as "Man-In-The Middle" (MITM) attack.

Let's consider that the data output from the channel is *y(k).* This output is changed or replaced to $y_1(k)$ in attack period. Relation between actual y*(k)* and manipulated measurement data $y_1(k)$ is given by equation (38) :

$$y(k) \neq y_1(k) \qquad \forall k \in K_a \text{ and } k_s \leq k \leq k_e \qquad (38)$$

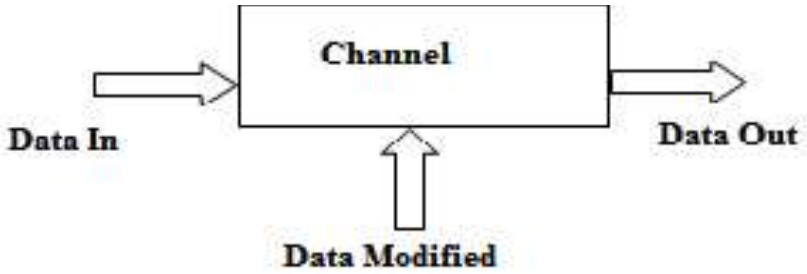


Figure 6. Data Attack on network.

4.3.2 Various Output Attacks

4.3.2.1 Stealth Attack

In stealth attack, output of the system is deviated in such a way that it is difficult to find out when the system has moved away from its normal behavior. System under attack remains operational even after attack.

Consider a general feedback controller. Assume that the output of the system is measured with sensor network made of 'p' sensors and their measurement vector '$y_i(k)$' as given by equation (39):

$$\vec{yi}(k) = \{y_1(k), y_2(k) \ldots . y_p(k)\} \quad , \quad \forall k \notin K_a \text{ and } y_i^{min} \leq y_i(k) \leq y_i^{max} \qquad (39)$$

y_i^{min} represent minimum sensor output. Let '$y_i(k)$' be modified to '$y_i^1(k)$' in attack period. '$y_i^1(k)$' denotes modified measurement of the i^{th} sensor at k^{th} time instant.

$$\vec{yi}^1(k) = \{y_1(k), y_2(k) \ldots . y_p(k)\} \quad , \quad \forall k \notin K_a \text{ and } y_i^{min} \leq y_i(k) \leq y_i^{max} \qquad (40)$$

A general model for the observed signal for k^{th} time instant is given by-

$$\vec{y}(k) = \vec{yi}(k), \quad \forall k \notin K_a \qquad (41)$$

$$= \vec{yi}^1(k), \quad \forall k \in K_a$$

For understanding let's consider a Triple Modular Redundant (TMR) architecture control system. Change in a single PLC from a group of three PLC will not affect control center decision, if final output is selected by median logic. Change will remain undetected unless someone examines the deviation in mean value across various control cycles.

4.3.2.2 Replay Attack

The attacker replays the recorded output for specified time period. In this attack, attacker is aware of all sensor reading and has the capability to inject an arbitrary control

input (u_k) into the system anytime. He can modify the measurement (y_k) by recording sufficient number of output (y) without providing any control input to it before attack and replaying it in attack period [31]. These types of attack are possible by either breaking a cryptography algorithm or inducing false sensor readings by disturbing local conditions of the distributed sensors. Replay attack can affect large system where there are many loosely coupled subsystems.

For analysis, consider a LTI system governed by state space equations (Equation. 1 and 2 from chapter 3). Let assume that the control signal 'u' is 'u'(k)' which is equal to the average of control signal for recorded period. 'u" as given by equation (42)-

$$u'(k) = \frac{\sum_{i=1}^{'n'} u(i)}{'n'} where N = \alpha * n \qquad (42)$$

Where 'N' represent number of samples that are replayed in the attack time period.

Hence the state space equation (Eq.1) is modified to equation (43).

$$x(k+1) = A\,x(k) + Bu'(k) + w_k \qquad \forall k \in K_a \qquad (43)$$

For understanding, consider Figure 7 where PLC output controls an air-conditioning (AC) system. Room temperature is recorded for specified time period and is provided as input to PLC irrespective of increase or decrease in temperature for attack duration 'K_a'. This leads in incorrect air flow and heating of the system rather than cooling effect.

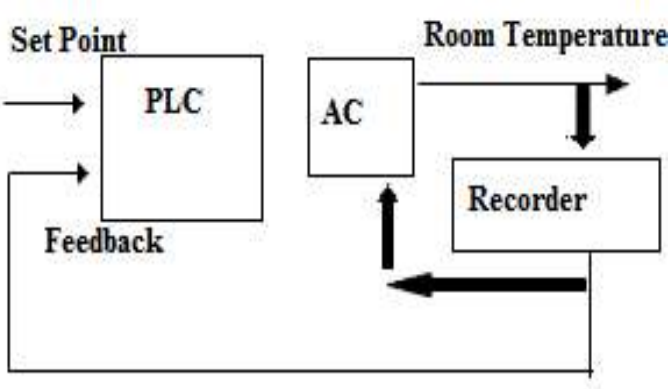


Figure 7. Replay Attack on Air conditioner.

4.3.2.3 Covert Attack

It is a close loop replay attack. Attacker never reveals its changes to the controlled device or to the controller and uses the feedback path to gain control on the system. E.g. Stuxnet attack on Iran's uranium enrichment plant was such type of covert attack.

Let's consider state space equations in attack duration given by equation (44) and (45).

$$x`(k+1) = (A+\mu)\ x(k) + Bu(k) + w_k, \qquad \forall k \in K_a \qquad (44)$$

$$y`(k) = (C+\delta)x(k) + v_k \quad . \qquad \forall k \in K_a \qquad (45)$$

States are modified by adding deviation 'μ' in matrix *'A'* or *'C'* or both for attack period 'K_a'. Similarly, attacker can add deviation 'φ' in matrix 'B' to give modified equation (46) as:-

$$x`(k+1) = A\,x(k) + (B+\varphi)u(k) + w_k \quad , \qquad \forall k\ in \in K_a \qquad (46)$$

'μ' and 'φ' doesn't results in the system deviation to unsafe zone and hence remains undetected.

4.3.2.4 Surge Attack

This attack is intended to cause maximum damage within a short duration of time. Maximum damage continues till the system does not achieve threshold value. Once the threshold is attained, the value of the output remains constant. Using SPRT knowledge from chapter (3) for threshold analysis, the threshold value '$\mathcal{I}$' in surge attack is as given by Eq (47).

$$\mathcal{I} = S_i(k) + \sqrt{\left(\left|\left|\widehat{y_i(k)} - y_i(k)\right|\right|\right)} - b_i \qquad (47)$$

Hence to stay at threshold, the attacker needs to solve the quadratic equation given by equation (47) [7].

Where '$S_i(k)$' represents CUSUM statistic for sensor 'i', $\widehat{y_i(k)}$ is observed measurement value while $y_i(k)$ represents actual value and 'b_i' is a small positive value added.

Example: - Consider a wind power plant where energy is generated based on the speed of wind drives rotor controlled by a PLC output. The generated current from rotor movement is then fed into the power grid by a transformer station. A sudden increase or decrease in speed of rotor by change in the PLC logic or in the PLC output, can cause change in energy production drastically and result in shut down of the plant.

4.3.2.5 BIAS Attack

In this attack, the attacker adds a small constant or 'bias' at every time step, as a result the output generated is more than the actual value by 'n' times of the bias value at n^{th} time instant. For e.g., consider a proportional controller is the system under control. Let $\tilde{y}$ be the actual output and with the addition of bias the output changes to $\hat{y}$ that exceed the actual value by a bias value 'c'. Hence the actual output value is given by Eq (48).

$$\tilde{y} = \hat{y} - c \quad (48)$$

For i^{th} sensor value Eq. (48) is rewritten as Eq.(49)

$$\widetilde{y_i} = \widehat{y_i} - c_i \in Y_i \quad (49)$$

Where Y_i denotes measurement value range (y_{min}, y_{max}) for i^{th} sensor. $\widetilde{y_i}$ is the actual output and $\widehat{y_i}$ is the observed output of i^{th} sensor.

Hence, SPRT equation modifies to.Eq. (50) in bias attack.

$$\mathcal{J} = S_i(k) + \sum_{k=0}^{n-1}(c_i) - n(b_i) \quad (50)$$

As per above said equation, the bias value is calculated and added in every step so as to drift the system in attack period. Bias value depends upon number of steps 'n'. If 'n' value is less, the bias generated is high resulting in maximum damage in short span and vice

versa for higher value of steps. If controller is replaced with PI or PID the bias value may varies. Integrator and differentiator may add bias value if connected to normal controller. (PI and PID analysis is kept out of our research scope.)

4.3.2.6 Geometric Attack

This attack is similar to bias attack, however, in this attack the attacker drifts the value very slowly at the beginning and maximizes the damage at the end (Alvaro A. et.al 2011) [7]. Geometric attack is based on geometric progression as given by Eq.(51).

$$\sum_{i,j=0}^{n} a_i\, r^j = a_0 r^1 + a_1 r^2 + \cdots \ldots \ldots a_i r^j \qquad (51)$$

$\sum_{i,j=0}^{n} a_i\, r^j$ represent second norm of difference between actual and estimate measurement output value.

Hence, bias (difference in actual and measured values) can be given by Eq. (52).

$$||\widehat{y_i(k)} - y_i(k)|| = \sum_{i,j=0}^{n} a_i\, r^j \qquad (52)$$

Threshold value for this attack modifies as given by Eq. (53):-

$$\mathcal{J} = S_i(k) + \sqrt{\left|\left|\sum_{i,j=0}^{n} a_i\, r^j\right)\right|} - (b_i) \qquad (53)$$

Hence to stay at threshold, the attacker needs to solve the quadratic equation given by equation (53) [7] similar to Equation (47).

Example: - consider a wind power plant. If the speed of the rotor is increased in geometric progression, the plant will still remain operational until the input does not move it into unsafe zone. Similarly, if an air conditioner is controlled by a PLC is subjected to geometric decrease in the temperature then the room temperature will start dropping slowly at the beginning and then drastically at the end. This action results damages to the devices that were to operate at a particular room temperature.

4.3.2.7 Deception Attack

In deception attack, the attacker modifies the states or replays the data such that the resultant output is different than the actual output (Alvaro A. et.al 2011) [5,6,7].

In this attack, 'y' and 'u' values are modified due to any of the attacks described (surge, geometric, bias etc).

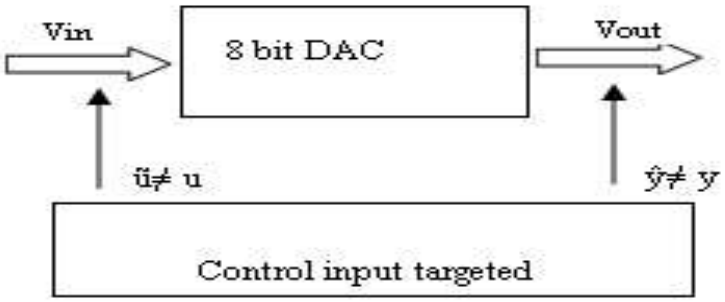


Figure 8. Deception attack on 8-bit DAC.

For illustrations assume the system under control is a digital to analog converter (DAC) as shown in Figure 8 with V_{in} and V_{out} as input and output voltages respectively. Modification of a single bit by an intruder can result in control input to D/A to change. Change in reference signal of DAC can completely change the output that is compared with this reference value.

4.3.2.8 Denial of Service Attack (DoS)

In DoS, the attacker prevents the legitimate user from gaining access to system or network. Attacker sends malicious data traffic in the attack period and may corrupt the controller software with a buffer overflow attack. The attacker prevents the actual signal from reaching the controller by either

a. Flooding of a network

b. Disrupting connections between them.

Attack model can be given for such system as mentioned in equation (54) and (55) for sensor and actuator signal output respectively.

$$\widehat{y'_1}(k) = \begin{cases} y_i(k), & \forall k \notin Ka \\ y'_i(k-1), & \forall k \in Ka \end{cases} \quad (54)$$

$$\widehat{u'_1}(k) = \begin{cases} u_i(k), & \forall k \notin Ka \\ u'_i(k-1), & \forall k \in Ka \end{cases} \quad (55)$$

The value at k-1 instant is considered as last known good value.

Equations (54) and (55) define a model of the system considering the classical "Last Known Good (LKG)" value which the system has latched. Consider an AC system controlled by a PLC system as shown in Figure 9. PLC system represents the controller while AC as the physical system. Let us assume that a switch controls the controller output and input from PLC. A kind of DoS attack would mean that the either of the switches remain open, causing systems to starve for input signal.

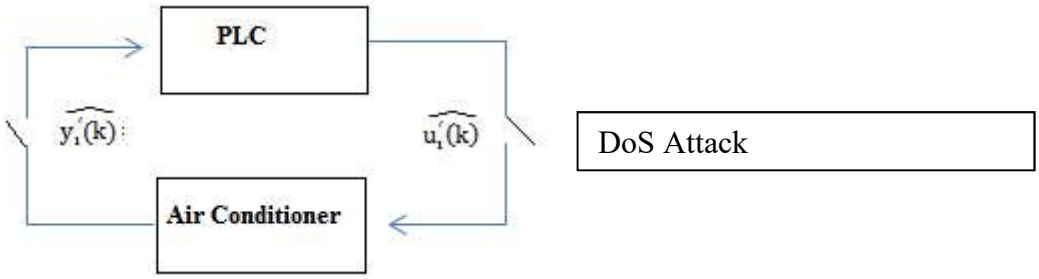


Figure 9. Denial of Service Attack.

4.3.2.9 Direct Attack

In this attack, attacker varies controller states as well as its output by subjecting controller input to uncertainty (Δ). Consider a general feedback system for analysis where controller output is given by *u(k)* at the 'kth' time instant. When the system is subjected to uncertainty (Δ) during attack period (K_a = *{k_s,....k_e}),* it gets added to the control signal and the resultant control signal output $\widehat{u'_1(k)}$ is given by Eq (56)

$$\widehat{u_i'(k)} = \begin{cases} u_i(k), & \forall k \notin Ka \\ u'_i(k) + \Delta, & \forall k \in Ka \text{ and } ks \leq k \leq ke \end{cases} \quad (56)$$

4.3.2.10 Min and Max Attack

In minimum attack, output of the sensor or actuator is subjected to minimum input value during attack period and vice versa for maximum attack.

Consider a general feedback system for analysis. Minimum attack model is given as below.

For sensor signal output ($\widehat{y_i^{min}(k)}$) is given by

$$\widehat{y_i^{min}(k)} = \begin{cases} y_i(k), & \forall k \notin Ka \\ y_i^{min}, & \forall k \in Ka \text{ and } ks \leq k \leq ke \end{cases} \quad (57)$$

Sensor output ($\widehat{y_i^{min}(k)}$) is equal to plant output $y_i(k)$ in normal state. During attack duration the sensor output is changed to minimum output $y_i^{min}(k)$.

For actuator signal output ($\widehat{u_i^{min}(k)}$) is given by

$$\widehat{u_i^{min}(k)} = \begin{cases} u_i(k), & \forall k \notin Ka \\ u_i^{min}, & \forall k \in Ka \text{ and } ks \leq k \leq ke \end{cases} \quad (58)$$

As per Eq (58) actuator output ($\widehat{u_i^{min}(k)}$) is equal to control signal output $u_i(k)$ in normal state. In attack duration it changes to $u_i^{min}(k)$.

For Maximum attack, the sensor and actuator output is as given as below.

For sensor signal output ($\widehat{y_i^{max}(k)}$) is

$$\widehat{y_i^{max}(k)} = \begin{cases} y_i(k), & \forall k \notin Ka \\ y_i^{max}, & \forall k \in Ka \text{ and } ks \leq k \leq ke \end{cases} \quad (59)$$

As per Eq (59) Sensor output ($\widehat{y_i^{max}(k)}$) is equal to plant output $y_i(k)$ when system is normal state. In attack duration the sensor output changes to maximum i.e. y_i^{max}.

For actuator signal output is ($\widehat{u_i^{max}(k)}$) given by

$$\widehat{u_i^{max}(k)} = \begin{cases} u_i(k), & \forall k \notin Ka \\ u_i^{max}, & \forall k \in Ka \text{ and } ks \leq k \leq ke \end{cases} \quad (60)$$

As per Eq. (60) actuator output ($\widehat{u_i^{max}(k)}$) is equal to control signal $u_i(k)$ in normal state and in attack duration it changes to max value i.e. $u_i^{max}(k)$.

4.3.2.11 Scaling Attack

In this attack output is scale by a factor 'α_i' in the attack duration [7].

$$\widehat{y_i^s(k)} = \begin{cases} y_i(k), & \\ \alpha_i(k)y_i(k) & \text{where } \alpha_i(k)y_i(k) \in Y_i \\ y_i^{min} & \text{where } \alpha_i(k)y_i(k) < y_i^{min} \\ y_i^{max} & \text{where } \alpha_i(k)y_i(k) > y_i^{max} \end{cases} \quad (61)$$

$y_i(k)$ *is output when* $\forall k \notin K_a$ *and* $\alpha_i(k)y_i(k)$ *when* $\forall k \in K_a and\ ks \leq k \leq ke$

$y_i(k)$ represent normal sensor output and ($\widehat{y_i^s(k)}$) in attack duration varies based on scaling factor $\alpha_i(k)$ at every time instant. If $\alpha_i(k)y_i(k)$ value is less than y_i^{min} then output considered is minimum value i.e.y_i^{min}. If $\alpha_i(k)y_i(k)$ value is greater than y_i^{max} then output considered is maximum value 'y_i^{max}'.

Actuator signal output ($\widehat{u_i^{min}(k)}$) is given by equation (62)

$$\widehat{u_i^s(k)} = \begin{cases} u(k), & \\ \alpha_i(k)u_i(k) & \text{where } \alpha_i(k)u(k) \in U_i \\ u_i^{min} & \text{where } \alpha_i(k)u_i(k) < u_i^{min} \\ u_i^{max} & \text{where } \alpha_i(k)u_i(k) > u_i^{max} \end{cases} \quad (62)$$

$u_i(k)$ *when* $\forall k \notin K_a$ *and* $\alpha_i(k)u_i(k)$ *when* $\forall k\ in \in K_a\ and\ ks \leq k \leq ke$

Actuator signal output ($\widehat{u_i^a(k)}$) in attack duration varies based on scaling factor $\alpha_i(k)$ at particular time instant. If $\alpha_i(k)u_i(k)$ value is less than u_i^{min} then output considered is minimum. If $\alpha_i(k)u_i(k)$ value is greater than u_i^{max} then output is high.

4.3.2.12 Additive Attack

In this attack sensor and actuator output is subjected to random value that gets added with the output during attack duration [7]. The attack model for additive attack can be given as-

For sensor signal output ($\widehat{y_i^s(k)}$) is given by equation (63)

$$\widehat{y_i^s(k)} = \begin{cases} y_i(k), \\ y_i(k) + \alpha_i(k) \quad \text{where } \alpha_i(k)y_i(k) \in Y_i \\ y_i^{min} \quad \text{where } \alpha_i(k)y_i(k) < y_i^{min} \\ y_i^{max} \quad \text{where } \alpha_i(k)y_i(k) > y_i^{max} \end{cases} \tag{63}$$

$y_i(k)$ *when* $\forall k \notin K_a$ *and* $\alpha_i(k)y_i(k)$ *when* $\forall k \in K_a$ *and* $k_s \leq k \leq k_e$

($\widehat{y_i^s(k)}$) is sensor output in attack duration. ($\widehat{y_i^s(k)}$) varies based on scaling factor $\alpha_i(k)$ at every time instant. If $\alpha_i(k)y_i(k)$ value is less than y_i^{min} then minimum output is considered. If $\alpha_i(k)y_i(k)$ value is greater than y_i^{max} then maximum output is considered. For rest of the attack duration scaling factor $\alpha_i(k)$ gets added to normal sensor signal output. For actuator signal output ($\widehat{u_i^a(k)}$) is given by Eq. (64)

$$\widehat{u_i^s(k)} = \begin{cases} u_i(k), \\ u_i(k) + \alpha_i(k) \quad \text{where } \alpha_i(k)u_i(k) \in U_i \\ u_i^{min} \quad \text{where } \alpha_i(k)u_i(k) < u_i^{min} \\ u_i^{max} \quad \text{where } \alpha_i(k)u_i(k) > u_i^{max} \end{cases} \tag{64}$$

$u_i(k)$ *when* $\forall k \notin K_a$ *and* $\alpha_i(k)u_i(k)$ *when* $\forall k \in K_a$ *and* $k_s \leq k \leq k_e$

Actuator signal output $\widehat{u_i^a(t)}$in attack duration varies based on scaling factor $\alpha_i(k)$. If $\alpha_i(k)u(k)$ value is less than u_i^{min} then minimum output is considered. If $\alpha_i(k)u_i(k)$ value is greater than u_i^{max} then maximum output is considered. For rest of the attack duration scaling factor $\alpha_i(k)$ get added to normal actuator signal output.

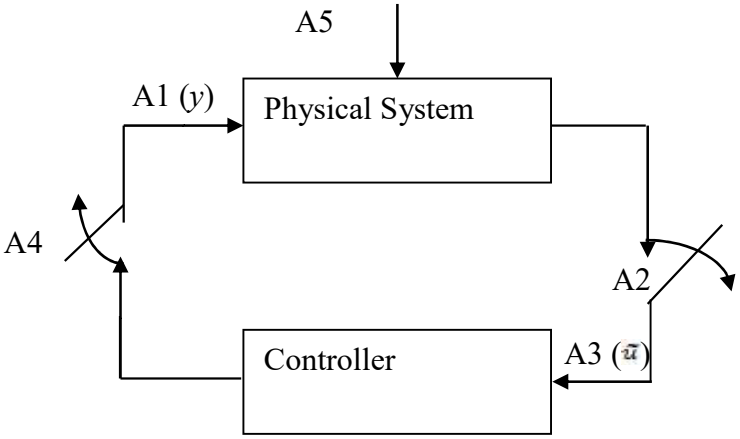


Figure 10. Examples of CPS Attack Models

Attacks described till now can be represented by an abstraction as shown in Figure 10, where we have a CPS with various sensors located across. A1, A2, A3, A4, and A5 are few sensors location where attacks can occur [7].

1. A1 and A3 represents deception attacks. The attacker launches these attacks by compromising few sensors located outside physical system. Signal (y) is targeted as shown as A1 or sensors outside controller (signals (u) is targeted as shown as A3).
2. A2 and A4 represents DoS attacks where actual signal from the controller is deprive from reaching the physical system and vice versa.
3. A5 represent a direct attack on plant due to the manipulation of physical devices directly.

4.3.2.13 State Attack Model

Consider a LTI system with A, B, and C being state space matrices. State model is given by Equation (1) and (2) [chapter 3].

$$x(k+1) = Ax(k) + Bu(k) \qquad (1)$$

$$y(k) = Cx(k) + Du(k) \qquad (2)$$

where $x(k) \in R^n$, $y(k)\, i \in R^p$, $A \in R^{nxn}$, $B \in R^{nxm}$, $C \in R^{pxn}$ and $D \in R^{pxm}$

State transition diagram for normal system is given by Figure 11. State changes from 'x_{k-1}' to 'x_k' to 'x_{k+1}' with changes in observed measurement vector being 'y_{k-1}' to 'y_k' to 'y_{k+1}'. In attack period, attacker manipulates any of the factors that results in the desired change in the state or output of the system (due to output attack). Figure 12 represents change in state transition, where measurement vector is changed from 'y_{k-1}' to 'y_k' *to* 'y^l_{k+1}' or state is varied from 'x_{k-1}' to 'x_k' to 'x^l_{k+1}'. 'y^l_{k+1}' represents manipulated measurement and 'x^l_{k+1}' *is* state of the system. In state space attack, it is assumed that an attacker has through knowledge of all state matrices as well as state vector *'x_k'* for all the time intervals. Attacker manipulate the next step *'x_{k+1}'* as well as *'y_k'* by varying current state 'x_k' and state matrices. For analyzing attack model consider a system made of 'n' sensor. State of each sensor at *'k^{th}'* time interval is given by '$x_i(k+1)$' and output measurement is given by 'y_i *(k+1)*'.

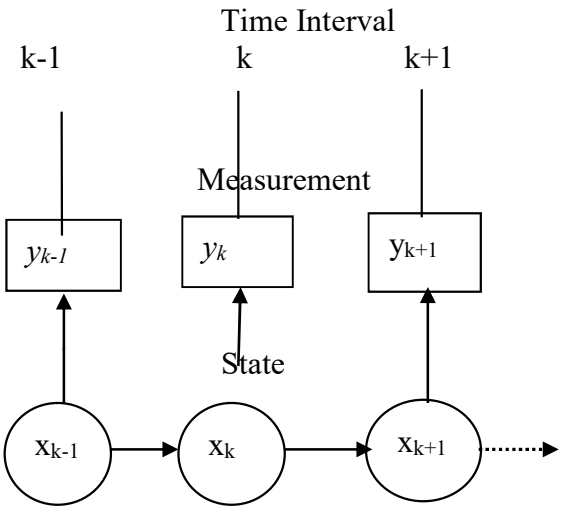


Figure 11. Normal State transition diagram

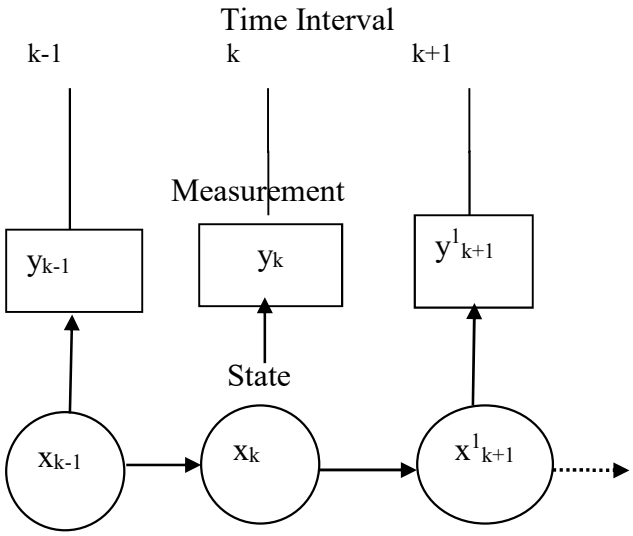


Figure 12. State transition diagram after attacker's manipulation

Various means by which states can be change are as listed below: -

1. **Variation in state space matrices** (A or B) causes change in next state ($x_i(k+1)$). Output ($y_i\ (k)$) may change at that next interval to $y_i(k+1)$ due to state ($x_i(k+1)$) change as shown in Eq. (65) and (66).

$$x_i(k+1) = A_1\, x(k) + B_1 u(k) \quad , \forall k \in K_a \tag{65}$$

$$y_i\ (k+1) = C x_i(k+1) \quad , \forall k \in K_a \tag{66}$$

2. **Variation in control input vector *u(k)*** causes change in next state value ($x_i(k+1)$) that impacts output measurement as shown in the Eq. (67).

$$x_i(k+1) = A\, x(k) + B u(k), \quad \forall k \in K_a \tag{67}$$

3. **Variation in current state vector ($x_i(k)$) to new state ($x_i^1(k)$)** causes next state ($x_i^1(k+1)$) as well as measurement value ($y_i(k)$) drift to new measurement ($y_i^1(k)$) as shown in Eq. (68) .

$$x_i^1(k+1) = A\, x_i^1(k) + B u(k), \quad \forall k \in K_a \tag{68}$$

4. Variation in noise parameter: -It's assumed that noise term is absent while analyzing the states of system. However, noise is always associated as shown in equations (69) and (70).

$$x_i(k+1) = A\,x_i(k) + Bu(k) + w_k, \qquad \forall k \in K_a \tag{69}$$

$$y_i(k) = Cx_i(k) + v, \qquad \forall k \in K_a \tag{70}$$

w_k and v_k are process and measurement noise respectively. Variations in noise parameter can results variation in system state.

4.3.2.14 Semantic Attack

Semantic attacks require deep knowledge of protocols, software, hardware and physical systems involved in an infrastructure. The more an attacker knows about the targets the better he can trigger the systems into inconsistent or dangerous states. A sequence attack is a specific type of semantic attacks. This attack concerns the misplacement of events within a sequence of ICS operations. There are two types of sequence attack: -

1. Order-based (messages or commands are sent with an incorrect/malicious order)
2. Time-based (messages or commands are sent with an incorrect/malicious timing). E.g. Water hammers effect.

Sequence-aware intrusion detection system (S-IDS) are used to detect specific type of semantic attack after studying sequences of events. Generic intrusion detection systems cannot recognize semantic attacks without any knowledge of the infrastructure and the physical processes under control. However, S-IDS can become unmanageable with large data set [12].

4.3.2.15 False Data Injection Attack

False Data injection attack is a special type of state attack. In this attack, either the sensor data or control input is falsified through injection with the intent of driving the system beyond operational parameters. The attacker aims to create a new attack vector $x_i^l(k)$ that results in wrong estimation of state variable(s) that remain undetected as shown in Figure 13. For false data injection attack analysis, it is assumed that system (as shown in the diagram) is equipped with a Kalman filter (Estimator), a controller and a detector for monitoring the innovation value change [28]. There are sensors that provide reading to state estimator so as to trigger change in controller value. Based on the attack vector selected, false data injection attack is categorized as:

1. **Random False Data Injection Attack:** - Attack vector is selected at random to insert arbitrary error into state variables estimates.

For analysis, let '$a = (a_1, a_2, \ldots, a_m)^T$' be an attack vector representing the malicious data added to the original measurement vector '$z = (z_1, z_2, \ldots, z_m)^T$'.

Let z_a (equal to z + a) represent the resulting modified measurement vector. Let 'x_{bad}' and 'x_o' represents the estimates of 'x' with manipulated measurements 'z_a' and original measurements 'z' respectively. Where $x_{bad} = 'x_o + c'$, and 'c' is the estimation error introduced by the attacker.

For random false data injection attack, the attack vector 'a' is selected at random so that output (z_a) is unpredictable.

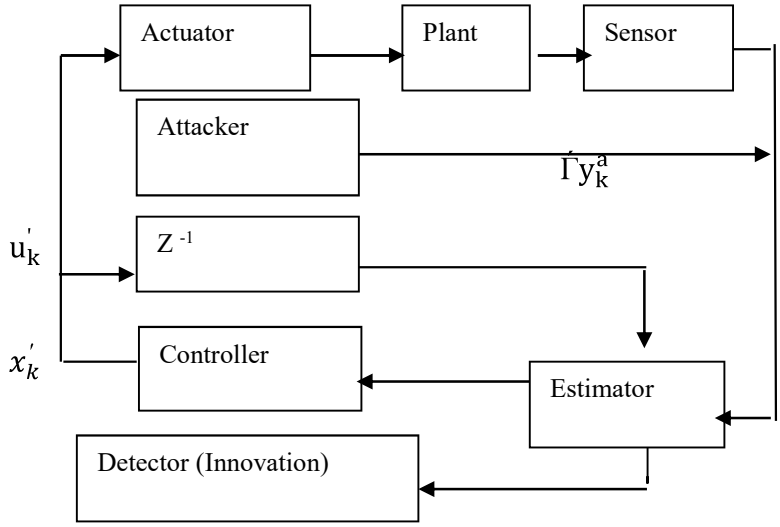


Figure 13. Schematic diagram of compromise Sensor in control Plant (Yao Liu et al. 2011)

2. **Targeted False Data Injection Attack**: -Attack vector injects only specific error/s into certain state variables (Yao Liu et.al.2011) [28]. A targeted attack is classified as constrained or unconstrained. In constrained attack, the attacker aims to find an attack vector that does not pollute the estimates of other state variables. However, in unconstrained attack the situation is vice versa (Yao Liu et.al.2011) [28]. For targeted false data injection attack, attack vector *($a = (a_1, a_2, \ldots, a_m)^T$)* is selected with predicted changes in *'z_a'*.

4.4 Representative Attack model for False Data Injection Attack

All the basic Kalman filter equations [20] as listed below (Eq. (71) to Eq. (80)) are assumed to hold good for designing of attack model.

$$\hat{x}_{0|1} = x_0 \tag{71}$$

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k \tag{72}$$

$$\hat{x}_{k+1|k+1} = A\hat{x}_{k+1|k} + K\left(y_{k+1} - \hat{x}_{k+1|k}\right) \tag{73}$$

$$P_k = AP_{k-1}A^T + Q \quad (74)$$

$$K = P_k C^T (CP_k C^T + R)^{-1} \quad (75)$$

$$y_k = C_k x_k + v_k \quad (76)$$

$$\tilde{x}_k = x_k - \hat{x}_k \quad (77)$$

$$e_k^- = z_k - C_k x_{k|k-1} \quad (78)$$

$$\hat{x}_k = A\hat{x}_{k|k-1} + K_k e_k \quad (79)$$

$$P_{k+1|k} = P_k - KCP_k \quad (80)$$

In analysis it is assume that the attacker is aware of all state matrices ('A', 'B' and 'C') as well as gain factor 'K'. Attack can occur by:-

i. Alteration in the output sensor value or

ii. Alteration of 'A' Matrix or

iii. Alteration of 'B' Matrix

A: Alteration in the output sensor value

1 Consider that the attacker has manipulated the sensor reading where there are 'n' sensors and attacker takes control of a sensor subset causing the measurement equation to change to y_k' given by Eq. (81).

$$y_k' = C x_k' + v_k + \acute{\Gamma} y_k^a \quad (81)$$

Where $\acute{\Gamma}$ is diagonal matrix that makes y_k^a order equal to $(C x_k' + v_k)$. It is a diagonal matrix of elements {v1.... vn} such that v1 = 1 iff 'i' is subset of bad sensors given by S_{bad}.

2 Kalman filter equations are modified due to sensor data variation as stated by Eq (82) to ((84).

$$y_k^{'} = Cx_k^{'} + v_k + \acute{\Gamma}y_k^a \quad (82)$$

$$x_{k+1}^{'} = Ax_k^{'} + Bu_k^{'} + Ke_{k+1}^{'} \quad (83)$$

$$e_{k+1}^{'} = y_{k+1}^{'} - C(Ax_k^{'} + Bu_k^{'}) \quad (84)$$

B: Alteration of 'A' Matrix

3 State transition matrix 'A' is modified to ' Amod' that causes change in state as describe below by Eq. (85) and (86).

$$\hat{x}_{k+1|k} = Amod\hat{x}_{k|k} + Bu_k \quad (85)$$

$$\hat{x}_{k+1|k+1} = Amod\hat{x}_{k+1|k} + K(y_{k+1} - C\hat{x}_{k+1|k}) \quad (86)$$

Hence, the difference between normal and compromised state is given by Eq. (87) and (88).

$$\Delta x_{k+1} = (A - Amod)\Delta x_k + Bu_k \quad (87)$$

$$\Delta e_{k+1} = \Delta y_{k+1} - C((A - Amod)\Delta x_k + Bu_k) \quad (88)$$

C: Alternation in 'B' Matrix

Output matrix 'B' is modified to 'Bmod' leading to changes in the state as described by Eq. (89) and (90)

$$\hat{x}_{k+1|k} = A\,\hat{x}_{k|k} + Bmod\,u_k \quad (89)$$

$$\hat{x}_{k+1|k+1} = A\,\hat{x}_{k+1|k} + K(y_{k+1} - C\hat{x}_{k+1|k}) \quad (90)$$

Hence, the difference between normal and compromised state is given by Eq. (91)

$$\Delta x_{k+1} = A\Delta x_k + (B - Bmod)u_k \quad (91)$$

Residue is changed to Eq. (92)

$$\Delta e_{k+1} = \Delta y_{k+1} - C(A\Delta\hat{x}_k + (B - B\,mod)\Delta u_k) \quad (92)$$

Kalman gain is changed to Eq .(93)

$$Kmodified = P_k C^T (CP_k C^T + R)^{-1} \quad (93)$$

4.5 Simulation and Results

In this section we have simulated various control attacks, that have been explained and modeled in the section 4.3. We have considered Four tank model to illustrate control theoretical cyber-attacks. The four-tank level control system is a typical control system with nonlinear, coupling and time delays characteristics, and can be used in simulation of multivariate industrial system. It can be used as a test bed so as to test the effects of the applications of various control theories (Daniel Perez Huertas (2011)[23]).

4.5.1 Four Tank Model

The system includes two inputs (speed of pump) and two outputs (level of two tanks), where the two outputs are controlled by two inputs as shown in Figure 14. 'h_i' is the level of water in tank '*i*' *(1, 2, 3 or 4).* 'v_1' and 'v_2' are the manipulated inputs (input voltage to the pumps), 'd_1' and 'd_2' are external disturbances representing flow out of tanks three and four. 'd_1' and 'd_2' are not considered in simulation and in nonlinear equation calculations. 'A_i' is the area of Tank 'i'. 'a_i' is the area of the pipe flowing out of tank 'i'. The ratio of water diverted to tank one rather than tank three is 'γ_1' and 'γ_2' is the corresponding ratio diverted from tank two to tank four. The outputs are 'y_1' and 'y_2' (voltages from level measurement devices).

State equation for four tank system is given by (Eq. (94) to Eq. (97)) [18].) Table 1 gives initial values of the parameters. It is a fourth order system as seen from equations (94 – 97) and hence comparing with standard fourth order equations (98 – 101) helps to provide state transition matrix.

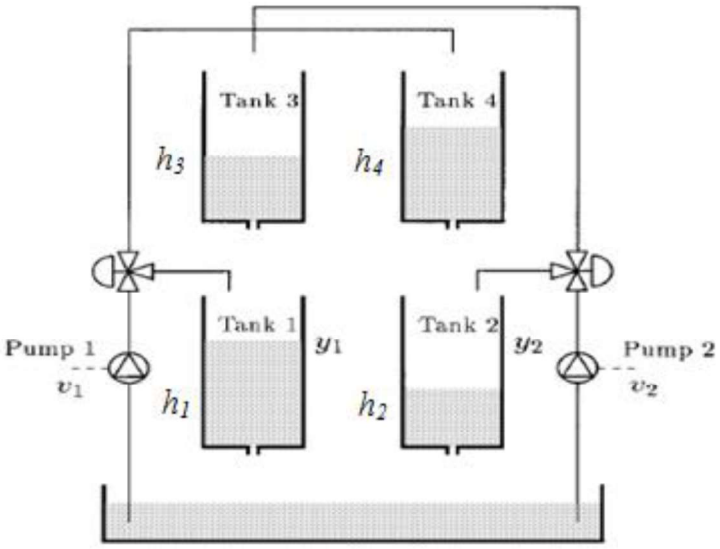


Figure 14. Four tank System

$$\frac{dh_1}{dt} = -\frac{a_1\left(\sqrt{2gh_1}\right)}{A_1} + \frac{a_3\sqrt{2gh_3}}{A_1} + \frac{\gamma_1 k_1 v_1}{A_1} \quad (94)$$

$$\frac{dh_2}{dt} = -\frac{a_2\left(\sqrt{2gh_2}\right)}{A_2} + \frac{a_4\sqrt{2gh_4}}{A_2} + \frac{\gamma_2 k_2 v_2}{A_2} \quad (95)$$

$$\frac{dh_3}{dt} = -\frac{a_3\left(\sqrt{2gh_3}\right)}{A_3} + \frac{(1-\gamma_2)v_2 k_2}{A_3} \quad (96)$$

$$\frac{dh_4}{dt} = -\frac{a_4\left(\sqrt{2gh_4}\right)}{A_4} + \frac{(1-\gamma_1)v_1 k_1}{A_4} \quad (97)$$

Consider a standard fourth order system given by Eq. (98) and (101)

$$\frac{dh_1}{dt} = a_{11}h_1 + a_{12}h_2 + a_{13}h_3 + a_{14}h_4 \quad (98)$$

$$\frac{dh_2}{dt} = a_{21}h_1 + a_{22}h_2 + a_{23}h_3 + a_{24}h_4 \quad (99)$$

$$\frac{dh_3}{dt} = a_{31}h_1 + a_{32}h_2 + a_{33}h_3 + a_{34}h_4 \quad (100)$$

$$\frac{dh_4}{dt} = a_{41}h_1 + a_{42}h_2 + a_{43}h_3 + a_{44}h_4 \quad (101)$$

State transition Matrix (A) for fourth order system govern by Eq.(98) and (101) is given by -

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \qquad A = \begin{bmatrix} -\frac{a_1(\sqrt{2gh_1})}{A_1} & 0 & \frac{a_3\sqrt{2gh_3}}{A_1} & 0 \\ 0 & -\frac{a_2(\sqrt{2gh_2})}{A_2} & 0 & \frac{a_4\sqrt{2gh_4}}{A_2} \\ 0 & 0 & \frac{a_3(\sqrt{2gh_3})}{A_3} & 0 \\ 0 & 0 & 0 & \frac{a_4(\sqrt{2gh_4})}{A_4} \end{bmatrix}$$

Comparing Eq. (94) to (97) with (98) to (101), State transition matrix of the four-tank system can be represented as shown above. B matrix is given as:

$$B = \begin{pmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_2)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{pmatrix}$$

All four levels of tanks are governed by time constant 'Ti' as given by Eq. (102): -

$$T_i = -\frac{A_i(\sqrt{2h_i(0)})}{a_i\sqrt{g}} \qquad (102)$$

Deviation in control plant behavior can be detected by innovation value. In all the mentioned attacks, it's assumed that the threat is due to an insider attack.

a_1,a_2	2.3	k_1	5.51
a_3,a_4	2.3	k_2	6.58
A_1,A_2,A_3,A_4	730	g	981
υ1(0)	60%	γ_1	0.333
υ2(0)	60%	γ_2	0.307
T_1	53.8	$h_1(0)$	14.1
T_2	48	$h_2(0)$	11.2
T_3	38.5	$h_3(0)$	7.2
T_4	31.1	$h_4(0)$	4.7

Table 1. Model parameters used for simulating four-tank system by Edward P. Gatzke et.al [18]. All units are in CGS.

We have used Matlab software for coding and simulation of four tank model and various control attacks as shown below: -

a) Normal system (T_i) innovation graph is as given in Figure 15.

For Normal plant, the controller output governed by Eq. (97) to (101 is passed through Kalman filter to generate innovation values (difference in the expected and actual output of four tank). As the system is stable the output is as generated in Figure 15.

At 100th interval innovation value attains (start of steady state response) == 0.0027

Convergence (Innovation) attained at 500th iteration: - 0.000734

Percentage of Innovation value changed: - 72.81

Number of iterations/cycles - 3

Total Time taken for execution - 265.234000 to 285 seconds.

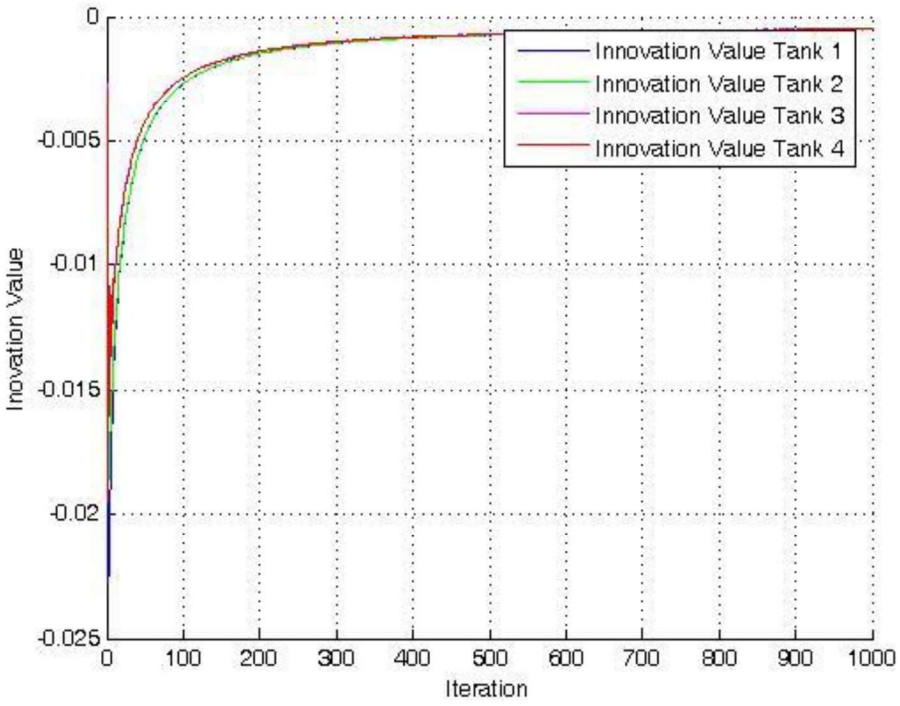


Figure 15. Normal Plant Innovation value graph

b) State Attack - (T_i Modified with orifice area 'a1' of tank *'1' of the four tank model)* as shown in Figure 16.

At 100th interval innovation value attains = -0.0076

Convergence (Innovation) attained at 500th iteration: -0.0021

Percentage of Innovation value changed: - 72.36

Number of iterations / cycles - 3

Total Time taken – 285 to 302.45 seconds.

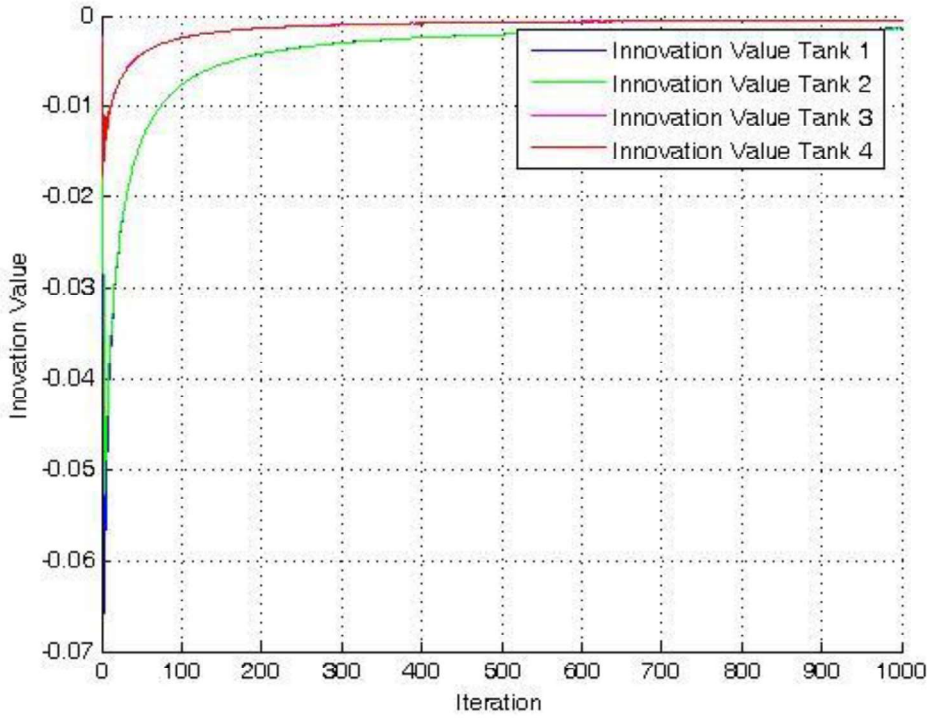


Figure 16. Plant Innovation value graph with orifice area a1 ***of tank '1' modified. (State Attack)***

c) State attack/ Input Attack – (Change in tank area 'A_i' of four tank model) as shown in Figure 17.

At 100th interval innovation value attains = - 0.0027

Convergence (Innovation) attained at 500th iteration: -0.000734

Percentage of Innovation value changed: - 72.814

Number of iterations/cycles - 2

Total Time taken for execution- 265 seconds.

Graph remains identical for all 'A_i' value change.

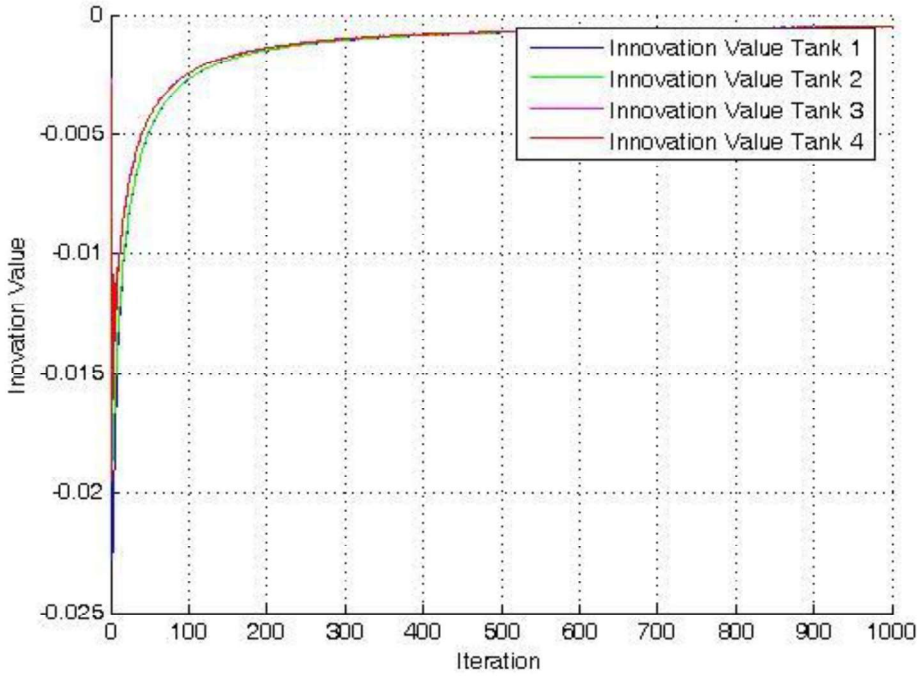


Figure 17. Plant Innovation value graph with Tank area 'A_i'of tank 'i' modified.

d) **Modifying input control vector 'u'** causes change in four-tank behavior. Innovation value graph generated as shown in Figure 18.

At 100th interval innovation value attains = -0.0027

Convergence (Innovation) attained at 500th iterations: - - 0.0007341

Percentage of Innovation value changed: - 72.811

Number of iterations - 2

Total Time taken - 300.328 seconds.

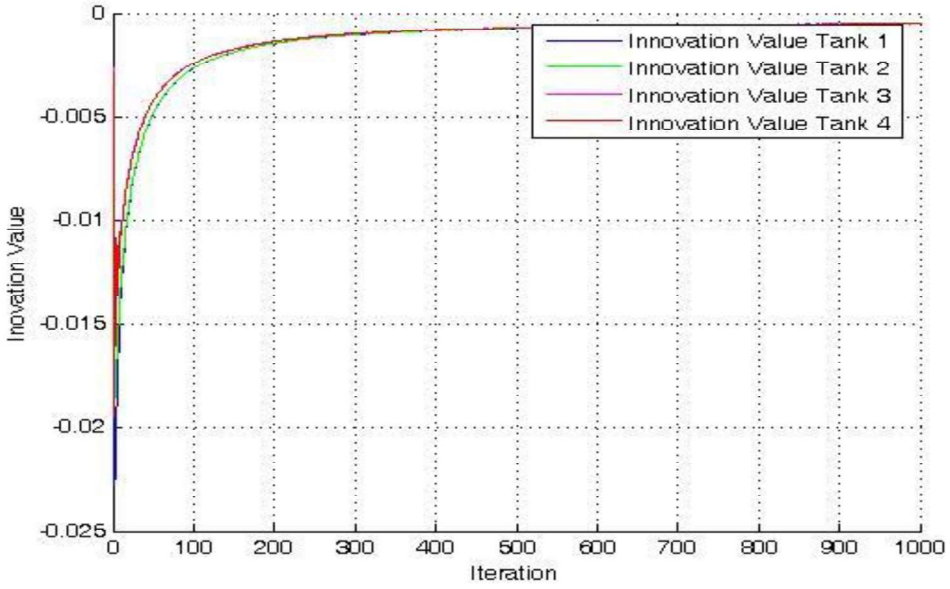


Figure 18. Plant Innovation value graph for change in input control vector 'u'

d) **Modifying noise parameter 'R' that affects Kalman gain of the Kalman filter connected to Four tank system**.

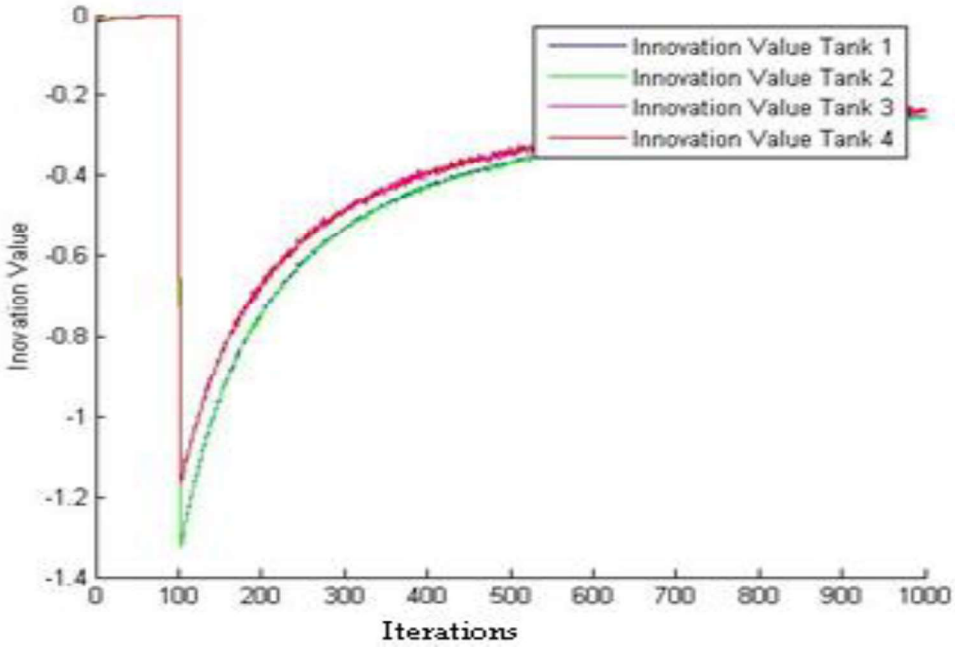


Figure 19. Plant Innovation value graph for change in Noise parameter 'R' that affects Kalman gain

Innovation value graph generated is as given in Figure 19 gives a drastic deviation to cause shutdown of the system. The key observations in change in innovation value captured in graphs (figure 15 to 19) are as listed below:

1. Change in orifice area 'a_i' of any tank causes greater impact in innovation value change.
2. A small change in tank area 'A_i' does not cause a remarkable change in innovation value. Hence, attacker has to change 'A_i' to a larger extent to achieve a noticeable change in output.
3. As 'B' matrix only depends on 'A_i' there is very less change in output innovation compared to normal behavior and difficult to detect by innovation value change.
4. Change in noise factor controlling Kalman gain (K) causes the system under control to deviate to larger extent in short time interval.

All the analysis presented above assumes that after getting innovation values, they are checked for causes of variation, i.e. whether it is from input or output end. Initially, it checks for 'B' and 'C' matrices changes and subsequently for 'A' matrix.

4.5.2 SPRT

Considering the innovation values of Kalman filter as I.I.D (Independent and Identically Distributed), any change in the innovation string can be diagnosed by use of SPRT knowledge, (SPRT explained in section [3.5]). SPRT has been analyzed for Four tank model for all the innovation values in Figure 20, 21, 22, 23 and 24 are for steady state.

SPRT is used to detect the presence or absence of a failure in the entire measurement as shown in Figure 20. Random fault was inserted at 505, 595 and 705 in normal four-tank plant set up. SPRT was tested for 1000 iteration with a window size of 10. Figure 20

shows values having non-zeroes across 500, 600 and 700 that indicate that the values are not within the acceptable limit. It indicates deviation from the normal behavior and possible chances of fault insertion. We have tried to capture fault when innovation values deviated 20% from its original zero mean values.

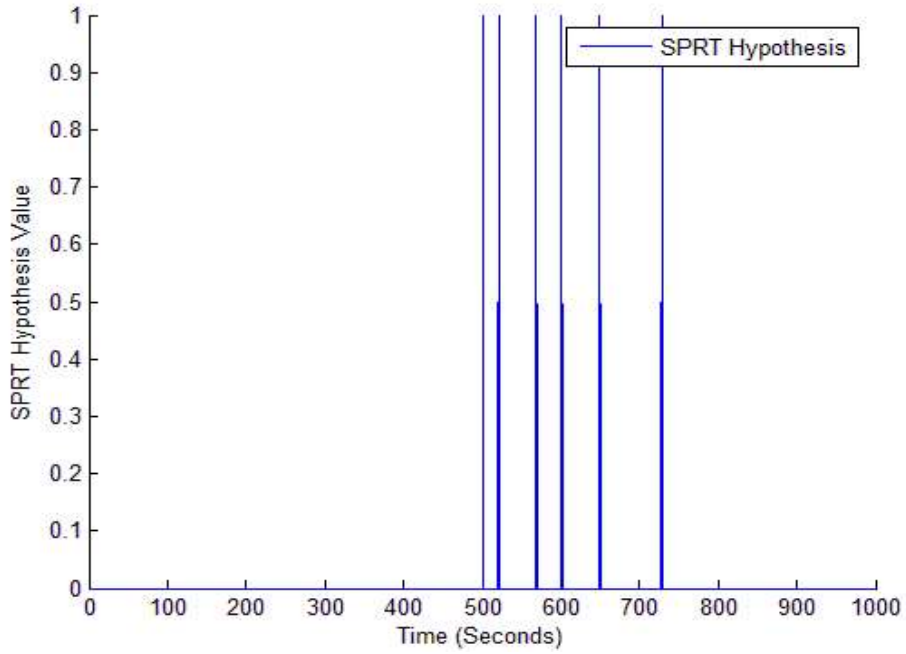


Figure 20. SPRT test for four tank control system subjected to random fault at interval 505, 595 and 705.

Simulations for SPRT were carried out after convergence, for a sliding window size of 100 as shown in the following Figure (21) to (24).

a. SPRT for Normal Plant (as shown in Figure 21)

For Normal plant (Four tank system model), the null hypothesis (as explained in section 3.5) holds true. Null hypothesis for normal plant is when there is no attack and innovation values are having constant (zero) mean and zero variance. The Figure 21 depicts SPRT for innovation values (Kalman filter output) for normal plant.

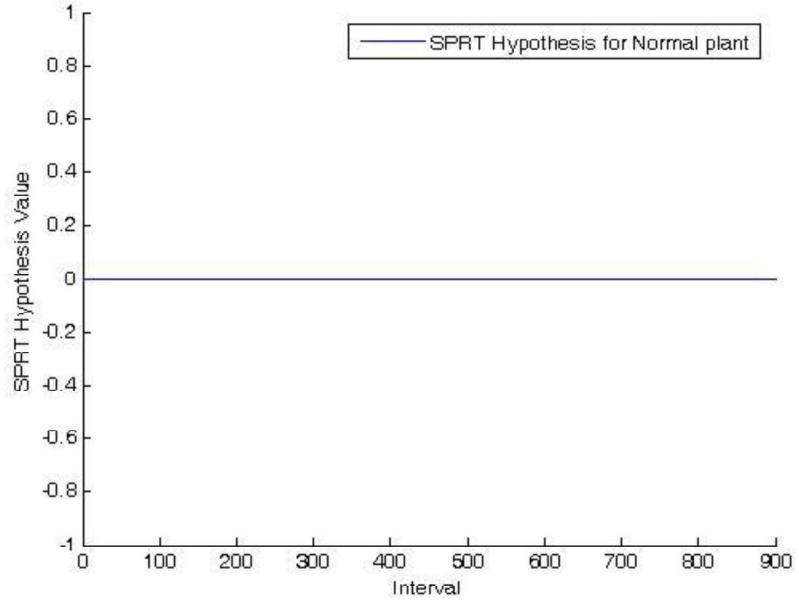


Figure 21. SPRT for Normal Plant

Mean and Variance of innovations = 0 and -5.6236 e -6

Total interval = 1000

Iteration = 2

Execution Time required = 570 seconds.

Result = There is no attack as SPRT value is equal to zero

b. SPRT for Plant with state attack as shown in Figure 22

Four tank system model is subjected to state attack (section 4.3.2.13) using simulation software (Matlab) for the innovation value output from the Kalman filter. SPRT for the said is as shown in Figure (22) with below details: -

Mean and Variance of innovations = 0 and 9.4909e-005

Total interval = 1000

Iteration = 3

Execution Time required = 410 seconds.

Result = There is a constant deviation from normal resulting in the SPRT value equal to 1

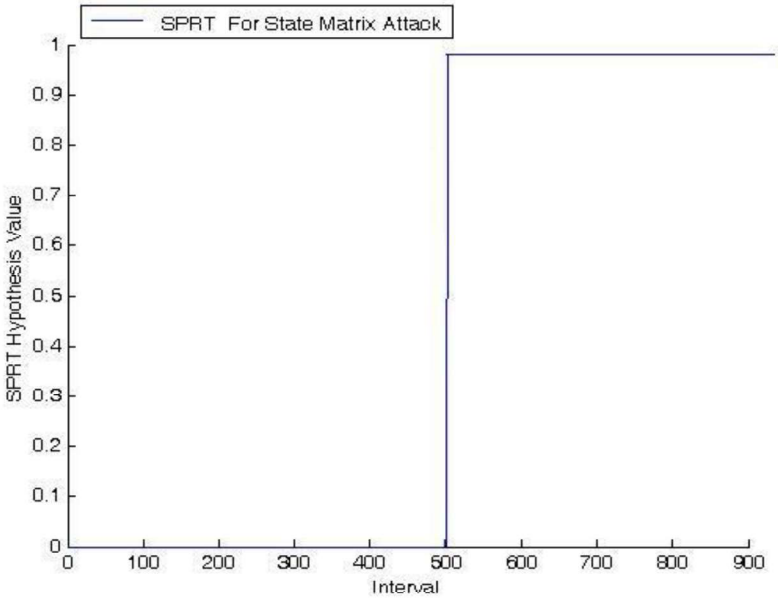


Figure 22. SPRT for State Attack

c. SPRT for Plant with Input attack as shown in Figure 23

Four tank system is subjected to input attack (section 4.3.1) using simulation software (Matlab) for the innovation value output from the Kalman filter. SPRT for the said is as shown in Figure (23) with below details: -

Mean and Variance of innovations = 0 and 5.6180e-006

Total interval = 1000

Iteration = 2

Execution Time required = 403 seconds.

Result = There are many peaks indicating possibility of attack.

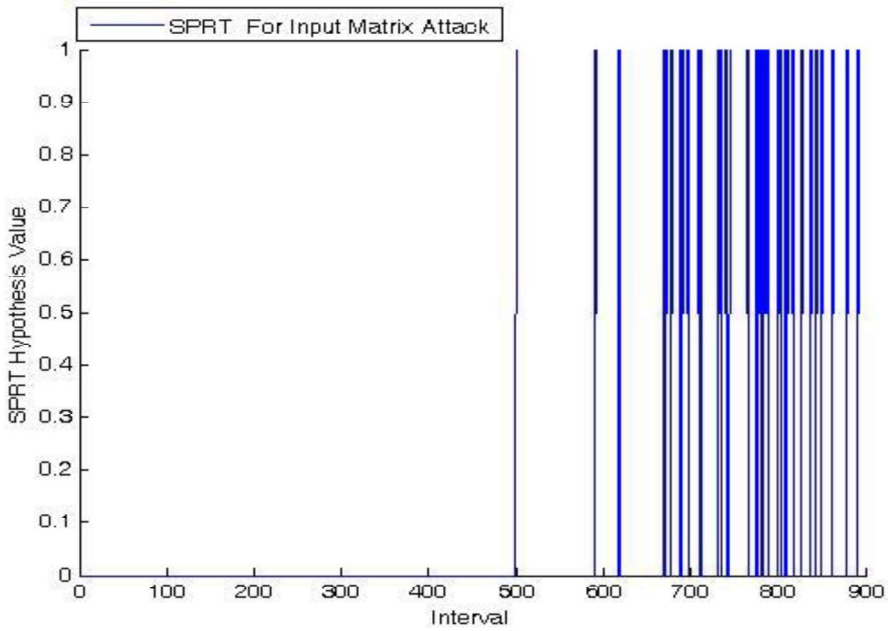


Figure 23. SPRT for Input Attack

d. SPRT for Plant with control attack as shown in Figure 24

Four tank system is subjected to the control attack (control matrix is varied) and simulated using simulation software (Matlab) for the innovation value output from the Kalman filter. SPRT for the said is as shown in Figure (24) with below details: -

Mean and Variance of innovations = 0 and 5.6236e-006

Total interval = 1000

Iteration = 2

Execution Time required = 404 seconds.

Result = the graph is identical to normal hence cannot be detected by SPRT technique.

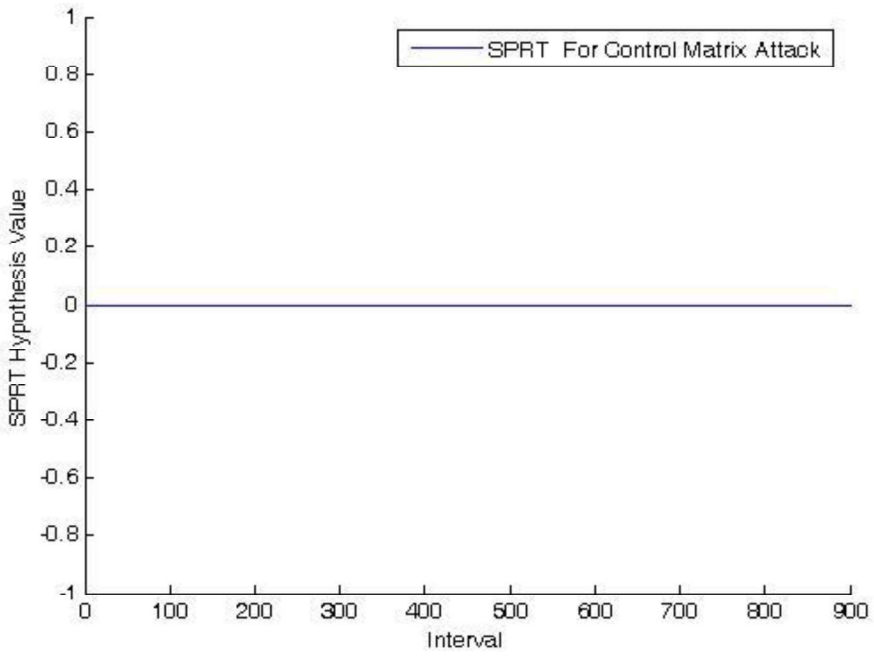


Figure 24. SPRT for Control Attack

e. Modifying Noise factor

Variation resulted due to noise cannot be detected by SPRT technique.

A SPRT technique has been used to identify various attacks as listed in the Table 2.

Sr.No	Figure Number	SPRT implementation Purpose
1	20	Identification of random faults.
2	21	Normal control plant behavior.
3	22	Indicates State Attack.
4	23	Indicates Input Attack.
5	24	Indicates Control Attack.

Table 2 :- Attacks analyzed using SPRT technique.

Following section further extend analysis using correlation and co-variance to indicate possible control aware attack.

4.5.3 Correlation

Correlation of the residue values collected from the Kalman filter controlling four-tank model can be used to indicate possible attack. Figure 25 and 26 indicate auto and cross correlation among residue values in various attack scenarios.

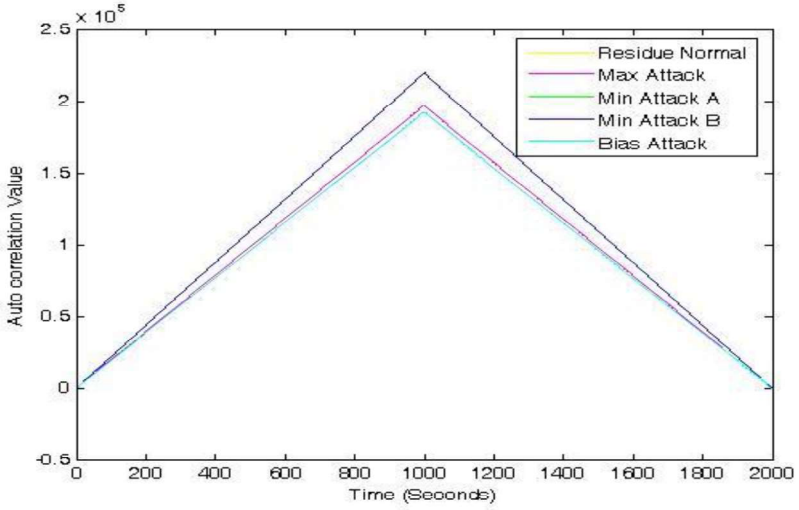


Figure 25. Autocorrelation in various attack scenarios

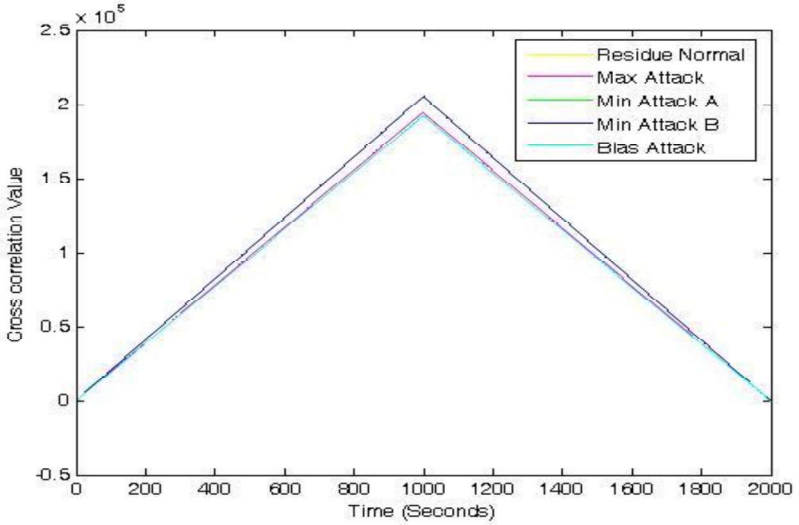


Figure 26. Cross-correlation in various attack scenarios mentioned in graph.

4.5.4 Covariance

Variance in the residue values controlling four tank model gives indication of attack. Figure 27 and 28 indicate auto and cross variance among residue values in various attack scenarios.

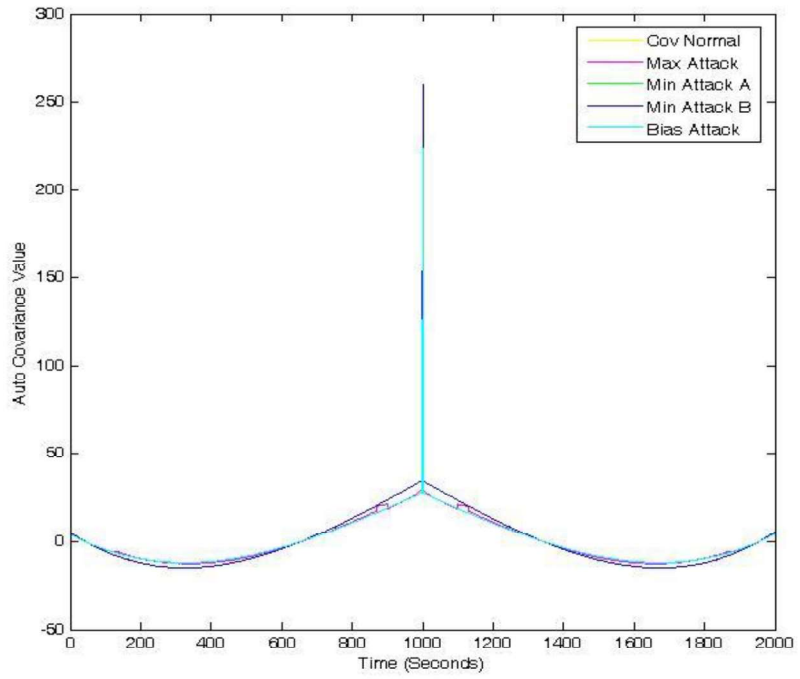


Figure 27. Autocovariance in various attack scenarios

As shown in Figure (27), the peak value varies for normal and each type of attack. Auto-covariance is first obtained for the normal plant for innovation value output from Kalman filter.

For various attacks, auto covariance graph (Figure (27) and cross covariance (Figure 28) varies with different peak giving indication that the system under control has compromise.

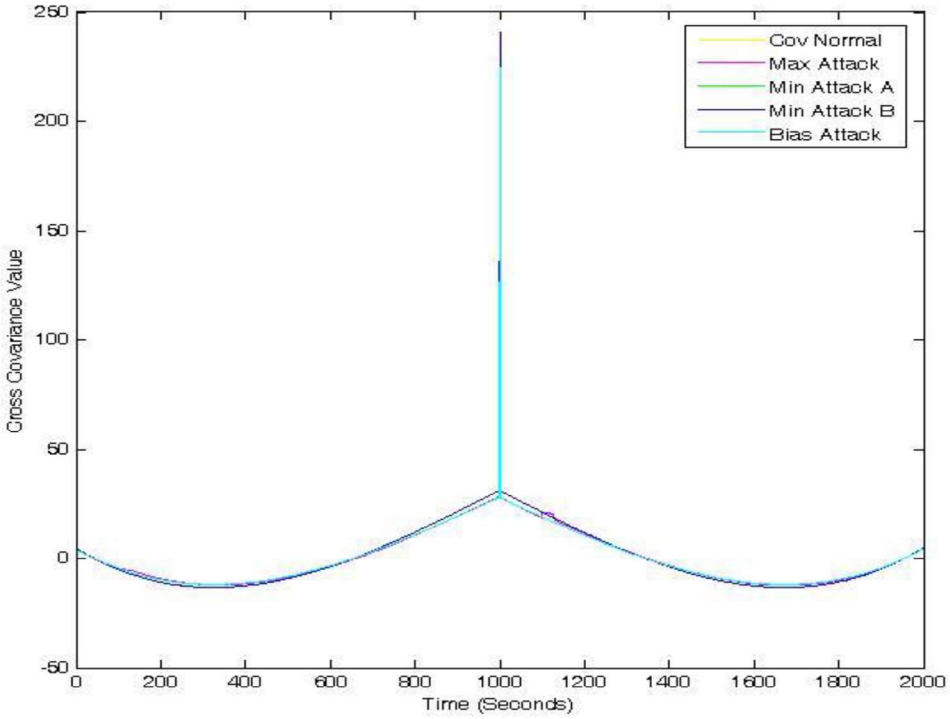


Figure 28. Cross covariance in various attack scenarios

4.5.5 Hinkley's 'CUSUM' Method

Extending the theory explained in section 3.8, we check for variation in mean value of the controller output by observing its deviations, with respect to its maximum of the cumulative sum by subjecting to suitable drift. Four tank model (as a CPS system and height of each tanks as controller output) is subjected to CUSUM test (for the innovation value output from Kalman filter that is provided with height of each tank as input). The resultant graph for CUSUM hypothesis value for innovation is as shown in Figure 29.

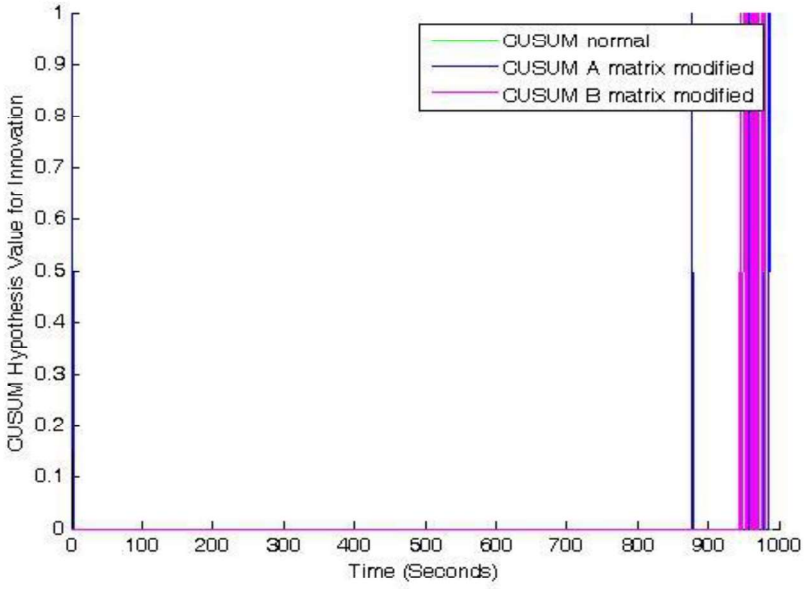


Figure 29. CUSUM test in various attack scenarios

4.6 Conclusion

Contribution: -

1. In this chapter we have identified various targeted control aware cyber-attacks. We have classified of various control aware attacks (targeted attacks) using state space equation.

2. We have also discussed techniques for securing control systems and showed the efficacy of the said techniques by incorporating changes as stated in attack model.

3. Our work demonstrated that the control attacks can be detected by innovation value change.

4. We have modeled various targeted attack using control equation knowledge from chapter 3 using Four tank model.

We have extended SPRT technique further to design security monitors in the following chapter. The proposed SPRT technique however has limitation in differentiating system

failure due to internal fault and attack. SPRT method is also not scalable to large distributed control system for which a detailed state space model is not available.

Chapter 5

Online Monitoring of a Cyber Physical System

Statistical techniques like SPRT, CUSUM, and GLR which are discussed in previous chapter are useful in analysis of small changes in controller output. In the current chapter, we extend the techniques using a analytic method based on controller output logs. Here we assume that controllers are realized in software-based systems with a support of data archiving. It is also assumed that enough data has been collected so that the statistics of its good performance recorded over time is good enough to classify as normal operation and possible abnormal behavior.

5.1 Introduction to Monitoring

Network based anomaly detection techniques are not enough to detect attacks on control systems as the access may be through a genuine access point but with the intent to change the behavior of the control loop. The latest information security tools alone are not sufficient for securing control systems. Securing the control system from possible targeted attacks on its computational elements requires a thorough diagnosis of its behavior against accepted normal behavior. Such diagnostics can be performed by a synchronous monitor (In synchronous monitor, the system under control is paused, waiting for the monitor to acknowledge back to the system before it can continue executing for every pulse/trace generated by the monitor) who reads the same set of

inputs as the controller and the output of the controller. It may be difficult to provide such systems online with the safety class controller due to complexity in providing comprehensive verification for regulatory requirements (regulatory requirement as per the implemented system like atomic energy regulatory board (AERB) designed requirement for nuclear power plant etc.). Hence, such systems need to be placed only in non-safety systems but with acquisition of data from safety systems. However, this would require large data sets that are being collected during the plant history.

The main contributions of this chapter are as follows:

1. Developing techniques to build a monitoring framework from data logs files.
2. Data mining-based computation techniques for design of online monitors such as
    a. Least Square Approximation method
    b. Computational Geometric method.
3. Experimental validation of monitors using an extensive simulation on a four-tank model.

5.2 Computational Analysis

For large historic or real time data, techniques such as SPRT, CUSUM are not sufficient and so we need computational methods that are able to handle huge log files. The following section explains the theory of computational method that is useful in designing monitor for analysis.

Various computational methods available are-

A. Euclidean distance

B. Mahalanobis distance

C. Pearson's correlation coefficient (ρ)

D. Lease Square approximation (LSA)

E. Computational geometric (Convexity) method

Least Square approximation (LSA) and Convexity (convex hull) methods are considered for our analysis.

5.2.1 Least Square approximation (LSA)

The method estimates parameters by minimizing the squared discrepancies between the observed data, and its expected values. For a given value of '*X,* ' the best prediction of '*Y* ' is given by:

$$Y = f(X) + noise$$

'Y' is regression function of '*X'* along with the noise term. It is estimated from *'n'* co-variables and their responses i.e. (x_1,y_1)……(x_n,y_n). LSA is the value that minimizes the squared distance between the vector *'Y'* and line governed by '*f(x)'* as given by Eq. (103).

$$LSA = \sum_{i=1}^{n}(y_i - f(x_i))^2 \qquad (103)$$

Constraints are checked for distance of the point from the governing line function *'f(x)'*. Consider that the line is specified by two points (x_1,y_1) and (x_2,y_2) as shown in Figure 30, then a vector perpendicular to the line is given by -

$$v = \begin{bmatrix} y_2 - y_1 \\ -(x_2 - x_1) \end{bmatrix}$$

Let 'r' be a vector from point (x_0,y_0) to the first point on the line '*f(x)'* and given by-

$$r = \begin{bmatrix} x_1 - x_0 \\ (y_1 - y_0) \end{bmatrix}$$

The distance ‘d’ from (x_0,y_0) to the line is given by projecting ’r’ onto ‘v’ as shown in Figure 30 and equation (104).

$$d = |\hat{v}.r| = \frac{(x_2-x_1)(y_1-y_0)-(x_1-x_0)(y_2-y_1)}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}} \qquad (104)$$

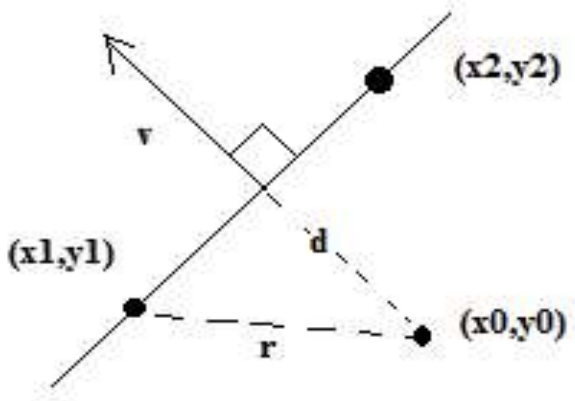


Figure 30. Distance of a point from line segment

5.2.2 Computational geometric (Convexity) method

Convex hull of a set of points ‘S’ is the smallest convex set that contains S. A convex hull is also known as convex envelope. Convexity is a geometric property used in computing the smallest convex shape termed as 'convex hull’ enclosing set of points. Given an ordered triplet of points (p, q, r) in the plane, it is said to have positive orientation if it defines a counterclockwise oriented triangle, negative orientation if it defines a clockwise oriented triangle, and zero orientation if they are collinear as shown in Figure 31. Orientation is defined as the sign of determinant given by-

$$\text{Orient}\,(\text{p},\text{q},\text{r}) = \begin{vmatrix} 1 & px & py \\ 1 & qx & qy \\ 1 & rx & ry \end{vmatrix}$$

Orientation is positive if p < q. It is zero for p = q, and negative for p> q as shown in Figure 31.

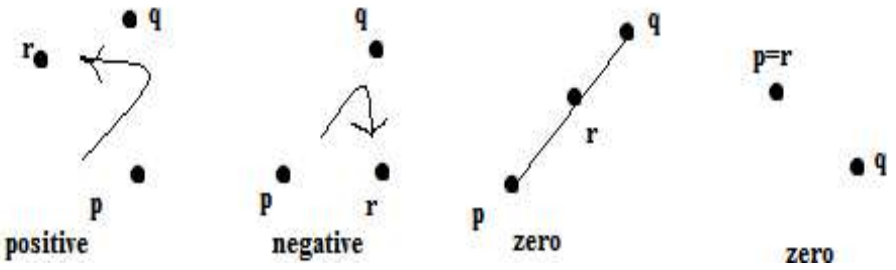


Figure 31: - Orientation of ordered triple of points (p, q, r).

Let's consider four points A, B, C, and D and a reference intersection point 'X' as shown in Figure 32. If a ball (circle) is drawn from 'X' to all points with radius r_1, r_2, r_3 and r_4 then the convex hull is subset of total area cover by all balls as given by Eq. (105).

$$conv(\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D}) \subseteq \bigcup_{i=1}^{4} B(\underset{ri}{\rightarrow}) \qquad (105)$$

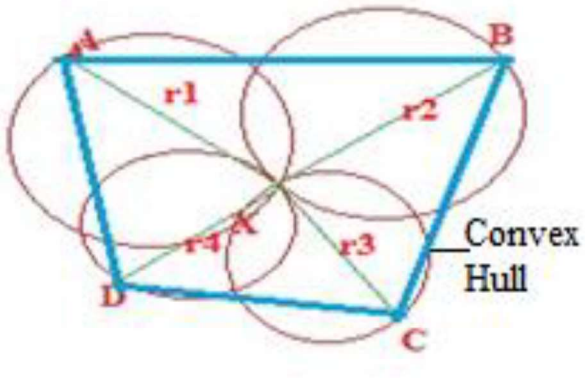


Figure 32. Convex hull bounded by circles.

For understanding consider an actual stream consists of data point sets $\{u_1,u_2,\ldots u_n\}$ and convex hull ($\vec{v}$) that governs the normal behavior of system given by Eq.(106).

$$\vec{v} = conv\ (u_1, u_2, \ldots u_n) \qquad (106)$$

Consider a new data stream $\{p_1,p_2,\ldots p_n\}$ and $\{q_1,q_2,\ldots q_n\}$ each represented by their convex hulls $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ respectively. If $\overrightarrow{v_1}$ satisfies condition mentioned in the following equation (Eq. (107)) then the data monitored is considered as safe as shown in Figure 33.

$$conv\ (p_1, p_2, \ldots p_n) \subseteq \overrightarrow{v} \qquad (107)$$

If the observed data is outside the desired convex hull and tolerance band, then it's considered unsafe because for any physical system, changes in the convex hull indicate a possible change in the state matrix that can lead the system under control to become uncontrollable.

Figure 33 illustrate that $\overrightarrow{v_2}$ representing $conv\ (q_1, q_2, \ldots q_n)$ is out of range of $\vec{v}$ and can be considered as an indication of attack on control system.

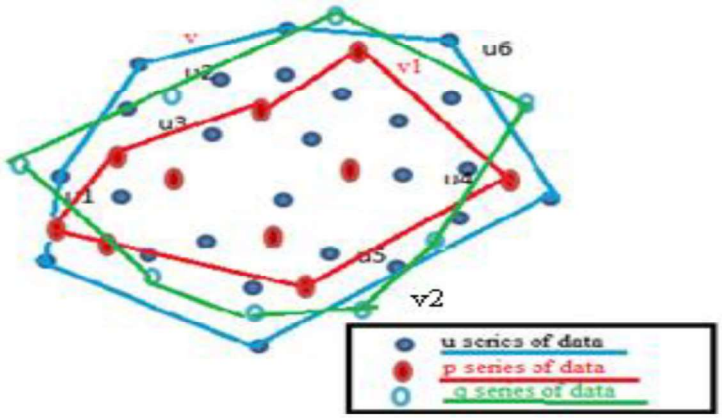


Figure 33. Convex Hull – Geometric Interpretation

5.2.3 Representation of convex hull as a set of constraint functions

Convex hull can be considered as a polygon created with the intersection of multiple lines. Every line in two-dimensional spaces is represented with an equation given by -

$$y = mx + c$$

where 'm' represents slope of the line and 'c' is a constant.

Convex hull needs to satisfy all the constrained functions of each line that bounds the hull as shown in Figure 34. For understanding consider an polygon ABCDE bounded by line

functions such as like x ≤ 0, y ≤ 0, x + y ≤ 5, y ≤ 4, 5x + 4y ≤ 20 and represented as figure 34.

$$f(x,y) = \begin{cases} x \leq 0 \\ y \leq 0 \\ x + y \leq 5 \\ y \leq 4 \\ 5x + 4y \leq 20 \end{cases}$$

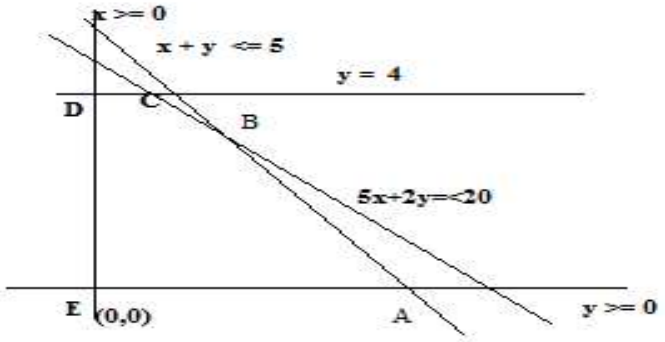


Figure 34:- Polygon as a set of constrained function.

5.3 Concept of Monitoring in CPS

Unauthorized access to the control system can lead to physical damage to the system. An attacker can intentionally (termed as targeted attack) or unintentionally (termed as non–targeted attack) manipulate controller characteristics. An attack can cause the observed system to behave abnormally. Parameters that control the controller output are subjected to various filter criteria to generate alerts.

Common design methodologies for monitor are as mentioned below: -

A. Single parameter Method - This is the simplest method for monitoring large data streams. Parameters that are monitored are average, threshold, min, max etc.

B. Set based Method - A set of variables and threshold functions over the sets are used.

C. Frequency Count Method - Frequency of the observed data is used as key factor in design. If frequency count rises above or falls below a predefined threshold, an alarm is triggered.

D. Feature Method – Control system properties are monitored based on their importance.

E. Computation Method - Controller output from a control plant is analyzed as streams of data. The method can be explained as –

Let us consider S = $\{s_1, s_2 \ldots s_n\}$ to be input data streams to the controller for which the corresponding controller output data stream is $\{v_1(t).....v_n(t)\}$. If we assign w_1, w_2...,w_n as positive weights to the controller data streams, we can define global statistic value (v(t)) as given by Equation (108) that remain constant over the sample set.

$$v(t) = \frac{\sum_{t=1}^{n} w_i v_i(t)}{\sum_{t=1}^{n} w_i} \quad (108)$$

If the derived controller output changes to $\{v'_i(0) \ldots\ldots v'_i(t)\}\}$, then the estimated output changes to $(e(t))$ as shown in equation (109).

$$e(t) = \frac{\sum_{t=1}^{n} w_i v'_i(t)}{\sum_{t=1}^{n} w_i} \quad (109)$$

The drift in the estimated value is given by $u_i(t)$ as shown in equation (110).

$$u_i(t) = e(t) - v(t) \quad (110)$$

Drift $(u_i(t))$ act as a monitoring parameter for the observed data stream.

An ideal monitoring system is designed from controller perspective has to support various features as listed.

5.4 Requirement of CPS Monitoring system

1. In the attack period, for an non-zero input '$u_1(k)$' the deviation in output '$y(k)$' to '$y_1(k)$' should be detectable as represented by following Equation (111) and (112).

$$x_1(k+1) = Ax(k) + Bu_1(k), \quad \forall k \in K_a \text{ and } ks \leq k \leq ke \qquad (111)$$

$$y_1(k) = Cx(k), \quad \forall k \in K_a \text{ and } ks \leq k \leq ke \qquad (112)$$

where $x_1(k+1)$, $x(k)$, $u(k) \in R^n$, and $u_1(k)$, $u_1(k) \in R^n$, and $u_1(k)$, $u_1(k) \neq 0$. $x_1(k+1)$ and $x(k)$ represents the next and current state of the system. *'k'* represents the time interval, while $y(k)$ represents the measurement output and $u(k)$ is the control input at k^{th} time interval.

2. If the monitor fails to detect the changes in the controlled device, then the attack remains undetectable.

3. Changes caused in the controller's state or control system output or both have to be detected by the monitor. Methods by which the *s*tate or output of the system can be varied are:

a) Variation in state space matrices,

b) Variation in control input vector '$u(k)$' ,

c) Variation in current state vector '$x_i(k)$' *to* ' $x_i^1(k)$' .

d) Variation in noise parameters.

a) **Variation in state space matrices** that causes change in next state '$x_i(k+1)$' and output '$y_i\ (k))$' as given by Eq. (113) and (114).

$$x_i(k+1) = A_1x_i(k) + B_1u(k), \quad \forall k \in K_a \qquad (113)$$

$$y_i(k) = Cx_i(k), \quad \forall k \in K_a \qquad (114)$$

b) **Variation in control input vector '*u(k)*'** result changes in the next state value $x_i(k+1)$ as given by Eq. (115).

$$x_i(k+1) = Ax_i(k) + Bu(k), \quad \forall k \in K_a \qquad (115)$$

c) **Variation in current state vector '$x_i(k)$' to '$x_i^1(k)$'** can cause next state to change to $x_i^1(k+1)$ as well as measurement value '$y_i(k)$' to drift to '$y_i^1(k)$' as given by Eq.(116) and (117).

$$x_i^1(k+1) = Ax_i^1(k) + Bu(k), \quad \forall k \in K_a \qquad (116)$$

$$y_i^1(k) = Cx_i^1(k), \quad \forall k \in K_a \qquad (117)$$

d) **Variation in noise parameter-**can cause changes in both controller's state and output measurement as given by Eq. (118) and (119).

$$x_i(k+1) = A_1x_i(k) + B_1u(k) + w_k, \quad \forall k \in K_a \qquad (118)$$

$$y_i(k) = Cx_i(k) + v_k, \quad \forall k \in K_a \qquad (119)$$

As shown in the above equations w_k and v_k are process and measurement noise that can be varied in the attack period.

4. Observation criterion for the monitor can fail for the mentioned scenarios such as-

a. Output is zero irrespective of any input value as shown by Eq. (120) and (121).

$$y_i(k) = 0, \quad \forall k \in K_a \; and \; u(k) \neq 0 \; or \; u(k) = 0 \qquad (120)$$

$$y_i(k) = 0, \quad \forall k \in K_a \; and \; u(k) \neq 0 \; or \; u(k) = 0 \qquad (121)$$

c. Output is identical irrespective of any input value, in attacks as well as in non-attack period as shown by Eq. (122).

$$y_i(k) = y(k), \quad \forall k \in K_a \; and \; u(k) \neq 0 \qquad (122)$$

d. The non-zero difference in the output value satisfies the threshold limit and remains undetectable as shown by Eq. (123).

$$y_1(k) - y(k) \neq 0 \; and \; y_1(k) - y(k) \leq threshold, \qquad (123)$$

$$\forall k \in K_a \; and \; u(k) \neq 0$$

d. If noise term (i.e.*(A, w_k))*' and '*E(C, v_k))*') are uncorrelated, attack can remains undetectable. '*E*' represents expected value.

General architecture of an anomaly-detecting controller that can act as a monitor is depicted in Figure 35.

5.5 Architecture of an anomaly-detecting controller

The controller delivers the next data if it is in safe zone or sends alarms to keep the operating plant in safe mode till appropriate corrective methods are implemented.

As shown in Figure 35, let $\underrightarrow{U(t)}$ be the control plant input based on the set point and feedback based on sensor measurements'. Let $\underrightarrow{X'(t)}$ and $\underrightarrow{X(t)}$ be plant and sensor outputs, respectively. Output of controller $\underrightarrow{U(t+1)}$ is validated by the monitor. The input $\underrightarrow{U(t+1)}$ at time t+1 is computed based on controller measurements. The monitor receives the measured values as well the outputs from the sensor. Let $\underrightarrow{y_1(t)}$ ...$\underrightarrow{y_n(t)}$ represents the estimated set of outputs corresponding to $\underrightarrow{U(t)}$ input and sensor output $\underrightarrow{X(t)}$. $\underrightarrow{U(t+1)}$ is considered safe if $\underrightarrow{U(t+1)}$ is in the convex hull of the monitor output.

The monitor analyzes the possible variations in $\underrightarrow{y_1(t)}$ ...$\underrightarrow{y_n(t)}$ with respect to the various parameters and computes an estimation of the correct output to the plant. Monitor output is preprogrammed for a desired system and hence if monitor output matches means the system is operating as per desired in safe zone. The new operational point becomes the next reference point, provided the controller output lies within the convex hull (the monitor generates a convex hull for the controller output). When the controller detects an

abnormal behavior, it decides to operate under the safe conditions until appropriate actions are taken.

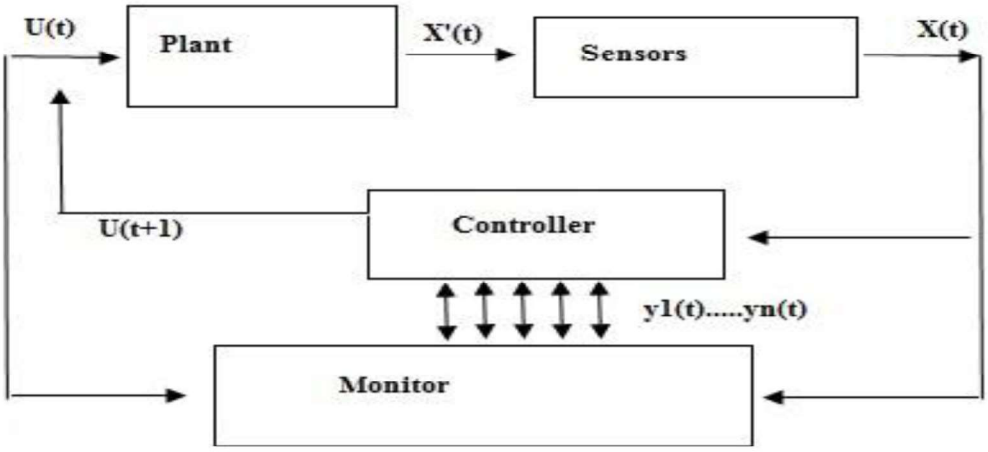


Figure 35. Anomaly detecting controller

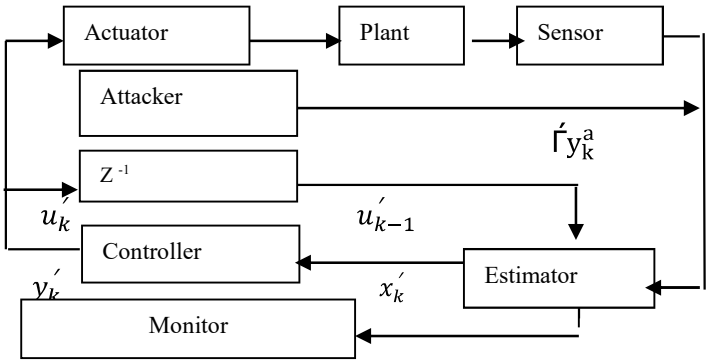


Figure 36:-False data injection attack

The proposed model shown in Figure 36 is mapped to a schematic diagram of a control plant, as shown in Figure 13, which depicts a control plant equipped with a Kalman filter (Estimator), a controller, and a monitor. The monitor observes innovation values (innovation is the output from Kalman filter that indicate the difference between the expected and actual value) value changes obtained by passing the controller output through a Kalman filter (Estimator). Based on the controller output, the next input to the

actuator is decided by the monitor. In the next subsections the designs of monitors based on the following are presented.

1. Least Square method (LSA);
2. Convex Hull;
3. Combination of LSA and Convex Hull.

5.6 Least Square Approximation (LSA)

The method is useful for continuous distance monitoring of the added new values to the governing function *'f(x)'*. An alarm is generated on constraint violations.

5.6.1 LSA Algorithm for 2-dimensional spaces (2D) point set

Input: A set S = $\{P_1, P_2, .. P_n\}$

Output: LSA line function ' *f(x)'*

1. Assume a LTI system that is defined by line function f*(x)* given by

   f(x) = y = ax+b as shown in Figure 37.

2. Compute 'a' and 'b' value.
3. Compute the maximum distances'' of the points from the line function '*f(x)'*.

If the distance 'd' is within the threshold, then the flag returns true (i.e. point is inside the convex hull) or false (i.e. point is outside the hull).

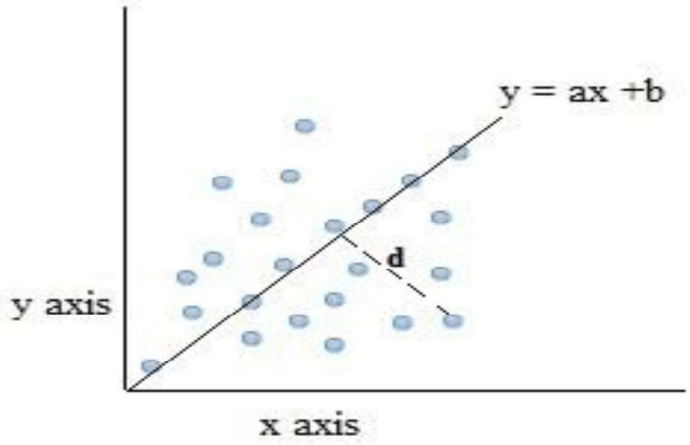


Figure37. Least square approximation approach

Figure 38 shows the LSA output for 100 random points. *'d'* represents maximum distance of the test point from the line function *f(x)*.

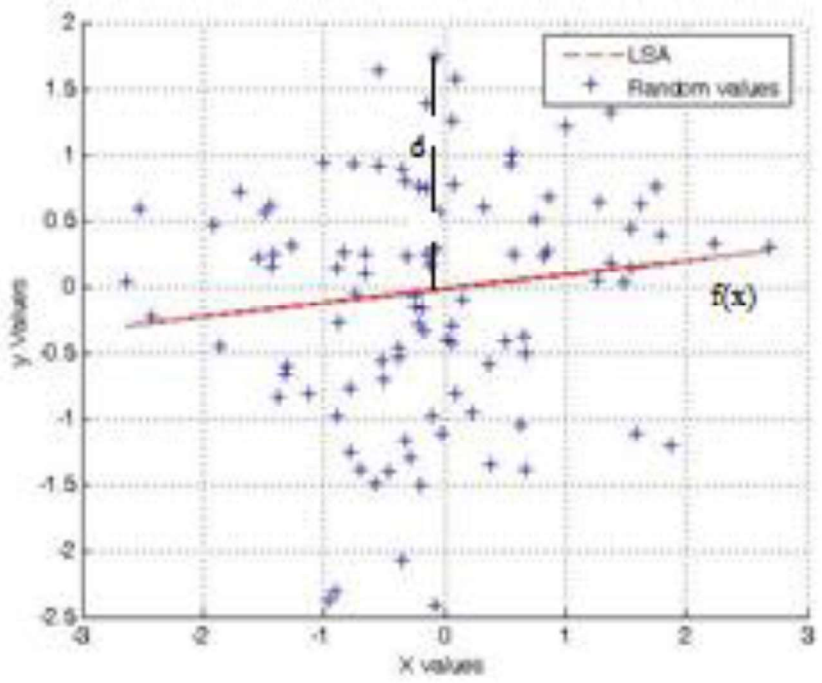


Figure 38. Test for new Single point added using LSA.

5.7 Convex Hull approach

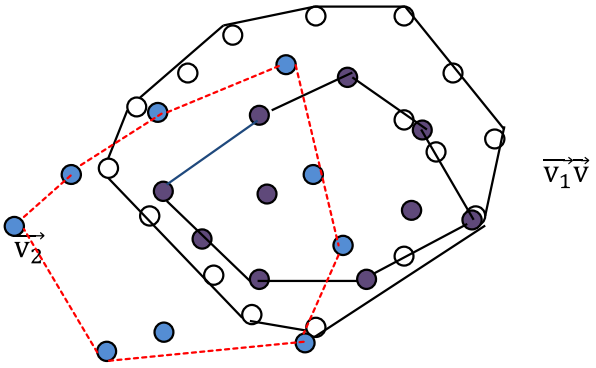
Convexity is a geometric property used in computing the smallest convex shape enclosing the point set. If the observed data are outside the convex hull and exceed the tolerance band, then it is considered to be unsafe and a possible indication of an anomaly. We have used a 2D model for analysis and demonstration.

Consider a stream of output from a controller consisting of data point set $\{u_1,u_2,\ldots u_n\}$. The convex hull ($\vec{v}$) that governs the normal behavior of the system is given by

$$\vec{v} = \text{conv}\,(u_1, u_2, \ldots u_n) \qquad (124)$$

Consider two new data streams $\{p_1, p_2, \ldots p_n\}$ and $\{q_1, q_2, \ldots q_n\}$, each represented by their convex hulls, $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ respectively. If $\overrightarrow{v_1}$ satisfies the condition mentioned in equation (125) then the monitored dataset is consider to be safe.

$$\text{conv}\,(p_1, p_2, \ldots p_n) \subseteq \overrightarrow{v_1} \qquad (125)$$



- $\vec{v}$ As convex hull of 'u' series of data
- $\overrightarrow{v_1}$ As convex hull of 'p' series of data
- $\overrightarrow{v_2}$ As convex hull of 'q' series of data

Figure 39. Convex Hull – Geometrical representation

Using 2D model as shown in figure 39, if the observed data '$\overrightarrow{v_2}$' is outside the convex hull '$\vec{v}$' and exceed the tolerance band then it can be considered as a indication of an possible anomaly.

Algorithm for the generating convex hull for a set of points is-

5.7.1 Convex Hull Algorithm for 2D point set based on Bentley-Faust-Preparata Algorithm

Input: A point set 'S' = {P_1,P_2,.. P_n}

Output: Convex hull C

Steps:

1. Sort 'S' according to increasing order of x- co-ordinate to find point with minimum and maximum x co-ordinate represented as $P_{min,y}$ and $P_{max,y}$.

2. Repeat step 1 for y co-ordinate to find point with minimum y co-ordinate ($P_{x,min}$) and maximum y co-ordinate ($P_{x,max}$) as shown in Figure 40.

3. Divide S into four sets of points such that

   A. L_1 = { $P_{min,y}$ …. $P_{x,max}$ }

   B. L_2 = { $P_{x,max}$ ….. $P_{max,y}$ }

   C. L_3 = { $P_{max,y}$ ….. $P_{x,min}$ }

   D. L_4 = { $P_{x,min}$…… $P_{min,y}$}

4. For L_1

   A. If ($\Delta x \neq 0$) calculate array of slope B[].

   B. For Max(B[]) add Point $P_{x,y}$ to C

C. Reset B

5. Repeat 4 for L_2, L_3 and L_4

6. Connect all point in C to generate convex hull [38]

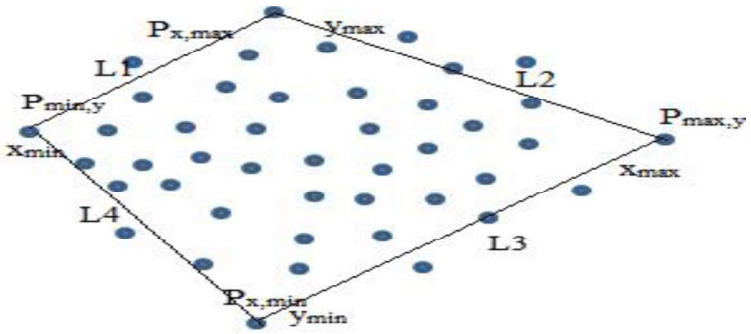


Figure 40. Sorted points bounded by L1, L2, L3 and L4.

The algorithm is tested for 100 random points using 'C' programming code as shown in Figure 41. Geometric representation of this convex hull is as shown in Figure 42.

```
Curve point are for 0 is [15.000000][12.000000]
Curve point are for 1 is [25.000000][43.000000]
Curve point are for 2 is [37.000000][61.000000]
Curve point are for 3 is [62.000000][60.000000]
Curve point are for 4 is [97.000000][58.000000]
Curve point are for 5 is [109.000000][51.000000]
Curve point are for 6 is [155.000000][-9.000000]
Curve point are for 7 is [165.000000][-28.000000]
Curve point are for 8 is [168.000000][-60.000000]
Curve point are for 9 is [164.000000][-64.000000]
Curve point are for 10 is [108.000000][-74.000000]
Curve point are for 11 is [39.000000][-30.000000]
Curve point are for 12 is [15.000000][12.000000]
```

Figure 41.Convex hull points for 100 random points using 'C' programming.

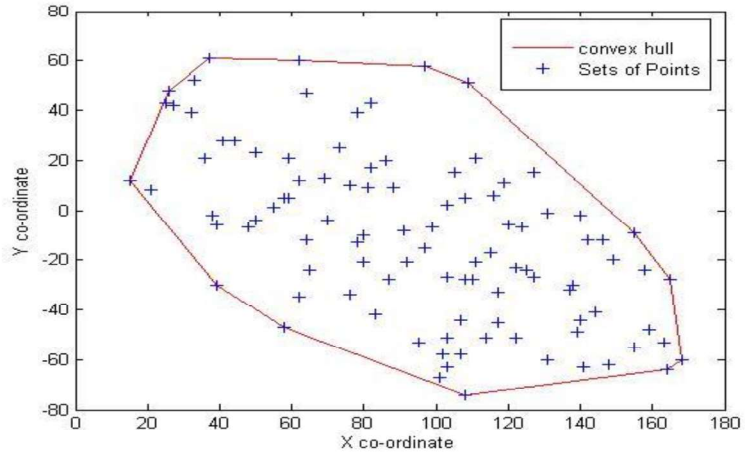


Figure 42. Convex hull generated for 100 random points

The convex hull algorithm is useful for the analysis of:

1. New single value added in the normal data set.
2. Set of values added in the data set.
3. New convex hull created by a set of added values intersecting a convex hull of the trained system.

5.7.2 New Single Point or Point-Set added in existing Convex Hull

1 Select a single point (in case of point set, average value is assumed) from a point array.

2 FOR EACH count ← 0 to size of (point array)

a. Test the point with each convex hull point set for slope and boundary conditions.

b. Return a flag inside or outside or on the boundary if

2.b.1 Slope is positive, Flag = inside.

2.b.2 Slope is negative, Flag = outside.

2.b.3 Point lies on the convex hull, Flag = boundary.

The algorithm output for 100 random points is as shown in Figure 43.

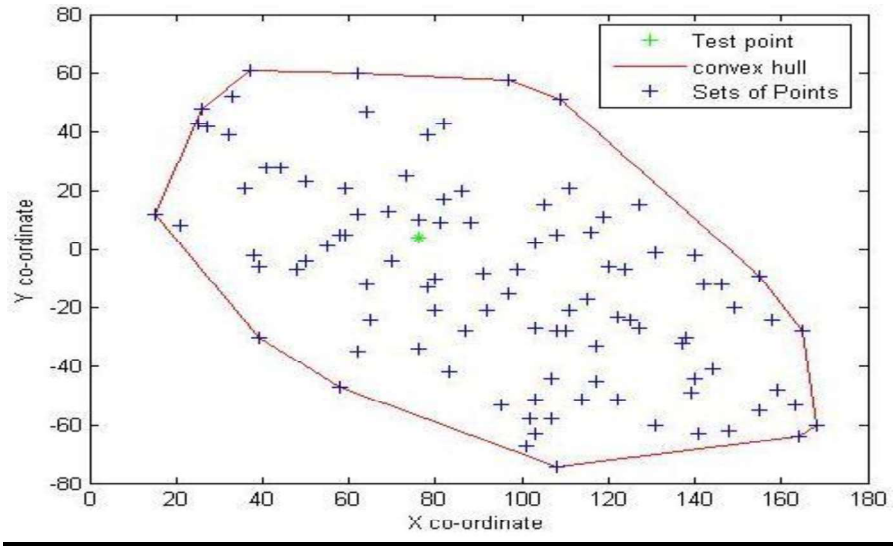


Figure 43. Test for new added random point in the existing Convex curve.

5.7.3 Algorithm for Intersection of Two Convex Hulls [40]

For a given two convex polygons '*P'={p(1), ... , p(m)}* and '*Q'={q(1), ... , q(n)}*, as shown in Figure 44, the algorithm (by rotating caliper method) for their intersection is given as:

1. Compute the vertices with the maximum '*y'* coordinate for both '*P'* and '*Q'*. If more than one vertex exists, take the vertex with greater *x* coordinates.
2. Construct horizontal lines through these points such that the polygons lie to their right.
3. Rotate both lines of support clockwise until one coincides with an edge. A new co-podal (co-podal mean parallel lines in same direction) pair ($p(i)$, $q(j)$) is determined. In the case of parallel edges, three co-podal pairs are determined.
4. For all valid co-podal pairs ($p(i)$, $q(j)$), check if $p(i\text{-}1)$, $p(i\text{+}1)$, $q(j\text{-}1)$, $q(j\text{+}1)$, all lie on the same side of the line joining ($p(i)$, $q(j)$). The co-podal pair is a bridge.

5. Repeat steps 3 and 4, until the lines of support reach their original position, as shown in Figure 44.
6. Construct the merged convex hull by joining the proper convex chains between consecutive bridges.

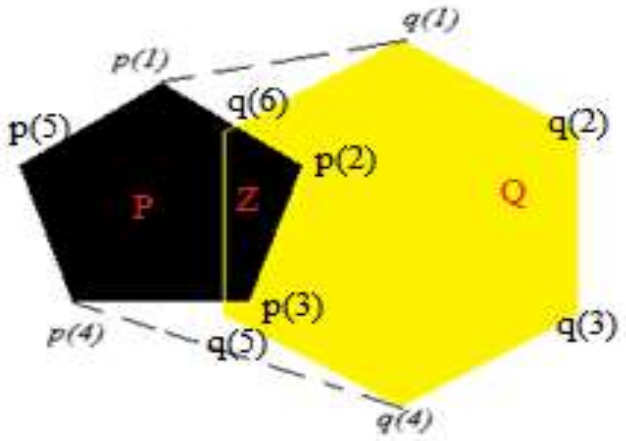


Figure 44. Intersection of two convex hulls.

Intersection of two convex hulls (C1 and C2) for random points can be simulated as shown in Figure 45.

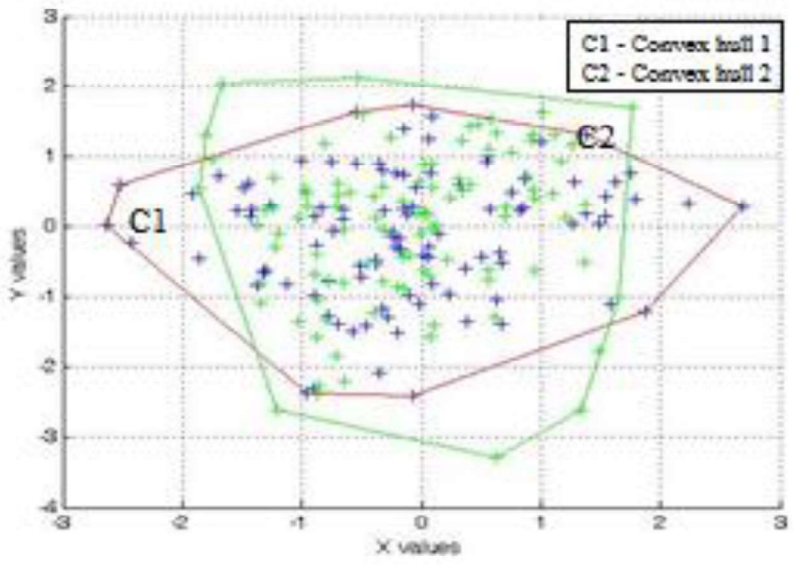


Figure 45. Intersection of two convex hulls C1 and C2.

Each convex hull is a represented using set of function that governs it. Convex hull is mathematically interpreted using 'R' (real) functions as explained in the following section.

5.7.4 Mathematical interpretation of convex hulls intersection using 'R' function

An 'R' function is a real valued function that is determined with sign and magnitude. 'R' functions are useful for describing geometric objects using a single or set of equations. Let's define a function S(x) along the real axis (X – axis) by Eq. (125)

$$S(x) = \begin{cases} 0 \; if \; x \leq 0 \\ 1 \; if \; x \geq 0 \end{cases} \quad (125)$$

'x' represents value on the X-axis.

R-function y = f(x) can be defined, if there exist a Boolean function Y = F(x) such that

S(f(x)) = F(S(x))

Consider two convex hulls 'P' and 'Q' as shown in figure 53. Each convex hull can be represented as 'R' – function. Let f(P) and f(Q) represent function governing the respective polygon 'P' and 'Q'.

1. As per set theory union between two or more polygons represent maximisation and intersection represent minimization respectively where polygon represents convex hulls.
2. Each polygon has Boolean properties according to R- function definition.
    a. Min (x1, x2) is a minimization R-function whose companion Boolean function is logical "and" i.e. (U), and

b. Max (xl, x2) is an maximization R-function whose companion Boolean function is logical “or” i.e. (V)

3. Intersection and union of convex hull can be represented as Eq. (126) and (127) respectively.

$$P \cap Q = \min(f(P), f(Q)) \quad (126)$$

$$P \cup Q = \max(f(P), f(Q)) \quad (127)$$

4. Above Equations simplifies using Boolean ‘AND’ operator as ‘+’ and ‘OR’ as ‘*’

$$P + Q = \min(f(P), f(Q)) + \max(f(P), f(Q)) \quad (128)$$

$$P * Q = \min(f(P), f(Q)) * \max(f(P), f(Q)) \quad (129)$$

5. Consider an quadratic equation for variable ‘z’ whose roots are min (f(P),f(Q)) and max (f(P),f(Q)) can be given by Eq.(130).

$$z^2 + (P + Q) * z + P * Q = 0 \quad (130)$$

6. Roots of this equation can be given by Eq.(131) and (132).

$$\text{Min}(f(P), f(Q)) = 1/2\,[f(P) + f(Q) - \sqrt{(f(P) + f(Q))^2}\,] \quad (131)$$

$$\text{Max}(f(P), f(Q)) = 1/2\,[f(P) + f(Q) + \sqrt{(f(P) + f(Q))^2}\,] \quad (132)$$

Hence Intersection of convex hull can be represented as Eq.(133)

$$P \cap Q = \min(f(P), f(Q)) = 1/2[f(P) + f(Q) - \sqrt{(f(P) + f(Q))^2}\,] \quad (133)$$

Union of convex hull is represented as Eq. (134)

$$P \cup Q = \max(f(P), f(Q)) = 1/2\,[f(P) + f(Q) + \sqrt{(f(P) + f(Q))^2}\,] \quad (134)$$

LSA method is used to validate convex hull intersection as explained in the following algorithm.

5.7.5 Algorithm using LSA and convex hulls approach

Input:-

Convex hulls '*P'* (representing a trained system output)

Convex hulls *'Q'* (representing new controller output).

Convex hull 'Z' (intersection of polygon 'P' and 'Q')

Output:-

Flag (True indicate safe, false indicate unsafe)

1. Generate Boundary point sets for convex hulls, P and *Q*.

2. Let center co-ordinate of 'P', 'Q' and 'Z' be $\{p_x p_y\}$, $\{q_x q_y\} and \{z_x z_y\}$ respectively.

3. Using LSA for each point $\{p_{ix},p_{iy}\}$ in boundary set 'P', calculate distance from center point by Eq. (135)-

$$d_{px} = \sqrt{((p_{ix} - p_x)^2 + (p_{iy} - p_y)^2} \quad (135)$$

d_{px} represent distance of boundary points.

$\{p_{ix}, p_{iy}\}$from center $\{p_x, p_y\}$.

4. Maximum distance value is computed for set 'P' by Eq. (136).

$$d_{pmax} = max\ (d_{px}) \quad (136)$$

5. Repeat step (4) and (5) for boundary point set for '*Q' and 'Z'*. Maximum distance computed is set as d_{qmax} and d_{zmax} respectively for 'Q' and 'Z'.

6. For all points $\{p_{ix},p_{iy}\} \in$ 'Q' $\notin$ 'Z' compute-

a) Distance of each point in 'Q' $\{q_{ix}, q_{iy}\}$ from center point $\{z_x z_y\}$ given by d_{qzx} satisfies condition

$$d_{qzx} \leq d_{zmax} \quad (137)$$

Return Flag = true.

b) If condition 6(a) fails check the distance of the point from *f(q)* given by d_{pqx}.

$$d_{pqx} \leq d_{pmax} \quad (138)$$

Return Flag = true

c) If 6(a) and 6(b) fails then the point can be considered unacceptable and returns Flag = false. This can be considered as an indication for a possible vulnerability.

5.8 Experiment and Simulation

5.8.1 Attack Scenario

Such control-oriented attacks (targeted attacks as explained in Chapter 4) would require extensive knowledge about the process behavior, dynamics, and control algorithms. The attack surface may be through the computer-based control system, if the attacker gains control to the computer system where the configuration parameters are stored. An example may be the reactor regulating system, which stores parameters for SPND coefficients, correction factors for thermal power, and look up tables.

5.8.2 Simulation

We have considered a four-tank model to illustrate a control theoretical cyber-attack and to design a monitor for detecting changes due to a targeted attack, as explained in the following section.

The four-tank level control system is a typical control system with nonlinear, coupling, and time delays characteristics, and can be used in the simulation of a multivariate industrial system. It can be used as a test bed to test the effects of the

applications of various control theories. The model is used for verification of computational algorithms, LSA, and the convex hull method.

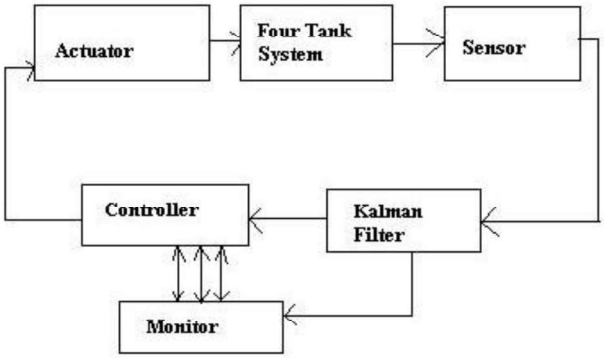


Figure 46. Schematic representation of four tank model in control plant

As shown in Figure 46, the four-tank system can be considered a controlled system and its output as an input to the Kalman Filter (estimator). Estimator output acts as an input to the monitoring system designed using LSA or the convex hull algorithm or both. We assume two dimensional spaces (2D) where two parameters, the 'difference in each tank height (cms)' and 'calculated height of each tank (cms)', are respectively considered in the 'x' direction and the 'y' direction.

In a non-attack (normal) mode, the monitor is trained with the normal data. Once trained in LSA and Convex hull approach, a monitor need to continuously validate the data for various attack scenarios like bias, geometric etc. For a given spectrum of input and output convex hulls are generated. Output is simulated after 500^{th} interval when the system reaches in its stable output stage (Kalman filter achieves convergence).

In our simulation monitoring parameters assumed are -

1. Estimated height of the individual tank in cms (as input) and

2. Difference in each tank height in cms (as output).

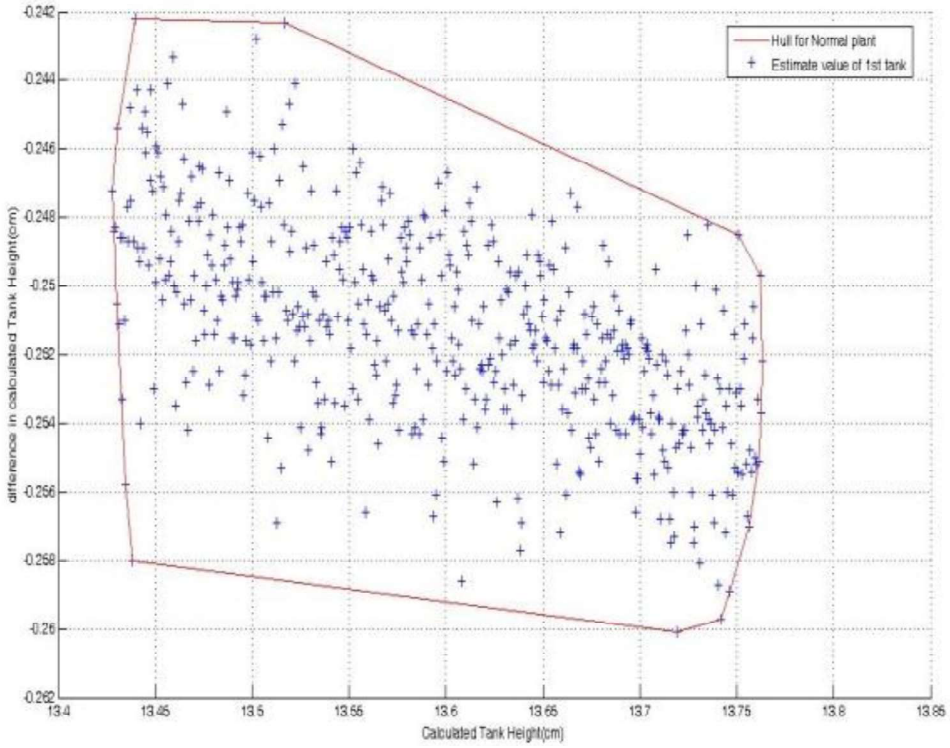


Figure 47. Convex hull for Normal plant.

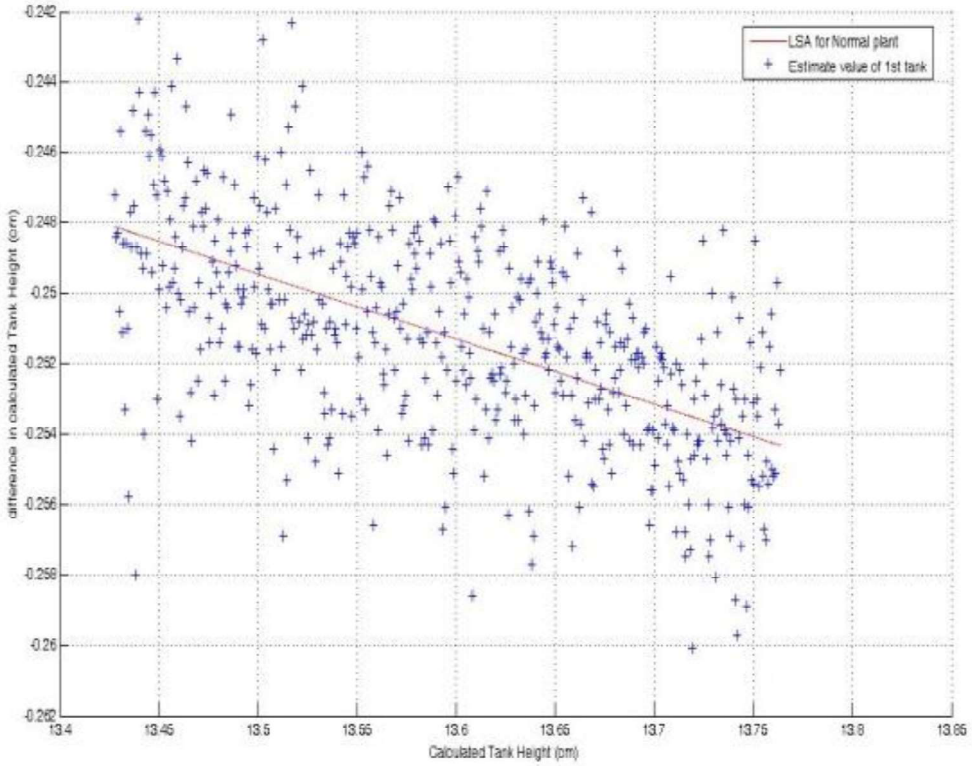


Figure 48. LSA for Normal plant.

For Normal trained plant the output is as shown in Figure 47 and 48 for Convex hull and LSA approach respectively. Four tank model is subjected to input data attack (explained in chapter 3). In this attack input signal *u(k)* is varied. The output generates convex hulls and LSA output as shown in Figure 49 and Figure 50 respectively.

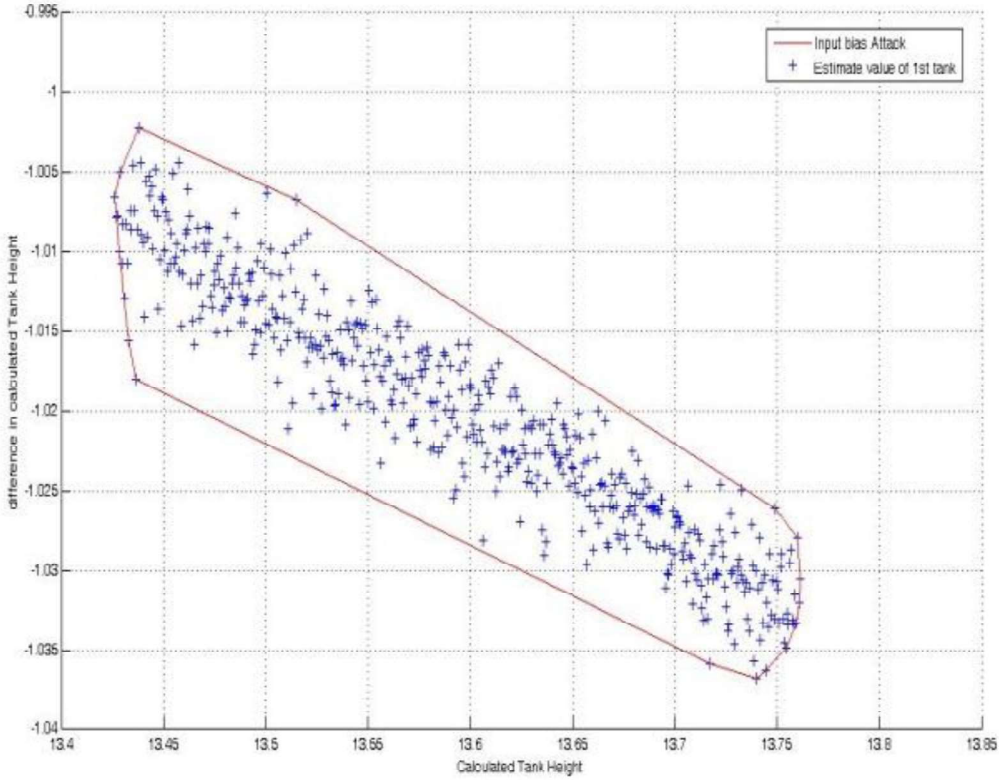


Figure 49. Convex hull for input bias attack.

Four tank model is subjected to bias attack (a type of targeted attack explained in chapter 3). State matrix is targeted in bias and various other attacks. The output generates convex hulls and LSA output as shown in Figure 51 and Figure 52 respectively. Figure 53 and 54 indicates output curves for Random bias attack (a type of targeted attack). Random input parameters are changed for this attack.

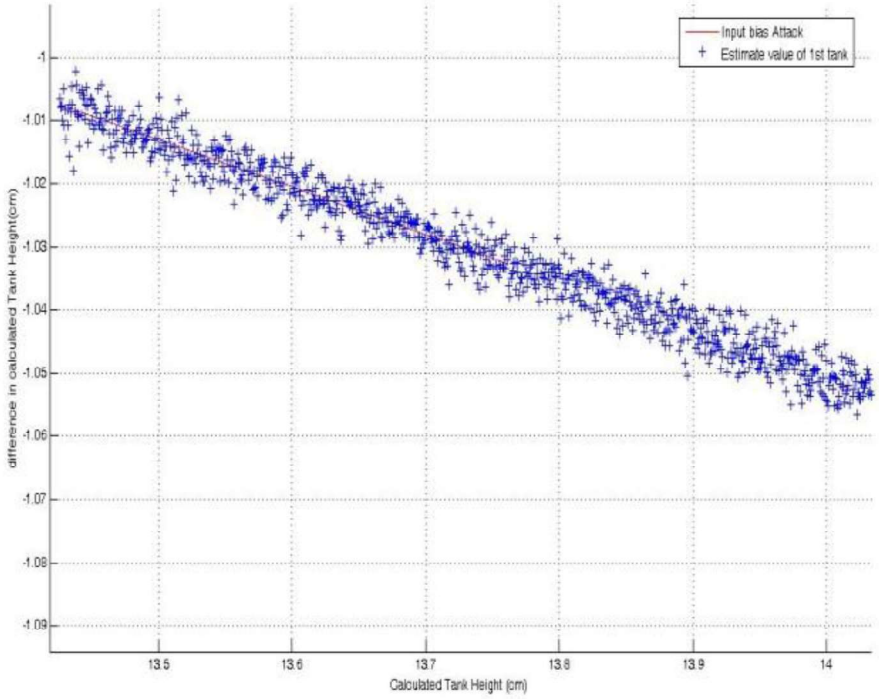


Figure 50. LSA for input bias attack.

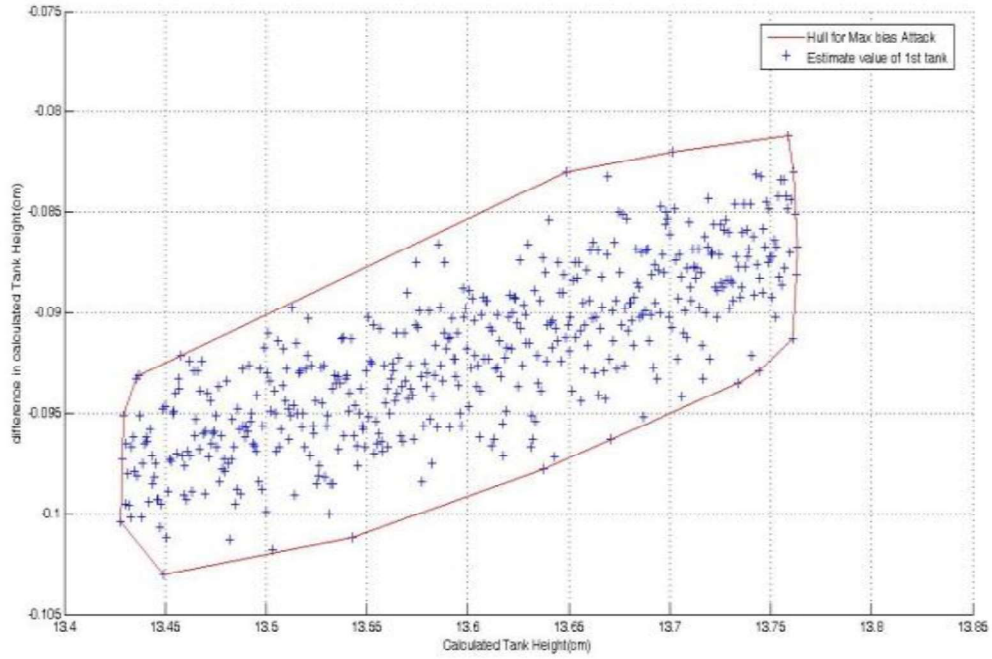


Figure 51. Convex hull for Maximum bias attack.

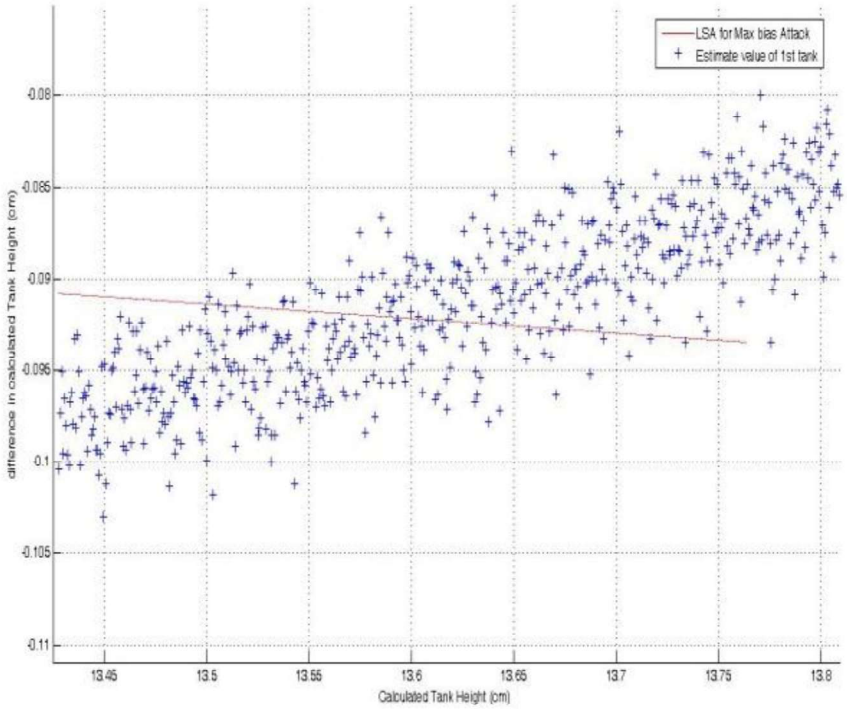


Figure 52. LSA for Maximum bias attacks.

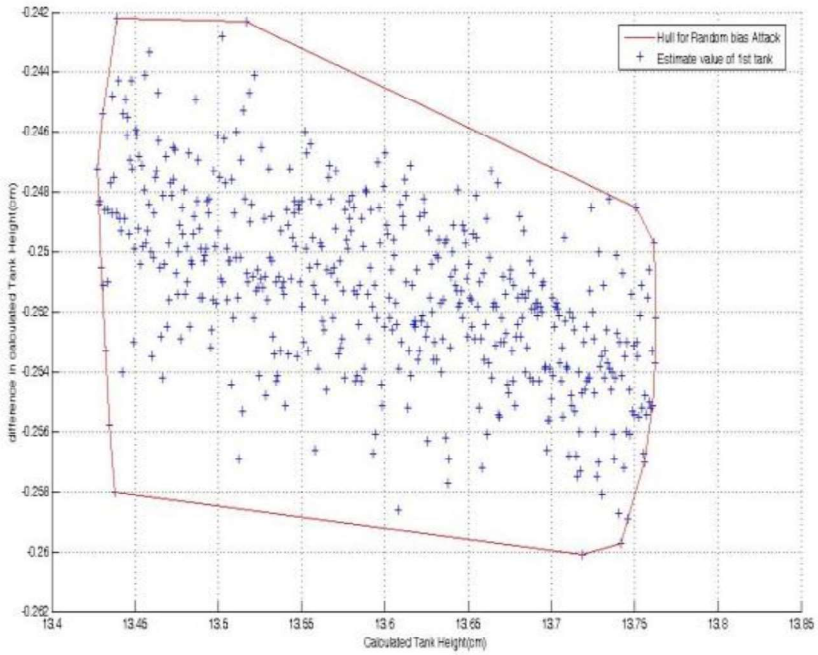


Figure 53. Convex hull for Random bias attack.

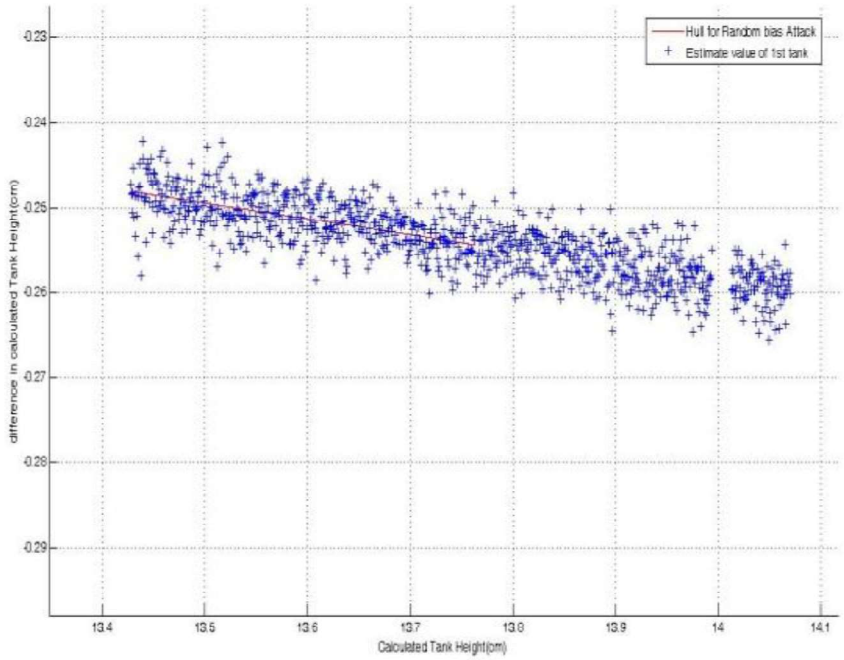


Figure 54. LSA for Random bias attack.

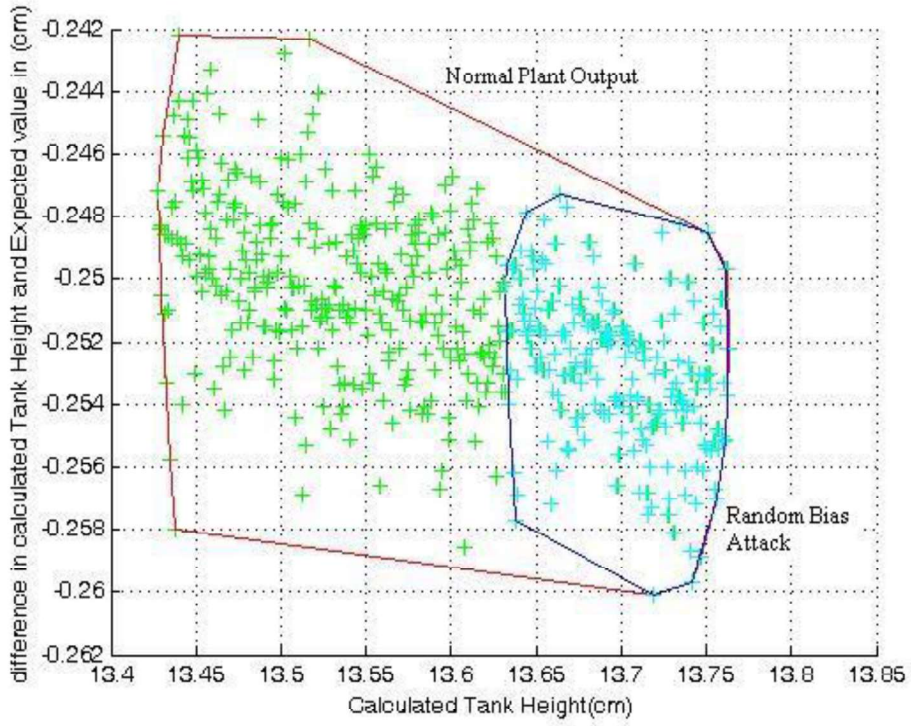


Figure 55. Intersection of convex hull for normal and for Random bias attack.

1. Figure 47 to 56 gives indication of convexity changes observed in monitoring parameters in various scenarios. Blue points in the graph symbolize various heights of Tank 1 and its measurement error at various time intervals (from 501th to 1000th interval).
2. Figure 55 indicates intersection of two convex hulls. One convex hull represents a normal trained output and second hull represents random bias attack. As both convex hulls almost overlap (random bias attack output is a subset of Normal plant output), it is difficult to detect the anomaly.

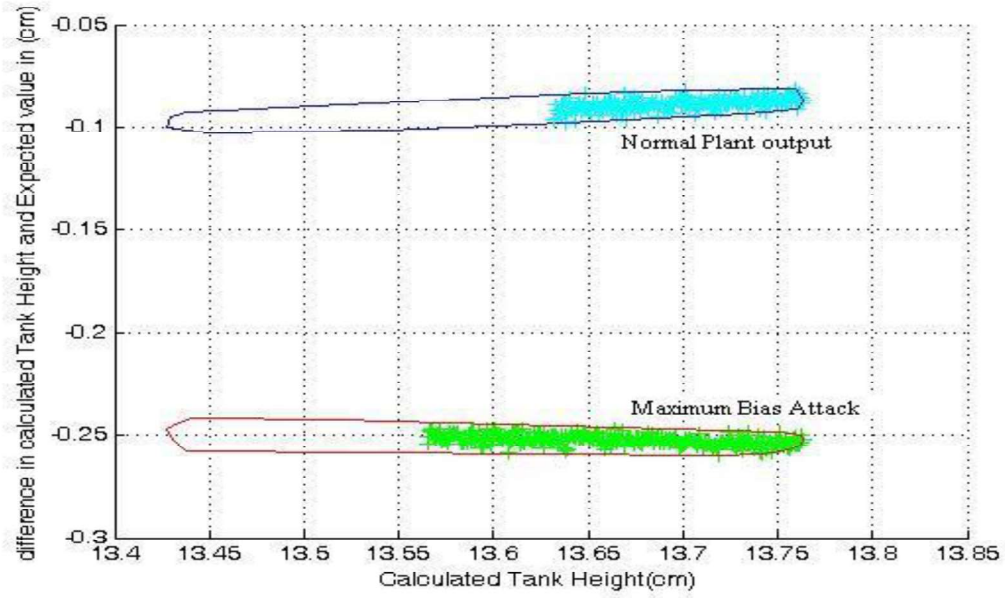


Figure 56. Non - Intersection of convex hull for normal, input bias and for minimum bias attack.

3. A completely new convex hull is generated in bias attack than in the normal plant as shown in Figure 56. For understanding Figure 56, consider an normal plant (example four tank model), the convex hull for normal operation indicate that the two output (height of tank 1 and tank 2) are having minimum difference in the

expected and actual values (0.1 cms) and the output values are more concentrated between 13.62 to 13.76cms. When the plant (four tank model) is subjected to maximum bias attack the difference between expected and actual value is more (0.25cms) and output values spread from 13.56 to 13.76 cms. Both the convex hulls (normal and Bias attack) does not overlap nor intersect. Hence this method is useful for attack detection.

4. Distance is calculated using LSA for various scenarios (Figure 58, 60, 62 and 64)-

Scenarios	Distance calculation using LSA(cms)
Normal plant	0.008175
Input bias attack	0.013723
Maximum bias attacks	0.0081765
Random bias attack	0.0073

Table 2:- LSA distance calculation

5.9 Conclusion

Online monitoring will play an increasingly important role in observing the control system behavior and to detect anomalous behavior. Application of such algorithm is demonstrated using a simulation model of the classical four tank model. The effectiveness of the algorithm along with LSA for anomaly detection is demonstrated. Computational geometric approaches such as LSA and convex hull methods have

advantages in post facto analysis of log data obtained from a SCADA server and to detect possible anomalies. Control system engineers are aware of various statistical techniques for fault detection in control systems; however, the application of such techniques in securing control systems against cyber-attacks has not been examined in detail. We tried to demonstrate such techniques in monitoring such anomalous behavior.

It is important to understand the difficulties of how such monitors can be synthesized in lower level controllers that would be able to apply such algorithms for the detection of anomalous behavior. Today's microcontrollers and FPGAs have good computing abilities to run such complex algorithms. Nuclear Power Plants have servers logging configurations for storing historic data and outputs of lower level controllers. An alternative to the online monitoring scheme would be to use such algorithms as offline monitors.

Chapter 6

Conclusion and Future Work

The research work demonstrated that targeted cyber-attacks in control systems are possible to be detected using statistical techniques. The thesis has provided a survey on the attack models and we have identified research challenges for securing control systems. We have implemented statistical techniques to detect such typical cyber-attacks on software implemented controllers and demonstrated the technique using a simulation on the classical four tank model.

We have discussed various statistical techniques like SPRT, CUSUM, and GLR etc., for detecting cyber-attacks and algorithms for securing control systems. We have also showed the efficacy of the techniques by incorporating changes as stated in attack model and demonstrated that these can be detected by innovation value change.

Online monitoring plays a vital role in observing the control system behavior and to detect anomalous behavior. Computational algorithm like LSA and geometric methods were designed and demonstrated using a simulated four tank model. The effectiveness of

the convex hull approach along with LSA for anomaly detection is tested. Computational geometric approaches such as LSA and convex hull methods have advantages in analyzing complex log data obtained from a SCADA server and to detect anomalies.

It is important to understand the difficulties of how such monitors can be synthesized in lower level controllers that would be able to apply such algorithms for the detection of anomalous behavior. Today's microcontrollers and FPGAs have good computing abilities to run such complex algorithms. An alternative to the online monitoring scheme would be to use such algorithms as offline monitors. It would be interesting to see how the algorithms could be realized efficiently.

References

1. Basseville, M. (1981) 'Edge detection using sequential methods for change in level – part II: sequential detection of change in mean', *EEE Trans. on Acoustics, Speech and Signal Processing*, February, Vol. ASSP-29, No. 1, pp.32–50.
2. Basseville, M. (1988) 'Detecting changes in signals and systems – a survey', *Automatica*, Vol. 24,No. 3, pp.309–326.
3. Basseville, M. and Nikiforov, I.V. (1993) *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Inc., Englewood Cliffs, NJ, April, ISBN 0-13-126780-9.
4. Byres, E., Ginter, A. and Langil, J. (2011) *How Stuxnet Spreads – A Study of Infection Paths in Best Practice Systems*, White Paper Abterra Technologies, pp.1–26.
5. Cardenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A. and Sastry, S. (2009) 'Challenges for securing cyber physical systems', *Workshop on Future Directions in Cyber-physical Systems Security*, DHS.
6. Cárdenas, A.A., Amin, S. and Sastry, S. (2008) 'Secure control: towards survivable cyber physical systems', *ICDCS '08*, pp.495–500.
7. Cárdenas, A.A., Amin, S., Liny, Z-S., Huangy, Y-L., Huangy, C-Y. and Sastry, S. (2011) 'Attacks against process control systems: risk assessment, detection, and response', *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ACM, USA,pp.355–366.
8. Case, M.J. (2007) 'US nuclear regulatory commission NRC information notice', in *Effects of Ethernet-Based, Non-Safety Related Controls on the Safe and Continued*

Operation of Nuclear Power Stations, IN 2007-15, pp.1–3.

9. Caselli, M., Zambon, E. and Kargl, F. (2015) 'Sequence-aware intrusion detection in industrial control systems', *CPSS '15 Proceedings of the 1st ACM Workshop on* Cyber-Physical System Security, pp.13–24, ISBN: 978-1-4503-3448-8.
10. Cervin Anton, (2003) , 'Integrated Control and Real-Time Scheduling', PhD thesis, Lund University.
11. Dan Henriksson, Anton Cervin, Karl-Erik Årzén, (2003),'TrueTime: Simulation of Control Loops Under Shared Computer Resources', *15th IFAC World Congress on Automatic Control, Barcelona*, Spain.
12. Dan Henriksson, Anton Cervin, Karl-Erik Årzén, (2003),'TrueTime: A Matlab/Simulink-based simulator for networked embedded control systems'. *Sweden*.
13. Eric Byres, Carter, P.E.J., Elramly, A. and Hoffman, D. (2002) 'Worlds in collision-ethernet and the factory floor', *Proceeding ISA Emerging Technologies Conference, Instrumentation Systems and Automation Society*, Chicago.
14. Espinosa, M.M. (2010) *Contributions to the Control of Networked Cyber-Physical Systems*, PhD thesis, University of California, USA.
15. Frederick M. Proctor and William P. Shackleford, (2001), 'Real-time Operating System Timing Jitter and its Impact on Motor Control', *Proceeding of the SPIE, Volume 4563*, pg: 10-16.
16. Genge, B., Graur, F. and Haller, P. (2015a) 'Experimental assessment of network designapproaches for protecting industrial control systems', *International Journal of Critical Infrastructure Protection*, December, Vol. 11, No. C, pp.24–38.

17. Genge, B., Haller, P. and Kiss, I. (2015b) ‘Cyber-security-aware network design of industrial control systems’, *IEEE Systems Journal*, Vol. PP, No. 99, pp.1–12, DOI:10.1109/JSYST.2015.2462715.
18. Gatzke Edward P., Edward S. Meadows, Chung Wang, Francis J. Doyle III,(2000),’Model Based Control of a Four-Tank System’, *Computers & Chemical Engineering, Elsevier*. Volume 24, Issues 2–7, 15 July 2000, 1503–1509.
19. Goodloe, A. and Pike, L. (2010) ‘Monitoring distributed real-time systems: a survey and future directions’, *NASA*, pp.1–50.
20. Greg Welch and Gary Bishop (2006) An Introduction to the Kalman Filter , UNC-Chapel Hill, TR 95-041, July 24, 2006
21. Hu, Fei. (2013), ‘Cyber-Physical Systems: Integrated Computing and Engineering Design’. *CRC Press, Taylor & Francis Group, N.Y, U.S.A*
22. Huang, Y-L., Cárdenas, A.A., Amin, S., Lin, Z-S., Tsai, H-Y. and Sastry, S. (2009) ‘Understanding the physical and economic consequences of attacks on control systems, *International Journal of Critical Infrastructure Protection*, Vol. 2, No. 3, pp.73–83.
23. Huertas, D.P. (2011) *Cyber-Security and Safety Analysis of Interconnected Water Tank Control Systems*, Stockholm, Sweden.
24. Jane W. S. Liu (2000), ‘Real-Time Systems’, Pearson Education.
25. Kiss, I., Genge, B. and Haller, P. (2015) ‘A clustering-based approach to detect cyber attacks in process control systems’, *Industrial Informatics (INDIN)*, pp.142–148, DOI:10.1109/INDIN.2015.7281725.

26. Kiss, I., Genge, B., Haller, P. and Sebestyen, G. (2014) ‘Data clustering-based anomaly detection in industrial control systems’, *Intelligent Computer Communication and Processing (ICCP)*,pp.275–281.
27. Lee, D. and Deepa, K. (2014) ‘Cyber attack detection in PMU measurements via the expectation maximization algorithm’, *Signal and Information Processing (GlobalSIP)*, IEEE, pp.223–227,DOI:10.1109/GlobalSIP.2014.7032111.
28. Liu, Y., Ning, P. and Reiter, M. (2011) ‘Generalized false data injection attacks against state estimation in electric power grids’, *ACM Transactions on Information and System Security*,Vol. 14, No. 1, pp.21–32.
29. Maybeck, P. (1979) *Stochastic Models, Estimation and Control*, Vol. 1, Chapter 3, pp.59–132.
30. Mo, Y. and Sinopoli, B. (2009) ‘Secure control against replay attacks’, *47th Annual Allerton Conference on Communication, Control, and Computing*, Illinois, USA.
31. Ogata, K. (2010) *Modern Control Engineering*, 5th ed., Prentice Hall, New Jersey, USA.
32. Pasqualetti Fabio, (2012),*'Secure Control Systems: A Control-Theoretic Approach to Cyber-Physical Security'*, PhD thesis, University Of California.
33. Pau Martí and Josep M. Fuertes, Gerhard Fohler, Krithi Ramamritham (2001), ‘*Jitter Compensation for Real-Time Control Systems*’, (RTSS 2001), London, ISBN:0-7695-1420-0.
34. Preparata F. P., S. J. Hong (1977), ‘*Convex hulls of finite sets of points in two and three dimensions*’, February 1977, ACM , Volume 20 Issue 2

35. Rrushi, J.L. (2009) *Composite Intrusion Detection in Process Control Networks*, PhD thesis, University of Milano, Milano, Italy.
36. Shapiro Vadim (1994), '*Real functions for representation of rigid solids, Computer Aided Geometric Design'*, pages 153-175
37. Sharfman Izchak, Assaf Schuster and Daniel Keren, (2013) '*A Geometric Approach to Monitoring Threshold functions over distributed data streams'*, SIGMOD '13, ACM ,USA.
38. Shyamasundar, R.K. (2013) 'Security and protection of SCADA: a big data algorithmic approach', *Proceedings of the 6th International Conference on Security of Information and Networks (SIN '13)*, ACM, New York, NY, USA, pp.20–27.
39. Slay, J. and Miller, M. (2007) 'Lessons learned from the Maroochy water breach', in *Critical Infrastructure Protection*, Volume 253 of the series IFIP International Federation for Information Processing, pp.73–82, Springer, USA.
40. Toussaint Godfried T. (1985), '*A simple linear algorithm for intersecting convex polygons'*, The Visual Computer, August 1985, Volume 1, Issue 2, pages 118-123
41. Von Solms, R. and van Niekerk, J. (2013) 'From information security to cyber security', *Computers and Security*, Vol. 38, pp.97–102, Elsevier Advanced Technology.
42. Willsky, A. (1976) 'A survey of design methods for failure detection in dynamic systems', *Automatica*, Vol. 12, No. 6, pp.601–611.
43. Yadegari, B. and Mueller, P. (2012) *The Stuxnet Worm*, Arizona University Teaching Notes, CSc566: Computer Security, pp.1–43.

44. Zhang, Q., Basseville, M. and Benveniste, A. (1994) ‘Early warning of slight changes in systems and plants with application to condition based maintenance’, *Automatica*, Special Issue on *Statistical Methods in Signal Processing and Control*, Vol. 30, No. 1, pp.95–114.

45. Zhu, Y. and Shasha, D. (2002) ‘StatStream: statistical monitoring of thousands of data streams in real time’, *Proceeding VLDB 2002 Proceedings of the 28th international conference on Very Large Data Bases*, Hong Kong, China, pp.358–369.