DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5423

ISSN No. 0976-5697

Volume 9, No. 1, January-February 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

IMPROVED PREDICTION STRATEGY USING PARTICLE SWARM OPTIMIZATION BASED ARTIFICIAL NEURAL NETWORK (IPS-ANN) CLASSIFIER FOR MALICIOUS NODE DETECTION FOR HYBRID P2P NETWORKS

E.Banumathi
Department of Computer Science
Dr.NGP Arts & Science College
Coimbatore,India

Dr.V.Jaiganesh
Professor,PG & Research Department
Department of Computer Science
Dr.NGP Arts & Science College
Coimbatore,India

Abstract:Mining data from the hybrid P2P networks is an ever demanding task for detecting malicious behavior of nodes. This research work aims to propose an improved prediction strategy using particle swarm optimization based artificial neural network (IPS-ANN) classifier. From the extensive literature study it is identified that only very few literatures are there for detecting malicious node activities. The simulation are carried out using MATLAB by configuring 1000 nodes with three different attack scenarios namely collusion attack, Sybil attacks and file polluter attacks. Detection accuracy, false positive rate and false negative rate are taken as the performance metrics for evaluating the efficiency of the proposed classifier and also compared with the existing mechanisms. Results proved that the proposed IPS-ANN classifier better than that of PeerMate , SMART , Outlier mining mechanisms.

Keywords: Data mining, Hybrid P2P networks, Malicious node, classifier, detection, MATLAB, Artificial neural network, particle swarm optimization.

I. INTRODUCTION

The Peer-to-Peer (P2P) systems are able to rapidly supply information exchange and sharing service by utilizing resources from participating peers, so have of late gained noteworthy recognition [2]. However, the open and anonymous nature of these systems often leads to a lack of responsibility for the content a peer puts on the network, opening the door to abuses by malicious peers [2]. Consequently, a major challenge for a large-scale P2P system is to establish a scalable trust framework without any pre-existing trust relationship among peers and without involving trusted third parties or authorities In a hybrid P2P network, all the nodes can be broadly grouped as super node and ordinary node [3]. Under each super node, there are several ordinary nodes with which the super node forms a subnet.

Each super node stores its neighbor super nodes' list [4], so as to ease the communications between subnets. In a subnet, the super node is responsible for managing the interaction data among the ordinary peers. Meanwhile, all the super nodes are in charge of managing the interaction data among the subnets. Based on the finding that each malicious peer has the specific characteristic of outlier [1] , this research work aims to present an improved prediction strategy using particle swarm optimization based artificial neural network (IPS-ANN) classifier for malicious node detection for hybrid P2P networks.

II. LITERATURE REVIEW

Xiaoliang et al. explained the importance of authentication in P2P networks [5]. These authors stated that anonymity searching allows attackers into P2P networks. To eliminate these attackers, the network requires trust and reliable authentication. Stefan Kraxberger et al. proposed trust transaction for the online service provision [6, 7]. The provision stipulates that before making a purchase decision, the buyer should check the trust value of the product from third parties, based on the feed back of their trust value. Alternatively, the buyer can purchase the item with the help of a reputation system. Jia Zhang proposed the establishment of trust based P2P network [8].

Accordingly, Jia highlighted that P2P anonymous communication needs the trust method. This is because the anonymous tunneling may not provide higher anonymity, but may provide lower performance. Takeda et al. proposed HDAM authentication method [9] Currently, P2P network suffers more from authentication, and it is for this reason that she has used HDAM method to increase reliability in P2P networks. Cheng et al. elucidated that, in the P2P framework, the service provider should authenticate the requested node based on the communication history. Cheng et al. also noted that the service provider should evaluate the trust value and finally establish if the node is trustworthy or not [10]. Gupta et al. developed a Reputation Aggregation method based on different gossip algorithms [11]. This technique is used to authenticate every node in the P2P network. The authentication is done with the help of reputation value obtained from every node.

Lu et al. explained that a Zero Knowledge-based authentication in P2P system identifies the fake trust (FT) [12]. In this procedure, a peer has never utilized its original ID. Thus, it creates a false name built in the one-way hash function. To empower the verification of the associates to guarantee complete security of the touchy data, another validation strategy is created on the premise of the Zero-

© 2015-19, IJARCS All Rights Reserved

Knowledge Proof. Dannewitz et al. proposed an Information Centric networking method [13] . The aim of this procedure is to develop an authenticated network infrastructure service for today's users. Bartram et al. proposed PGP (Pretty Good Privacy) [14]. The author asserts that PGP is an authentication and reliable information exchange in P2P networks.

For the PGP, the trusted database is used to authenticate every other node in the network. Additionally, Jayaraj et al. proposed an Efficient, RR based Password-Authentication key protocol [15]. This protocol is used to authenticate the peers in RR protocol. The nodes can send the information when they know the password of particular nodes since the password information is stored in a trusted database. Another study by Wright et al. investigated the attackers by the damaged group members that disgrace the anonymity of all protocols over time [16] . Based on this analysis, they have proved that when an exacting initiator continues to communicate with an exacting responder, the attacker can easily identify the source and responder nodes. Hence, the attackers can take the valid information from those nodes and consequently degrade the network performance.

Bi et al. examined the idea of employing split manufacturing method in Radio Frequency (RF) circuit protection and also developed qualitative security evolution method [17–19]. The exploration conducted by Bi et al. revealed that Intellectual Property (IP), piracy, and hardware Trojans are becoming the main hardware security threats. They have designed three sample circuit structures namely; camouflaging gates, power regulators and polymorphic gates. They have also proved the high-efficiency IP piracy deterrence and circuit protection with the help of the designed circuit structures.

It is noteworthy to mention a very recent and relevant work by Wahab et al., [23] in which authors have proposed a SVM based distributed classification framework in clustered Vehicular Networks (VANETS) for the detection of malicious nodes in a cooperative manner. Selected monitoring vehicles exchange their observations about the credibility of inter-cluster relay nodes and use these observations to learn a SVM classifier in an incremental and online fashion. Final class labels are propagated among cluster heads to take required action.

III. PROPOSED WORK

A. *Overview of ANN*

In the ANN prediction model, a multilayer feed forward network with one hidden layer is used. The input layer has n nodes, the hidden layer has H nodes and the output layer has O nodes. The transfer function for the hidden node is the sigmoid function, and the output transfer function is a linear activation function. The output of the jth hidden node is computed as follows:

$$(y_j) = 1/(1+\exp(-(\sum_{i=1}^{n} x_i - \theta_j))), \text{ j= 1, 2…, H} \qquad (1)$$

Where w_{ji} is the weights that connect the ith input node to the jth hidden node, θ_j is the threshold of the hidden layer, x_i is the ith input, and y_j is the output of the hidden layer. The output from the kth output layer is computed as follows:

$$z_k = \sum_{j=1}^{H} (y_j) \text{ k } = 1,2,\ldots\ldots,O \qquad (2)$$

Where w_{kj} is the weights that connect the j^{th} hidden node to the kth output node. The maximum number of hidden nodes in the network is computed using $(2n+1)$, where n is the number of input nodes in the network, which corresponds to the features of the dataset. The accuracy based on the learning error E is computed as follows:

$$E_k = \sum_{i=1}^{O} (Z_1^k - C_i^K)^2 \qquad (3)$$

Where z_i^K is the obtained output from the network and C_i^K is the target output. E is the difference between the target output and the obtained output. The fitness value of the training sample is computed as follows:

$$Fitness\ (X_i) = E_k \qquad (4)$$

The gradient error in the network with respect to the weight increment and weight update is computed using equation (5) and (6).

$$\Delta w_{ji} = (z_i - c_i) \qquad (5)$$

$$\Delta w_{jinew} = w_{jiold} + \Delta w_j \qquad (6)$$

Where $\Delta w_{ji}$is the change in weight that connect the hidden node and the input node. The bias (θ_j) in the network is incremented and updated using equation (7) and (8).

$$\Delta\theta_j = (z_i - c_i) \qquad (7)$$

$$\theta_j = \theta_j + \Delta \qquad (8)$$

Where $\Delta\theta_j$ is the change in bias and η is the learning rate.

B. *Particle Swarm Optimization*

In PSO, particles placed at random positions in the search space is d- dimensional and the particle i of the swarm can be represented by a d-dimensional position vector $X_i = (X_{i1} + X_{i2} + \ldots\ldots, X_{iD})$. The velocity of the particle can be represented as $V_i = (V_{i1} + V_{i2} + \ldots\ldots,)$. During the search process, the position of a particle is guided by two factors. That is, local best $(P_{i,})$ and global best (P_{gbest}). The best visited position for the particle by itself is $P_{i,} = (P_{i1,2,}\ldots\ldots, P_{iD})$ The $P_{i,}$ of the particle can be updated in the next generation as given inequation (9).

$$P_{i,,t}(t+1) = \begin{cases} P_{i,best} & if \quad f(X_i(t+1)) > f(P_{i,best}) \\ x_i(t+1) & else \end{cases} \qquad (9)$$

Where NP is the size of the swarm. The position of the best particle in the swarm is denoted by $P_{gbest} = (P_{g1,2,}\ldots\ldots, P_{gD})$, and is computed using equation (10)

$$P_{gbest} = (P_{i,\text{best}}) \ldots \qquad (10)$$

For each generation the position of the particle and its velocity in a PSO can be updated using the following equations:

$$(t+1) = \omega.V(t) + C_1\varphi_1(P_{i,best} - X_i) + C_2\varphi_2(P_{gbest} - X_i) \qquad (11)$$

$$(t+1) = X(t) + V_i(t+1) \qquad (12)$$

Where $V\ (t+1)$ is the velocity of the particle. $Xi(t+1)$ is the position of the particle, C_1 and C_2 are the cognitive and social learning parameter, ω is the inertia weight, $\varphi_1$and$\varphi_2$are random numbers uniformly distributed in [0,1], $i = 1,2,\ldots,NP$.

© 2015-19, IJARCS All Rights Reserved

The linearly decreasing inertia weight is computed using equation (13).

$$\omega = \omega_{max} - (\omega_{max} - \omega_{min})\left(\frac{iter}{iter_{\max}}\right) \quad (13)$$

Where $iter$ denotes the current iteration number and $iter_{max}$is the maximum number of iterations. The value of ω is between 0.9 to 0.3. Larger inertia weight is assigned to the particle during the initial search which gradually reduces as the search proceeds in further iterations.

In this research work, the best mutation operation is computed using the linearly non- increasing weight values instead of random weights. The linearly non-increasing weight values improves exploration in search space. The modified best mutation strategy is computed using equation (14).

$$Y_{i,} = X_{best1,\text{gen}} + F(X_{best2,gen} - X_{best3,gen}) \quad (14)$$

where F is a scaling factor which is used in controlling the amplification of the differential variation $\in$ [0,1]. X_{best1}, X_{best2}, X_{best3} are linearly non-increasing particles in the population.

$$Norm\text{ali}(X) = \frac{\text{V - E}_{\min}}{\text{E}_{\max} - E_{\min}} (E_{new\_max} - E_{new\_min}) + E_{new\_min} \quad (15)$$

where E_{min} and $E_{max}$is the minimum and maximum values of an attribute, A. The normalized $E_{new\_max,\_min}$ within the range of [0,1].

C. *Improved Prediction strategy using Particle Swarm Optimization based ANN Classifier for Malicious Node Detection Model for Hybrid p2p Networks*

The IPS-ANN algorithm is proposed to train the ANN classifier to improve the prediction accuracy. The given Q number of training samples with n dimensional input patterns are mapped onto the corresponding target output z^k. The input patterns are represented as $x^k = \{x_i^k = 1,2\ldots n\}$. The objective is to find the function f with a global optimum $Gbest$ at a faster convergence rate. The total number of nodes in the hidden layer H is computed as follows:

$$H = (2n + 1) \quad (16)$$

In the IPS-ANN algorithm, the NN weights are considered as particles. The weights from w_{ji} and w_{kj} and two biases from input layer and hidden layer take part in the IPS-ANN process. The group of particles is called a swarm. The size of the swam $NP \in$ [0,1] is computed using (17):

$$NP = (n + 1) * H + (H + 1) * O \forall i = 1,2\ldots NP \quad (17)$$

where n is the total number of inputs. H is the total number of hidden nodes. O is the total number of output nodes.

Step 1: Initialize the parameters learning rate (η), maximum generations (Gen_{max}), minimum error (ε), swarm size (NP), mutation rate (F),inertia weight (ω_{max},ω_{min}),cognitive and social learning parameter(C_1,C_2),maximum velocity(V_{max}). The weights are initialized randomly to a value between 0 and 1.

Step 2: The any training pattern from the normalized dataset with n number of features is applied to the input layer X, whose size is equal to n. Then the data from each input (x_i, = 1,2…n) node of that pattern propagates to the hidden layer.

Step 3: Linearly non increasing weights are assigned on the link connecting input layer to the hidden layer (w_{ji}) and the hidden to the output layer(w_{kj}). The weights are obtained from the modified best mutation operation.

Step 4: The feed forward operation is performed using equation (1-4) to find the local best (l_{best}) values. Among the local best values $\{l_{i,} = 1,2\ldots Q\}$ the minimum fitness value is the global best G_{best} = min($l_{i,best}$). After finding the global and local best, for each generation the particles' position and velocity are updated.

Step 5: The gradient descent BP algorithm is applied to the global best, when it is greater than the fitness value. The network propagates back to change the weights (Δw_{ji},Δw_{kj}) and bias ($\Delta\theta_j$) to minimize the error. The change in weights and bias values are computed.

Step 6: Compute the fitness of the weight tuned, back propagated network. If the fitness value is less than 10, it terminates the process. Otherwise, the following steps are carried out.

Step7: The back propagated weight values and bias values $\{\Delta w_{ji1}, \Delta w_{ji2}\ldots\Delta w_{jin},\Delta w_{kj1},\Delta w_{kj2}\ldots\Delta w_{kjn},\Delta\theta_j \ldots NP\}$ are stored and used for the next generation process. The procedure presented from step 4 to step 6 is repeated.

Step8: If the global best value for the current iteration is greater than that of the previous iteration ($[Gen] < G[Gen + 1]$), then the training process is terminated as it may tend to overfit. Otherwise, the following step is carriedout.

Step 9: The above steps (3-7) belongs to the training procedure. Train the ANN till the global best value is reached. At the end of the training, the test set is used to test the generality of the IPS-ANN classifier.

IV. SIMULATION RESULTS

Simulations are carried out using MATLAB. Each simulation runs 10 times, and the average value is reported as the simulation result. Without the loss of generality, commonly used false positive rate (FPR, i.e. the ratio of peers that are normal but considered as malicious to all the normal peers) and false negative rate (FNR, i.e. the ratio of peers that are malicious but considered as normal to all the malicious peers) as the criterion [20, 21] to assess the performance of our model.True positive is the normal activity correctly identified as normal activity. False positive is the attack behavior incorrectly identified as normal activity. True negative is the attack behavior correctly identified as attack behavior. False negative is the normal activity incorrectly identified as attack behavior. The simulation has been performed with three different attack scenario in 500 nodes deployed for hybrid P2P network.

FPR = FP / (FP + TN)

FNR = FN / (FN + TP)

Attacks	Number of injected attacks
Collusion attacks	1846

© 2015-19, IJARCS All Rights Reserved

Sybil attacks	713
File polluter attacks	1264

Table1. Performance Analysis of False Positive Rate, False Negative Rate and Accuracy

	TP	TN	FP	FN	FPR	FNR	Detection Accuracy
Collusion attacks							
PeerMate [20]	1204	216	221	205	50.57	14.55	76.92
SMART [21]	1286	185	184	191	49.86	12.93	79.69
Outlier Mining [22]	1399	162	133	152	45.08	9.80	84.56
IPS-ANN	1445	178	102	121	36.43	7.73	87.92
Sybil attacks							
PeerMate [20]	337	213	91	72	29.93	17.60	77.14
SMART [21]	342	232	78	61	25.16	15.14	80.50
Outlier Mining [22]	361	242	58	52	19.33	12.59	84.57
IPS-ANN	379	244	41	49	14.39	11.45	87.38
File Polluter attacks							
PeerMate [20]	822	147	114	181	43.68	18.05	76.66
SMART [21]	854	171	98	141	36.43	14.17	81.09
Outlier Mining [22]	896	182	87	99	32.34	9.95	85.28
IPS-ANN	912	199	71	82	26.30	8.25	87.90

© 2015-19, IJARCS All Rights Reserved

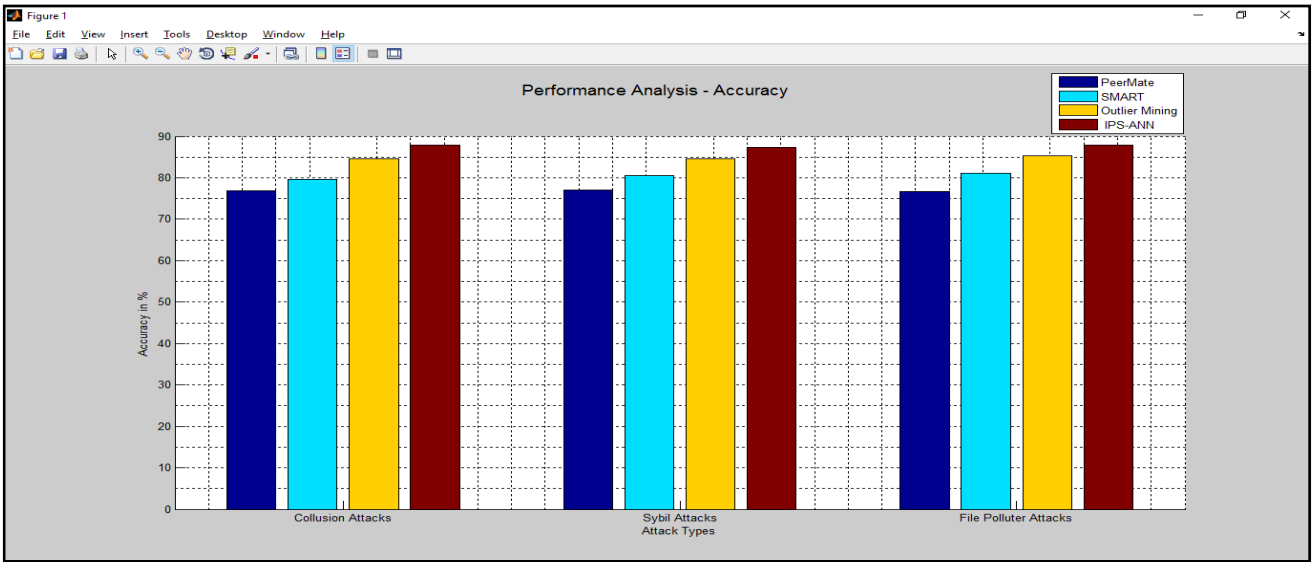

Figure1. Performance Analysis – Accuracy

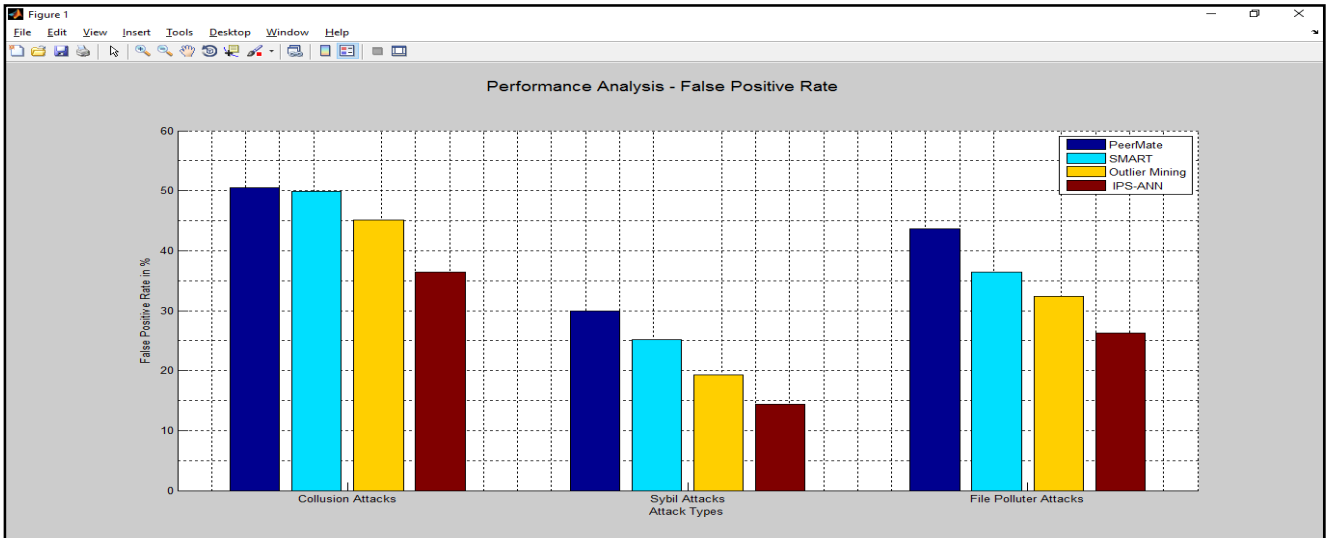

Figure 2. Performance Analysis – False Positive Rate

© 2015-19, IJARCS All Rights Reserved

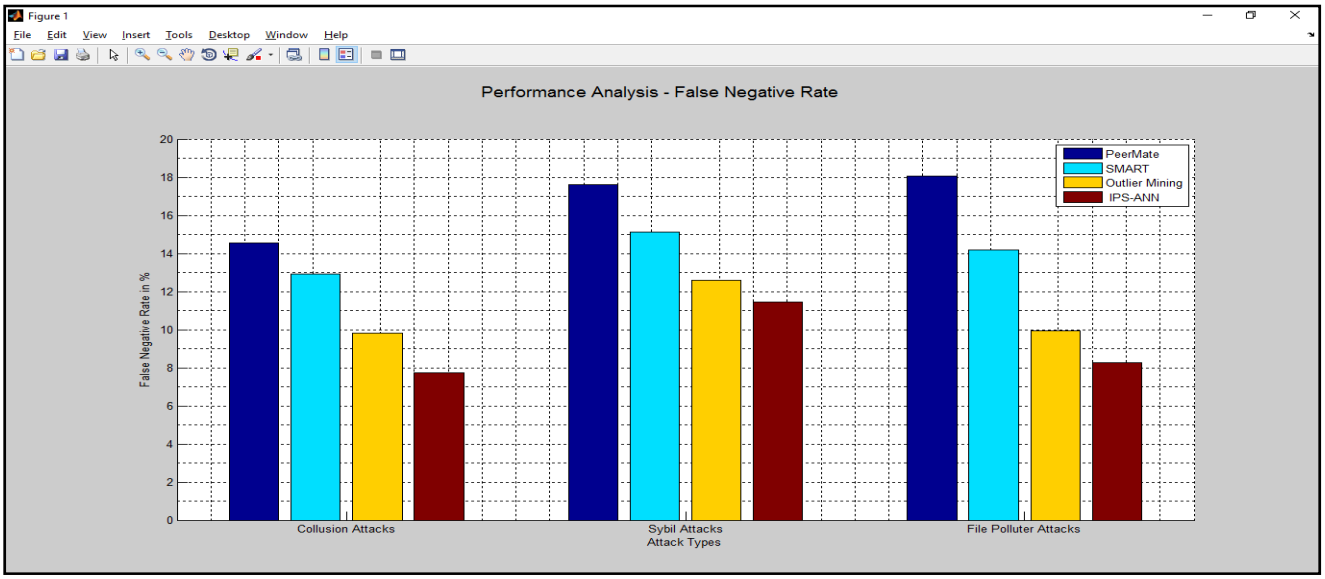

Figure 3. Performance Analysis – False Negative Rate

From the results it is evident that the proposed IPS-ANN classifier correctly detected collusion attacks, Sybil attacks and file polluter attacks with the accuracy 87.92%, 87.38% and 87.90% respectively. Also the false positive ratio is reduced upto 36.43%, 14.39% and 26.30% for the collusion attacks, Sybil attacks and file polluter attacks respectively. In addition to that the false negative ratio is reduced to 7.73%, 11.45% and 8.25% respectively. The results showed that the proposed IPS-ANN classifier outperforms than that of the existing mechanisms namely PeerMate [20], SMART [21], Outlier mining [22] mechanisms. The matlab results are also presented for the same in Fig.1, Fig.2 and Fig.3.

V. CONCLUSIONS

Data mining nowadays expanded its scope to discover hidden information / knowledge from the network datasets. In real time scenario like hybrid P2P networks the intrusion is prevailing as a common one. This research aim to propose a classifier that would detect the several attacks such as collusion attacks, Sybil attacks and file polluter attacks. Hence an improved prediction strategy using particle swarm optimization based artificial neural network (IPS-ANN) classifier for malicious node detection for hybrid P2P Networks. Simulations are carried out using MATLAB and the obtained results prove that the proposed IPS-ANN performs better detection in terms accuracy, false positive ratio and false negative ratio.

VI. REFERENCES

[1] Q. Lian , Z. Zhang , M. Yang , et al. , An empirical study of collusion behavior in the maze P2P file-sharing system, in: ICDCS 27th International Conference on Distributed Computing Systems, Toronto, Canada, IEEE, 2007, pp. 1–10.

[2] J. Li, mSSL: a framework for trusted and incentivized peer-to-peer data sharing between distrusted and selfish clients, Peer-to-Peer Netw. Appl. 4 (2011) 325–345.

[3] C. TIan, S. Zou, D. Wang, S. Cheng, A new trust model based on recommendation evidence for P2P networks, Chin. J. Comput. 31 (2) (2008) 270–281.

[4] Xu Ke, Shen Meng, Ye Mingjiang, A model approach to estimate peer-to-peer traffic matrices, in: Proceedings of IEEE INFOCOM, 2011.

[5] Wang X , Yang L , Sun X , Han J , Liang W , Huang L . Survey of anonymity and authentication in p2p networks. In: Proceedings of the IH, Inf Technol J 2010;9(6):1165–71.

[6] Kraxberger S , Martin Pirker R , Guijarro EP , Garcia Millan G . Trusted identity management for overlay networks. Lecture Notes in Computer Science, vol 7863. Springer publication; 2013. p.16–30.

[7] Audun J, Roslan I, Colin AB. A survey of trust and reputation systems for online service provision. Decision Support Systems 2007;43(2):618–44.

[8] Zhang J , Haixin D , Wu L , Jianping W . Anonymity analysis of P2P anonymous communication systems. Comput Commun 2011;34(3):358–66 .

[9] Takeda A , Chakraborty D , Kitagata G , Hashimoto K , Shiratori N . A new scalable distributed authentication for p2p network and its performance evaluation. 12th WSEAS international conference on computers, Greece,.2008 Oct; 2008. 7(10;1628-37).

[10] Cheng W , Zhen HT . Correlation trust authentication model for peer-to-peer networks. In: Advanced materials research, vol. 756. Trans Tech Publications; 2013. p. 2237–42.

[11] Gupta R , Singh YN . Reputation aggregation in peer-to-peer network using differential gossip algorithm. IEEE Trans Knowl Data Eng 2015;27(10):2812–23 .

[12] Lu L , Han J , Hu L , Huai J , Liu Y , Lionel M N . Pseudo trust: zero-knowledge based authentication in anonymous peer-to-peer protocols. In: 2007 IEEE International Parallel and Distributed Processing Symposium. IEEE; 2007. p. 1–10.

[13] Dannewitz C , Kutscher D , Ohlman B , Farrell S , Ahlgren B , Karl H . Network of information (NetInf)-an information-centric networkin architecture. Comput Commun 2013;36(7):721–35 .

[14] Linda Ruth B, Chesser MM, Sawadsky N, Schumacher SJ, Blackstock M. Peer-to-to Peer authentication for real time

© 2015-19, IJARCS All Rights Reserved

collaboration, 2008, U.S. Patent 7,392,375, issued June 24, 2008.

[15] Jayaraj V , Sharmila R . An efficient RR based pasword-authentication key agreement protocol. Wireless Commun 2016;8(8):309–11.

[16] Wright MK , Adler R , Levine BN , Shield sC . The predecessor attack: an analysis of a threat to anonymous communications systems. ACM Trans Inf Syst Security 2004;7(4):489–522 .

[17] Bi Y , Jiann-Shiun Y , Yier J . Split manufacturing in radio-frequency designs. In: Proceedings of the International Conference on Security and Management (SAM). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp); 2015.

[18] Bi Y , Jiann S , Yuan YJ . Beyond the interconnections: split manufacturing in RF designs. Electronics 2015;4(3):541–64 .

[19] Bi Y , Gaillardon P-E , Hu XS , Niemier M , Yuan J-S , Jin Y . Leveraging emerging technology for hardware security-case study on silicon nanowire fets and graphene symfets. In: 2014 IEEE 23rd Asian test symposium. IEEE; 2014. p. 342–7.

[20] X.L. Wei , T. Ahmed , M. Chen , et al. , PeerMate: a malicious peer detection algorithm for P2P systems based on MSPCA, in: ICNC International Conference on Computing, Networking and Communications, HI, IEEE Commun, 2012, pp. 815–819 .

[21] X.L. Wei , J.H. Fan , M. Chen , et al. , SMART: a subspace based malicious peers detection algorithm for P2P systems, Int. J. Commun. Netw. Inf. Secur. 5 (1) (2013) 1–8.

[22] X. Meng, S. Ren, An outlier mining-based malicious node detection model for hybrid P2P networks, Computer Networks, 108, (2016), 29-39.

[23] Wahab, O. A., Mourad, A., Otrok, H., & Bentahar, J. (2016). Ceap: Svm-based intelligent detection model for clustered vehicular ad hoc networks. Expert Systems with Applications, 50 , 40–54.

© 2015-19, IJARCS All Rights Reserved