Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Confirmation of Theoretical Results Regarding Control Theoretic Cyber Attacks on Controllers

Hemangi Gawand *, A.K Bhattacharjee **, Kallol Roy **

** Homi Bhabha National Institute, BARC, India,*
(hemangi.gawand@gmail.com),
*** Reactor Control Division, BARC, India,*
(anup@barc.gov.in),
***Research Reactor Maintenance Division, BARC, India,*
(kallolr @barc.gov.in)

Abstract: National critical infrastructures like power plants, grids, water distribution system etc employ a hierarchy of controllers exchanging data over a network. They employ sophisticated control algorithm implemented in software. Various researches have examined the attack scenarios in such embedded control systems from control theoretic perspectives. In this paper we revisit these theoretical attacks and postulate that such attacks could be detected by statistical techniques and hence may be used to design security monitors. The postulate is confirmed by simulation results

Keywords: False data injection attack, Cyber Physical System (CPS), Kalman filter, Sequential probability ratio test (SPRT)

1 INTRODUCTION

Cyber physical systems are hierarchy of computer elements running a discretized control algorithm electrically connected to physical sensors and control elements. The notion of stored program and data with communication over networks make such system susceptible to various security threats. In this paper, we are concerned with attacks originating from an attacker with deep knowledge of the control algorithm and access to the CPS network. Such attacks have been described by various researchers [2, 3, 4, and 6]. In this paper, we provide simulation results of such attack scenarios. We argue that such attacks are possible to be detected by statistical techniques as commonly used in diagnostic framework.

The paper is organized as follows: section 2 includes a short discussion on related work followed by an explanation of the cyber physical systems. In section 4 and 5, a presentation of the various attack models and their theoretical basis are explained. Section 6 includes discussion on the simulation results, followed by a brief discussion on conclusions in section 7.

2 RELATED WORK

A significant amount of research effort has been carried out to analyze, detect and handle failures in control systems. Byres et. al. [2] provides an insight and modalities of Stuxnet attack, its mean and method of spread to desired control device. Alvaro et. al in [3] demonstrates the threat to control system by Stuxnet that reprograms controller to behave out of specified boundaries. It also analyses various control system attack by using Tennessee Eastman plant example. Alvaro et.al. [4,5] have also shown that by incorporating knowledge of physical system under control, it is possible to detect the change in behavior of the targeted control system. They have classified those attacks as targeted and non-targeted attacks. Yu-Lun Huang et.al.[6] describe an approach for developing threat models for attack on control system that he called as false data injection attack. Yao Liu and Sinopoli in [4] have studied the estimation scheme in power grid. A generalized likelihood ratio test to detect dynamics or sensor jump is proposed by Willsky et. al.[10]. Jones et.al. [11] has discussed about sensor and controller failure model.

The contribution of this paper is in studying the theoretical results with simulation results. One of the focus is in the application of statistical techniques to detect such typical cyber-attacks on software implemented controllers. We also propose to use such techniques to design monitoring algorithms [14].

3 CONTROL LOOPS IN CYBER PHYSICAL FRAMEWORK

Control systems are computer based systems that monitor and control physical processes. They are made up of sensors, computational and

Copyright © 2013 IFAC

communication capabilities. Figure 1 shows a simplified CPS network.

Data received by actuator causes necessary action on physical system. Sensors measures physical system states and transmits to distributed controllers. Controller in turn performs a control action (by hardware e.g. PID or by computing an algorithm) whose output is transmitted to the physical actuator. A control action is a reactive process and failure of any non- redundant sensor or actuator can break the reactive action which may cause irreparable damage to the system under control.

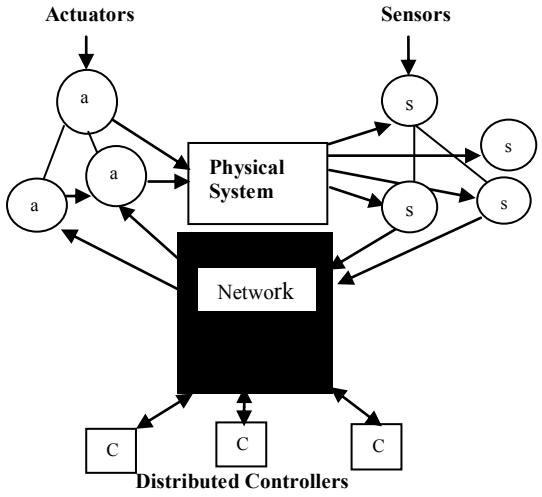

Figure 1:- The general architecture of cyber physical systems [3]

4 VARIOUS PROPOSED CYBER PHYSICAL ATTACKS

An attack on a CPS can cause physical damage to system under control by manipulating the controller characteristic parameters. It can lead to destruction and loss of human life by targeting critical infrastructure. Stuxnet attack was one of such attack that was targeted on uranium enrichment program of Iran. It was developed with complete awareness of process and sabotaged critical systems. "Data Storm" attack [8] is also well-known CPS attack.

In general, such attacks on CPS can be broadly classified as:-

1. **Non Targeted Attacks**: - In this attack, attacker is unaware of the damage that is going to be caused by his act.
2. **Targeted Attacks**: - In this attack the attacker is aware of the targeted control system and the strategy is well planned. Stuxnet [1], Maroochy Shire incident [7] are few examples of targeted attacks. Targeted Attack can be further classified based on input, output or state of the system altered.
    1. Input Data Attack
    2. Output Data Attack
    3. State Attack.

A detailed discussion on the scenarios to cause input, output or state attacks by alteration of output sensors or by alteration of state matrices (A, B or C) is given in section 5. False Data injection attack is one such attack that is explained using this attack models.

4.1 False Data injection Attack

It is an output attack method. In this attack, the attacker aims to create a new attack vector x'_k that can result in wrong estimation of state variable/s that can remain undetected as shown in figure 2. For false data injection attack analysis it is assumed that system is equipped with a Kalman filter, controller and a detector for monitoring innovation value change as shown in figure 2. There are sensors that provide reading to state estimator 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 selected is random
2. **Targeted false data injection attack**: - Attack vector injects specific error into certain state variables [11].

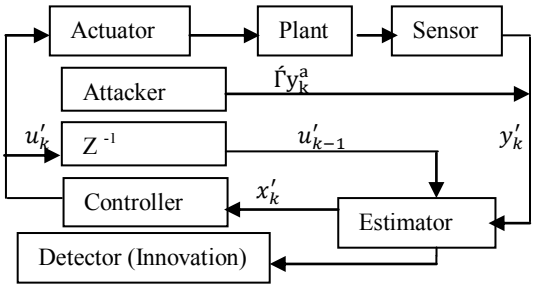

Figure 2:- Schematic diagram of compromise Sensor in control Plant [6]

5 REPRESENTATIVE ATTACK MODEL

All basic Kalman filter equations from (1) to (12) are assumed to hold good for attack model design in below section [9].

$$x_{k+1} = Ax_k + B u_k + w_k \quad (1)$$

$$y_k = Cx_k + v_k \quad (2)$$

$$\hat{x}_{0|1} = x_0 \quad (3)$$

$$\hat{x}_{k+1|k} = A\,\hat{x}_{k|k} + Bu_k \quad (4)$$

$$\hat{x}_{k+1|k+1} = A\,\hat{x}_{k+1|k} + K\left(y_{k+1} - \hat{x}_{k+1|k}\right) \quad (5)$$

$$P_k = AP_{k-1}A + Q \quad (6)$$

$$K = P_k\,C^T(CP_kC^T + R)^{-1} \quad (7)$$

$$y_k = C_k\,x_k + v_k \quad (8)$$

$$\tilde{x}_k = x_k - \hat{x}_k \quad (9)$$

$$e_k^- = z_k - C_k x_{k|k-1} \quad (10)$$
$$\hat{x}_k = A\hat{x}_{k|k-1} + K_k e_k \quad (11)$$
$$P_{k+1|k} = P_k - KC P_k \quad (12)$$

5.1 Alteration of output sensor values

1 Consider that attacker has manipulated sensor reading as shown in figure 2 from y_k to y_k' . Also assume that the attacker knows all state matrices ('A','B' and 'C') as well as gain factor 'K'.

2 Consider that there are 'n' sensors and attacker take control of sensor subset causing the measurement equation changes to y_k'

$$y_k' = C x_k' + v_k + \acute{\Gamma} y_k^a \quad (13)$$

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

3 Equations explained in section 5 are modified due to sensor data variation as stated below.

$$y_k' = Cx_k' + v_k + \acute{\Gamma} y_k^a \quad (14)$$
$$x_{k+1}' = Ax_k' + Bu_k' + Ke_{k+1}' \quad (15)$$
$$e_{k+1}' = y_{k+1}' - C(Ax_k' + Bu_k') \quad (16)$$

5.2 Alteration of 'A' Matrix

4 State transition matrix 'A' is modified to '$Amod$' by the attacker that causes change in State as describe below.

$$\hat{x}_{k+1|k} = Amod\hat{x}_{k|k} + Bu_k \quad (17)$$
$$\hat{x}_{k+1|k+1} = Amod\hat{x}_{k+1|k} + K(y_{k+1} - C\hat{x}_{k+1|k}) \quad (18)$$

5 Hence the difference between normal and compromised state is given by

$$\Delta x_{k+1} = (A - Amod)\Delta x_k + Bu_k \quad (19)$$
$$\Delta e_{k+1} = \Delta y_{k+1} - C((A - Amod)\Delta x_k + Bu_k) \quad (20)$$

5.3 Alternation in 'B' Matrix

5 Output matrix 'B' is modified to 'Bmod' by the attacker that causes change in State as describe below.

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bmod\, u_k \quad (21)$$
$$\hat{x}_{k+1|k+1} = A\hat{x}_{k+1|k} + K(y_{k+1} - C\hat{x}_{k+1|k}) \quad (22)$$

Hence the difference between normal and compromised state is given by

$$\Delta x_{k+1} = A\Delta x_k + (B - Bmod)u_k \quad (23)$$

Residue is changed to

$$\Delta e_{k+1} = \Delta y_{k+1} - C(A\Delta\hat{x}_k + (B - B\,mod)\Delta u_k) \quad (24)$$

Kalman gain is changed to

$$Kmodified = P_k C^T (CP_k C^T + R)^{-1} \quad (25)$$

Four tank Model is used to simulate attack model as explained in section 6.

6 SIMULATION APPROACH FOR CONFIRMATION OF THE CONJECTURE

6.1 Four tank Model

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 use as a test bed to test the effects of the applications of various control theories. The system includes two inputs (speed of pump) and two outputs (level of two tanks), where two outputs is controlled by two inputs as shown in figure 3.

'hi' is the level of water in tank *'i'(1, 2, 3 or 4)* and *'v1'* and *'v2'* are the manipulated inputs (pump speeds), *'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 non linear equation calculation. *'Ai'* 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 'μ_1' and 'μ_2' is the corresponding ratio diverted from tank two to tank four.

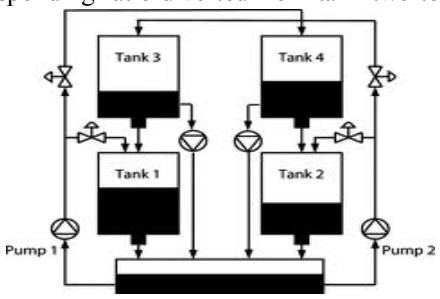

Figure 3:- Four tank System

Nonlinear model equations used for simulation [12] are as below. Table 1 gives initial values of the parameters.

$$\frac{dh_1}{dt} = -\frac{a_1(\sqrt{2gh_1})}{A_1} + \frac{a_3\sqrt{2gh_3}}{A_1} + \frac{\mu_1 k_1 v_1}{A_1} \quad (26)$$
$$\frac{dh_2}{dt} = -\frac{a_2(\sqrt{2gh_2})}{A_2} + \frac{a_4\sqrt{2gh_4}}{A_2} + \frac{\mu_2 k_2 v_2}{A_2} \quad (27)$$
$$\frac{dh_3}{dt} = -\frac{a_3(\sqrt{2gh_3})}{A_3} + \frac{(1-\mu_2)v_2 k_2}{A_3} \quad (28)$$
$$\frac{dh_4}{dt} = -\frac{a_4(\sqrt{2gh_4})}{A_4} + \frac{(1-\mu_1)v_1 k_1}{A_4} \quad (29)$$

State transition Matrix is given as below:-

$$A = \begin{bmatrix} -\frac{a_1(\sqrt{2gh_1})}{A_1} & 0 & \frac{a_3\sqrt{2gh_3}}{A_1} & 0, \\ 0 & -\frac{a_2(\sqrt{2gh_2})}{A_2} & 0 & \frac{a_4\sqrt{2gh_4}}{A_2}, \\ 0 & 0 & \frac{a_3(\sqrt{2gh_3})}{A_3} & 0, \\ 0 & 0 & 0 & -\frac{a_4(\sqrt{2gh_4})}{A_4} \end{bmatrix}$$

6.2 Simulation

Four tank model is simulated using following details. All four levels of tanks are governed by time constant Ti given as below:-

$$T_i = -\frac{A_i(\sqrt{2h_i(0)})}{a_i\sqrt{g}} \quad (30)$$

Deviation in control plant behavior can be detected by innovation value change as described in section 6.2.1. In all this mentioned attacks it's assumed that threat is due to 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%	μ_1	0.333
υ2(0)	60%	μ_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 [15]. All units in CGS.

6.2.1 Innovation Value change

Change in output data or State transition 'A' and 'B' matrix causes change in innovation value as stated below. Δe_{k+1} is difference in innovation value.

$$\Delta e_{k+1} = \Delta y_{k+1} - C((A - Amod)\Delta x_k + Bu_k) \quad (31)$$

$$\Delta e_{k+1} = \Delta y_{k+1} - (C)(A\Delta \hat{x_k} + (B - Bmod)\Delta u_k) \quad (32)$$

6.2.2 Simulation of Innovation Value change

Change in T_i values causes state matrix 'A' to change as shown below.

a) Normal system $T_i$is as given by equation 29 and innovation graph is as given in figure 4.
b) T_iModified with orifice area $'a_1'$of tank *'1'* generates innovation values as given by figure 5.

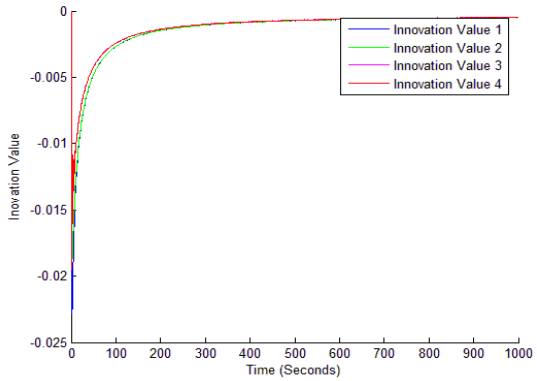

Figure 4:- Normal Plant Innovation value graph

c) For T_i modified with Area $'A_3'$ of tank *'3'* generates innovation values as given by figure 6.
d) Change in tank area 'A_i' causes change in 'B' matrix behavior as shown in figure 7.

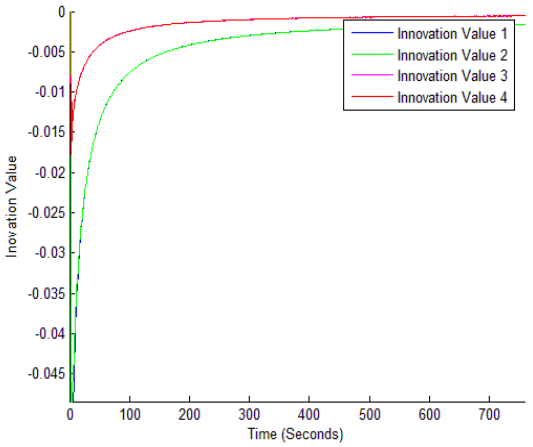

Figure 5:- Plant Innovation value graph with orifice area $'a_1'$of tank '1'modified

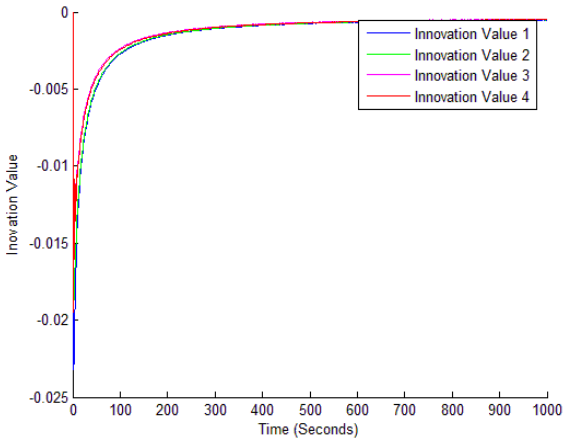

Figure 6:- Plant Innovation value graph with Tank area $'A_3'$of tank '3'modified

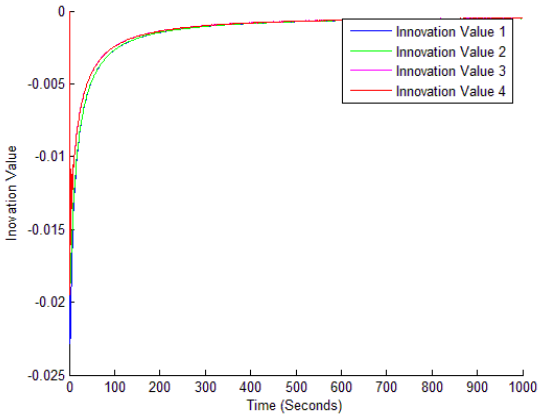

Figure 7:- Plant Innovation value graph with Tank area $'A_i'$of tank 'i' modified. Graph remains identical for all $'A_i'$ value change.

The key observations in Change in innovation value captured in graph are discussed below.

1. Change in orifice area $'a_i'$of any tank causes greater impact in innovation value change.
2. Small change in tank area $'A_i'$ does not cause remarkable change in innovation value. Hence attacker has to change $'A_i'$ to larger extend to get remarkable 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.

All the analysis approach presented above assumes that after getting innovation value it's checked for causes of variation is from input or output end. Firstly it checks for output causes by 'B' and 'C' matrices changes and finally for 'A' matrix change.

Cases where innovation value changes are undetectable SPRT technique can be used as explained in 6.2.2.

6.2.2 Sequential Probability Ratio Test

Sequential probability ratio test (SPRT) is a test to decide between two statistical hypotheses called as null and alternative hypothesis. SPRT is based on considering 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 parameter ^ as below.

$$\hat{} = \frac{\max_{Ho} f(x1.....xn|\mu,\rho)}{\max_{Ho\ U\ Ha} f(x1.....xn|\mu,\rho)} \quad (32)$$

H_o and H_a are decision criteria as per value of ^. H_o (null hypothesis) is true if ^< λ (threshold value). H_o symbolizes no Attack. H_a(alternate hypothesis) is true if ^> λ (threshold value). H_a symbolizes there has been change in behavior that indicates possibility of attack. Assuming probability density function is Gaussian; Threshold value can be expressed as below.

$$\hat{} = \frac{\max_{Ho} \prod f(x1.....xn|\mu,\rho)}{\max_{Ho\ U\ Ha} \prod f(x1.....xn|\mu,\rho)} \quad (33)$$

$$\hat{} = \frac{\max_{Ho} \prod_{i=1}^{n} \frac{1}{\rho\sqrt{2\prod}} e^{-\frac{(xi-\mu)^2}{2\rho^2}}}{\max_{Ho\ U\ Ha} \prod_{i=1}^{n} \frac{1}{\rho\sqrt{2\prod}} e^{-\frac{(xi-\mu)^2}{2\rho^2}}} \quad (34)$$

To find the MLE we need to take the derivative with respect to both of the parameters for denominator as shown in equation (34).

$$lik = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(xi-\mu)^2}{2\rho^2}} \quad (35)$$

Taking log and then derivative with respect to μ and ρ for equation 35 we get:-

$$\frac{\partial}{\partial\mu} = \widehat{\mu_{MLE}} = \bar{x} \quad (36)$$

$$\frac{\partial}{\partial\rho} = \sqrt{\frac{\sum_{i=1}^{n}(xi-\mu)^2}{n}} \quad (37)$$

We need to find maximum value for numerator, in equation (34). We get below equations:-

$$If\ \bar{x} \leq \mu_0\ then\ \widehat{\mu_{HoMLE}} = \bar{x},$$

$$\widehat{\rho_{HoMLE}} = \frac{\sum_{i=1}^{n}(xi-\bar{x})^2}{n} \quad (38)$$

$$If\ \bar{x} > \mu_0\ then\ \widehat{\mu_{HoMLE}} = \mu_0,$$

$$\widehat{\rho_{HoMLE}} = \frac{\sum_{i=1}^{n}(xi-\mu)^2}{n} \quad (39)$$

SPRT was tested with four tank systems for 1000 time interval for window size of 10. Figure 8 shows values having non-zeroes across 500, 600 and 700 which indicate that the values are not in the acceptable level. It also shows a deviation from normal behavior and possible chances of fault insertion. We have tried fault capture when innovation values deviated 20% from its original zero mean values.

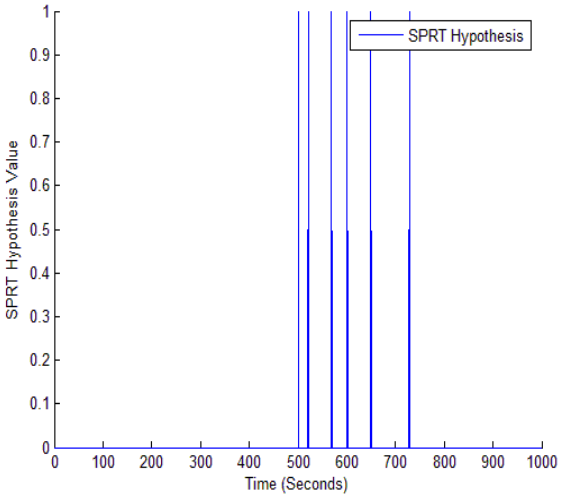

Figure 8:- SPRT test for four tank control system tested.

Wald SPRT is used to detect the presence/absence of a failure in the entire measurement as shown in figure 8. Wald SPRT knowledge can be extended to design dynamic monitor as describe in section 6.2.3 for change detection and isolation.

6.2.3 Multiple Model Approach for Monitor design

Multiple model method uses bank of filter made of multiple Kalman filter each having different behavior as per its state matrices. Figure 9 describes multiple model approach. Filter input in

the figure is equivalent to estimator output shown in figure 2. The filter input is process by a bank of filters consisting of normal filter and multiple Kalman filters. Filter input is checked with the respective innovation values of each Kalman filter in decision box. Final innovation value is calculated in decision logic and given to Filter output.

The multi-model approach is based on Bayesian decision. It has been used for fault detection and isolation and we have shown that it can be also used in detecting a control theoretic cyber-attack.

Attack detection Monitors can be constructed to detect the change in behavior of system under observation using methods such as:-

1. Change in mean and variance
2. Filter derivative method
3. CUSUM method [13]

The strategy on the Monitor design and architecture is guided by [14] and planned to be investigated in our future work.

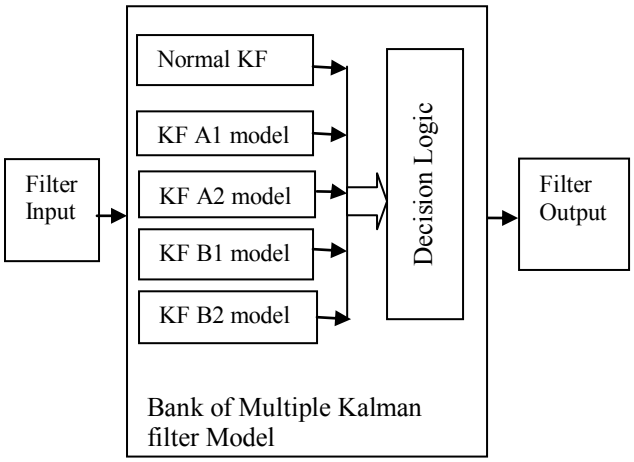

Figure 9 :- Multiple Kalman Filter Model

7 CONCLUSION AND FUTURE SCOPE

In this work we identified research challenges for securing control systems. We showed that by incorporating changes as stated in attack model on four tank system model that we were able to simulate the attacks that were detected by innovation value change and SPRT technique. Four tank system model was assumed to behave like control plant and was a helpful tool to simulate control plant behavior in attack state.

SPRT technique was further extended to design monitor that will be extended in our future work. The study will also be extended for LQG domain.

REFERENCES

1. N. Paulauskas, E. Garsva,(2006), "Computer System Attack Classification", *Automation Robotics,* pages 84 -88
2. Eric Byres, Andrew Ginter and Joel Langil, (2011), "How Stuxnet Spreads – A Study of Infection Paths in Best Practice Systems".
3. Alvaro A. Cárdenas, Saurabh Amin, Zong-Syun Liny, Yu-Lun Huangy, Chi-Yen Huangy and Shankar Sastry, March 22–24, 2011, "Attacks Against Process Control Systems: Risk Assessment, Detection, and Response", *ASIACCS '11.*
4. Yu-Lun Huang c, Alvaro A. Cárdenas a , Saurabh Amin , Zong-Syun Lin , Hsin-Yi Tsai ,Shankar Sastry, 2009, "Understanding the physical and economic consequences of attacks on control systems", *International Journal of critical infrastructure protection* , pages 73-83
5. Alvaro A. Cárdenas, Saurabh Amin, Shankar Sastry, 2008, "Secure control: Towards survivable cyber physical systems,", ICDCS '08, pages 495–500.
6. Yao Liu, Peng Ning, and Michael Reiter , 2011 , "Generalized False Data Injection Attacks against State Estimation in Electric Power Grids", *ACM Transactions on Information and System Security (TISSEC)*. Volume 14, Issue 1.
7. Slay, J. and Miller, M. (2007), "Lessons learned from the Maroochy Water Breach", in *'Critical Infrastructure Protection'*, *Springer*, pg. 73–82.
8. US Nuclear Regulatory Commission NRC Information Notice (2007), "Effects of Ethernet-based, non-safety related controls on the safe and continued operation of nuclear power stations"
9. Peter Maybeck, (1979), *Stochastic Models, Estimation and control,* volume 1, Chapter 3.
10. A. Willsky, (1976) "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pages. 601–611
11. H. L. Jones, (1973), "Failure detection in linear systems," Ph.D. dissertation, M.I.T., Cambridge, Massachusetts.
12. Edward P. Gatzke, Edward S. Meadows, Chung Wang, Francis J. Doyle III,(2000), "Model Based Control of a Four-Tank System", *Elsevier*.
13. Michele Basseville and Igor V. Nikiforov, (1993), "Detection of Abrupt Changes:- Theory and Application", *Prentice Hall*.
14. Alwyn Goodloe and Lee Pike,(2010), "Monitoring Distributed Real-Time Systems: A Survey and Future Directions", NASA.