Chapter 3

Model based Security Verification of Cyber-Physical System Based on Petrinet: A Case Study of Nuclear Power Plant

This chapter proposes a systematic approach to model a secure CPS design and evaluate of alternative designs for the same. Since attacks are the events that occur randomly and consequently adds uncertainty in system behavior even in the presence of defensive measures. A deterministic model cannot model and analyse this effect quantitatively. Moreover, an in-depth security analysis requires the identification of a standard set of evaluation metrics. We provide the design-time methodology to map and analyze system security qualitatively and quantitatively using Stochastic Petri nets (SPN)[98] and their fundamental properties. The proposed theoretical framework exploits the power of SPN to model the stochastic nature of the system in the presence of external threats and to provide mathematical support for structural and behavioral analysis to validate the effect of mitigation against the security vulnerabilities. The effectiveness of the proposed methodology is evaluated using a Nuclear power plant (NPP) case study. The proposed methodology may benefit

the practitioners and academicians to understand the modeling power of a widely adopted Petri net and its properties to filter out broad security issues at the modeling level by analyzing the model correctness properties and identifying the security threats and their impact in the early phases of the system life cycle to apply the mitigation.

Outline: The rest of the chapter is organized as follows. Section 3.1 presents the formal description of CPS. Section 3.2 proposes a methodology for security modeling and analysis. Section 3.3 evaluates the proposed approach using a case study of the digital feedwater controller system as an NPP subsystem. Section 3.4 presents discussion. Section 3.5 summaries the chapter.

3.1 Formal Description of CPS

A CPS is represented as a 3-tuple $< S, C, A >$ where S is a set of sensors, C is a set of controllers, A is a set of actuators. The dynamic behavior of CPS is formally presented as equations (3.1) and (3.2) defined in [109]

$$sm_x(t+1) = a \times sm_x(t) + b \times c_x(t) \tag{3.1}$$

$$s_x(t) = c \times sm_x(t) \tag{3.2}$$

where a, b, c are constants, $s_x(t) \in S$ denotes the measurement of sensor x , $c_x(t) \in C$ denotes the output of controller i and $sm_x(t)$ is operational state of system at time t. The intended system functionality is represented as a set of such operational states and denoted as a set $\alpha = \{sm_1, ...sm_i...sm_x\}$.

3.2 Proposed Methodology

The section presents a description of the proposed methodology for security modeling and analysis of CPS. The methodology includes five phases as shown in FIGURE 3.1 and discussed as follows

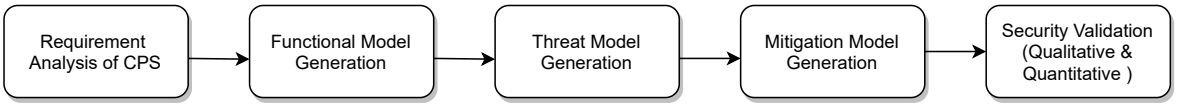

FIGURE 3.1: Proposed security modeling and analysis methodology

3.2.1 Requirement Analysis

This phase identifies the technical, in term of identified components, functional and non-functional requirements of a CPS. The identification of system components and their interactions to deliver the intended functionality with the security requirements are the output of this phase.

3.2.2 Functional Model Generation

This phase models the interaction among system components to deliver the intended functionality using SPN.

3.2.3 Threat Model Generation

This phase identifies the sensor network's vulnerabilities and specifies the type of threats that can exploit them to intrude and compromise the CPS. An active threat γ targets security goals $\beta_\alpha(t)$ to the extent of its strength $s_\gamma(v,t)$ which uses the existing vulnerability v at time t to disturb the intended functionality in threat duration δt as denoted in equation (3.3) and (3.4)

$$\beta'_{\alpha}(t) = \beta_{\alpha}(t) - s_{\gamma}(v,t) \tag{3.3}$$

such that

$$\alpha'(t) = \alpha(t) + \beta'_{\alpha}(t) \tag{3.4}$$

where α' is deviation in intended functionality.

The attacker's strength is calculated as

$$s_{\gamma}(v,t) = \frac{\text{Number of successful attacks}}{\text{Total number of attack attempts}} \tag{3.5}$$

More specifically, these threats identify and exploit the vulnerabilities within the sensors, distributed controllers, or communication channels and target the availability and integrity attributes to compromise system state $sm(t)$. Denial of Service (DoS) threat jams network traffic or exhausts the computational resources of sensor network devices to block the communication between sensors and controllers as represented in equation (3.6)

$$s'_x(t + \delta t - t) = 0 \tag{3.6}$$

Integrity threats [109] tamper the x^{th} sensor data either in additive or scaled form and described as equation (3.7) and equation (3.8)

$$s'_x(t+1) = s_x(t+1) \pm \sigma(t+1) \tag{3.7}$$

$$s'_x(t+1) = s_x(t+1) \times \sigma(t+1) \tag{3.8}$$

where $\sigma(t)$ is disturbance factor for sensor values. Thus,the exploitation of process variables and system states leads to system failure as described in equation (3.9).

$$sm_j(t + \delta t) \in FS \tag{3.9}$$

where $sm_j(t + \delta t)$ is representing state change from state $sm_j(t)$ after δt time and $FS = \overline{sm_n(t)}$ is set of failed states.

This malicious behavior is modeled as a threat model to reflect how the probable threats use different attack vectors to exploit the existing system vulnerabilities and force the system to reach undesirable states.

3.2.4 Mitigation Model Generation

Security mitigation is applied to retaliate or neutralize the effect of the threat represented as in equation (3.10)

$$\beta''_\alpha(t) = \beta'_\alpha(t) + s_\theta(t) \tag{3.10}$$

where θ is a threat mitigation, $s_\theta(t)$ is strength of applied mitigation at time t such that $s_\theta(t) \geq s_\gamma(v, t)$ to restore $\beta''_\alpha = \beta_\alpha$ that prevents violation of system's correctness properties.

$$s_\theta(t) = \frac{\text{No. of unsuccessful attacks}}{\text{Total number of attack attempts}} \tag{3.11}$$

This defensive behavior is modeled as a mitigation model using SPN.

3.2.5 Security Validation

Each successful run of the model ensures that the system is executing without any conflict and security violations. This phase evaluates model behavior qualitatively as well as quantitatively on the basis of different security metrics. The qualitative analysis shows what the effect of threats is, and quantitative analysis demonstrates how much the applied mitigation reduces the attack probability.

3.2.5.1 Qualitative Analysis

This phase qualitatively analyzes the effect of threats on system behavior using several behavioral (marking dependent) and structural (marking independent) metrics. The definition of these metrics is stated as follows

1. Coverability : The metric is closely related to reachability to verify the dynamic nature of the system. A coverable marking is a minimum marking needed to enable a transition t or make it L1-live. Formally, in an SPN (N, M_0) marking M is coverable if $M'(p) \geq M(p)$ where marking $M' \in R(M_0)$ [103].

   For security analysis, this property is useful to verify whether the system exhibits all the desirable and specified functional behavior (operational, recovered or repair) or prone to transit in any undesirable (compromised and irrecoverable) state by executing the possible firing sequences.

2. Boundedness : Boundedness condition is represented as $M(p) \leq k$. A SPN (N, M_0) is said to be k-bounded if for any marking $M \in R(M_0)$, atmost k number of tokens may reside in each place p where, k is a finite positive number. A 1-bounded net (N, M_0) $i.e$ $k = 1$ is considered safe irrespective of what transition sequence is firing [103].

   Thus, boundedness identifies the existence of overflow by verifying if the spurious tokens are generating in places and are greater than the specified k tokens to recognize that the system is behaving erroneously.

3. Liveness : An SPN is said to live for an initial marking if, for every marking belonging to the reachability set, it is possible to fire all the transitions at least once by some firing sequence. Mathematically, an SPN (N, M_0) is called Live with respect to an initial marking M_0 , if $\forall M_n \in R(M_0)$, it is possible to fire all the transitions at least once by some firing sequence σ [103].

   This is one of the elementary metrics to identify if an attacker is trying to prevent the legitimate transition to fire in a firing sequence to force the system

to transit in deadlock state (L_0) by exploiting the existing system vulnerability to disturb the L_1-live firing sequence.

4. Reversibility : This metric advocates that the system returns to its initial or home marking in each model's successful run. A Net (N, M_0) model is called reversible if $\forall(M_n) \in R(M_0), M_0 \in R(M_n)$ or $M' \in R(M_n)$ where M' is home marking covered by initial marking M_0 [103].

   Hence, reversibility is useful to detect the type of threat and verify in case of security failures if the designed system can be recovered from the compromised or failed state.

5. Persistence : An SPN (N, M_0) is persistent if an enabled transition can not be disabled without firing that are independent of each other for any state [103]. Hence, for any two enabled transitions, the firing of one cannot disable the other, and it remains enabled in the next states until it fires. Consequently, the representation of state space is reduced while preserving all deadlocks and the existence of infinite firing sequences. From a security point of view, the metric is useful to identify the DoS attack.

6. Synchronic Distance : The synchronic distance can be seen as a metric to specify the mutual dependence degree between two transitions or events. Synchronic distance sd_{ij} between two transitions t_i and t_j is defined as [103]

$$sd_{ij} = max|\bar{\sigma}(t_i) - \bar{\sigma}(t_j)| \tag{3.12}$$

   where $\bar{\sigma}(t_i)$ is firing frequency of t_i in a firing sequence σ.

7. Fairness : Transitions t_i and t_j are called bounded-fair if there exists a maximum bound l, where l is a finite positive number, on either of transition firing without others is not being fired [103].

   The fairness metric increases the system design's confidence by reducing uncertainty by verifying if the system meets the expected conditions and satisfies the

rules as per requirement and system specifications. This avoids the starvation condition or indefinite delay in the system.

8. Conservativeness : The metric is a special case of structural boundedness. An SPN is conservative if the number of tokens is conserved. Net (N, M_0) is said to be conservative if there exist an n-vector v, where n is the number of places and $\forall p, v(p) > 0$ such that the weighted sum of tokens $(M_n)^T v = (M_0)^T v = c$, where c is a constant and M_n is each reachable marking from M_0 [103].

   The property is useful to ensure the conservation of resources allotted to the CPS as there may be multiple constraints on resources related to cost, energy, or the count.

9. Repetitiveness : An SPN (N, M_0) is repetitive if, for a marking M_0 and a large firing sequence σ, every transition fires infinitely frequently.

   The metric verifies whether the system is unbiased.

10. Consistency : An SPN (N, M_0) is called to be consistent if there exists a firing sequence σ that makes M_0 reachable from M_0 itself such that each transition fires at-least-once [103].

    Consistency is a special case of repetitiveness that can be seen as a condition to verify that system can execute all possible operations as specifications. It ensures that the system is not biased or maliciously forced to perform some target tasks only.

3.2.5.2 Quantitative Analysis

Security mitigation is designed to prevent or delay the attack by increasing the cost (*i.e.* time and effort) against it. For quantitative analysis, Probability of Successful Attack (PSA) is a dominant metric as it reduces as the mitigation strength increases. The phase quantitatively analyzes the effect of applied security mitigation

by calculating the steady-state probabilities through solving the linear equations [103].

$$\pi Q = 0, \sum \pi_i = 1 \tag{3.13}$$

where, π is probability distribution, Q is transition rate matrix, π_i is the steady-state probability of being in state M_i. The transition rate matrix is calculated as

$$Q = \begin{cases} -\sum_{i \neq j} \lambda_{ij}, & i = j \\ |q_{ij}|, & i \neq j \end{cases}$$

where q_{ij} represents the transition rate of arriving at j^{th} state from i^{th} state that depends on which transition is being fired. In our threat mitigation model, the probability of firing a malicious transition $t_{k\gamma}$ to change the correct system marking and can be calculated as

$$p\{t_{k\gamma}|M_j\} = \frac{\lambda_{k\gamma}(M_j)}{\sum_{i:t_i \in E(M_j)} \lambda_i(M_j)} \tag{3.14}$$

where $t_{k\gamma} \in E(M_j)$ and rate of malicious transition $\lambda_{k\gamma}$ is calculated as

$$\lambda_{k\gamma} = 1 - s_\theta(t) \tag{3.15}$$

3.3 Case study

The proposed methodology is applied to the case study of Digital Feed Water Controller System (DFWCS) developed by NUREG/CR-6942 [5]. In a nuclear power generation plant, the main function of DFWCS is to maintain the Steam Generators (SG) water level within specified setpoint ranges (usually ±2 inch of a setpoint *i.e.* 0) under changing power demands. DFWCS serves two SGs, with each controlled by its own digital feedwater controller represented in FIGURE 3.2.

3.3.1 Requirement identification and analysis:

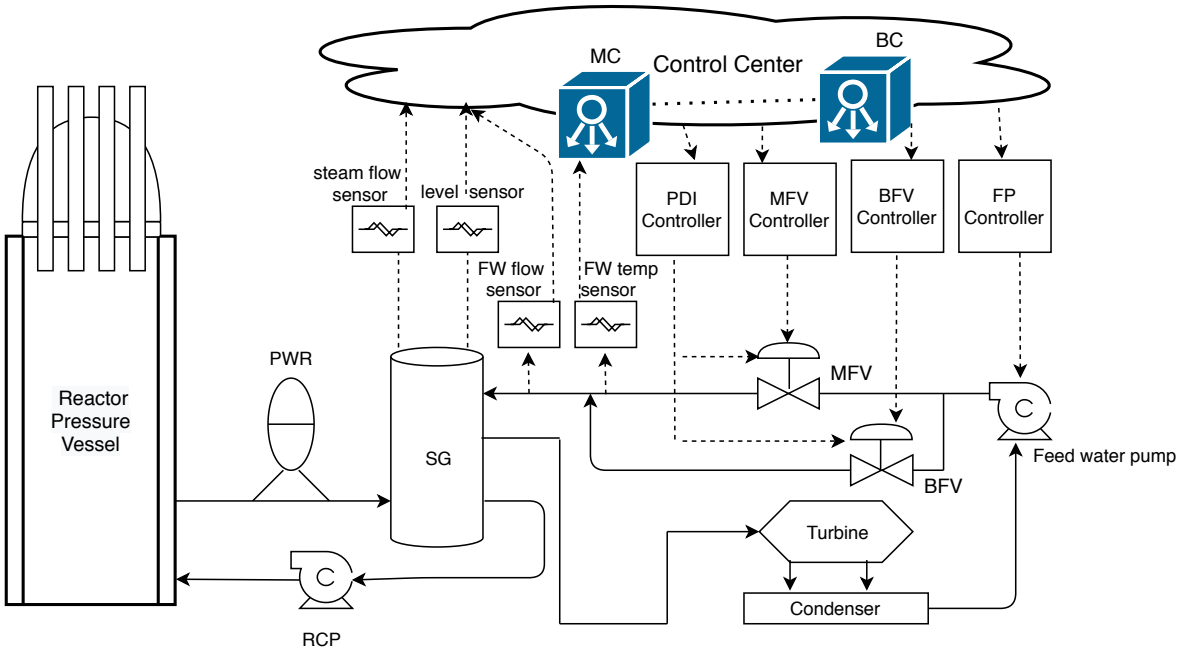

FIGURE 3.2: Architectural view of feed water controller [4]

The system is divided into physical and cyber layers to provide control capabilities. The physical layer includes Pressure Water Reactor (PWR), SG, Turbine, Condenser, pumps (Reactor Coolant Pump (RCP), Feedwater Pump (FP)), and valves. The valve list includes Main Feedwater Valve (MFV), Bypass Feedwater

TABLE 3.1: Operational mode of DFWCS

Operational mode	**Description**
Low power automatic mode	Reactor operates between 2% to 15% reactor power where BFV regulates the feed water flow.
High power automatic mode	Reactor operates between 15% to 100% reactor power where MFV and FP regulate the feed water flow.
Automatic transition from low to high power mode	When nutron flux increases at a threshold point where high mode is essential, MC signals MFV to open and BFV to close
Automatic transition from high to low power mode	When nutron flux decreases at a threshold point where low mode is essential, MC signals MFV to close and BFV to open

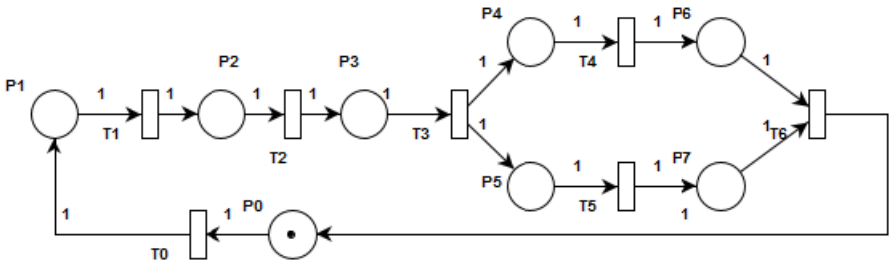

FIGURE 3.3: Functional model of DFWCS (FMA)

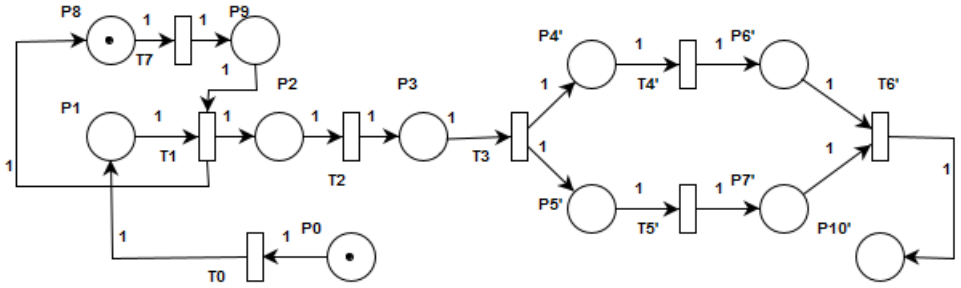

FIGURE 3.4: Integrity attack on sensor data (FMIT)

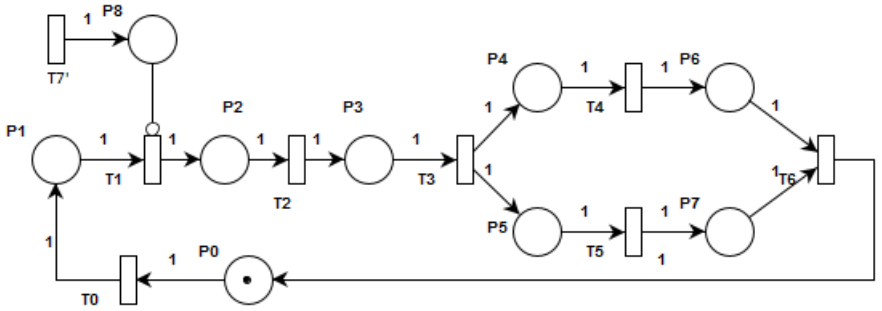

FIGURE 3.5: DoS attack on communication channel (FMDT)

Valve (BFV), Pressure Differential Indicator (PDI) to perform the controlled process. The cyber layer collects information about the state of the controlled process through the use of several sensors to measure feedwater level, neutron flux, feedwater flow, steam flow, and feedwater temperature. Main computer (MC) and backup computer (BC) in the control center (CC) process this environment information and generate the output signal for BFV, MFV, PDI, FP controllers. These controllers forward the same signal to their respective controlled devices to perform the required actions. Although, these signals can be overridden by operators if a fault in MC is being diagnosed. Thus, DFWCS digital controller is responsible for regulating the

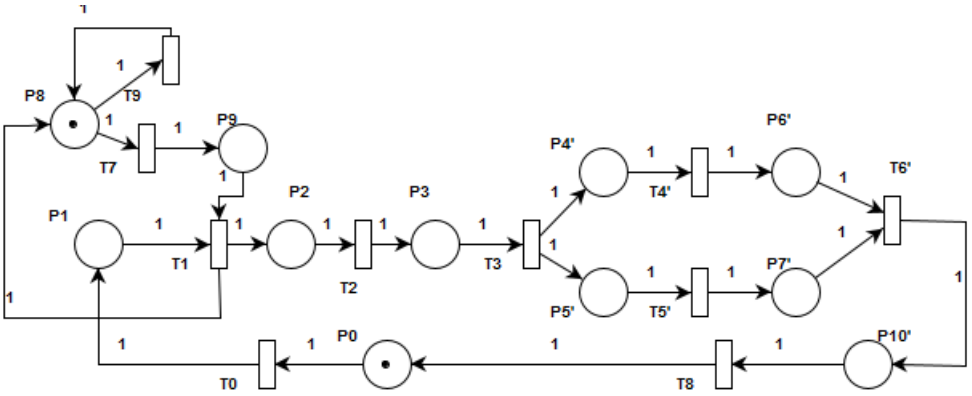

FIGURE 3.6: Functional model of DFWCS with security measure (FMM)

flow of feed water to SG to maintain the constant water level in it. From the operational perspective, DFWCS operates in different modes depending on the power generated in the primary system, as mentioned in TABLE 3.1. Although, we are just considering the scenario where the system operates in high power automatic mode for security modeling and analysis. In this mode, the SG level drops if the increment in steam flow will not match with a change in feed water flow accordingly. This results in reactor trip condition as the water level in SG becomes too high (30 inches above from setpoint) or falls too low (less than 24 inches than setpoint).

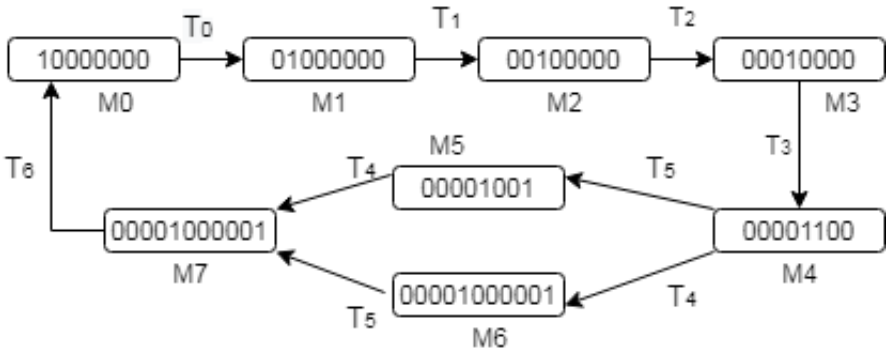

FIGURE 3.7: Reachability graph of FMA

3.3.2 Functional Model Generation

The functional model of DFWCS demonstrates the use case of maintaining the SG water level in high power mode. The functional model in the absence of attack

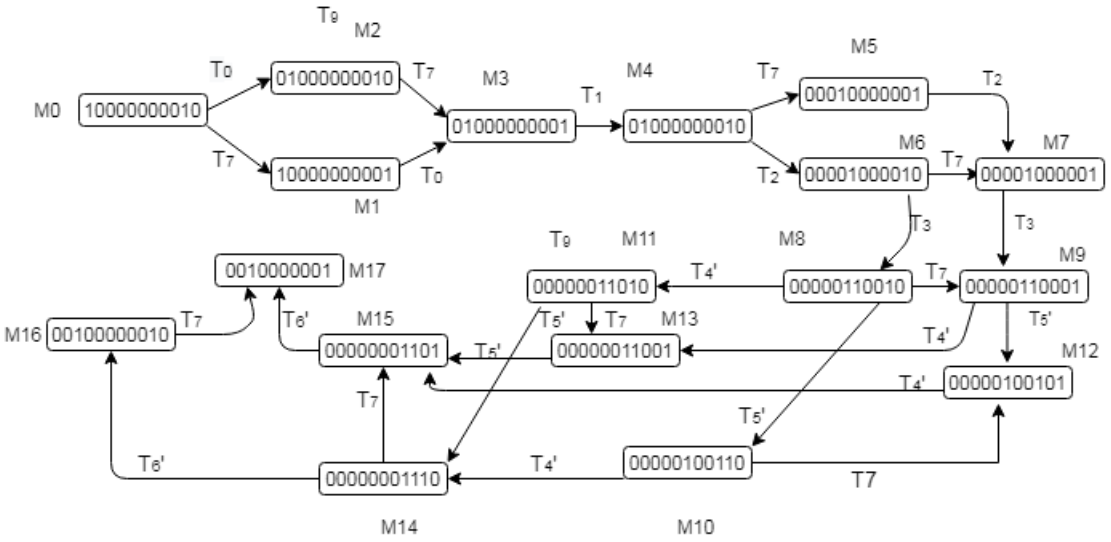

FIGURE 3.8: Reachability graph of FMIT

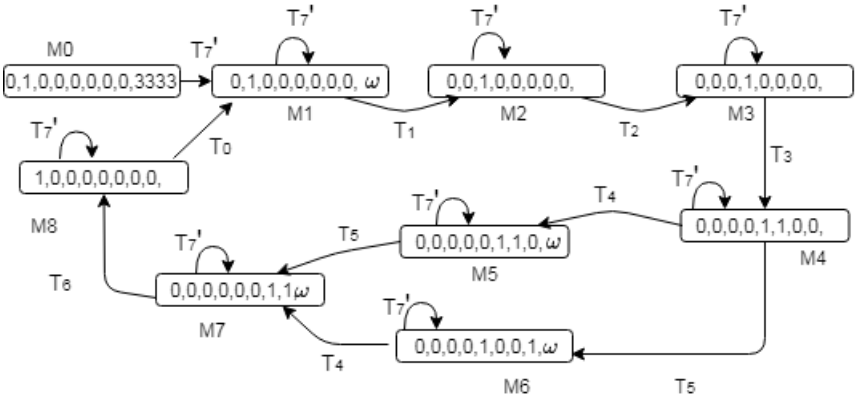

FIGURE 3.9: Reachability graph of FMDT

(FMA) is shown in FIGURE 3.3, where the system starts in working condition with no failed components. The description of corresponding places and transitions is given in TABLE 3.2 and TABLE 3.3, respectively.

The token in place P_0 shows that the process starts when SG feed water is in a normal state. Transition T_0 fires when the feedwater level starts decreasing, and the system reaches to water level out of range state P_1. T_1 fires to send the level sensors value to MC and this value is received by MC for analysis denoted by a token in P_2. T_2 fires to show that MC analyses received data and verify the setpoints and process reaches to comparative diagnosis completed state P_3. T_3 fires to send control directives to MFV and FP controllers after comparative diagnosis by MC and token

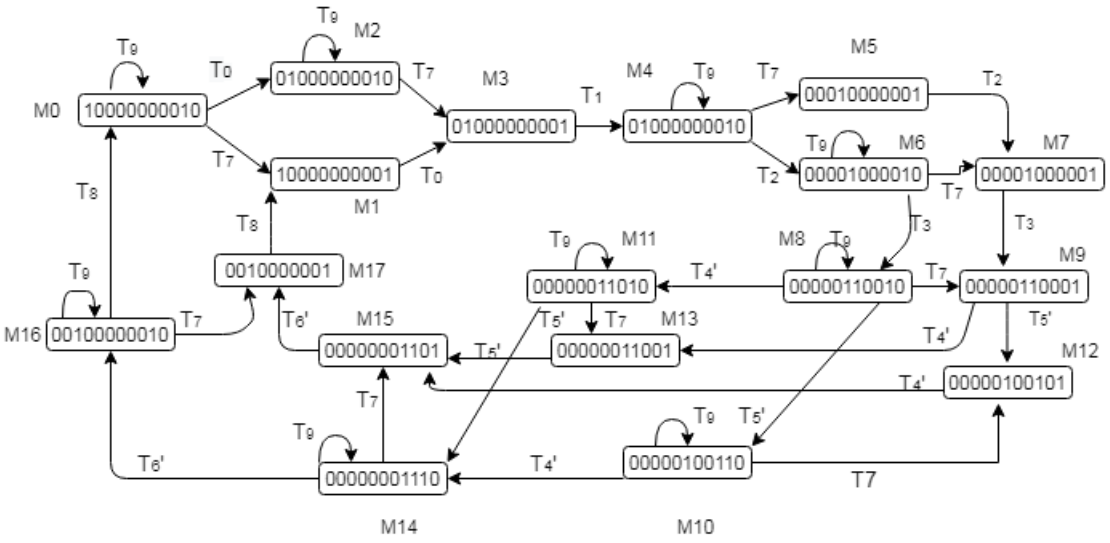

FIGURE 3.10: Reachability Graph of FMM

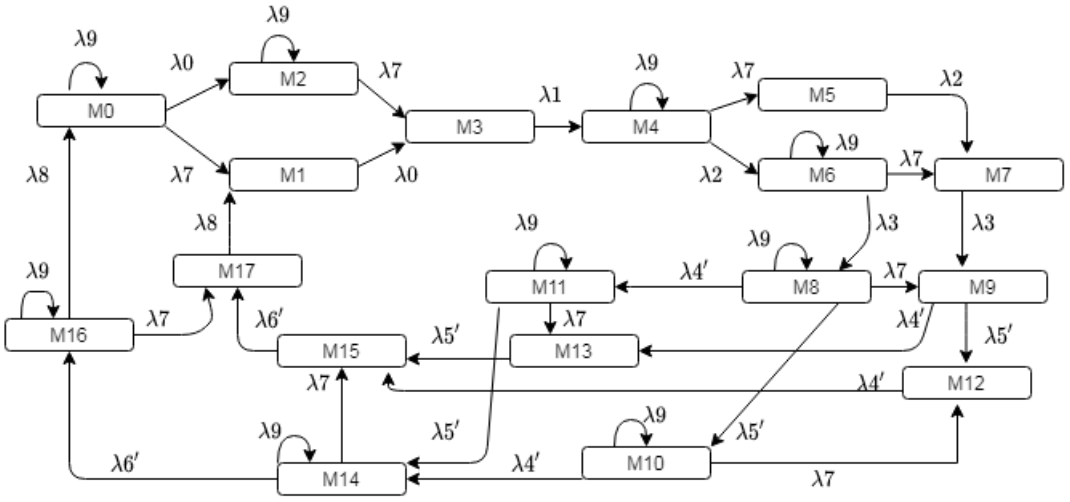

FIGURE 3.11: Markov chain corresponding to FMM reachability graph

is deposited in both places P_4 and P_5 to represent MC signals are interpreted by MFV and FP controller. The token reaches to state P_6, and P_7 where MFV position is adjusted and FP speed is increased on firing transition T_4 and T_5 to send the signal for adjusting the position of MFV and increase the speed of FP. As a result, transition T_6 fires for increasing the feed water supply to maintain its normal range in SG.

$$
Q_{FMM=}\begin{array}{c}\\ M0\\ M1\\ M2\\ M3\\ M4\\ M5\\ M6\\ M7\\ M8\\ M9\\ M10\\ M11\\ M12\\ M13\\ M14\\ M15\\ M16\\ M17\end{array}
\begin{array}{c}
\begin{array}{cccccccccccccccccc} M0 & M1 & M2 & M3 & M4 & M5 & M6 & M7 & M8 & M9 & M10 & M11 & M12 & M13 & M14 & M15 & M16 & M17 \end{array}\\
\begin{bmatrix}
-\lambda_0-\lambda_7 & \lambda_7 & \lambda_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -\lambda_0 & 0 & \lambda_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -\lambda_7 & \lambda_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -\lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -\lambda_7-\lambda_2 & \lambda_7 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -2\lambda_3 & \lambda_3 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_3 & 0 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_7-\lambda_4'-\lambda_5' & \lambda_7 & \lambda_5' & \lambda_4' & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_4'-\lambda_5' & 0 & 0 & \lambda_5' & \lambda_4' & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_7-\lambda_4' & 0 & \lambda_7 & 0 & \lambda_4' & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_7-\lambda_5' & 0 & \lambda_7 & \lambda_5' & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_4' & 0 & 0 & \lambda_4' & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_5' & 0 & \lambda_5' & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_7-\lambda_6' & \lambda_7 & \lambda_6' & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_6' & 0 & \lambda_6'\\
\lambda_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_7-\lambda_8 & \lambda_7\\
0 & \lambda_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_8
\end{bmatrix}
\end{array}
$$

FIGURE 3.12: Transition rate matrix corresponding to FMM

3.3.3 Threat Model Generation

The sensor network is compromised by exploiting the existing network-level vulnerabilities (insecure communication channel, remote access to the enterprise network, or node spoofing). It is assumed that the attacker is persistent and performs an attack with probability one (i.e., whenever it has a chance). He has knowledge of control

TABLE 3.2: Place Description of FMA

Place	**Description**
P0	SG Feed water in normal range
P1	Water level out of range
P2	Data received by MC
P3	Comparative diagnosis completed
P4	MC signal interpreted by MFV controller
P5	MC signal interpreted by FP controller
P6	MFV Position adjusted
P7	FP speed increased

TABLE 3.3: Transition Description of FMA

Transition	**Description**
T0	Feed water level starts decreasing
T1	Level sensors sending the observed value to MC
T2	MC analyses received data
T3	MC sending control directives to MFV and FP controllers
T4	MFV controller signals MFV to adjust position
T5	FP controller signals FP to increase speed
T6	Increasing feed water flow

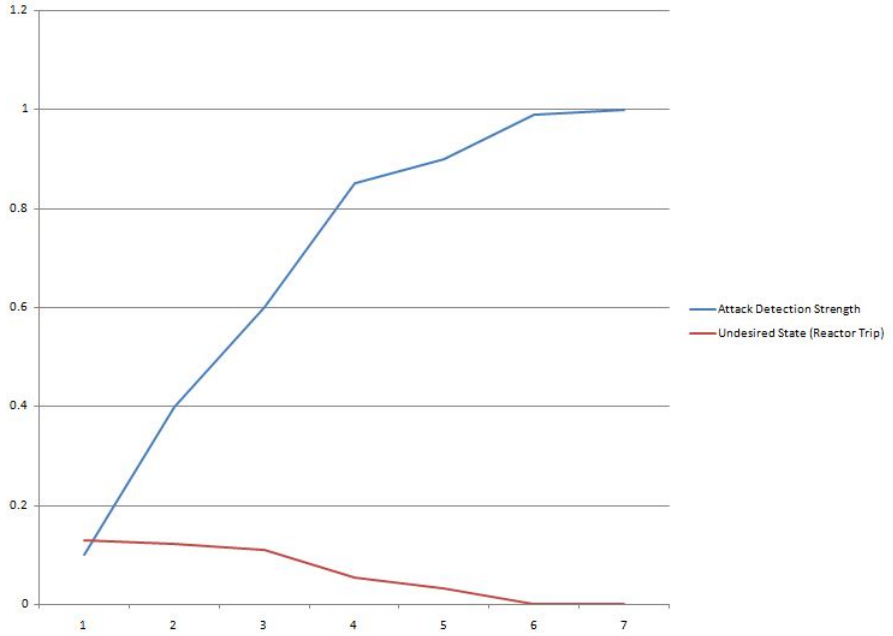

FIGURE 3.13: Effect of readjustment of mitigation strategy

and operational information such as network topology and process parameters. He intercepts the sensor-controller communication network to temper the communication data or signals, jam the communication channels, and exhausts the computational resources of the system to target the integrity and availability attributes of DFWCS. For that different attacks such as ARP spoofing, false data/command injection attacks, Emergency-stop abuse, or replay attacks may be launched to intrude the sensor networks and compromise DFWCS successfully.

In our threat model, possible attacks are modeled as malicious events or transitions responsible for altering the system model's behavior that violates its correctness properties. Almost every transition is the same as in the functional model, although the pre and post conditions are reformulated to reflect the effect of the attack.

FIGURE 3.4 shows the system model under integrity attack through the sensor network (FMIT), and FIGURE 3.5 shows the system model under DoS attack through a communication channel between the sensor and the control center (FMDT). In the FMIT model, initially, the tokens at place P_0 and P_8 denote that SG feedwater is in the normal range and the attacker tries to intrude the system. As both transition T_0

and T_7 are enabled, any of the events may occur first. Firing T_0 denotes feed water level decreasing and the system reaches the water level out of range state P_1. Firing of T_7 denotes the attacker successfully performs man-in-the-middle attack [91] for false data injection and token reaches to place $P9$, which denotes the communication network is compromised. When both place P_1 and P_9 have tokens transition T_1 fires and deposits a token at place P_8 which represents the attacker again try to attempt another attack and at place P_2 which denotes the level sensors sending the values through the compromised channel where the attacker tempers sensor values to falsify the state of water out of range (decreasing) as increasing water level range and MC received this tempered value. Transition T_2 enables and fires to analyse the data and reached to comparative diagnosis completed state P_2. Firing of Transition T_3 represents the incorrect control directives are sent to MFV and FP controllers. Tokens at place P'_4 and P'_5 represent incorrect signals are interpreted by MFC and FP. Firing of T'_4 and T'_5 denotes neither MFV controller signals MFV to adjust position nor FP controller signals FP to increase speed. Hence, even water is out of range, MC does not signal MFV and FP controllers to adjust the MFV position and increase FP speed to match the feed water flow with the stream flow. As a result, water level decreased below the critical threshold level *i.e.* system reaches to an undesired reactor trip state P'_{10}. In FMDT, the tokens at place P_0 represent SG feedwater is in the normal range and firing of transition T_0 deposits token in P_1. As both transition T_1 and T'_7 are enabled, any of the events may occur first. The firing of T_1 denotes level sensors sending the state values, and no DoS attack is performed yet. Transition T'_7 is proactive transitions that enable always. Firing T'_7 denotes the malicious script execution by an attacker to successfully perform man-in-the-middle attack and generate illegitimate packets (flooding attacks) to exhaust the network bandwidth. The token at places P_8 represents illegitimate packets are generated to jam the network. The inhibitor arc between P_8 and T1 represents the loss of communication between sensor-and MC by inhibiting the legitimate data packet (token) coming from P_1, making MC unable to take control decisions for δt time.

3.3.4 Mitigation Model Generation

The behavior of security mitigation (FMM) is modeled in FIGURE 3.6. This retaliates the attack for preventing or reducing the reactor trip condition rate (as modeled in FMIT). As a security mitigation measure, a Network intrusion detection and response system (NIDRS) is applied to detect and respond against the unusual behavior of sensor nodes in sensor networks. In the FMM model, initially, the tokens at place P_0 and P_8 denote that SG feedwater is in the normal range and the attacker tries to intrude the system. As the transition T_0, T_7, and T_9 are enabled, any of the events may occur. The firing of T_0 denotes feedwater level decreasing, and the system reaches the water level out of range state P_1. The firing of T_9 denotes NIDRS detects and responds to the attack attempt, and this forces the attacker to start afresh. Firing of T_7 denotes NIDRS fails to detect the false data injection attack, and token reaches to place $P9$ to represent communication network is compromised and undetected. The place P_2, P_3, P'_4, P'_5, P'_6, P'_7 and P'_{10} are same as in FMIT. Similarly, transitions T_2, T_3, T'_4, T'_5 and T'_6 are same as in FMIT. A manual recovery starts when transition T_8 fires to bring back the system in operational mode.

The NIDRS combines a voting approach along with behavior analysis. To detect integrity attacks, it uses a voting approach where it collects the values from redundant sensor nodes and compares the data to verify the unusual behavior of sensor nodes. To detect flooding attacks, the NIDRS monitors the rate of an incoming packet from each sensor and network device and analyzes the usage of resources like power, bandwidth, and memory. The responses are generated in the form of alerts and blocking of suspects nodes. In case of NIDRS fails to detect and respond to the attack, the system reaches an undesired state, and manual recovery starts to bring the system back to the initial condition. The strength of NIDRS $s_\theta(t)$ is readjusted by configuring its false-positive rate (FPR) where an acceptable behavior is considered as an attack and false-negative rates (FNR) where NIDRS fails to identify an attack. Lower FPR and FNR results in higher attack detection capabilities that can

be obtained by applying the appropriate design rules, enriching and updating its knowledge base.

3.3.5 Security Metrics Validation

In this phase, FMA, FMIT, FMDT, FMM models are evaluated qualitatively and quantitatively.

3.3.5.1 Qualitative Analysis

Qualitative analysis identifies the effect on correctness metrics of system model under integrity (FMIT) and DoS threat (FMDT) compared to a functional system in the absence of attack and defense (FMA). A comparison of correctness metrics corresponding to FMA, FMIT, FMDT, and FMM is given in TABLE 3.4. Here, 1 denotes a particular security metric is satisfied and 0 represents its violation.

1. Coverability : The reachability/coverability graphs corresponding to FMA and FMIT and FMDT are shown in FIGURE 3.7, FIGURE 3.8, and FIGURE 3.9 respectively. Reachability graph of FMA shows the system represents the correct functional behavior as it reaches to all the specified Marking corresponding to marking $\{P_0,\ P_1,\ P_2,\ P_3,\ P_4,\ P_5,P_6\}$. The Reachability graph of FMIT is generated corresponding to marking $\{P_0,\ P_1,\ P'_{10},\ P_2,\ P_3,\ P'_4,\ P'_5,\ P'_6,\ P'_7,\ P_8,\ P_9\}$ which represents the attacked system takes a different path than the correct operational path to transit in compromised Marking M_{17}. Reachability graph of FMIT is generated corresponding to marking is generated corresponding to places $\{P_0,\ P_1,\ P_2,\ P_3,\ P_4,\ P_5,\ P_6,\ P_7,\ P_8\}$ to shows a DoS attack leads to infinite communication delays by sending illegitimate packets to exhaust the network resources.

2. Boundedness : The reachability graph of FMA shows that FMA is 1-bounded and safe. As integrity attack only temper the values to change the operational

path without generating spurious tokens, FMIT is also 1-bounded and safe. The reachability graph of FMDT shows that FMDT is unbounded and unsafe as a huge number of spurious tokens are generated that may be responsible for exhausting system resources.

3. Liveness : In the case of safety-critical CPSs like NPP, the violation of liveness property may result in a deadlock situation and unable to reach in initial or any home state. Hence, liveness is useful to ensure whether DFWCS is deadlock-free and not going in any dead state. FMA and FMDT are live, Although FMIT is not live as there is no transition exists between marking M_{17} to M_0.

4. Reversibility : In the case of threat-free condition, all the markings of FMA are reachable from the initial marking and vice versa. However, this metric is violated in FMIT and FMDT.

5. Persistence : The persistence property is satisfied for FMA and FMIT. It is violated for FMDT as the firing of transition T_7' disables the transition T_1.

6. Synchronic distance : For FMA, each transition is mutually dependent. Hence this property is satisfied and denoted as 1. In an active attack, a threat agent disrupts functional dependencies. For instance, the Firing sequence of FMIT (σ_{FMIA}) is $T_0T_1T_2T_3T_4'T_5'T_6'$ which does not include the firing of T6 and sd_{06} is ∞. In the case of FMDT, where a possible firing sequence is $T_0T_7'...T_7'...$ *i.e.* T_7' is firing infinitely. In such situation sd_{07} is also ∞ as mentioned in TABLE 3.4. Hence, while designing critical CPSs defender tries to increase the synchronic

TABLE 3.4: Security metrics evaluation table of FMA, FMIT, FMDT

Security Metrics	**FMA**	**FMIT**	**FMDT**
Coverability	1	0	0
Boundedness	1	1	0
Safe	1	1	0
Liveness	1	0	1
Reversibility	1	0	0
Persistence	1	1	0
Synchronic distance	1	∞	∞
Fairness	1	0	0
Conservativeness	1	1	0
Repetitiveness	1	0	0
Consistency	1	0	0

distance between operational and probable compromised state by applying the preventive defense states to either fail attack attempt or add delay or increase attack costs.

7. Fairness : The fairness property is preserved in FMA as each transition has equal chances to fire. The metric is violated in FMIT and FMDT as, In FMIT, the system does not return to its initial state. DoS threat violates the fairness condition due to proactive transition T_7', as shown in FMDT.

8. Conservativeness : The conservativeness is preserved in FMA and FMIT as there is no token loss or generation in both the model. Although, the property is not preserved in FMDT, where the weighted sum of tokens is not the same for every marking.

9. Repetitiveness : The repetitiveness metric is preserved in FMA. The property is violated in FMIT and FMDT as in FMIT as after reaching to marking M_{17}, no transitions fire and in FMDT, due to frequent firing of $T8$, other transitions in σ_{FMDT} cannot fire frequently.

10. Consistency : As consistency is a special case of repetitiveness, hence, it is compromised in both FMIT and FMDT and only preserves for FMA.

    The qualitative analysis is based on the structural and behavioral attributes of the PN model. Hence, applicable to any external active security attack.

3.3.5.2 Quantitative Analysis

To analyze the effect of security mitigation, a Markov Chain (MC) is generated corresponding to the reachability graph of the FMM model and shown in FIGURE 3.11. We perform a sensitivity analysis to analyze the effect of adjusting the strength of NIDRS. The firing rates corresponding to each transition are calculated using mean firing delay of respective transitions ($\lambda_m = 1/d_m$) [103]. The mean firing delay for each transition is based on expert elicitation approach, system specifications, and experiences from similar projects as mentioned in several studies [90, 133, 28, 126,

127]. In our case the mean firing delay of the transitions that propagate the effect of the attack in FMM is computed as 1, i.e., $d_0 = d_1 = d_2 = d_3 = d'_4 = d'_5 = d'_6 = d_8 = 1$. Hence, the firing rates corresponding to these transitions are calculated as $\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 = \lambda'_4 = \lambda'_5 = \lambda'_6 = \lambda_8 = 1$. To perform sensitivity analysis, the values of $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda'_4, \lambda'_5, \lambda'_6, \lambda_8$ remain constant and only λ_7, λ_9 vary to reflect the effect of readjusting the $s_\theta(t)$ on PSA in each run. λ_9 represents the strength of NIDRS $s_\theta(t)$ and obtained using equation 3.11. Its value start with 0.1 and further suitable increments are done as shown in TABLE 3.5 to observe the effect of readjusting the λ_9 on the probability of the system being in an undesired state M_{17}. The value of λ_7 is calculated using equation (3.15) which is placed in the equation 3.14 to obtain the probability that NIDRS fails to detect and respond the active attack. For each run, the steady-state probabilities of FMM are calculated by solving equation (3.13) where

$\pi_{FMM} = [\ \pi_0\ ,\pi_1,\ \pi_2,\ \pi_3,\ \pi_4,\pi_5,\ \pi_6,\ \pi_7,\ \pi_8,\ \pi_9,\ \pi_{10},\ \pi_{11},\ \pi_{12},\ \pi_{13},\ \pi_{14},\ \pi_{15},\ \pi_{16},\ \pi_{17}]$, $Q = Q_{FMM}$ and

$$\pi_0+\pi_1+\pi_2+\pi_3+\pi_4+\pi_5+\pi_6+\pi_7+\pi_8+\pi_9+\pi_{10}+\pi_{11}+\pi_{12}+\pi_{13}+\pi_{14}+\pi_{15}+\pi_{16}+\pi_{17} = 1$$

TABLE 3.5 shows the effect on steady state probabilities of each state when the strength of λ_9 varies in each run. The values of π_{17} shows the probability to reach state M_{17} reduces when the $s_\theta(t)$ increases and it gives optimal result of $\pi_{17} = 0.00001$ when $\lambda_9 = 0.999$ and $\lambda_7 = 0.001$. The result corresponding to TABLE 3.5 is plotted as a graph in FIGURE 3.13 for better visualization.

The quantitative analysis is demonstrated on the case study specifically with respect to NIDRS as security mitigation and the results show that it reduces the probability of the system being successfully attacked and reached an undesired state. Although, the probability of the system not being in an undesired state highly depends on how effective the defense system is (*i.e.* on NIDRS strength). Similarly, it will be applicable to other external active attacks. The state-space analysis of the model is performed with PIPE v4.3 Tool [33].

TABLE 3.5: Effect of readjustment of mitigation strength on steady-state probabilities λ_9

	π_0	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}	π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	π_{16}	π_{17}
λ_9=0.1	.00195	.13109	.00217	.13304	.07002	.06302	.03685	.09619	.01271	.05381	.00669	.00669	.05983	.05983	.00704	.126	.00371	.12934
λ_9=0.2	.00279	.13008	.00349	.13287	.07382	.05905	.04101	.09186	.01465	.05179	.00814	.00814	.0583	.0583	.00904	.12383	.00502	.12785
λ_9=0.4	.00605	.12594	.01009	.13199	.08249	.0495	.05156	.08043	.01983	.04616	.01239	.01239	.0536	.0536	.01549	.1165	.00968	.012231
λ_9=.6	.01423	.11436	.03558	.12859	.09185	.03674	.06561	.06298	.02734	.03696	.01953	.01953	.04477	.04477	.02789	.1007	.01992	.10867
λ_9=0.8	.03374	.07709	.16872	.11084	.09236	.01847	.07697	.03387	.03499	.02043	.02916	.02916	.02626	.02626	.04859	.06224	.04049	.07034
λ_9=0.9	.04175	.03591	.41752	.07766	.0706	.00706	.06419	.01348	.03056	.00827	.02779	.02779	.01105	.01105	.05052	.02714	.04593	.03174
λ_9=0.99	.00926	.00062	.92592	.00988	.00978	.0001	.00968	.00019	.00482	.00012	.00477	.00477	.00017	.00017	.00945	.00043	.00935	.00053
λ_9=0.999	.00099	.00001	.99251	.001	0	.001	0	.0005	0	.0005	.0005	0	0	.00099	0	.00043	.00099	.00001

3.4 Discussion

The proposed approach includes qualitative and quantitative analysis for security evaluation. However, the success of proposed method highly depends on the correct estimation of model parameters such as firing rates of transitions and strength of security measure, which is based on system specifications, expert elicitation approach, and experiences from similar projects as mentioned in section 3.3.5.2. We have done sensitivity analysis as shown in TABLE 3.5 for demonstrating the probability of successful attacks highly depend on the strength of mitigation, which can be adjusted by applying the appropriate design rules, enriching and updating its knowledge base.

3.5 Summary

CPSs are vulnerable to different sophisticated cyber threats. Consideration of security in the early stage of the system development cycle helps deliver a more robust and cost-effective system. This chapter has made an attempt to explain model-based security verification at the sensor network level using SPN and performed a design-time analysis based on several identified system and security metrics. The qualitative analysis shows that different type of threats attempts to violate a subset of system and security metrics according to their malicious behavior by using different attack vectors. The quantitative analysis shows the impact of attack and responsive mitigation methodology on the system and adjusting mitigation strength in response to attacker significantly reduces the probability of a successful attack. Hence, the proposed methodology may help academicians and system analysts filter out broad security issues at the modeling level by checking the model correctness properties. However, as system availability is one of the major requirement of system design, there is need of formal specification and analysis of the combined effect of preventive and responsive measures on system availability. Hence, in next chapter, we have presented a formal model to analyse the combined effect of applying preventive and responsive measures on system availability.