doi:10.5937/jaes14-10176 Paper number: 14(2016)1, 355, 75 - 83

DETECTION OF SYBIL'S ACROSS COMMUNITIES OVER SOCIAL INTERNET OF THINGS

A. Meena Kowshalya,
Government College of Technology,Coimbatore, Tamil Nadu, India
M.L.Valarmathi,
Government College of Technology,Coimbatore, Tamil Nadu, India

Social Internet of Things is a new paradigm that integrates Internet of things and Social Networks. Several challenges exist in building Social Internet of Things (SIoT). Very limited research has been carried out in the past 7 years to build a reliable Social Internet of Things community. A major threat with Social Things is Sybil attacks. Since SIoT is comprised of autonomous objects/nodes, tracking fake identities is an open problem. This paper proposes a new mechanism to identify Sybil's in communities of Social Internet of Things. This paper aims at (i) identifying communities among Social Internet of Things using Community_Infer algorithm. Using the properties of Social Networks and ACO heuristics various communities among the Social Internet of Things were identified. (ii) The communities are checked for existence of Sybil's. The algorithm Detect_Sybil detects and classifies the number of Sybil's in each communities. Compared to existing schemes the proposed method classifies communities accurately with a high modularity score.

Keywords: Node attributes similarity, Ant Colony Optimization, Community detection, Sybil detection

INTRODUCTION

Internet of Things has enabled integration of new heterogeneous technologies and communication standards. Social Networks (SN) are revolutions beyond Internet of Things (IoT). The convergence of IoT and Social Networks has paved way for a new paradigm called Social Internet of Things where not only humans collaborate but also objects. Objects in SIoT are autonomous and controlled via owners. SIoT environment is such an example for the 3rd ICT wave where object interact and collaborate to achieve the desired task. One of the many visions of IoT is to make objects not only smarter but also socially conscious. The authors in [06] defined Social Internet of Things as collection of intelligent objects. [25] has reported evolutions, applications, architecture, challenges and solutions for Internet of Things. Establishing a successful Social Internet of Things community is a complex task. Challenges like Data Management, Data discovery, Interoperability, Trust Management, Security, Privacy, Heterogeneity, and Fault Tolerance has to be handled. Table 1 illustrates the difference between a Social network and a SIoT. The table show what humans do using a social network and what object does in a SIoT.

SIoT objects mimic human behavior. [18] Suggests four kinds of social relationships between objects. The things can be engaged in (a) Parental object relationship defined by the same manufacturer among similar objects. (b) Co-location or Co-work objects relationship defined by objects which periodically or constantly reside in the same place. (c) Social object relationship established when objects come into contact when their owners do. (d) Ownership objects relationship established by objects of same owners. To establish a SIoT network, relationship and behavior of objects play an important role.

Table 1: Social network Vs Social Internet of Things

Social Network	Social Internet of Things
Humans increase popularity and become visible	Objects publish information or services
Find new friends	Find new services and updates information
Get context information	Get environmental information / characteristics

* *Southwest State University, st. 50 October, d.94, Kursk, Russia;*
swsu.ee@gmail.com

[01] Proposed a widely accepted classification of social relationship among humans. The authors describe four relational structures namely communal sharing, equality matching, authority ranking and market pricing used to generate social actions. Community sharing can be associated with objects that have collective relevance ie., not relevant individually. Equality sharing represents relationship between objects that maintains their individuality and exchange information between objects that they consider as equals. Authority ranking relationship established between objects of different hierarchical level, for example a RFID reader and tag/ Master slave relationship. Market pricing relationship represents objects that work with other objects for mutual benefits. SIoT enables not only about human- human, human-object/node but also object/node – object/node. The term object and node are interchangeably used in the paper, literally SIoT environments means node to be an object. The authors in [09] proved that the SIoT network is navigable and efficient service search can be done without affecting the application performance. This paper focuses on communal sharing and identifies communities in Social Internet of Things. Since the environment is open, major threat with such networks are Sybil attacks where many fake identities of objects are created to disrupt the reliability of network communication. This paper also detects Sybil's in the identified communities. The main contributions of this paper are:

I. The communities in a Social Internet of Things are detected using the Community_ Infer algorithm. The algorithm uses Attributes of nodes to classify them as communities. The Ant Colony Optimization algorithm is employed along with node attribute similarity to efficiently detect communities.

II. The Sybil_Detect algorithm detects number of Sybil nodes in the communities. Since honest nodes are fast mixing, there exists a cut between the Sybil region and the honest region. This simple strategy is used to detect the Sybil nodes in the communities.

To the best of our knowledge this is the first work that detects Sybil nodes in a community. Compared with the existing schemes the Community detection algorithm and the Sybil detection algorithms are faster and accurate. The rest of the paper is organized as follows: Section 2 presents the related work, Section 3 discusses the proposed Community_Infer algorithm, Section 4 presents the proposed Sybil_Detect algorithm, and Section 5 discusses experimental results followed by Conclusion.

RELATED WORKS

Community Detection using ACO

This section discusses the recent works in community detection using ACO. The authors in [20] defined a cluster by comparing it to a random graph. The degree distribution of the random graph was the same as that of the original graph. Community detection algorithms fall into two main categories namely modularity based and non modularity based. This paper focus only on modularity based algorithms. Modularity based algorithms are further divided into agglomerative based and divisive based. Divisive/ Agglomerative algorithms divide/aggregate nodes in a network into groups based on dissimilarities/ similarities. The authors in [24] proposed a computationally less expensive linear time algorithm to detect communities in large scale networks. [16] used a greedy heuristics to maximize the modularity score and obtain quality clustered communities. As an extension of [16], the authors in [03, 17, and 13] proposed improvements over the existing schemes that aimed to detect community structures based on quality modularity score. This paper uses the modularity concept of the existing algorithms and implements Ant colony Optimization heuristic technique for detection of communities. The collective behavior of ants to forage food resulted in communication between them by using their pheromone (chemicals deposited). This pheromone marks the path from the nest to the food and the same pheromone acts as a guide to other ants for the shortest path food. This technique is being used in Social Internet of Things where nodes/objects establish communication with each other directly or indirectly. To predict the social behavior of objects and to classify them ACO algorithm can be used. [08] proposed the first Ant Colony Optimization algorithm called the Ant system (AS). Several versions of ACO algorithms are proposed in the literature. The way pheromones are updated and the search procedures differ from algorithm to algorithm.

Sybil Identification

Recent literature has proposed many solutions to Sybil attacks. These approaches have been classified into three categories namely Sybil prevention methods, Social network based Sybil detection methods and Social network based Sybil tolerance methods. Sybil prevention approaches tries to reduce the number of false identities. Such approaches require a trusted central authority [22]. Social network based Sybil detection scheme can be further divided into three categories.

1. Trust based Sybil detection
2. Community based Sybil detection
3. Rank based Sybil detection

[12, 11], [19] are the Trust based Sybil detection schemes that successfully identify and limit malicious identities created in a social network. Existing community based algorithms such as [03], [17] can be used to detect Sybil nodes in a social network. The only issue with these approaches is their faster mixing time. Sybil ranking mechanisms output a trust score for each node in a social network and differentiate Sybil and honest nodes by their trustworthiness. [13] uses degree normalization and trust propagation to detect Sybil nodes. Social network based Sybil tolerance schemes such as [04] relies on the graph structure features and application specific information of the network based on the above classification. Few recent Sybil defense mechanisms have been compared in table 2. [12, 11] proposed SybilGuard and SybilLimit, [10, 19, 27] proposed SybilInfer, SybilResilient and SybilDefender. All these protocols perform well when the topology is small and the algorithms are not computationally efficient. There is a need for a protocol that is computationally efficient and scalable that can handle millions of users.

Sybil Guard [12] is a novel protocol that uses the concept of a cut between Sybil users and honest users and exploits this property to bind the number of identities that a malicious user can create. This protocol relies on fast mixing property whereas the current social networks are not fast mixing. SybilLimit [11] is a defense mechanism against Sybil attacks using social networks. Compared to [12] which accepts O(√n logn) Sybil nodes per attack edge, [11] accepts only O(log n) Sybil nodes per attack edges. SybilLimit leverages independent instances of random routes to find short routes. Instead of exploiting intersection of edges between nodes it exploits intersection of edges which shows a significant improvement over SybilGuard. SybilInfer[10] uses Markov chain to transform a graph G into G' in order to maintain a uniform node distribution. It proves that the presence of Sybil nodes increases the mixing time of the graph. [19] is an improved protocol which is decentralized and outperforms [11] in terms of number of false negatives.[27] is a predecessor of [19] maintaining O(log n) guarantee under sum up settings. Sybilproof [05] creates one DHT node for every node in the network for user identification. The performance of [05, 04] are the same, [04] uses user feedback and a centralized design to identify Sybil's.

3. Community Detection using ACO algorithm.

The Social Internet of Things is represented as a graph G(V,E) where V is the set of vertices representing the nodes/objects, |V| = n and E represents the number of links between nodes |E| = e. The algorithm works by recursively dividing the graph into two. As a result, the resultant graph obtained is G1 and G2. Let the converged binary solution of ACO be c*. The original graph is partitioned only if the modularity value is high. The modularity score is calculated according to [20] as shown in equation 1.

$$M(x) = \frac{1}{2|e|} \sum_{i=0}^{n} \sum_{j=1}^{n} \mathrm{bij}\ \delta(x_i, x_j) \quad (1)$$

Table 2: Comparison of existing Sybil Defense mechanism

Protocol	Honest Region	Technique	Design
SybilGuard [09]	Unbalanced	Random walk	Decentralized
SybilLimit [10]	Unbalanced	Random walk	Decentralized
SybilInfer [11]	Unbalanced	Random walk	Centralized
SybilResilient [12]	Balanced	Random walk and Breadth first search	Decentralized
SybilDefender [13]	Balanced	Random walk	Decentralized
Sybilproof [14]	Unbalanced	Random walk Distributed Hash Table	Decentralized
Ostra [15]	Unbalanced	User feedback	Centralized

b_{ij} is the excess number of edges between i and j ie., actual number of edges – expected number of edges.

All nodes are assigned a set of pheromone values τ = ($τ_1$, $τ_2$, …., $τ_n$) τi ant assigned to node i to group 1. As the iteration progresses, ants generates their solution based on their pheromone levels. The pheromone vector is updated to get better solution quality. The process is repeated until τ converges to the solution c*. The solution construction of the algorithm is simple. For each iteration t, every ant x constructs a solution solxt using τ. The probability of constructing the solution is given by [20] as

$$P(sol_i^{xt} = 1) = \tau_i \text{ and } P(sol_i^{xt} = 0) = 1 - \tau_i \quad (2)$$

The pheromone gets updated according to the equation

$$\tau \rightarrow \tau + \rho\ (c^* - \tau) \quad (3)$$

where ρ is the evaporation factor. According to [06] it is proved that the pheromone values converge after a finite series of iterations. At the end of each iteration, convergence is checked.

Community_Infer algorithm

```
Algorithm: Community_Infer
initialize $T$=0, $τ_i$=0.4, $N_{ant}$, $T$
while t < $T$
for all ants do
    solxt = construct solution (τ)
end for
Update pheromone
if convergence
    reset $τ$
Calculate node attribute similarity
Classify nodes into communities
```

Node attributes similarity

Node attribute based similarity is an efficient way of producing meaningful weights when compared to clustering coefficient similarity (CCS) and common neighbor similarity (CNS). For classifying the communities of the Social Internet of Things, node based attribute similarity can be employed since CCS and CNS are based on the network topology [15].

```
for each node i=1,2,3….,n do
    for each node j=1,2,… neighbors(i) do
        w(i,j)=0
            for each node attribute a do
                if a is nominal and i.a - j.a then
                    w(i,j) = w(i,j) + 1
                else if a is continuous then
                    w(i,j) = w(i,j) + 1 – α | i.a – j.a |
                end if
            end for
        end for
end for
```

Sybil Node Detection:

Sybil nodes are not as easy to detect and they disrupt the reliability of the network. Let G(V,E) be a graph consisting of V vertices and E edges. V is number of nodes in the network community and E represents the links between them. An initiator node is chosen to select two neighboring nodes randomly. The graph G(V,E) is comprised of set of imitator nodes v and their neighboring nodes. The algorithm for Sybil node detection is shown.

Detect_Sybil algorithm

```
Detect_Sybil (v)
Input: Imitator node v
Output: Honest nodes neighboring v
 1. Select the init node N, let N chooses neighbors randomly
 2. Initialize the community G(V,E)
 3. Compute the position of nodes using the trilateration
 4. For G(V,E)
    a. Perform Random walk through the community
    b. Find set of honest seeds using direct and indirect opinion
 5. For each Hn
    a. Find nodes n whose frequency f >3 (threshold)
    b. Compute mean(ni) and standard deviation stddev(ni)
 6. If mean – ni > stddev
        then u is Sybil
else
    u is honest
```

The position of every node can be found using trilateration. The algorithm performs a random walk to find the number of honest seeds. This could be done by direct opinion and indirect recommendation of neighboring nodes. For each honest seed whose frequency is greater than 3, the mean and standard deviation are calculated. If the difference between mean and total honest seeds is greater than the standard deviation then the node is Sybil node. The edge between this node and the other honest node is labeled as attack edge. Thus this algorithm is repeated to find all Sybil nodes in the network. Direct and indirect opinions are in the range [0, 1].

EXPERIMENTAL RESULTS

Jazz Musicians [21] and Zachary's Karate club [26] was used for the experiments. SIoT is a dynamic environment where objects mimic human behavior, hence the basic assumption here is that things/objects move when their owners move. SWIM (Small World in Motion) was used to generate events of objects. Thus the datasets were modified accordingly to SIoT environment. [21] is the collaboration network between Jazz musicians. Each node is a Jazz musician and an edge denotes that two musicians have played together in a band. [26] is a Social network of friendships between 34 members of a karate club at a US university in the 1970. Figure 1 and Figure 2 shows the degree distribution and cumulative degree distribution for Jazz musician's dataset. The local clustering coefficient for the same is shown in figure 3. Figure 4 shows the layout of jazz musicians where the algorithm has detected 4 communities' namely white and black communities and their smaller sub communities. Figure 5 and 6 shows the degree distribution and cumulative degree distribution for Zachary karate club. Figure 7 shows the local clustering coefficient for the same. Figure 8 represents the layout of the karate club and the detected communities. Due to dispute among node 1 and 34 the network spilt initially into two and later the algorithm detected 3 new sub communities. Table 3 illustrates the properties of the datasets. Graphml was used to generate graphical layouts of networks.

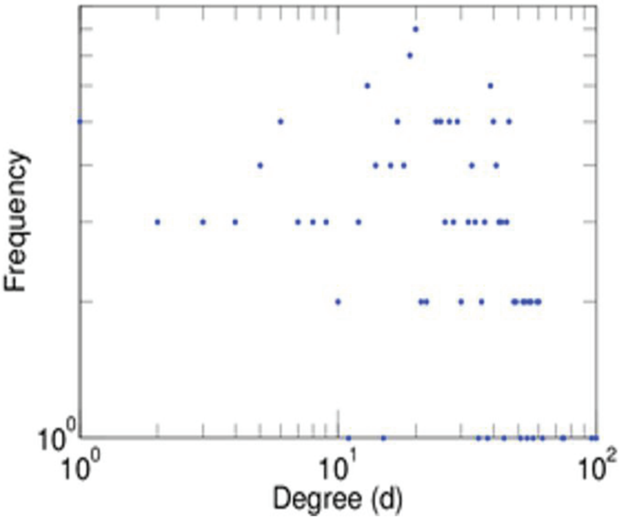

Figure 1: Degree distribution for Jazz musicians

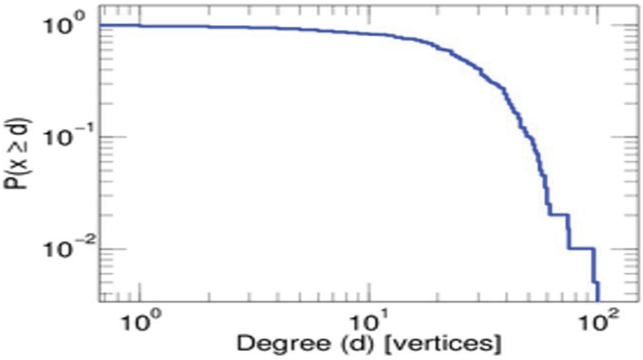

Figure 2: Cumulative degree distributions for Jazz musicians

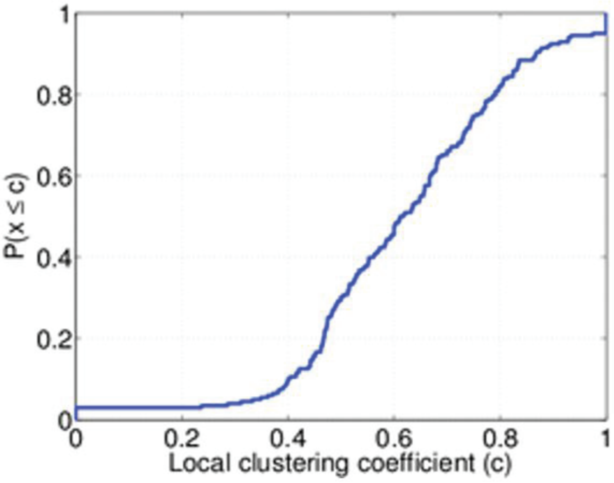

Figure 3: Local clustering Coefficient for Jazz musicians

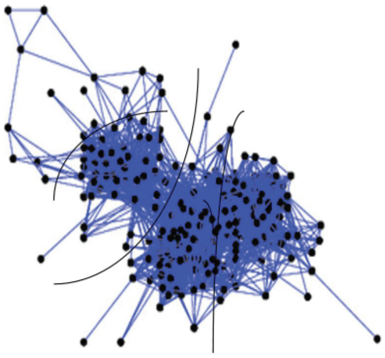

Figure 4: Four communities of Jazz musicians

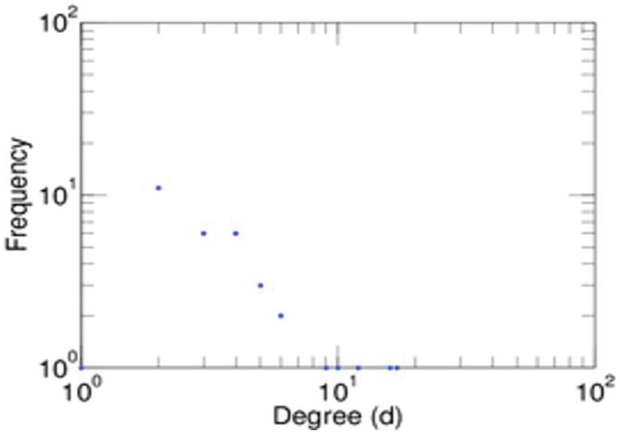

Figure 5: Degree distribution for Karate club

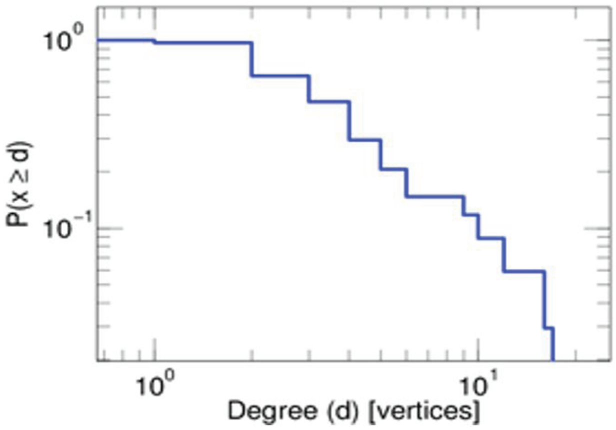

Figure 6: Cumulative degree distributions for Karate club

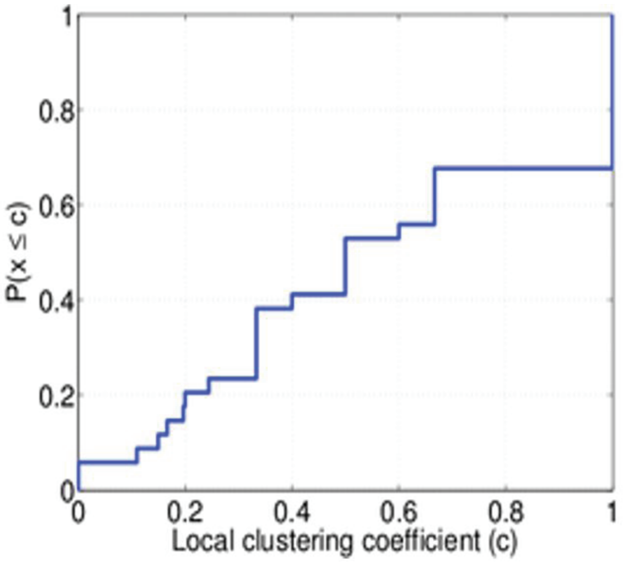

Figure 7: Local clustering Coefficient for Karate club

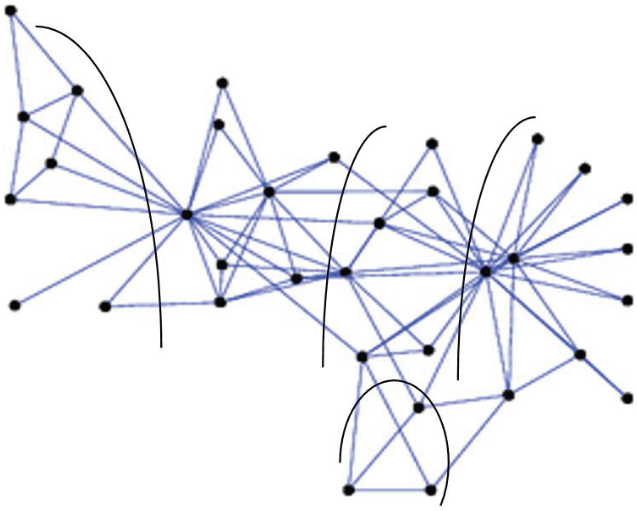

Figure 8: Five communities detected for Karate club

Table 3: Properties of Jazz musicians and Karate Club

Property	Jazz Musicians	Karate Club
Vertices	198 vertices (musicians)	34 vertices (members)
Edges	2,742 edges (collaborations)	78 edges (ties)
Average degree	27.697 edges / vertex	4.5882 edges / vertex
Clustering coefficient	52.0%	25.6%
Diameter	6 edges	5 edges
Mean shortest path length	2.21 edges	2.44 edges

Table 4: Time taken by Community_Infer algorithm

Network	Nodes	Time taken (seconds)
Karate Club	34	0.1 approximately
Jazz musicians	198	1.12

Table 4 shows the comparison of time taken by Community_Infer algorithm for Karate Club and Jazz musicians.

Table 5 presents the comparison of modularity scores and number of community detected across various algorithms. The proposed community_Infer algorithm detected 4 communities in both the networks and the modularity score of Jazz musicians showed substantial improvement over existing algorithms. To detect Sybil's in the community, one Sybil node in karate club and four Sybil's in Jazz musicians were induced at random. All other nodes are the honest nodes. The algorithm Detect_Sybil detected 1 Sybil for Karate club correctly and 5 Sybil's for Jazz musicians instead of 4. One honest node was identified as Sybil in Jazz musician's network. This false positive could be eliminated by improving the detection mechanism in the future. The running time of the algorithm was O(log n) and found to be efficient since Sybil detection was done inside communities. Existing algorithms for Sybil detection [12, 11, 10, 19, 27, 05, and 04] detect Sybil's on a whole network or sub network. This is the first attempt to find Sybil's based on Community of Interest of users. More security can also be imposed using attributed based encryption schemes [02].

CONCLUSION

Social objects can be grouped easily based on their interest. Community detection is very crucial in Social Internet of Things. This paper detects communities in the SIoT environment using the Ant Colony Optimization heuristics and the node attribute similarity property. Sybil attacks are major threats in the current social environments. This paper eliminates Sybil nodes within communities. Compared to the existing schemes the proposed scheme detects communities efficiently with a high modularity score. The proposed Community_Infer has produced a modularity score of 0.419 and 0.617 for Karate Club and Jazz musicians respectively. Number of Communities detected was 4 and 5 for Karate Club and Jazz musicians respectively. The Detect_Sybil Algorithm detected accurately the 1 Sybil node for Karate Club and 5 Sybil nodes for Jazz musicians. Only 4 nodes in Jazz musicians were Sybil's. This false positive could be eliminated in the future by more accurate and powerful Sybil Detection schemes. Also, the same can be tested for large scale social networks.

Table 5: Comparison of modularity values and detected communities across various algorithms

Network	Nodes	External optimization [25]		Spectral partitioning [19]		ACO [24]		Community_ Infer	
		Modularity	Community	Modularity	Community	Modularity	Community	Modularity	Community
Karate Club	34	0.419	4	0.419	2	0.419	4	0.419	4
Jazz musicians	198	0.445	5	0.445	4	0.595	4	0.617	5

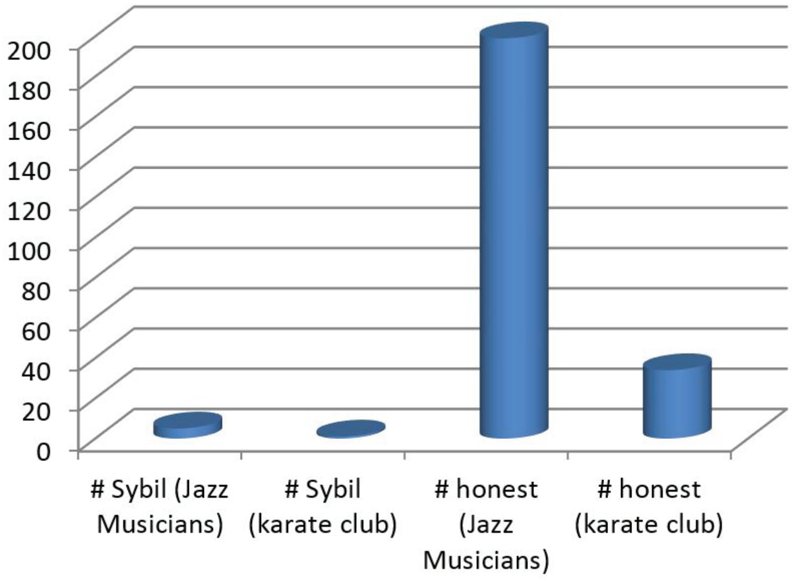

Figure 9: Number of Sybil's and Honest nodes of Karate Club and Jazz Musicians

REFERENCES

1) A. P. Fiske, "The four elementary forms of sociality: framework for a unified theory of social relations," Psychological review, vol. 99, 1992.
2) A.MeenaKowshalya, Dr.M.L.Valarmathi, "Secure and Efficient Ciphertext Policy Attribute Based Encryption without key Escrow Problem", International Journal of Emerging Engineering Research and Technology, Vol 2, No 2, pp:126-130, 2014.
3) Aaron Clauset,M.E.J.Newman, Cristopher Moore, "Finding community structure in very large networks", Phys. Rev. E 70, 066111 ,2004.
4) Alan Mislove , Ansley Post , Peter Druschel , Krishna P. Gummadi, Ostra, " leveraging trust to thwart unwanted communication", Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, pp.15-30, April 16-18, 2008, San Francisco, California
5) Alice Cheng , Eric Friedman,"Sybilproof reputation mechanisms", Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, August 22-22, 2005, Philadelphia, Pennsylvania, USA
6) Bao.F and Chen.I.R, 2012 Dynamic Trust Management for the Internet of Things Applications , International Workshop on Self Aware IoT, San Jose, California, USA, pp:1-6.
7) Blum C, Dorigo M, the hypercube framework for Ant Colony Optimization, IEEE Transactions on Systems, Man, and Cybernetics, Vol 34, No 2, pp : 1161 – 1172, 2004.
8) Dorigo, M. Maniezzo, V. Colorni, A, "Ant system: optimization by a colony of cooperating agents", IEEE Transactions on Systems, Man, and Cybernetics, Vol:26 , No:1, 1996.
9) Dr. M. L. Valarmathi, A.Meena Kowshalya, "Improved Network Navigability and Service Search in Social Internet of Things (SIoT)", International Journal of Research and Scientific Innovation, Vol 3, No 11, pp: 75-77, 2015. 28
10) G. Danezis and P. Mittal. SybilInfer: Detecting Sybil Nodes using Social Networks. In NDSS, 2009
11) H. Yu , P. B. Gibbons , M. Kaminsky and F. Xiao "SybilLimit: A near-optimal social network defense against sybil attacks", Proc. 2008 IEEE Symp. Security and Privacy, pp.3 -17 2008
12) Haifeng Yu, Kaminsky, M, Gibbons, P.B, Flaxman, A.D., "SybilGuard: Defending Against Sybil Attacks via Social Networks ",IEEE/ACM Transactions on Networking, Vol:16, Issue:3, pp:576 – 589, June 2008.
13) J Ruan, W Zhang, "An efficient spectral algorithm for network community discovery and its applications to biological and social networks", Seventh IEEE International Conference on Data Mining, pp: 643 – 648, 2007.
14) J. Duch, A. Arenas," Community detection in complex networks using Extremal Optimization", Physical Review E, vol. 72, 027104, 2005
15) Karsten Steinhaeuser, Nitesh V Chawla, "Community Detection in a Large Real-World Social Network", Social Computing, Behavioral Modeling, and Prediction, Springer, pp 168-175, 2008.
16) M. E. J. Newman, "Fast algorithm for detecting community structure in networks", Phys. Rev. E 69, 066133 ,2004
17) M. E. J. Newman, "Modularity and community structure in networks", Proceddings of National Acadamey of Sciences of United States of America, vol. 103 no. 23, pp: 8577–858 , 2006.
18) M. Nitti, R. Girau, and L. Atzori, "Trustworthiness Management in the Social Internet of Things," IEEE Transactions on Knowledge and Data Management, Vol. 26, Issue No. 5, 2014, pp. 1-11.
19) N. Tran, J. Li, L. Subramanian, and S. Chow. Optimal Sybil-resilient Node Admission Control. In INFOCOM, Apr. 2011.
20) Newman MEJ, Girvan M, "Finding and evaluating community structure in networks, Phys Rev E 69:026113
21) Pablo M. Gleiser and Leon Danon. Community structure in jazz. Advances in Complex Systems, 6(4):565-573, 2003.
22) R.Levien, A.Aiken, " Attack Resistant Trust Metrics for Public Key Certificates", in Proceedings of 7th USENIX Security Symposium, pp. 229-242, Texas, 1998.
23) Supreet Reddy Mandala et al, " Clustering Social Networks using Ant Colony Optimization", Operational Research ,Vol 13, No 1, pp 47-65. April 2013.

24) Usha Nandini Raghavan, Reka Albert, Soundar Kumara, "Near linear time algorithm to detect community structures in large-scale networks", Physics Review E 76, 036106 (2007)

25) Valarmathi.M.L, Kowshalya.M, Aarthi.M, 2015 Research Challenges in the Social Internet of Things (SIoT) – A Survey, Proceedings of National Conference on Science Research and Information Technology :128 – 133

26) W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research 33, 452-473 ,1977

27) Wei Wei ; Fengyuan Xu ; Tan, C.C. ; Qun Li, "SybilDefender: Defend against sybil attacks in large social networks", Proceeding of IEEE InfoComm, PP: 1951 – 1959, March 2012.

28) Yan ,Lihua Yin, "A Security Routing Mechanism against Sybil Attacks in Mobile Social Networks", APWeb Workshops,pp.325-332, Springer, 2014.

Paper sent to revision: 05.02.2016.

Paper ready for publication: 09.03.2016.