IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.

www.ijiset.com

ISSN 2348 – 7968

Mitigating Sybil Attacks by Using Sybil Based Defense Mechanism In Large Social Networks

S.Krishnaveni[1],
Research Scholar, Department of Computer Science, Hindusthan College of Arts and Science, Bharathiar University,Cbe-28.

A.V.Senthil Kumar[2]
Director, Department of MCA, Hindusthan College of Arts and Science, Bharathiar University,Cbe- 28.

Mohamed Adnane Mellah[3], Abdelmalek Amine[4], Reda Mohamed Hamou[5]
GeCoDe Laboratory, Department of Computer Science, Tahar Moulay University of Saida,Saida,Algeria.

ABSTRACT

Due to the improvement of the technologies the usage of the social networks in nowadays also increase .It becomes important to maintain the distributed systems in several types of communication in the social network and perform their action in online social networks .Due to the lacking of the communication the centralized authority becomes one of the important issues in the distributed system ,the several number attacks have been performed in the distributed system one of the major important attack is Sybil attack .In this attack the fake information are created by attacker and try to find the information of the user . In order to solve these problem in earlier work proposes a Sybil Defender, it efficiently identification of the Sybil nodes from Sybil community in online social networks. The major problem of this existing work it becomes not capable to differentiate Sybil users from valid ones in the small non-Sybil regions .It becomes major important problem in the existing work, in order to solve this problem in this work proposed a novel methods called Sybil Shield. The proposed methods efficiently detects the sybil user from both sybil and non sybil community in the social networks for distributed system and a it reduces the falsely detected sybil nodes in the sybil and non sybil community system ,improved performance results when compared to the existing sybil defend methods.

Keywords: Distributed systems, Sybil Shield, Multi-community Social network, Non-Sybil community, Sybil community, Sybil attacks and non Sybil attacks

1. INTRODUCTION

Due to the improvement of the technologies the usage of the social networks in nowadays also increase .It becomes important to maintain the Distributed systems in several types of communication in the social network, but this category of the distributes system are easily damaged by sybil attacks [1], in which an challenger generate numerous fake identities, and cooperation the consecutively of the system .Some of the other types of the attack also occurred in the social networks [2-3].

Recently, the identification of the sybil nodes in the social networks [4-5] becomes one of the interesting research area in the distributed system . Since in the online social network community the different user among the various places share their personal information to maintain a relationship among each them in efficient manner .To maintain the relationship among them it becomes important to keep their personal information in secure manner also. To perform this process the identification of the sybil attacks plays major important. The social networks links between one user to another user in the networks have been created the network in the form of graphical representation.
In order to decrease the number of attacks in the social networks all of the existing methods assume that the relationship among the user can be maintained by the creation of the trust value among them ,but the trust value

calculation is not applicable to all of the areas in the online social networks [6] ,it is applied to most frequently visited web pages or online social networks [7]. It provides good performance but it is computationally exhaustive and cannot balance to networks through millions of nodes. In earlier work several number of the works have been performed based on the network topologies methods such as SybilGuard [8] and SybilLimit [7].Both of these methods detect Sybil node based on the suspected node results in network topologies only non sybil nodes are not supported in these works but their inter-connections can be multi-hop.

If the size of the social network becomes large the above mentioned work are not supported since these methods are computationally expensive and less probability of the identification of the sybil nodes in the sybil community and non sybil community regions are not focused in this work ,in order to solve this issue and support non community region identification for sybil attack ,in this paper proposed system solves and capability to differentiate Sybil users from valid ones in the small non-Sybil regions. Sybil attacks from non Sybil region are easily identified based on the selection of the verification node and the agent node in the distributed system in the earlier works. As a result the sybil and non sybil user are exactly identified in both sybil and non sybil regions in the distributed system for online social networks .,it also solves the differentiation problems of the Sybils from non-Sybils successfully.

2. BACKGROUND STUDY

In recent One of the interesting way to detect the Sybil attacks in social networks is based on the usage of the network topologies .In earlier work several number of the works have been performed based on the network topologies methods such as SybilGuard [8] and SybilLimit [7].Both of these methods detect Sybil node based on the suspected node results in network topologies .In order to identify Sybil nodes in the social network topologies it randomly select the route to search sybil nodes and calculate precompute permutation methods for each nodes in the randomly selected node ,but it becomes finds sybil node for selected random route only instead of consideration of the entire network topologies and insider attacks are not founded in the topologies

It suffers from identification of the false nodes in both positive and negative manner for each nodes in the network and becomes more time complexity this problem is overcome in the sybilimit [7] methods it also having some major disadvantages .The problem of these methods is solved in [9] .GateKeeper [9] is one of the another type of the Sybil defender mechanism that largely depends on the statement of the random expander based methods for social networks. This is a well-built theory which has not been authenticating through earlier study. It suffers from high false positive and negative rates may not successfully discover Sybil nodes for real time online social networks.

Sybil identification mechanism also supported for centralized manner also in earlier work [10], based on the Bayesian learning framework with probability value estimation for each and every one of the node in the social network .When compare to existing sybil identification methods proposed system achieves less false rate ,but it becomes higher computation cost for online social networks it is applicable to only 30 nodes ,if the number of nodes is increases the computation cost of the system is high.

Shortest path route selection with Sybil node identification and detection also proposed in earlier work [11].The work calculates the shortest path between each one of the nodes in the network and determines the Sybil attacks within the network for smaller size of the networks. Trusted certification node based mechanism also useful to improve the Sybil node identification system through the creation of the centralized authority between one nodes to other nodes in the social networks, such as type of the verification process is supported by either hardware device [12] or a digital number [13-15

3. PROPOSED EFFICIENT SYBIL AND NON SYBIL IDENTIFICATION BASED MULTI COMMUNITY METHODOLOGY

In order to reduce or identification of the Sybil attack plays major important role in social network system .In existing this have been carried out by using Sybil Defender. Since the distributed environment are vulnerable easily by the sybil nodes since it affects entire information of the neighborhood nodes and collects the information of all nodes in the distributed networks, it consists of several number of steps to detect sybil nodes in the networks those steps are sybil community detection, and two approaches to limiting the attacks of sybil nodes. The major problem presented in the existing Sybil identification methods is that the Sybil population is lightly associated to each and every one of the non- Sybil population due to lack of trust, other than their inter-connections can be multi-hop. Non sybil

community are not supported in this work, it is solved by using proposed work.

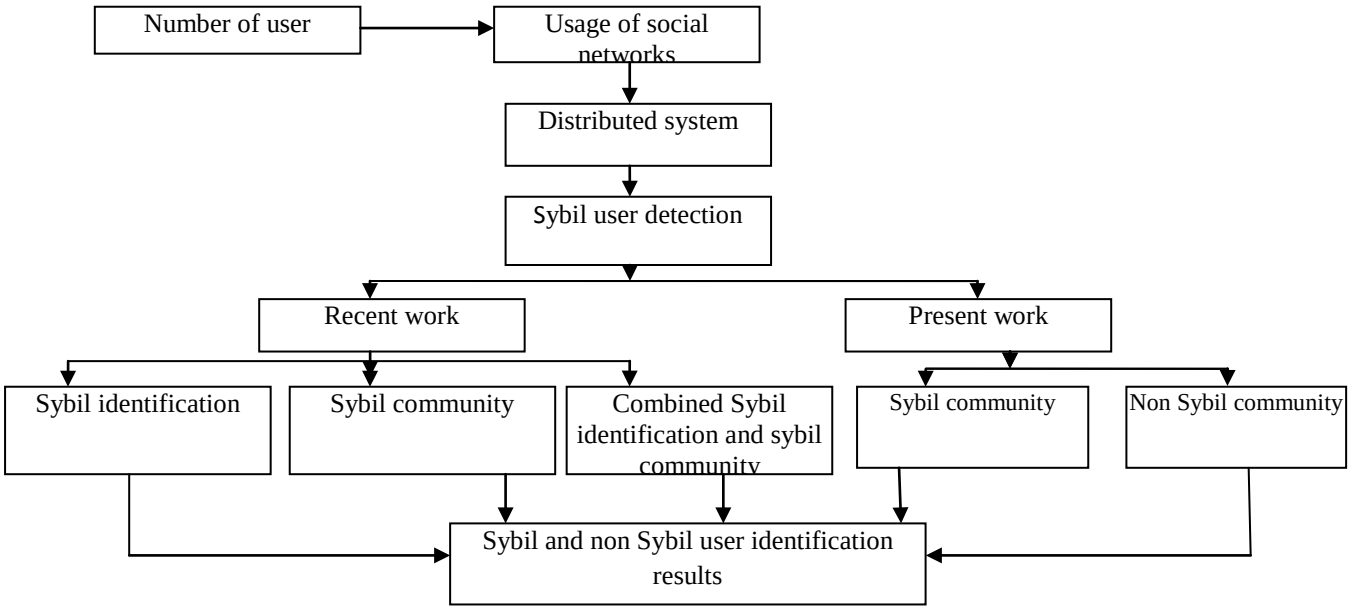

Figure 1: Architecture of the entire system

Social network of the proposed work have been represented in the form of graphical manner $G(V, E)$ a infer node u as input, and results becomes sybil or non sybil node detection result u.In existing work consists of two major works is specified in algorithm 1 and 2. In first part of the work the graph based representation of the social network is considered as input, maximum number of length for each and every node in the social network also determined through $l_s = \log n$ along with frequency level calculation R random walks with smaller level of the frequency count, then computes the mean and standard deviation for each and every result $< l;\ mean;\ stdDeviation >$.

Algorithm 1: Preprocessing (G,h)

1. $J = \{h\}$
2. **For** $i = 1\ to\ f$ **do**
3. Perform a random walk with length $l_s = \log n$ originating from h
4. J = J ∪ { the ending node of the random walk}
5. **End for**
6. l = l_{min}
7. **while** l<= l_{max} **do**
8. **for** $i = J.first\ ()\ to\ J.last()$ **do**
9. Perform R random walks with length l originating from node i
10. Get n_i as the number of nodes with frequency no smaller than t
11. **End for**
12. Output $< l, mean(\{ n_i : i\ \epsilon J\}), stdDeviation(\{n_i : i\ \epsilon J\}) >$
13. $l = l + 100$
14. **end while**

Algorithm 2

1. **l** = l_0
2. **while** l <= l_{max} **do**
3. Perform R random walks with length l originating from u
4. M = the number of nodes whose frequency is no smaller than t
5. Let the tuple corresponding to length l in the outputs of algorithm 1 be $< l, mean, stdDeviation >$
6. **If** $mean - m > stdDeviation * \alpha$ **then**
7. Output u is Sybil
8. End the algorithm
9. **End if**
10. $l = l * 2$
11. **end while**
12. output u is honest

In algorithm 2 decides whether the selected nodes is suspected node or malicious node through the calculation of the length values $l = l_0$ from algorithm one and the results of the algorithm 1 for each nodes $stdDeviation * \alpha(\alpha = 20)$,then consider the corresponding node as Sybil node or else it becomes non Sybil nodes and repeat the process it reaches maximum value until the Sybil node is identified l is larger than l_{max}.

Sybil Community Detection Algorithm

In this step it identification and detection of the Sybil nodes for entire community user (or) nodes presented in the social network based representation graph . The algorithm divide the Sybil region into smaller Sybil region to exact identification of the Sybil user in the social network graph based representation using network simulation tool, if the values of the Sybil node one S_1 is smaller than the Sybil region then it is identified as non Sybil region or else is considered as Sybil user ,similarly it is applied for Sybil user two. The system is applied to entire network through the random walk length estimation method in Algorithm 3 and it is applied to Algorithm 4. Initiate the algorithm with R random walks for each number of the Sybil nodes and calculation the ratio of lifeless walks; previous to they attain the required length. If it is smaller than the β it becomes close to one it is considered as the Sybil node. Repeat this procedure until the value of the node is no lesser than β.

Algorithm 3. Walk Length Estimation (G, s)

1. l = $l_0/2$
2. $dead\ Walk\ Ratio\ = 0$
3. **while** dead $Walk\ Ratio\ < \beta$ **do**
4. $l\ = l * 2$
5. dead $Walk\ Num\ = 0$
6. **for** $i\ = 1\ to\ R$ **do**
7. Perform a partial random walk originating from s with length l
8. **If** the partial random walk is dead before it reaches l hops **then**
9. $Dead\ Walk\ Num++$
10. **End if**
11. **End for**
12. $Dead\ Walk\ Ratio\ = dead\ Walk\ Num/R$
13. **End while**
14. Output l

Algorithm 4: Sybil Region Detection $(G, S, l\ from\ Alg3.)$

1. Set the frequency of all the nodes to be 0
2. **For** $I\ = 1\ to\ R$ **do**
3. Perform a partial random walk originating from node s with length l
4. $S.frequency++$
5. **For** $j\ = 1\ to\ l$ **do**
6. Let the j^{th} hop of the partial random walk be node k
7. $K.frequecy++$
8. **End for**
9. **End for**
10. Traversed list = Sort the traversed nodes by their frequency in decreasing order
11. $Counter\ = 0$
12. $S\ = \phi$
13. **Do**
14. $Counter\ = conductance(s)$
15. **For** $I\ = traversedList.first()\ to\ traversed\ List.last()$ **do**
16. **If** node i ϵ S **then**
17. Continue
18. If $conductance\ (\{i\} \cup S) <= conductance(s) then$
19. $S\ = \{i\} \cup S$
20. While $(counter\ >\ conductance(S))$
21. Output S

In algorithm 4 perform the Sybil community identification system based on the random walk theory in partial manner ,it may not produces best results in all social network ,in order to overcome these problems in this Algorithm 5 ,it considers suspected node results from algorithm 3 u. Instead of performing the random walk theory based Sybil identification the algorithm directly detects the Sybil nodes based on the conductance measure. it perform the sorting based methods to sort the nodes frequency count value of each and every node in the social network graph theory ,and simultaneously adds a new nodes or users to detect the Sybil nodes region and non Sybil region in the network

Algorithm 5: $Combo\ (G, u, tuples\ from\ Alg.1)$

1. $l = l_0$
2. **while** $l <= l_{max}$ **do**
3. Perform R random walks with length l originating from u
4. m = the number of nodes whose frequency is no smaller than t
5. Let the tuple corresponding to length l in the outputs of algorithm 1 be $< l, mean, stdDeviation >$
6. **If** $mean - m > std\ Deviation * \alpha$ **then**
7. Output u is Sybil
8. $Traversed\ list =$ $Sort\ the\ traversed\ nodes\ by\ their\ frequency\ i$
9. $Counter = 0$
10. $S = \phi$
11. **Do**
12. $Counter = conductance\ (S)$
13. **for** $I =$ $traversed\ List.first\ ()\ to\ traversed\ List.Last()$ **do**
14. **if** node iϵ S **then**
15. continue
16. **if** $conductance(\{i\} \cup S) <=$ $conductance(S)$ **then**
17. $S = \{i\} \cup S$
18. **While** $(counter > conductance\ (S))$
19. Output S
20. End the algorithm
21. **End if**
22. $l = l * 2$
23. **end while**
24. output u is honest

In above mentioned algorithm it doesn't detect non Sybil region ,in order to overcome these problem in this work present an new sybilshield mechanism to detect Sybil and non Sybil region in their community along with consideration of the verification node V to find the suspect node S, the verifies nodes accepts the suspected node if both nodes are intersected with each other ,otherwise verification nodes rejects the suspected node S. In order to avoid the misleading behavior of the nodes in the network introduce the agent walk ,In algorithm 6 route based agent walk is performed for each and every one of the honest communities, V accepts S if their if it contains at least one intersection. For each and every one of the nodes in the social network graph the algorithm calculate the routing table through the randomly selected neighbor's nodes. Routing tables are designed through indiscriminate permutation, demonstrating a one-to-one mapping beginning received edges to leaving edges. Routing table is well-known to contain the subsequent properties:

(1) Convergence property: if the dissimilar two random routes passes through destination nodes from same edges for each input nodes the output or the outgoing messages of the nodes have reach the same node and contains same information .

(2) Back-traceable property: if the two corresponding route nodes from earlier results consist of the same outgoing message they come from same incoming number of edges for each and every route

These two properties becomes easier to find the Sybil and non Sybil attacker information in the social network graph theory ,negative results of the Sybil attacker node to verifier node is reduce based on the calculation of the t_V with distance $d_V = 2$ to present a good substitution. In this work the length of the each and every one of the route for each nodes in the social network graph is measured based on the parameter w,it is intitated by user by calculation of the probability values for each nodes in the network within the specified community without consideration attack regions ,so it is easily identifies the attacker region in exact manner false positive rate of the system is increased through the calculation of honest nodes in their community.

Algorithm 6 : Initial Verification

for $i = 1\ to\ d_V$ do $/*\ d_V : V's\ degree\ */$
V performs random route along its i[th] edge;
for $j = 1\ to\ d_S$ do $/*\ d_S: S's\ degree\ */$
S performs random route along its j[th] edge;
Check whether an intersection exits by V's i[th] random route and S's j[th] random route and record the result;
End
if Intersection percentage is no less than t_S then
V accepts S along its i[th] edge;
Else
V rejects S along its i[th] edge;
End
End

if Along all $V's$ edges, the accepting ratio is no less than t_V then
V accepts S;
Else
V finds agents for further authentication;
End

In order to confirm their results of the verifier node to detect the sybil or suspected node results in the algorithm 7 we need to check the suspected node once again and honest nodes information also verified once again by calculation of the probability value to each and every nodes in their community to other community. In algorithm 7 shows the entire decision making results of the Sybil and non Sybil user identification system .The algorithm 7 it starts with the verification node along with the distance value or length value of w, pickup hop count value of the node as agent for current process. In step 2 both the verification node and the agent nodes are initiated simultaneously to calculate their state of the nodes. if both node have different community it may not traverse same path ,their belongs to Sybil community and another one of the node belongs to non Sybil community and enter one another's community ,the probability value of these two interconnected nodes becomes very less when compare to same community probability value ,the verified node is considered as outlier and the agent is considered as the valid node ,it continues until the length of the hop count w_V reaches the last ending point in the route. The length of the verifier node is compared based on the n_w threshold value and it becomes less $(n_w < w_V)$,it is also verified with node degree d_V , with number of efficient agents n_a is not higher than the d_V $(1 \leq n_a \leq d_V)$.

Algorithm 7: Agents Discovering

for $i = 1\ to\ d_V$ do d_V : $V's\ degree\ */$
Repeat
V performs random route along its i[th] edge with length w_V ;
V picks the last hop of the random route as an agent A;
V verifies A;
if V accepts A then
V increases its random route length by w_V ;
Else
V accepts A as a valid agent; break;
End
until $om\ route\ length > n_w w_V$;
End
n_a valid agents are found, where $n_a \leq d_V$

After the identification of the Sybil and non Sybil in the social network the verification of the nodes plays major important to detect the Sybil and non Sybil user in their community using the following Algorithm 8. Assume that the degree of each and every nodes in the social network d_A and nodes in the network at least contains one intersection point the route is accepted by system or else it is not accepted by the system ,honest community of the Sybil nodes and non Sybil nodes contains at least $\frac{1}{2}$ In this case, the opposition would manage the indiscriminate way of the Sybil agent depending on the individuality each and every nodes in the social network .

Algorithm 8: Agent-Aid Verification

for $i = 1\ to\ na$ do
for $j = 1\ to\ d_A$ do /* d_A: current agent's degree
i[th] agent performs random route along its j[th] edge;
i[th] agent verifies S by Algorithm 1 and records its accept/reject decision;
End
End
if Among n_a agents, the accepting ratio is no less than t then
V accepts S;
Else
V rejects S;
End

4. EXPERIMENTAL RESULTS

In this section finally measure the performance of the proposed non Sybil region identification system with existing Sybil attacks identification (Sybil detection) method based on the parameters like collision rate, routing load, end to end delay ,packet delivery ratio, Falsely detected nodes between the existing and proposed Sybil and non Sybil region identification system .The proposed methods achieves best results than the existing system since it additionally identifies Sybil and non Sybil regions community in the social network by generation of the random nodes through the network simulator in the networks .The comparison results of the methods and the parameter results are shown in the following Figure 2,3,4,

ISSN 2348 – 7968

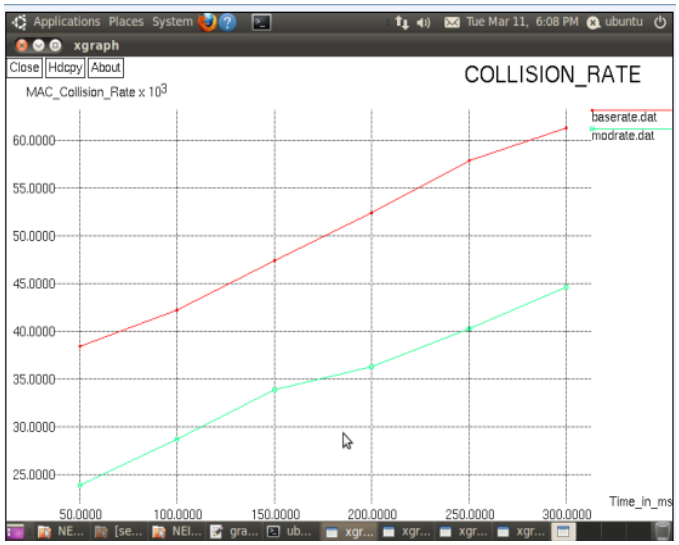

Figure 2: Collision rate vs methods

In the Figure 2 measure the collision rate of the Sybil Defender and sybilshield ,if the usage time to complete social network searching is completed then the collision rate of the proposed sybilshield is less when compare to existing methods.

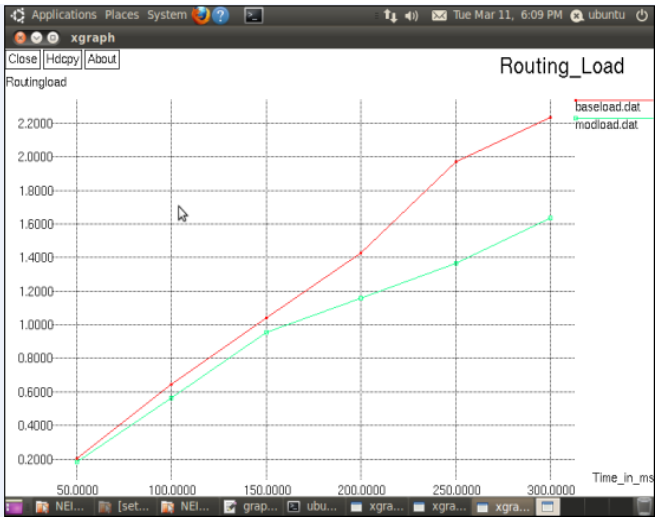

Figure 3: Routing load vs methods

In the Figure 3 measure the routing load of the Sybil Defender and sybilshield ,if the usage time to complete social network searching is completed then the routing load of the proposed sybilshield is less when compare to existing methods.

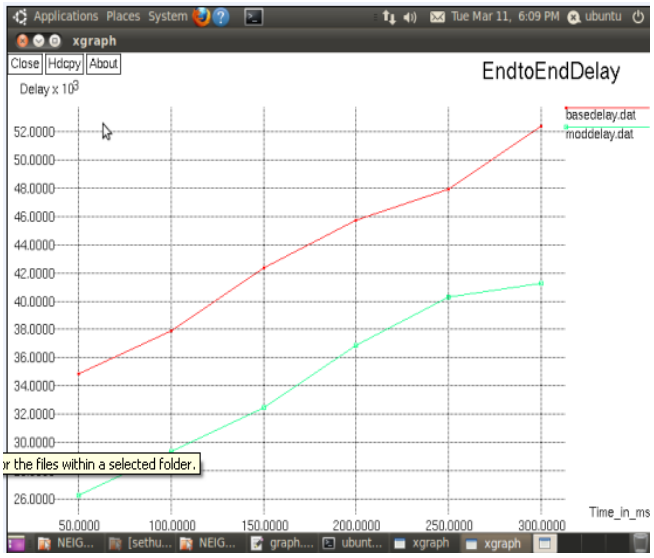

Figure 4: End to end delay vs methods

In the Figure 4 measure the end to end delay of the Sybil Defender and sybilshield ,if the usage time to complete social network searching is completed then the end to end delay of the proposed sybilshield is less when compare to existing methods.

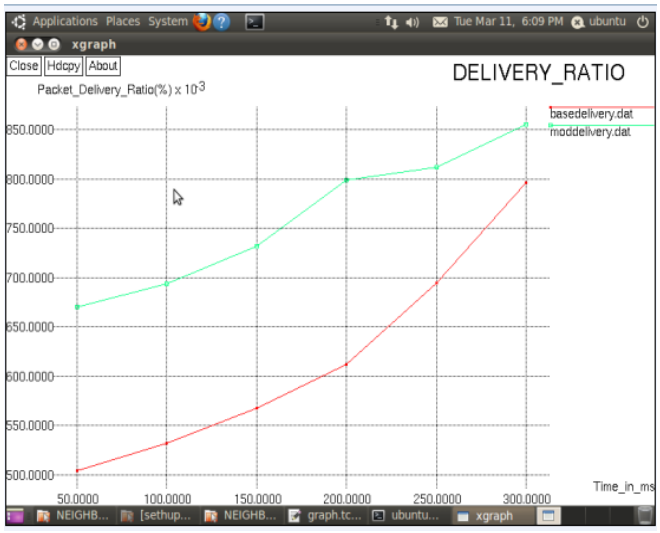

Figure 5: Packet delivery ratio vs methods

In the Figure 5 measure the packet delivery ratio of the Sybil Defender and sybilshield ,if the usage time to complete social network searching is completed then the end to end delay of the proposed sybilshield is high when compare to existing methods.

ISSN 2348 – 7968

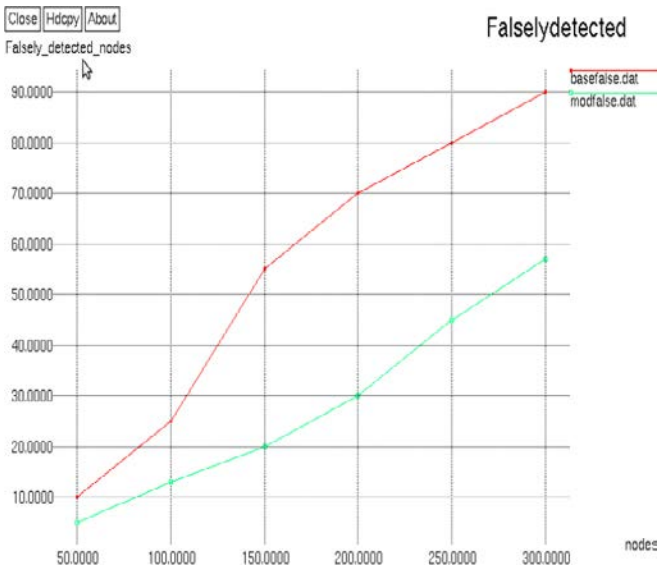

Figure 6: Falsely detect nodes vs methods

In the Figure 6 measure Falsely detected nodes of the Sybil Defender and sybilshield , if the number of nodes increases the falsely detected node level increases in existing system when compare to proposed system .

TABLE 1: PERFORMANCE COMPARISON RESULTS OF THE EXISTING AND PROPOSED METHODS

Parameters	**Number of time in ms**	**50**	**100**	**150**	**200**	**250**	**300**
Collision rate (10^3)	**Base rate**	38361.75	42156.67	47367.34	52345.12	57844.57	61234.35
	Proposed rate	23863.5	28645.73	33865.21	36231.667	40234.233	44596.578
End to end delay (10^3)	**Base delay**	34840	37895	42367	45737	47898	52376
	Proposed delay	50 26271	29375	32458	36843	40262	41232
Packet delivery ratio	**Based delivery**	0.50432	0.532221	0.56743	0.61234	0.69456	0.79662
	Proposed delivery	0.66996	0.69362	0.73227	0.79863	0.812236	0.85525
Node level	**Number of nodes**	**50**	**100**	**150**	**200**	**250**	**300**
False rate	**Base false**	10	25	55	70	80	90
	Proposed false	5	13	20	30	45	57
Honest level	**Base honest**	5	13	20	30	45	57
	Proposed honest	2	6	11	15	23	27

5. CONCLUSION AND FUTURE WORK

5.1. CONCLUSION

In this paper proposed an efficient Sybil Defender, mechanism to detect the Sybil and non Sybil regions for social networks or network users in the Sybil user as well as non Sybil users in the system .Proposed Sybil defender mechanism additionally detects the Sybil through the community based Sybil mechanism for single and numerous community based social networks which rejects the falsely detected nodes and improved identification results of the Sybil and non Sybil user when compare to the existing protocol through the identification of the misconduct and truthful nodes. In this proposed methods additionally non-Sybil regions also identified and the number of misbehavior attacks created in middle of the social networks becomes less when compare to existing methods with multi community mechanism by using the creation of the network nodes using the simulation tool NS2,it additionally includes an agent to collect the information the information of the neighborhoods nodes to remove the truthfulness nodes in the social network ,thus reject the misbehavior nodes in the network and selects the another path (route) randomly until the best route is found.

Experimentation results shows that the proposed sybilshield achieves higher detection rate, less collision rate ,high packet delivery ratio ,less end to end delay

5.2. FUTURE WORK

In future work we apply the existing SybilShield methods into the more real social network data among dissimilar structures and further improve the efficiency of our SybilShield algorithm.

REFERENCES

1. J. R. Douceur. **The sybil attack**. *In IPTPS*, 2002
2. K. Xing and X. Cheng. **From time domain to space domain: Detecting replica attacks in mobile ad hoc networks.** *In IEEE INFOCOM*, 2010.
3. K. Xing, F. Liu, X. Cheng, and D. H. Du. **Realtime detection of clone attacks in wireless sensor networks**. *In IEEE ICDCS*, 2008
4. G. Danezis and P. Mit. **Sybilinfer: Detecting sybil nodes using social networks**. *In NDSS*, 2009.
5. N. Tran, J. Li, L. Subramanian, and S. S.M. Chow. **Optimal sybilresilient node admission control.** *In IEEE INFOCOM*, 2011.
6. L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. **All your contacts are belong to us: automated identity theft attacks on social networks**. *In WWW*, 2009.
7. **The top 500 sites on the web.** *http://www.alexa.com/topsites*.
8. H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. **Sybillimit: A nearoptimal social network defense against sybil attacks**. *In IEEE Symposium on Security and Privacy*, 2008.
9. H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. **Sybilguard: defending against sybil attacks via social networks**. *In SIGCOMM*, 2006.
10. N. Tran, J. Li, L. Subramanian, and S. S.M. Chow. **Optimal sybilresilient node admission control**. *In IEEE INFOCOM*, 2011
11. G. Danezis and P. Mit. **Sybilinfer: Detecting sybil nodes using social networks**. *In NDSS*, 2009.
12. L. Xu, S. Chainan, H. Takizawa, and H. Kobayashi. **Resisting Sybil attack by social network and network clustering.** *In SAINT*, 2010.
13. J. Newsome, E. Shi, D. Song, and A. Perrig, **The sybil attack in sensor networks: analysis & defenses**, *in Proc. of ACM IPSN*, 2004, pp. 259–268.
14. J. Ledlie and M. Seltzer, **Distributed, secure load balancing with skew, heterogeneity and churn**,*in Proc. of IEEE INFOCOM*, vol. 2, 2005, pp. 1419–1430.
15. G. Mathur, V. Padmanabhan, and D. Simon, **Securing routing in open networks using secure trace route**, *in Technical report MSR-TR-* 2004-66, Microsoft Research, **2004.**