International Journal of Electrical and Computer Engineering (IJECE)
Vol. 7, No. 3, June 2017, pp. 1228~1239
ISSN: 2088-8708, DOI: 10.11591/ijece.v7i3.pp1228-1239

Evaluation of Feature Reduction using Pricipal Component Analysis and Sequential Pattern Matching for Manet

M. Reji, P. C. Kishore raja, Bhagyalakshmi M
Departement of Electronics and Communication Engineering, Saveetha University, India

Article Info

Article history:

Received Aug 12, 2016
Revised Nov 10, 2016
Accepted Nov 24, 2016

Keyword:

MANET
PCM
SPM

ABSTRACT

In Mobile Ad hoc Networks (MANETs) there are some security problems because of portability, element topology changes, and absence of any framework. In MANETs, it is of extraordinary significance to identify inconsistency and malignant conduct. With a specific end goal to recognize malignant assaults by means of interruption identification frameworks and dissect the information set, we have to choose some components. Thus, highlight determination assumes basic part in recognizing different assaults. In the writing, there are a few recommendations to choose such elements. For the most part, Principal Component Analysis (PCA) breaks down the information set and the chose highlights. In this paper, we have gathered a list of capabilities from some cutting edge works in the writing. Really, our reproduction demonstrates this list of capabilities identify inconsistency conduct more precise. Likewise, interestingly, we utilize PCA for investigating the information set. In contrast to PCA, our results show Sequential pattern mining (SPM) cannot be affected by outlier data within the network. The normal and attack states are simulated and the results are analyzed using NS2 simulator.

Corresponding Author:

M. Reji,
Departement of Electronics and Communication Engineering,
Saveetha University,
India.
Email: rejime@gmail.com

1. INTRODUCTION

Portable Ad-hoc Network (MANET) is an unstructured remote framework that can be developed quickly, every center point is silly and free in the fundamental administration. In MANET, center points can incorporate to the framework or isolate from it at whatever point. Thusly, there I no central control on the framework for the center points to take after. Interference disclosure models were introduced by Denning in 1987 and rather are another development.All through the previous decades, we have understood that interruption anticipation techniques (e.g. cryptography, confirmation and so forth) can't ensure and secure our system appropriately. Thusly, interruption recognition frameworks (IDS) could secure the systems against assaults from pernicious hosts. In the system layer, different directing conventions require the collaboration between portable hubs; thus, brings about various vulnerabilities in MANET. Principal component Analysis (PCA) is utilized as a part of MANET to dissect the chose highlights [3]. Then again, PCA is a semi-regulated way to deal with distinguish peculiarity and it needs an unadulterated secured system amid framing the benchmark profile. Notwithstanding, on account of versatility of the MANET hubs and utilizing Ad hoc On demand Distance Vector (AODV) directing convention, we never could make sure the learning time of PCA has been secured totally. It ought to be noticed that AODV never validates hubs inside the system. Rather than wired systems, there is no information set in MANET so as to learn pattern profile in semi-regulated calculations essentially. In this way, we need to utilize unsupervised calculations to gather

information free of any information set from system. Consequently, we need to utilize vigorous PCA keeping in mind the end goal to utilize unsupervised approach and shaping the gauge profile more precise for abnormality recognition.

2. RELATED WORK

To deal with the extended information security threats, various sorts of security sorts of rigging have been used as a part of the immense scale framework. These supplies make packs of security events. It's especially difficult to obtain the security state of the whole framework completely while going up against an overabundance of alert information. To settle this issue, various request about had displayed the possibility of condition care into web security structure. Bass was the principle who brought this thought into framework and present the system security perception plot in light of multi-sensor data blend [1] proposes another component choice calculation called Optimal Feature Selection calculation in view of Information Gain Ratio and acquire the exactness .Ayman I proposes erasing superfluous and excess elements fabricates a quicker preparing and testing procedure, to have less asset utilization and in addition to keep up high recognition rate[2] Dr. Saurabh Mukherjeea propose strategy Feature Vitality Based Reduction Method, to recognize vital lessened information highlights. We apply one of the effective classifier guileless bayes on lessened datasets for interruption recognition [3] Vetrichelvi Rajaram PCA is utilized to investigate the chose highlights. This is on account of excess and unimportant components frequently diminish execution of the discovery framework [4]. Fang Lan propose a structure for system security circumstance mindfulness taking into account learning discovery [5] Mohammad K. Houri Zarch use strong PCA for breaking down the information set rather than PCA in MANET [6]. Srilatha Chebrolua cross breed design for joining diverse element choice calculations for true interruption identification

3. ANAMOLY DETECTION ENGINE

Consistently, different methodologies have been introduced keeping in mind the end goal to recognize interruption in the system by method for Principal Component Analysis (PCA). In this segment we survey on main segment investigation and successive example mining.

a. Principal Component Analysis

Essential segment investigation (PCA) was concocted in 1901 by Karl Pearson. PCA includes a scientific method that changes various conceivably related variables into an arrangement of estimations of straightly uncorrelated variables called central parts. PCA is the most across the board technique for information pressure and representation [4]. Fundamental point of interest of PCA is that once you have found these examples in the information, you pack the information, i.e. by diminishing the quantity of measurements, without much loss of data. By and large, PCA tries to give us the most imperative hub, express the disseminating of information, by discovering relationship between different component.

Steps for Principal component Analysis:

a) Taking the whole dataset ignoring the class labels
b) Computing the d-dimensional mean vector
c) Computing the Scatter Matrix
d) Computing the Covariance Matrix (alternatively to the scatter matrix)
e) Computing eigenvectors and corresponding eigenvalues
f) Transforming the samples onto the new subspace

b. Sequential Pattern Matching

Steps for Sequential Pattern Matching:

a) Once the feature selection process gets completed, sequence of the features is formed for all nodes in the neighborhood.
b) In the formed sequence each row represents the nodes and each field represents the feature of the particular nodes
c) The reference pattern of the each sequence field is formed by estimating the min and max bound values based on the average value and difference value of individual sequence.
d) Sequential pattern matching process is performed by checking for all nodes for all available features with the reference feature
e) During the pattern matching process, the matched features are consecutively compared with reference without mismatching for identifying the strong sequential match
f) If a strong sequence match is found then sequential pattern is checked for semi sequence by validating the LP point.
g) Else if sequence is found then the sequence is classified as sequential pattern.

h) Else the pattern is identified as non-sequential pattern. Based on this matching anomaly nodes are identified

4. FEATURE SELECTION

Needless to say feature selection methodology plays a critical role in data analysis in order to detect different attacks in MANETs. Features should be able to describe the behavior of the network precisely. Moreover, if new attacks are defined in the future, it can be also detected by these proper features. Thus, choosing right and decent features in MANETs helps us to know more about the behavior of our network from different aspects. On the other hand, there are many works that tried to define and select different features to analyze and detect various attacks. Huang et al, use 141 features for describing the normal behavior of protocol. Cabrera et al,, use 28 features for describing the normal behavior of AODV. Also, Nakayama et al, use 14 features for detecting anomaly in the AODV protocol. Moreover, Zhang et al, have collected some features related to the normal behavior of network from Medium Access Control (MAC) layer, network layer, and application layer. Most of them select traffic features and take advantage from control messages. Huang et al, , defined 132 traffic features for normal behavior of network by considering some issues like the number of send, receive, drop, and forward the control packets in 5 seconds, 60 seconds, and 900 seconds time slots. Nakayama et al, have mined 14 features from the RREQ, RERR, and RREP control packets. Actually, we have used these features in our feature set. It is of great importance to monitor and use control packets in order to detect the attacks. A lot of attacks including RREQ flooding, RERR flooding and isolation affect the traffic of control packets directly. Therefore, it is of great importance to monitor these types of features. However, by analysis of some attacks like tunneling, wormhole, and rushing, we come up with this idea that traffic features cannot provide us profound guarantee to detect all kinds of attacks although they are necessary.

In this work, with review of literature, we have selected the best features that can explain changes in the routing table properly: To Identify the wormhole attacks the following features are chosen,

a) Route change in percentage (RCP)
b) Hops Changes of all the routes (HCR).
c) Sequence number field changes
d) Maximum hop count field changes
e) Average sequence number
f) Average hop count
g) Packet drop

Route change in percentage (RCP) = (|P2 – P1| + |P1 − P2|)/|P1|. |P1| indicates the number of elements in P. (P2 – P1) means the newly increased routing entries during the time interval (t2 – t1), and (P1 – P2) means the deleted routing entries during (t2− t1). They together represent the changes of routing entries in (t2 – t1).

Hops Changes of all the routes (HCR) = (H2 – H1)/H1. (H2 – H1) indicates the changes of the sum of hops of all routing entries during the time interval (t2 – t1).

In addition, we have selected other features that monitor the routing table changes more accurate: The maximum sequence number field changes of entries of active routes in the routing table. The maximum hop count field changes of entries of active routes in the routing table. Average of differences between sequence number field of RREQ and RREP source node and sequence number field of routing table entrance packet for the node

Average of differences between hop count field of RREQ and RREP source node and hop number field of routing table entrance packet for the node.

5. SIMULATION RESULTS

In this section we present the simulation results and show how our collected features and algorithm help us to detect attacks more accurately. We have utilized the well-known Network Simulator version 2 (ns-2) for our simulation. In this simulation two Scenarios are considered:

5.1. Scenario: I Variation with Node

In the first subsection the importance of feature selection will be described in first scenario. In addition, we evaluate our collected real time network feature. We propose our evaluation on using sequential pattern mining and by means of that we have provided an unsupervised algorithm. Actually, we will provide a comparison between PCA and SPM and show the advantages of using SPM. The performance parameters like:

FP (False Positives): The number of normal events being predicted as attacks
FN (False Negatives): The number of attack events incorrectly predicted as attacks
TP (True Positives): The number of attack events correctly predicted as attack
Throughput, packet delivery ratio are computed by varying the no of nodes and the detection ratio is obtained. The variations in nodes are vary according to the Table 1.

Table 1. Variation with Nodes

Environment Size	1000*1000
Number of Node	50-90
Traffic Type	CBR
Mobility Model	RANDOM WAY MOBILITY
	25sec
Pause Time	AODV
Routing Protocol	200s
Simulation Time	2
No of attacker	

5.1.1. False positive

FP (False Positives): Refer to the number of normal events being predicted as attacks. The graph between nodes and false positive are shown in the Figure1 by using Table 2.

Table 2. False positive with Nodes

Nodes	PCA	SPM
50	4	4
60	6	3
70	6	3
80	7	4
90	10	5

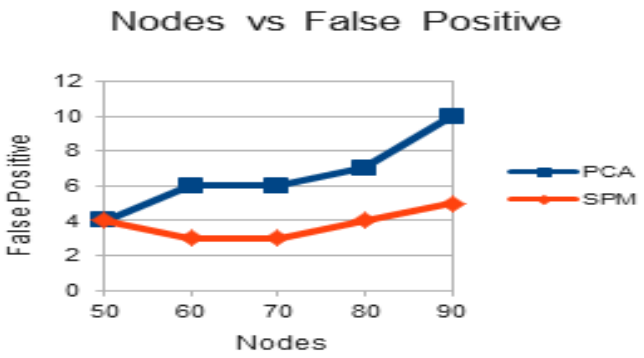

Figure 1. Nodes vs False Positive

5.1.2. False negative

FN (False Negatives): The number of attack events incorrectly predicted as attacks. The graph between nodes and false positive is shown in the Figure 2 by using Table 3.

Table 3. False Negative with Nodes

Nodes	PCA	SPM
50	2	1
60	2	1
70	1	0
80	0	0
90	0	0

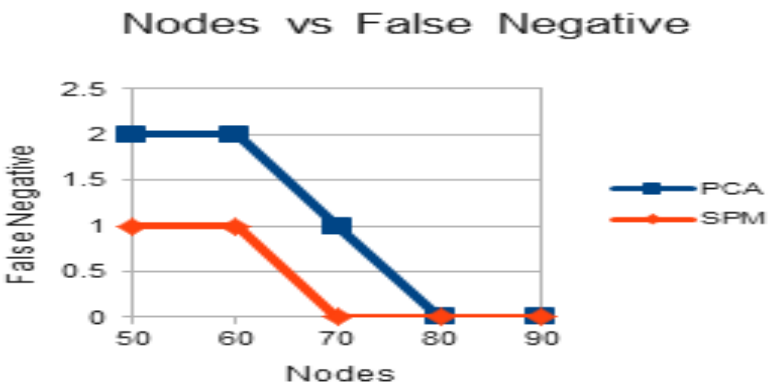

Figure 2. Nodes vs False Negative

5.1.3. Delay

A delay in network will identify the time delay of the data transferring from one to another. The graph between delay and the nodes is shown in the Figure 3 by using Table 4

Table 4. Delay with Nodes

Nodes	PCA	SPM
50	5.36251	4.3417
60	10.9751	6.63515
70	1.96633	0.939719
80	0.496987	0.394072
90	13.4948	6.83613

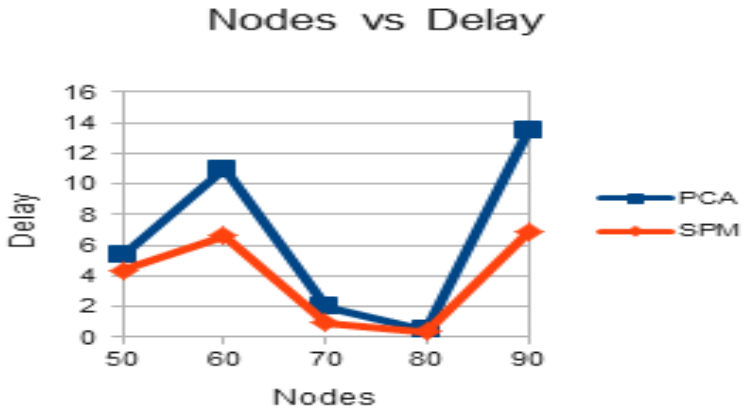

Figure 3. Nodes vs Delay

5.1.4. Detection Ratio

The graph between nodes and detection ratio is shown in the Figure 4 by using Table 5.

Table 5. Detection Ratio with Nodes

Nodes	PCA	SPM
50	0.66666667	0.83333333
60	0.66666667	0.83333333
70	0.83333333	1
80	1	1
90	1	1

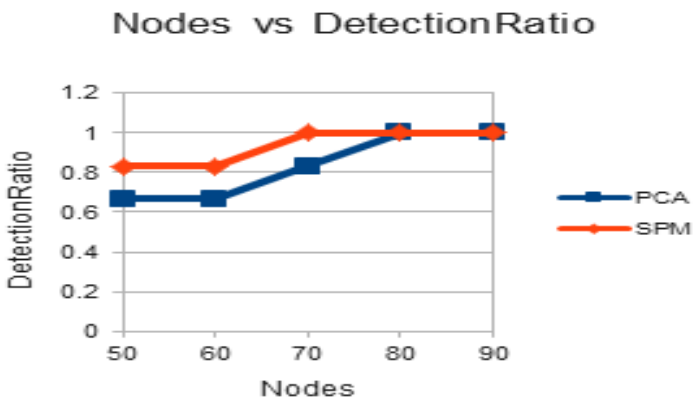

Figure 4. Nodes vs Detection Ratio

5.1.5. Packet Delivery Ratio

The data packets delivered ratio in PCA and SPM is shown in the Figure 5 by using Table 6

Table 6. Packet delivery ratio with Nodes

Nodes	PCA	SPM
50	71.8009	87.225
60	69.1213	89.1213
70	72.5629	77.9558
80	87.6819	93.2494
90	80.1884	82.8116

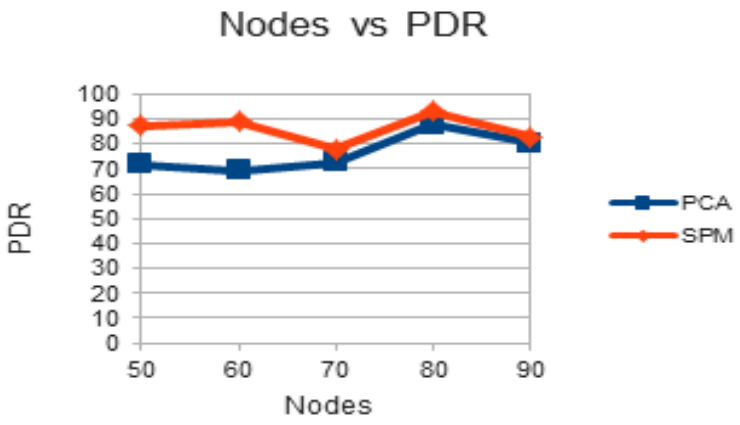

Figure 5. Nodes vs PDR

5.1.6. Throughput

The throughput is the amount of data moved from one place to another in the given time period is shown in the Figure 6 by using Table 7.

Table 7. Throughput with Nodes

Nodes	PCA	SPM
50	172896	159045
60	77373.1	97432.8
70	244060	289791
80	291940	302687
90	101254	142597

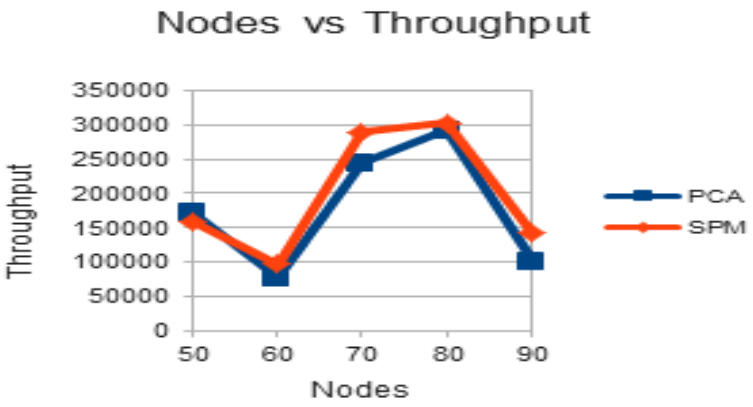

Figure 6: Nodes vs Throughput

5.2. Scenario: II VARIATION WITH ATTACKER NODE

Actually, we will provide a comparison between PCA and SPM and show the advantages of using SPM. The performance parameters like

FALSE POSITIVE: Indicates the number of normal events successfully labeled as normal.

FN (False Negatives): The number of attack events incorrectly predicted as normal.

Throughput, packet delivery ratio is computed by varying the attacker nodes and the detection ratio is obtained by using the Table8.

Table 8: Variation with attacker node

Environment Size	1000*1000	
Number of Node	90	
Traffic Type	CBR	
Mobility Model	RANDOM MOBILITY	WAY
Pause Time	25sec	
Routing Protocol	AODV	
Simulation Time	200s	
No of attacker	2-10	

5.2.1. False positive

FALSE POSITIVE: Indicates the number of normal events successfully labeled as normal. The False positive vs attacker is shown in the Figure 7 by using Table 9.

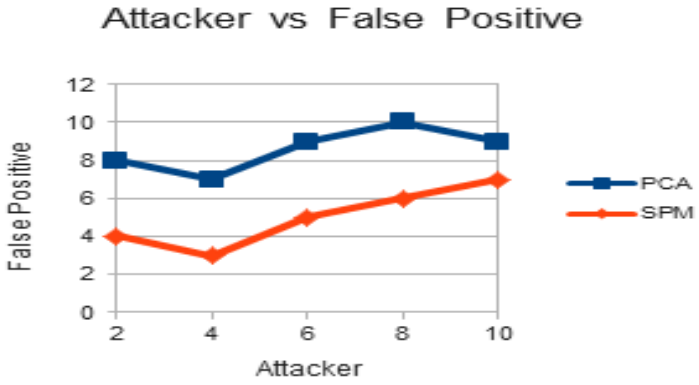

Figure 7. Attacker with false positive

Table 9. False positive with attacker node

Attackers	PCA	SPM
2	8	4
4	7	3
6	9	5
8	10	6
10	9	7

5.2.2. False negative

FN (False Negatives): The number of attack events incorrectly predicted as normal. The attacker vs False negative is shown in the Figure 8 by using Table 10.

Table 10. False negative with attacker node

Attackers	PCA	SPM
2	1	0
4	2	2
6	2	0
8	2	1
10	3	1

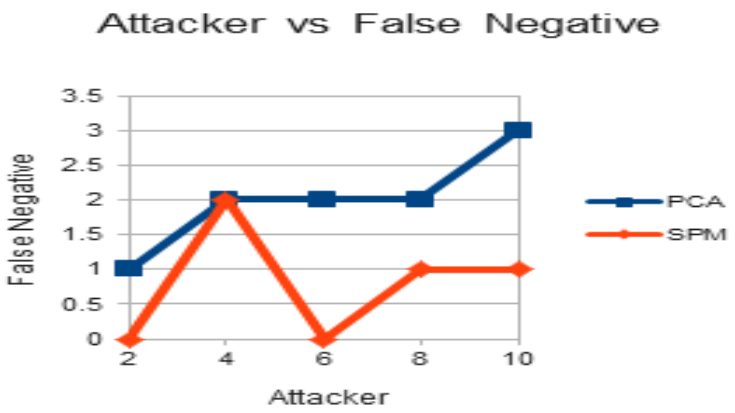

Figure 8. Attacker vs False Negative

5.2.3. Detection Ratio

The detection ratio with attacker is shown in the Figure 9 by using Table 11.

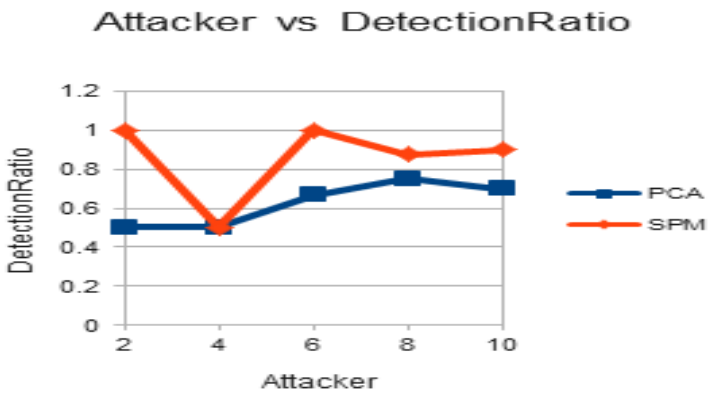

Figure 9. Attacker vs Detection ratio

Table 11. Detection Ratio with attacker node

Attackers	PCA	SPM
2	0.5	1
4	0.5	0.5
6	0.66666667	1
8	0.75	0.875
10	0.7	0.9

5.2.4. Packet Delivery Ratio

The packet delivery ratio with attacker is shown in the Figure 10 by using Table 12.

Table 12. PDR with Attacker node

Attackers	PCA	SPM
2	30.0253	25.7983
4	35.8603	29.8874
6	27.303	23.3747
8	39.1224	19.102
10	38.64	32.345

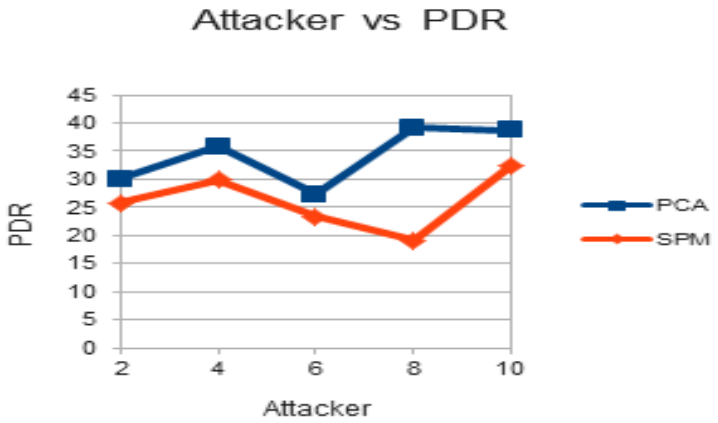

Figure 10. Attacker vs PDR

5.2.5. Throughput

The throughput with attacker for PCM and SPM is shown in the Figure 11 by using Table 13.

Table 13 Throughput with Attacker node

Attackers	PCA	SPM
2	218147	301514
4	141107	187605
6	304181	298723
8	82711.9	153944
10	156915	182045

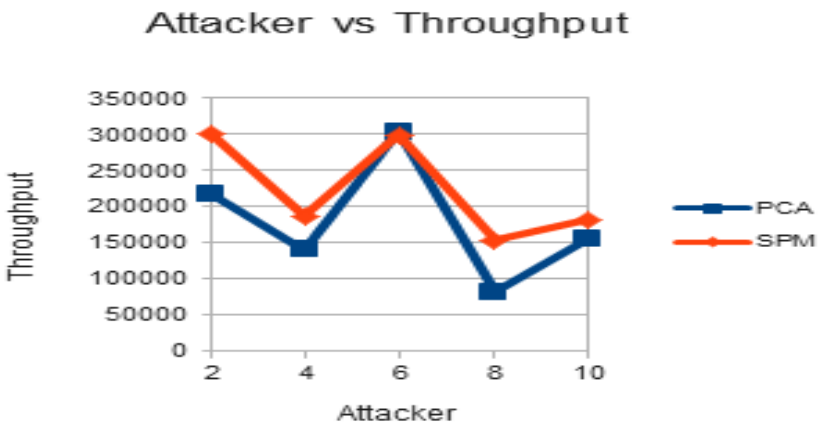

Figure 11. Attacker vs throughput

5.2.6. Reduction Ratio

The reduction ratio with attacker for PCM and SPM is shown in the Figure 12 by using Table 14

Table 14. Reduction ratio with attacker node

Attackers	PCA	SPM
2	0.177852	0.182748
4	0.215495	0.155613
6	0.203744	0.183589
8	0.201198	0.184252
10	0.196881	0.138408

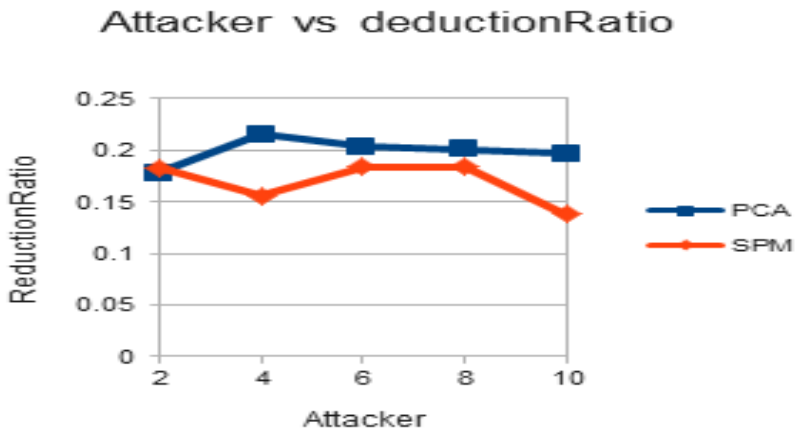

Figure 12. Attacker vs reduction ratio

6. CONCLUSION AND FUTURE WORK

In this paper, we have contemplated the security of MANETs by method for interruption location shield. The significance of selecting nice components for interruption identification frameworks has been clarified in points of interest. We utilized ns-2 to reenact our situations. The outcomes demonstrated our components can distinguish a great deal more assaults either by applying PCA or by applying SPM. By means SPM we could have an unsupervised calculation that distinguishes peculiarity more exact. Really, SPM can shape the benchmark profile even by presence of vindictive hubs in the learning stage. Moreover, we plan to propose a plan with a specific end goal to recognize and find the foe in a MANET. This will be accounted for in a future work.

REFERENCES

[1] Senthilnayaki, Balakrishnan, Venkatalakshmi K, Kannan A, "Intrusion Detection System Using Feature Selection and Classification Technique". *International Journal of Computer Science and Application* (*IJCSA*) Volume 3 Issue 4, November 2014.

[2] Ayman I. Madboul, Amr M. Gody, Tamer M. Barakat, "Relevant Feature Selection Model using Data Mining for Intrusion Detection System", *International Journal of Engineering Trends and Technology* (*IJETT*) – Volume 9 Number 1.

[3] Dr. Saurabh Mukherjeea, Neelam Sharmaa, "Intrusion Detection using Naive Bayes Classifier with Feature Reduction Procedia Technology", 4 (2012) 119 – 1280 - Mar 2014.

[4] Vetrichelvi Raja, ramVanitha, Dr.G.Mohankumar, "Feature Analysis for Intrusion Detection in Mobile Ad-hoc Networks", *IJCSNS International Journal of Computer Science and Network Security*, Vol. 10 No. 9, September 2010.

[5] Fang Lan A, *"Framework for Network Security Situation Awareness Based on Knowledge Discovery"*, 2nd International Conference on Computer Engineering and Technology 2010.

[6] Mohammad K. Houri Zarch, "An Unsupervised Anomaly Detection Engine with an Efficient Feature set for AODV 2010".

[7] Srilatha Chebrolua, Ajith Abrahama B, Johnson P. Thomasa, "Feature Deduction and Ensembledesign of Intrusion Detection Systems", *Computers & Security (2004)*

[8] M. Alikhani, M. Ahmadi Livani, and M. Abadi, "Dynamic Anomaly Detection by using Incremental Approximate PCA in AODV -based MANETs," *Journal of AI and Data MiningJournal of AI and Data Mining*, vol. 1, 2013.

[9] S. McCanne, S. Floyd, K. Fall, and K. Varadhan, "Network Simulator ns-2," ed, 1997.

[10] F. Barani and M. Abadi, *"An ABC-AIS Hybrid Approach to Dynamic Anomaly Detection in AODV-Based MANETs,"* in Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on, 2011, pp. 714-720.

[11] S. Verboven and M. Hubert, "LIBRA: a MATLAB Library for Robust Analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, pp. 127-136, 2005.

[12] Wang Ruilian and Shengjian, "Comprehensive Evaluation to Distribution Network Planning Schemes Using Principal Component Analysis Method", *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol 12 No 8, 2014, pp. 5897-5904.

[13] Alok Mukherjee, Arabinda Das, "Identification and Classification of Power System Faults using Ratio Analysis of Principal Component Distances", *TELKOMNIKA Indonesian Journal of Electrical Engineering* Vol 12 No 11, 2014, pp. 7603-7612.

BIOGRAPHIES OF AUTHORS

Mr. Reji. M received his post graduate degree from College of Engineering, Guindy, India. Currently pursuing Ph.D on Intrusion Detection in Wireless Sensor Networks in Saveetha University and working as an Assistant Professor in Saveetha School of Engineering, Saveetha University.

Dr. Kishore Raja P.C, received the doctorate degree in the field of wireless security from Anna University, Chennai His area of interest includes Wireless security, intrusion detection systems and IOT. Currently, he is working as Professor and the Head of the Department of Electronics & Communication Engineering at Saveetha School Of Engineering, Saveetha University.

Bhagyalakshmi M is currently pursuing B.E. Electronics and Communication Egineering in Saveetha University. Area of interest includes adhoc networks and security