iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54732

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Leveraging Social Networks for P2P Content-Based File Sharing in Disconnected Manets

Dr. S. Sahaya Tamil Selvi
M. Sc., M.Phil., MBA, SET, Ph. D.,
Associate Professor and Head, Department of Computer Science, St. Joseph's College for Women, Tirupur, Tamil Nadu India.

Abstract: Mobile Ad Hoc Networks (MANETs) consist of digital devices, nodes are constantly moving, forming disconnected form of network with exploiting chances offered by immediate circumstances without reference to a general plan node encountering. Such transient network connections pose a challenge for the development of P2P MANETs. Traditional methods supporting P2P MANETs are flooding-based, advertisement-based and social networking method. This paper proposes a P2P content-based file sharing system, namely SPOON, for disconnected MANETs. The system uses an interest extraction algorithm to derive a node's interests from its files for content-based file searching. For efficient file searching, SPOON groups common-interest nodes that frequently meet with each other as communities.
Keywords: Ad-hock network, MANETs, Content-Based File Sharing, Social Networks.

I. INTRODUCTION

Mobile Ad Hoc Networks (MANETs) has become one of the most prevalent areas of research in the recent years because of the challenges it pose to the related protocols. MANET is the new emerging technology which enables users to communicate without any physical infrastructure regardless of their geographical location, that's why it is sometimes referred to as an ―infrastructure less network.

The proliferation of cheaper, small and more powerful devices make MANET a fastest growing network. An ad-hoc network is self-organizing and adaptive.

Device in mobile ad hoc network should be able to detect the presence of other devices and perform necessary set up to facilitate communication and sharing of data and service. Ad hoc networking allows the devices to maintain connections to the network as well as easily adding and removing devices to and from the network. Due to nodal mobility, the network topology may change rapidly and unpredictably over time. The network is decentralized, where network organization and message delivery must be executed by the nodes themselves. Message routing is a problem in a decentralize environment where the topology fluctuates. While the shortest path from a source to a destination based on a given cost function in a static network is usually the optimal route, this concept is difficult to extend in MANET[1].

The set of applications for MANETs is diverse, ranging from large-scale, mobile, highly dynamic networks, to small, static networks that are constrained by power sources. Besides the legacy applications that move from traditional infrastructure environment into the ad hoc context, a great deal of new services can and will be generated for the new environment. MANET is more vulnerable than wired network due to mobile nodes, threats from compromised nodes inside the network, limited physical security, dynamic topology, scalability and lack of centralized management. Because of these vulnerabilities, MANET is more prone to malicious attacks.

II. MANET VULNERABILITIES

Vulnerability is a weakness in security system. A particular system may be vulnerable to unauthorized data manipulation because the system does not verify a user's identity before allowing data access. MANET is more vulnerable than wired network. Some of the vulnerabilities are as follows:-

A. Lack of Centralized Management

MANET doesn't have a centralized monitor server. The absence of management makes the detection of attacks difficult because it is not east to monitor the traffic in a highly dynamic and large scale ad-hoc network. Lack of centralized management will impede trust management for nodes.

B. Resource Availability

Resource availability is a major issue in MANET. Providing secure communication in such changing environment as well as protection against specific threats and attacks, leads to development of various security schemes and architectures. Collaborative ad-hoc environments also allow implementation of self-organized security mechanism.

C. Scalability

Due to mobility of nodes, scale of ad-hoc network changing all the time. So scalability is a major issue concerning security. Security mechanism should be capable of handling a large network as well as small ones.

D. Cooperativeness

Routing algorithm for MANETs usually assumes that nodes are cooperative and non-malicious. As a result a malicious attacker can easily become an important routing agent and disrupt network operation by disobeying the protocol specifications.

E. Dynamic Topology

Dynamic topology and changeable nodes membership may disturb the trust relationship among nodes. The trust may also be disturbed if some nodes are detected as compromised. This dynamic behavior could be better protected with distributed and adaptive security mechanisms.

F. Limited Power Supply

The nodes in mobile ad-hoc network need to consider restricted power supply, which will cause several problems. A node in mobile ad-hoc network may behave in a selfish manner when it is finding that there is only limited power supply.

G. Bandwidth Constraint

Variable low-capacity links exists as compared to wireless network which are more susceptible to external noise, interference and signal attenuation effects.

H. Adversary Inside the Network

The mobile nodes within the MANET can freely join and leave the network. The nodes within network may also behave maliciously. This is hard to detect that the behavior of the node is malicious. Thus this attack is more dangerous than the external attack. These nodes are called compromised nodes.

I. No Predefined Boundary

In mobile ad- hoc networks we cannot precisely define a physical boundary of the network. The nodes work in a nomadic environment where they are allowed to join and leave the wireless network [2].

III. SECURITY GOALS

Security involves a set of investments that are adequately funded. In MANET, all networking functions such as routing and packet forwarding, are performed by nodes themselves in a self-organizing manner. For these reasons, securing a mobile ad -hoc network is very challenging [3].

The goals to evaluate if mobile ad-hoc network is secure or not are as follows:

1) *Availability:* Availability means the assets are accessible to authorized parties at appropriate times. Availability applies both to data and to services. It ensures the survivability of network service despite denial of service attack.
2) *Confidentiality:* Confidentiality ensures that computer-related assets are accessed only by authorized parties. That is, only those who should have access to something will actually get that access. To maintain confidentiality of some confidential information, we need to keep them secret from all entities that do not have privilege to access them. Confidentiality is sometimes called secrecy or privacy.
3) *Integrity:* Integrity means that assets can be modified only by authorized parties or only in authorized way. Modification includes writing, changing status, deleting and creating. Integrity assures that a message being transferred is never corrupted.

4) *Authentication:* Authentication enables a node to ensure the identity of peer node it is communicating with. Authentication is essentially assurance that participants in communication are authenticated and not impersonators. Authenticity is ensured because only the legitimate sender can produce a message that will decrypt properly with the shared key.
5) *Non Repudiation:* Non repudiation ensures that sender and receiver of a message cannot disavow that they have ever sent or received such a message .This is helpful when we need to discriminate if a node with some undesired function is compromised or not.
6) *Anonymity:* Anonymity means all information that can be used to identify owner or current user of node should default be kept private and not be distributed by node itself or the system software.
7) *Authorization:* This property assigns different access rights to different types of users. For example a network management can be performed by network administrator only.

IV. ATTACKS IN MANET

Securing wireless ad-hoc networks is a highly challenging issue. Understanding possible form of attacks is always the first step towards developing good security solutions. Security of communication in MANET is important for secure transmission of information. Absence of any central co-ordination mechanism and shared wireless medium makes MANET more vulnerable to digital/cyber-attacks than wired network there are a number of attacks that affect MANET. These attacks can be classified into two types:

A. *External Attack*

External attacks are carried out by nodes that do not belong to the network. It causes congestion sends false routing information or causes unavailability of services.

B. *Internal Attack*

Internal attacks are from compromised nodes that are part of the network. In an internal attack the malicious node from the network gains unauthorized access and impersonates as a genuine node. It can analyze traffic between other nodes and may participate in other network activities.

C. *Impersonation*

If the authentication mechanism is not properly implemented a malicious node can act as a genuine node and monitor the network traffic. It can also send fake routing packets, and gain access to some confidential information.

D. *Eavesdropping*

This is a passive attack. The node simply observes the confidential information. This information can be later used by the malicious node. The secret information like location, public key, private key, password etc. can be fetched by eavesdropper.

E. *Routing Attacks*

The malicious node make routing services a target because it's an important service in MANETs. There are two flavors to this routing attack. One is attack on routing protocol and another is attack on packet forwarding or delivery mechanism. The first is aimed at blocking the propagation of routing information to a node. The latter is aimed at disturbing the packet delivery against a predefined path.

V. COMPONENTS OF MOBILE AD HOC NETWORKS

By leveraging the properties of social networks, social network-based P2P content-based file sharing is proposed in disconnected mobile adhoc Networks (SPOON) with four components as shown in Figure 1.

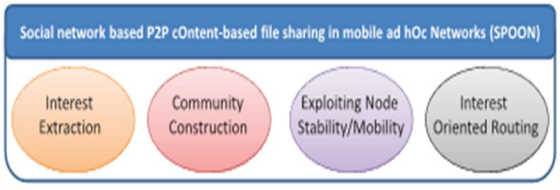


Fig 1: Components of SPOON

1) Based on P2, an interest extraction algorithm is proposed to derive a node's interests from its files. The interest facilitates queries in content-based file sharing and other components of SPOON.

2) A collective of nodes is referred that share common interests and meet frequently as a community. According to P3, a node has high probability to find interested files in its community. If this fails, based on P1, the node can rely on nodes that frequently travel to other communities for file searching. Thus, the community construction algorithm is proposed to build communities to enable efficient file retrieval.

3) According to P1, a node role assignment algorithm is proposed that takes advantage of node mobility for efficient file searching. The algorithm designates a stable node that has the tightest connections with others in its community as the community coordinator to guide intra community searching. For each known foreign community, a node that frequently travels to it is designated as the community ambassador for intercommunity searching.

4) An interest-oriented file searching and retrieval scheme is proposed that utilizes an interest-oriented routing algorithm (IRA) and three of the above components. Based on P3, IRA selects forwarding node by considering the probability of meeting interest keywords rather than nodes. The file searching scheme has two phases: Intra- and intercommunity searching. In the former, a node first queries nearby nodes, then relies on coordinator to search the entire home community. If it fails, the intercommunity searching uses an ambassador to send the query to a matched foreign community. A discovered file is sent back through the search path or the IRA if the path breaks.

SPOON is novel in that it leverages social network properties of both node interest and movement pattern. First, it classifies common-interest and frequently encountered nodes into social communities. Second, it considers the frequency at which a node meets different interests rather than different nodes in file searching[4][5].

VI. THE P2P FILE SHARING MODEL

The P2P file sharing model makes large-scale networks a blessing instead of a curse, in which nodes share files directly with each other without a centralized server. Wired P2P file sharing systems have already become a popular and successful paradigm for file sharing among millions of users. The successful deployment of P2P file sharing systems and the aforementioned impediments to file sharing in MANETs make the P2P file sharing over MANETs (P2P MANETs in short) a promising complement to current infrastructure model to realize pervasive file sharing for mobile users[6][7].

As the mobile digital devices are carried by people that usually belong to certain social relationships, in this paper, we focus on the P2P file sharing in a disconnected MANET community consisting of mobile users with social network properties. In such a file sharing system, nodes meet and exchange requests and files in the format of text, short videos, and voice clips in different interest categories[8].

SPOON is novel in that it leverages social network properties of both node interest and movement pattern. First, it classifies common-interest and frequently encountered nodes into social communities. Second, it considers the frequency at which a node meets different interests rather than different nodes in file searching. Third, it chooses stable nodes in a community as coordinators and highly mobile nodes that travel frequently to foreign communities as ambassadors. Such a structure ensures that a query can be forwarded to the community of the queried file quickly [9][10].

A. Trace Data Analysis

The proposed system Tt is used to denote the time length of the trace, and define the total meeting time of two nodes as the sum of the time length of each encountering. By regarding a community as a group of nodes in which each node has total meeting time larger than Tt=4 with at least half of all nodes in the community, we detected eight communities from the trace. Then each node's average number of shared interested tracks with other members is calculated in its own community Ci (0 □□ i< 8), and with nodes in all other communities, respectively. Finally, the average values of all nodes in each community are calculated and shown in Table 6.1.

From the table, it is seen that for each community, nodes have higher average number of shared interested tracks with same community nodes than with nodes from other communities. Note that we used a relatively loose community creation requirement that each node only needs to have a high contact frequency with half of nodes in a community. With a stricter requirement and a more sophisticated clustering method, nodes in the same community would share more interested tracks. Above traces verify the previously observed social properties and support the basis for SPOON that nodes with common interests tend to meet frequently.

Table 6.1 Average Number Of Shared Interested Tracks

Community Ci	Ave. # of shared interests with nodes in Ci	Ave. # of shared interests with nodes not in Ci
1	1.50	0.99
2	0.83	0.69
3	1.17	0.79
4	1	0.39
5	1.93	0.94
6	0.33	0.21
7	1.1	0.71
8	1	0.33

B. Interest Extraction

Without loss of generality, it is assumed that node contents can be classified to different interest categories. It was found that users usually have a few file categories that they query for files frequently in a file sharing system. Specifically, for the majority of users, 80 percent of their shared files fall into only 20 percent of total file categories. Like other file sharing systems [8], it is considered that a node's stored files can reflect its file interests. Thus, SPOON derives the interests of a node from its files. Table 2 lists the notations used in this section. To derive its interests, a node infers keywords from each of its files using the document clustering technique. Specifically, a node derives a file vector for each of its files from its metadata. For file fi, we denote its file vector by vi =(t1,wit1, t2,wit2, t3, wit3 ,..., tm,witm), in which tk and wik (1 □□ k □□m) denote a keyword and its weight that represents the importance of the keyword in describing the file.

Here formula used is

$$w_{it_k} = 1 + log(n_{t_k}), \quad \textbf{(1)}$$

$$w_{it_k} = w_{it_k} \Big/ \sum_{q=1}^{m} w_{it_q}. \quad \textbf{(2)}$$

Then the following formula is used to calculate the similarity between v1 and v2:

$$sim(v_1, v_2) = \frac{\sum_{k=1}^{m'} w_{1_k} * w_{2_k}}{\sqrt{\sum_{k=1}^{m'} w_{1_k}^2} * \sqrt{\sum_{k=1}^{m'} w_{2_k}^2}}, \quad \textbf{(3)}$$

where m0 is the total number of common keyword and w1k and w2k represent the weights of the kth common keyword of the two vectors, respectively.

After retrieving the file vector of each of its files, a node classifies its files to derive its interest groups[11].

C. Community Construction

Social network theory reveals that people with the same interest tend to meet frequently [48]. By exploiting this property, SPOON classifies nodes with common interests and frequent contacts into a community to facilitate interest-based file searching, as introduced in Algorithm 1 and 2. Nodes with multiple interests belong to multiple communities. The community construction can easily be conducted in a centralized manner by collecting node interests and contact frequencies from all nodes to a central node. However, considering that the proposed system is for distributed disconnected MANETs, in which timely information collection and distribution is nontrivial, we further propose a decentralized method to ensure the adaptivity of SPOON in real environment.

When two nodes, say N1 and N2, meet, they consider two cases for community creation: 1) they do not belong to any communities, and 2) at least one of them is already a member of a community. In the first case, they calculate the similarity between each pair of their interest vectors using (3). A pair of interest groups, say Gi and Gj with interest vectors vi and vj, is called matched interest group when W(Gi)W(Gj) im(vi, vj) TG,where TG is a predefined threshold. The purpose of taking into account the weight of each interest group is to eliminate the noise of interest groups with a small number of files and achieve better interest clustering. If N1 and N2 have at least one pair of matched interest group, and their contact frequency, F(N1,N2), is higher than the top h_1 percent highest encountering frequencies in either node, the two nodes form a new community. The keywords in their matched interest groups and corresponding weights constitute the community vector (vC) of the community[12].

D. Catch Memory and Web services Methods

Fixed number of servers, the Bass diffusion model suggests that as the number of interested users increases, the delay in service perceived by users should first increase, and then gradually decrease.

Informally, the servers would be over-whelmed for a period of time before being able to cope with the "crowd"—a characteristic of the so called "flash crowd" effect. It is important to note that even if we reduce the time-scale of events by a constant factor (so that interest grows more steeply), the order of magnitude of the delays (relative to) would be unchanged.

Exact performance will depend on the exogenous constants parameterizing the system, such as installed server capacity

The file sharing efficiency is increased using cache concept both in server and client nodes. This method supports larger and more disconnected networks environments also.

Input: search_in_file, client_node, server_node

Output: Search_out_file

Get search_in_file←in_file

If search_in_file == client_node then

Return search_out_file

Else if

Return 0

Else

Search_out_file←server_node End End End

VII. CONCLUSION

The difficulty in distributing the content in the server is eliminated by using this application. It reduces the server bandwidth to consistent amount. The end users need not wait for server in downloading the content since the P2P application gets the content from available clients.

A good documentation of user-friendly features had been incorporated in the system.

1) The system has been introduced to eliminate human error.

2) To minimize the time consumption and design & development work.

REFERENCES

[1] Y. Huang, Y. Gao, K. Nahrstedt, and W. He, "Optimizing File Retrieval in Delay-Tolerant Content Distribution Community," Proc. IEEE 29th Int'l Conf. Distributed Computing Systems (ICDCS '09), 2009.

[2] J. Reich and A. Chaintreau, "The Age of Impatience: Optimal Replication Schemes for Opportunistic Networks," Proc. Fifth Int'l Conf. Emerging Networking Experiments and Technologies (CoNEXT '09), 2009.

[3] V. Lenders, M. May, G. Karlsson, and C. Wacha, "Wireless Ad Hoc Podcasting," ACM SIGMOBILE Mobile Computing and Comm. Rev., vol. 12, pp. 65-67, 2008.

[4] F. Li and J. Wu, "MOPS: Providing Content-Based Service in Disruption-Tolerant Networks," Proc. IEEE 29th Int'l Conf. Distributed Computing Systems (ICDCS '09), 2009.

[5] R. Zhou, K. Hwang, and M. Cai, "Gossiptrust for Fast Reputation Aggregation in Peer-2-Peer Networks," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 9, pp. 1282- 1295, Sept. 2008.

[6] C. Boldrini, M. Conti, and A. Passarella, "ContentPlace: Social-Aware Data Dissemination in Opportunistic Networks," Proc. 11th Int'l Symp. Modeling, Analysis and Simulation Wireless and Mobile Systems (MSWiM '08), 2008.

[7] M. Papadopouli and H. Schulzrinne, "A Performance Analysis of 7DS: A Peer-to-Peer Data Dissemination and Prefetching Tool for Mobile Users," Proc. IEEE Sarnoff Symp. Digest Advances in Wired and Wireless Comm., 2001.

[8] J.B. Tchakarov and N.H. Vaidya, "Efficient Content Location in Wireless Ad Hoc Networks," Proc. IEEE Int'l Conf. Mobile Data Management (MDM '04), 2004.

[9] A. Iamnitchi, M. Ripeanu, E. Santos-Neto, and I. Foster, "The Small World of File Sharing," IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 7, pp. 1120-1134, July 2011.

[10] H. Schu¨tze and C. Silverstein, "Projections for Efficient Document Clustering," Proc. 20th Ann. Int'l ACM Conf. Research and Development in Information Retrieval (SIGIR '07), pp. 74-81, 1997.

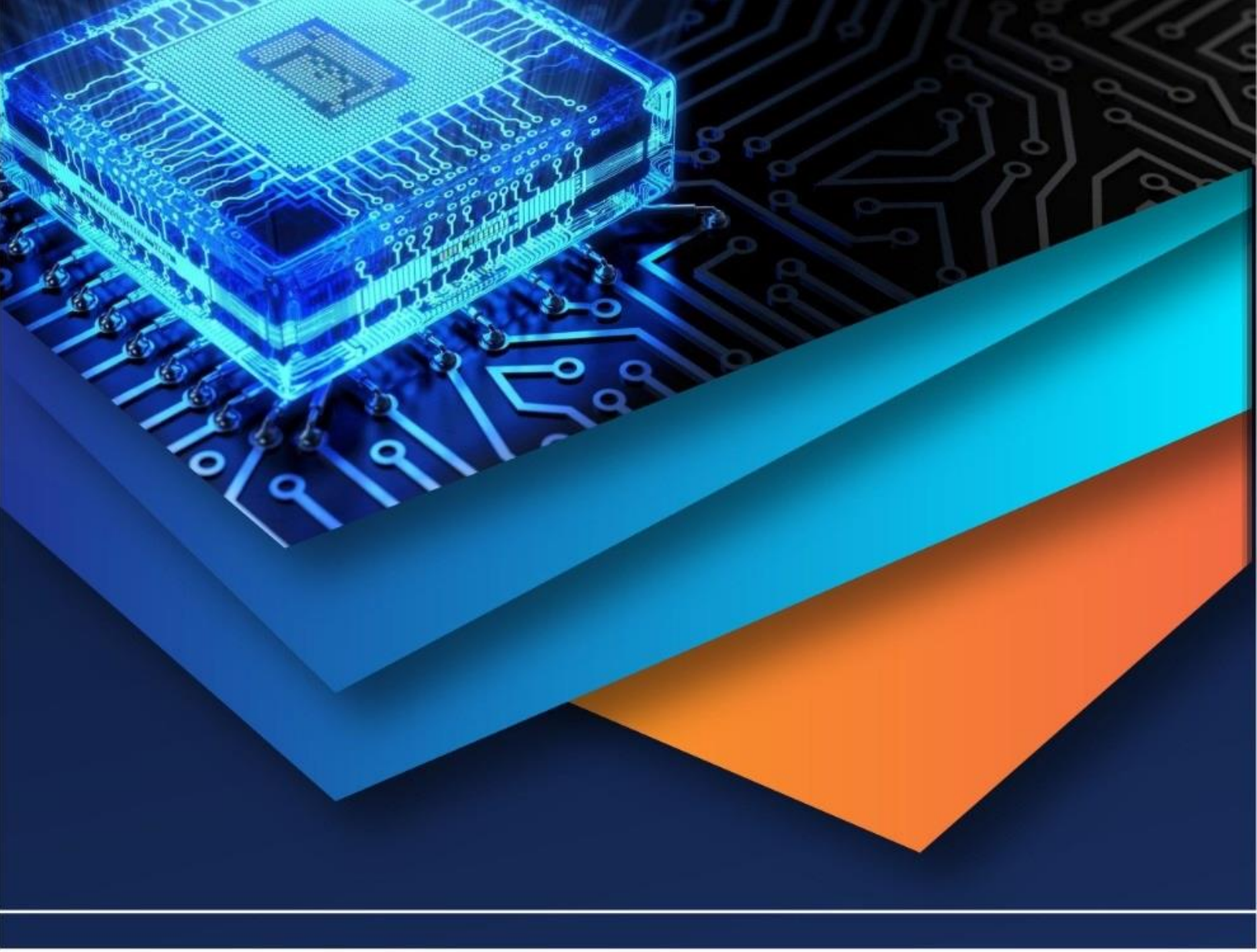
[11] P. Bonacich, "Factoring and Weighting Approaches to Status Scores and Clique Identification," J. Math. Sociology, vol. 2, pp. 113-120, 1972.

[12] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient estimation of word representations in vector space", 2013.

[13] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar et al., "Universal sentence encoder", 2018.

[14] B. Barz and J. Denzler, "Hierarchy-based image embeddings for semantic imagretrieval", *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 638-647, 2019.

[15] Y. A. Malkov and D. A. Yashunin, "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs", *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 42, no. 4, pp. 824-836, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY