

Mathematical Model Validation of Search Protocols in MP2P Networks

Ajay Arunachalam^{1,*}, Vinayakumar Ravi², Moez Krichen³, Roobaea Alroobaea⁴ and Saeed Rubaiee⁵

¹Centre for Applied Autonomous Sensor Systems (AASS), Örebro University, Örebro, Sweden

²Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

³Faculty of CSIT, Al-Baha University, Saudi Arabia ReDCAD Laboratory, University of Sfax, Tunisia

⁴Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif, 21944, Saudi Arabia

⁵Department of Industrial and Systems Engineering, College of Engineering, University of Jeddah, Jeddah, Saudi Arabia

*Corresponding Author: Ajay Arunachalam. Email: ajay.arunachalam08@gmail.com

Received: 08 January 2021; Accepted: 08 February 2021

Abstract: Broadcasting is a basic technique in Mobile ad-hoc network (MANET), and it refers to sending a packet from one node to every other node within the transmission range. Flooding is a type of broadcast where the received packet is retransmitted once by every node. The naive flooding technique, floods the network with query messages, while the random walk technique operates by contacting the subsets of every node's neighbors at each step, thereby restricting the search space. One of the key challenges in an ad-hoc network is the resource or content discovery problem which is about locating the queried resource. Many earlier works have mainly focused on the simulation-based analysis of flooding, and its variants under a wired network. Although, there have been some empirical studies in peer-to-peer (P2P) networks, the analytical results are still lacking, especially in the context of P2P systems running over MANET. In this paper, we describe how P2P resource discovery protocols perform badly over MANETs. To address the limitations, we propose a new protocol named ABRW (Address Broadcast Random Walk), which is a lightweight search approach, designed considering the underlay topology aimed to better suit the unstructured architecture. We provide the mathematical model, measuring the performance of our proposed search scheme with different widely popular benchmarked search techniques. Further, we also derive three relevant search performance metrics, i.e., mean no. of steps needed to find a resource, the probability of finding a resource, and the mean no. of message overhead. We validated the analytical expressions through simulations. The simulation results closely matched with our analytical model, justifying our findings. Our proposed search algorithm under such highly dynamic self-evolving networks performed better, as it reduced the



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

search latency, decreased the overall message overhead, and still equally had a good success rate.

Keywords: Search protocols; random walk; MANET; P2P networks; P2P MANET; mathematical model; peer-to-peer; wireless *ad hoc* network; flooding; mobile P2P; unstructured; P2P; NS-2; network simulator

Abbreviations

DHT:	Distributed Hash Table
CAN:	Content Addressable Network
NS-2:	Network Simulator-2
P2P:	Peer-to-Peer
MANET:	Mobile <i>Ad hoc</i> Network
RW:	Random Walk
TTL:	Time To Live

1 Introduction

MANETs are dynamic mobile ad-hoc wireless networks that use multi-hop routing. The nodes in such networks can communicate using layer-3 (Network Layer) routing in case they are not connected at layer-2 (Data Link Layer) directly. The resource discovery process also referred to as content discovery/resource search is very challenging in such networks as there is a continuous movement of nodes. It is the process to find or locate information resources/objects. Previously, traditional search techniques like random walk and flooding were extensively employed for the resource discovery process. In flooding, the source node transmits the packet to all the other nodes in the network. Contrary to this, the packet is randomly transmitted to a few nodes in the network in the random walk approach. Although, both approaches have some disadvantages, they are used in MANETs as it suits the self-organizing nature of the network. Several previous research works have studied the effectiveness of peer-to-peer (P2P) resource discovery approaches for wired networks. The effectiveness of several content searching techniques is tested for the P2P network. But, due to problems related to energy consumption, mobility, infrastructure deficiency, and churn their performance is not validated against MANETs. The design of the resource discovery technique should be based on the characteristics of the environment where it is intended to work. In short, a resource discovery approach intended for wired context may not always scale well in a wireless environment and vice versa. To design an effective resource discovery scheme, one must consider the following features as given in [Tab. 1](#).

Service discovery in MANETs is also another challenging issue. Service discovery in layman terms is the automatic detection of devices, and services offered by these devices on a network. Authors in [1] studied the effectiveness of the random walk approach for MANETs and they have discussed its issues like unicast transmission disadvantage, multiple query walk, valid termination check parameter, next-hop selection criteria, etc. A brief survey of various searching techniques is given in [2]. P2P networks are very common, and they constitute a majority of internet traffic [3]. They are known for their flexibility and distributive network properties and P2P based systems are employed over the internet for services like video conferencing applications, torrent applications, information retrieval systems, telephony, file sharing, etc. A P2P network can be classified based on the structural and behavioral characteristics as purely decentralized architecture (which can be structured or unstructured) and hybrid architecture (that can be centralized indexing

or decentralized indexing). In structured P2P systems the arrangement of the peers are based on strict well organized rules, while in the unstructured P2P systems the peers connect to each other in a random fashion. The hybrid architecture is a mixture of peer-to-peer and client-server models. The centralized indexed hybrid P2P systems maintain a central index server, wherein the files and user information are stored in directory-based order. Each peer maintains a connection to the index server. In the decentralized indexed hybrid P2P systems there is no centralized index server, but here some nodes are given a more important role than the others called the super nodes, which maintain the indexes for the information shared by their local connections. The super nodes themselves are connected through a pure P2P network. We compare the different architectures of peer-to-peer systems based on features like scalability, flexibility, robustness, and manageability. We provide a schematic analysis as shown in [Tab. 2](#). The unstructured pure P2P system where the blind search techniques are used is not scalable, due to the overwhelming message exchanges. While, in a structured system, one can overcome the scalability limitation with intelligent DHT-based search strategies. The other important feature in a pure P2P system is the joining and leaving of the peers, i.e., the flexibility. The unstructured systems lack manageability as each peer is its own controller. But then such systems are robust since failure of any particular node doesn't impact the overall system.

Table 1: Features required for the resource discovery method and their definition

Feature	Definition
Robust	The ability to cope up well with the change in network conditions.
Scalable	How effectively the resource discovery method can handle the growth in network size.
Fault-tolerance	Ability of the resource search scheme to operate properly even in case of an event failure.
Reliable	Ability of the method to be overall good in any network condition.
Dynamicity	Ability of the search method to cope up with node movement and signal attenuation.
Adaptability	It is the ability of the resource discovery technique to manage network dynamics and network load.
False positive error	Ability of the search technique to exactly locate the resource if available.
Security	It is the critical aspect of the MP2P network that includes denial of different attacks incurred due to the injection of malicious nodes, authorization, authentication etc.
Load balancing	It is the fair distribution of workload across the nodes in the network intended to improve maximum utilization of the resources and lower node overhead while utilizing the network and mobile resources effectively.

Table 2: P2P architecture comparison

Characteristics	Unstructured P2P	Structured P2P	Centralized hybrid P2P	Decentralized hybrid P2P
Scalable	N	Y	N	Y
Flexible	Y	N	N	Y
Robust	Y	Y	N	Y
Manageable	N	Y	Y	Y

Centralized architecture is used for most mobile communication even today and the bottleneck issue created by them can be overcome by employing the P2P platform [4]. This P2P based approach suits wired and wireless networks including MANETs. Combining P2P network properties in mobile *ad hoc* networks is coined as P2P MANETs or Mobile P2P networks. The first mobile peer-to-peer service was a dating client. The main idea was that people with the same interest in meeting other people could specify their interests and needs by using a mobile application. Recently the trend is shifting towards social mobile networking and cooperative networking which are being implemented in this field. The MP2P encompasses different architectures such as point to point, meshed networks, and cooperative networks. Some of the key research areas for MP2P networks include data retrieval, service discovery, caching, data dissemination, query processing, routing, network architectures, resource discovery, etc. Efficient P2P application and schemes for cellular environment is still an open research area. Due to similarities in P2P and MANETs, a P2P overlay can easily be formed over Mobile *Ad hoc* Network. An overlay network is formed by communication between the peers wherein each link of overlay corresponds to a path in an underlay physical network. But, at the same time, their direct combination also poses difficulties due to differences in the operating layer, transmission mechanism, and rapid mobility in MANETs [5–7].

To clearly understand these networks (P2P, MANETs, P2P MANETs) we compare their similarities and differences. We summarize the differences and similarities of these networks in detail as given in [Tabs. 3](#) and [4](#) respectively.

Table 3: Differences between P2P, MANET and MP2P network

Differences	Peer-to-Peer	MANET	MP2P
Nodes connection	Wired and direct	Wireless, multihop using relay nodes (indirect communication)	Wireless, peers and <i>ad hoc</i> nodes (indirect communication)
Network reliability	High	Low	Low
Frequent disconnection	Low	High	High
Structure	Virtual overlay and physical structure are separate	Logical and physical structure corresponds	Overlay network resembles physical underlay network
Operative layer	Application layer	Network layer	Cross-layered
Transmission mechanism	Physical unicast	Physical broadcast	Virtual broadcast over physical broadcast (broadcast over broadcast)
Routing	Continues until TTL is 0	Continues until the destination is found	Mixed mode (layered, cross-layered, integrated design)

Table 4: Similarities between P2P, MANET and MP2P network

Similarities	Peer-to-Peer	MANET	MP2P
Topology	Flat, dynamic changing topology	Flat and dynamic changing topology	Flat, hierarchical and dynamic changing topology
Basic architecture	Decentralized	Decentralized	Decentralized, hybrid
Establishment of connection	Hop-by-hop via TCP links	Hop-by-hop via radio links	Hop-by-hop via radio links
Range of connection	Unlimited	Limited by communication range of the node	Limited by communication range of the node
Node reliability	Low	Low	Low
Basic routing mechanism	Physical unicast, flooding	Physical broadcast, flooding	Broadcast, broadcast over broadcast, flooding

In such a distributed environment, resource discovery is a key issue. Unstructured P2P networks mostly rely on the flooding technique. As a preliminary study, we evaluated the unstructured P2P searching techniques over MANET [8,9]. According to our study, the pure random walk approach consumes more battery power, increases network overhead, has high search latency, and lowered hit rate when compared to other unstructured protocols. It is essential to take the underlay topology into consideration to improve the efficiency of the random walk protocol over the mobile *ad hoc* network [10]. To reduce the overhead in the classic random walk protocol, we proposed cross-layered 1-hop addressed resource discovery architecture that is more suitable for P2P-MANETs as shown in Fig. 1.

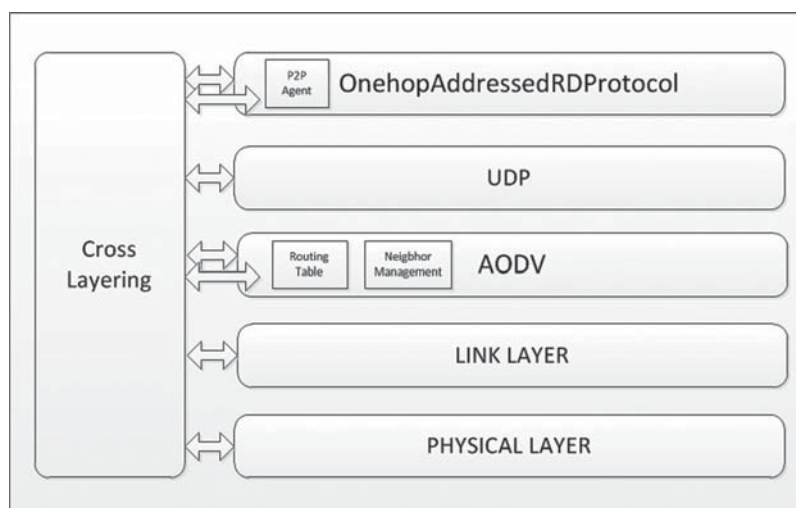


Figure 1: Cross-layered addressed search architecture

For resolving the node neighbors & its density, we proposed a method for P2P application over MANETs, and the source code for which is publicly available in SourceForge¹ & the corresponding P2P network resource discovery simulation over MANETs with our proposed node discovery technique is made publicly available in GitHub².

The rest of this paper is structured as follows. In Section 2, we review the existing work related to analytical based analysis for P2P and MANETs. Section 3 discusses the problem statement of mathematical modeling under mobile P2P networks. Analysis of flooding-based search procedures, and our proposed scheme are introduced in Section 4, and their corresponding performance measures are derived. In Section 5, we summarize our theoretical results, along with simulation-based validation, and finally, conclude in Section 6.

2 Related Works

In this section, we discuss the various relevant works that estimate the performance of resource discovery protocols analytically. Although there are many empirical studies and simulation-based analyses of flooding-based schemes and its variants, the mathematical results are still lacking especially in the context of P2P networks over MANET. Quickly and efficiently locating a resource is a major issue in unstructured and decentralized P2P networks. Techniques such as probabilistic forwarding, random walk, and flooding are extensively used in such unstructured architecture. Barjini et al. [11] analytically review many search techniques that are based on a flooding approach for an unstructured P2P network. They measure the coverage growth rate and traffic loads for each scheme. They introduce a metric, i.e., the critical value which is the ratio of the number of redundant messages over the coverage growth rate. According to their study and simulation results, the probabilistic limit-based flooding schemes (i.e., teeming, modified breadth-first search, random walk, etc.) have better performance than the TTL limit-based flooding schemes (such as local indices, expanding ring, iterative deepening, etc.). Bisnik et al. [12] present an analytical expression to measure the performance metrics of random walk protocol in unstructured P2P networks in terms of TTL of walkers, count of walkers, and demand of the resource. Their work focuses on a wired P2P network. Dimakopoulos et al. [13] study the performance of teeming, random path, and flooding search schemes, in P2P systems. Their study considers two scenarios, i.e., resource requests in the presence of popular and unpopular resources. Further, they assume that a cache is used to store the details of the resources and their corresponding resource providers at each node. To overcome the drawbacks observed in flooding and random walk techniques, Lin et al. [14] propose a technique that is the combination of both schemes. In the context of searching, it performs flooding for a short-term search and follows the random walk technique for a long-term search. They use Newman's work [15] for the random graph and adopt generating functions to model distribution degrees. An alternative to the flooding scheme is the gossip protocol, where every node transmits the message to a subset of its neighbors in a random fashion based on some probability. In [16], costs of Gnutella's flooding-based broadcasting and the classic gossip protocol are studied by varying network size and the average number of neighbors. Further, they calculate the bandwidth required for each node while using the flooding technique. They propose a deterministic gossip-based protocol and compare their performance with the flooding technique. Ferretti [17] proposes a mathematical model for the gossip-based resource discovery protocol in unstructured P2P overlay networks.

¹ <https://sourceforge.net/projects/neighbordiscovery/>

² <https://github.com/ajayarunachalam/Neighbor-Discovery>

Their analysis is based on complex network theory. They analyze and evaluate the count of nodes receiving the query and the amount of query hits. Specific work to model and optimize random walk protocol over MANET is presented in [18]. They modeled their method using a queuing system called G-network which has positive & negative kinds of customers. Further, they used the gradient descent method to optimize their method which uses three parameters such as consumption of energy, response time, and hit rate as part of the cost function. In the above work, they justify their findings with theoretical results but haven't simulated and evaluated the protocol in a realistic environment. Most of the aforementioned works are evaluated over a wired and fixed scenario. Further, in the case of analysis and evaluation of P2P content searching techniques under MANET, there is a lack of empirical study that considers the underlay network topology. Most of the earlier contributions are limited to the experimental/simulation based algorithmic evaluations only. This was the primary motivation behind the inception of this study especially in context to P2P MANETs. In this paper, unstructured P2P resource discovery protocols under the MANET are modeled by applying the probability and queuing theory concept. Further, our theoretical results are also validated through simulation.

3 Mathematical Modeling

An Overlay Graph $OG = (V, E)$ is used to represent a P2P overlay network, where V is the count of nodes, and E is the count of links. The vertices and the edges of this graph represent the participating peers and virtual links, respectively. Therefore, OG can be explicitly defined as, $OG = (V, E)$ where $|V| = p$ and $link(i, j) \in E$ if i is a neighbour of j , and vice versa, and p is the count of peers. The underlay structure is formed by ad-hoc and peer nodes. In MANETs, the P2P platform is implemented as an overlay network at the application layer, formed by communications between the peers wherein each link of the overlay is supported by a path in the underlay physical network. In our analysis, the underlay structure is considered as a connected graph $CG = (V, E)$ where $|V| = WN$ and $link(i, j) \in E$ if and only if $D_{ij} \leq T_x$ where WN is the count of wireless nodes, and D_{ij} is the Euclidean distance between them. If the Euclidean distance between 2 nodes is less than T_x (transmission range of the node), then it means that they are in the transmission range of each other. Fig. 2. depicts the above explanation, where each virtual link of the overlay network is supported by a path in the underlying physical network.

We consider an underlay network of N nodes with varying degrees, i.e., each node has a different number of d neighbors. Our model focuses on a flat topology without any hierarchy, i.e., not requiring any super node or caching mechanism. To simplify the model, we assume that there is only one type of resource x with one node holding it, and there is only one querying node. But this situation may be more complex where the network can have different types of resources with many nodes holding it and many querying nodes. In short, the resource discovery process is how a node S that needs x finds the respective node that holds the resource.

We assume a random network topology of N mobile nodes where only one node provides the resource x , and there is a single node S searching for that resource. Each node knows its d neighbors. We use the work done in [13] as the starting point for further investigation. Tab. 5. contains the summary of the notations along with their descriptions that will be used in the model.

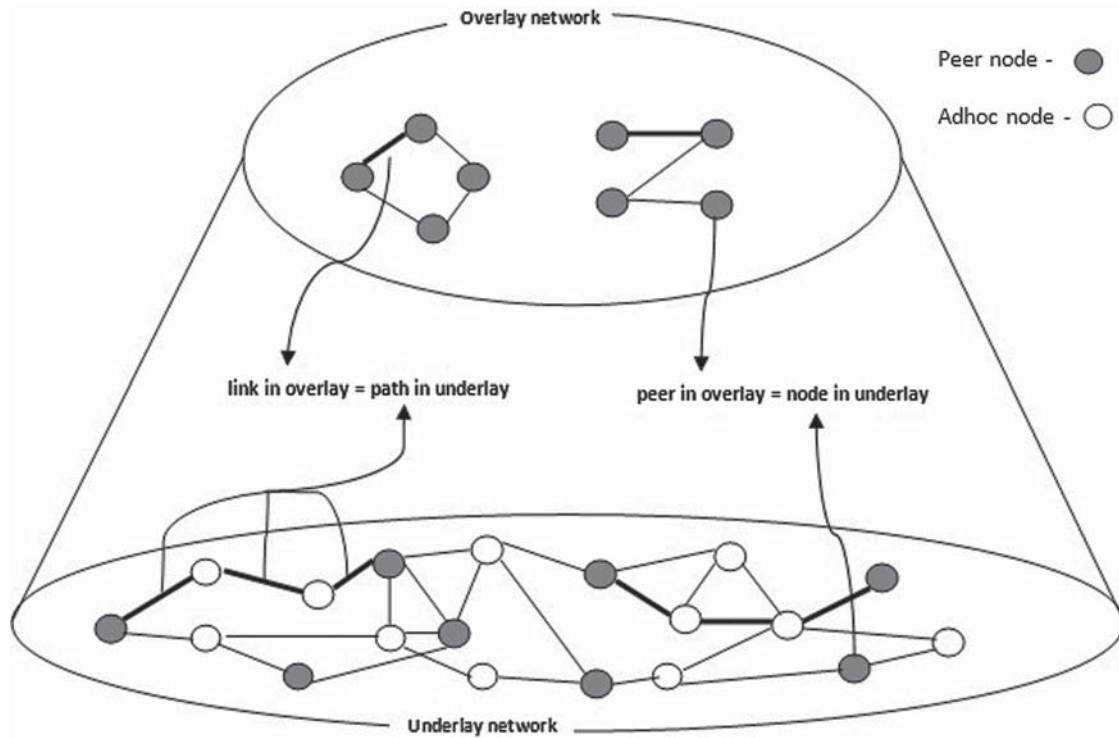


Figure 2: P2P overlay formations

Table 5: Notations used & derived

N	is the number of nodes in the network
d	is the average number of neighbors
x	is the resource to be searched
t	is the maximum number of search steps
F	is the probability that a resource is known to a given node
$a = 1 - F$	is the probability that a resource is not known to a given node
\underline{Q}_t	is the probability of finding resource within t steps
\overline{S}_t	is the average number of steps needed to find a resource
\overline{M}_t	is the average number of generated messages

To limit the broadcast storm, the search is bounded to the maximum t number of steps (i.e., TTL value). Let us assume that F is the probability of resource x being known to a specific node. A node can know a resource only when it holds that resource.

$$F = 1 - P[\text{the node doesn't hold } x]$$

Let n_x be the count of nodes offering the resource “ x ”. Since a resource x is offered by only one node, we have $n_x = 1$.

$$P[\text{node holds } x] = \frac{n_x}{N} = \frac{1}{N} \quad (1)$$

$$F = 1 - \left[1 - \left(\frac{n_x}{N} \right) \right] = 1 - \left[\frac{N - n_x}{N} \right] = 1 - [a], \quad \text{where } a = \frac{N - n_x}{N} \quad (2)$$

Let a be the probability that a node is not aware of the resource. So, $a = 1 - F$. Let Q_t be the probability of finding a resource within t steps. We assume that within t steps, a resource can be located. Hence, the mean of steps needed to find the resource x is given as,

$$\bar{S}_t = t - \frac{1}{Q_t} \sum_{i=0}^{t-1} Q_i \quad (3)$$

The derivation of Eq. (3) is given at the end of the section. In the equation, Q_t depends on the searching strategy. We consider flooding, random walk and our proposed addressed random walk methods which are depicted as a d-ary tree that unfolds as the search progresses, shown in Fig. 3. A tree representation may not be perfect for representing such wireless network topology as many nodes may overlap each other's range and, multiple nodes may communicate with a single node. Yet, the tree representation can visualize the search scenario pictorially.

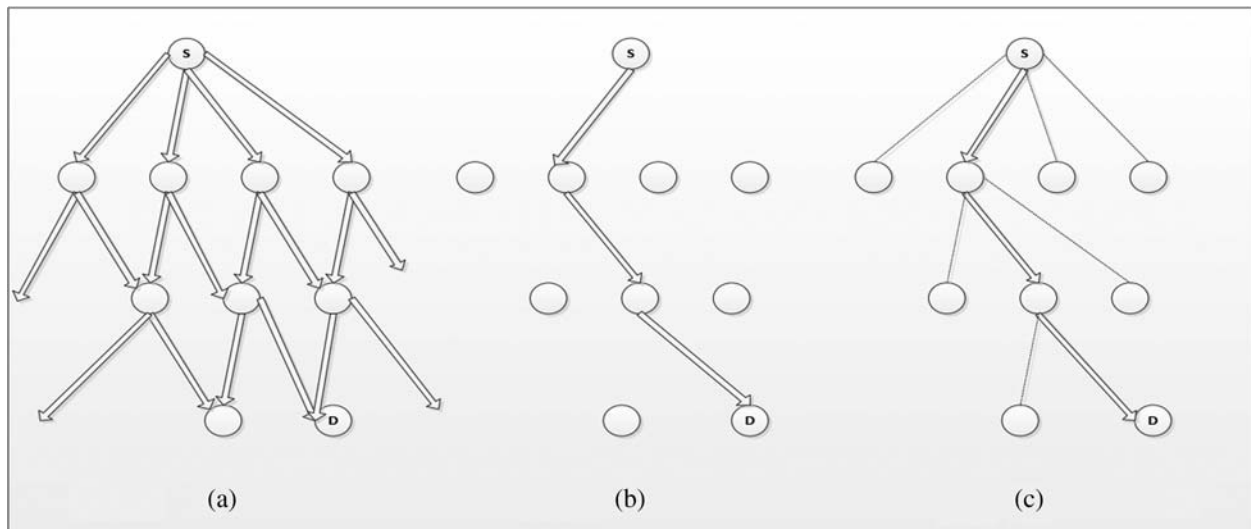


Figure 3: Comparison of different flooding-based search strategies, and our proposed scheme (a) flooding (b) random walk (c) addressed random walk

The focus of this article is on studying many decentralized resource discovery schemes. In flooding, when a resource is required by a node, it will communicate with its neighboring nodes and further those nodes will communicate with their neighbors & so on. This process will be repeated until every node in the network is communicated as shown in Fig. 3a. The flooding method can find a resource without any hierarchy or prior specific information about the system, thus an ideal candidate in such dynamically evolving networks. One of the disadvantages of this pure flooding method is that it increases the network traffic exponentially. Therefore, another variation of flooding called random walk is considered. This method restricts the search space by restricting the node to communicate only one of its neighbors in a random fashion as presented

in Fig. 3b. The limitation of this method is that it takes more steps to find a resource. Both the above-discussed schemes have their drawbacks and limitations. From our preliminary study [4,5], we observe that the random walk scheme under MANET increases the network overhead, battery power consumption, bandwidth usage, search latency, etc. To overcome these issues, we propose a simple and lightweight resource discovery technique to suit well under MANET where the node propagates the inquiring message to exactly one of their neighbors at each hop, but at the same time end up unfolding different virtual paths concurrently, as shown in Fig. 3c.

Derivation of Eq. (3): Let s_i is the probability of finding a resource in i th step exactly. The probability of finding it within $t \geq 1$ steps is formulated as:

$$Q_t = \sum_{i=0}^t s_i$$

This implies that $s_t = Q_t - Q_{t-1}$. Under the assumption that within t steps, the resource is found, the probability of finding it in i th step exactly is given by $\frac{s_i}{Q_t}$ and the mean of steps is given by:

$$\overline{S}_t = \frac{1}{Q_t} \sum_{i=1}^t i s_i = \frac{1}{Q_t} (Q_{t-1} \overline{S}_{t-1} + t s_t)$$

Substituting for s_t , we obtain:

$$Q_t \overline{S}_t = Q_{t-1} \overline{S}_{t-1} + t Q_t - t Q_{t-1}$$

Let $g_t = Q_t \overline{S}_t$, the recurrence takes the following form: $g_t = g_{t-1} + t Q_t - t Q_{t-1}$, with the boundary of $g_0 = S_0 Q_0 = 0$

Evaluating g_t we get,

$$g_t = \sum_{i=1}^t i Q_i - \sum_{i=1}^t i Q_{i-1} = t Q_t - \sum_{i=0}^{t-1} Q_i$$

Since, $g_t = Q_t \overline{S}_t$, Eq. (3). follows,

$$\overline{S}_t = t - \frac{1}{Q_t} \sum_{i=0}^{t-1} Q_i$$

4 Performance Analysis

4.1 Flooding Resource Discovery Algorithm

In flooding, a node that requires the resource will flood the message to its neighbors which in turn retransmits it to their neighbors (Refer Algorithm 1). This will continue until the node that is holding the required resource is found or the TTL expires. For better understanding, the nodes in the network are represented as a d-ary tree. To visualize the situation, refer to Fig. 4., where the resource requesting node S is the root and node D holds the resource being searched. Fig. 4a shows a sample network where flooding technique is used while its corresponding rough d-ary tree

representation is shown in Fig. 4b. Within consecutive search steps, the tree will contact at most d^i different nodes in the i th level.

Algorithm 1: Flooding in P2P applications running over MANETs

Procedure flooding(m) {
 querying node floods the message m
 message m received for the first time
 broadcast(m)
 }

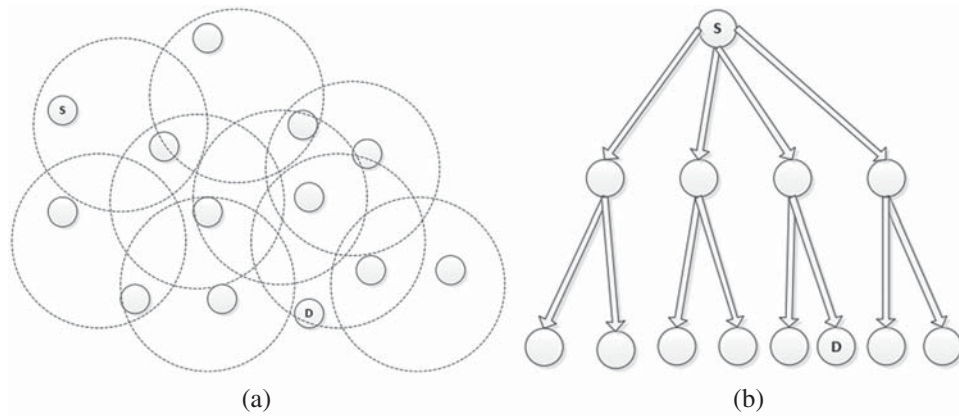


Figure 4: Resource discovery using naïve flooding technique. (a) Searching using broadcast by flooding technique (b) d-ary tree representation

Let $Q_t^{(F)}$ be the probability to find the resource x within t steps for the flooding algorithm. F is the probability that a resource is known to a node and ‘a’ is the probability that a resource is not known to a node. Hence, $a = 1 - F$.

At ‘0’ step, the probability of not locating the resource x is given as,

$$Q_0^{(F)} = 1 - a = F \tag{4}$$

$$1 - Q_0^{(F)} = a \tag{5}$$

$1 - Q_1^{(F)}$ = Probability of not locating the resource within “1” Step

$1 - Q_1^{(F)}$ = Probability of not locating in 0 step * Probability of not finding in Step 1

$$1 - Q_1^{(F)} = a * (1 - Q_0^{(F)})^d$$

Similarly,

$$1 - Q_2^{(F)} = \text{Probability of not locating in 0 steps} * \text{Probability of not locating in Step 1} *$$

Probability of not locating in Step 2

$$1 - Q_2^{(F)} = a * (1 - Q_0^{(F)})^d * (1 - Q_1^{(F)})^d$$

Continuing this way, if we unfold the subsequent inquiring node's neighbors until the t steps, then the probability of not locating resource x is generated as,

$$1 - Q_t^{(F)} = a * (1 - Q_{t-1}^{(F)})^d \quad (6)$$

From Eq. (4), it is clear than the initial boundary condition is $Q_0^{(F)} = 1 - a = F$. The probability Q_{t-1} occurs only if it finds the required resource until the $t-1$ levels of the subtree.

Therefore,

$$Q_t^{(F)} = 1 - a * (1 - Q_{t-1}^{(F)})^d \quad (7)$$

Now replace $1 - Q_t^{(F)}$ as q_t , so Eq. (6) becomes, $q_t = a * (q_{t-1})^d$. Solving further we get,

$$q_t = a^{\frac{d^{t+1}-1}{d-1}}$$

Now evaluating for different values of t ,

$$q_0 = a$$

$$q_1 = a^{d+1}$$

$$q_2 = a \left(a^{d+1} \right)^d = a^{d(d+1)+1}$$

$$q_3 = a \left(a^{d(d+1)+1} \right)^d = a^{(d+1)(d^2+1)}$$

$$q_4 = a^{(d^2+1)(d+1)(d+1)}$$

Therefore, generalized as, $q_t = a^{\frac{d^{t+1}-1}{d-1}}$

Hence Eq. (7) becomes,

$$Q_t^{(F)} = 1 - a^{\frac{d^{t+1}-1}{d-1}} \quad (8)$$

Next, we determine the average number of steps required to locate the resource. In general, the average steps needed are given as in Eq. (3). Now substituting Eqs. (8) in (3), we get the average number of steps required to find the resource within t steps for the flooding algorithm for $t \geq 1$.

$$\overline{S}_t^{(F)} = t - \frac{1}{Q_t^{(F)}} * \sum_{i=1}^{t-1} Q_i^{(F)}$$

$$\begin{aligned} \overline{S_t^{(F)}} &= t - \frac{1}{Q_t^{(F)}} * \sum_{i=1}^{t-1} \left(1 - a^{\frac{d^{i+1}-1}{d-1}}\right) \\ \overline{S_t^{(F)}} &= t - \frac{1}{Q_t^{(F)}} * \left(\sum_{i=1}^{t-1} 1 - \sum_{i=1}^{t-1} a^{\frac{d^{i+1}-1}{d-1}}\right) \\ \overline{S_t^{(F)}} &= t - \frac{t}{Q_t^{(F)}} + \sum_{i=0}^{t-1} a^{\frac{d^{i+1}-1}{d-1}} \end{aligned} \tag{9}$$

$$\overline{S_t^{(F)}} = t - \frac{t}{Q_t^{(F)}} + \frac{1}{Q_t^{(F)}} \sum_{i=1}^t a^{\frac{d^i-1}{d-1}} \tag{10}$$

We now compute the average number of messages that will be generated for the flooding technique. In flooding, the messages are flooded through the network. A 1-hop broadcast with the message will be transmitted by the node S . Once the message is received by the 1-hop neighbors, the processed messages list will be updated with the newly received message's sequence ID. This is done to avoid retransmitting the same message again in case if the message is received again through a different path. Now, if the required resource is held by a neighboring node D , an exclusive multi-hop reply will be transmitted by node D through the path resolved by routing protocol to the source S . However, even after the transmission of the resource reply message, the request message may be forwarded meaninglessly by all the other nodes that have received the request until every node in the network is reached. So, under MANETs, “ d ” messages will be generated by “ d ” neighboring nodes of the root, plus for the internal subtrees of those neighboring nodes. Thus,

$$\overline{M_t^{(F)}} = a(d + dm(t-1)) \tag{11}$$

where, $m(t-1)$ denotes the transmission along the subtree T with $(t-1)$ levels. In a subtree T , if the resource x is located at the root, a response will be sent to the querying node or else d messages will be generated by the child nodes plus the generated messages internally for each respective child subtrees T with $(t-2)$ levels.

Symbolically it means,

$$m(t-1) = F + (d + dm(t-2)) * (1 - F)$$

$$m(t-1) = 1 - a + ad + adm(t-2)$$

which has initial boundary condition of $m(0) = 1$. After evaluating the recurrence, we get,

$$m(t-1) = (ad)^{t-1} + \frac{(ad)^{t-1} - 1}{ad - 1} (ad + 1 - a) \tag{12}$$

Now substituting Eq. (12) in Eq. (11) and evaluating further we have,

$$\overline{M_t^{(F)}} = a + \frac{(c^t - 1)(2c - a)}{c - 1}, \quad c = ad, t \geq 1 \tag{13}$$

4.2 Random Walk Resource Discovery Algorithm

In the random walk algorithm under MANET, the querying node transmits the message to a specific neighboring node n that is selected in a random fashion from a list of neighbors resolved from the routing layer information over a unicast transmission (Refer Algorithm 2). For such a random walk scenario, be p randomly unfolding paths to reach the destination, i.e., the node sends the request to $p \geq 1$ of its neighboring nodes and each of those neighboring nodes will unfold a random path of which one possibility is shown in Fig. 5.

Algorithm 2: Random walk in P2P applications running over MANETs

Procedure randomwalk(m) {
 querying node resolves neighbours from routing layer
 select random node n from the neighbour list
 message m received for the first time
 unicast(m)
 }

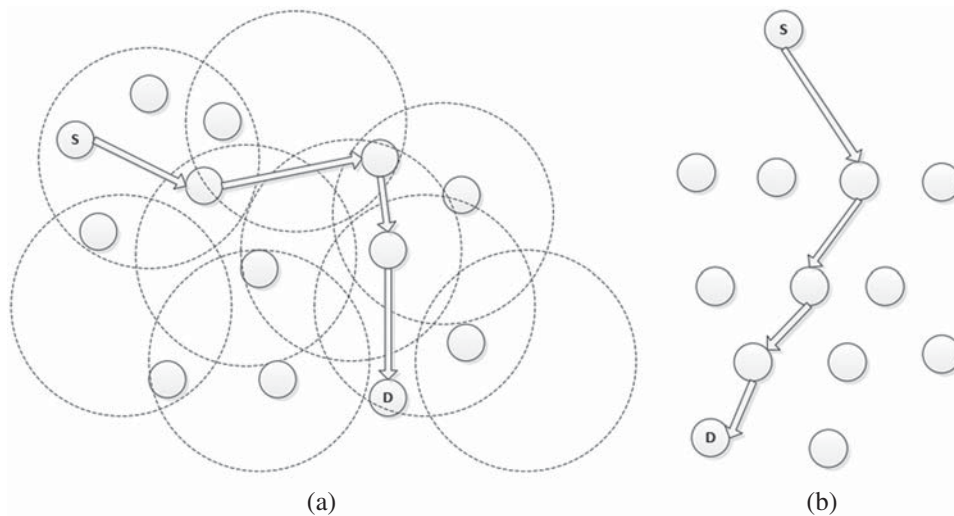


Figure 5: Resource discovery using random walk technique. (a) Searching the network using classic random walk protocol (b) d-ary tree representation

Let $Q_t^{(p)}$ be the probability of finding a resource within t steps for the random walk algorithm where p is one of the randomly unfolding paths of length $t-1$. F is the probability that a resource is known to a node and a is a probability that a resource is not known to a node. Hence, $a = 1 - F$. At “0”, step the probability of not locating x is given as,

$$1 - Q_0^{(p)} = 1 - F = a$$

$$1 - Q_t^{(p)} = \text{Probability ('x' is not occur in 't' step)}$$

$1 - Q_t^{(p)}$ = Probability ('x' is not occur in '0' step) and Probability ('x' is not occur in '1' to 't' step)

$$1 - Q_t^{(p)} = (1 - F) (1 - q_{t-1})^p \quad (14)$$

where $(1 - F)$ is the probability that a resource is not known to a given node and $(1 - q_{t-1})$ is the probability of not finding the resource in one of the "p" randomly unfolding paths of length $t - 1$.

After further evaluation we get,

$$Q_t^{(p)} = 1 - a^{pt+1} \quad (15)$$

The average number of steps needed in general to locate the resource x is given in Eq. (3). Now substituting Eq. (15) in Eq. (3) we get the average number of steps needed to find a resource within t steps for the random walk algorithm for $p \geq 1, t \geq 1$.

$$\overline{S_t^{(p)}} = t - \frac{1}{Q_t^{(p)}} \sum_{i=0}^{t-1} 1 - a^{pi+1}$$

which further is evaluated to,

$$\overline{S_t^{(p)}} = \frac{a - (1 + t - ta^p) a^{pt+1}}{(1 - a^p) (1 - a^{pt+1})} \quad (16)$$

Finally, we compute the avg. generated messages in the random walk scheme under MANET which is different from that as in the wired network. The unicast transmission uses the routing layer information in a standard random walk algorithm under MANET. But the issue is that the nodes move continuously in MANETs. Therefore, there will be a rapid change of neighbors of all nodes over time. So, such transmission may often lead to failure since the topology of the network changes rapidly. Also, this leads to frequent re-route discovery which increases the message overhead. Hence, the overall message generated will be much more than the normal for each path due to the failure thereby incurring frequent re-route discovery at every hop. In such a dynamic scenario, there will be p messages generated by d neighbors for every p of its children, plus the messages generated in each of the p paths.

$$\overline{M_t^{(p)}} = (1 - F) (dp + dpm(t - 1)) \quad (17)$$

where, $m(t - 1)$ denotes the transmissions that take place along a path p of $t - 1$ nodes.

For such a path p , the reply will be unicasted through the path resolved by the routing protocol if resource x is located at the root node. Further, the query will be terminated at that moment since the received message also contains the address S and the process will get terminated or else a message will be generated while forwarding to next node of the path plus the message generated inside the subpath p' with $t - 2$ nodes rooted at the next node. This gives the recurrence relation:

$$m(t - 1) = 1 * F + (1 + m(t - 2)) * (1 - F) = am(t - 2) + 1$$

where, $m(0) = 1$. On solving the recurrence relation, we get:

$$m(t-1) = \frac{a^t - 1}{a - 1} \quad (18)$$

Now substituting Eq. (18) in Eq. (17) it gives:

$$\overline{M}_t^{(p)} = adp + adp \left(\frac{a^t - 1}{a - 1} \right) \quad (19)$$

4.3 Addressed Broadcast Random Walk (ABRW) Resource Discovery Algorithm

To reduce the overhead in the classic random walk protocol, we proposed a cross-layered 1-hop addressed random walk resource discovery protocol [10] which is more suitable for P2P MANETs as shown in Fig. 1. Our technique is based on the addressed broadcast transmission mechanism in which the querying node will pass the request to a particular random node over a 1-hop addressed broadcast transmission, i.e., the request is forwarded to that node, while the neighboring nodes within the hop range can overhear and receive the request, but only the addressed node will process the request further and continue the resource discovery (Refer Algorithm 3). The P2P agent at the application layer will process it differently based on the message content. In short, every neighbor node will receive the message at its application layer, but inside the message, there will be the nodeID for which that message is addressed to. So, that node will only try to forward the message again. For our addressed random walk algorithm also, there will be a p randomly unfolding paths that can be followed to reach the destination. Our scheme is different from the standard random walk protocol as even the d overhearing neighbors can also see or receive the message, and hence there is a high probability of reaching the destination as the intermediate nodes can also reply if they have the resource x along the path p . And only p messages will be generated for its p children, as the addressed node will only continue with the resource discovery process further, i.e., the node transmits the message to $p \geq 1$ of its neighbors where each such neighbor unfolds a sub-tree with d neighbors of length $t - 1$ where every d neighbors can also overhear the message. Let Q_t^{dp} be the probability of locating a resource within t steps for our proposed scheme where p is one of the randomly unfolding paths and for every such p children it unfolds a sub-tree of length $t - 1$ with their d neighbors shown in Fig. 6.

Algorithm 3: Addressed broadcast random walk in P2P applications running over MANETs

Procedure addressedrandomwalk(m) {
 querying node resolves neighbours from routing layer
 select random node n from the neighbour list
 message m received for the first time
 1-hop addressedbroadcast(m)
 }

Let F be the probability that a resource is known to a node, and a is the probability that a resource is not known to a node. Hence, $a = 1 - F$. At step "0", the probability of not locating x is given as,

$$1 - Q_0^{(dp)} = 1 - F = a$$

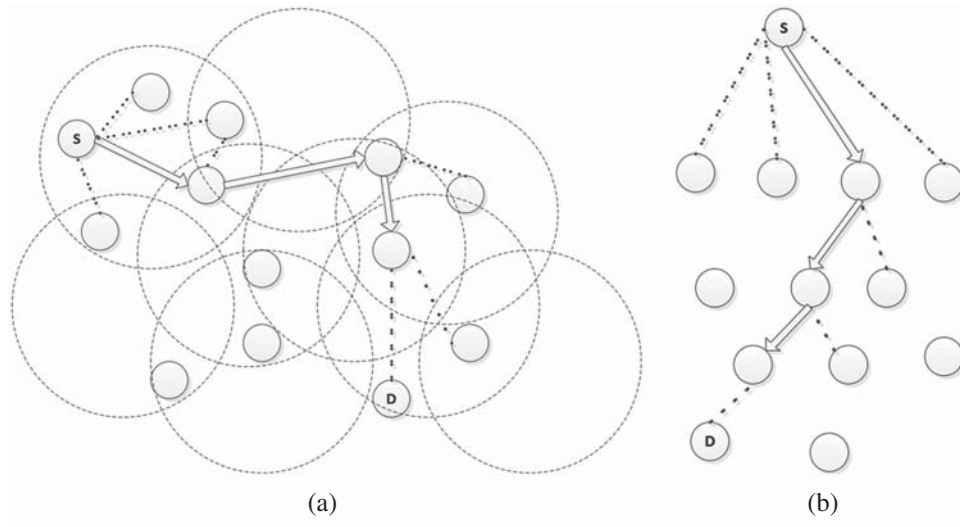


Figure 6: Resource discovery using addressed broadcast random walk technique. (a) Searching the network using addressed broadcast random walk technique (b) corresponding d-ary tree representation

Now,

$1 - Q_t^{(dp)}$ = Probability that 'x' is not found in 't' step

$1 - Q_t^{(dp)}$ = (Probability that 'x' is not found in '0' step) * (Probability that 'x' is not occur in '1' to 't' step)

$$1 - Q_t^{(dp)} = (1 - F) (1 - P_{t-1})^p \tag{20}$$

where, $1 - P_{t-1}$ is the probability of not locating the resource in one of p randomly unfolding sub-tree of length $t - 1$ with d neighbors.

$$1 - P_{t-1} = a (1 - P_{t-2})^d$$

$$P_{t-1} = 1 - a^{\frac{d^{t-1}-1}{d-1}}$$

$$1 - P_{t-1} = a^{\frac{d^{t-1}-1}{d-1}} \tag{21}$$

Now substituting Eqs. (21) in (20) we have,

$$1 - Q_t^{(dp)} = (1 - F) * a^{p \left(\frac{d^{t-1}-1}{d-1} \right)}$$

$$1 - Q_t^{(dp)} = a * a^{p \left(\frac{d^{t-1}-1}{d-1} \right)}$$

$$1 - Q_t^{(dp)} = a^{p \left(\frac{d^{t-1}-1}{d-1} \right) + 1}$$

Hence,

$$Q_t^{(dp)} = 1 - a^{p \left(\frac{d^{t-1}-1}{d-1} \right) + 1} \tag{22}$$

In our proposed random walk technique, the probability of finding x is high at each p children along the path, as even the d neighbors can respond immediately. Hence, the number of steps required to find the resource will be minimal when compared to the other discussed techniques.

The average number of steps needed to find the resource “ x ” can be found using Eq. (3), now substituting Eqs. (22) in (3) we have,

$$\overline{S}_t^{(dp)} = t - \frac{1}{Q_t^{(dp)}} \sum_{i=1}^{t-1} Q_i^{(dp)}$$

which further evaluates to,

$$\overline{S}_t^{(dp)} = t - \frac{1}{Q_t^{(dp)}} \sum_{i=0}^{t-1} \left(1 - a^p \left(\left(\frac{d^i - 1}{d - 1} \right) + 1 \right) \right) \quad (23)$$

Next, we compute the average number of generated messages for our proposed algorithm. In our resource discovery scenario, the nodes reply directly to the querying node. If x is not offered by the addressed node then there will be one fixed message generated while forwarding to every addressed child node and the message generated in each of the p paths.

$$\overline{M}_t^{(dp)} = (1 - F) (1 + pm(t - 1)) \quad (24)$$

where, $m(t - 1)$ are the transmission occurring along the path p of $t - 1$ nodes.

For such a path p , if resource x is found in its addressed node or any of its d neighbors of the broadcasting node then it will generate a reply to the querying node which is unicasted back using the path resolved by the routing protocol, and the query gets terminated only if the reply was from the addressed node or else a message will be generated while forwarding to next addressed node of the path plus the message generated inside the subpath p' with $t - 2$ nodes rooted at the next node. This gives the recurrence relation:

$$m(t - 1) = 1 * F + (1 + m(t - 2)) * (1 - F) = am(t - 2) + 1$$

where, $m(0) = 1$, since the last node receiving the message will always reply to the querying node whether it knows x or not. On solving the recurrence relation, we get:

$$m(t - 1) = \frac{a^t - 1}{a - 1} \quad (25)$$

Now substituting Eq. (25) in Eq. (24) it gives:

$$\overline{M}_t^{(dp)} = a + ap \left(\frac{a^t - 1}{a - 1} \right) \quad (26)$$

5 Theoretical Analysis & Model Validation

We compare the three search strategies analytically with the following settings. The network consists of $N = 100$ nodes, and $n_x = 1$ (as only one node holds the resource “x”). We consider for two different node degrees, i.e., sparse ($d = 4$) and dense ($d = 6$) scenarios. We evaluate with $p = 1$ and $p = 4$ paths for both the random walk protocol and our proposed method. We measure the probability of not finding the resource within “t” steps ($1 - Q_t$), the average number of steps required (\overline{S}_t), and the number of messages generated within “t” steps (\overline{M}_t) for each of the algorithms. The results are shown in Fig. 7.

The mathematical results are summarized in Tab. 6. for each of the compared search techniques with our proposed protocol.

We validate the search strategy models by comparing the theoretical analysis and simulation results. To verify the proposed theoretical analysis, we perform a series of simulations using NS-2. In the simulation, upon initialization a network area is set with N nodes, each having d random neighbors. The maximum speed of the nodes is set to 20 m/s. The resource x is assigned to a random node. The transmission range is set at 250 m, and the simulation time is 200 sec. A node randomly initiates a resource request for the resource. Pause time is set to 20 s. We verify for the varying node degrees, i.e., sparse ($d = 4$) and dense ($d = 6$) scenarios. Fig. 8 shows the comparisons between the simulation results (patterned lines) and analytical results (unpatterned lines) for the number of messages generated, the number of steps required, and the probability of resource being not found, i.e., the flag denoting the status of unsuccessful search. For each of the search schemes, an average of 30 scenarios is considered. The simulation results closely match the analytical results in most of the cases.

Table 6: Performance Measures ($t \geq 1, p \geq 1, Q_0 = 1 - a$)

Metric	Flooding	Random walk	Addressed random walk
Q_t	$1 - a^{\frac{d^{t+1}-1}{d-1}}$	$1 - a^{pt+1}$	$1 - a^{p\left(\frac{d^t-1}{d-1}\right)+1}$
\overline{S}_t	$t - \frac{t}{Q_t} + \frac{1}{Q_t} \sum_{i=1}^t a^{\frac{d^i-1}{d-1}}$	$\frac{a-(1+t-ta^p)a^{pt+1}}{(1-a^p)(1-a^{pt+1})}$	$t - \frac{1}{Q_t^{(dp)}} \sum_{i=0}^{t-1} \left(1 - a^{p\left(\frac{d^i-1}{d-1}\right)+1}\right)$
\overline{M}_t	$a + \frac{(c^t-1)(2c-a)}{c-1}, c = ad$	$adp + adp \left(\frac{a^t-1}{a-1}\right)$	$a + ap \left(\frac{a^t-1}{a-1}\right)$

As a part of our future work, it would be interesting to investigate adapting the P2P communication design for mobile intelligence of things as an alternative to internet access. Security is very critical in such intelligence of things architectures [19–22]. As of now, we have not focused on the security attacks scenario within the discovery process in general. As a part of our continued research effort, we will certainly investigate this dimension as part of our future works.

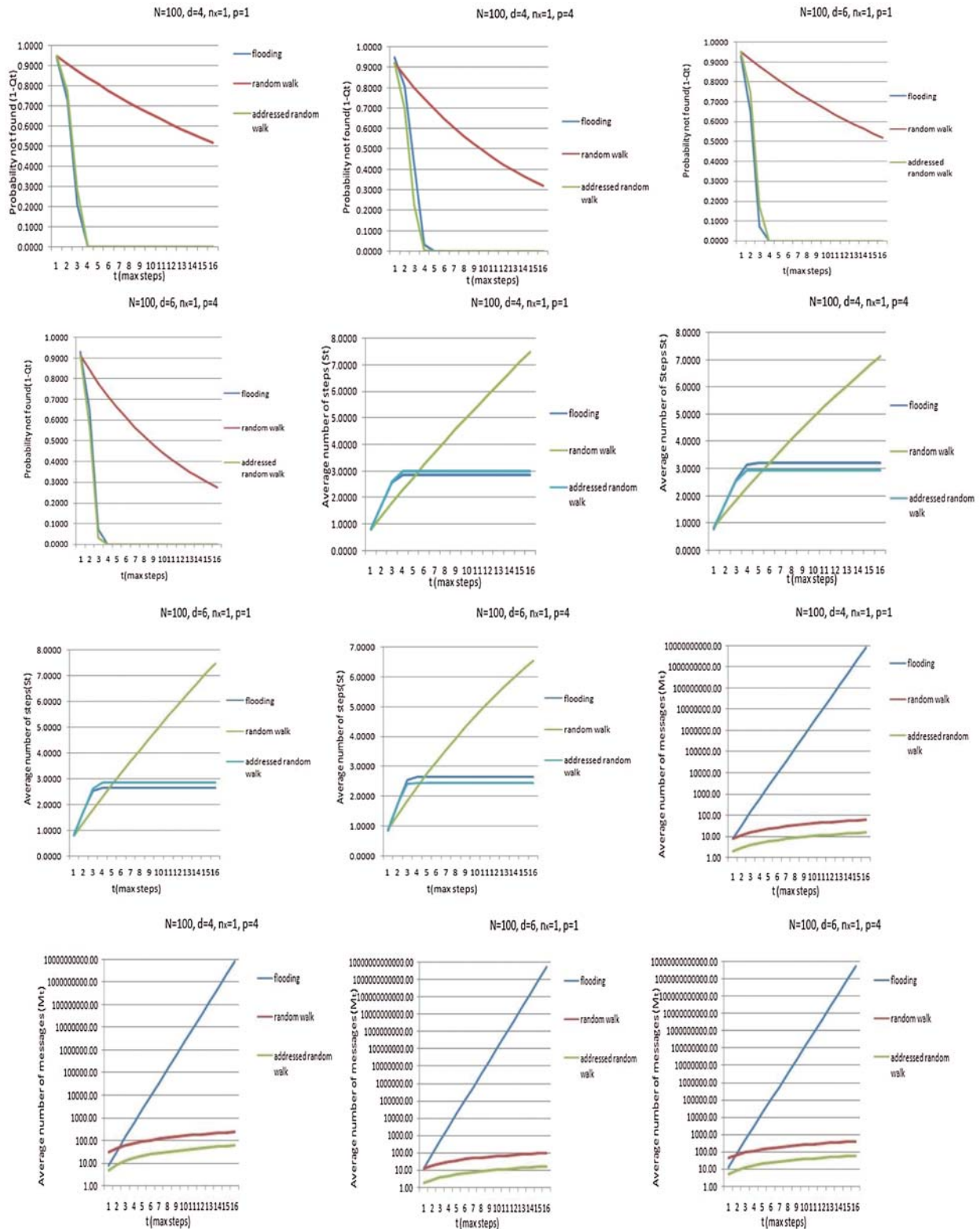


Figure 7: Comparisons of the search techniques using the metrics: probability of not finding the resource, average number of steps to locate the resource, and average number of messages generated

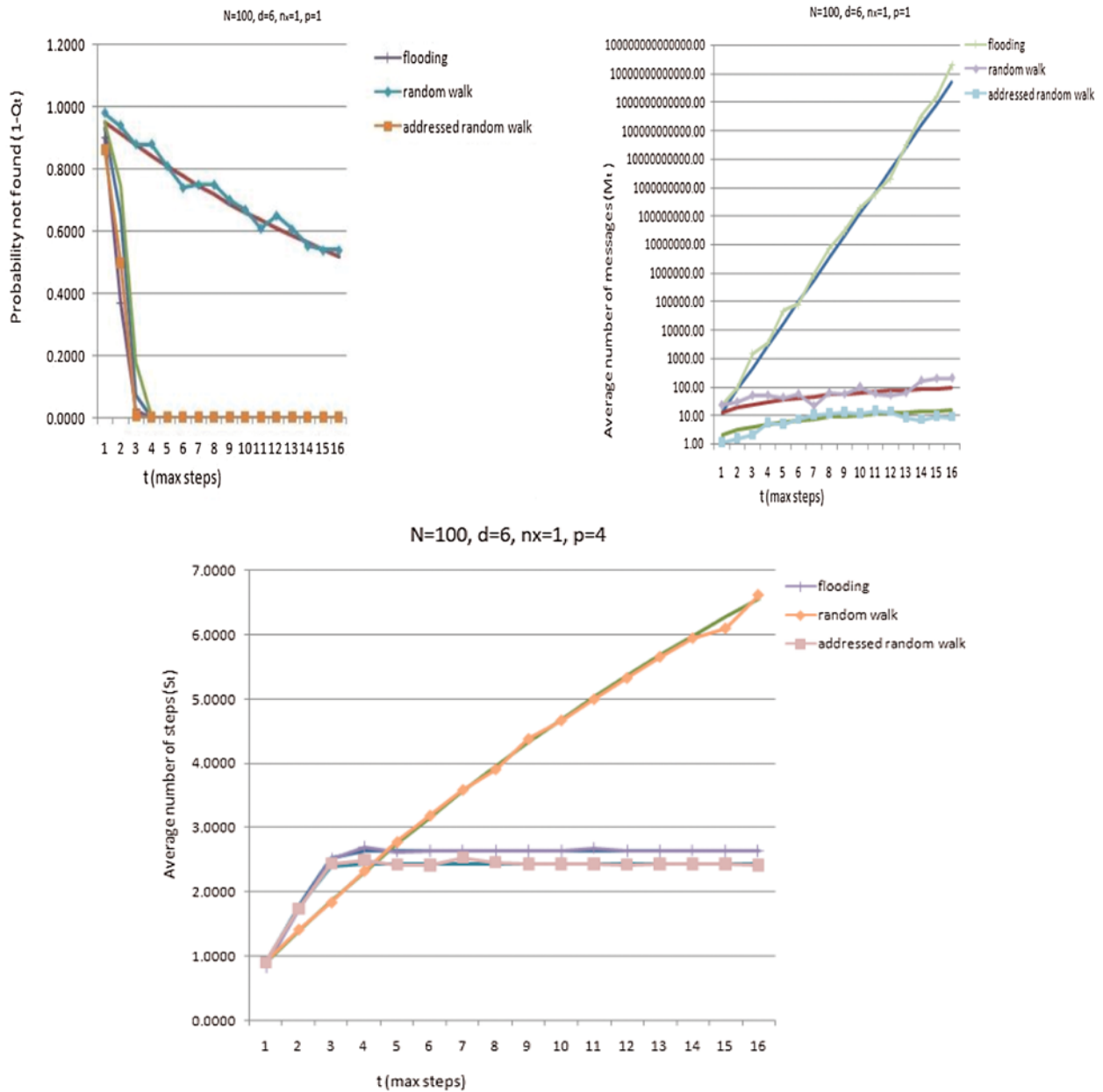


Figure 8: Simulation (dotted lines) and analytical (plain lines)

6 Conclusion

The contributions of this article are as follows:- 1) Theoretical modeling of the popular search algorithms, and deriving key analytical measures for evaluating their performances are done. 2) Mathematical modeling of our proposed addressed broadcast resource discovery protocol is done, focused towards optimizing random walk algorithm over MANETs. Such, evaluation in context to mobile Peer-to-Peer networks is not done before. 3) Simulation based validation of the

analytical/theoretical estimations is done, for justifying the theoretical results. This paper largely focuses on the general issues of resource discovery under a highly dynamic mobile P2P network. Specifically, the performance of well-known and widely used benchmark techniques, i.e., random walk, and flooding algorithms are studied. Flooding scheme tries to locate the resource in an aggressive manner by visiting almost all the nodes and thereby suffers from scalability issues due to the generation of an enormous amount of query messages. On the contrary, the random walk scheme searches conservatively, but it still generates a huge amount of query messages at each hop, and consumes more power under MANETs, while further increasing the search latency. To alleviate these problems, we propose an improved random walk scheme suitable for P2P MANETs, i.e., an addressed random walk protocol which is a hybrid cross-layered strategy of flooding and random walk techniques aimed to suit well for such highly dynamic self-evolving networks. We present an analytical model to estimate the performance of each strategy based on the metrics that include the probability of finding a resource, the mean no. of steps needed to locate a resource, and the mean no. of messages generated during the resource discovery process. We further validate the mathematical models through simulations. The results justify our findings, as our proposed scheme reduces the search latency, decreases the message overhead, and equally has a good hit rate. To further verify the robustness of our proposed technique, as future work, we plan to test the algorithm under different mobility models especially the Reference Point Group Mobility (RPGM) with sparse & dense network scenarios, to reaffirm the efficacy.

Acknowledgement: We are thankful to all the collaborating partners.

Funding Statement: Taif University Researchers Supporting Project number (TURSP-2020/36), Taif University, Taif, Arabia Saudi.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

- [1] M. A. Noor, R. Beraldi and R. Baldoni, "Identifying open problems in random walk based service discovery in mobile," in *Proc. IICS*, Bonn, NW, Germany, pp. 91–102, 2010.
- [2] A. Helmy, "Efficient resource discovery in wireless adhoc networks: Contacts do help," in *Resource Management in Wireless Networking Series: Network Theory and Applications*, 1st ed., vol. 16. Boston, MA, USA: Springer, pp. 419–471, 2005.
- [3] R. Schollmeier, I. Gruber and F. Niethammer, "Protocol for peer-to-peer networking in mobile environments," in *Proc. ICCN*, Dallas, TX, USA, pp. 121–127, 2003.
- [4] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham and S. Shenker, "Making gnutella-like p2p systems scalable," in *Proc. ATAPCC*, Karlsruhe, BW, Germany, pp. 407–418, 2003.
- [5] A. Arunachalam and O. Sornil, "Minimizing redundant messages and improving search efficiency under highly dynamic mobile p2p network," *Journal of Engineering Science & Technology Review*, vol. 9, no. 1, pp. 23–35, 2016.
- [6] A. Arunachalam and O. Sornil, "A broadcast based random query gossip algorithm for resource search in non-DHT mobile peer-to-peer networks," *Journal of Computers*, vol. 28, no. 1, pp. 209–223, 2017.
- [7] A. Arunachalam, "Rock, paper, scissors game based model for content discovery in P2P MANETs," *Wireless Personal Communications*, vol. 113, no. 1, pp. 1–16, 2020.
- [8] A. Arunachalam and O. Sornil, "An analysis of the overhead and energy consumption in flooding, random walk and gossip based resource discovery protocols in MP2P networks," in *Proc. ICACC*, Haryana, HR, India, pp. 292–297, 2015.

- [9] A. Arunachalam and O. Sornil, "Issues of implementing random walk and gossip based resource discovery protocols in P2P MANETs & suggestions for improvement," in *Proc. ICRTC*, Delhi, India, pp. 509–518, 2015.
- [10] A. Arunachalam and O. Sornil, "Reducing routing overhead in random walk protocol under MP2P network," *International Journal of Electrical & Computer Engineering*, vol. 6, no. 6, pp. 3121–3130, 2016.
- [11] H. Barjini, M. Othman, H. Ibrahim and I. N. Udzir, "Shortcoming, problems and analytical comparison for flooding-based search techniques in unstructured P2P networks," *Peer-to-Peer Networking and Applications*, vol. 5, no. 1, pp. 1–13, 2012.
- [12] N. Bisnik and A. A. Abouzeid, "Optimizing random walk search algorithms in P2P networks," *Computer Networks*, vol. 51, no. 6, pp. 1499–1514, 2007.
- [13] V. Dimakopoulos and E. Pitoura, "On the performance of flooding-based resource discovery," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 11, pp. 1242–1252, 2006.
- [14] T. Lin, P. Lin, H. Wang and C. Chen, "Dynamic search algorithm in unstructured peer-to-peer networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 5, pp. 654–666, 2008.
- [15] M. E. Newman, S. H. Strogatz and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review*, vol. 64, no. 2, pp. 1–17, 2001.
- [16] M. Portmann and A. Seneviratne, "Cost-effective broadcast for fully decentralized Peer-to-Peer networks," *Computer Communications*, vol. 26, no. 11, pp. 1159–1167, 2003.
- [17] S. Ferretti, "Searching in unstructured overlays using local knowledge and gossip," in *Proc. Complex Networks*, V, Bologna BO, Italy, pp. 63–74, 2014.
- [18] H. Babaei, M. Fathy and M. Romoozi, "Modeling and optimizing random walk content discovery protocol over mobile ad-hoc networks," *Performance Evaluation*, vol. 74, no. 2, pp. 18–29, 2014.
- [19] M. Krichen, O. Cheikhrouhou, M. Lahami, R. Alroobaea and A. Jmal, "Towards a model-based testing framework for the security of internet of things for smart city applications," in *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. vol. 224. Cham: Springer, 2018.
- [20] M. Krichen and R. Alroobaea, "A new model-based framework for testing security of IOT systems in smart cities using attack trees and price timed automata," in *ENASE 2019: Proc. of the 14th Int. Conf. on Evaluation of Novel Approaches to Software Engineering*, Heraklion, Crete, Greece, pp. 570–577, 2019.
- [21] M. Krichen and R. Alroobaea, "Towards optimizing the placement of security testing components for internet of things architectures," in *IEEE/ACS 16th Int. Conf. on Computer Systems and Applications*, Abu Dhabi, UAE, pp. 1–2, 2019.
- [22] M. Krichen, M. Lahami, O. Cheikhrouhou, R. Alroobaea and A. Jmal, "Security testing of internet of things for smart city applications: A formal approach," in *Smart Infrastructure and Applications. EAI/Springer Innovations in Communication and Computing*. Cham: Springer, 2020.